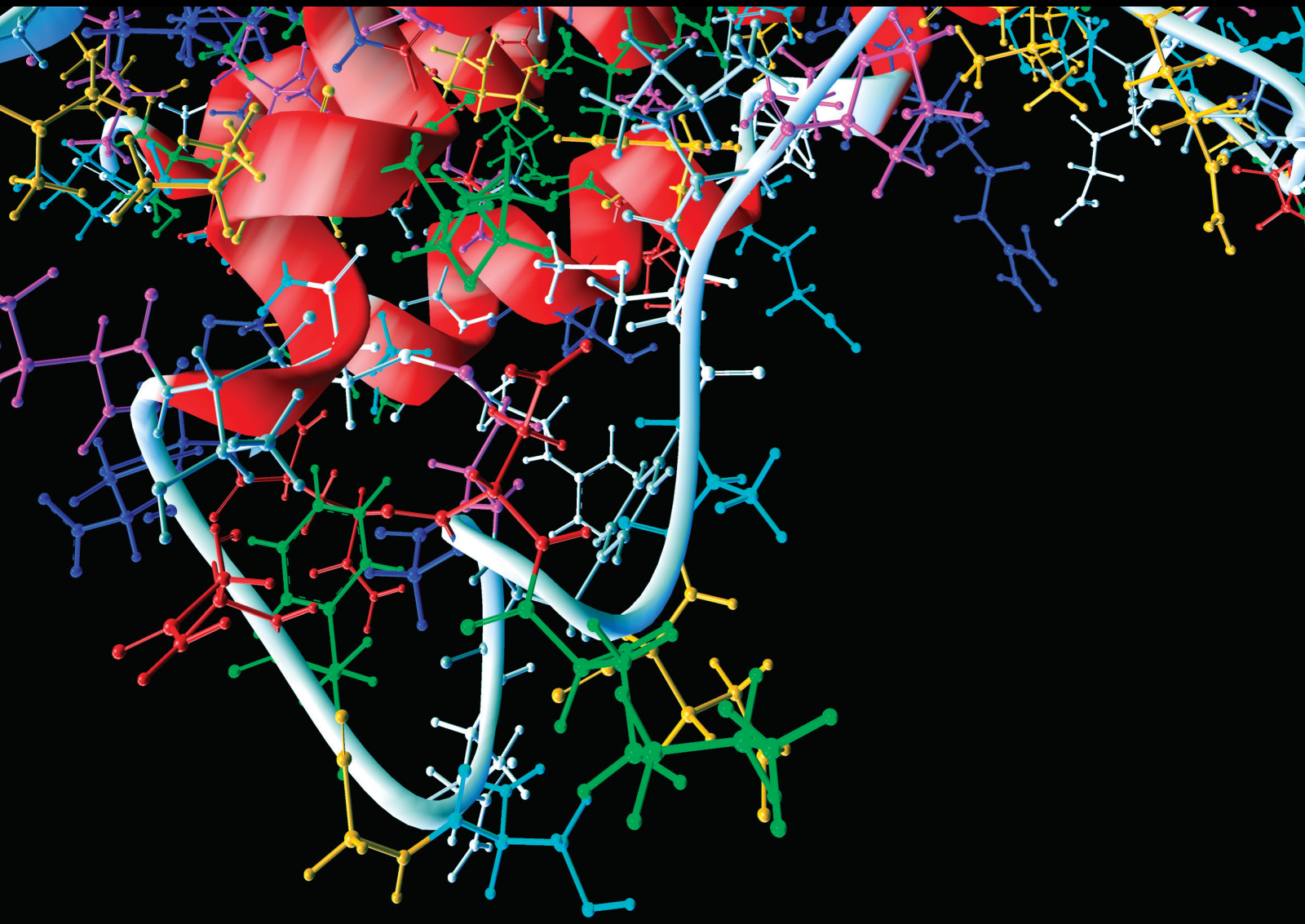
Computational and Mathematical Methods in Medicine

Machine Learning and Computational Modelling for Clinical Decision Making

Lead Guest Editor: Mario Cesarelli
Guest Editors: Giovanni D'Addio, Paolo Gargiulo, Alberto Cuocolo, and Marianna Amboni



Machine Learning and Computational Modelling for Clinical Decision Making

Computational and Mathematical Methods in Medicine

Machine Learning and Computational Modelling for Clinical Decision Making

Lead Guest Editor: Mario Cesarelli
Guest Editors: Giovanni D'Addio, Paolo Gargiulo, Alberto Cuocolo, and Marianna Amboni

Contents

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 9452157, 9 pages
https://doi.org/10.1155/2022/9452157

Research Article

SC-Dynamic R-CNN: A Self-Calibrated Dynamic R-CNN Model for Lung Cancer Lesion Detection

Xun Wang,[1] **Lisheng Wang**,[1] **and Pan Zheng**[2]

[1]*China University of Petroleum, China*
[2]*University of Canterbury, New Zealand*

Correspondence should be addressed to Pan Zheng; pan.zheng@canterbury.ac.nz

Received 6 September 2021; Revised 24 February 2022; Accepted 6 March 2022; Published 28 March 2022

Academic Editor: Huiling Chen

Lung cancer has complex biological characteristics and a high degree of malignancy. It has always been the number one "killer" in cancer, threatening human life and health. The diagnosis and early treatment of lung cancer still require improvement and further development. With high morbidity and mortality, there is an urgent need for an accurate diagnosis method. However, the existing computer-aided detection system has a complicated process and low detection accuracy. To solve this problem, this paper proposed a two-stage detection method based on the dynamic region-based convolutional neural network (Dynamic R-CNN). We divide lung cancer into squamous cell carcinoma, adenocarcinoma, and small cell carcinoma. By adding the self-calibrated convolution module into the feature network, we extracted more abundant lung cancer features and proposed a new regression loss function to further improve the detection performance of lung cancer. After experimental verification, the mAP (mean average precision) of the model can reach 88.1% on the lung cancer dataset and it performed particularly well with a high IoU (intersection over union) threshold. This method has a good performance in the detection of lung cancer and can improve the efficiency of doctors' diagnoses. It can avoid false detection and miss detection to a certain extent.

1. Introduction

Cancer is the second leading cause of human death in the world, and its mortality and morbidity are increasing year by year. According to the data of the World Health Organization (WHO), cancer has led to 9.6 million deaths in 2018 and lung cancer ranks first, with 1.76 million deaths [1]. Compared with other cancers, the biological characteristics of lung cancer are very complex and it has a short onset time and high malignancy, which makes lung cancer still the number one "killer" of cancer [2, 3]. The main reason for the high morbidity and mortality is that the diagnosis and treatment methods of lung cancer are still at an early stage, so it is urgent to refine and improve the diagnosis methods of lung cancer.

At present, histopathological examination is the standard for pathological diagnosis of tumors, which can only be performed on tissue specimens such as surgical resection or needle biopsy. However, the tissue specimens obtained are invasive and susceptible to specimen sampling. To assist diagnostic doctors in their work and improve the efficiency of cancer diagnosis, the computed tomography (CT) [4] has been widely used in the intelligent diagnosis of medical images, becoming a powerful tool to comprehensively capture the characteristics of cancer. Computer-aided detection systems are mostly machine learning algorithms such as support vector machines, which are usually used to detect and classify tumors [5, 6]. However, they are usually limited by the assumptions made during the definition of elements and still have drawbacks such as a complex process, parameter setting based on experience, and strong dependence. For example, lung cancer detection results depend on the quality of segmentation results and the effectiveness of extracted features.

In recent years, artificial neural networks, especially deep neural networks, have made remarkable achievements in many fields of intelligent medicine [7–9]. This learning algorithm is driven by big data, excavates rules from a large amount of data, and then classifies and judges unknown phenomena [10–16]. The continuous accumulation of medical data provides powerful materials and tools for intelligent

screening and diagnosis of cancer. Zhang et al. [17] used a convolutional neural network to extract deep features and combine them with shallow features to achieve the classification of ovarian cancer. In addition, Wu et al. [18] used the deep convolutional neural network based on AlexNet to realize the classification of ovarian cancer pathological images and the accuracy rate of the model achieved 78.2%. Tajbakhsh and Suzuki [19] used an artificial neural network and convolutional neural network to test the benign and malignant classification of pulmonary nodules in CT images, and the experiment found that the performance of the convolutional neural network was better than the other types of artificial neural network in the lung lesion and tumor classification task.

With the development of the field of intelligent medical treatment, the types of diseases are increasing and the complexity of the pathological relationship between diseases is also increasing, so the requirements of a deep neural network are more and more strict. At present, mainstream object detection algorithms in deep learning are mainly based on two types: the first is a one-stage detection algorithm, which includes Yolo [20] and RetinaNet [21]; the performances of those methods are fast yet not accurate. As a representative of the one-stage algorithm, the Yolo series runs fast. It divides an image into multiple cells of the same size, predicts the category of each cell, and gives the category confidence of the bounding box. The other is a two-stage object detection algorithm, such as Fast R-CNN [22], Faster R-CNN [23], and Mask R-CNN [24]. The first stage of this algorithm takes the CT image as the input and generates the region of interest through the algorithm. The second stage is to use the output of the first stage to further classify and regress the bounding box. Although the detection accuracy of the two-stage object detection algorithm is better than the one-stage object detection algorithm, high-quality samples contribute significantly less to the network during the training process. Zhao et al. [25, 26] proposed a Cascade R-CNN network based on Faster R-CNN to solve the problem that high-quality samples contribute less to training in object detection. Through the Cascaded R-CNN network, each R-CNN network is set with different IoU thresholds. In this way, the accuracy of each network output has been improved to a certain extent and the output of the previous R-CNN network can be used as the input of the next high-precision network. Finally, the accuracy of the network will gradually improve. In addition, in order to solve the imbalance of object detection in the training process, Pang et al. [27] proposed a Libra R-CNN network, which paid attention to the problems of the sample layer, feature layer, and target layer, and balanced the imbalance through the overall balanced design. Zhang et al. [28] drew lessons from the idea of Cascade R-CNN and proposed Dynamic R-CNN, which further solved the problem of inconsistencies between training processes.

In addition to the network's architecture, the quality of feature map extraction also greatly affects the accuracy of object detection. In most computer vision tasks, it is helpful to establish a long-distance dependency mechanism for feature map extraction. One way to model remote dependencies is to use a spatial pool or convolution operator with a large kernel window. Some typical examples, such as PSPNet [29], employ multiple spatial pool operators of different sizes to capture multiscale contexts. There is a lot of work [30–32] using a large convolution kernel or extended convolution for long-term context aggregation. By introducing an adaptive response calibration operation, SCNet [33] constructs multiscale feature representation in the building block and greatly improves the prediction accuracy.

In this study, the histologic types of lung cancers that we are looking at are adenocarcinoma, squamous cell carcinoma, and small cell carcinoma. The first two types are the major types of lung cancer of non-small cell lung cancer (NSCLC) which takes 85% to 90% of all lung cancer cases. Small cell carcinoma constitutes 10% to 15% of lung cancers [34]. The percentage of different lung cancer types objectively causes the imbalance of the image data collected. Some data preprocessing procedure is conducted to resolve its impact on our SC-Dynamic R-CNN development. The types of lung cancers studied in this research bear high-level significance and real-life value in medical practices.

To improve the detection accuracy of lung cancer, a new lung cancer detection algorithm based on Dynamic R-CNN [28] is proposed in this paper. We divide the collected datasets into three categories: adenocarcinoma, squamous cell carcinoma, and small cell carcinoma, and amplified the data of squamous cell carcinoma and small cell carcinoma by an oversampling method. Next, we implement the SCNet [33] module into the Dynamic R-CNN network, which can fully extract lesion features. In addition, we propose a new loss function, DBS L1 loss, which further improves the contribution of high-quality samples to training. After experimental verification, we found that our algorithm has a great improvement in the detection of lung cancer compared with other advanced algorithms.

2. Materials and Methods

2.1. Materials. This paper's dataset was taken from the Shandong Provincial Hospital and Shandong Provincial Third Hospital in Shandong, China. The datasets include 34056 pathological images on 261 patients, and the lesion location was marked by professional radiologists. According to the radiologist's annotation, we selected 3442 images of lung cancer with lesions.

The data selected are firstly divided into three categories, namely, adenocarcinoma, squamous cell carcinoma, and small cell carcinoma. In this paper, we use "Adenocarcinoma," "Squamous carcinoma," and "small cell carcinoma" to represent these three categories. Among the pathological types of lung cancer, adenocarcinoma is the most common and there is little data on other types of cancer, which leads to the imbalance towards the number of samples of different types of lung cancer. The dataset of lung cancer is distributed as follows:

Figure 1 shows that there are 2273 samples of adenocarcinoma, 845 samples of squamous carcinoma, and 324 samples of small cell carcinoma. To more objectively train the

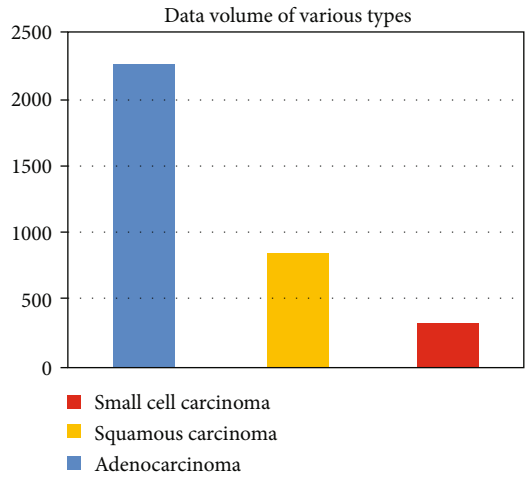


Figure 1: The distribution of lung cancer CT image data of different types.

method, we would like to have the datasets of different cancer types in a similar size; hence, for the small size of cancer-type datasets, we expanded the size of the dataset by oversampling methods. It is noticeable that the number of adenocarcinoma data samples is about three times of squamous carcinoma and eight times of small cell carcinoma, and therefore, the latter two minority class datasets were oversampled 3 times and 8 times of their original size to match the majority class, i.e., adenocarcinoma.

Different from conventional oversampling approaches, e.g., random oversampling and synthetic minority oversampling technique (SMOT), for image data, we can synthesize samples using image processing techniques, e.g., spatial transformation including flipping, shearing, and rotating [35], gamma transformation, histogram equalization, and other methods to enhance the dataset [36]. An example of an image enhancement result is shown in Figure 2.

2.2. Methods. We present the next new method for robust lung cancer lesion detection in CT studies that uses Dynamic R-CNN trained on our dataset. To achieve accurate detection of lung cancer lesions, we use Dynamic R-CNN as the baseline network and use the self-calibrated convolutions to replace the traditional convolution. Besides that, we proposed a new regression loss function which is better than the loss function in Dynamic R-CNN.

We first present an overview of the method and then describe in detail its components. To make the paper self-contained, we describe all steps of the extended method.

2.2.1. Model. Figure 3 shows the flow diagram of our method. The structure of the SC-Dynamic R-CNN network is similar to Faster R-CNN [23]. It is composed of two modules. The first module is a deep fully convolutional network that proposes regions, which is called the region proposal network (RPN) module. The RPN module is aimed at detecting multiple objects in a single image. The second module is the detector that uses the proposed regions, namely, Box_Head. After the Box_Head, there are two loss functions: classification loss function and regression loss function. But unlike Faster R-CNN [23], SC-Dynamic R-CNN can adjust the label assignment criteria and the shape of regression loss function automatically during training that makes better use of the training samples. In order to enhance the ability of feature representation of lung cancer, SC-Dynamic R-CNN adds SCNet [33] to the RPN module. Except that, the loss function of Dynamic R-CNN has been optimized for getting a better detection result of lung cancer.

As shown in Figure 3, initially, the lung cancer images are resized to 512×512 pixels for the training phase. The resize images are subsequently fed to the region proposal network (RPN) to get the proposed region. Next, the proposed regions are classified and regressed by the Box_Head module. Eventually, the classification and regression results are fed into the corresponding loss function and as the parameter update of the network. We use softmax loss as the classification loss, and regression loss uses our newly proposed loss function, the details of which will be described in the next section.

To better exploit the dynamic property in the training stage, SC-Dynamic R-CNN uses a lower IoU threshold to better accommodate these imperfect proposals in the second-stage training (Figure 3(a)). As the training goes, the quality of proposals is continuously improved. Therefore, we can increase the threshold to better use them to train a high-quality detector, so the network can be more discriminative at higher IoU. Dynamic label assignment can be formulated as follows:

$$\text{Label} = \begin{cases} 1, & \text{if } \max \text{ IoU}(b, G) \geq T_{\text{now}}, \\ 0, & \text{if } \max \text{ IoU}(b, G) < T_{\text{now}}, \end{cases} \tag{1}$$

where T_{now} stands for the current IoU threshold. In order to realize the dynamic property that the distribution of proposals changes over time during the training process, the dynamic label assignment will automatically update based on the proposal's statistics. Specifically, SC-Dynamic R-CNN first calculates the IoUs I between the proposals and its target ground truth and then selects the maximum value of K_I from I as the threshold T_{now}. As the training goes, the IoUs I between the proposal and its target ground truths will increase gradually and so does the updated threshold T_{now}.

In addition, according to the conclusion of Dynamic R-CNN [28], with the improvement of IoU threshold, the quality of positive samples will be further improved. As a result, the contribution of high-quality samples will be further decreased, which will greatly limit the overall performance. Based on the method of Dynamic R-CNN, we have improved its regression loss function and obtained more accurate results which are described in the next section.

2.2.2. DBS L1 Loss. According to the conclusion of Dynamic R-CNN [28], with the improvement of the sample quality, its contribution will gradually decrease. As a result, Dynamic R-CNN adds a factor α based on the Smooth L1 loss function. The network adjusts the loss function by adjusting the value of the factor α. With the increase of factor α, the

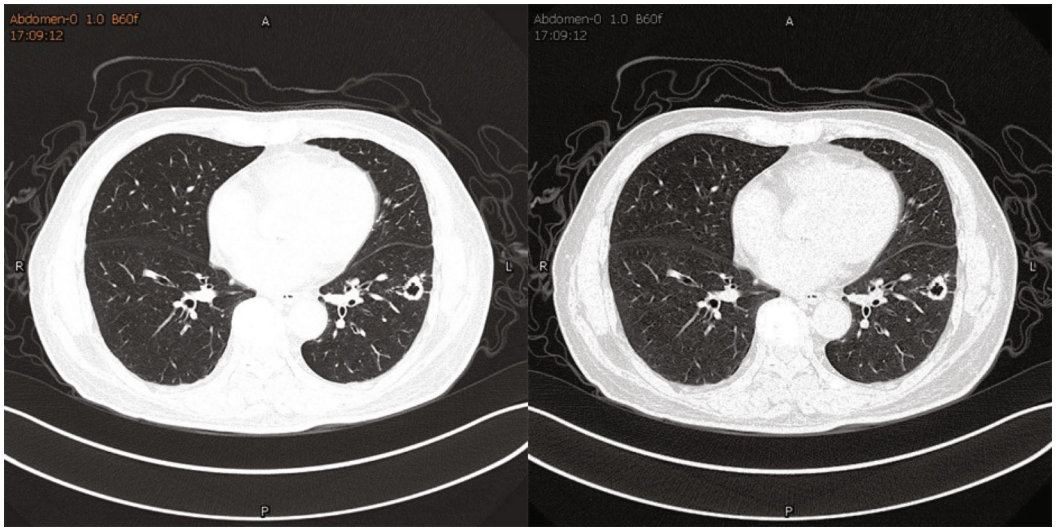


Figure 2: The transaxial view of the enhanced lung cancer image data.

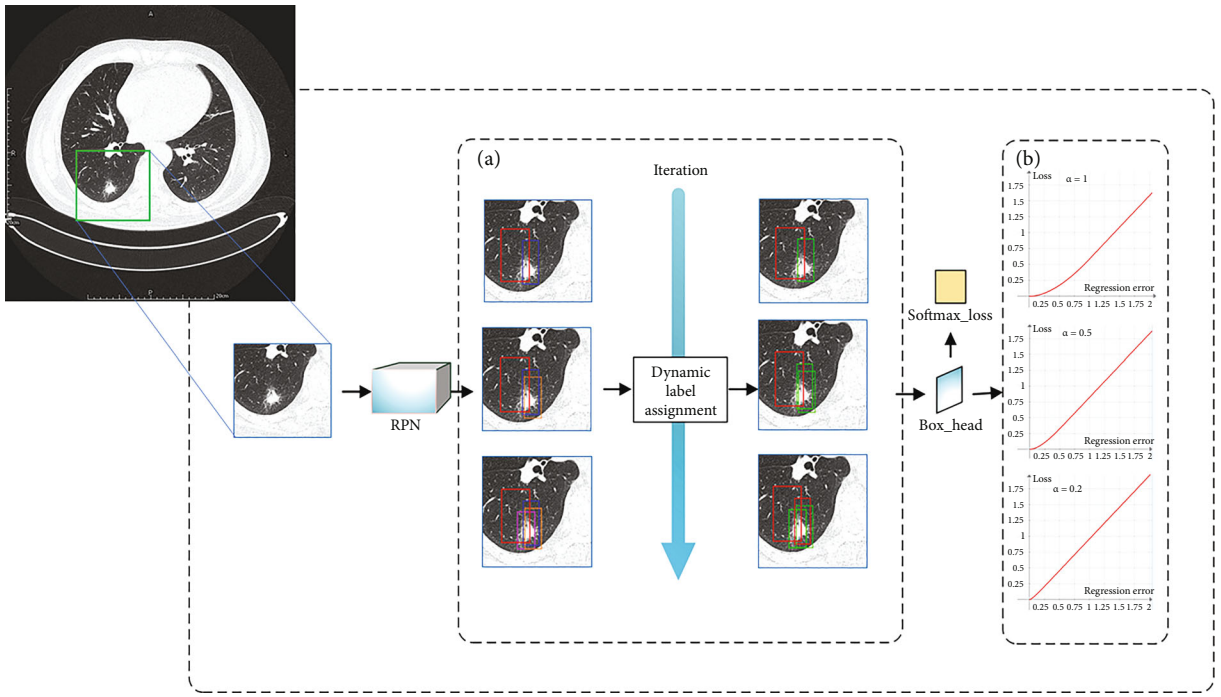


Figure 3: The overall structure of the proposed SC-Dynamic R-CNN.

gradient of high-quality sample training will increase gradually, so the contribution to the network will be increased. The regression loss function of Dynamic R-CNN is shown as follows:

$$\mathrm{DSL}(x, \alpha_{\mathrm{now}}) = \begin{cases} \dfrac{0.5|x|^2}{\alpha_{\mathrm{now}}}, & \text{if } |x| < \alpha_{\mathrm{now}}, \\ |x| - 0.5\alpha_{\mathrm{now}}, & \text{otherwise}, \end{cases} \tag{2}$$

where the α_{now} will decrease with the training, as shown in Figure 2.

But the loss function can be further improved. Taking Libra R-CNN [27] as a reference, we improve the Dynamic R-CNN loss function and further improve the contribution of high-quality samples to training. The improved DBS L1 loss can be formulated as follows:

$$\mathrm{DBSL}(x, \alpha_{\mathrm{now}}) = \begin{cases} \dfrac{\alpha_{\mathrm{now}}}{b}(b|x| + 1)\ln(b|x| + 1) - \alpha_{\mathrm{now}}|x|, & \text{if } |x| < \alpha_{\mathrm{now}}, \\ |x| + C, & \text{otherwise}. \end{cases} \tag{3}$$

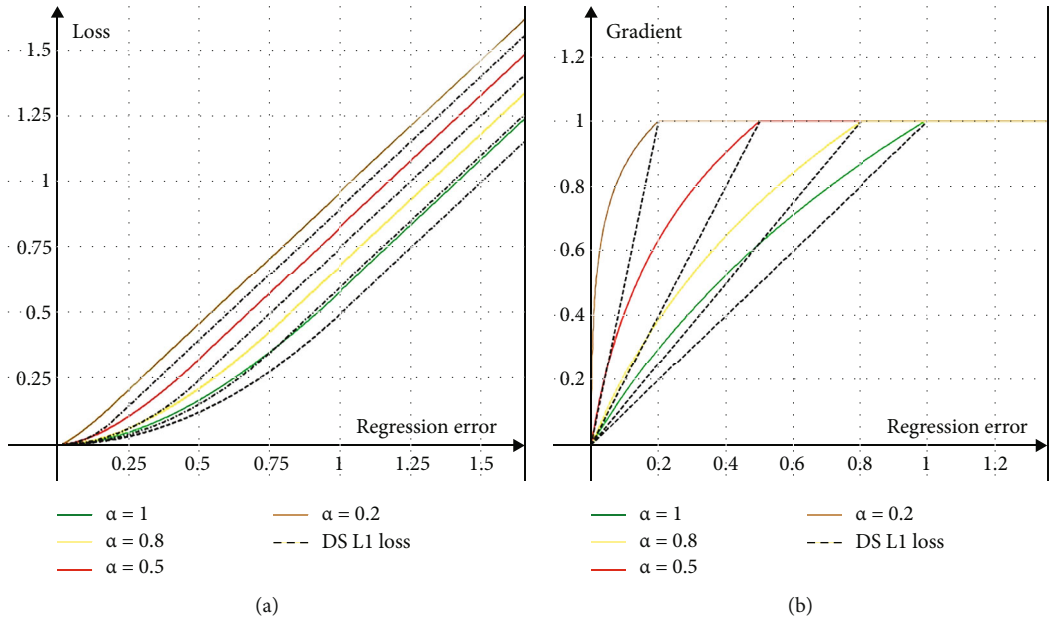


Figure 4: The curves for (a) loss and (b) gradient of our regression loss with different α. α is set to default as 1.0.

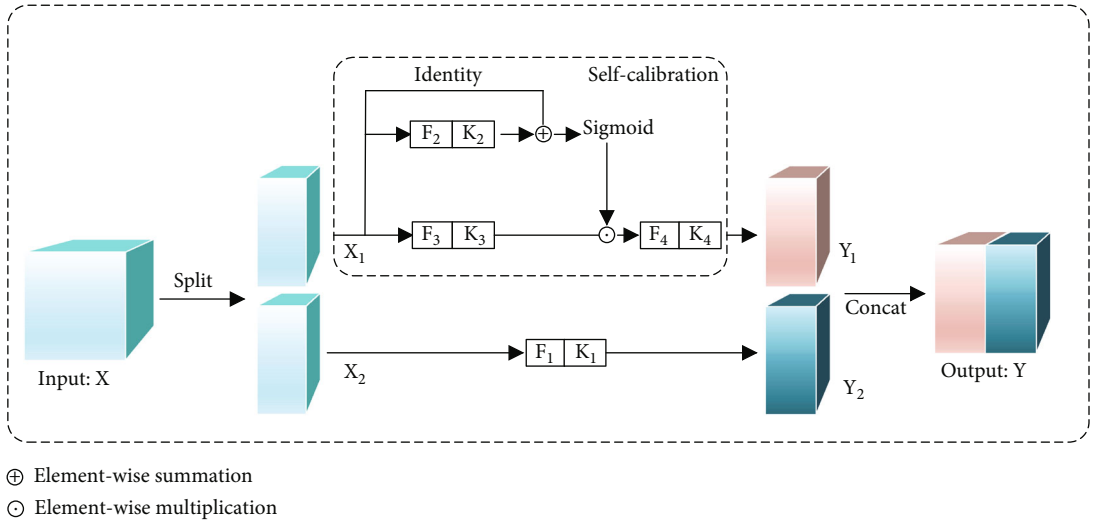


Figure 5: The overall structure of SCNet.

where b and C are constants and their values are constrained by the factor α.

Similar to the dynamic label assignment process in Dynamic R-CNN [28], DBS L1 loss first obtains the regression label E between proposals and their target ground truths. Then, we select the K_α minimum value from E to update the factor α in the equation.

As shown in Figure 4, with the continuous reduction of factors in DBS L1 loss, the contribution of high-quality samples to training increases gradually. Clearly, the DBS L1 loss is superior to DS L1 loss, which greatly improves the recognition accuracy of lung cancer lesions.

2.2.3. Self-Calibration. Conventional 2D convolution is still used to calculate the convolution in Dynamic R-CNN [28]. But in conventional 2D convolution, each output feature map is generated by the same formula, which results in the convolutional filters learning similar patterns. In addition, the fields of view for each spatial location in the convolution feature transformation can only be controlled by the size of the predefined convolution kernel. As a result, the discrimination of the lung cancer feature map will be decreased. In order to enhance the ability of feature representation of lung cancer lesions and identify lung cancer lesions more accurately, SCNet [33] is used in SC-Dynamic R-CNN instead of traditional 2D convolution.

As shown in Figure 5, the shape of the given group of the filter is (C, C, k_h, k_w), where C is the number of channels and k_h and k_h are the spatial height and width, respectively. SCNet first separates it into four portions, each of which is responsible for different functionality. The separated filter is expressed by $\{K_i\}_{i=1}^{4}$, and the size of each filter is

Table 1: Comparisons with different models on our lung cancer dataset.

Method	Backbone	Adenocarcinoma		Squamous		Small cell		mAP
		AP_{50}	AP_{75}	AP_{50}	AP_{75}	AP_{50}	AP_{75}	
RetiNanet [21]	ResNet-50	87.7%	67.8%	89.7%	79.9%	88.1%	77.8%	81.8%
SSD [38]	ResNet-50	80.7%	61.4%	89.2%	78.4%	86.2%	77.3%	78.9%
Faster R-CNN [23]	ResNet-50	81.6%	62.5%	90.5%	80.3%	89.7%	79.4%	80.1%
Libra R-CNN [27]	ResNet-50	81.9%	71.4%	89.9%	81.5%	89.3%	83.2%	82.9%
Cascade R-CNN [25]	ResNet-50	82.7%	73.5%	90.1%	82.9%	90.1%	84.9%	84.0%
SC-Dynamic R-CNN	ResNet-50	**91.6%**	**77.3%**	**91.5%**	**88.2%**	**91.4%**	**88.6%**	**88.1%**

Table 2: Results of each component in SC-Dynamic R-CNN on val set.

Backbone	FPN	DBS L1 loss	SCNet	AP50	AP75	mAP
ResNet-50	√			90.1%	80.7%	85.4%
ResNet-50	√	√		90.6%	83.6%	87.1%
ResNet-50	√	√	√	91.5%	84.7%	88.1%

$(C/2, C/2, k_h, k_w)$. The input X will be divided into two parts before entering the self-calibrated convolutional network, which represents by X_1 and X_2, where X_1 will conduct self-calibration through $\{K_2, K_3, K_4\}$ to produce Y_1. At the same time, X_2 will be manipulated by K_1 and produce Y_2. Finally, Y_2 will be connected to Y_1 to generate the final output Y.

In order to collect the context information of each spatial location effectively, SCNet conducts convolution feature transformation in two different scale spaces. Firstly, input X_1 will be performed with average pooling operation:

$$T_1 = \text{AvgPool}(X_1). \tag{4}$$

Then, the obtained T_1 maps the intermediate references from the small-scale space to the original feature space by a bilinear interpolation operator. The specific formula is as follows:

$$X_1' = \text{Up}(F_2(T_1)) = \text{Up}(T_1 * K_2), \tag{5}$$

where "$*$" denotes convolution and Up $(\cdot)$ is a bilinear interpolation operator. The calibration operation can be formulated as follows:

$$Y_1' = F_3(X_1) \cdot \sigma\left(X_1 + X_1'\right), \tag{6}$$

where $F_3(X_1) = X_1 * K_3$, "$\cdot$" denotes element-wise multiplication, and σ is the sigmoid function. After the calibration operation, Y'_1 needs to be operated by the following formula to get the final output:

$$Y_1 = F_4\left(Y_1'\right) = Y_1' * K_4. \tag{7}$$

In our model, SCNet is used to replace the convolutional 2D convolution, which considers the context around each spatial location, avoids the information irrelevant to the lesion partly, and also improves the recognition accuracy of lung cancer lesions.

3. Experiments

3.1. Evaluation Metrics. To evaluate the performance of the proposed SC-Dynamic R-CNN on the image data that we have, we utilize a set of prevalent performance metrics for object detection, which are AP_{50}, AP_{75}, and mAP. AP_{50} and AP_{75} are average precision with IoU (intersection over union) thresholds of 50% and 75%. The mAP is mean average precision. The reason to choose more than one threshold is to eliminate possible evaluation biases and provide more objective evaluation results. We have partitioned our data into three groups, namely, training set, validation set, and test-dev set. The proposed Dynamic R-CNN variant is trained and validated with the training set and validation set.

The final results are reported on the test-dev set. It is worth noting that our mAP averages AP_{50} and AP_{75} for each category as a whole. Generally speaking, the better the detection effect of the model, the higher the value of mAP.

3.2. Implementation Details. For truthful comparisons, all experiments are implemented using PyTorch and mmdetection [37]. And the experiments are carried out in the operating environment of Ubuntu 16.04 OS with 6 ×Intel(R) Core(TM) i7-7700 CPU, using an NVIDIA GeForce RTX 2080 GPU for training. The test experiments use the same configuration. The input image size of each network is 512 × 512 pixels unless noted. We train detectors with 12 epochs with an initial learning rate of 0.01. The SGD momentum is set to be 0.9, and weight decay is with a value of 0.0001. All other hyperparameters follow the settings in mmdetection [37] if not specifically noted.

3.3. Main Results. In the experimental results of this paper, we used "Adenocarcinoma," "Squamous," and "small cell," to represent adenocarcinoma, squamous cell carcinoma, and small cell carcinoma, respectively.

The detection results obtained under different models are shown in the following table:

There are five contemporary methods used to compare and benchmark the results of our proposed SC-Dynamic R-CNN. The five methods are ReinaNet [21], SSD [38], Faster R-CNN [23], Libra R-CNN [27], and Cascade R-CNN [25]. These methods are among the most popular object

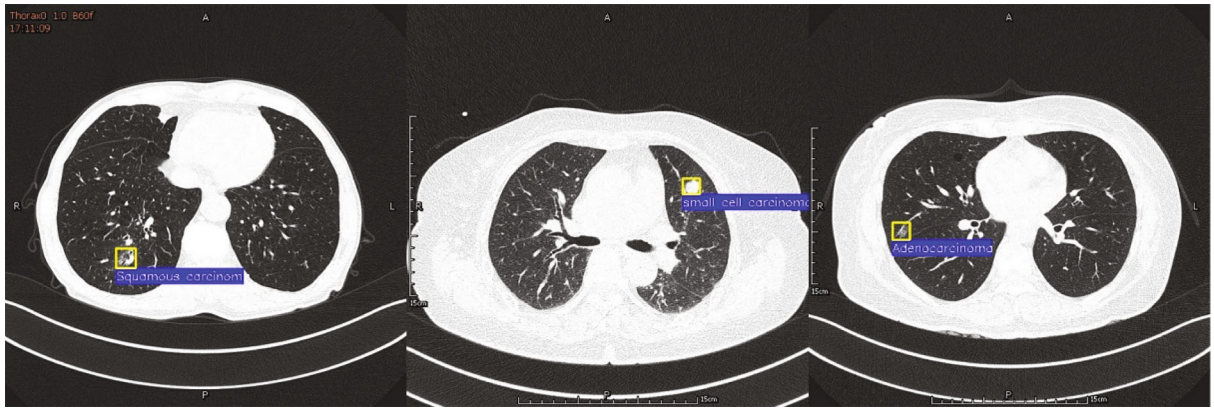


Figure 6: SC-Dynamic R-CNN detection effect diagram. Adenocarcinoma test results (left), small cell carcinoma test results (center), and squamous carcinoma test results (right).

detection neural network algorithms. The same lung cancer image data, training set, validation set, and test-dev set are used for a fair comparison. The performance of our proposed methods against the five popular methods is presented in Table 1.

The result shows that SC-Dynamic R-CNN achieves 88.1% mAP with ResNet-50, which is 8 points higher than the FPN-based Faster R-CNN baseline. As a one-stage detection network, RetinaNet and SSD achieved 81.5% and 78.9% mAP, respectively, whose accuracy is inferior to our method.

Moreover, SC-Dynamic R-CNN is much better than other networks at AP_{75}. This is because SC-Dynamic R-CNN can train better results by constantly increasing the IoU threshold. Although Cascade R-CNN also achieves good results in detection, our network is higher than Cascade R-CNN no matter being at AP_{50} or AP_{75} and our mAP is 4.1 points higher than that of Cascade R-CNN.

Our proposed method demonstrates a decent level of effectiveness and robustness. The performance accuracy of our method is consistent even with different IoU thresholds. The reason why our method surpasses other methods in term of accuracy is due to the novel enhancement implemented in the previous Dynamic R-CNN algorithm. During the training phase, the proposed variant is able to automatically adjust the label assignment criteria and the shape of regression loss function so that the training set is better utilized. Another distinctive improvement is to integrate self-calibration mechanism to the RPN of the previous methods and it helps CNN generate more discriminative representations and ultimately enhances the overall performance of the variant.

3.4. Ablation Experiment. To show the effectiveness of each proposed component, we report the overall ablation studies in Table 2.

These results show the effectiveness and robustness of our method.

(1) DBS L1 loss: compared with Dynamic R-CNN, DBS L1 loss improves the mAP of lung cancer detection from 85.4% to 87.1%. This proves that our proposed module has better performance than the Dynamic R-CNN loss module. Results in higher IoU metrics like AP_{75} are hugely improved, which validates the effectiveness of changing the loss function to compensate for the high-quality samples during training

(2) SCNet: when we replace the traditional convolution with SCNet, the mAP of lung cancer detection is improved from 87.1% to 88.1%. Compared with Dynamic R-CNN with DBL L1 loss, AP_{50} and AP_{75} increased by 0.9 points and 1.1 points, respectively, after adding SCNet. This also proves the effectiveness of SCNet for lung cancer detection

The experimental results of SC-Dynamic R-CNN are shown in the following figure:

As shown in Figure 6, this paper used the SC-Dynamic R-CNN model to detect lung cancer lesions and achieved good results. This fully demonstrates that our proposed model has greatly improved the recognition effect of lung cancer lesions.

4. Conclusion

To solve the problem that the biological characteristics of lung cancer were complex and difficult to detect, we proposed the SC-Dynamic R-CNN network. First, we extended the lung cancer dataset with the oversampling method and obtained the balanced dataset. Then, we added the self-calibrated convolution module to the Dynamic R-CNN network and proposed a new regression loss function, DBS L1 loss. This algorithm solves the problem of false detection and miss detection to a certain extent and greatly improves the detection accuracy of lung cancer. After experimental verification, the new algorithm achieves 88.1% mAP on the lung cancer dataset and it performed particularly well on high IoU threshold (such as AP_{75}). In the next work, we will try to further improve the accuracy of the network and verify the broad applicability of the model in cancer detection.

In future, it is always worthwhile to solve this issue with some other intelligence algorithms and the bio-inspired computational methods, such as monarch butterfly optimization (MBO) [39], earthworm optimization algorithm

(EWA) [40], elephant herding optimization (EHO) [41], moth search (MS) algorithm [42], slime mould algorithm (SMA) [43], hunger games search (HGS) [44], Runge Kutta optimizer (RUN) [45], colony predation algorithm (CPA) [46], Harris hawks optimization (HHO) [47], and Spiking neural P(SN-P) systems with learning [48].

Data Availability

The data can be provided upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (nos. 61972416, 61873280, and 61873281) and the Natural Science Foundation of Shandong Province (no. ZR2019MF012).

References

[1] "World Health Organization Cancer," 2018, https://www.who.int/news-room/fact-sheets/detail/cancer.

[2] R. Siegle, D. Naishadham, and A. Jemal, "Cancer statistics," *A Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 7–30, 2018.

[3] J. Lortet-Tieulent, I. Soerjomataram, J. Ferlay, M. Rutherford, E. Weiderpass, and F. Bray, "International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women," *Lung Cancer*, vol. 84, no. 1, pp. 13–22, 2014.

[4] C. Jacobs, E. M. van Rikxoort, E. T. Scholten et al., "Solid, part-solid, or non-solid," *Investigative Radiology*, vol. 50, no. 3, pp. 168–173, 2015.

[5] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1/3, pp. 389–422, 2002.

[6] C. L. Huang, H. C. Liao, and M. C. Chen, "Prediction model building and feature selection with support vector machines in breast cancer diagnosis," *Expert Systems with Applications*, vol. 34, no. 1, pp. 578–587, 2008.

[7] Y. Xu, Z. Jia, L. B. Wang et al., "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC Bioinformatics*, vol. 18, no. 1, pp. 1–17, 2017.

[8] A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[9] R. Vivanti, L. Joskowicz, N. Lev-Cohain, A. Ephrat, and J. Sosna, "Patient-specific and global convolutional neural networks for robust automatic liver tumor delineation in follow-up CT studies," *Medical and Biological Engineering and Computing Journal of the International Federation for Medical and Biological Engineering*, vol. 56, no. 9, pp. 1699–1713, 2018.

[10] J. Zhou, X. Zhang, and Z. Jiang, "Recognition of imbalanced epileptic EEG signals by a graph-based extreme learning machine," *Wireless Communications and Mobile Computing.*, vol. 2021, pp. 1–12, 2021.

[11] J. Zhang, J. Yu, S. Fu, and X. Tian, "Adoption value of deep learning and serological indicators in the screening of atrophic gastritis based on artificial intelligence," *The Journal of Supercomputing*, vol. 77, no. 8, pp. 8674–8693, 2021.

[12] Z. Wu, H. Zhu, G. Li et al., "An efficient Wikipedia semantic matching approach to text document classification," *Information Sciences*, vol. 393, no. 393, pp. 15–28, 2017.

[13] Z. Wu, L. Lei, G. Li et al., "A topic modeling based approach to novel document automatic summarization," *Expert Systems with Applications*, vol. 84, no. 84, pp. 12–23, 2017.

[14] R. Wang, Z. Wu, J. Lou, and Y. Jiang, "Attention-based dynamic user modeling and deep collaborative filtering recommendation," *Expert Systems with Applications*, vol. 188, no. 188, article 116036, 2022.

[15] G. Xu, Z. Wu, G. Li, and E. Chen, "Improving contextual advertising matching by using Wikipedia thesaurus knowledge," *Knowledge and Information Systems*, vol. 43, no. 3, pp. 599–631, 2015.

[16] Z. Wu, S. Shen, X. Lian, X. Su, and E. Chen, "A dummy-based user privacy protection approach for text information retrieval," *Knowledge-Based Systems*, vol. 195, no. 195, article 105679, 2020.

[17] L. Zhang, J. Huang, and L. Liu, "Improved deep learning network based in combination with cost-sensitive learning for early detection of ovarian cancer in color ultrasound detecting system," *Journal of Medical Systems*, vol. 43, no. 8, p. 251, 2019.

[18] M. Wu, C. B. Yan, H. Liu, and Q. Liu, "Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks," *Bioscience Reports*, vol. 38, no. 3, 2018.

[19] N. Tajbakhsh and K. Suzuki, "Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs," *Pattern Recognition*, vol. 63, pp. 476–486, 2017.

[20] J. Redmon and A. Farhadi, 2018, Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[21] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[22] R. Girshick, *Fast R-CNN*, 2015, Computer Ence.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[25] Z. Cai and N. Vasconcelos, "Cascade R-CNN delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6154–6162, 2017.

[26] S. Wang, L. Wang, L. Wang, Z. Yu, X. Zhao, and X. Wang, "AT-Cascade R-CNN: a novel attention-based cascade R-CNN model for ovarian cancer lesion identification," *International Journal of Adaptive and Innovative Systems*, vol. 3, no. 1, pp. 74–86, 2021.

[27] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: towards balanced learning for object detection," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR*, pp. , 2020821–830, 2020.

[28] H. Zhang, H. Chang, B. Ma, N. Wang, and X. Chen, "Dynamic R-CNN: towards high quality object detection via dynamic training," in *European Conference on Computer Vision*, Springer, Cham, 2020.

[29] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[30] Q. Hou, M. M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. Torr, "Deeply supervised salient object detection with short connections," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3203–3212, 2017.

[31] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters — improve semantic segmentation by global convolutional network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4353–4361, 2017.

[32] Y. Fisher and K. Vladlen, "Multi-scale context aggregation by dilated convolutions," vol. 2, 2016.

[33] J. J. Liu, Q. Hou, M. M. Cheng, C. Wang, and J. Feng, "Improving convolutional networks with self-calibrated convolutions," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10096–10105, 2020.

[34] S. Gilad, G. Lithwick-Yanai, I. Barshack et al., "Classification of the four main types of lung cancer using a microRNA-based diagnostic assay," *The Journal of Molecular Diagnostics*, vol. 14, no. 5, pp. 510–517, 2012.

[35] I. Pitas, *Digital Image Processing Algorithms and Applications*, John Wiley & Sons, 2000.

[36] R. Liu, L. O. Hall, K. W. Bowyer, D. B. Goldgof, R. Gatenby, and K. B. Ahmed, "Synthetic minority image over-sampling technique: how to improve AUC for glioblastoma patient survival prediction," in *In2017 IEEE international conference on systems, man, and cybernetics SMC*, pp. 1357–1362, 2017.

[37] K. Chen, J. Wang, J. Pang et al., "MMDetection: open mmlab detection toolbox and benchmark," 2018, http://arxiv.org/abs/1906.07155.

[38] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *InEuropean conference on computer vision*, 2016.

[39] G. G. Wang, S. Deb, and Z. Cui, "Monarch butterfly optimization," *Neural Computing and Applications*, vol. 31, no. 7, pp. 1995–2014, 2019.

[40] G. G. Wang, S. Deb, and L. D. Coelho, "Earthworm optimisation algorithm: a bio-inspired metaheuristic algorithm for global optimisation problems," *International Journal of Bio-Inspired Computation*, vol. 12, no. 1, pp. 1–22, 2018.

[41] G. G. Wang, S. Deb, and L. D. Coelho, "Elephant herding optimization," *International Symposium on Computational and Business Intelligence*, pp. 1–5, 2015.

[42] G. G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, no. 2, pp. 151–164, 2018.

[43] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: a new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020.

[44] Y. Yang, H. Chen, A. A. Heidari, and A. H. Gandomi, "Hunger games search: visions, conception, implementation, deep analysis, perspectives, and towards performance shifts," *Expert Systems with Applications*, vol. 177, no. 177, article 114864, 2021.

[45] I. Ahmadianfar, A. A. Heidari, A. H. Gandomi, X. Chu, and H. Chen, "RUN beyond the metaphor: an efficient optimization algorithm based on Runge Kutta method," *Expert Systems with Applications*, vol. 181, no. 181, article 115079, 2021.

[46] J. Tu, H. Chen, M. Wang, and A. H. Gandomi, "The colony predation algorithm," *Journal of Bionic Engineering*, vol. 18, no. 3, pp. 674–710, 2021.

[47] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, no. 97, pp. 849–872, 2019.

[48] T. Song, L. Pan, T. Wu, P. Zheng, M. D. Wong, and A. Rodríguez-Patón, "Spiking neural P systems with learning functions," *IEEE Transactions on Nanobioscience*, vol. 18, no. 2, pp. 176–190, 2019.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 1633858, 18 pages
https://doi.org/10.1155/2022/1633858

Research Article

Breast Tumor Classification Using Intratumoral Quantitative Ultrasound Descriptors

Sabiq Muhtadi

Department of Electrical and Electronic Engineering, Islamic University of Technology, Gazipur, Bangladesh

Correspondence should be addressed to Sabiq Muhtadi; sabiqmuhtadi@iut-dhaka.edu

Received 5 October 2021; Revised 15 February 2022; Accepted 23 February 2022; Published 7 March 2022

Academic Editor: Mario Cesarelli

Breast cancer is a global epidemic, responsible for one of the highest mortality rates among women. Ultrasound imaging is becoming a popular tool for breast cancer screening, and quantitative ultrasound (QUS) techniques are being increasingly applied by researchers in an attempt to characterize breast tissue. Several different quantitative descriptors for breast cancer have been explored by researchers. This study proposes a breast tumor classification system using the three major types of intratumoral QUS descriptors which can be extracted from ultrasound radiofrequency (RF) data: spectral features, envelope statistics features, and texture features. A total of 16 features were extracted from ultrasound RF data across two different datasets, of which one is balanced and the other is severely imbalanced. The balanced dataset contains RF data of 100 patients with breast tumors, of which 48 are benign and 52 are malignant. The imbalanced dataset contains RF data of 130 patients with breast tumors, of which 104 are benign and 26 are malignant. Holdout validation was used to split the balanced dataset into 60% training and 40% testing sets. Feature selection was applied on the training set to identify the most relevant subset for the classification of benign and malignant breast tumors, and the performance of the features was evaluated on the test set. A maximum classification accuracy of 95% and an area under the receiver operating characteristic curve (AUC) of 0.968 was obtained on the test set. The performance of the identified relevant features was further validated on the imbalanced dataset, where a hybrid resampling strategy was firstly utilized to create an optimal balance between benign and malignant samples. A maximum classification accuracy of 93.01%, sensitivity of 94.62%, specificity of 91.4%, and AUC of 0.966 were obtained. The results indicate that the identified features are able to distinguish between benign and malignant breast lesions very effectively, and the combination of the features identified in this research has the potential to be a significant tool in the noninvasive rapid and accurate diagnosis of breast cancer.

1. Introduction

According to the World Health Organization (WHO) fact-sheet, breast cancer is the world's most prevalent form of cancer, with a staggering 7.8 million patients being diagnosed in the 5-year period between 2016 and 2020 [1]. It was the most commonly diagnosed form of cancer, as well as the second leading cause of cancer-related deaths for women in 2020 [2]. Early diagnosis of breast cancer is crucial to the survival of patients due to its role in treatment selection as well as prediction of response to therapy [3].

Ultrasound imaging has established itself as an important noninvasive screening technique for breast cancer [4]. It retains a significant advantage over other modalities such as mammography due to its nonionizing nature, low costs, and high portability. Furthermore, ultrasound imaging can improve tumor detection during breast cancer diagnosis by as much as 17% [5], as well as reduce the number of nonessential biopsies by 40% [6]. However, ultrasound imaging suffers from system and operator dependency [7, 8] which negates its reproducibility. Furthermore, conventional ultrasound imaging procedures are qualitative in nature, and thus radiological evaluation of ultrasound B-mode images relies heavily on the diagnostic experience of the radiologist.

Quantitative ultrasound (QUS) techniques represent a domain of ultrasound imaging procedures which extract various quantitative measures of tissue microstructure [9, 10]. Unlike conventional ultrasound imaging techniques,

QUS procedures are independent of the system and operator related factors [11, 12] and as a result are highly reproducible. Furthermore, QUS techniques can provide an indication of diagnosis without the need for expert evaluation and thus have the potential for rapid diagnosis of conditions such as breast cancer. The utility of QUS techniques has been established over multiple areas, such as differentiation between benign and malignant thyroid tissues [13], detection of prostate cancer [14, 15], and characterization of carotid plaques [16]. Several different quantitative parameters have also been explored by researchers with regard to characterization of breast tissue.

QUS spectroscopy involves extraction of spectral parameters from the attenuation-corrected normalized power spectrum of raw ultrasonic radiofrequency signals. Lizzi et al. [17, 18] proposed the linear parameterization of this normalized power spectrum in order to extract the spectral slope, spectral intercept, and midband fit of ultrasound echoes. These features provide a measure of shape, size, concentration, and power of acoustic scatterers and have been applied for both diagnosis of breast lesions [19, 20], as well as noninvasive evaluation of response to chemotherapy [21, 22] with notable success.

The statistics of the acquired ultrasound envelope signal can be modelled as a probability density function (PDF) in order to analyze the scattering properties of soft tissue. Several well-known statistical distributions may be utilized in this regard to model the statistics of the envelope, and two popular distributions which are applied to model scattered signals from the breast are the Nakagami and homodyned K distribution. The Nakagami distribution was proposed for the modelling of ultrasonic backscatter by Shankar [23]. Several approaches have been proposed by researchers for the classification of breast lesions using the characteristics of the Nakagami distribution. The parameters of the distribution have been analyzed for their potential as quantitative descriptors of breast cancer by themselves [24], through compounding approaches [25], in conjunction with the parameters of other distributions such as the K distribution [26], as well as in conjunction with other types of quantitative descriptors such as entropy and texture [27, 28]. The homodyned K distribution was proposed for the modelling of ultrasound echoes by Dutt and Greenleaf [29] and later modified by Hruska [30] and Hruska and Oelze [31]. The homodyned K distribution parameters have been applied in conjunction with breast imaging reporting and data system (BIRADS) descriptors as well as shear wave elasticity (SWE) features for the classification of breast lesions [32, 33].

Tumors are known to exhibit heterogeneities in physiology, microenvironment, and metabolism, which is significant for the characterization of cancer [34–37]. These heterogeneities may be quantified using texture analysis techniques [38]. In the context of ultrasonic B-mode images, texture analysis provides an indication of gray-level transitions by analyzing the spatial relationships between neighboring pixels in an image, and this is useful for evaluating the differing textures exhibited by benign and malignant masses [20]. With this rationale, texture analysis techniques applied to ultrasound scans have been utilized by several studies for the characterization of breast lesions [39–42].

This study proposes a breast tumor classification system that utilizes the three major types of QUS features used by researchers to characterize breast lesions: spectral features, envelope statistics features, and texture features. To my knowledge, no other research works have evaluated the features analyzed in this study simultaneously for breast cancer diagnosis. A total of 16 different features were extracted from ultrasound patient data for evaluation across two different datasets, of which one is balanced, and the other is severely imbalanced. Holdout validation was used to split the balanced dataset into 60% training and 40% testing sets, and feature selection in the form of sequential forward selection (SFS) was applied to the training set to identify the subset of features most relevant to the classification of benign and malignant breast tumors. The performance of the identified features was evaluated on the test set, where a maximum classification accuracy of 95% and an area under the receiver operating characteristics curve (AUC) of 0.968 were obtained. The performance of the identified relevant features was further validated on the imbalanced dataset, where a hybrid resampling strategy was firstly utilized to create an optimal balance between benign and malignant samples. A maximum classification accuracy of 93.01%, sensitivity of 94.62%, specificity of 91.4%, and AUC of 0.966 were obtained. The results indicate that the identified features are able to distinguish between benign and malignant breast lesions very effectively, and the combination of the features identified in this research work has the potential to be a significant tool in the noninvasive rapid and accurate diagnosis of breast cancer.

2. Materials and Methods

2.1. Description of Datasets

2.1.1. OASBUD Dataset. The Open Access Series of Breast Ultrasonic Data (OASBUD) [43] was utilized in this study. It consists of ultrasound radiofrequency (RF) data of 100 breast lesions of patients at the Oncology Institute in Warsaw. Among these, 52 were malignant lesions, and 48 were benign. All malignant lesions were histologically assessed by core needle biopsy. 37 out of the 48 benign lesions were also histologically assessed; the remaining 13 did not qualify for a biopsy but were observed by a radiologist over a 2-year period. The ultrasound data was recorded at the Department of Ultrasound, Institute of Fundamental Technological Research Polish Academy of Sciences, and the study was approved by the Institutional Review Board (IRB). Patients were examined by a radiologist with 18 years of experience, following the BI-RADS guidelines as well as the Polish Ultrasound Society standards. For each lesion, two individual longitudinal and transverse scans were recorded using an Ultrasonix SonixTouch Research ultrasound scanner with an L14-5/38 linear array transducer and a center frequency of 10 MHz. Each scan consisted of 512 RF lines, and the signals were digitized using a 40 MHz sampling

frequency. The region of interest (ROI) for each individual scan was indicated by the radiologist.

2.1.2. ATL Dataset. The ultrasound data from ATL's premarket approval (PMA) IRB-approved study undertaken in 1994 [19] was also used for this research. It consists of ultrasound RF data of breast lesions from 130 patients. Among these, 104 were benign and 26 were malignant, all histologically assessed by core needle biopsy. The ultrasonic data was recorded at three clinical sites, Thomas Jefferson University, University of Cincinnati, and Yale University, during routine ultrasonic examinations of patients scheduled for biopsy. The tumors were examined by an experienced radiologist using a Phillips Ultrasound UM-9 HDI scanner, with an L10-5 linear array transducer and a center frequency of 7.5 MHz. The L10-5 transducer was used at a default power level and a single transmit focal length, as selected by the operator. All standard ultrasonic breast examination procedures were maintained during the examination. Multiple views were selected by the radiologist for every lesion, which included at least a radial and an antiradial view. The signals were digitized by interfacing a Spectrasonics Inc. (King of Prussia, PA) acquisition module using a 20 MHz sampling frequency and an effective dynamic range of 14 bits. Time-gain-control (TGC) data was obtained before each scan, and the acquired data was corrected for TGC before processing. As can be observed, the dataset contains quite a high imbalance ratio between benign and malignant cases (4 : 1).

2.2. Feature Extraction. Three types of features were extracted from patient ultrasound scans for use in this study: spectral features, envelope statistics features, and texture features. All processing codes were written in MATLAB™ (The MathWorks, Inc., Natick, MA).

2.2.1. Spectral Features. Spectral features were obtained from parametric images formed using spectrum analysis parameters [18, 44, 45]. A Hamming window of length 2.4 mm was applied to the RF data of each ultrasound patient scan. The power spectrum of the windowed RF data was then computed using the Fourier transform and expressed in dB. Linear regression was applied to the power spectrum over the 6 dB bandwidth of the signal. This regression analysis yields the slope (SL) of the regression line, the value at midpoint (MBF) of signal bandwidth, and the intercept at zero frequency (INT). Images of these parameters were formed by progressively sliding the Hamming window over each RF data with an overlap of 87.5% and repeating the above sequence.

The linear regression line which approximates the normalized power spectrum can be expressed as

$$P(f) = I + sf, \tag{1}$$

where f, s, and I represent frequency, SL, and INT, respectively. Thus, the MBF can be expressed as

$$M = I + sf_0, \tag{2}$$

with f_0 representing center frequency of the usable bandwidth.

The presence of frequency-dependent attenuation affects the MBF and SL values obtained during analysis [19]. To compensate for this, the attenuation (in dB) is assumed to vary linearly with frequency, and this approximation is validated through the findings of Alam et al. [19] and Bamber [46] on the invariance of intercept in the presence of attenuation. For this study, the MBF and SL were corrected as follows:

$$M_\alpha = P_\alpha(f_0) = I - (s - 2\alpha d) f_0, \tag{3}$$

$$s_\alpha = (s - 2\alpha d), \tag{4}$$

where α represents the effective attenuation coefficient and d represents the depth of the intervening tissue. The value of the attenuation coefficient α was set to 1.0 dB/MHz-cm, based on the attenuation coefficient for muscle reported by Mast [47].

Figure 1 illustrates the three types of spectral parametric images (MBF, INT, and SL) that are formed from ultrasound RF data. The mean and standard deviation of pixel values from the intratumoral region of these parametric images were used in this study for the classification of breast cancer.

2.2.2. Envelope Statistics Features. Ultrasonic pulses moving through tissue are subject to scattering due to artifacts located within the tissue, which are aptly termed as "scatterers." Consequently, the backscattered ultrasonic echo signal received at the transducer can be viewed as the superposition of scattered signals from individual scatterers within the tissue [48]. Application of a statistical distribution model to this backscattered ultrasound envelope can provide information related to tissue microstructure. Two such statistical distribution models that effectively describe the scattering characterization of ultrasound echo signals from breast tissue are the Nakagami distribution [23] and the homodyned K distribution [31].

(1) Homodyned K Distribution. The homodyned K distribution is an analytically complex model; however, it is more versatile than models such as the Rayleigh distribution and the K distribution [49]. The probability density function (pdf) $H(A)$ of the homodyned K distribution is expressed in the form of an improper integral [29] as follows

$$H(A) = A \int_0^\infty x J_0(sx)\, J_0(Ax) \left(1 + \frac{x^2\sigma^2}{2\mu}\right)^{-\mu} dx, \tag{5}$$

where J_0 is a zero-order Bessel function of the first kind, s^2 is the coherent signal energy, σ^2 is the diffuse signal energy, and μ is a measure of the effective number of scatterers in the target cell. The ratio of the coherent to diffuse signal can be used as a derived parameter $k = s/\sigma$ to define the periodicity in scatterer locations. The parameters k and μ are believed to provide an accurate description of tissue scattering properties [49].

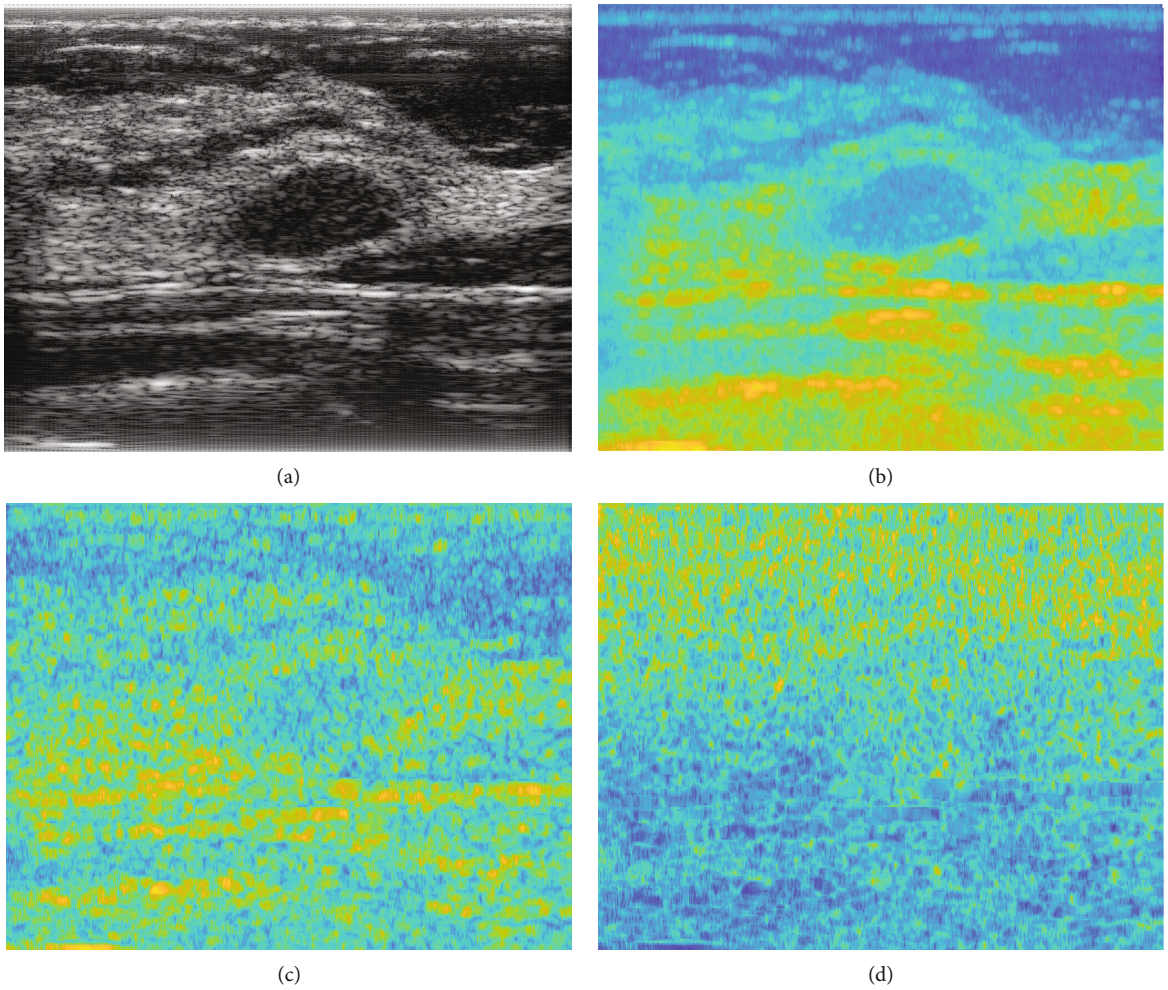


Figure 1: (a) Ultrasound B-mode image and corresponding (b) midband fit (MBF) parametric image, (c) spectral intercept (INT) parametric image, and (d) spectral slope (SL) parametric image.

The homodyned K parameter estimation technique outlined by Hruska et al. [31] was utilized for this study. This technique uses the signal-to-noise ratio (SNR), skewness, and kurtosis of fractional order moments to estimate the parameters of the homodyned K distribution.

A third parameter, the diffuse-to-total signal power ratio [50] $h = 1/(k+1)$, is also defined. The parameters μ, k, and h were estimated by fitting the homodyned K distribution to all samples within the tumor region of each ultrasound envelope image, and these parameters were then utilized for the classification of breast lesions.

(2) Nakagami Distribution. The Nakagami distribution [51] was introduced by Nakagami (1943, 1960) in the context of wave propagation. It is far less analytically complex than the homodyned K distribution. The pdf $N(A)$ of the ultrasonic backscattered envelope under the Nakagami distribution model is given by

$$N(A) = \frac{2m^m A^{2m-1}}{\Gamma(m)\Omega^m} e^{-\frac{mA^2}{\Omega}} U(A). \tag{6}$$

Here, $\Gamma(.)$ and $U(.)$ represent the Euler gamma function and the unit step function, respectively.

The Nakagami distribution has two parameters, expressed as follows:

$$m = \frac{\left[E\left(R^2\right)\right]^2}{E\left[R^2 - E\left(R^2\right)\right]^2}, \tag{7}$$

$$\Omega = E\left[R^2\right], \tag{8}$$

where R represents the ultrasonic backscattered envelope and m is referred to as the shape parameter, providing information about envelope statistics. In the case of the Nakagami distribution, it is constrained such that $m \geq 0.5$ [51], in which case it is referred to as the Nakagami parameter. Ω is a scaling parameter.

The similarity between the Nakagami distribution and the K distribution may be used to define a third parameter of the Nakagami distribution. The K distribution has a

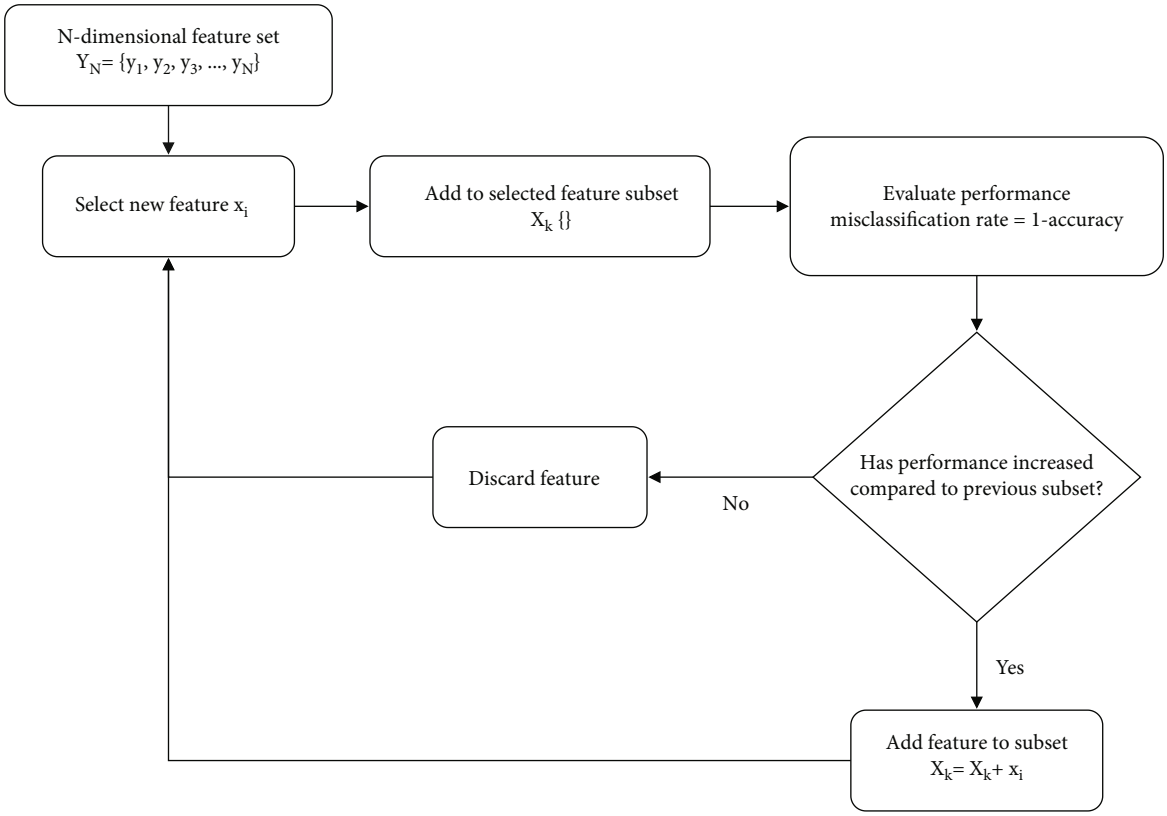


Figure 2: Flowchart of sequential forward selection (SFS) algorithm.

Table 1: Feature values for benign and malignant cases and statistical significance of features in the OASBUD dataset.

Feature	Benign values	Malignant values	p value	Statistical significance
Mean of MBF	88.25 ± 6.84	92.06 ± 9.25	<0.05	*
Standard deviation of MBF	4.61 ± 0.86	5.08 ± 1.19	<0.05	*
Mean of INT	95.19 ± 4.25	91.8 ± 6.56	<0.05	*
Standard deviation of INT	14.98 ± 0.78	15.32 ± 0.76	<0.05	*
Mean of SL	−4.54 ± 0.84	−4.52 ± 1.01	>0.05	~
Standard deviation of SL	2.09 ± 0.12	2.12 ± 0.08	>0.05	~
k (homodyned K)	0.73 ± 0.08	0.84 ± 0.23	<0.001	**
μ (homodyned K)	0.18 ± 0.1	0.28 ± 0.26	<0.001	**
h (homodyned K)	0.58 ± 0.03	0.55 ± 0.08	<0.001	**
m (Nakagami)	0.54 ± 0.06	0.67 ± 0.17	<0.001	**
Ω (Nakagami)	408736.82 ± 134753.77	189811.2 ± 166031.05	<0.001	**
α (Nakagami)	219.74 ± 55.23	119.24 ± 98.33	<0.001	**
Contrast	2.51 ± 1.11	2.01 ± 1.02	<0.05	*
Correlation	0.59 ± 0.06	0.61 ± 0.07	<0.05	*
Energy	0.135 ± 0.09	0.18 ± 0.10	<0.05	*
Homogeneity	0.67 ± 0.07	0.7 ± 0.08	<0.05	*

Table 2: Feature values for benign and malignant cases and statistical significance of features in the ATL dataset.

Feature	Benign values	Malignant values	p value	Statistical significance
Mean of MBF	76.44 ± 18.55	77.71 ± 11.84	>0.05	~
Standard deviation of MBF	6.26 ± 1.51	6.53 ± 1.52	>0.05	~
Mean of INT	64.77 ± 14.01	64.11 ± 6.93	>0.05	~
Standard deviation of INT	13.32 ± 1.09	12.91 ± 0.54	>0.05	~
Mean of SL	−3.28 ± 1.5	−3.45 ± 0.69	>0.05	~
Standard deviation of SL	1.73 ± 0.14	1.73 ± 0.09	>0.05	~
k (homodyned K)	0.35 ± 0.14	0.5 ± 0.1	<0.001	**
μ (homodyned K)	0.16 ± 0.13	0.26 ± 0.21	<0.05	*
h (homodyned K)	0.75 ± 0.08	0.67 ± 0.04	<0.001	**
m (Nakagami)	0.33 ± 0.09	0.44 ± 0.09	<0.001	**
Ω (Nakagami)	5116.67 ± 6622.77	1692.06 ± 997.2082	<0.05	*
α (Nakagami)	31.21 ± 16.11	16.46 ± 6.6	<0.001	**
Contrast	3.55 ± 1.64	2.97 ± 0.8	>0.05	~
Correlation	0.4 ± 0.07	0.38 ± 0.04	>0.05	~
Energy	0.17 ± 0.06	0.17 ± 0.07	>0.05	~
Homogeneity	0.66 ± 0.06	0.67 ± 0.05	>0.05	~

cumulative distribution expressed as

$$F_K(r) = \frac{2b}{\Gamma(M)}\left(\frac{br}{2}\right)^M K_{M-1}(br) r \geq 0 \quad M \geq 0, \tag{9}$$

where M provides a measure of the effective number of scatterers in the target cell and b is a scaling parameter. The parameters of the K distribution can be expressed in terms of the Nakagami distribution [24]:

$$M = \frac{2m}{1-m}, \tag{10}$$

$$b = 2\sqrt{\frac{2m}{\Omega(1-m)}}. \tag{11}$$

Using this relationship, a parameter α can be defined, where $\alpha = 1/b$, or

$$\alpha = \frac{1}{2}\sqrt{\frac{\Omega(1-m)}{2m}}, \tag{12}$$

where α is defined as the effective cross-section of scatterers in the target cell [24].

The parameters m, Ω, and α were estimated by fitting the Nakagami distribution to all samples within the tumor region of the ultrasound envelope image, and these parameters were then utilized for the classification of breast lesions.

2.2.3. Texture Features. The texture of the ultrasound envelope images was quantified using gray-level cooccurrence matrix (GLCM) techniques. GLCM techniques quantify texture by evaluating the spatial relationship between neighboring pixels in an image [41]. A GLCM matrix is created by calculating how often a pixel with gray-level intensity value i occurs adjacent to a pixel with the value j. Let $P(i, j)$ denote the GLCM matrix representing the probability of having neighboring pixels with gray-level intensities i and j in the ultrasound image. Let μ and σ denote the mean and standard deviation for row i or column j of the GLCM matrix. The following four parameters may be defined from such a matrix

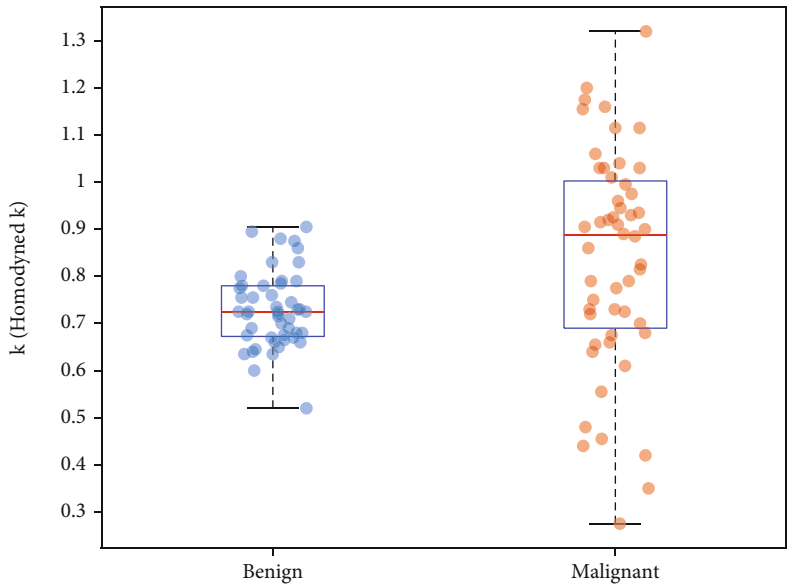
$$\text{Contrast} = \sum_{i,j} |i-j|^2 P(i,j), \tag{13}$$

$$\text{Correlation} = \frac{1}{\sigma_i \sigma_j}\sum_{i,j}(i-\mu_i)\left(j-\mu_j\right)P(i,j), \tag{14}$$

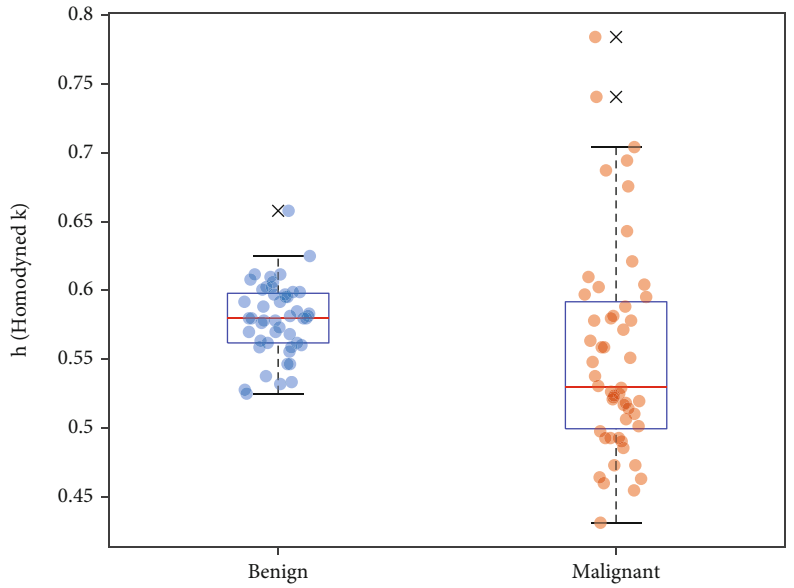
$$\text{Energy} = \sum_{i,j} P^2(i,j), \tag{15}$$

$$\text{Homogeneity} = \sum_{i,j}\frac{P(i,j)}{1+|i-j|}. \tag{16}$$

Contrast represents a measure of gray-level variations in the parametric image. Correlation provides an indication of the linear correlation between neighboring pixels. Energy quantifies textural uniformity between neighboring pixels, and homogeneity represents a measure of the incidence of pixel pairs of different intensity within the parametric image. To extract GLCM features, an ROI composed of the minimum bounding rectangular area around the tumor of each ultrasound envelope image was formed, similar to the procedure followed by [41]. The full range of gray levels in each ROI was linearly scaled into 16 discrete gray levels. GLCM matrices were then formed at five interpixel distances, 1, 2,



(a)



(b)

Figure 3: Continued.

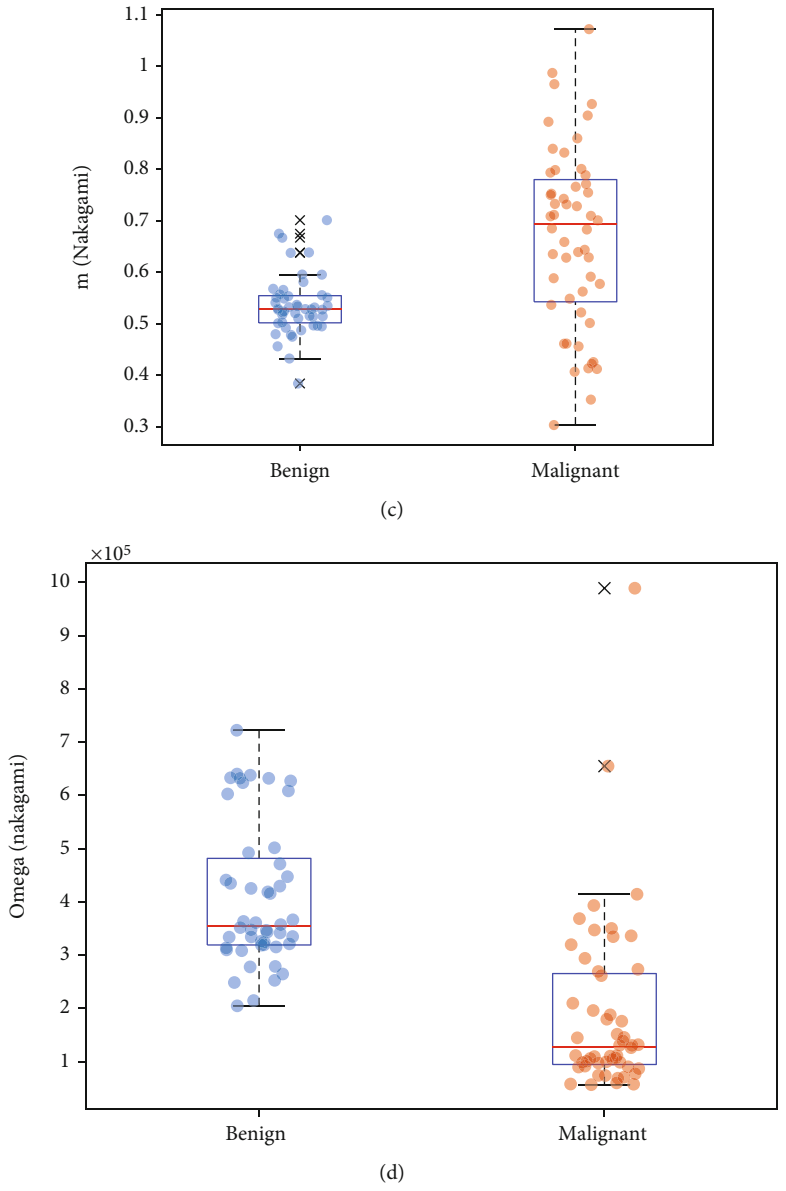


Figure 3: Box and scatter plots of (a) k (homodyned K), (b) h (homodyned K), (c) m (Nakagami), and (d) Ω (Nakagami) values from the OASBUD dataset.

Table 3: Classification performance of the four selected features on the testing portion of the OASBUD dataset.

Classifier	Classification accuracy	Sensitivity	Specificity	AUC	95% CI
KNN	92.5%	95%	90%	0.963	0.823~0.997
SVM	87.5%	85%	90%	0.968	0.878~0.995
RF	95%	95%	95%	0.959	0.797~0.993

3, 4, and 5 pixels, and at four angular directions, 0°, 45°, 90°, and 135°, and the four GLCM features were calculated from each of the GLCM matrices. All four texture features were averaged over distances and angular directions to obtain final values for each patient and then used for classification of breast lesions.

2.3. Resampling of ATL Dataset. As noted before, the ATL dataset contains a high level of imbalance between benign and malignant cases (4 : 1). A hybrid resampling strategy is applied in order to mitigate the imbalance between the classes. The number of majority class instances are firstly reduced using undersampling to decrease the imbalance

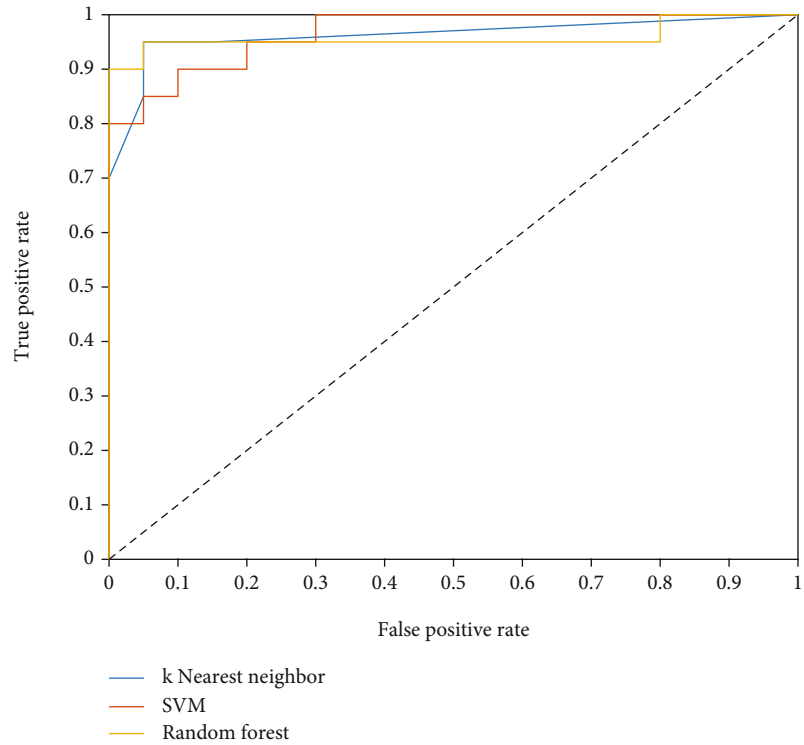


Figure 4: ROC curves obtained by the three classifiers using the 4 selected features in the testing portion of the OASBUD dataset.

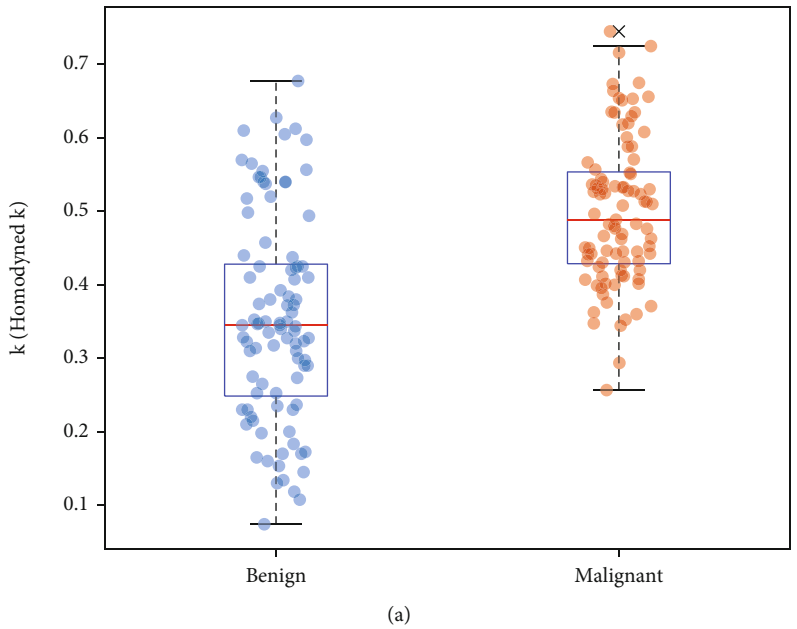
ratio between classes. Oversampling is then performed to generate new minority class samples in order to balance the dataset. The hybrid strategy creates an optimal balance between the classes and ensures the quality of the resampled data. Synthetic minority oversampling (SMOTE) [52] is used for oversampling, while Tomek links [53] are used for undersampling. They are described below.

2.3.1. Smote. SMOTE is an oversampling technique that is used to synthesize minority class instances based on their nearest neighbors and is frequently applied to address class imbalance in the medical domain [54]. Consider an k -dimensional dataset with samples of x_i, where $x_i = (i = 1, 2, 3, \cdots..n)$ and k represents the number of features. Let A represent the majority class with c samples and B represent the minority class with d samples, such that $c + d = n$ and $c \geq d$. SMOTE processes the dataset as follows: (i) for each minority class sample $b_i\,(i = 1, 2, \cdots, d)$, identify its T nearest neighbors, (ii) select a sample b_j from the T nearest neighbors of b_i and generate a synthetic data sample $p_i = x_i + (x_j - x_i) \times \lambda$, where $\lambda \epsilon [0, 1]$ is a random number, (iii) repeat s_i times to obtain s_i new synthetic samples of b_i. In this work, a T value of 5 was used.

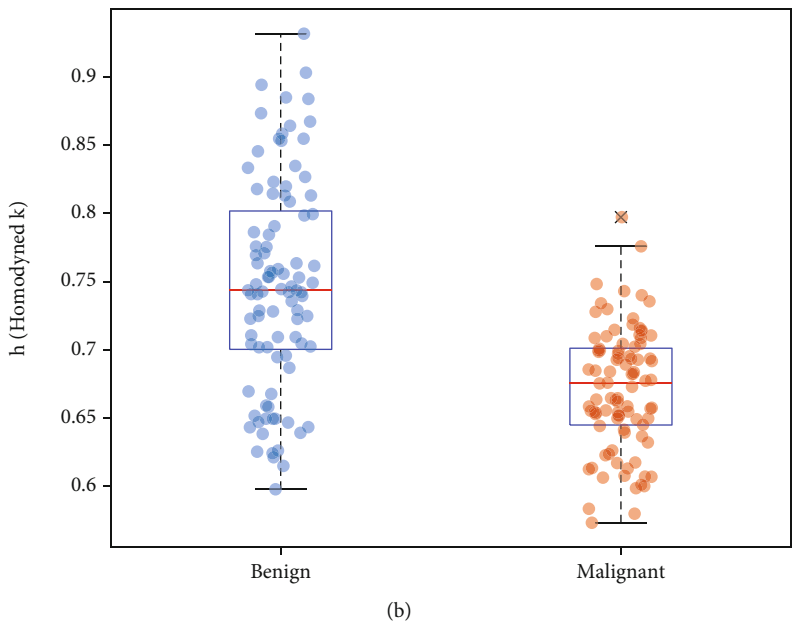
2.3.2. Tomek Links. Tomek link is an undersampling method that is used to eliminate majority instances from the dataset whenever a "Tomek link" is found. Let b_i denote a sample from the minority class and a_i denote a sample from the majority class. Then b_i and a_i are said to form a Tomek link pair if there is no sample x_k such that $d(b_i, x_k) < d(b_i, a_i)$, where d is used to represent distance between two samples. In this instance, the majority sample a_i is eliminated as a process of under sampling.

2.4. Sequential Forward Selection. Sequential forward selection (SFS) is a wrapper method that adds relevant features to the selected feature subset over multiple iterations on the basis of an evaluation criterion. The process begins with an empty subset of selected features. In the first iteration the model is trained using each feature individually, and the best performing feature is identified based on the evaluation metric and added to the selected feature subset. In the second iteration, the model is trained using pairings of the already selected feature along with each of the remaining features. The performance of each pair is analyzed using the evaluation metric, and the feature that achieves the best performance when paired with the first feature is added to the selected feature subset, but only if the performance of the pair is higher than the performance of the best individual feature in terms of the evaluation criterion. This process is repeated over multiple iterations until no improvement in the evaluation criterion is obtained by adding more features. The misclassification rate was used as the evaluation criterion in this study. Figure 2 illustrates a flowchart of the SFS process.

2.5. Performance Evaluation. A total of 16 features were extracted from the intratumoral region of ultrasound scans in both OASBUD and ATL datasets: (i) mean of MBF, (ii) standard deviation of MBF, (iii) mean of INT, (iv) standard deviation of INT, (v) mean of SL, (vi) standard deviation of SL, (vii) k (homodyned K), (viii) μ (homodyned K), (ix) h



(a)



(b)

Figure 5: Continued.

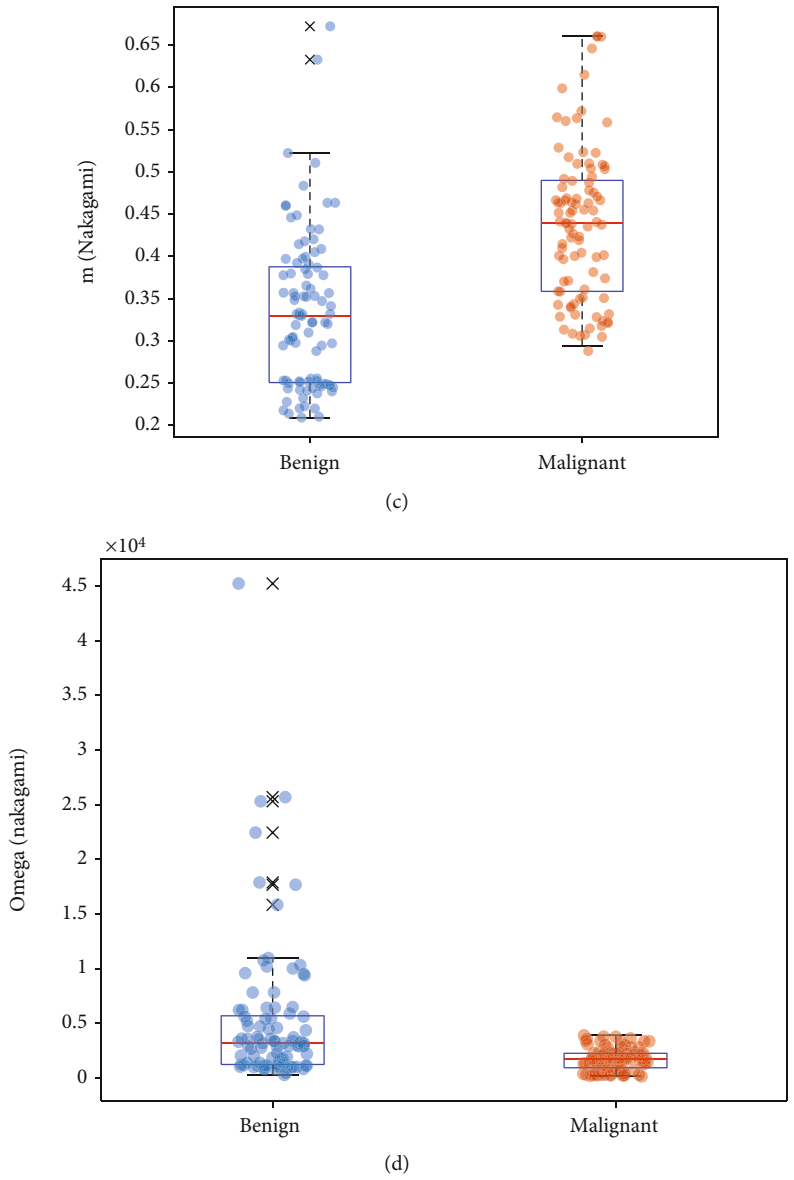


Figure 5: Box and scatter plots of (a) k (homodyned K), (b) h (homodyned K), (c) m (Nakagami), and (d) Ω (Nakagami) values from the ATL dataset.

Table 4: Classification performance of the four selected features in the ATL dataset without resampling.

Validation	Classifier	Classification accuracy	Sensitivity	Specificity	AUC	95% CI
10-fold CV	k-NN	79.23%	34.62%	90.38%	0.805	0.705~0.879
	SVM	84.62%	42.31%	95.2%	0.895	0.803~0.946
	RF	85.38%	65.38%	90.38%	0.849	0.748~0.92
LOOCV	k-NN	78.462%	30.77%	90.38%	0.855	0.753~0.918
	SVM	84.62%	42.31%	95.2%	0.892	0.811~0.944
	RF	87.3%	53.84%	92.3%	0.856	0.758~0.916

(homodyned K), (x) m (Nakagami), (xi) Ω (Nakagami), (xii) α (Nakagami), (xiii) contrast, (xiv) correlation, (xv) energy, and (xvi) homogeneity. Most lesions in both datasets were scanned at multiple intersecting scan planes, thereby providing complementary data for a given lesion. If a lesion had multiple scans, each quantitative feature value for multiple scans of a specific lesion was averaged to arrive at a single number. A two-sided Wilcoxon rank sum test (95%

Table 5: Classification performance of the four selected features in the ATL dataset with SMOTE.

Validation	Classifier	Classification accuracy	Sensitivity	Specificity	AUC	95% CI
10-fold CV	k-NN	87.2%	94.23%	79.81%	0.948	0.895~0.966
	SVM	82.21%	82.69%	81.73%	0.909	0.857~0.942
	RF	87.98%	92.31%	83.68%	0.956	0.903~0.972
LOOCV	k-NN	88.94%	82.69%	95.2%	0.959	0.921~0.982
	SVM	82.69%	82.69%	82.69%	0.909	0.859~0.942
	RF	89.42%	88.46%	90.38%	0.948	0.903~0.971

Table 6: Classification performance of the four selected features in the ATL dataset with SMOTE-Tomek.

Validation	Classifier	Classification accuracy	Sensitivity	Specificity	AUC	95% CI
10-fold CV	k-NN	88.17%	93.55%	82.8%	0.943	0.90~0.97
	SVM	80.65%	80.65%	80.65%	0.909	0.858~0.947
	RF	93.01%	94.62%	91.4%	0.966	0.928~0.984
LOOCV	k-NN	86.6%	93.55%	79.57%	0.955	0.917~0.979
	SVM	84.95%	83.87%	86.02%	0.917	0.856~0.95
	RF	91.4%	93.55%	89.25%	0.964	0.93~0.985

confidence) was performed on each of the extracted features in both datasets to assess statistical significance between benign and malignant groups. The purpose of the statistical test was solely to demonstrate discrimination capability of the extracted features.

The OASBUD dataset was used to determine the relevant features for classification of breast lesions as it contains a healthy balance between benign and malignant cases. Holdout validation was utilized to split the OASBUD dataset into 60% training and 40% testing sets. SFS was applied on the training set to identify the best performing features, and the performance of these features was evaluated using the test set. Three different algorithms were used for classification: (i) K-nearest neighbor (KNN) with Mahalanobis distance and a K value of 5, (ii) support vector machine with linear kernel (SVM), and (iii) random forest (RF). KNN predicts the class of an unknown data sample based on the class of the "K" nearest samples through a majority voting scheme. SVM identifies a linear hyperplane in the feature space that maximizes the margin between the classes and distinctly classifies the data samples. RF is a robust bagging algorithm that uses an ensemble of decision trees to classify random subsets of the training samples and makes a final classification prediction through majority voting.

The ATL dataset was used to validate the performance of the identified relevant features and ensure transferability. Due to limited number of samples, the ATL dataset could not be used as a completely independent test set. However, both 10-fold stratified cross-validation (SCV) and leave-one-out cross-validation (LOOCV) were utilized to evaluate the performance of the features on the ATL dataset, as both of these methods are appropriate for performance evaluation of smaller datasets. Furthermore, the ATL dataset contains a high imbalance ratio (4 : 1 between negative and positive samples). To mitigate this, SMOTE and hybrid SMOTE-Tomek resampling techniques were applied on the ATL dataset, and the performance of the features with and without sampling was analyzed. SMOTE by itself increased the number of positive (malignant) samples from 26 to 104, to provide a completely balanced scenario. Meanwhile, the SMOTE-Tomek procedure reduced the number of negative samples (benign) from 104 to 93 and increased the number of positive samples from 26 to 93, again providing a completely balanced scenario.

Classification results are evaluated by analyzing the receiver operating characteristic (ROC) curve, in particular the area under the curve (AUC), sensitivity, specificity, and accuracy. AUC is a single scalar value which ranges between 0 and 1 (1 indicating significant performance) representing the predictive performance of a classification task. Accuracy is the ratio of the total number of correct predictions to the total number of instances in a classification task. Sensitivity is a measure of correctly classified positive instances (malignant cases), and specificity is a measure of correctly classified negative instances (benign cases). MATLAB™ (The MathWorks, Inc., Natick, MA) was used to develop all models and evaluate all performance metrics.

3. Results

Table 1 denotes the mean and standard deviation of all features in the OASBUD dataset for benign and malignant cases, as well as the p value and level of statistical significance of the features. Statistical significance is divided into three levels based on p value: not statistically significant ($p \geq 0.05$) indicated by "~," statistically significant ($p < 0.05$) indicated by "∗," and extremely significant ($p < 0.001$) indicated by "∗∗." Table 2 similarly denotes mean and standard deviation feature values for benign and malignant cases in the ATL dataset, as well as statistical significance of the features.

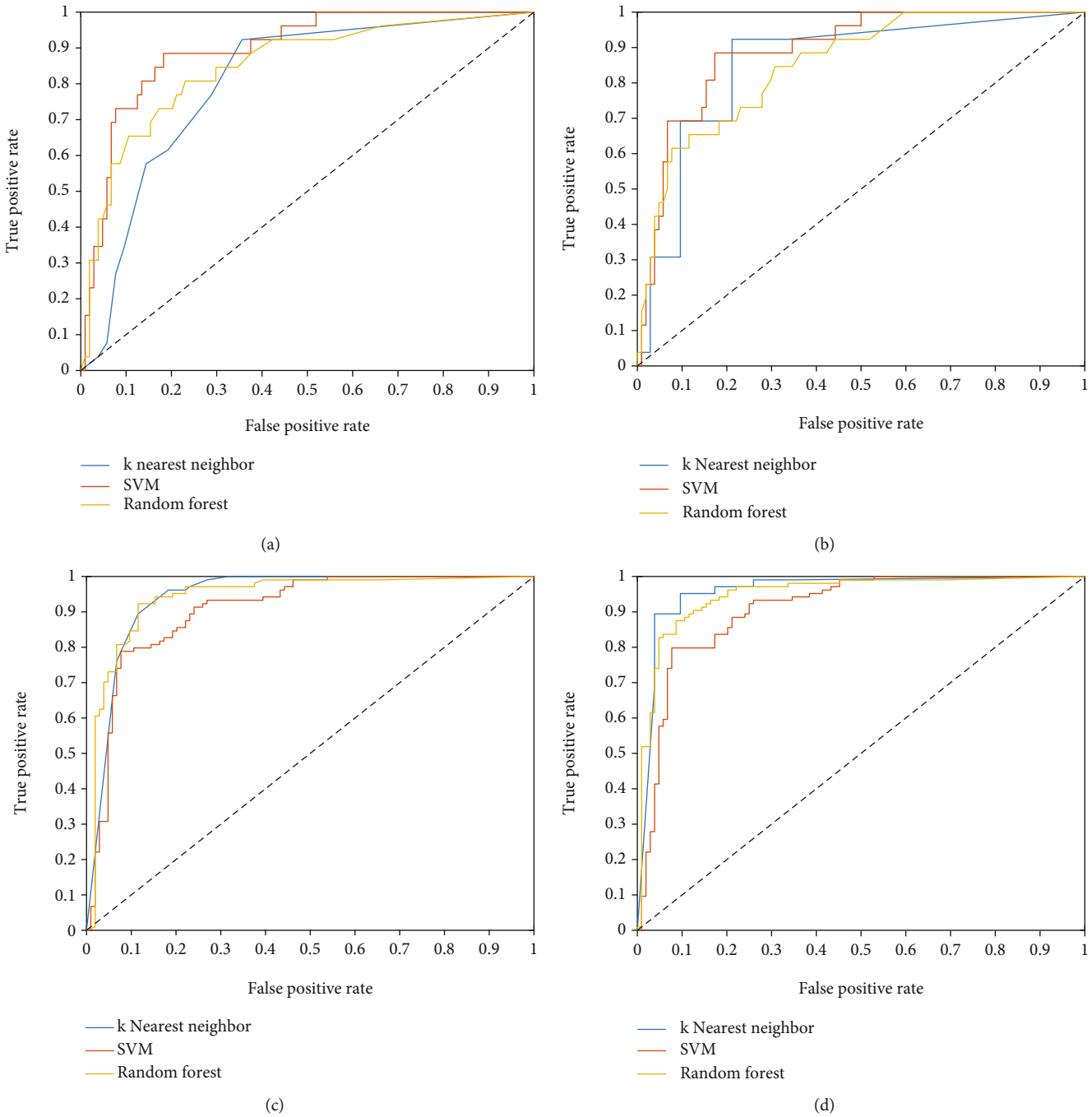


Figure 6: Continued.

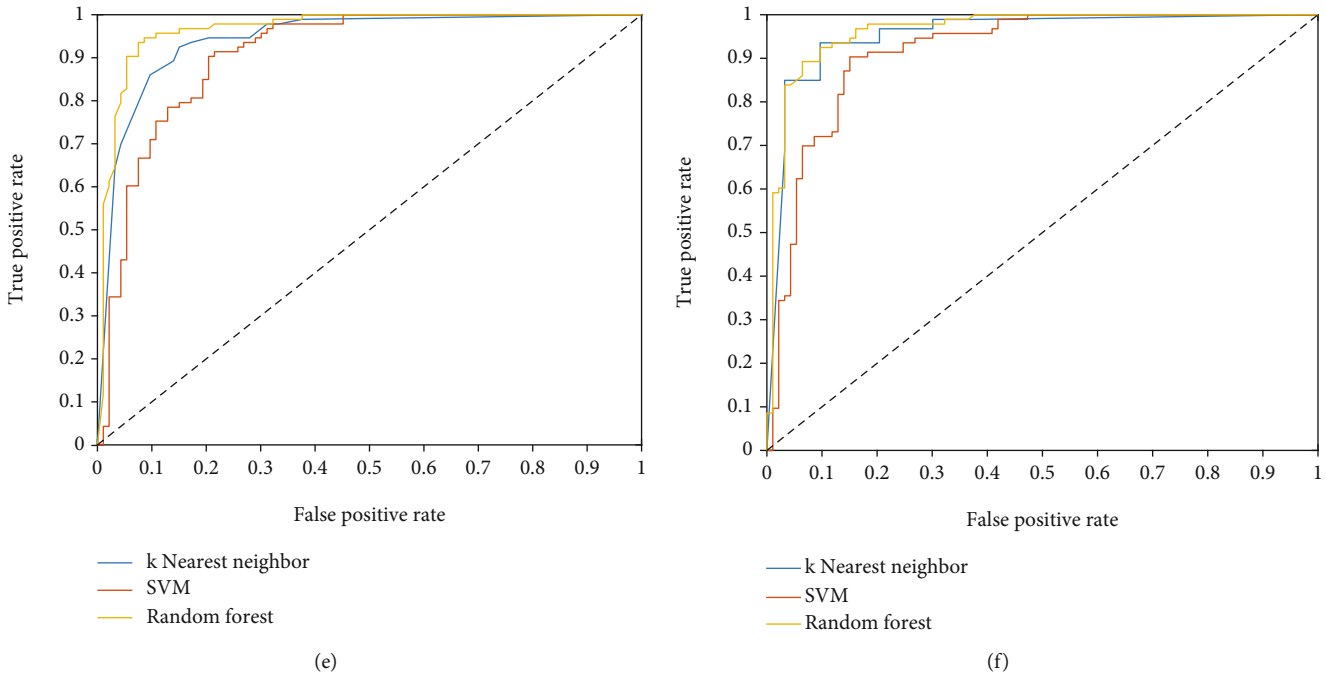


Figure 6: ROC curves obtained by the three classifiers using the 4 selected features in (a) unsampled ATL data with 10-fold SCV, (b) unsampled ATL data with LOOCV, (c) ATL data with SMOTE applied and 10-fold SCV, (d) ATL data with SMOTE applied and LOOCV, (e) ATL data with hybrid SMOTE-Tomek applied and 10-fold SCV, and (f) ATL data with hybrid SMOTE-Tomek applied and LOOCV.

Table 7: Comparison of classification performance of existing multiparametric QUS methods for breast lesion characterization and performance parameters obtained in this study.

	Parameters used	Classification accuracy	Sensitivity	Specificity	AUC
Hsu et al. [55]	Standard deviation of shortest distance (SS), contrast, and Nakagami m	89.4%	92.5%	86.3%	0.96
Klimonda et al. [28]	Contrast, correlation, energy, and homogeneity	91%	93%	88%	0.94
This study	Homodyned k, homodyned h, Nakagami m, and Nakagami Ω	93.01%	94.62%	91.4%	0.966

SFS applied on the training split of the OASBUD dataset identified 4 out of the 16 features as the most significant to breast cancer diagnosis:

(i) k (homodyned K)

(ii) h (homodyned K)

(iii) m (Nakagami)

(iv) Ω (Nakagami)

Figure 3 illustrates the representative box and scatter plots of these four features from the OASBUD dataset.

Table 3 denotes the performance parameters obtained by the three classifiers on the testing portion of the OASBUD dataset using the 4 selected features. Figure 4 illustrates the ROC curves obtained by the three classifiers.

Figure 5 illustrates the representative box and scatter plots of the four selected features from the ATL dataset.

Table 4 denotes the performance parameters obtained by the three classifiers on the unsampled ATL data using the 4 selected features with both 10-fold SCV and LOOCV. Table 5 provides the performance parameters for the ATL dataset after SMOTE was applied, and Table 6 provides the performance parameters after hybrid SMOTE-Tomek was applied.

Figure 6 illustrates the ROC curves obtained by the three classifiers on the unsampled and resampled instances of the ATL dataset using both validation schemes.

4. Discussion

This study proposes a breast tumor classification system using the three major types of intratumoral QUS descriptors. A total of 16 different QUS parameters are extracted from the intratumoral region of breast ultrasound RF scans, consisting of spectral features, envelope statistics features, and

texture features. Sequential forward selection was utilized to identify the most relevant subset of features for breast cancer diagnosis.

Analyzing the statistical significance of each of the 16 features extracted from the OASBUD dataset (Table 1), it can be clearly seen that the envelope statistics features (homodyned K features: k, μ, and h and Nakagami features: m, Ω, and α) are more statistically significant than spectral features or texture features for distinguishing between benign and malignant samples. A similar scenario is observed in Table 2, where the envelope statistics features were found to be more statistically significant than the other types of extracted features for the ATL dataset.

The OASBUD dataset was used to identify the most relevant QUS features for the classification of breast lesions, as the proportion of positive and negative classes is similar. Using a balanced dataset enables feature selection techniques to identify key features that can distinguish between the positive and negative class effectively without bias towards any specific class. All four features selected by the SFS algorithm were related to envelope statistics. Thus, the feature selection algorithm seems to be selecting the most statistically relevant features for breast cancer diagnosis. Specifically, two features were chosen from the homodyned K distribution, and two features were chosen from the Nakagami distribution. Thus, a significant finding of this study is that envelope statistics features are able to segregate between breast lesion types more effectively than the spectral and texture features analyzed in this study. A hypothesis for this may be the fact that envelope statistics are able to describe the subresolutional properties of tissue better than spectral analysis and provide more distinguishing capability than features obtained from analyzing the spatial relationships between pixels in ultrasound envelope images.

Analyzing the performance parameters obtained on the testing portion of the OASBUD dataset using the four selected features (Table 3), it can be observed that all three classifiers obtained similar AUC of around 0.96. In terms of classification accuracy, sensitivity, and specificity, the SVM classifier obtained slightly lower performance than the KNN or RF classifiers. The best performance was clearly obtained using the RF classifier, with a classification accuracy of 95%, sensitivity of 95%, and specificity of 95%.

The ATL dataset was used to validate the performance of the identified relevant features. However, due to the limited number of samples in this study, the ATL dataset could not be used as an independent test set to classify models trained only by the OASBUD dataset. Two validation schemes were utilized to demonstrate that the performance does not suffer from any bias. Both 10-fold SCV and LOOCV are established validation schemes for validation of smaller datasets.

As mentioned before, the ATL dataset contains a high imbalance ratio between positive and negative cases. The impact of this can be observed from the performance parameters provided in Table 4. All three classifiers inadvertently became biased towards the negative class (which represented the majority), as observable by the very low sensitivity values and very high specificity values. For both 10-fold SCV and LOOCV, the KNN classifier provided the poorest performance. The best performance was obtained by the RF classifier using 10-fold SCV, with a moderate sensitivity of 65.38%, accuracy of 85.38%, and AUC of 0.8711.

Application of SMOTE introduced a large number of synthesized positive samples (representing the minority class). This significantly improved performance, particularly in terms of sensitivity (Table 5). The KNN classifier and the RF classifier obtained the highest sensitivity using 10-fold SCV: 94.23% and 92.31%, respectively. However, there was a disparity between the sensitivity and specificity values in these two cases, with both classifiers also correspondingly obtaining lower specificity measures. Thus, applying SMOTE by itself may introduce bias towards the positive minority class, particularly for highly imbalanced cases such as the ATL dataset where a large number of samples need to be synthesized.

To account for this, a hybrid SMOTE-Tomek procedure is utilized, which firstly reduces majority class instances to decrease the imbalance ratio between the classes and then performs oversampling. This approach ensures quality of resampled data, as the number of samples needed to be synthesized is lower. Analyzing Table 6, it can be observed that the disparity between sensitivity and specificity is much lower than those obtained in Table 5, particularly for the two cases discussed above. The best performance was obtained by the RF classifier, with a classification accuracy of 93.01%, sensitivity of 94.62%, specificity of 91.4%, and AUC of 0.9660 obtained using 10-fold SCV and classification accuracy of 91.4%, sensitivity of 93.55%, specificity of 89.25%, and AUC of 0.9640 obtained using LOOCV. Both cases represent significant performance for breast tumor characterization. The results obtained are compared with two recent multiparametric QUS studies for breast cancer in Table 7.

It should be noted that the procedure for acquisition of envelope statistics features differed in this work from other literature. In general, envelope statistics features are estimated by fitting the statistical distribution (i.e., Nakagami or homodyned K) at several small windows spanning the ROI [27, 28, 33]. Following this, the statistical parameters for each distribution (i.e., Nakagami m, Nakagami α, and homodyned k) are estimated at each window, and the final feature value is taken as the average parameter value across all the windows [27, 28, 33]. This methodology reduces impact of signal attenuation at different depths. However, in this study, rather than using windows, the statistical distribution model (both Nakagami and homodyned K) was fit on all samples within the tumor region, and the envelope statistics features were acquired correspondingly from this. This methodology was chosen at it fits the distribution model on a larger pool of samples (i.e., all the samples within the tumor), which ensures a more stable estimation of the statistical parameter for each distribution. However, it does not take into account signal attenuation like the methodology discussed previously, and future studies may analyze the impact of this on breast tumor characterization.

This study has a few limitations. Firstly, it utilizes a limited amount of patient data. Ideally, such a study should utilize a large pool of ultrasound RF data, apply feature

selection on a large training set, and validate performance on a significant testing set. Although two datasets were utilized in this study, they were not mixed. The two datasets were acquired at a difference of about 20 years, and thus, the quality of ultrasound signals in the OASBUD dataset should be far superior to those present in the ATL dataset. This may be a likely cause for the difference in feature values for the two datasets (Tables 1 and 2). Furthermore, a concern with the ATL dataset is the sampling frequency utilized during data collection. Generally, sampling frequency is chosen to be about 4 times higher than the transducer central frequency [56]. The 20 MHz sampling frequency used for a transducer central frequency of 7.5 MHz may lead to loss of information. It should be noted that this condition was met in case of the OASBUD dataset, which used a 40 MHz sampling frequency for a transducer central frequency of 10 MHz. Thus, rather than combining the two datasets physically, the datasets were combined artificially, where the recently acquired OASBUD dataset was used to identify relevant features, and the ATL dataset was used to validate the performance of the identified features. Another limitation of this study is the large imbalance present in the ATL dataset, which necessitates the application of resampling techniques. In an ideal scenario, sampling should not be applied to the test set, as the characteristics of the test set should coincide with medical data available in the real world where imbalance is very prevalent. However, without sampling, the classifiers used in this research become very strongly biased towards the positive majority class and provide poor sensitivity as highlighted in Table 4. This is unacceptable, as correctly identifying malignant cases is of crucial importance. The resampling techniques used in this paper were intended to display that, in a case where the positive and negative classes are fairly balanced, the identified features will be able to distinguish between benign and malignant lesions very effectively. This objective is achieved considering the significant improvement in performance, particularly in terms of sensitivity, after resampling techniques were used to balance the ATL dataset (Tables 5 and 6). Another issue is the undersampling approach that was utilized. The Tomek link technique removes benign samples in the feature space that are close to malignant samples, which may inevitably translate to overly optimistic results. However, in this study, Tomek links was not applied on the ATL dataset by itself, but rather as part of the hybrid SMOTE-Tomek strategy. The purpose of Tomek links in this framework was to act as a data cleaning method and remove overlapping samples created after application of SMOTE, rather than simply removing benign samples that were originally present in the dataset. Such techniques are commonly utilized after application of SMOTE in order to prevent overgeneralization. Next, the spectrum of ultrasonic signals acquired during evaluation of spectral features are not only dependent on tissue properties but also on the two-way transfer function of the transducer and the ultrasonic module (system effects), the beam properties corresponding to the two-way range dependent diffraction function (diffraction effects) and acoustic attenuation [23]. As most lesions analyzed in this study lie at similar depths (2-3 cm), system and diffraction effects will not significantly affect the acquired spectrum analysis parameters, and hence, these effects were not accounted for in this study. However, acoustic attenuation was considered, as it is known to significantly affect SL and MBF values obtained from ultrasound images [23]. Furthermore, this study opted sequential forward selection (SFS) to identify the most relevant texture features, as it is a relatively simple wrapper technique which has been shown to be very effective [57]. Future studies may analyze more robust selection algorithms such as fuzzy rough set-based selection procedures [58] or ensemble selection approaches [59].

5. Conclusion

This study proposes a breast lesion classification system using the three major types of intratumoral QUS descriptors that can be extracted from ultrasound radiofrequency (RF) data. A total of 16 QUS features corresponding to spectral features, envelope statistics features, and textural features were extracted from ultrasound patient data. Four features from envelope statistics were identified as the most significant by feature selection. These four features were able to distinguish between tumor types with a high level of accuracy across two datasets. This demonstrates the capability of the identified features in characterization of benign and malignant breast lesions, and the combination of features identified in this research work has the potential to aid the diagnostic procedure associated with noninvasive screening and diagnosis of breast tumors. The scope of this study can be further enhanced by incorporating more advanced feature selection procedures, incorporating more patient data, and including other types of features in the analysis, for instance more advanced texture features obtained from gray-level run length matrix (GLRLM) and gray-level size zone matrix (GLSZM) techniques, as well as statistical features such as information entropy.

Data Availability

The OASBUD dataset is publicly available via the Zenodo repository (10.5281/zenodo.545928), while the ATL dataset can be obtained through the corresponding author upon reasonable request.

Conflicts of Interest

There are no conflicts of interest.

Acknowledgments

The author expresses gratitude to Dr. S. Kaisar Alam, Centre for Computational Biomedicine Imaging and Modelling, Rutgers University, the State University of New Jersey, for providing one of the datasets and contributing in this research.

References

[1] World Health Organization (WHO), "Breast cancer," 2021, https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[2] H. Sung, J. Ferlay, R. L. Siegel et al., "Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[3] E. Senkus, S. Kyriakides, S. Ohno et al., "Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 26, pp. v8-v30, 2015.

[4] F. J. Gilbert and K. Pinker-Domenig, "Diagnosis and staging of breast cancer: when and how to use mammography, tomosynthesis, ultrasound, contrast-enhanced mammography, and magnetic resonance imaging," in *Diseases of the Chest, Breast, Heart and Vessels 2019-2022: Diagnostic and Interventional Imaging*, J. Hodler, R. A. Kubik-Huch, and G. K. von Schulthess, Eds., Springer, Cham, 2019.

[5] M. K. Feldman, S. Katyal, and M. S. Blackwood, "US Artifacts," *Radiographics*, vol. 29, no. 4, pp. 1179–1189, 2009.

[6] J. L. Jesneck, J. Y. Lo, and J. A. Baker, "Breast mass lesions: computer-aided diagnosis models with mammographic and sonographic descriptors," *Radiology*, vol. 244, no. 2, pp. 390–398, 2007.

[7] W. A. Berg, J. D. Blume, J. B. Cormack, and E. B. Mendelson, "Training the ACRIN 6666 investigators and effects of feedback on breast ultrasound interpretive performance and agreement in BI-RADS ultrasound feature analysis," *American Journal of Roentgenology*, vol. 199, no. 1, pp. 224–235, 2012.

[8] R. F. Brem, M. J. Lenihan, J. Lieberman, and J. Torrente, "Screening breast ultrasound: past, present, and future," *American Journal of Roentgenology*, vol. 204, no. 2, pp. 234–240, 2015.

[9] M. L. Oelze and J. Mamou, "Review of quantitative ultrasound: envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 2, pp. 336–351, 2016.

[10] J. Mamou and M. L. Oelze, Eds., *Quantitative Ultrasound in Soft Tissues*, Springer, Netherlands, Dordrecht, 2013.

[11] E. J. Boote, J. A. Zagzebski, E. L. Madsen, and T. J. Hall, "Instrument-independent acoustic backscatter coefficient imaging," *Ultrasonic Imaging*, vol. 10, no. 2, pp. 121–138, 1988.

[12] L. X. Yao, J. A. Zagzebski, and E. L. Madsen, "Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors," *Ultrasonic Imaging*, vol. 12, no. 1, pp. 58–70, 1990.

[13] R. J. Lavarello, W. R. Ridgway, S. S. Sarwate, and M. L. Oelze, "Characterization of thyroid cancer in mouse models using high-frequency quantitative ultrasound techniques," *Ultrasound in Medicine & Biology*, vol. 39, no. 12, pp. 2333–2341, 2013.

[14] K. C. Balaji, W. R. Fair, E. J. Feleppa et al., "Role of advanced 2 and 3-dimensional ultrasound for detecting prostate cancer," *Journal of Urology*, vol. 168, no. 6, pp. 2422–2425, 2002.

[15] E. J. Feleppa, "Ultrasonic tissue-type imaging of the prostate: implications for biopsy and treatment guidance," *Cancer Biomarkers*, vol. 4, no. 4-5, pp. 201–212, 2008.

[16] T. Noritomi, B. Sigel, V. Swami et al., "Carotid plaque typing by multiple-parameter ultrasonic tissue characterization," *Ultrasound in Medicine & Biology*, vol. 23, no. 5, pp. 643–650, 1997.

[17] F. L. Lizzi, M. Astor, T. Liu, C. Deng, D. J. Coleman, and R. H. Silverman, "Ultrasonic spectrum analysis for tissue assays and therapy evaluation," *International Journal of Imaging Systems and Technology*, vol. 8, no. 1, pp. 3–10, 1997.

[18] F. L. Lizzi, M. Astor, E. J. Feleppa, M. Shao, and A. Kalisz, "Statistical framework for ultrasonic spectral parameter imaging," *Ultrasound in Medicine & Biology*, vol. 23, no. 9, pp. 1371–1382, 1997.

[19] S. K. Alam, E. J. Feleppa, M. Rondeau, A. Kalisz, and B. S. Garra, "Ultrasonic multi-feature analysis procedure for computer-aided diagnosis of solid breast lesions," *Ultrasonic Imaging*, vol. 33, no. 1, pp. 17–38, 2011.

[20] A. Sadeghi-Naini, H. Suraweera, W. T. Tran et al., "Breast-lesion characterization using textural features of quantitative ultrasound parametric maps," *Scientific Reports*, vol. 7, no. 1, p. 13638, 2017.

[21] L. Sannachi, H. Tadayyon, A. Sadeghi-Naini et al., "Non-invasive evaluation of breast cancer response to chemotherapy using quantitative ultrasonic backscatter parameters," *Medical Image Analysis*, vol. 20, no. 1, pp. 224–236, 2015.

[22] A. Sadeghi-Naini, L. Sannachi, H. Tadayyon et al., "Chemotherapy-response monitoring of breast cancer patients using quantitative ultrasound-based intra-tumour heterogeneities," *Scientific Reports*, vol. 7, no. 1, p. 10352, 2017.

[23] P. Mohana Shankar, "A general statistical model for ultrasonic backscattering from tissues," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 47, no. 3, pp. 727–736, 2000.

[24] P. M. Shankar, V. A. Dumane, J. M. Reid et al., "Classification of ultrasonic B-mode images of breast masses using Nakagami distribution," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 2, pp. 569–580, 2001.

[25] P. M. Shankar, V. A. Dumane, C. W. Piccoli, J. M. Reid, F. Forsberg, and B. B. Goldberg, "Classification of breast masses in ultrasonic b-mode images using a compounding technique in the Nakagami distribution domain," *Ultrasound in Medicine & Biology*, vol. 28, no. 10, pp. 1295–1300, 2002.

[26] P. M. Shankar, V. A. Dumane, T. George et al., "Classification of breast masses in ultrasonic B scans using Nakagami and K distributions," *Physics in Medicine and Biology*, vol. 48, no. 14, pp. 2229–2240, 2003.

[27] Z. Klimonda, K. Dobruch-Sobczak, H. Piotrzkowska-Wróblewska, P. Karwat, and J. Litniewski, "Quantitative Ultrasound of Tumor Surrounding Tissue for Enhancement of Breast Cancer Diagnosis," in *Bioinformatics and Biomedical Engineering*, I. Rojas and F. Ortuño, Eds., Springer, Cham, 1st edition, 2018.

[28] Z. Klimonda, P. Karwat, K. Dobruch-Sobczak, H. Piotrzkowska-Wróblewska, and J. Litniewski, "Breast-lesions characterization using quantitative ultrasound features of peritumoral tissue," *Scientific Reports*, vol. 9, no. 1, p. 7963, 2019.

[29] V. Dutt and J. F. Greenleaf, "Ultrasound echo envelope analysis using a homodyned K distribution signal model," *Ultrasonic Imaging*, vol. 16, no. 4, pp. 265–287, 1994.

[30] D. P. Hruska, *Improved Techniques for Statistical Analysis of the Envelope of Backscattered Ultrasound Using the Homodyned K Distribution, [M.S. thesis]*, Dept. Elect. Comput. Eng., Univ. Illinois at Urbana-Champaign, Urbana, IL, USA, 2009.

[31] D. P. Hruska and M. L. Oelze, "Improved parameter estimates based on the homodyned K distribution," *Ferroelectrics and Frequency Control*, vol. 56, no. 11, pp. 2471–2481, 2009.

[32] I. Trop, F. Destrempes, M. el Khoury et al., "The added value of statistical modeling of backscatter properties in the management of breast lesions at US," *Radiology*, vol. 275, no. 3, pp. 666–674, 2015.

[33] F. Destrempes, I. Trop, L. Allard et al., "Added value of quantitative ultrasound and machine learning in BI-RADS 4-5 assessment of solid breast lesions," *Ultrasound in Medicine & Biology*, vol. 46, no. 2, pp. 436–444, 2020.

[34] F. Davnall, C. S. P. Yip, G. Ljungqvist et al., "Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?," *Insights Into Imaging*, vol. 3, no. 6, pp. 573–589, 2012.

[35] A. Heindl, S. Nawaz, and Y. Yuan, "Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology," *Laboratory Investigation*, vol. 95, no. 4, pp. 377–384, 2015.

[36] J. P. B. O'Connor, C. J. Rose, J. C. Waterton, R. A. D. Carano, G. J. M. Parker, and A. Jackson, "Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome," *Clinical Cancer Research*, vol. 21, no. 2, pp. 249–257, 2015.

[37] D. Sengupta and G. Pratx, "Imaging metabolic heterogeneity in cancer," *Molecular Cancer*, vol. 15, no. 1, p. 4, 2016.

[38] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[39] B. S. Garra, B. H. Krasner, S. C. Horii, S. Ascher, S. K. Mun, and R. K. Zeman, "Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis," *Ultrasonic Imaging*, vol. 15, no. 4, pp. 267–285, 1993.

[40] Y. Y. Liao, P. H. Tsui, C. H. Li et al., "Classification of scattering media within benign and malignant breast tumors based on ultrasound texture-feature-based and Nakagami-parameter images," *Medical Physics*, vol. 38, no. 4, pp. 2198–2207, 2011.

[41] W. Gomez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1889–1899, 2012.

[42] A. V. Alvarenga, W. C. A. Pereira, A. F. C. Infantosi, and C. M. Azevedo, "Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images," *Medical Physics*, vol. 34, no. 2, pp. 379–387, 2007.

[43] H. Piotrzkowska-Wróblewska, K. Dobruch-Sobczak, M. Byra, and A. Nowicki, "Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions," *Medical Physics*, vol. 44, no. 11, pp. 6105–6109, 2017.

[44] F. L. Lizzi, M. Greenebaum, E. J. Feleppa, M. Elbaum, and D. J. Coleman, "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," *The Journal of the Acoustical Society of America*, vol. 73, no. 4, pp. 1366–1373, 1983.

[45] E. J. Feleppa, F. L. Lizzi, D. J. Coleman, and M. M. Yaremko, "Diagnostic spectrum analysis in ophthalmology: a physical perspective," *Ultrasound in Medicine & Biology*, vol. 12, no. 8, pp. 623–631, 1986.

[46] J. C. Bamber, "Ultrasonic properties of tissue," in *Ultrasound in Medicine*, F. A. Duck, A. C. Baker, and H. C. Starritt, Eds., pp. 57–88, Institute of Physics (IOP) Publishing, Bristol, 1998.

[47] T. D. Mast, "Empirical relationships between acoustic parameters in human soft tissues," *Acoustics Research Letters Online*, vol. 1, no. 2, pp. 37–42, 2000.

[48] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," *IEEE Transactions on Sonics and Ultrasonics*, vol. 30, no. 3, pp. 156–163, 1983.

[49] F. Destrempes and G. Cloutier, "A critical review and uniformized representation of statistical distributions modeling the ultrasound echo envelope," *Ultrasound in Medicine & Biology*, vol. 36, no. 7, pp. 1037–1051, 2010.

[50] F. Destrempes, E. Franceschini, F. T. H. Yu, and G. Cloutier, "Unifying concepts of statistical and spectral quantitative ultrasound techniques," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 488–500, 2016.

[51] M. Nakagami, *The m-Distribution A General Formula of Intensity Distribution of Rapid Fading, in: Statistical Methods in Radio Wave Propagation*, Elsevier, 1960.

[52] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[53] I. Tomek, "Two modifications of CNN," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, pp. 769–772, 1976.

[54] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," *Journal of Biomedical Informatics*, vol. 90, p. 103089, 2019.

[55] S. M. Hsu, W. H. Kuo, F. C. Kuo, and Y. Y. Liao, "Breast tumor classification using different features of quantitative ultrasound parametric images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 4, pp. 623–633, 2019.

[56] L. Svilainis and V. Dumbrava, "Ultrasonic data acquisition: sampling frequency versus bandwidth," *Ultragarsas*, vol. 29, pp. 29–33, 1998.

[57] A. Newaz and S. Muhtadi, "Performance improvement of heart disease prediction by identifying optimal feature sets using feature selection technique," *2021 International Conference on Information Technology (ICIT)*, 2021, Amman, Jordan, 2021.

[58] X. Zhang, C. Mei, D. Chen, and J. Li, "Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy," *Pattern Recognition*, vol. 56, pp. 1–15, 2016.

[59] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: homogeneous and heterogeneous approaches," *Knowledge-Based Systems*, vol. 118, 2017.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 7631271, 11 pages
https://doi.org/10.1155/2022/7631271

Research Article

Intelligent Diagnosis Method for New Diseases Based on Fuzzy SVM Incremental Learning

Shi Song-men

China Pharmaceutical University, Nanjing 211198, China

Correspondence should be addressed to Shi Song-men; mss_98@sohu.com

Received 28 September 2021; Revised 24 November 2021; Accepted 13 December 2021; Published 13 January 2022

Academic Editor: Giovanni D Addio

The diagnosis of new diseases is a challenging problem. In the early stage of the emergence of new diseases, there are few case samples; this may lead to the low accuracy of intelligent diagnosis. Because of the advantages of support vector machine (SVM) in dealing with small sample problems, it is selected for the intelligent diagnosis method. The standard SVM diagnosis model updating needs to retrain all samples. It costs huge storage and calculation costs and is difficult to adapt to the changing reality. In order to solve this problem, this paper proposes a new disease diagnosis method based on Fuzzy SVM incremental learning. According to SVM theory, the support vector set and boundary sample set related to the SVM diagnosis model are extracted. Only these sample sets are considered in incremental learning to ensure the accuracy and reduce the cost of calculation and storage. To reduce the impact of noise points caused by the reduction of training samples, FSVM is used to update the diagnosis model, and the generalization is improved. The simulation results on the banana dataset show that the proposed method can improve the classification accuracy from 86.4% to 90.4%. Finally, the method is applied in COVID-19's diagnostic. The diagnostic accuracy reaches 98.2% as the traditional SVM only gets 84%. With the increase of the number of case samples, the model is updated. When the training samples increase to 400, the number of samples participating in training is only 77; the amount of calculation of the updated model is small.

1. Introduction

The acceleration of the pace of modern life and the aggravation of the pollution of air, water, and other living resources cause the increase of incidence disease rate [1]. Although the construction of medical conditions has made great progress, it is still stretched in the face of such a large population base. Medical staff and patients are facing great pressure.

With the development of artificial intelligence technology, it has been widely used in various fields and achieved very good results [2–5]. In the medical field, intelligent diagnosis and treatment have become a powerful tool and a hot spot [6, 7]. Machine learning, with its powerful data processing and mining ability, has become the main research direction of intelligent diagnosis and treatment: neural network, Bayesian network, random forest, support vector machine, and other methods have been applied to the exploration of this problem [8–12]. Particularly with the advent of the era of big data, the deep learning method [13] shows strong advantages.

Although these methods are effective, high-precision, the treatment often needs a lot of data support. At present, there are many channels for data acquisition. However, people often ignore a problem: the diagnosis of new diseases, especially those with strong infectivity. Coronavirus disease 2019 (COVID-19), which broke out in December 2019, is a very typical example of virus pneumonia. In the early stage of new crown, if the disease cases can be diagnosed quickly and accurately, the difficulty and cost of disease control will be greatly reduced. Many artificial intelligence methods have been adopted to help diagnosis [14–19]. Reference [20] proposed a deep migration learning method based on DenseNet201 to judge whether the patient is infected with COVID-19. A convolutional neural network model based on a multitask learning model was proposed in Reference [21] to realize COVID-19 detection and refinement of patient severity. Wang and Wong [22] proposed a CNN network model based on ResNet (COVID net). The model predicts normal, bacterial infection, non-COVID-19 viral

infection, and COVID-19 viral infection, the accuracy is higher than 80%, and the computational complexity is less than 250 million times of multiplication and addition. Narin et al. propose three different deep learning models based on ResNet50, Inception V3, and Inception-ResNet v2 [23] to detect COVID-19 from X-ray images. All these studies have achieved high diagnostic accuracy, but they are based on large sample conditions. However, the initial case samples are very few. In this paper, we analyze this problem and study the intelligent diagnosis in the early stage of new diseases with few case samples.

This problem faces two challenges: (1) there are few sample data; (2) after the disease develops, the new case samples are added and the diagnostic model needs to be updated. In the machine learning method, SVM can deal with the classification problem well under the condition of small samples. Therefore, this paper selects the SVM method to solve this problem and learns the newly collected case samples through incremental learning, constantly updates and improves the diagnostic model, and improves the diagnostic ability [11, 12]. However, every time the diagnostic model of standard SVM is updated, all samples need to be retrained, which costs a lot of storage and calculation. To solve this problem, many scholars have proposed some SVM incremental learning methods. These methods mainly include three ideas: support vector, Karush Kuhn Tucker (KKT) condition, and the geometric features [24–29]. From the perspective of support vector idea, only support vectors have an impact on the solution [25, 26], so we only need to pay attention to the support vector. From the perspective of geometric features [28, 29], all the support vectors are at the boundary of the classification hyperplane. On this basis, in this paper, we intend to find out the support vector set and boundary sample set related to the SVM model, abandon most samples, ensure the classification accuracy of the model, and reduce the cost of calculation and storage.

In actual cases, some disease symptoms are similar, which brings difficulties to diagnosis. These features often exist as noise points. SVM adopts the same punishment method for all data points in the training process, which makes the training model more sensitive to noise and outliers. This situation will be more obvious when the number of samples is relatively small. In the incremental learning method we intend to adopt, most of the samples will be omitted in the sample updating process. In this case, if the traditional SVM training diagnosis model is still used, once some noise points or outliers with large deviation appear in the new samples, it may lead to a large deviation of the classification hyperplane, resulting in the possibility of a significant decline in the diagnosis effect. In this case, it is necessary to reduce the sensitivity of noise points and outliers. Lin and Wang proposed fuzzy support vector machine (FSVM) by introducing fuzzy membership function into standard SVM [30]. By giving different penalties to different samples, the influence of these points is weakened and the classification accuracy is improved.

To sum up, to solve the limited number of samples in the initial stage of new diseases, SVM is adopted. As the new samples are collected, this paper updates the diagnosis model in real time through the incremental learning method of SVM. At the same time, in order to reduce the impact of noise points on the model, the fuzzy membership function is introduced. It is hoped that these methods can improve the accurate diagnosis of new diseases and improve the diagnostic accuracy.

The rest of this paper is organized as follows: Section 2 introduces the SVM incremental learning method; the sample updating and the calculation method of fuzzy membership are proposed. In Section 3, the effectiveness of the proposed algorithm is verified by the banana dataset. In Section 4, an intelligent diagnostic application analysis was conducted using COVID-19 data; we discuss and compare the outcomes by experiment and analysis. Finally, conclusions are drawn and future directions are discussed in Section 5.

2. SVM Incremental Learning Method

The key of intelligent diagnosis is classification. The SVM method has the advantages of fast solution speed and strong generalization ability in solving small sample, nonlinear, and high-dimensional problems [31]. This is in line with the data characteristics in the early stage of new diseases. In this paper, SVM is selected as the diagnosis algorithm to realize the accurate diagnosis in the early stage of new diseases.

2.1. The Classification Principle of SVM. Assumes that the sample space of the case is $S = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in R^n, y_i = \pm 1, i = 1, \cdots, l\}$, where $\mathbf{x}_i$ is the feature vector of the disease and y_i is the corresponding state identification value (1 represents the target disease, and -1 represents the nontarget disease). SVM classifies disease diagnosis into convex quadratic programming shown in

$$\begin{aligned} \min \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \cdots, l, \end{aligned} \tag{1}$$

where $\mathbf{w}$ is the normal vector corresponding to the optimal classification hyperplane and C is the penalty factor. The greater the C value, the greater the penalty for misclassification samples. ξ is a relaxation variable, which represents the distance from the sample points between the classification boundaries to the respective classification boundaries. The dual problem of equation (1) is shown in

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad & \sum_{i=1}^{l} y_i\alpha_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2\cdots, l, \end{aligned} \tag{2}$$

where α_i is the Lagrange multiplier. When $(\mathbf{x}_i, y_i)$ satisfies the Karush Kuhn Tucker condition given in equation (3), the corresponding optimal solution of equation (2) is

$$\begin{cases} \alpha_i = 0 \Rightarrow y_i f(\mathbf{x}_i) \geq 1, \\ 0 < \alpha_i < C \Rightarrow y_i f(\mathbf{x}_i) = 1, \\ \alpha_i = C \Rightarrow y_i f(\mathbf{x}_i) \leq 1. \end{cases} \tag{3}$$

The samples that violate the KKT condition (corresponding $\alpha_i \neq 0$) constitute the SV set, and the diagnostic model trained by the full sample set is shown in

$$\mathbf{y} = \operatorname{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b\right]. \tag{4}$$

From the discrimination results of equation (4), we can see that the SVM-based diagnosis model only relates to the support vector set (SV set). The SV set is equivalent to the complete set.

2.2. Training Sample Set Update. Through the analysis above, we can get that for the SVM diagnosis method, the model is only related to the SV set. This is also applicable in the sample updating process of incremental learning. In the training process, we can simplify the updating process as long as we find the SV set in advance. For the SVM classification, the intuitive geometric interpretation is shown in Figure 1.

In Figure 1, solid dots and hollow dots represent two types of samples, respectively; H is the optimal classification hyperplane; H_1 and H_2 are the classification boundary hyperplanes parallel to H and passing through the nearest samples in two classes. *Margin* is the interval between classification boundaries. The positions of the SV set are mainly concentrated on the classification boundary and between the two classification boundaries, that is, the points identified by the red "o" in the figure.

Since only SV sets contribute to the classification hyperplane, the non-SV samples should be deleted during model update, which can reduce the amount of computation. For the newly added samples, if all samples are outside the classification boundary, meaning all the new samples are non-SV, they have no contribution to the diagnostic model. The added samples between two classification boundaries are usually new SVs. Due to these new samples, the previous classification boundary will be deflected, which will make some original non-SV_S transform into SV_S [28]. According to the geometric distribution of SV, the samples that may be transformed to SV are usually distributed near the classification boundary. Therefore, when updating the model, we need to take these sample points into account in addition to the original SV set. In this way, we can divide the updated sample set into three parts: the new sample set S_n, the original SV set, and the sample set S_h near the classification boundary.

The sample set near the classification boundary:

$$S_h = \left\{ (\mathbf{x}_i, y_i) \left| \begin{array}{l} \mathbf{x}_i \in R^n, y_i = \pm 1, \\ 1 < \sum_{i=1}^{l} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b < c, i = 1, \cdots, l \end{array} \right. \right\} \tag{5}$$

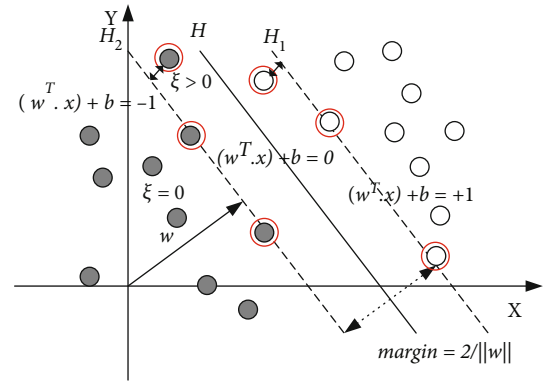


Figure 1: SVM classification diagram.

where c is a constant; the number of selected samples of S_h can be adjusted through the setting of c. The smaller c is, the smaller size of the updated S_h. It can obtain faster training speed and simpler diagnostic model. However, it may lead to the loss of key information and reduce the diagnostic accuracy. On the contrary, the larger the size of the updated S_h, the higher the accuracy will be obtained. For incremental learning of model update, the update speed is determined by the number of training samples involved in the update. The smaller the number of training samples, the faster the update speed. Here, we further cut the updated sample set: in the new sample set, we use equation (5) to find the boundary samples. When updating the diagnostic model, only three sets of S_n, SV, and S_h need to be retrained.

2.3. Fuzzy Support Vector Machine. FSVM first assigns membership values to the samples in the training set according to their importance in classification. The training sample set after evaluation is $S' = \{(\mathbf{x}_1, y_1, \mu_1), (\mathbf{x}_2, y_2, \mu_2), \cdots, (\mathbf{x}_l, y_l, \mu_l)\}$. $\mu_i \in [\varepsilon, 1]$ is the fuzzy membership of $(\mathbf{x}_i, y_i)$; ε is a sufficiently small positive number. Training S' with SVM, the optimization problem in equation (1) transforms into the optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} \mu_i \xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, 2, \cdots, l. \end{aligned} \tag{6}$$

Compared with the standard SVM, FSVM uses weighted error measurement $\mu_i \xi_i$ to reduce the ξ_i in classification to a certain extent. As the outliers and noise in the sample are often within the classification boundary and have large relaxation variable values, FSVM weakens the influence of outliers and noise by assigning them small fuzzy membership, so as to avoid overfitting and improve the generalization of the diagnostic model. The core problem of FSVM is to give different fuzzy membership degrees to different sample points. In this paper, the fuzzy membership is determined by the relationship between points, classification hyperplane, and classification boundary. The samples are

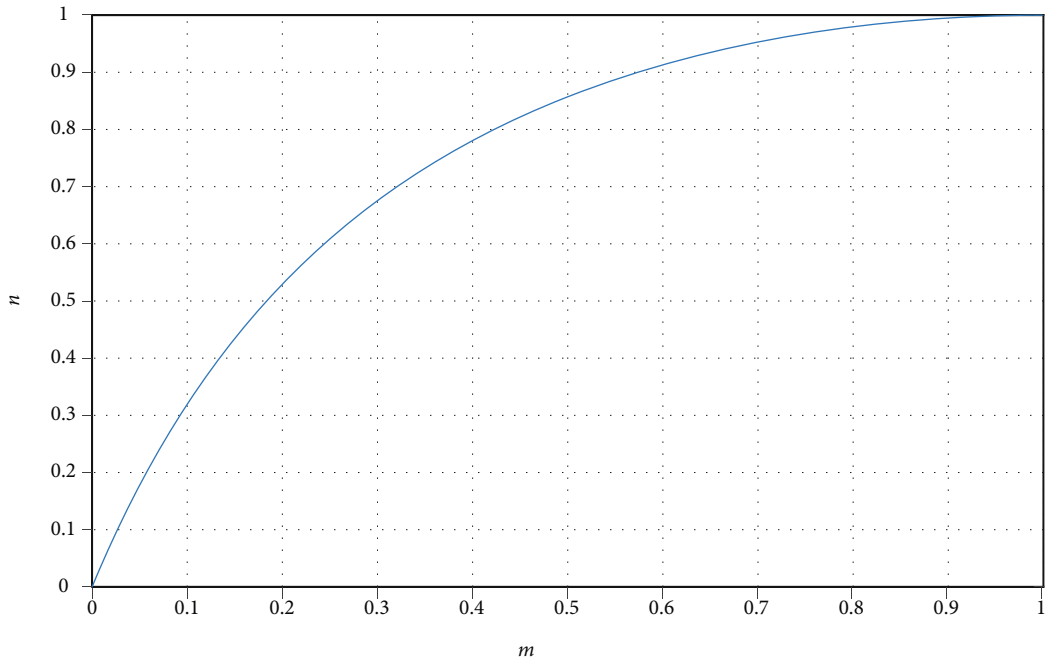


Figure 2: Fuzzy membership function.

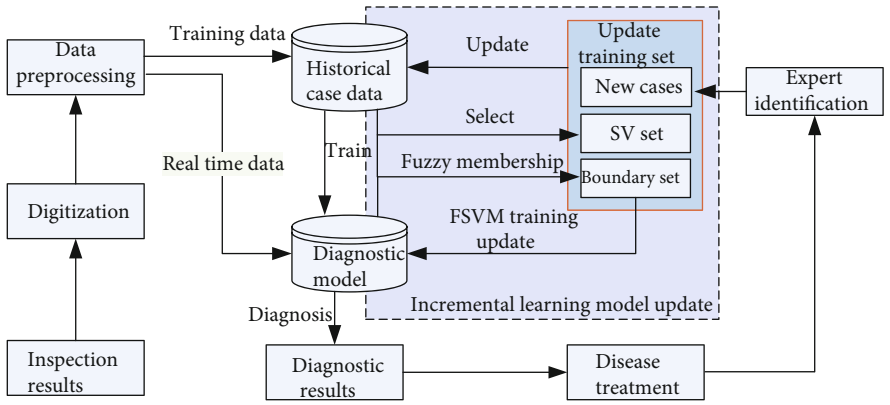


Figure 3: Intelligent diagnosis process based on FSVM incremental learning.

divided into inside boundaries and outside boundaries. The samples outside the boundary can be considered as determined sample, and the fuzzy membership degree is set 1. If the sample points are misclassified, we give its fuzzy membership a very small value ε (here $\varepsilon = 0.0001$). The correctly classified samples between the boundaries are the samples that we should focus on. Consider the interval between the sample points and the optimal classification hyperplane. The farther the interval, the greater the probability that they belong to this category. The interval of the optimal classification hyperplane of sample points m can be expressed as

$$m = \left| \sum_{i=1}^{l} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \right|. \tag{7}$$

The final fuzzy membership function is

$$\mu = \left| \left(\frac{2}{(m-1)^2} - 1 \right)^{-1} - 1 \right|. \tag{8}$$

The function diagram of fuzzy membership function is shown in Figure 2.

The fuzzy membership function is constructed by the generalized bell membership function model. When m changes from 0 to 1, the first half u increases rapidly because it is close to the optimal classification hyperplane and away from the sample class; the latter half is close to the classification boundary, and the increase of value u tends to be gentle.

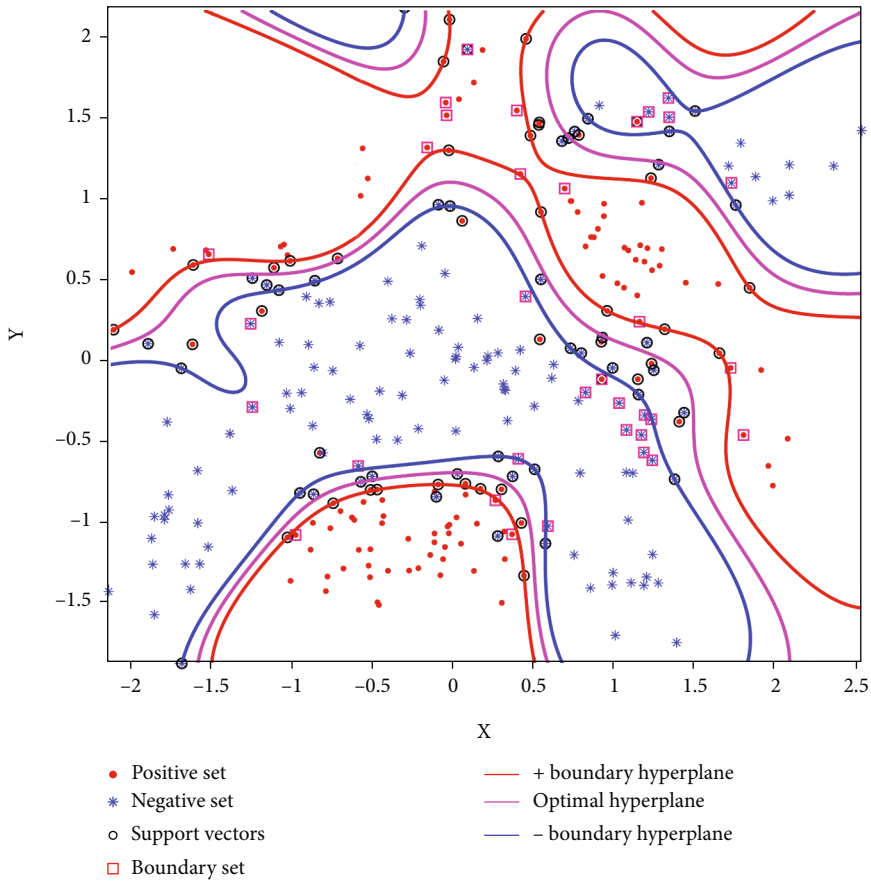


Figure 4: The classification result of the banana dataset.

2.4. Diagnostic Process. The processing process of the intelligent diagnosis method based on FSVM incremental learning is shown in Figure 3.

Firstly, the initial model is established by the previously collected historical case database and is used to judge whether it is the target disease and give specific diagnosis and treatment suggestions; according to the recovery of patients, the misdiagnosed and missed cases in the initial diagnosis results are analyzed. After being identified by experts, they are input into the historical sample database as incremental samples. When the incremental samples accumulate to a certain number, the model update program is triggered; in the model updating stage, the SV set and boundary sample set are extracted from the historical sample database according to the diagnosis results of historical samples, and they are added together with the boundary samples and boundary samples in the new sample set as a new training sample set, and a new diagnosis model is obtained by giving different fuzzy membership degrees to different samples for FSVM training. The intelligent diagnosis system incorporated into the model update forms a closed-loop self-learning system, which is conducive to the continuous correction and improvement of the diagnosis model and enhances the SVM's ability to diagnose new diseases.

3. Algorithm Verification

In order to verify the effectiveness of the proposed method in this paper, we select the typical two-dimensional nonlinear separable dataset banana dataset in benchmark dataset [32] and verify the performance of the algorithm by updating the classification samples of the dataset and analyzing the classification results. The experimental environment is Xeon (R) 3.3G CPU, 8G memory, Windows 7 system, and MATLAB 2018b. RBF kernel function is selected for SVM training. Its parameters are determined by cross-validation and grid search method and c set to 1.5.

Firstly, we analyze the sample update results of the algorithm. Figure 4 shows the classification of the banana dataset (the initial training set includes 150 sample points of positive and negative classes) under initial training. The purple, green, and yellow contours in Figure 4 represent the positive class sample classification boundary, the optimal classification hyperplane, and the negative class classification boundary, respectively. "o" represents the set of support vectors, and "□" represents the set of boundary samples found. From the figure, we can see that the support vector sets are distributed within and on the classification boundary, and the boundary sample sets are near the classification boundary. This verifies the previous analysis is correct. These samples contain all the classification boundary information.

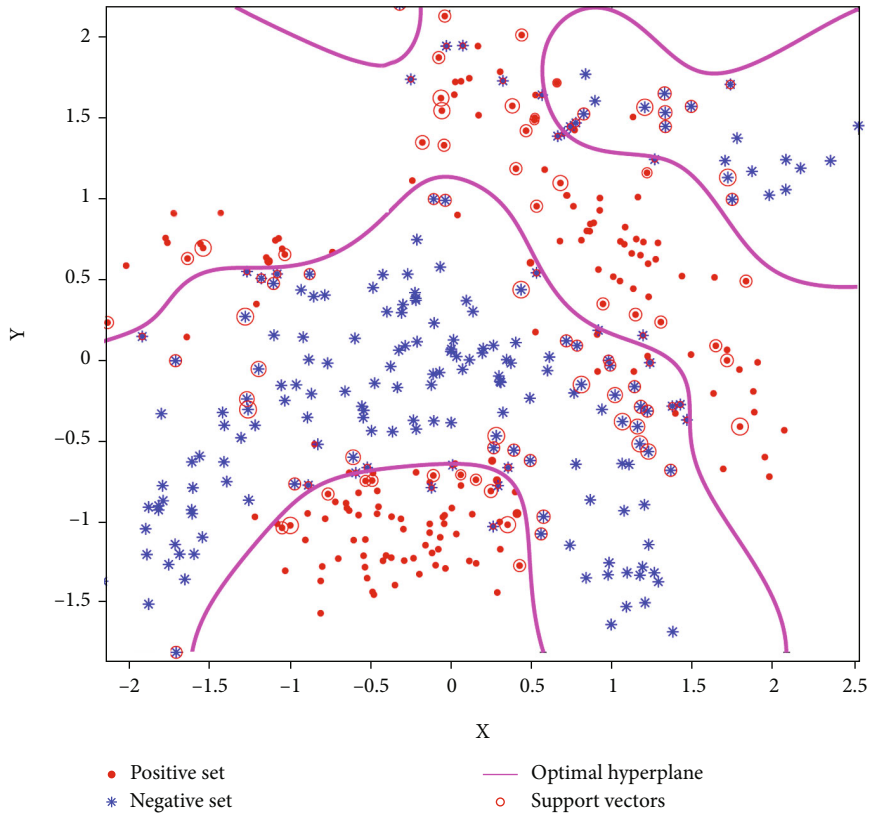


Figure 5: Updated set and fuzzy membership.

Next, the rationality of the updated fuzzy membership function is verified. We added 50 positive and negative samples, respectively. Figure 5 shows the updated dataset and their fuzzy membership. The red "o" point in the figure is all the updated sample sets (including the boundary sample set found earlier, support vector set, and new sample set after clipping). The size of "o" represents the value of fuzzy membership. It should be noted that for better display effect, the minimum size of "o" is set to the size of 8 labels in MATLAB. It can be clearly seen from the experimental results that the sample point "o" at the classification boundary is the largest. The closer the point between the two classification boundaries is to the optimal classification hyperplane, the smaller its value. In this way, their influence on the classification model can be reduced during training, and the misclassified sample points are the smallest, which means that they can be almost ignored during training. In this way, FSVM can improve the generalization of the classification model.

Finally, we analyze the classification performance of the updated classification model after adding new samples. Compare the nonupdated classification hyperplane, the SVM incremental learning updated model, and the classification model by adding fuzzy membership. Figure 6 shows the classification results of the three methods. From the results, we can see that the classification model has changed after incremental learning. This is mainly because the addition of new training samples affects the original classification hyperplane after incremental learning. The newly obtained classification hyperplane is more accurate than the original classification hyperplane. Compared with the classification hyperplane obtained by the two update methods, FSVM can effectively reduce the impact of noise points and outliers by giving different penalty coefficients to different training samples. For example, as can be seen from the local amplification part in Figure 6, the optimal classification hyperplane is biased to the right due to the influence of two ".". The FSVM can effectively modify the classification hyperplane and improve generalization.

Further, we continuously add the sample set to the training set, with each increase of 100 samples (positive class 50, negative class 50). The updated model is tested on banana_test_2. The specific classification accuracy results are shown in Table 1.

From the experimental results in Table 1, we can see that if there is no incremental learning, the accuracy of the classification model is only 86.4%. However, through incremental SVM learning, the model is continuously optimized with the update of the sample set, and finally, the classification accuracy of 89.8% is achieved. Through the introduction of fuzzy factor, the generalization of the model is further improved, and the classification accuracy can reach 90.4%. In the process of model updating, the size of training set will affect the training time and the timeliness of model updating. The proposed method in this paper can filter the initial samples and only select the support vector set and boundary sample set; the number of samples participating in the update is very small. So the update speed of the model is also relatively fast.

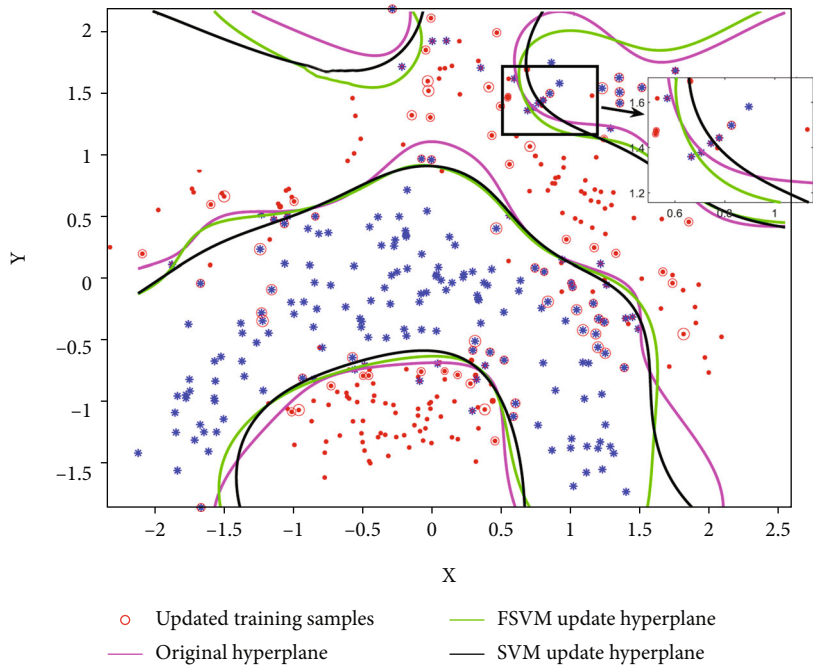


Figure 6: Results of FSVM incremental learning classification.

Table 1: Comparison of classification accuracy of the banana dataset.

Training set		Test set	Classification accuracy (%)	
			SVM	FSVM
Initial set	Positive 150, negative 150		86.4	
Increment 1	Positive 50, negative 50		88.2	88.4
Increment 2	Positive 50, negative 50	Banana_test_2	88.6	89.1
Increment 3	Positive 50, negative 50		89.4	90.2
Increment 4	Positive 50, negative 50		89.8	90.4

4. Application Analysis of Intelligent Diagnosis

In order to verify the effectiveness of the proposed algorithm, an intelligent diagnostic application analysis was conducted using COVID-19 data provided by our affiliated hospital. The dataset includes two types of samples: COVID-19 and non-COVID-19. The total number of samples was 571, including 357 non-COVID-19 and 212 COVID-19.

The sample includes a total of 37 features. The first two features are patient ID and category. Excluding these two-dimensional features, the remaining 35 dimensions are the features we use for diagnosis. Overall, it includes physiological features, biochemical examination results, and CT image characteristics. The physiological features include ambulatory blood pressure, pulmonary hypertension, heart rate, SpO_2, body temperature, and respiratory rate. The biochemical examination mainly includes the following features: the number of white blood cells, the percentage of lymphocytes, creatine kinase, alanine aminotransferase, aspartate aminotransferase, high-sensitivity C-reactive protein, and erythrocyte sedimentation rate. The CT image features mainly include density, shape, lesion distribution, interstitial thickening, thickening of vascular bundle in the lesion, cord focus, and pleural effusion. From the overall data, the ratio of non-COVID-19 and COVID-19 is about 6 : 4. We assume that 100 groups of cases were collected in the initial stage, of which 60 groups are non-COVID-19 and 40 groups are COVID-19. The selection method is random. Incremental learning is performed every 100 samples. Finally, the remaining 171 groups were used as the sample set for the test. The experimental results are shown in Table 2.

From the experimental results, we can see that the diagnostic accuracy is gradually improved with the increase and improvement of the sample set, which is similar to the previous experiments on the banana dataset. In the initial sample set, there are only 100 cases; the diagnostic accuracy of the diagnostic model has reached 84.0%, which reflects the advantages of the SVM method in dealing with small samples. The classification accuracy of the same data trained by the BP neural network is only 74.6%, which is far lower than that of the SVM method. If incremental learning is not used to update the sample set and diagnostic model, the diagnostic accuracy of such model is far from enough,

Table 2: Diagnosis results.

Training set		Test set	Diagnostic accuracy(%)	
			SVM	FSVM
Initial set	Non 60, COVID-19 40	Non 118, COVID-19 53	84.0	
Incremental 1	Non 57, COVID-19 43		87.6	88.8
Incremental 2	Non 64, COVID-19 36		91.1	93.4
Incremental 3	Non 63, COVID-19 37		95.9	98.2

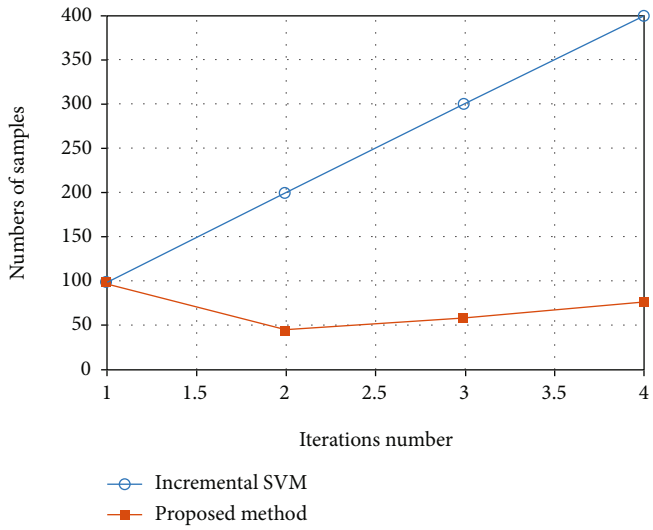


Figure 7: Size of FSVM incremental training sample.

which means that more than 15% of patients will be misdiagnosed. When only the SVM method is used to update the model, the diagnostic accuracy has been greatly improved due to the new case samples, and the final diagnostic accuracy can reach 95.9%. However, the existence of wild points and noise points may reduce the generalization of the model. Through the fuzzy processing of these sample points, FSVM can effectively reduce their impact on the classification hyperplane of the model and improve the generalization. The experimental results show that the incremental learning model processed by FSVM can improve the diagnosis accuracy to 98.2% and further verify the advantages of FSVM incremental learning.

Another key problem to be considered in incremental learning is the update speed. With the increase of samples, if the update speed is too slow and does not have real-time performance, the whole incremental learning method will not be applied to practice. For SVM learning, the speed of training depends on the number of samples participating in training. If the updated training samples are not clipping, the storage and calculation cost of the system will get higher. The method proposed in this paper discards the useless sample points on the basis of the previous model. Figure 7 shows the number of training samples trained by SVM without clipping and the number of training samples updated by the method in this paper. From the figure, we can see that without clipping, the number of samples gradually increases with the change of model update iteration and has reached 400 by the fourth update. The proposed method makes necessary selection every time, and the number of sample points is relatively stable. Even in the fourth generation, the number of samples participating in training is only 77. It can be seen that the method in this paper can greatly reduce the amount of calculation and ensure the efficiency of updating.

Table 3: Confusion matrix of diagnosis problem.

Predictive	Actual	
	COVID-19	Non-COVID-19
COVID-19	TP	FP
Non-COVID-19	FN	TN

Next, we analyze the accuracy of the overall diagnosis of the model. Here, the confusion matrix is chosen to judge the binary classification problem. The confusion matrix is shown in Table 3.

According to Table 3, the recall rate and accuracy rate of the diagnostic model can be calculated as follows:

$$\begin{aligned} \text{recall rate} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{accuracy rate} &= \frac{\text{TP}}{\text{TP} + \text{FP}}. \end{aligned} \tag{9}$$

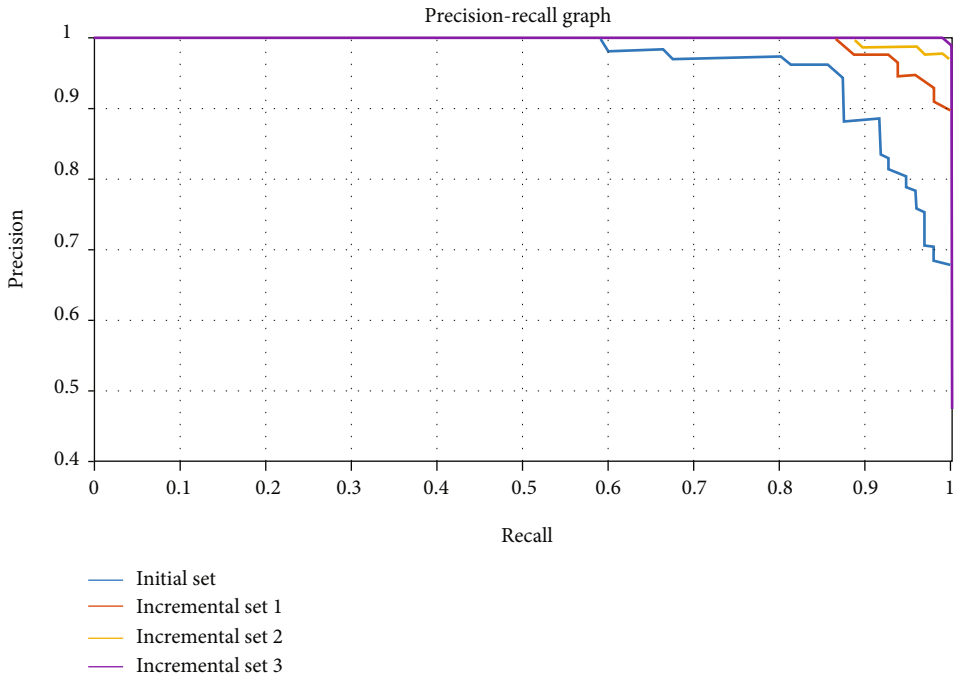


Figure 8: The PR curve of FSVM incremental learning diagnosis.

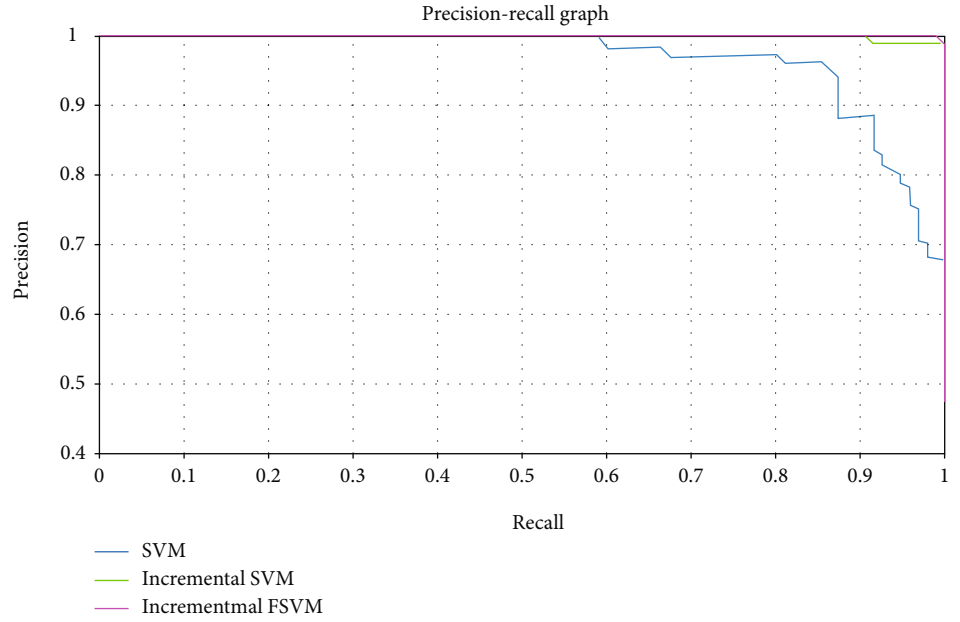


Figure 9: The PR curve of FSVM incremental learning and other diagnosis methods.

In the diagnostic problem, the recall rate can represent the proportion of correctly identified positive cases in all confirmed cases. This measures the recognition ability of the diagnostic model for new diseases. The accuracy rate is oriented to the training model, which represents the proportion of confirmed cases identified by the model. The two measure the diagnostic performance of the diagnostic model from different angles. We generally combine the two to draw precision-recall (PR) curves to investigate the diagnostic model. The PR curves under different samples and methods are shown in Figures 8 and 9.

When the PR curve is closer to the upper right, it indicates that the performance of the model is better. When comparing different models, if the PR curve of one model is completely covered by the PR curve of another model, it indicates that the performance of the latter is better than the former. From the experimental results in Figure 8, we can see that with the increase of samples and the update of

the diagnostic model, the PR curve of the new diagnostic model gradually approaches to the right and up, and the PR curve updated each time can completely cover the previous curve, which also shows that the proposed FSVM incremental learning method can effectively improve the performance of the diagnostic model. From Figure 9, compared with the SVM without update and the SVM incremental method, the FSVM incremental learning method can also cover the other two methods, and the curve obtained is more upper right than the other two methods, which also shows that the FSVM incremental learning diagnosis method proposed in this paper can obtain better diagnosis effect.

5. Conclusion

The diagnosis of new diseases is a challenging problem in intelligent diagnosis and treatment with machine learning. In order to solve the problem of few sample cases, the SVM method is selected in this paper. At the same time, incremental learning is used to update the sample database and diagnostic model. Incremental learning is an important means to ensure that the knowledge-based intelligent diagnosis method can adapt to the increase of samples. According to the basic principle of the SVM method, this paper determines the sample set related to the model, mainly including support vector set, boundary sample set, and new sample set, in which boundary sample set solves the problem of support vector transformation. In order to solve the problem that the influence of noise points and outliers on the diagnosis results increases after the number of samples is reduced, the FSVM method is used in the process of model updating. Experiments show that the proposed method not only effectively simplifies the incremental training set but also effectively improves the training efficiency while ensuring the diagnosis accuracy. The addition of fuzzy membership also effectively improves the generalization of the model. The research of this paper can provide a new idea for the application of machine learning method in the field of intelligent medical diagnosis, especially in the early stage of new diseases and the real-time update of the diagnosis model. Future directions include continuing to improve sensitivity and accuracy for COVID-19 and other new disease infections as new data is collected, as well as extend the proposed method to risk stratification for survival analysis, predicting risk status of patients and so on.

Data Availability

Data will be available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Q. Liu, W. Xu, S. Lu et al., "Landscape of emerging and re-emerging infectious diseases in China: impact of ecology, climate, and behavior," *Frontiers of Medicine*, vol. 12, no. 1, pp. 3–22, 2018.

[2] T. Jin, H. Ding, H. Xia, and J. Bao, "Reliability index and Asian barrier option pricing formulas of the uncertain fractional first-hitting time model with Caputo type," *Chaos, Solitons and Fractals*, vol. 142, p. 110409, 2021.

[3] H. Cui, Y. Guan, H. Chen, and W. Deng, "A novel advancing signal processing method based on coupled multi-stable stochastic resonance for fault detection," *Applied Sciences*, vol. 11, no. 12, p. 5385, 2021.

[4] W. Deng, S. Shang, X. Cai et al., "Quantum differential evolution with cooperative coevolution framework and hybrid mutation strategy for large scale optimization," *Knowledge-Based Systems*, vol. 224, article 107080, 2021.

[5] W. Deng, J. Xu, X.-Z. Gao, and H. Zhao, "An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–10, China, 2020.

[6] S. Tao and X. Xiulin, "Medical big data analysis and clinical application based on machine learning," *Software guide*, vol. 18, pp. 1–5, 2019.

[7] X. Zhang, "New concept of the development of modern medicine: make full use of the internet, large data, and artificial intelligence," *Zhongguo Fei Ai Za Zhi*, vol. 21, no. 3, pp. 141-142, 2018.

[8] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[9] M. Kotti, L. D. Duffell, A. A. Faisal, and A. H. McGregor, "Detecting knee osteoarthritis and its discriminating parameters using random forests," *Medical Engineering & Physics*, vol. 43, pp. 19–29, 2017.

[10] L. Zhi, S. Guoming, and L. Mingyu, "Syndrome differentiation model of coronary heart disease based on attribute weighted naive Bayes," *Journal of Guangxi Normal University*, vol. 26, no. 4, pp. 67–70, 2008.

[11] K. Zhang, W. Lu, and P. Marziliano, "Automatic knee cartilage segmentation from multi-contrast MR images using support vector machine classification with spatial dependencies," *Magnetic Resonance Imaging*, vol. 31, no. 10, pp. 1731–1743, 2013.

[12] T. Yuchi and H. Liang, "A medical data analysis model based on SVM," *Journal of Northeast Normal University*, vol. 47, no. 1, pp. 77–82, 2015.

[13] S. Chae, S. Kwon, and D. Lee, "Predicting infectious disease using deep learning and big data," *International Journal of Environmental Research and Public Health*, vol. 15, no. 8, p. 1596, 2018.

[14] S. A. Lauer, K. H. Grantz, Q. F. Bi et al., "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application," *Annals of Internal Medicine*, vol. 172, no. 9, pp. 577–582, 2020.

[15] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing," *Radiology*, vol. 296, no. 2, pp. 1–5, 2020.

[16] M. Chung, A. Bernheim, X. Y. Mei et al., "CT imaging features of 2019 novel coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.

[17] A. Amyar, R. Modzelewski, and S. Ruan, *Multi-task deep learning based CT imaging analysis for COVID-19: classification and segmentation [EB/OL]*, 2021, http://www.medrxiv.org/content/10.1101/2020.04.16.20064709v1.full.pdf.

[18] S. Wang, Y. Zha, W. Li et al., "A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis," *European Respiratory Journal*, vol. 56, no. 1, pp. 1–31, 2020.

[19] F. Jiang, L. Deng, L. Zhang, Y. Cai, C. W. Cheung, and Z. Xia, "Review of the clinical characteristics of coronavirus disease 2019 (COVID-19)," *Journal of General Internal Medicine*, vol. 35, no. 5, pp. 1545–1549, 2020.

[20] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 15, pp. 5682–5689, 2020.

[21] M. Goncharov, M. Pisov, A. Shevtsov et al., "CT-based COVID-19 triage:deep multitask learning improvesjoint identification and severity quantification," *Medical Image Analysis*, vol. 71, article 102054, 2021.

[22] L. Wang and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Scientific Reports*, vol. 10, no. 1, 2020.

[23] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, 2021.

[24] N. A. Syed, H. Liu, and K. Sung, "Handling concept drifts in incremental learning with support vector machines," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '99*, pp. 317–321, San Diego, California, USA, 1999.

[25] T. Poggio and G. Cauwenberghs, "Incremental and decremental support vector machine learning," *Advances in Neural Information Processing Systems*, vol. 13, no. 5, pp. 409–412, 2021.

[26] C. H. Li, K. W. Liu, and H. X. Wang, "The incremental learning algorithm with support vector machine based on hyperplane-distance," *Applied Intelligence*, vol. 34, no. 1, pp. 19–27, 2011.

[27] J. Xu, C. Xu, B. Zou, Y. Y. Tang, J. Peng, and X. You, "New incremental learning algorithm with support vector machines," *IEEE Transactions on Systems Man and Cybernetics Systems*, vol. 49, no. 11, pp. 2230–2241, 2018.

[28] X. Wang, C. Wu, D. Bai, and H. Zhang, "A fast svm incremental learning algorithm based on the central convex hulls algorithm," *Global Congress on Intelligent Systems*, vol. 3, pp. 472–475, 2009.

[29] D. Wang, H. Qiao, B. Zhang, and M. Wang, "Online support vector machine based on convex hull vertices selection," *Transactions on Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 593–609, 2013.

[30] C. F. Lin and S. D. Wang, "Fuzzy support vector machines," *IEEE Transaction on Neural Network*, vol. 13, no. 2, pp. 464–471, 2002.

[31] VAPNIK, *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 2000.

[32] https://pan.baidu.com/s/1o5EsPZ21p_cR2CUoElNW-g.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 2794888, 11 pages
https://doi.org/10.1155/2021/2794888

Research Article

A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric Measurements: A Cross-Sectional Study in Sarawak, Malaysia

Soo See Chai,[1] **Whye Lian Cheah**,[2] **Kok Luong Goh**,[3] **Yee Hui Robin Chang**,[4] **Kwan Yong Sim**,[5] **and Kim On Chin**[6]

[1]*Faculty of Computer Science and Information Technology, University of Malaysia Sarawak (UNIMAS), Malaysia*
[2]*Department of Community Medicine and Public Health, Faculty of Medicine and Health Sciences, University of Malaysia Sarawak (UNIMAS), Malaysia*
[3]*School of Science and Technology, International University College of Advanced Technology Sarawak (i-CATS University College), Malaysia*
[4]*Faculty of Applied Sciences, Universiti Teknologi MARA, Cawangan Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*
[5]*School of Engineering, Faculty of Engineering, Computing and Science, Swinburne University of Technology Sarawak Campus, Malaysia*
[6]*Faculty Computing and Informatics, Universiti Malaysia Sabah (UMS), Malaysia*

Correspondence should be addressed to Soo See Chai; sschai@unimas.my

Received 28 June 2021; Revised 28 September 2021; Accepted 13 November 2021; Published 7 December 2021

Academic Editor: Giovanni D Addio

This study outlines and developed a multilayer perceptron (MLP) neural network model for adolescent hypertension classification focusing on the use of simple anthropometric and sociodemographic data collected from a cross-sectional research study in Sarawak, Malaysia. Among the 2,461 data collected, 741 were hypertensive (30.1%) and 1720 were normal (69.9%). During the data gathering process, eleven anthropometric measurements and sociodemographic data were collected. The variable selection procedure in the methodology proposed selected five parameters: weight, weight-to-height ratio (WHtR), age, sex, and ethnicity, as the input of the network model. The developed MLP model with a single hidden layer of 50 hidden neurons managed to achieve a sensitivity of 0.41, specificity of 0.91, precision of 0.65, *F*-score of 0.50, accuracy of 0.76, and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) of 0.75 using the imbalanced data set. Analyzing the performance metrics obtained from the training, validation and testing data sets show that the developed network model is well-generalized. Using Bayes' Theorem, an adolescent classified as hypertensive using this created model has a 66.2% likelihood of having hypertension in the Sarawak adolescent population, which has a hypertension prevalence of 30.1%. When the prevalence of hypertension in the Sarawak population was increased to 50%, the developed model could predict an adolescent having hypertension with an 82.0% chance, whereas when the prevalence of hypertension was reduced to 10%, the developed model could only predict true positive hypertension with a 33.6% chance. With the sensitivity of the model increasing to 65% and 90% while retaining a specificity of 91%, the true positivity of an adolescent being hypertension would be 75.7% and 81.2%, respectively, according to Bayes' Theorem. The findings show that simple anthropometric measurements paired with sociodemographic data are feasible to be used to classify hypertension in adolescents using the developed MLP model in Sarawak adolescent population with modest hypertension prevalence. However, a model with higher sensitivity and specificity is required for better positive hypertension predictive value when the prevalence is low. We conclude that the developed classification model could serve as a quick and easy preliminary warning tool for screening high-risk adolescents of developing hypertension.

1. Introduction

The mortality rate of heart and blood vessel disease is increasing globally. Among the diverse risk factors, hypertension turns out to be the most contributing element for this specific noncommunicable disease, particularly for premature cardiovascular disease [1]. Coronary heart disease, stroke, heart failure, dementia, aneurysm, and renal failure are some consequences that are closely linked to hypertension [2, 3]. In addition, hypertension was found to raise the severity and mortality rate of COVID-19 by around 2.5 times especially in elderly patients who are more than 60 years old [4].

Hypertension is characterized as blood pressure ≥ 140 mm Hg systolic and/or ≥90 mm Hg diastolic for adults, and its prevalence has become a worldwide health burden. In adolescents, hypertension is interpreted as blood pressure of ≥130 mm Hg systolic and/or ≥80 mm Hg diastolic [5]. Due to the global widespread of obesity and physical inactivity in children and adolescents, hypertension in this group has become an increasing health problem, yet often overlooked [6]. It was discovered that the risk factor levels of cardiovascular disease from children and adolescents persist into adulthood, which in turn increases the probability of heart and blood vessel disease events later in life [7]. Therefore, the prediction of adolescents at risk of hypertension before adulthood is crucial to implement better prevention and control programs [8]. Furthermore, childhood and adolescence are the crucial stages for hypertension control and prevention prior to any further clinical symptoms related to hypertension-associated cardiovascular disease [9]. The prevalence of hypertension was reported to be 24.5% among adolescents in Malaysia in a recent study [10].

Anthropometric indices are gradually trusted by scientists to be the mandatory factors in identifying the risk of heart disease [11]. The use of anthropometric indices promises a simple, inexpensive, efficient, and reliable initial screening technique for hypertension [12]. Many anthropometric indices are used to define obesity-associated hypertension. These include the most commonly used body mass index (BMI), waist circumference (WC), weight-to-hip ratio (WHR), and weight-to-height ratio (WHtR) [13]. Nonetheless, research shows that the predictive powers of anthropometric measures for hypertension are countries and ethnicities dependent [14].

The emergence of machine learning (ML) in the medical field has revealed the insight of new techniques for hypertension prediction. ML techniques could be used as an early prediction for hypertension disease and could serve as a supporting tool or second opinion in assisting medical doctors in making timely decisions [15]. Artificial neural network (ANN) models have shown to be a powerful ML technique and exhibited great success in disease prediction and classification [16]. Although the ANN has been extensively used to investigate risk factors for hypertension, the utilization of anthropometric, demographic, and lifestyle indices as the estimator for hypertension prediction did not outperform prediction models that use biomedical estimators. Furthermore, current research work using ML did not report how meaningful or clinically useful a classifier might be when looking at the prevalence of hypertension for a population. Therefore, there is a need to bridge this research gap by understanding whether the use of simple anthropometric is feasible for hypertension prediction and how clinically beneficial the developed model is.

In two earlier works [17, 18], the prevalence of hypertension in Sarawak adolescents and its relationships with anthropometric indices were analyzed using multivariate logistic regression and the stepwise logistic regression statistical approach. This research work is an evolution of the previous two studies by focusing on the use of an artificial neural network model. The purpose of this research is fourfold: (a) investigate which anthropometric indices are important for adolescents hypertension prediction, (b) develop an artificial neural network model for hypertension prediction focusing on the use of anthropometric indices based on a cross-sectional research work conducted in Sarawak, Malaysia, (c) analyze whether hypertension in adolescents could be reliably predicted using anthropometric indices, and (d) assess how clinically beneficial the developed model is.

2. Related Work

Many researchers have implemented ANN models for hypertension prediction, and some of these recent researches are [19–30]. Among these, Bani-Salameh et al. [26] developed a multilayer perceptron (MLP) neural network model with six inputs: age, weight, fat ratio, blood pressure, alcohol, and smoking; one hidden layer and one output layer of hypertension and nonhypertension classes were implemented to train and test a sample size of 760 patients. They managed to achieve a correct classification rate of 68.7% with a measured Area Under the Receiver Operating Characteristic (ROC) curve (AUC) of 0.618. In addition, the authors compared the classification results of the MLP model with the *k*-nearest neighbour (KNN) and Support Vector Machine (SVM) and concluded that MLP outperformed these two models. The analysis on the independent variables revealed that blood pressure was the most important variable while smoking was the least significant variable.

In another study by López-Martínez et al. [27], a three-layered ANN model with rectified linear activation function (ReLU) in the hidden layers to classify hypertension and nonhypertension patients using sex, race, body mass index (BMI), kidney disease, and diabetes as the input features was implemented. A large imbalance sample size of 24,434 with 60.71% nonhypertensive and 30.29% hypertensive patients was used. The ANN model implemented with seven inputs, 3 hidden neuron layers with 64, 32, and 16 nodes, respectively, and 2 outputs managed to produce classification results with a sensitivity of 40%, specificity of 87%, precision of 57.8%, and AUC of 0.77. In their earlier work [28], a logistic regression model was used on the data set from the same source but smaller size (19,709), and they achieved classification results with a sensitivity of 77%, specificity of 68%, precision of 32%, and AUC of 73% (95% CI [0.70–0.76]). Although the total number of samples used was slightly smaller in [28] as compared to their work in [27],

it showed that the use of the ANN model could produce a better classification result.

A gradient descent backpropagation neural network model with four hidden units and 0 momentum value produced the best AUC (0.67), specificity (88%), sensitivity (30.6%), and precision (57.43%) results in the research work by Sakr et al. [29]. The features used included age, metabolic equivalents (METS), resting systolic blood pressure, peak diastolic blood pressure, resting diastolic blood pressure, coronary artery disease, the reason for the test, history of diabetes, percentage of heart rate achieved, race, history of hyperlipidemia, aspirin use, and hypertension response. The total number of patients was 23,095 with ages ranged between 17 and 96.

A study focusing on predicting the systolic and diastolic blood pressure of archers aged between 13 and 20 was carried out using an ANN model in [30] using a small sample size of 50 targets. The ANN model used only the calf circumference as the input variable. They reported the results for systolic and diastolic blood pressure prediction in terms of R^2 (0.95, 0.95), mean absolute percentage error (MAPE) (0.05, 0.06), means of mean absolute error (MAE) (6.55, 4.44), and root mean square error (RMSE) (78.05, 35.51).

There are other earlier studies [31–35] that utilized the ANN for hypertension classification, and each of these research works exhibited their cost and values. From the most recent and relevant research mentioned above, it could be concluded that the use of anthropometric indices together with sociodemographic and lifestyle parameters is beneficial as an initial screening for hypertension. As self-reported diabetes and hypertension are not reliable [36] and the lifestyle parameters reporting are subjective [37], in our work, only simple anthropometric measurements together with sociodemographic data are used as the features to predict cases of hypertension. We want to look into how basic anthropometric measurements combined with sociodemographic data may be used to predict hypertension in adolescents and which variables contribute to predicting hypertension. The classification results derived from this study would reveal whether hypertension in adolescents could be predicted accurately using the anthropometric indices. Several performance assessment measures, such as ROC, AUC, sensitivity, specificity, accuracy, RMSE, MAE, and MAPE, were provided as a way to benchmark the constructed models in the aforementioned review. However, the question of whether the developed model is significant, particularly in terms of clinical utility in a population with a given hypertension prevalence, remains unanswered.

3. Method

3.1. Data Source and Study Population. A cross-sectional study assessing the blood pressure of secondary school children aged between 13 and 17 years in Sarawak was carried out for 7 months from 9 March 2016 to 27 September 2016. Ethical approval was obtained from the Medical and Ethical Committee of Universiti Malaysia Sarawak (UNIMAS/TNC (AA)-03.02/06-11 Jld.3(1)) and the Ministry of Education Malaysia.

Sarawak is the largest state in Malaysia located on the island of Borneo. According to the Department of Statistics Malaysia [38], in the year 2019, the population in Sarawak is estimated to be 2.81 million with more than 40 subethnic groups. Each of these subethnic groups has its own language, lifestyle, and culture [17]. Iban, Chinese, Malay, Bidayuh, Melanau, and Orang Ulu are among the six major subethnics in Sarawak.

A total of 19 schools participated in this study with 14 of these schools classified as rural while the other 5 schools were classified as urban. For each school, a class was randomly chosen from each of the schooling levels of secondary one to secondary six. Only participants without physical and mental disability, no prediagnosed hypertension, and sickness that might lead to secondary hypertension were enrolled in the study. Data collection was carried out by a team of trained laboratory personnel. According to the Ministry of Education, the total number of students aged 13 to 17 in Sarawak in February 2014 was 200,130. Equation (1) is used to compute the required sample size (s) for a finite group [39]:

$$s = \frac{X^2 NP(1-P)}{d^2(N-1)} + X^2 P(1-P), \tag{1}$$

where X is the z-score for 99% confidence interval (2.58), N is the population size (200130), P is the population proportion (assume to be 0.5 as this would produce the maximum sample size), and d is the degree of accuracy or margin of error (0.028).

According to the calculations, a sample size of 2124 was required.

Sociodemographic information comprising the age, sex, and ethnicity of each participant was recorded. Next, the trained personnel would gather the anthropometric data from the participants. The anthropometric data collection was done using a SECA body meter and portable weighing scale. During weighing, the participants were asked to take off their footwear. In addition, it was ensured that the participants only wore their school uniforms during this process. For height measurements, the participants were requested to stand upright with no footwear on a flat surface with their back of the heels and occiput against the equipment. The weight and height were recorded to the precision of 0.1 kg and 0.1 cm, respectively. For waist circumference, measurements were taken using a plastic nonelastic tape placed at the midpoint of the last rib and the top of the hip bone (iliac crest).

The body mass index (BMI) was computed using the height and weight data provided by dividing the participant's weight (kg) by the squared height (m^2). The indices of waist-to-height ratio (WHtR) were calculated based on the ratio of the waist circumferences (WC) (cm) to height (cm). Conicity index (CI), an anthropometric measurement that is used to assess central adiposity, is calculated using

$$\text{Conicity index (CI)} = \frac{\text{waist circumference (m)}}{0.109 \times \sqrt{\text{body weight (kg)/height(m)}}}. \tag{2}$$

A digital blood pressure monitor was used for blood pressure measurements. The participants were requested to rest for 5 minutes to ensure that there was no exercise before the measurement. In addition, the participants were also checked to ensure that they did not consume any caffeine or medication before the measurement. For each participant, two measurements were taken. There was an interval of one minute between these two measurements. If the differences between these two readings were more than 5 mm Hg, a third reading would be taken. A third reading would also be taken when a participant was found to be prehypertension or hypertension. The average of these readings would be calculated as the final blood pressure reading for each of the participants. The participants were categorized into prehypertension, hypertension, and normal following the 4th report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents [40] where the cut-off point was based on age, sex, and height.

A total of 2461 sample data with a slightly higher number of females ($n = 1428$, 58%) compared to the number of males ($n = 1033$, 42%) was collected. This sample size is higher than the needed minimum sample size determined using Equation (1) and hence represents the Sarawak adolescent population. The mean age of the participants was 14.5 $\pm$ 1.50 years. In terms of ethnicity, the participants were mostly Iban, followed by Malay, Chinese, Bidayuh, and other ethnicities. The sociodemographic data included the age, sex, location, ethnicity, and whether the parent(s) was/were hypertensive of the participants, which is shown in Table 1. Most of the participants were from rural areas (74.2%). Referring to Table 2, the males had higher mean weight, height, and waist circumferences (WC), whereas the females showed higher mean body mass index (BMI) and waist-to-height ratio (WHtR). Both sexes exhibited the same mean C index. In terms of hypersensitivity (Table 3), it was found that more males were in the prehypertension and hypertension categories comparing to the females.

3.2. Methodology Design and Implementation. In this study, a multilayer perceptron feedforward neural network was designed and developed in the SAS Visual Data Mining and Machine Learning (VDMML) environment. The overall process of this classification procedure is shown in Figure 1. The detail of each step is presented below.

3.3. Data Partitioning. The statistical properties of the training, validation, and testing data play a vital role in ANN prediction and classification. The data set is partitioned into three subsets: 60% training, 30% validation, and 10% testing. For the original data set, the prehypertensive and hypertensive categories are grouped as one, resulted in binary output variables (normal and hypertensive) [32]. With this grouping, the total numbers of hypertensive and normal categories are 741 (30.1%) and 1720 (69.9%), respectively. Stratified random sampling according to the hypertensive and normal group ratio was done for the training, validation, and testing data sets. The distribution of the data obtained is shown in Table 4.

Table 1: Sociodemographic data collected.

(a)

	n	%
Sex		
Male (M)	1033	42.0
Female (F)	1428	58.0
Ethnicity		
Iban	737	29.9
Malay	681	27.7
Chinese	475	19.3
Bidayuh	256	10.4
Other	312	12.7
Location		
Urban	634	25.8
Rural	1827	74.2
Parents hypertension history		
One of the parents	448	18.2
Both parents	80	3.3
No	1933	78.5

(b)

Age	Min	Max	Mean	Standard deviation
Male (M)	12	17	14.4	1.48
Female (F)	12	17	14.5	1.51

3.4. Variable Selection. In SAS VDDML environment, using the Fast Supervised Selection method, a set of input variables that mutually explain the maximum amount of variance contained in the target variable is chosen. The Fast Supervised Selection approach, which utilizes the Bayesian Information Criterion, penalises larger models more strongly and favours smaller models as a way of completing the selection process. With the cumulative variance cut-off set to 1.0, the Fast Supervised Selection process ends when the selected variables can explain this proportion of the overall variation. Table 5 shows the proportion of the variance explained by these five selected parameters. From the total 11 input variables (sex, ethnicity, location, parents' hypertension history, age, weight, height, BMI, WC, WHtR, and C index), 5 parameters are selected: age, sex, ethnicity, weight, and WHtR.

3.5. Feature Extraction. In this part of the procedure, new feature(s) would be produced using the five variables obtained from the previous variable selection stage. The newly created features would capture the central characteristics of the selected data set and represent this data set in a lower-dimensional space. Principal Component Analysis (PCA) is a simple and most popular nonparametric method of obtaining the most relevant information from redundant or noisy data [41], and the new set features are called Principal Components (PCs). Using the PCA procedure, the features weight and WHtR variables are combined as a new variable, named Principal Component 1 (PC1). The use of

Table 2: Anthropometric data of the participants.

	Male ($n = 1033$)				Female ($n = 1428$)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Weight (kg)	24.4	121.8	55.5	14.78	21.2	109.4	51.0	12.80
Height (m)	1.3	1.8	1.6	0.08	1.24	1.78	1.5	0.06
BMI (kg/m^2)	13.3	43.1	21.3	4.72	13.1	43.5	21.6	4.78
WC (cm)	51.5	125.0	71.3	11.56	50.0	655.0	70.2	18.46
WHtR	0.3	0.7	0.4	0.07	0.3	4.1	0.5	0.11
C index	0.8	1.4	1.1	0.07	0.8	11.0	1.1	0.27

Table 3: Blood pressure profile of the participants.

Sex	Male ($n = 1033$)		Female ($n = 1428$)	
Blood pressure	n	%	n	%
Prehypertension	199	19.3	125	8.8
Hypertension	232	22.5	185	13.0
Normal	602	58.3	1118	78.3

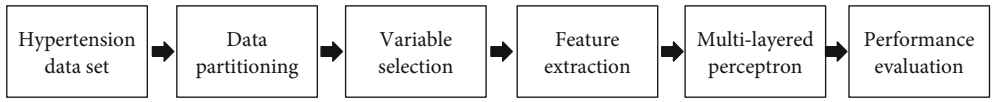


Figure 1: Overall classification procedures implemented in this study.

Table 4: Distribution of data using a stratified sampling method according to the ratio of hypertensive and normal groups.

Partition	Normal		Hypertensive		Total ($N = 2461$)	
	n	%	n	%	n	%
Training	1032	69.9	445	30.1	1477	60.0
Validation	516	69.9	222	30.1	738	30.0
Testing	172	69.9	74	30.1	246	10.0

Table 5: Proportion of variance explained for the five selected parameters through Fast Supervised Selection method.

Parameter	Proportion of variance explained
Weight	0.2314
Sex	0.2540
Ethnic	0.2614
WHtR	0.2657
Age	0.2682

PCA for feature extraction to reduce the feature dimension is well documented for clinical studies utilizing electronic healthcare records in [42]. The body weight is significantly correlated to the waist circumference [43]. The WHtR holds additional information on the height. Using the PCA process, a new feature (PC1) that captured the essential features from these two variables was created. Therefore, the final input features are reduced from five to four.

3.6. Artificial Neural Network Model. A multilayer perceptron neural network model with four input features, a single hidden layer of 50 hidden neurons, and one output layer is developed for the classification of hypertension and normal targets. The trial-and-error approach, which is a commonly utilized method [44], was applied in this study to determine the hidden neurons in the neural network model. Single-layer feedforward neural network possesses the universal approximation property [45]. Figure 2 shows the network architecture of the developed model.

The input variables are normalized using the z-score normalization method. The model properties are summarized in Table 6. Early stopping with five stagnations is carried out to avoid overtraining and to reduce training time. The model uses the Limited-Memory Broyden Fletcher Goldfarb Shanno (LBFGS), one of the quasi-Newton methods, that requires less computer memory.

4. Performance Evaluation

In this study, a few performance evaluation metrics are calculated to assess the performance of the developed multilayer perceptron model on hypertensive and normal patient classification.

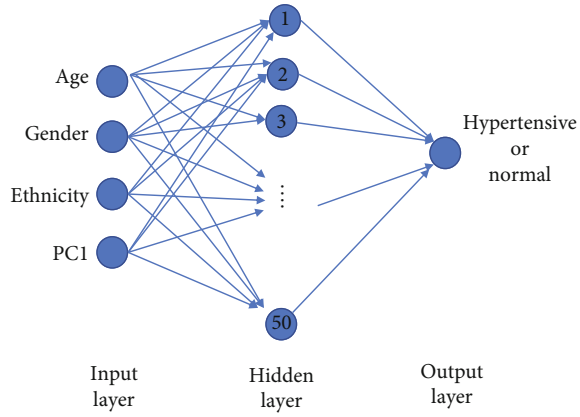


Figure 2: Multilayer perceptron model developed in this research study.

Table 6: Multilayer perceptron parameter settings.

Parameter	Value
Input dimension	4
Number of output classes	2
Number of hidden layers	1
Hidden layer dimension	50
Hidden layer activation function	tanh
Momentum	0
Learning rate	0.0010
Optimization method	LBFGS (Limited-Memory Broyden Fletcher Goldfarb Shanno)

In general, the performance of the binary classifier is grounded on the calculation of the following four parameters:

(i) True positive (TP) is defined as the number of hypertensive adolescents who are classified as hypertensive

(ii) False negative (FN) is defined as the number of hypertensive adolescents who are classified as normal

(iii) False positive (FP) is defined as the number of normal adolescents who are classified as hypertensive

(iv) True negative (TN) is defined as the number of normal adolescents who are classified as normal

Using these four parameters, the sensitivity, specificity, precision, F-score, accuracy, misclassification rate, Receiver Operating Characteristic (ROC) Curve, and Area Under the ROC Curve (AUC) are calculated.

4.1. Bayes' Theorem. The sensitivity and specificity of a classifier can be used to assess its validity. However, these two performance indicators do not accurately reflect how well the model performs for a certain population given the incidence of a specific condition. In order to evaluate how relevant or therapeutically beneficial a test could be for a population, we need underlying information about the predicted incidence or prevalence of a disease. Bayes' Theorem is useful to explain this [46]. The formula of Bayes' Theorem is

$$P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(B)}, \tag{3}$$

where $P(A)$ is the unconditional probability of the disease in the population, i.e., prevalence; $P(B)$ is the unconditional probability of the classifier/test returning positive; $P(B \mid A)$ denotes the chances of event B given that event A occurring; and $P(A \mid B)$ is the posterior probability which denotes the chance of A happening given B.

5. Results

5.1. Multilayer Perceptron Model Performance. The distribution of the actual classification results for training, validation, and testing data sets are presented using the confusion matrix in Tables 7–9. Using the confusion matrix, the performance metrics of the developed multilayer perceptron model are presented in Table 10. From this table, it can be seen that the developed model managed to achieve a classification accuracy of 76% with 65% precision. The sensitivity and the specificity of the model are 0.41 and 0.91, respectively, while the AUC is 0.75. It should be noted that the cut-off point used for all these matrices is 0.5. The ROC for the training, validation, and testing data sets are shown in Figures 3, 4, and 5. The similar shape of the ROC in these figures indicates that the multilayer perceptron model did not overfit the data during training; i.e., the model demonstrated comparable predictive capability in the training, validation, and testing data sets. In other words, the developed model is well-generalized. This aligns with the similar sensitivity and specificity values achieved for these three sets of data, as shown in Table 10.

5.2. Variable Importance. A classification tree model is used to determine the variable importance in predicting the output variable. This is done in two steps. During the first step, the variable importance of each variable is calculated based on the change of Residual Sum of Square (RSS) when a split is found at a node. The maximum variable importance value is found from these values. In the second step, the relative variable importance value for each variable is calculated by dividing the variable importance by the maximum variable importance value. The detailed calculation of RSS could be found in [47]. Table 11 shows the variable importance and relative variable importance values of the four extracted features in this study. The classification results obtained without the feature extraction process in the multilayer perceptron neural network model developed in this study are included in Supplementary 1.

Table 7: Confusion matrix obtained using training data.

	Actual	
	Hypertensive	Normal
Prediction		
Hypertensive	204	66
Normal	241	966

Table 8: Confusion matrix obtained using validation data.

	Actual	
	Hypertensive	Normal
Prediction		
Hypertensive	99	44
Normal	123	472

Table 9: Confusion matrix obtained using testing data.

	Actual	
	Hypertensive	Normal
Prediction		
Hypertensive	30	16
Normal	44	156

Table 10: Classification results obtained for training, validation, and testing data sets.

Performance metrics	Training	Validation	Testing
Sensitivity	0.46	0.45	0.41
Specificity	0.94	0.91	0.91
Precision	0.76	0.69	0.65
F-score	0.57	0.54	0.50
Accuracy	0.79	0.77	0.76
Misclassification rate	0.21	0.23	0.24
AUC	0.82	0.79	0.75

5.3. Reliability Test Using Bayes' Theorem. According to a study in year 2018, hypertension of secondary students in Sarawak was 30.1% [18]. The population of adolescents in Sarawak was 200130. Use Bayes' Theorem formula in Equation (3):

Event A denotes the prevalence of adolescent hypertension in Sarawak: $P(A) = 0.301$.

Use the sensitivity and specificity of the model developed in this study: sensitivity = 0.41 and specificity = 0.91.

Event B denotes the unconditional probability that our test coming up positive, which would include both true positive and false positive using our test. To calculate the total true positive (TTP) of our test (N_{TP}),

$$N_{TP} = \text{hypertension prevalence} \times \text{total population} \times \text{sensitivity} = 0.301 \times 200130 \times 0.41 = 24698. \quad (4)$$

In order to calculate the total false positive (TFP) (N_{FP}),

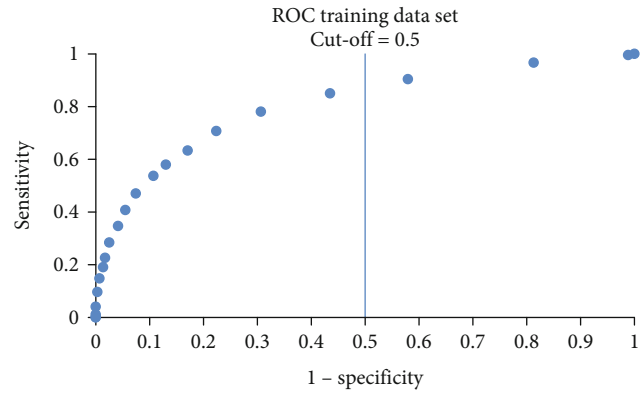


Figure 3: ROC of the training data set.

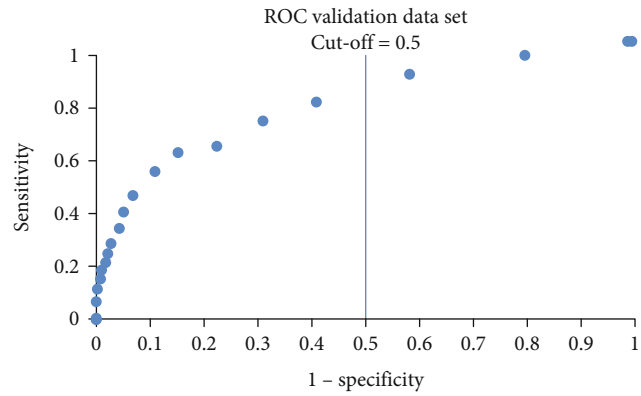


Figure 4: ROC of the validation data.

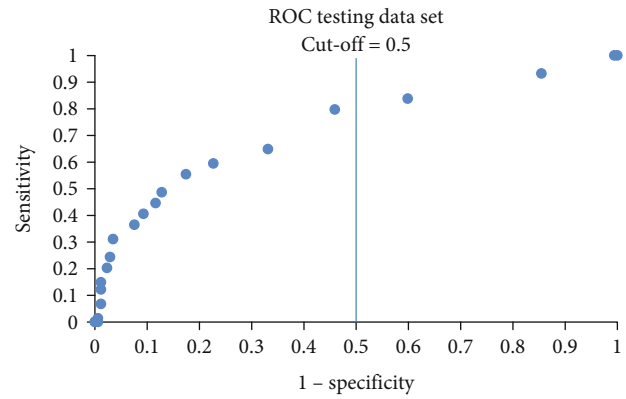


Figure 5: ROC of the testing data.

Table 11: Variable importance.

Variable	Variable importance	Relative variable importance
PC1	246.67	1.00
Sex	37.88	0.15
Age	33.02	0.13
Ethnicity	32.44	0.13

$$N_{\mathrm{FP}} = \text{probability of not having hypertension} \times \text{total population} \times (1 - \text{specificity}) = (1 - 0.301) \times 200130 \times (1 - 0.91) = 12590.$$
$$\text{Therefore, the total positive (TP) from our test} = N_{\mathrm{TP}} + N_{\mathrm{FP}} = 24698 + 12590 = 37288. \tag{5}$$

With this, $P(B) = 37288/200130 = 0.1863$.
From Equation (3):

$$P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(B)}. \tag{6}$$

$P(A \mid B)$ denotes the likelihood that an adolescent will have hypertension if our model indicates that he or she is hypertensive. $P(B \mid A)$ defines the likelihood of receiving a positive result, regardless of whether it is a true-positive or false-positive. As a result, $P(B \mid A)$ represents our sensitivity.

$$P(A \mid B) = \frac{0.301 \times 0.41}{0.1863} = 0.662 = 66.2\%. \tag{7}$$

This indicates that an adolescent diagnosed with hypertension using our method has a 66.2% likelihood of being hypertensive.

6. Discussion

In this paper, a single hidden layer multilayer perceptron neural network was developed to model the hypertension classification problem in adolescents in Sarawak, Malaysia. The study manages to prove the claim that a multilayer neural network with one single hidden layer can model a broad range of challenges in the clinical domain.

A comparison of the performance of the developed model with the above-mentioned research is presented in Table 12. From the performance metrics obtained, it could be seen that the classification capability of the developed model (AUC = 0.75) is compatible with the use of deep learning for hypertension classification by López-Martínez et al. [27] (AUC = 0.77). Our model performs slightly better for all the other performance metrics. Besides the model developed by Bani-Salameh et al. [26] that did not report on the model's specificity, the other models, including the model developed in this research work, are better in terms of the models' specificity than the sensitivity. In other words, all these models are better at classifying normal patients than correctly classifying hypertensive patients. This could be the result of the imbalance data set used, which is higher in percentage of occurrence of normal than hypertensive patients.

Comparing to the model architecture developed by López-Martínez et al. [27], our network architecture is smaller (3 layers of 64 nodes, 32 nodes and 16 nodes, respectively, vs. single layer of 50 nodes). In addition, the percentage of normal (69.71%) and hypertensive patients (30.29%) used in [27] is similar to the ratio used in our study (30.1% normal and 69.9% hypertensive).

Another significant contribution of this research work is that only simple anthropometric measurements and sociodemographic data were collected during the cross-sectional study, i.e., age, sex, ethnicity, location, parent(s) hypertension history, weight, height, waist circumferences, and blood pressure. The variable selection process in the methodology in this study had selected age, sex, ethnicity, weight, and WHtR parameters as the input for the multilayer perceptron model. All the other research works included personal medical history data and lifestyle parameters. For example, smoking and kidney conditions were required in [27]; family history, history of hyperlipidemia, and coronary artery bypass graft in [29]; and diabetes data in [26, 27, 29]. Yet, self-reported diabetes and other medical history conditions are not reliable [36], and the lifestyle parameters reporting are subjective [37].

The analysis on the variable importance reveals that PC1, which is a new feature transformed from the weight and WHtR variable, is the most important feature for the classification of hypertensive and normal patients, followed by sex, age, and ethnicity. The use of ANN with these simple anthropometric measurements and sociodemographic data demonstrates the potential of the usage of the simple measurements for hypertension detection. However, as the predictive powers of anthropometric measures for hypertension are countries and ethnicities dependent [14], further studies on the use of these parameters on other geographical locations would better validate the usefulness of these inputs.

From the performance metrics presented for the training, validation, and testing data sets in Table 10, it could be concluded that the developed model is well-generalized. That is, the model can handle the unseen data. This is proved by the almost equal values obtained for the performance metrics of the training, validation, and testing data sets. This property is important in ensuring the usefulness of the model in real-life situation.

In terms of reliability, focusing whether the developed classifier is sufficiently trustworthy to be used in a clinical context, the prevalence of hypertension in a particular population should be taken into consideration. While a highly accurate classifier may be beneficial in populations with a greater prevalence of hypertension, it would be less instructive in populations with lower hypertension rates. In our work, if an adolescent is diagnosed with hypertension using our model, he or she has a 66.2% likelihood of having hypertension. Using Bayes' Theorem, we further examine our model with different adolescent hypertension prevalences of 10% and 50% in Sarawak. The results are summarized in Table 13. With a lower prevalence (10%), the model only managed to conclude a 33.6% chance of an adolescent of having hypertension. For a higher prevalence of 50%, the model could better conclude (82.0%) an adolescent of

Table 12: Performance metrics comparison.

Our model	Sensitivity	Specificity	Precision	F-score	Accuracy	AUC
	0.41	0.91	0.65	0.50	0.76	0.75
López-Martínez et al. [27]	0.40	0.87	0.58	0.47	0.73	0.77
Bani-Salameh et al. [26]	0.69	—	0.68	0.68	0.68	0.62
Sakr et al. [29]	0.31	0.88	0.57	0.39	—	0.67

Table 13: Model reliability testing using Bayes' Theorem for different prevalence and sensitivity levels. The model reliability on current hypertension prevalence in Sarawak adolescents is highlighted.

Prevalence	Sensitivity	Specificity	TTP	TFP	TP	$P(B)$	$P(A \mid B)$
10%	41%	91%	8205	16210	24415	0.1220	33.6%
50%	41%	91%	41026	9005	50031	0.2500	82.0%
30.1%	65%	91%	39155	12590	51745	0.2586	75.7%
30.1%	90%	91%	54215	12590	66805	0.3338	81.2%
30.1%	41%	91%	24698	12590	37288	0.1863	66.2%

having hypertension. When the sensitivity of the model is improved to 90% and the specificity remains at 91%, at 30.1% of hypertension in the Sarawak adolescents population, the model may yield an 81.2% likelihood of an adolescent having hypertension.

7. Conclusions

In this research work, a multilayer perceptron neural network with one hidden layer of 50 hidden neurons was developed. The proposed model incorporating the variable selection and feature extraction procedure managed to improve the classification accuracy of the hypertension classification problem focusing on adolescents in Sarawak, Malaysia. The primary contribution of this study effort is the smaller designed network architecture, consisting of three layers with five inputs at the input layer, one hidden layer of fifty hidden neurons, and one output layer, for improved classification accuracy utilizing simple anthropometric measures and sociodemographic data. Furthermore, we demonstrated that if an adolescent tests positive for hypertension, the established model can predict that he or she has a 66.2% likelihood of developing hypertension. This model, which combines basic and straightforward anthropometric measures with sociodemographic data, i.e. age, sex, ethnicity, weight, and WHtR, is clinically useful for Sarawak adolescents with a hypertension prevalence of 30.1%.

Although the performance of the developed model is encouraging, the model could not serve as a clinical decision-making tool for diagnosing hypertensive patients. Nevertheless, the classification result could function as an early warning mechanism to alert patients on the possibility of being hypertensive.

The process to develop a multilayer perceptron neural network for adolescent hypertension classification is outlined clearly in this work. The knowledge gained in designing, developing, implementing, testing, and analyzing the network model is valuable in a future work to build an early warning tool for hypertension prediction. Such an early warning tool could serve as a cheap, simple, and rapid screening mechanism in helping the public on identifying the risk of hypertension, especially in settings when blood pressure monitoring equipment is not available. The developed model could only predict a 66.2% likelihood of an adolescent having hypertension, which is insufficient for the model to be utilized as a clinical decision-making tool. As a result, further research into the utilization of anthropometric data for hypertension prediction using machine learning algorithms is necessary. Furthermore, it would be necessary to assess whether additional training data will enhance the accuracy of the constructed model. This might be accomplished by data augmentation to generate additional data or through data collection.

Data Availability

The data underlying the results presented in the study are available upon request to the coauthor of the paper Dr. Cheah Whye Lian upon request by email (wlcheah@unimas.my).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The project is funded under the University of Malaysia, Sarawak (UNIMAS), Cross Disciplinary Grant (F08/CDRG/1832/2019). The authors would like to express their gratitude to Dr. Cheah Whye Lian and her colleagues for collecting the data for this study.

Supplementary Materials

A more detail description of the data used in this study could be found in [17, 18]. Supplementary 1: results obtained using the multilayer perceptron model without feature extraction process. *(Supplementary Materials)*

References

[1] A. Pourshams, S. G. Sepanlou, K. S. Ikuta et al., "The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017," *The Lancet Gastroenterology & Hepatology*, vol. 4, no. 12, pp. 934–947, 2019.

[2] S. R. Daniels, "Understanding the global prevalence of hypertension in children and adolescents," *JAMA Pediatrics*, vol. 173, no. 12, pp. 1133-1134, 2019.

[3] D. Drozdz and K. Kawecka-Jaszcz, "Cardiovascular changes during chronic hypertensive states," *Pediatric Nephrology*, vol. 29, no. 9, pp. 1507–1516, 2014.

[4] G. Lippi, J. Wong, and B. M. Henry, "Hypertension in patients with coronavirus disease 2019 (COVID-19): a pooled analysis," *Polish Archives of Internal Medicine*, vol. 130, no. 4, pp. 304–309, 2020.

[5] B. Falkner, "Monitoring and management of hypertension with obesity in adolescents," *Integrated blood pressure control*, vol. Volume 10, pp. 33–39, 2017.

[6] M. Riley, A. K. Hernandez, and A. L. Kuznia, "High blood pressure in children and adolescents," *American Family Physician*, vol. 98, no. 8, pp. 486–494, 2018.

[7] M. Oikonen, J. Nuotio, C. G. Magnussen et al., "Repeated blood pressure measurements in childhood in prediction of hypertension in adulthood," *Hypertension*, vol. 67, no. 1, pp. 41–47, 2016.

[8] S. Kalantari, D. Khalili, S. Asgari et al., "Predictors of early adulthood hypertension during adolescence: a population-based cohort study," *BMC Public Health*, vol. 17, no. 1, pp. 1–8, 2017.

[9] S. A. Lule, B. Namara, H. Akurut et al., "Blood pressure risk factors in early adolescents: results from a Ugandan birth cohort," *Journal of Human Hypertension*, vol. 33, no. 9, pp. 679–692, 2019.

[10] L. JK, C. XP, L. L et al., "Prevalence and factors associated with hypertension among adolescents in Malaysia," *IIUM Medical Journal Malaysia*, vol. 18, no. 1, 2019.

[11] Q. Nguyen Minh and M. H. Nguyen Vo, "Anthropometric indexes for predicting high blood pressure in Vietnamese adults: a cross-sectional study," *Integrated blood pressure control*, vol. Volume 13, pp. 181–186, 2020.

[12] C. J. Ononamadu, C. N. Ezekwesili, O. F. Onyeukwu, U. F. Umeoguaju, O. C. Ezeigwe, and G. O. Ihegboro, "Comparative analysis of anthropometric indices of obesity as correlates and potential predictors of risk for hypertension and prehypertension in a population in Nigeria," *Cardiovascular Journal of Africa*, vol. 28, no. 2, pp. 92–99, 2017.

[13] M. Yazdi, F. Assadi, M. Qorbani et al., "Validity of anthropometric indices in predicting high blood pressure risk factors in Iranian children and adolescents: CASPIAN-V study," *The Journal of Clinical Hypertension*, vol. 22, no. 6, pp. 1009–1017, 2020.

[14] Y. Khader, A. Batieha, H. Jaddou, M. El-Khateeb, and K. Ajlouni, "The performance of anthropometric measures to predict diabetes mellitus and hypertension among adults in Jordan," *BMC Public Health*, vol. 19, no. 1, pp. 1–9, 2019.

[15] K. Arun Bhavsar, A. Abugabah, J. Singla, A. Ali AlZubi, A. Kashif Bashir, and Nikita, "A comprehensive review on medical diagnosis using machine learning," *Computers, Materials and Continua*, vol. 67, no. 2, pp. 1997–2014, 2021.

[16] N. Shahid, T. Rappon, and W. Berta, "Applications of artificial neural networks in health care organizational decision-making: a scoping review," *PloS one*, vol. 14, no. 2, article e0212356, 2019.

[17] W. L. Cheah, C. T. Chang, H. Hazmi, and G. W. F. Kho, "Using anthropometric indicator to identify hypertension in adolescents: a study in Sarawak, Malaysia," *International Journal of Hypertension*, vol. 2018, Article ID 6736251, 7 pages, 2018.

[18] W. Feng Grace Kho, W. L. Cheah, and H. Hazmi, "Elevated blood pressure and its predictors among secondary school students in Sarawak: a cross-sectional study," *Central European Journal of Public Health*, vol. 26, no. 1, pp. 16–21, 2018.

[19] Z. Assaghir, A. Janbain, S. Makki, M. Kurdi, and R. Karam, "Using neural network to predict the hypertension," *International Journal of Science & Engineering Development Research*, vol. 2, no. 2, 2017.

[20] E. W.-Y. Kwong, H. Wu, and G. K.-H. Pang, "A prediction model of blood pressure for telemedicine," *Health Informatics Journal*, vol. 24, no. 3, pp. 227–244, 2018.

[21] D. LaFreniere, F. Zulkernine, D. Barber, and K. Martin, "Using machine learning to predict hypertension from a clinical dataset," in *2016 IEEE symposium series on computational intelligence (SSCI)*, Athens, Greece, 2016IEEE.

[22] H. Zhao, Z. Ma, and Y. Sun, "A hypertension risk prediction model based on BP neural network," in *2019 International Conference on Networking and Network Applications (NaNA)*, Daegu, Korea (South), 2019IEEE.

[23] M. A. J. Tengnah, R. Sooklall, and S. D. Nagowah, "A predictive model for hypertension diagnosis using machine learning techniques," in *Telemedicine Technologies*, pp. 139–152, Elsevier, 2019.

[24] L. Wang, W. Zhou, Y. Xing, and X. Zhou, "A novel neural network model for blood pressure estimation using photoplethesmography without electrocardiogram," *Journal of healthcare engineering*, vol. 2018, Article ID 7804243, 9 pages, 2018.

[25] S. Yang, W. S. W. Zaki, S. P. Morgan, S. Y. Cho, R. Correia, and Y. Zhang, "Blood pressure estimation with complexity features from electrocardiogram and photoplethysmogram signals," *Optical and Quantum Electronics*, vol. 52, no. 3, pp. 1–16, 2020.

[26] H. Bani-Salameh, S. M. Alkhatib, M. Abdalla et al., "Prediction of diabetes and hypertension using multi-layer perceptron neural networks," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 12, no. 2, 2021.

[27] F. López-Martínez, E. R. Núñez-Valdez, R. G. Crespo, and V. García-Díaz, "An artificial neural network approach for predicting hypertension using NHANES data," *Scientific Reports*, vol. 10, no. 1, pp. 1–14, 2020.

[28] F. López-Martínez, A. Schwarcz, E. R. Núñez-Valdez, and V. Garcia-Diaz, "Machine learning classification analysis for a hypertensive population as a function of several risk factors," *Expert Systems with Applications*, vol. 110, pp. 206–215, 2018.

[29] S. Sakr, R. Elshawi, A. Ahmed et al., "Using machine learning on cardiorespiratory fitness data for predicting hypertension: the Henry Ford ExercIse Testing (FIT) Project," *PLoS One*, vol. 13, no. 4, article e0195344, 2018.

[30] R. M. Musa, M. Z. Suhaimi, A. P. Majeed, M. R. Abdullah, S. M. Mat-Rasid, and M. H. Hassan, "The application of artificial neural networks in predicting blood pressure levels of youth archers by means of anthropometric indexes," in *International Conference on Movement, Health and Exercise*, Springer, Singapore, 2019.

[31] K. Pytel, T. Nawarycz, L. Ostrowska-Nawarycz, and W. Drygas, "Anthropometric predictors and artificial neural networks in the diagnosis of hypertension," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, Lodz, Poland, 2015.

[32] N. A. Bainn, *Hypertension Predictive Model with a Neural Network Approach: A Case Study of Kumasi Metropolis*, Kwame Nkrumah University Of Science And Technology, 2012.

[33] S. Huang, Y. Xu, L. Yue et al., "Evaluating the risk of hypertension using an artificial neural network method in rural residents over the age of 35 years in a Chinese area," *Hypertension Research*, vol. 33, no. 7, pp. 722–726, 2010.

[34] C. Wang, L. Li, L. Wang et al., "Evaluating the risk of type 2 diabetes mellitus using artificial neural network: an effective classification approach," *Diabetes Research and Clinical Practice*, vol. 100, no. 1, pp. 111–118, 2013.

[35] A. Kupusinac, R. Doroslovački, D. Malbaški, B. Srdić, and E. Stokić, "A primary estimation of the cardiometabolic risk by using artificial neural networks," *Computers in Biology and Medicine*, vol. 43, no. 6, pp. 751–757, 2013.

[36] M. Ning, Q. Zhang, and M. Yang, "Comparison of self-reported and biomedical data on hypertension and diabetes: findings from the China Health and Retirement Longitudinal Study (CHARLS)," *BMJ Open*, vol. 6, no. 1, p. e009836, 2016.

[37] E. Modey Amoah, D. Esinam Okai, A. Manu, A. Laar, J. Akamah, and K. Torpey, "The role of lifestyle factors in controlling blood pressure among hypertensive patients in two health facilities in urban Ghana: a cross-sectional study," *International Journal of Hypertension*, vol. 2020, Article ID 9379128, 8 pages, 2020.

[38] Department of Statistics, Malaysiahttps://www.dosm.gov.my/v1/index.php?r=column/cone&menu_id=clJnWTlTbWFHdmUwbmtSTE1EQStFZz09.

[39] R. V. Krejcie and D. W. Morgan, "Determining sample size for research activities," *Educational and Psychological Measurement*, vol. 30, no. 3, pp. 607–610, 1970.

[40] B. K. Poh, A. N. Jannah, L. K. Chong, A. T. Ruzita, M. N. Ismail, and D. McCarthy, "Waist circumference percentile curves for Malaysian children and adolescents aged 6.0–16.9 years," *International Journal of Pediatric Obesity*, vol. 6, no. 3-4, pp. 229–235, 2011.

[41] S. Cateni, M. Vannucci, M. Vannocci, and V. Colla, "Variable selection and feature extraction through artificial intelligence techniques," in *Multivariate analysis in management, engineering and the Science*, pp. 103–118, IntechOpen, 2013.

[42] Z. Zhang and A. Castelló, "Principal components analysis in clinical studies," *Annals of translational medicine*, vol. 5, no. 17, p. 351, 2017.

[43] N. Miyatake, S. Matsumoto, M. Miyachi, M. Fujii, and T. Numata, "Relationship between changes in body weight and waist circumference in Japanese," *Environmental Health and Preventive Medicine*, vol. 12, no. 5, pp. 220–223, 2007.

[44] R. A. Sarker, S. M. Elsayed, and T. Ray, "Differential evolution with dynamic parameters selection for optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 5, pp. 689–707, 2014.

[45] N. J. Guliyev and V. E. Ismailov, "On the approximation by single hidden layer feedforward neural networks with fixed weights," *Neural Networks*, vol. 98, pp. 296–304, 2018.

[46] G. M. Chan, "Bayes' theorem, COVID19, and screening tests," *The American Journal of Emergency Medicine*, vol. 38, no. 10, pp. 2011–2013, 2020.

[47] "The TREESPLIT procedure: variable importance," *SAS Visual Statistics 8.3: Procedures 2018*https://documentation.sas.com/doc/en/casstat/8.3/casstat_treesplit_details20.htm.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 3551756, 9 pages
https://doi.org/10.1155/2021/3551756

Research Article

A Comparison among Different Machine Learning Pretest Approaches to Predict Stress-Induced Ischemia at PET/CT Myocardial Perfusion Imaging

Rosario Megna,[1] **Mario Petretta**,[2] **Roberta Assante**,[3] **Emilia Zampella**,[3] **Carmela Nappi**,[3] **Valeria Gaudieri**,[3] **Teresa Mannarino**,[3] **Adriana D'Antonio**,[3] **Roberta Green**,[3] **Valeria Cantoni**,[3] **Parthiban Arumugam**,[4] **Wanda Acampa**,[1,3] **and Alberto Cuocolo**[3]

[1]*Institute of Biostructure and Bioimaging, National Council of Research, Naples, Italy*
[2]*IRCCS-SDN, Naples, Italy*
[3]*Department of Advanced Biomedical Sciences, University Federico II, Naples, Italy*
[4]*Department of Nuclear Medicine, Central Manchester Foundation Trust, Manchester, UK*

Correspondence should be addressed to Rosario Megna; rosario.megna@ibb.cnr.it

Received 30 July 2021; Revised 29 October 2021; Accepted 15 November 2021; Published 27 November 2021

Academic Editor: Huiling Chen

Traditional approach for predicting coronary artery disease (CAD) is based on demographic data, symptoms such as chest pain and dyspnea, and comorbidity related to cardiovascular diseases. Usually, these variables are analyzed by logistic regression to quantifying their relationship with the outcome; nevertheless, their predictive value is limited. In the present study, we aimed to investigate the value of different machine learning (ML) techniques for the evaluation of suspected CAD; having as gold standard, the presence of stress-induced ischemia by ^{82}Rb positron emission tomography/computed tomography (PET/CT) myocardial perfusion imaging (MPI) ML was chosen on their clinical use and on the fact that they are representative of different classes of algorithms, such as deterministic (Support vector machine and Naïve Bayes), adaptive (ADA and AdaBoost), and decision tree (Random Forest, rpart, and XGBoost). The study population included 2503 consecutive patients, who underwent MPI for suspected CAD. To testing ML performances, data were split randomly into two parts: training/test (80%) and validation (20%). For training/test, we applied a 5-fold cross-validation, repeated 2 times. With this subset, we performed the tuning of free parameters for each algorithm. For all metrics, the best performance in training/test was observed for AdaBoost. The Naïve Bayes ML resulted to be more efficient in validation approach. The logistic and rpart algorithms showed similar metric values for the training/test and validation approaches. These results are encouraging and indicate that the ML algorithms can improve the evaluation of pretest probability of stress-induced myocardial ischemia.

1. Introduction

Artificial intelligence has assumed a consolidated role in numerous fields and also in the healthcare and research and development. Machine learning (ML), an application of artificial intelligence that refers to computational algorithms designed to learn from experience, has been used successfully for diagnosis, prognosis, and drug development [1–4]. Among the recommendations for ML implementation in clinical research, there is data normalization, feature selection, parameter tuning, and independent validation [5, 6].

In the field of cardiology, the search for methods for obtaining reliable pretests probability of disease has been underway for some time [7]. These tools should assist the physician in making decisions about referring patients for

Table 1: Clinical characteristics of cohort according to MPI outcome.

	Normal ($n = 2002$)	Ischemic ($n = 501$)	P value
Age, n (%)			<0.001
<55	777 (39)	84 (17)	
55-65	603 (30)	146 (29)	
>65	622 (31)	271 (54)	
Male gender, n (%)	881 (44)	334 (67)	<0.001
Body mass index ≥30, n (%)	1024 (51)	258 (52)	0.93
Chest pain, n (%)			<0.001
Typical	678 (34)	114 (23)	
Atypical	256 (13)	87 (17)	
Noncardiac*	1068 (53)	300 (60)	
Diabetes, n (%)	479 (24)	187 (37)	<0.001
Dyspnea, n (%)	446 (22)	139 (28)	<0.05
Family history of CAD, n (%)	945 (47)	199 (40)	<0.005
Hypertension, n (%)	1361 (68)	401 (80)	<0.005
Hyperlipidemia, n (%)	1210 (60)	343 (69)	<0.005
Smoking, n (%)	557 (28)	144 (29)	0.72
Diagnostic question, n (%)§			<0.001
Diagnostic evaluation	1642 (82)	370 (74)	
Presurgery evaluation	360 (18)	131 (26)	

*Considering noncardiac patients as the reference. §Considering diagnostic evaluation patients as the reference.

examination. Usually, for the prediction of coronary artery disease (CAD), traditional risk factors, such as age, gender, chest pain, and comorbidity related to cardiovascular diseases, such as hypertension, diabetes, and hyperlipidemia, are considered. These variables are analyzed by logistic regression to quantifying their relationship with the outcome of the exam and obtaining predictions for new patients [8–11]. However, the models obtained by these studies do not show a great performance, probably due to the declining prevalence of CAD and because the evaluation for CAD has shifted to older patients, more women, and more patients with atypical symptoms than in previous decades [12]. Including in the model, other clinical, laboratory, and instrumental characteristics could improve prediction accuracy; however, adding variables may be expensive and time-consuming and also incorrectly reclassify patients with suspected CAD. Using publicly available dataset, it has been recently reported that ML algorithms have high accuracy to detect the presence of CAD [13]. Yet, if the application of more complex algorithms on traditional risk factor may optimize the estimation of pretest probability of CAD, it remains to be defined. In the present study, we aimed to investigate the potential of different ML techniques for the evaluation of suspected CAD, having as gold standard the presence of stress-induced ischemia by ^{82}Rb positron emission tomography/computed tomography (PET/CT) myocardial perfusion imaging (MPI).

In summary, the main contributions of this work include the following:

(1) A comparison of the value of several ML algorithms in predicting the presence of stress-induced ischemia by noninvasive cardiac imaging

(2) We selected ML algorithms based on their use in the medical field and on the fact that they are representative of different classes of algorithms, such as deterministic, adaptive, and decision tree

The rest of this paper is organized as follows. Section 2 describes the method with detailed information of datasets and ML techniques used. Section 3 describes the results. The discussion is presented in Section 4 followed by the conclusions in Section 5.

2. Materials and Methods

2.1. Study Design and Eligibility. Our cohort included a total of 2503 consecutive patients, who underwent cardiac ^{82}Rb PET/CT for suspected CAD as part of their diagnostic program between June 2010 and October 2019. Patients with known CAD and patients with acute coronary syndrome were excluded. A patient was considered to have known CAD at the time of imaging based on a provided history of previously diagnosed atherosclerotic coronary disease, history of myocardial infarction (chest pain or equivalent symptom complex, positive cardiac biomarkers, or typical electrocardiographic changes), history of percutaneous coronary intervention, or history of coronary artery bypass grafting. For patients undergoing more than one PET/CT study, only the earliest procedure was considered. All patients were part of ongoing prospective dedicated database [14]. This study complies with the Declaration of Helsinki. The review committee of our institution approved this study (Ethics Committee, University Federico II, protocol number 110/17), and all patients gave informed consent.

2.2. Clinical Definitions. Chest pain was classified according to the American College of Cardiology/American Heart Association 2002 guideline update on exercise testing [15]. Patients were considered as having diabetes if they were receiving treatment with oral hypoglycemic drugs or insulin. A family history of premature CAD was defined as a diagnosis of CAD in a first-degree relative prior to or at 55 years of age. Hypertension was defined as a blood pressure > 140/90 mm Hg or use of antihypertensive medication. Hyperlipidemia was defined as total cholesterol level > 6.2 mmol/L or treatment with cholesterol lowering medication. Smoking history was defined as prior or current tobacco use. Body mass index (BMI) was dichotomized with cut-off to 30, according to obesity definition.

2.3. PET/CT Imaging. As a routine preparation for ^{82}Rb cardiac PET/CT, patients were asked to discontinue taking methylxanthine containing foods or beverages for 24 hours. Scans were acquired using a Biograph mCT 64-slice scanner (Siemens Healthcare). Rest and stress cardiac PET/CT

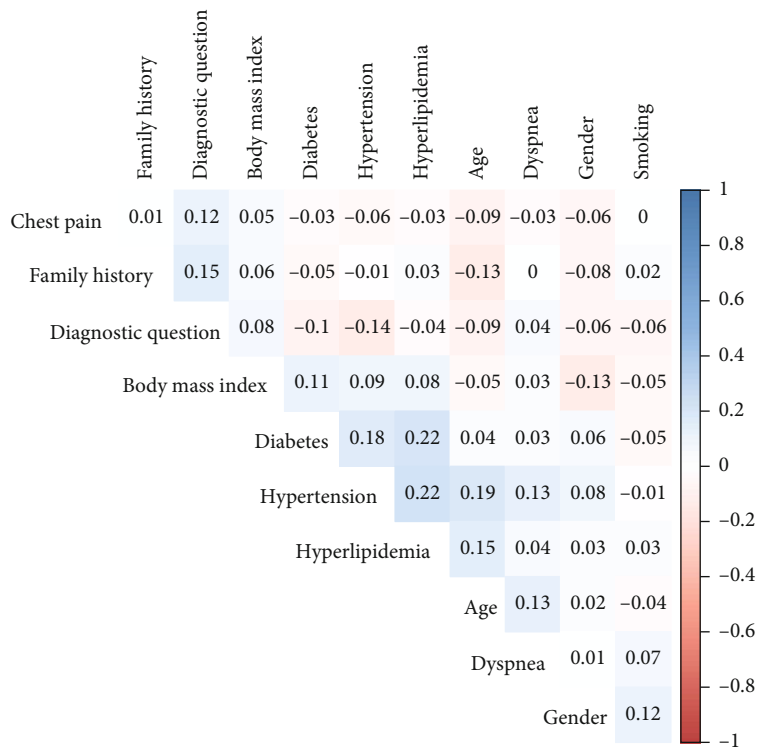


Figure 1: Correlation matrix of the features used. The matrix elements are displayed in hierarchical clustering order. The numbers indicate the Spearman ρ coefficient between two features.

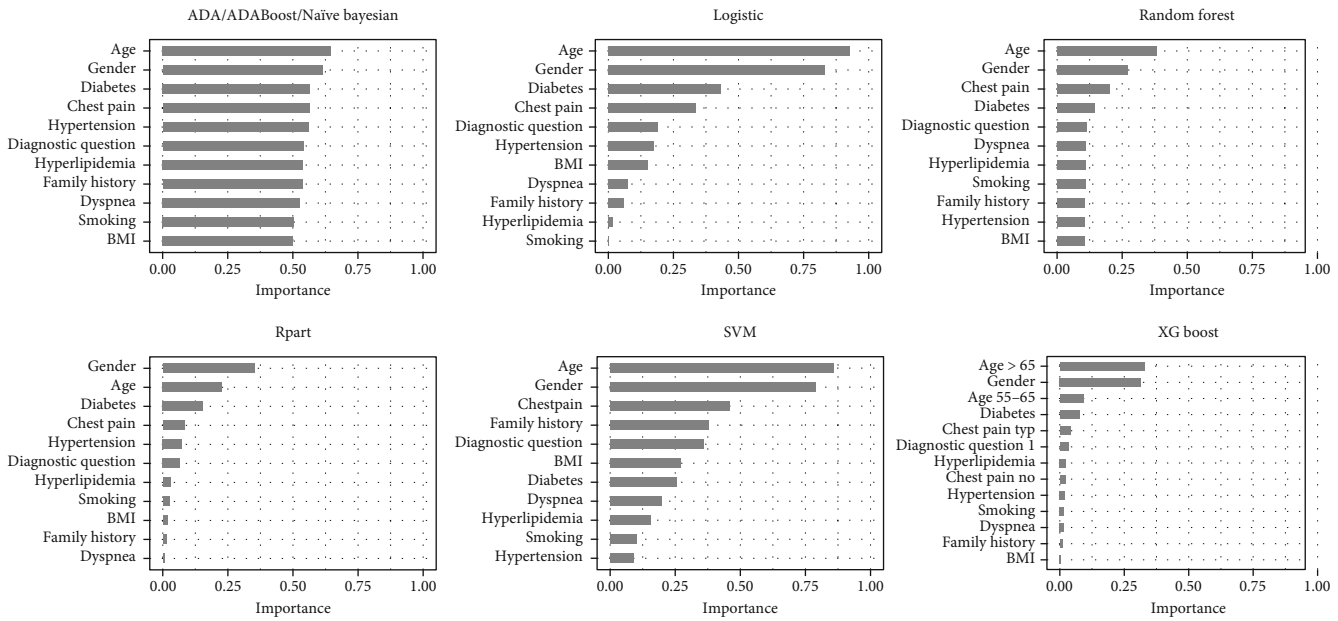


Figure 2: Importance of the features for each ML algorithm. ADA, AdaBoost, and Naïve Bayesian features importance were grouped into a single bar plot as the values for the two adaptive algorithms turned out to be equals, and Naïve Bayesian values differed with them by less than 5%.

images were acquired as follows: scout CT was performed to check patient position, and low-dose CT (0.4 mSv; 120 kVp; effective tube current, 26 mA [11-mAs quality reference]; 3.3 seconds) was performed for attenuation correction, during normal breathing before and after PET acquisitions. For both rest and stress images, 1110 MBq of ^{82}Rb was injected intravenously with a 7-minute list-mode PET acquisition. Dynamic PET acquisition was started at rest followed by adenosine pharmacologic stress ($140\,\mu g \times kg^{-1} \times min^{-1}$ for 4.5 minutes, with tracer administration between 2 and 2.5

Table 2: Values used for tuning of parameters for each ML technique.

	Parameter	Parameter space	Chosen value
ADA	Number of trees	10, 25, 50, 100, 200	25
	Max tree depth	5, 10, 20, 50	10
	Learning rate	0.001, 0.005, 0.01, 0.05, 0.1, 0.5	0.01
AdaBoost	Number of trees	10, 25, 50, 100, 200	50
	Method	AdaBoost.M1, real AdaBoost	AdaBoost.M1
Logistic	Family	Binomial	Binomial
Naïve Bayes	Laplace correction	0, 0.5, 1.0	0
	Distribution type (kernel)	True, false	False
	Bandwidth adjustment	0.01, 0.05, 0.1, 0.5, 1.0	0.1
Random Forest	Number of randomly selected predictors	3, 5, 10, 20	10
Rpart	Minimum number of observations in a node	10, 15, 30	15
	Minimum number of observations in any leaf node	3, 5, 10	5
	Max tree depth	3, 5, 10, 20	10
	Complexity parameter of the tree	0.0001, 0.001, 0.01, 0.1	0.001
SVM	Kernel	Linear, radial, sigmoid	Sigmoid
	Parameter needed for sigmoid	0.05, 0.1, 0.25, 0.5	0.1
	Cost	0.5, 1, 2, 5	1
XGBoost	Number of trees	25, 50, 100, 200	100
	Max tree depth	5, 10, 20	10
	Learning rate	0.001, 0.005, 0.01, 0.05, 0.1, 0.5	0.01
	Subsamples	0.5, 0.75, 1	1

Table 3: Metrics obtained from the ML techniques, evaluated on training/test and validation approaches.

	Training/test ($n = 2003$)				Validation ($n = 500$)			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUROC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUROC (%)
ADA	88	48	97	90	76	26	89	68
AdaBoost	89	67	95	95	71	23	87	66
Logistic	80	5	98	72	80	7	98	75
Naïve Bayes	77	23	91	70	80	27	92	73
Random Forest	89	51	98	93	75	21	89	65
Rpart	82	27	96	75	76	17	91	70
SVM	72	13	87	61	77	21	91	65
XGBoost	83	27	97	83	77	18	92	69

minutes). Rest and stress dynamic images were reconstructed into 26-time frames (12×5 seconds, 6×10 seconds, 4×20 seconds, and 4×40 seconds; total, 6 minutes) using the vendor standard ordered subsets expectation maximization 3D reconstruction (2 iterations, 24 subsets) with 6.5 mm Gaussian postprocessing filter. In addition, the images were corrected for attenuation using the low-dose CT. The heart rate, systemic blood pressure, and 12-lead ECG were recorded at baseline and throughout the infusion of adenosine. An automated software program (e-soft, 2.5, QGS/QPS, Cedars-Sinai Medical Center, Los Angeles, CA) was used to calculate the scores (summed stress score, summed rest score, and summed difference score) incorporating both the extent and severity of perfusion defects, using the standardized segmentation of 17 myocardial regions [16, 17]. A summed difference score ≥ 2 was considered ischemic.

2.4. Statistical Analysis. Statistical analysis was performed using the *R* software, version 3.6.2 (The *R* Foundation for Statistical Software, Vienna, Austria). Two-sided *P* values <0.05 were considered statistically significant. The dataset

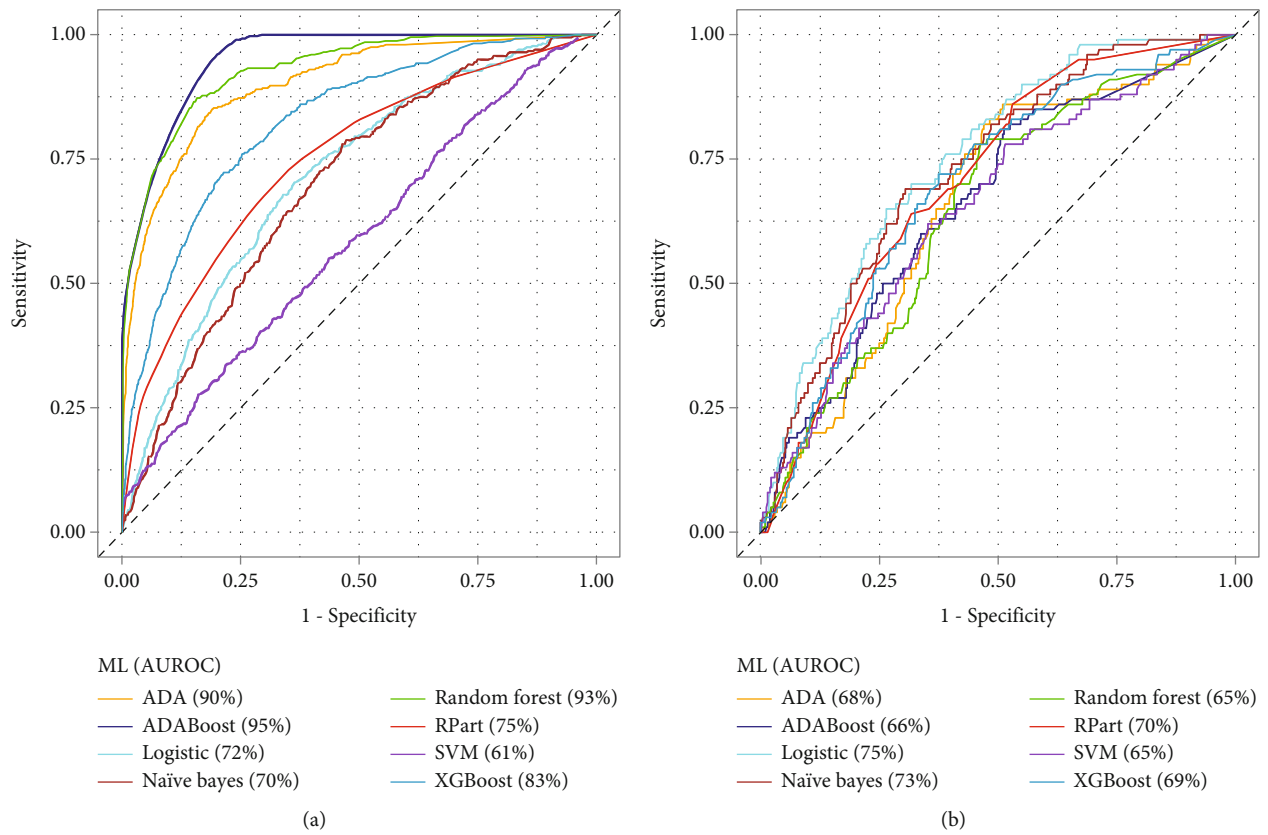


Figure 3: Comparison among the ROC curves of the eight ML techniques considered. The ML performances are reported separately for the training/test approach (a) and validation approach (b). Parenthesis are reported the AUROC values.

consisted of 11 features, of which 10 demographic or clinical variables (age, gender, BMI, typical or atypical chest pain, diabetes mellitus, dyspnea, family history, hypertension, hyperlipidemia, smoking), and the diagnostic question with two categories: diagnostic or presurgery evaluation. Age and BMI continuous variables were categorized (<55, 55-65, >65 years, and BMI < 30); then, all data were expressed as percentages. Differences between groups were analyzed by χ^2 test. The correlation among features was tested by Spearman ρ coefficient, embedded in the corrplot package. This nonparametric test is appropriate to evaluate the correlation between categorical variables and to find redundant features. Data in input to ML algorithms were normalized. Sensitivity, specificity, and accuracy were computed using the *c*onfusionMatrix function embedded in the caret package. Sensitivity evaluated how good a ML is for detecting the positive patients (i.e., ischemic according to MPI results), and its numeric value was obtained by ratio between the number of patients correctly assessed as positive by ML and the number of positive patients. Specificity evaluated the negative patients (i.e., normal according to the MPI results), and it was calculated by ratio between the number of patients correctly assessed as negative by ML and the number of negative patients. Accuracy measured how correctly a ML identified and excluded a given condition, and it was obtained from the ratio between the number of patients correctly assessed by ML and the total number of patients. Receiver operating characteristic curve is a graphic presentation of the relationship between sensitivity and specificity, whereas the area under this curve provides a measurement of the correct evaluation of ML with respect a random classifier. The areas under the receiver operating characteristic (AUROC) curves were computed by the *pROC* package.

2.5. ML Techniques. For the comparison presented in this study, we selected supervised ML algorithms, appropriate to categorical data for a binary response. We used the algorithms developed in *R*. ADA is a classification tree based on adaptive algorithms, used to fit a variety stochastic boosting. This algorithm can be used in conjunction with other types of learning procedures to improve performance. The output of these procedures, called weak learners, is combined into a weighted sum that represents the final output of the boosted classifier [18]. AdaBoost is a classifier similar to ADA, differing from this for the AdaBoost.M1 algorithm implemented by Freund and Schapire [19]. Logistic algorithm used in this study is a part of generalized linear models [20]. This classifier was chosen as a reference because adopted in clinical statistical analysis, with categorical or numerical data and dichotomous response. The equation assumed a linear relationship between the predictor variables x_i and the log odds

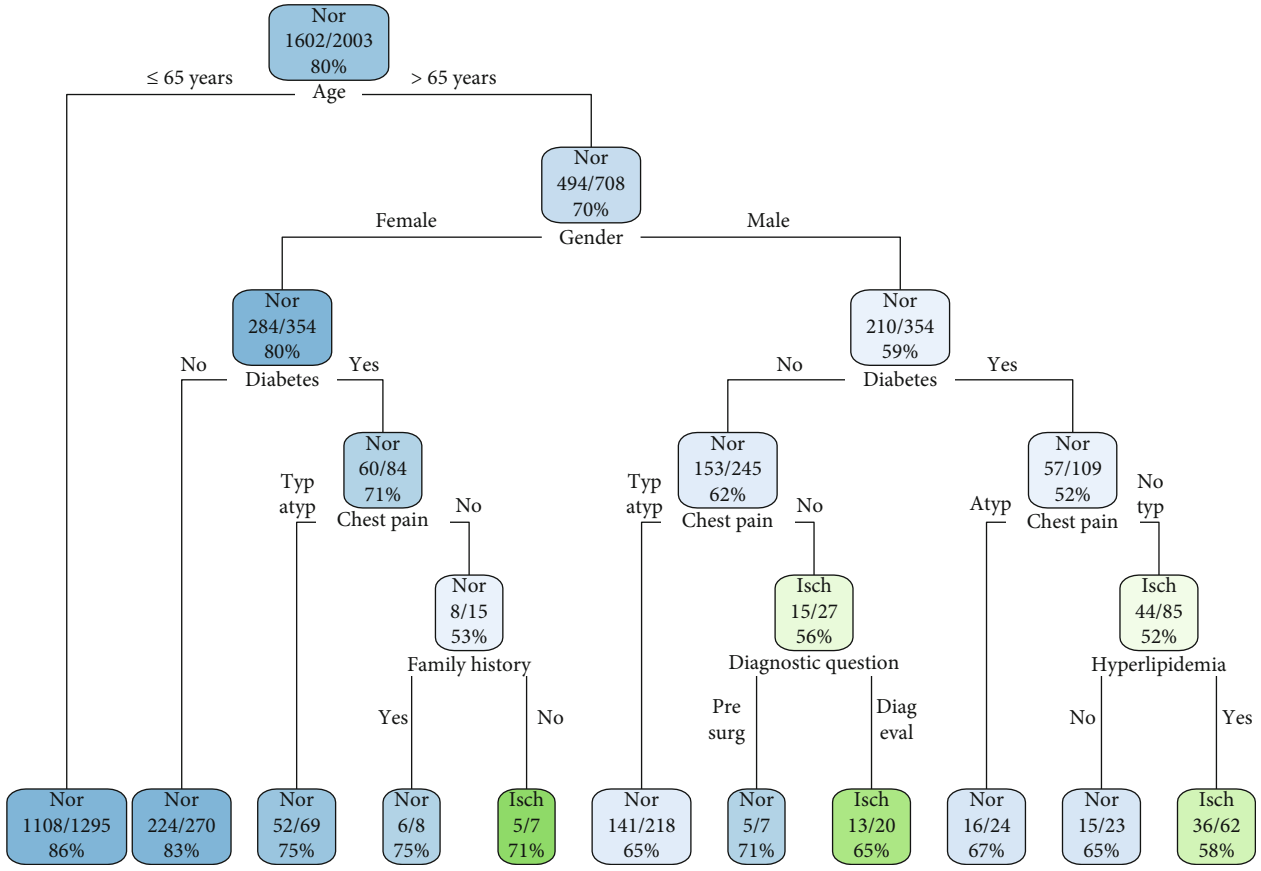


Figure 4: Decision tree obtained by rpart algorithm. Each node or leaf is reported the prevalence concerning MPI outcome (nor: normal; isch: ischemic), the ratio between the number of prevalent and total patients, and the relative percentage.

(in term of probability p) of the event, as follows:

$$\log \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i. \quad (1)$$

Then, the β coefficients are determinates, with β_0 representing the particular case with all variables equal to zero. The Naïve Bayes is a probabilistic classifier based on the Bayes' theorem. This algorithm requires a strong (naïve) independence assumption between the features [21]. Random Forest is an algorithm based on an ensemble learning method for classification and regression that operate by constructing a multitude of decision trees at training time. The procedure returns as output the class that is the mode of the classes (for classification) or average prediction (for regression) of the individual trees [22]. Rpart is a decision tree algorithm that works by splitting in two parts the dataset recursively. For each step, the split is obtained considering the feature that results in the largest possible reduction in heterogeneity of the outcome variable [23]. Support vector machine (SVM) is an algorithm that constructs hyperplanes in a high-dimensional space, which can be used for classification and regression [24]. SVM is a robust prediction method that can efficiently perform nonlinear classifications, by appropriate kernels. XGBoost is a scalable end-to-end tree boosting method, based on a sparsity-aware algorithm for sparse data and weighted quantile sketch for approximate tree learning [25].

2.6. Approaches Used for the ML Evaluation. To testing the ML performances, the data were split randomly into two parts: training/test (80%) and validation (20%). For the training/test of data, we applied a 5-fold cross-validation method, repeated 2 times. With this subset, we performed the tuning of free parameters for each algorithm. For both training/test and validation, we computed accuracy, sensitivity, specificity, and AUROC.

2.7. Hardware and Software Characteristics. For this study, we used a common personal computer equipped with a 2.2 GHz Intel i3-2330 quad-core processor, 8 GB of RAM, and a 0.5 TB SSD. The operating system was a Windows 10, whereas the scripts in *R* programming code were obtained developing inhouse software.

3. Results

Demographic and clinical characteristics of study population according to normal or ischemic MPI response are summarized in Table 1. All features, except BMI and smoking, were statistically significant to χ^2 test.

Figure 1 shows the Spearman correlation coefficients matrix of features. All the found absolute values were <0.25, highlighting only weak correlations among features. The cluster with higher correlation among features was obtained by diabetes, hypertension, and hyperlipidemia ($\rho = 0.22$). The very low correlation values demonstrated the absence of redundant features.

Figure 2 reports the feature importance for each algorithm. We observed the same feature importance values for ADA and AdaBoost algorithms, whereas small differences (<5%) were found between these procedures and the Naïve Bayes ML. Therefore, we reported a unique bar plot for these three algorithms. In general, the most important features were age and gender, followed from diabetes or chest pain. We also observed relevant differences among features importance of most of ML algorithms, except for the two adaptive and Naïve Bayesian algorithms. In fact, for these three algorithms, the importance values were comprised between 0.50 and 0.65, whereas for the logistic algorithm, we obtained larger interval of values from 0.001 to 0.93.

Table 2 summarizes the space parameters and the value chosen for the tuning of ML. Parameters were tested using a 5-fold cross-validation, repeated 2 times, targeted to maximize the C-index. Among all tested setting for each algorithm, we chose the combination with higher sensitivity to balance the result performances.

Table 3 shows the C-statistics results of the ML algorithms, for training/test and validation approaches. In general, the performances in training/test approach were better than of the validation approach. Due to unbalanced dataset, specificity resulted greater than sensitivity. For all metrics, the best performance in training/test was observed for AdaBoost ML. The Naïve Bayes ML resulted to be more efficient in validation approach. ML based on traditional logistic algorithm showed a low sensitivity and similar performance for the training/test and validation approaches. Figure 3 shows a graphical comparison among the ROC curves of the ML algorithms, for both training/test and validation approaches.

Figure 4 shows the tree generated from the rpart algorithm. To make the decision tree easier to read, the max depth was fixed to 5. The first spit was on age and for younger patients (≤65 years), without any node until the terminal leaf, where a prevalence of normal MPI of 86% was observed. For older patients (>65 years), the algorithm calculated the gender node, with a percentage of normal MPI of 70%. The split in this node, related to the female gender, was followed by diabetes, chest pain, and family history of CAD.

4. Discussion

At best of our knowledge, this is the first study comparing the value of several ML algorithms in predicting the presence of stress-induced ischemia by ^{82}Rb PET/CT cardiac imaging. We selected eight ML algorithms based on their clinical use and on the fact that they are representative of different classes of algorithms, such as deterministic (e.g., SVM), adaptive (e.g., ADA), and decision tree (e.g., rpart). The results indicate that by adaptive (ADA and AdaBoost) and Random Forest algorithms, AUROC curve was ≥90% in training/test phase.

As input features for the ML algorithms, we considered demographic data and traditional cardiac risk factors. No significant correlations were detectable between variables, a necessary condition for features selection in ML techniques and for data processing. The feature importance is an important step for ML techniques. In our study aside from demographic characteristics, diabetes and chest pain resulted to be the most useful features for predicting stress-induced ischemia by PET/CT. This result confirms another study based on SPECT, where the feature importance, obtained by logistic regression, was the following: gender, age, and chest pain [26]. Noteworthy, features (*BMI* and smoking) showing not significant χ^2 statistic resulted relevant at ML analysis. Indeed, ML algorithms may capture the subtle value of features apparently not significant at conventional analysis.

The ML algorithms showed a variable accuracy (72%-89%) by training/test phase, with low sensitivity and high specificity. This latter finding probably reflects the unbalanced dataset between normal and abnormal MPI and is in agreement with the observation that, in the contemporary pretest probability of CAD, noninvasive imaging tests have greater ruling out that ruling in capabilities [12]. Also, the AUROC values were very wide (61%-95%), with better performances for ADA, AdaBoost, and Random Forest. By these ML algorithms, we obtained the greater values of sensitivity. However, these better performances were lower in the validation set, probably due to the ensemble of weakly solutions and a high number of decision trees elaborated during the training/test phase for each of the three ML algorithms. For XGBoost, we observed a similar performance to these three algorithms, but a lower sensitivity. The Naïve Bayes and SVM resulted to have more generalized performances by the two approaches, with lightly better results by validation phase. The logistic and rpart algorithms showed similar metric values for the training/test and validation approaches.

The logistic technique, taken as a reference, did not result particularly performant with respect to the other ML algorithms. In particular, the value of sensitivity was the lowest, probably explainable with the unbalanced dataset. However, the AUROC resulted higher with respect to a similar study (AUROC = 64%) based on clinical risk factors, single-photon emission computed tomography imaging, and logistic regression [10].

As an example of a tool for decision-making, we reported the tree obtained by rpart. From a graphic point of view, it is immediate to verify the effect of age and *gender* on the construction of the decision tree. For younger patients, there is a prevalence of normal MPI, without further ramifications. Otherwise, a gender split is observed, followed in both cases by the split of diabetes and chest pain, with a larger complexity for the male gender.

Previous studies used ML algorithms in cardiology [27], but at the best of our knowledge, no study evaluated this approach to estimate the pretest probability of an ischemic response to PET/CT. In a study based, an XGBoost ML

was developed in a large series of symptomatic patients to predict pretest probability of obstructive CAD on coronary computed tomography angiography. The ML model had significantly higher discrimination (AUROC = 81%), as compared to traditional models, with a good sensitivity (91.9%) but a low (38.8%) specificity. This study was used a 10-fold cross-validation approach but and no independent validation dataset [28]. In another study [29], a SVM algorithm was used to determine the diagnostic value of joint PET myocardial perfusion and metabolic imaging for predicting obstructive coronary artery disease in symptomatic patients with available coronary angiography. The study included only 88 patients, most of them with known CAD. The joint PET evaluation improves had a good performance (AUROC = 86%), and the SVM algorithm outperformed the other methods evaluated. In a study [30], including a total of 16,120 patients, ML improved one-year risk discrimination in predicting durable left ventricular assist devices as compared to logistic regression (C-index 71% vs. 69%, $P < 0.001$); however, calibration metrics were comparable. Globally, these studies confirm limited value of current clinical models to accurately predict the presence of myocardial ischemia at stress MPI [31].

5. Conclusions

The results of this study performed in a large series of patients with suspected CAD demonstrate that the classification based on demographic and cardiovascular risk factors has a limited value in validation phase for predicting an ischemic response by ^{82}Rb PET/CT in patients with suspected CAD. We selected eight ML algorithms that are implemented by different software packages and can be used by other researchers on their MPI data. Other ML algorithms, such as monarch butterfly optimization [32], earthworm optimization algorithm [33], elephant farming optimization [34, 35], moth search algorithm [36], slime mould algorithm [37], and Harris hawks optimization [38], can also be used to predict stress-induced ischemia by MPI and should be tested in future studies. In conclusion, the role of other clinical and instrumental characteristics, as well as developing and perfecting more complex algorithms to improve the prediction of stress-induced ischemia by MPI, remains a work in progress.

Data Availability

The data used in this study are available from the corresponding author on a reasonable request.

Ethical Approval

This study was approved by the Ethics Committee of the University of Naples Federico II, and written informed consent was obtained from each participant.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Authors' Contributions

Rosario Megna, Mario Petretta, and Alberto Cuocolo conceptualized the study and drafted the manuscript. Rosario Megna, Roberta Assante, Emilia Zampella, Carmela Nappi, Valeria Gaudieri, Teresa Mannarino, Adriana D'Antonio, Roberta Green, Valeria Cantoni, Parthiban Arumugam, and Wanda Acampa collected and analyzed the data. All the authors revised and commented on the paper and approved the final version of the manuscript.

Acknowledgments

We would like to express our gratitude to the staff of the Division of Nuclear Medicine for their excellent technical support.

References

[1] P. Chen, P. C. Chen, Y. Liu, and L. Peng, "How to develop machine learning models for healthcare," *Nature Materials*, vol. 18, no. 5, pp. 410–414, 2019.

[2] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[3] A. Rajkomar, J. Dean, I. Kohane, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.

[4] R. Megna, A. Cuocolo, and M. Petretta, "Applications of machine learning in medicine," *Biomedical Journal of Scientific & Technical Research*, vol. 20, no. 5, pp. 15350–15352, 2019.

[5] L. M. Stevens, B. J. Mortazavi, R. C. Deo, L. Curtis, and D. P. Kao, "Recommendations for reporting machine learning analyses in clinical research," *Circulation: Cardiovascular Quality and Outcomes*, vol. 13, no. 10, article e006556, 2020.

[6] C. Ricciardi, R. Cuocolo, R. Megna, M. Cesarelli, and M. Petretta, "Machine learning analysis: general features, requirements and cardiovascular applications," *Minerva Cardiology and Angiology*, 2021.

[7] G. A. Diamond and J. S. Forrester, "Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease," *New England Journal of Medicine*, vol. 300, no. 24, pp. 1350–1358, 1979.

[8] T. S. Genders, E. W. Steyerberg, M. G. Hunink et al., "Prediction model to estimate presence of coronary artery disease: retrospective pooled analysis of existing cohorts," *BMJ*, vol. 344, no. jun12 1, article e3485, 2012.

[9] J. Reeh, C. B. Therming, M. Heitmann et al., "Prediction of obstructive coronary artery disease and prognosis in patients with suspected stable angina," *European Heart Journal*, vol. 40, no. 18, pp. 1426–1435, 2019.

[10] R. Megna, R. Assante, E. Zampella et al., "Pretest models for predicting abnormal stress single-photon emission computed tomography myocardial perfusion imaging," *Journal of Nuclear Cardiology*, 2019.

[11] R. Megna, C. Nappi, V. Gaudieri et al., "Diagnostic value of clinical risk scores for predicting normal stress myocardial

perfusion imaging in subjects without coronary artery calcium," *Journal of Nuclear Cardiology*, 2020.

[12] L. E. Juarez-Orozco, A. Saraste, D. Capodanno et al., "Impact of a decreasing pre-test probability on the performance of diagnostic tests for coronary artery disease," *European Heart Journal - Cardiovascular Imaging*, vol. 20, no. 11, pp. 1198–1207, 2019.

[13] A. Akella and S. Akella, "Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution," *Future Science OA*, vol. 7, no. 6, article FSO698, 2021.

[14] R. Megna, M. Petretta, B. Alfano et al., "A new relational database including clinical data and myocardial perfusion imaging findings in coronary artery disease," *Current Medical Imaging*, vol. 15, no. 7, pp. 661–671, 2019.

[15] Committee Members, R. J. Gibbons, G. J. Balady et al., "ACC/AHA 2002 guideline update for exercise testing: summary Article," *Circulation*, vol. 106, no. 14, pp. 1883–1892, 2002.

[16] D. S. Berman, A. Abidov, X. Kang et al., "Prognostic validation of a 17-segment score derived from a 20-segment score for myocardial perfusion SPECT interpretation," *Journal of Nuclear Cardiology*, vol. 11, no. 4, pp. 414–423, 2004.

[17] H. J. Verberne, W. Acampa, C. Anagnostopoulos et al., "EANM procedural guidelines for radionuclide myocardial perfusion imaging with SPECT and SPECT/CT: 2015 revision," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 42, no. 12, pp. 1929–1940, 2015.

[18] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[19] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148–156, Morgan Kaufmann, 1996.

[20] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall/CRC, London, UK, 2nd ed edition, 1989.

[21] A. McCallum and N. Kamal, "A comparison of event models for Naive Bayes text classification," in *AAAI-98 workshop on learning for text categorization*, Madison, Wisconsin, July 1998.

[22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Routledge, Wadsworth, 1984.

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[25] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco, California, USA, 2016.

[26] R. Megna, E. Zampella, R. Assante et al., "Temporal trends of abnormal myocardial perfusion imaging in a cohort of Italian subjects: relation with cardiovascular risk factors," *Journal of Nuclear Cardiology*, vol. 27, no. 6, pp. 2167–2177, 2020.

[27] R. Cuocolo, T. Perillo, E. De Rosa, L. Ugga, and M. Petretta, "Current applications of big data and machine learning in cardiology," *Journal of Geriatric Cardiology*, vol. 16, no. 8, pp. 601–607, 2019.

[28] Z. H. Hou, B. Lu, Z. N. Li et al., "Machine learning for pretest probability of obstructive coronary stenosis in symptomatic patients," *JACC Cardiovascular Imaging*, vol. 12, no. 12, pp. 2584–2586, 2019.

[29] F. Wang, W. Xu, W. Lv et al., "Evaluation of the diagnostic value of joint PET myocardial perfusion and metabolic imaging for vascular stenosis in patients with obstructive coronary artery disease," *Journal of Nuclear Cardiology*, 2020.

[30] A. Kilic, D. Dochtermann, R. Padman, J. K. Miller, and A. Dubrawski, "Using machine learning to improve risk prediction in durable left ventricular assist devices," *PLoS One*, vol. 16, no. 3, article e0247866, 2021.

[31] T. S. Dunn 2nd and F. G. Hage, "Stress myocardial perfusion imaging: can we tell the results without doing the test?," *Journal of Nuclear Cardiology*, 2020.

[32] Y. Feng, S. Deb, G. G. Wang, and A. H. Alavi, "Monarch butterfly optimization: a comprehensive review," *Expert Systems with Applications*, vol. 168, article 114418, 2021.

[33] G. G. Wang, S. Deb, and L. D. S. Coelho, "Earthworm optimization algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *International Journal of Bio-Inspired Computation*, vol. 1, no. 1, p. 1, 2015.

[34] G. G. Wang, S. Deb, X. Z. Gao, and L. D. S. Coelho, "A new metaheuristic optimisation algorithm motivated by elephant herding behaviour," *International Journal of Bio-Inspired Computation*, vol. 8, no. 6, p. 394, 2016.

[35] M. A. Elhosseini, R. A. El Sehiemy, Y. I. Rashwan, and X. Z. Gao, "On the performance improvement of elephant herding optimization algorithm," *Knowledge-Based Systems*, vol. 166, pp. 58–70, 2019.

[36] G. G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, no. 2, pp. 151–164, 2018.

[37] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: a new method for stochastic optimization," *Future Generation Computer Systems.*, vol. 111, pp. 300–323, 2020.

[38] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 6401105, 10 pages
https://doi.org/10.1155/2021/6401105

Research Article

Application of Bayesian Decision Tree in Hematology Research: Differential Diagnosis of β-Thalassemia Trait from Iron Deficiency Anemia

Mina Jahangiri,[1] **Fakher Rahim**,[2] **Najmaldin Saki**,[2] **and Amal Saki Malehi**[2,3]

[1]*Ph.D. Student, Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran*
[2]*Thalassemia & Hemoglobinopathy Research Center, Research Institute of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran*
[3]*Department of Biostatistics and Epidemiology, Faculty of Public Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran*

Correspondence should be addressed to Amal Saki Malehi; amalsaki@gmail.com

Received 2 June 2021; Revised 21 September 2021; Accepted 11 October 2021; Published 9 November 2021

Academic Editor: Giovanni D Addio

Objective. Several discriminating techniques have been proposed to discriminate between β-thalassemia trait (βTT) and iron deficiency anemia (IDA). These discrimination techniques are essential clinically, but they are challenging and typically difficult. This study is the first application of the Bayesian tree-based method for differential diagnosis of βTT from IDA. *Method*. This cross-sectional study included 907 patients with ages over 18 years old and a mean (±SD) age of 25 ± 16.1 with either βTT or IDA. Hematological parameters were measured using a Sysmex KX-21 automated hematology analyzer. Bayesian Logit Treed (BLTREED) and Classification and Regression Trees (CART) were implemented to discriminate βTT from IDA based on the hematological parameters. *Results*. This study proposes an automatic detection model of beta-thalassemia carriers based on a Bayesian tree-based method. The BLTREED model and CART showed that mean corpuscular volume (MCV) was the main predictor in diagnostic discrimination. According to the test dataset, CART indicated higher sensitivity and negative predictive value than BLTREED for differential diagnosis of βTT from IDA. However, the CART algorithm had a high false-positive rate. Overall, the BLTREED model showed better performance concerning the area under the curve (AUC). *Conclusions*. The BLTREED model showed excellent diagnostic accuracy for differentiating βTT from IDA. In addition, understanding tree-based methods are easy and do not need statistical experience. Thus, it can help physicians in making the right clinical decision. So, the proposed model could support medical decisions in the differential diagnosis of βTT from IDA to avoid much more expensive, time-consuming laboratory tests, especially in countries with limited recourses or poor health services.

1. Introduction

Iron deficiency anemia (IDA) and β-thalassemia trait (βTT) are the two most common hypochromic microcytic anemia. βTT is more prevalent in the Mediterranean region, in specific geographical areas, including the Caspian Sea and Persian Gulf regions; the 10% prevalence was reported [1]. The differential between βTT from IDA is crucial for preventing iron overload and related complications caused by misdiagnosis and inaccurate treatment [2].

Differentiation of β-thalassemia trait from iron deficiency anemia is also essential for premarital counseling in developed countries; for patients with microcytic anemia, complete blood count (CBC), in conjunction with hemoglobin variant analysis by high-performance liquid chromatography (HPLC), is interpreted to differentiate iron deficiency

from thalassemia traits. Then, iron studies and molecular testing are also performed. Hemoglobin electrophoresis, serum iron, and ferritin levels are considered to make a definitive differential diagnosis between βTT and IDA [3–5].

However, in low-resource settings where HPLC and molecular testing are not available, different studies proposed discrimination indices to distinct between βTT and IDA. These indices have been defined to quickly discriminate between IDA and βTT and avoid more time-consuming and expensive methods. Mentzer [3], Shine and Lal [4], England and Fraser [5], RBC [6], Srivastava and Bevington [7], Ricerca et al. [8], Green and King [9], Bessman and Feinstein (RDW) [10], Gupta et al. [11], Jayabose et al. (RDWI) [12], Telmissani-MCHD [13], Telmissani-MDHL [13], Huber-Herklotz [14], Kerman I [15], Kerman II [15], Sirdah et al. [16], Ehsani et al. [17], Keikhaei [18], Nishad et al. [19], Wongprachum et al. [20], Dharmani et al. [21], Pornprasert et al. [22], Sirachainan et al. [23], Bordbar et al. [24], Matos et al. [25], Janel (11T) [26], CRUISE Index [27], and Index26 [27] are all hematological discrimination indices used for discriminating between the IDA and the βTT. However, these indices were obtained empirically and have an inconsistent performance for differential diagnosis of βTT and IDA in the same patient [28]. On the other hand, sometimes, the same indices showed different discrimination power in varied age groups [29, 30].

Recently, the accessibility of powerful statistical software has provided data mining techniques for health-related data. Many studies have proposed advanced statistical methods and data mining techniques such as decision tree methods [31] for differential diagnostic between βTT and IDA to avoid much more expensive, time-consuming, and complicated laboratory procedures and nonsatisfactory hematological indices in discriminating between βTT and IDA [32–38]. [32, 35–39]. Urrechaga, Aguirre, and Izquierdo [39] used multivariable discriminant analysis for differential diagnosis of microcytic anemia. Wongseree et al. [37] implemented neural network and genetic programming for thalassemia classification. Dogan and Turkoglu [35] proposed a decision tree for detecting iron deficiency anemia from hematology parameters.

Jahangiri et al. [32] used classic decision-tree-based methods for constructing a differential diagnosis scheme and investigating the performance of several tree-based methods for the differential diagnosis of βTT from IDA. Decision trees have advantages over traditional statistical methods like discriminant analysis and generalized linear models (GLMs). The main advantage of tree-based methods is a tree structure that makes it easy to interpret the clinical data and be accepted by medical researchers and clinicians. CART is one of the best-known classic tree algorithms. However, this algorithm suffers from some problems such as greediness, instability, and bias in split rule selection. Bayesian tree approaches were proposed to solve the greediness of the CART algorithm. The greedy search algorithm has disadvantages such as limit the exploration of tree space, the dependence of future splits to previous splits, generate optimistic error rates, and the inability of the search to find a global optimum [40]. Also, the Bayesian approaches can quantify uncertainty and explore the tree space more than classic tree approaches. Bayesian approaches combine prior information with observations, unlike classic tree methods (these methods use only observations for data analysis). The Bayesian approaches define prior distributions on the components of classic tree methods and then use stochastic search algorithms through Markov Chain Monte Carlo (MCMC) algorithms for exploring tree space [41–47]. So, in the last two decades, many studies have developed Bayesian Treed Generalized Linear Models. These models fit a parametric model such as GLMs instead of using constant models in each tree node. So, these treed algorithms create smaller trees than tree models and improve the tree's interpretation [43].

This paper aims to compare the Bayesian Treed Generalized Linear Models and CART for the differential diagnosis of βTT from IDA based on simple laboratory test results. The outcome variable of the present study is qualitative, so we must use the Bayesian Logit Treed (BLTREED) algorithm for discrimination between these two disorders. This Bayesian treed model fits the logistic regression model in each tree node for data prediction and uses the Metropolis-Hastings algorithm for exploring tree space.

2. Material and Methods

2.1. Criteria for Selecting Patient Groups. In this study, a total of 907 patients aged over 18 years old diagnosed with IDA ($n = 370$) or βTT ($n = 537$) were selected. The mean (±SD) age of the patients was 25 ± 16.1 years. Most of the patients ($n = 592$ (65%)) were women, and 315 (35%) were men.

CBC analysis of EDTA-K2 anticoagulated blood samples was performed using the Sysmex KX-21 automated hematology analyzer (Japan) to measure differential parameters. Hematological parameters like hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), Red Blood Cell Distribution Width (RDW), Mean Corpuscular Hemoglobin Concentration (MCHC), and Red Blood Cell count (RBC) were measured for all patients.

2.2. Inclusion Criteria. In the IDA group, patients had hemoglobin (Hb) levels less than 12 and 13 g/dl for women and men, respectively. Mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) were below 80 fl and 27 pg for both sexes, respectively, and for men, ferritin of <28 ng/ml was considered as IDA. In the βTT group, patients had an MCV value below 80 fl. Patients with HbA2 levels of >3.5% were considered as βTT carriers.

2.3. Exclusion Criteria. In the IDA group, the patients who had mutations associated with αTT (3.7, 4.2, 20.5, MED, SEA, THAI, FIL, and Hph) were excluded. For the βTT group, patients with αTT confirmed by mutations in the molecular analysis were excluded. All patients with malignancies or inflammatory/infectious diseases were also excluded.

2.4. Ethical Consideration. This study was approved and supported by the Ethical committee affiliated with the Ahvaz Jundishapur University of Medical Sciences (AJUMS), Ahvaz, Iran. Written informed consent was filled before the enrollment.

2.5. Machine Learning Analysis. Tree-based machine-learning methods are valuable tools in data mining techniques. These methods empower predictive models and could provide a

solution for constructing the diagnostic test with high accuracy [48, 49]. Tree-based models do not need any assumptions about the functional form of the data.

One of the advantages of these methods is the graphical presentation of results that make them easy to interpret and no need for statistical experience for the understanding result of models [50–53]. Tree-based models also were constructed based on Bayesian algorithms. Chipman et al. proposed the Bayesian approach of the CART model (BCART) with defining a prior distribution. Chipman et al. also developed the Bayesian Logit Treed (BLTREED) model as an extension of BCART. The BLTREED model fits a logistic regression model for data prediction in the terminal nodes [43, 54].

2.5.1. Bayesian Logit Treed (BLTREED) Model. The Bayesian approach (BCART) was implemented by using a prior distribution on the two components (Θ, T) of the CART model; T is a binary tree with $\mathcal{K}$ terminal nodes or tree with size $\mathcal{K}$, and $\Theta = (\theta_1, \theta_2, \cdots, \theta_{\mathcal{K}})$ is the parameter set in the terminal nodes ($\theta_i = p_{ij}$, $i = 1, \cdots, \mathcal{K}$, $j = 1, \cdots, N$: the number of distinct classes of the response variable and p_{ij} shows the probability of the jth class of response variable in ith terminal node). The joint posterior distribution of parameters and tree structure was as the following equation:

$$p(\Theta, T) = p(\Theta|T)p(T), \tag{1}$$

where $p(T)$ and $p(\Theta \mid T)$ show the prior distributions for tree and parameters in terminal nodes, respectively.

Usually, the Bayesian approach defines prior distributions as unknown; so, tree structure and parameters in terminal nodes were considered unknown [42]. BCART was extended by fitting a parametric model such as a logistic regression model for data prediction and describing the conditional distribution of $Y|X$ in each terminal node [43, 54]. In the BLTREED model, the conditional distribution of $Y|X$, unlike the BCART model, depends on X $(Y \mid X \sim f(Y \mid X, \theta_i))$ and also by fitting sophisticated model at terminal nodes (by fitting logistic regression model for data prediction in each terminal node), smaller trees and more interpretable were generated. In the BLTREED model, one subset of X can be used to generate the tree and other subsets were used to fit models in terminal nodes (these subsets can be joint and/or disjoint). In the Bayesian approach, $\theta_i = B_i$ shows the regression coefficients for the logistic model fitted in an ith terminal node.

The recursive stochastic process using a tree-generating stochastic process for tree growing $(p(T))$ is as follows [42, 43]:

(1) Start from T that has only a root node (terminal node η)

(2) Calculate the probability for splitting node η as follows:

$$P_{\text{Split}} = \alpha\left(1 + d_\eta\right)^{-\beta}, \tag{2}$$

where d_η is the depth of the node η, α is the base probability of tree growth of splitting a node, and β is the rate that determines the propensity to split decreases with increased tree size.

Actually, α and β are parameters that control the shape and size of trees, and these parameters provide a penalty to avoid an overfitting model

(3) If the node η splits into left and right nodes according to the distribution of $p_{\text{RULE}}(\rho \mid \eta, T)$, then let T as the newly created tree from step 3 and reapply steps 2 and 3 to the new children nodes

The BLTREED model was fitted based on standardized data. So, the same prior distribution can be used independently for parameters in the terminal nodes, and they were considered a multivariate normal distribution with zero mean and variance matrix proportional to the identity for these parameters [43, 54].

Posterior distribution function $p(T \mid X, y)$ was computed by combining the marginal likelihood function $p(Y \mid X, T)$ and tree prior $p(T)$ as follows:

$$P(T \mid X, y) \propto p(y \mid X, T)p(T). \tag{3}$$

In this study, no informative priors were considered. The priors were uniform on variables at a particular node, and all possible splits for variables.

Where $p(Y \mid X, T)$ is as follows:

$$\begin{aligned} P(Y \mid X, T) &= \int p(y \mid X, \Theta, T)p(\Theta \mid T)\, d\Theta \\ &= \prod_{i=1}^{\mathcal{K}} \int \prod_{h=1}^{n_i} p(y_{ih} \mid x_{ih}, B_i)p(B_i)dB_i, \end{aligned} \tag{4}$$

which $p(y \mid X, \Theta, T)$, (y_{ih}, x_{ih}), and n_i show the data likelihood function, observed values for hth observation in ith node, and the number of observations in ith node, respectively. The integral of equation four has no closed form, so the Laplace approximation was used to solve it [43, 54].

Chipman et al. [42, 43] utilize a Metropolis-Hastings algorithm to simulate equation (3) for finding trees with the high posterior distribution. The Metropolis-Hastings algorithm simulates a Markov chain sequence of trees, namely, $T^0, T^1, T^2, \cdots$.

The simulation algorithm was implemented with multiple restarts for reasons mentioned in Chipman et al. [42, 43].

Table 1: Comparison between hematological parameters of study groups using the Mann–Whitney U test (data are presented as median (IQR)).

	βTT ($n = 537$)	IDA ($n = 370$)	P
MCV (fl)	62 (5.4)	72.2 (9.7)	<0.001
MCH (pg)	19.6 (1.8)	21.9 (4.2)	<0.001
Hb (g/dl)	11 (1.6)	10.5 (2.6)	<0.001
RDW (%)	15.7 (1.7)	15.7 (3.3)	0.94

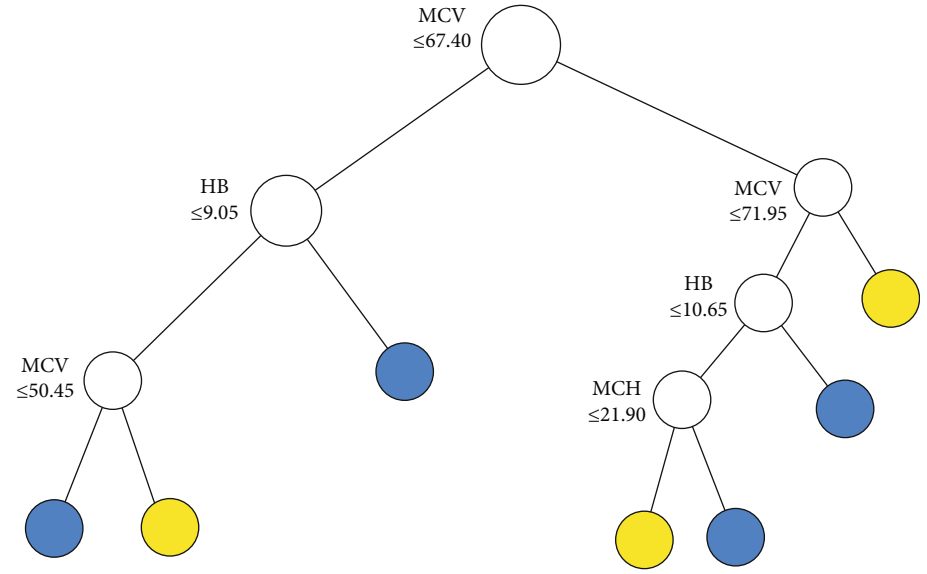


Figure 1: The tree structure of the CART algorithm based on the Gini index (blue terminal node: βTT and yellow terminal node: IDA).

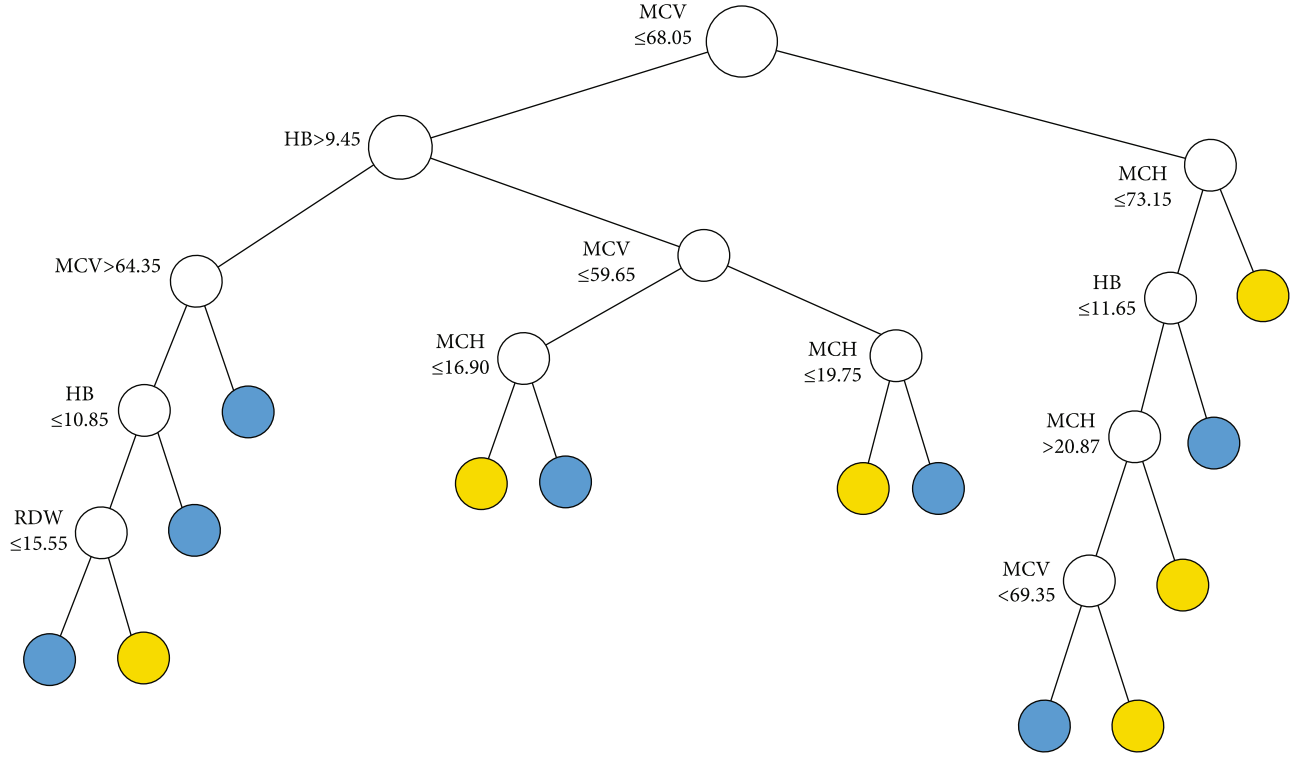


Figure 2: The tree structure of the CART algorithm based on the entropy index (blue terminal node: βTT and yellow terminal node: IDA).

2.5.2. Classification and Regression Trees (CART). Breiman et al. proposed the CART model [55]. The CART algorithm generates a tree using a binary recursive partitioning, and the tree-generating process contains four steps: (1) tree growing: tree growth is based on a greedy search algorithm, and this algorithm generates a tree by sequentially choosing splitting rules. The CART algorithm uses traditional splitting functions for choosing splitting rules (entropy and Gini index). (2) Tree-growing process continues until none of the nodes can split. (3) Tree pruning: this tree algorithm uses the cost-complexity pruning method for tree pruning to avoid overfitting. This pruning method generates a sequence of pruned trees, and each tree in this sequence is an extension of previous trees. (4) Best tree selection: CART uses an

independent test dataset or cross-validation to estimate the prediction error of each tree and then selects the best tree with the lowest estimated prediction error.

2.6. Data Analysis. The BLTREED model and classic CART algorithm based on the two splitting functions like entropy and Gini index (after that, we named the CART method-based Gini index as CART1 and CART method-based entropy as CART2) were fitted by using predictor variables such as hemoglobin (Hb), mean cell volume (MCV), mean cell hemoglobin (MCH), and red cell distribution width (RDW) for differential diagnosis of βTT from IDA.

The BLTREED model fitted using eight restarts with 6000 iterations per restart and a prior standard deviation of 20 for the logit coefficients [54]. For determining the pair of (α, β), the BLTREED model was fitted with two choices, 0.5 and 0.95 for the α parameter, and four choices for β (a range 0.5-2 by step 0.5), then select the pair of (α, β) that generate the best tree with smallest FNR.

Based on the acceptable method of cross-validation in machine learning studies, for assessing the performance of the three models, the dataset was split randomly in the ratio 2 : 1 into a training and a test dataset, respectively, using a stratified random sample to ensure equal allocation of presences and absences (for a classification tree). The model was then fit to the training dataset, and the set of the best trees was determined. For each tree, the posterior predictive distribution was computed for both the training data and the test dataset; this was implemented for each iteration of the BLTREED algorithms, thus incorporating the uncertainty of the model parameters and the data in the evaluation of models. Finally, the predictive performances were calculated based on the confusion matrix of the posterior predictive distribution for both the training and the test dataset [43, 47, 54, 56, 57].

Differential performance of the Bayesian classification tree and CART was evaluated using criteria such as sensitivity (TPR), specificity (TNR), false-negative rate (FNR) and false-positive rate (FPR), positive predictive value (PPV) and negative predictive value (NPV), positive likelihood ratio (PLR) and negative likelihood ratio (NLR), accuracy, Youden's index, and the area under the curve (AUCROC). AUCROC represents the degree of separate ability showing how much the machine learning model can distinguish between the classes (IDA and βTT); actually, it is a global measure of diagnostic accuracy. A perfect classification algorithm has an AUCROC = 1. The interpretation of the AUCROC is described as follows: AUCROC > 0.9: excellent differentiation, AUCROC > 0.8: very good differentiation, AUCROC > 0.7: good differentiation, AUCROC > 0.6: sufficient differentiation, AUCROC > 0.5: bad differentiation, and AUCROC < 0.5: classification method is not useful for discriminating between IDA and β TT [58, 59]. Criteria such as Youden's index, accuracy, PLR, NLR (an excellent diagnostic test has NLR < 0.1 and PLR > 10), and AUC take both sensitivity and specificity into consideration, so that can present the performance of the model more accurately than other criteria. In addition, AUC values were compared using DeLong et al. method [60]. A P value < 0.05 was considered a statistically significant difference.

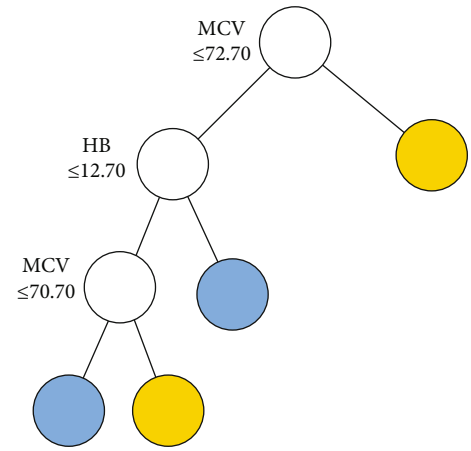


Figure 3: Decision tree for the BLTREED model ($\alpha = 0.95$, $\beta = 1$, Log integrated likelihood = 123.43) (blue terminal node: βTT and yellow terminal node: IDA).

Table 2: Confusion table of the BLTREED model and CART algorithm for training dataset and test dataset.

Dataset	Algorithm	Disease status	TP	FP	FN	TN	(TP+TN)
Training	BLTREED	βTT	363	25	13	234	597
		IDA	234	13	25	363	
	CART1	βTT	366	46	10	213	579
		IDA	213	10	46	366	
	CART2	βTT	358	23	18	236	594
		IDA	236	18	23	358	
Test	BLTREED	βTT	155	8	6	103	258
		IDA	103	6	8	155	
	CART1	βTT	160	33	1	78	238
		IDA	78	1	33	160	
	CART2	βTT	159	12	2	99	258
		IDA	99	2	12	159	

2.7. Software. Data were analyzed by free software (http://gsbwww.uchicago.edu.fac.robert.mcculloch.research.code.CART.index.html) based on Chipman et al. (2002) that was developed for fitting BLTREED model, R 3.0.3 used for fitting CART algorithm (package rpart), computing performance measures (package ePiR and package pROC), and splitting data to training dataset and test dataset (package caTools).

3. Results

A total of 537 patients were diagnosed as βTT with an average of age (±SD) 22 ± 16.4 including 299 (56%) women and 238 (44%) men, while 370 patients (mean of age (±SD): 29 ± 14.6) were diagnosed as IDA including 293 (79%) women and 77 (21%) men. Table 1 shows the median and interquartile range (IQR) of laboratory parameters as predictor variables across the type of hypochromic microcytic anemia (βTT and IDA).

Table 3: Sensitivity (TPR), specificity (TNR), false-positive rate (FPR), false-negative rate (FNR), positive predictive value (PPV), negative predictive value (NPV), accuracy, Youden's index, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) of the BLTREED model in prediction of IDA and βTT groups and their 95% exact confidence interval for training and test dataset.

	BLTREED		CART1		CART2	
Accuracy measure	Training dataset	Test dataset	Training dataset	Test dataset	Training dataset	Test dataset
TPR	97 (94, 98)	96 (92, 99)	97 (95, 99)	99 (97, 100)	95 (93, 97)	99 (96, 100)
TNR	90 (86,94)	93 (86, 97)	82 (77, 87)	70 (61, 79)	91 (87, 94)	89 (82, 94)
FNR	3 (2, 6)	4 (1, 8)	3 (1, 5)	1 (0, 3)	5 (3, 7)	1 (0, 4)
FPR	10 (6,14)	7 (3, 14)	18 (13, 23)	30 (21, 39)	9 (6, 13)	11 (6, 18)
PPV	94 (91, 96)	95 (91, 98)	89 (85, 92)	83 (77, 88)	94 (91, 96)	93 (88, 96)
NPV	95 (91, 97)	94 (88, 98)	96 (92, 98)	99 (93, 100)	93 (89, 96)	98 (93, 100)
Youden's index	87 (80, 92)	89 (78, 95)	80 (72, 85)	70 (57, 79)	86 (80, 91)	88 (77, 94)
Accuracy	94 (92,96)	95 (91, 97)	91 (89, 93)	87 (83, 91)	93 (91, 95)	95 (91, 97)
PLR	10 (7, 14)	13.36 (7, 26)	5.48 (4, 7)	3.34 (2, 4)	10.72 (7, 16)	9.14 (5, 16)
NLR	0.04 (0.02, 0.07)	0.04 (0.02, 0.09)	0.03 (0.02, 0.06)	0.01 (0, 0.06)	0.05 (0.03, 0.08)	0.01 (0, 0.06)

Table 4: The area under ROC curve (AUC) of BLTREED and CART algorithms in the prediction of IDA and βTT groups for training and test dataset (SE: standard error of AUC; CI: confidence interval).

	BLTREED		CART1		CART2	
	Training dataset	Test dataset	Training dataset	Test dataset	Training dataset	Test dataset
AUC	0.99	0.98	0.93	0.94	0.97	0.97
SE	0.003	0.009	0.011	0.015	0.006	0.011
95% CI	(0.98, 0.99)	(0.96, 0.99)	(0.90, 0.95)	(0.91, 0.97)	(0.96, 0.99)	(0.95, 1)
P value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

The tree structure of CART1, CART2, and BLTREED models is shown in Figures 1–3, respectively. The first split of the three methods of classification trees was based on MCV, which showed that MCV has a higher importance value in differentiation between the βTT and the IDA. Another predictor that was used as the second splitting variable in tree structure was HB. According to the presented trees, the BLTREED model produced a smaller tree size and was more interpretable than the CART algorithm (Figures 1 and 2). This model showed values of MCV ≤ 72.6 screening the βTT patients. The BLTREED model extracted four homogenous subgroups for differentiating between the βTT and the IDA (Figure 3).

The predictive performance of models in differentiation between βTT and IDA was calculated based on the confusion matrix (Table 2). The BLTREED model, CART1, and CART2 trees showed the high TPR, TNR, PPV, NPV, Youden's Index, and accuracy in differentiation between βTT and IDA (Table 3). However, the BLTREED model had a higher accuracy and Youden's index other than CART1 and CART2.

In addition, all the models have NLR < 0.1 that three classification tree algorithms have good diagnostic accuracy for discriminating the patients. Table 4 shows the AUCs of the three tree models from ROC analysis that were statistically significant ($P < 0.001$) and revealed that all three classification methods had an excellent diagnose accuracy (AUC > 0.9: excellent differentiation) in differentiation between the βTT and the IDA. In addition, Figure 4 displays the receiver operating characteristic curves of the BLTREED model, CART1, and CART2 algorithms for the test dataset, and the comparisons of AUC values between the models. According to the exhibited figure, there was no significant difference between the methods ($P > 0.05$).

4. Discussion

In this paper, we used the BLTREED model as the differential diagnostic tool for thalassemia diagnosis. In addition, we compare the predictive performance of the BLTREED model

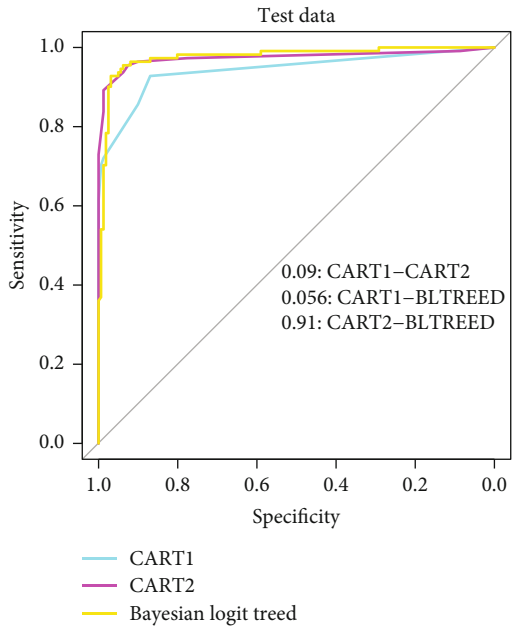


Figure 4: Receiver operating characteristic curves of BLTREED and CART algorithms in the prediction of IDA and βTT groups for test dataset.

as a Bayesian decision tree with the CART algorithm. It is the first study that uses the BLTREED model in the hematological data.

The Bayesian decision tree was used to solve uncertain problems of conventional tree-based methods [43, 54, 61]. This model was implemented by using Hb, MCV, MCH, and RDW as independent variables.

Our dataset included 537 (59%) patients with βTT and 293 (41%) patients with IDA. However, there was not any degree of relative imbalance between the IDA and βTT classes. [62, 63].

Based on our result, MCV and Hb were the main predictor parameters in differential diagnostic, and it showed that the patient with βTT has lower values of MCV.

In previous studies that used the different conventional decision trees for differential diagnosis βTT from IDA, the first split of all algorithms was based on MCV. They also concluded that MCV was a significant predictor variable in the discrimination of IDA and βTT [32, 36]. The performance of the BLTREED model that was evaluated using sensitivity, specificity, false-negative and positive rate, and positive and negative predictive value exhibited the high performance of the differential diagnosis of βTT from IDA. In addition, positive likelihood ratio, negative likelihood ratio, accuracy, and Youden's index showed that BLTREED has good diagnostic accuracy for discriminating the patients. It was indeed classified as 96% of βTT patients. Furthermore, AUC as an overall performance index showed excellent and significant accuracy (99, 98) in training and test data, respectively, in differential diagnostic of βTT and IDA. BLTREED has also generated a tree with a smaller size, and it is more interpretable other than the CART algorithms and indicated better diagnostic performance.

Our study has a limitation, which should be considered. The investigated patients have included just IDA and βTT cases and excluded concomitant diseases and αTT cases. Therefore, considering αTT patients in the study would affect the performance of the presented models and changed the interpretation of the result. Particularly when only simple hematologic parameters are used like in the present study, it may be difficult to distinguish αTT from βTT.

Other studies that used different data mining techniques and decision trees based on the frequentist approach of fitting revealed the high performance and accuracy but lower than our result [32, 34–36, 38]. In many studies which had imbalanced datasets, Oversampling Technique (SMOTE) was applied for handling this problem [34, 64].

The BLTREED model improves the classification performance by solving the uncertainty of previous models [43, 54]. The diagnostic performance of the BLTREED was better than other discrimination methods (classification trees or hematological discrimination indices) in past studies for differentiating βTT from IDA. These studies are as follows: Setsirichok et al. used a C4.5 decision tree, naïve Bayes (NB) classifier, and multilayer perceptron (MLP) for classifying eighteen classes of thalassemia abnormality [38]. Bellinger et al. used classification algorithms like the J48 decision tree, support vector machines (SVM), k-nearest neighbors (k-NN), MLP, and NB for differentiating between βTT, IDA, and cooccurrence of these disorders. In this study, the imbalanced dataset was a cause for the weaker performance [34]. AlAgha et al. compared the diagnostic performance of different classification algorithms such as J48, k-NN, artificial neural networks (ANN), and NB for classifying β-thalassemia carriers. They showed that SMOTE helped decrease the problem of highly imbalanced class distribution and consequently improved the predictive performance [64]. Jahangiri et al. utilized classification tree algorithms such as CHAID, E-CHAID, CART, QUEST, GUIDE, and CRUISE for differential diagnosis of βTT from IDA. They indicated that the CRUISE algorithm has the best diagnostic performance similar to the present study, but this classic algorithm uses the greedy algorithm for tree generating and cannot explore the tree space more than the Bayesian tree approaches. Also, many studies compared the diagnostic performance of hematological discrimination indices, and BLTREED showed better performance in comparison to them [16–19, 23, 25–30, 65–80].

5. Conclusion

In the present study, the BLTREED model showed excellent diagnostic accuracy for differentiating βTT from IDA. According to the advantages of Bayesian tree-based methods like generating a small and more interpretable tree, and lack of uncertainty of different conventional decision trees, this method can be helpful along with other laboratory parameters for discriminating between these two anemia disorders. Also, understanding tree-based methods are easy and do not need statistical experience. So, it can help physicians in making the right clinical decision.

Abbreviations

βTT: β-Thalassemia trait
IDA: Iron deficiency anemia
MCV: Mean corpuscular volume
MCH: Mean corpuscular hemoglobin
RDW: Red Blood Cell Distribution Width
MCHC: Mean corpuscular hemoglobin concentration
RBC: Red blood cell
BLTREED: Bayesian Logit Treed
TPR: Sensitivity
TNR: Specificity
FNR: False-negative rate
FPR: False-positive rate
NPV: Negative predictive value
PPV: Positive predictive value
PLR: Positive likelihood ratio
NLR: Negative likelihood ratio.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethical Approval

This study was approved by the Ethics Committee of Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran (IR.AJUMS.REC.1395.456).

Disclosure

This paper is part of the thesis of Mina Jahangiri, MSc student of Biostatistics (no. U-95095).

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

ASM and MJ performed the conception and design, analysis and interpretation of the data, and drafting of the article. FR and NS performed the conception and design, collection and assembly of data, and drafting of the article. All authors approved the final version of the article for submission.

Acknowledgments

This paper was supported by the vice chancellor for Research Affairs of Ahvaz Jundishapur University of Medical Sciences.

References

[1] A. Batebi, A. Pourreza, and R. Esmailian, "Discrimination of beta-thalassemia minor and iron deficiency anemia by screening test for red blood cell indices," *Turkish Journal of Medical Sciences.*, vol. 42, no. 2, pp. 275–280, 2012.

[2] L. Hallberg, "Iron requirements," *Biological Trace Element Research*, vol. 35, no. 1, pp. 25–45, 1992.

[3] W. Mentzer, "Differentiation of iron deficiency from thalassaemia trait," *The Lancet.*, vol. 301, no. 7808, p. 882, 1973.

[4] I. Shine and S. Lal, "A strategy to detect β-thalassaemia minor," *The Lancet.*, vol. 309, no. 8013, pp. 692–694, 1977.

[5] J. England and P. Fraser, "Differentiation of iron deficiency from thalassaemia trait by routine blood-count," *The Lancet.*, vol. 301, no. 7801, pp. 449–452, 1973.

[6] G. G. Klee, V. F. Fairbanks, R. V. Pierre, and M. B. O'sullivan, "Routine erythrocyte measurements in diagnosis of iron-deficiency anemia and thalassemia minor," *American Journal of Clinical Pathology*, vol. 66, no. 5, pp. 870–877, 1976.

[7] P. Srivastava and J. Bevington, "Iron deficiency and/or thalassaemia trait," *The Lancet.*, vol. 301, no. 7807, p. 832, 1973.

[8] B. Ricerca, S. Storti, G. d'Onofrio et al., "Differentiation of iron deficiency from thalassaemia trait: a new approach," *Haematologica*, vol. 72, no. 5, pp. 409–413, 1986.

[9] R. Green and R. King, "A new red cell discriminant incorporating volume dispersion for differentiating iron deficiency anemia from thalassemia minor," *Blood Cells*, vol. 15, no. 3, pp. 481–495, 1989.

[10] J. D. Bessman and D. Feinstein, "Quantitative anisocytosis as a discriminant between iron deficiency and thalassemia minor," *Blood*, vol. 53, no. 2, pp. 288–293, 1979.

[11] A. D. Gupta, C. Hegde, and R. Mistri, "Red cell distribution width as a measure of severity of iron deficiency in iron deficiency anaemia," *The Indian Journal of Medical Research*, vol. 100, pp. 177–183, 1994.

[12] S. Jayabose, J. Giamelli, O. Levondoglu Tugal, C. Sandoval, F. Ozkaynak, and P. Visintainer, "# 262 differentiating iron deficiency anemia from thalassemia minor by using an RDW-based index," *Journal of Pediatric Hematology/Oncology*, vol. 21, no. 4, p. 314, 1999.

[13] O. A. TELMISSANI, S. KHALIL, and G. T. ROBERTS, "Mean density of hemoglobin per liter of blood: a new hematologic parameter with an inherent discriminant function," *Laboratory Hematology*, vol. 5, pp. 149–152, 1999.

[14] A. R. Huber, C. Ottiger, L. Risch, S. Regenass, M. Hergersberg, and R. Herklotz, "Thalassemie-syndrome: klinik und diagnose," Schweiz Med Forum, 2004.

[15] N. KOHAN and M. Ramzi, "Evaluation of sensitivity and specificity of Kerman index I and II in screening beta thalassemia minor," 2008.

[16] M. Sirdah, I. Tarazi, E. Al Najjar, and H. R. Al, "Evaluation of the diagnostic reliability of different RBC indices and formulas in the differentiation of the β-thalassaemia minor from iron deficiency in Palestinian population," *International Journal of Laboratory Hematology*, vol. 30, no. 4, pp. 324–330, 2008.

[17] M. Ehsani, E. Shahgholi, M. Rahiminejad, F. Seighali, and A. Rashidi, "A new index for discrimination between iron deficiency anemia and beta-thalassemia minor: results in 284 patients," *Pakistan journal of biological sciences: PJBS.*, vol. 12, no. 5, pp. 473–475, 2009.

[18] B. Keikhaei, "A new valid formula in differentiating iron deficiency anemia from ß-thalassemia trait," *Pakist J Med Sci.*, vol. 26, pp. 368–373, 2010.

[19] A. A. N. Nishad, A. Pathmeswaran, A. Wickremasinghe, and A. Premawardhena, "The Thal-index with the BTT prediction.exe to discriminate β-thalassaemia traits from other microcytic anaemias," *Thalassemia Reports*, vol. 2, no. 1, 2012.

[20] K. Wongprachum, K. Sanchaisuriya, P. Sanchaisuriya, S. Siridamrongvattana, S. Manpeun, and F. P. Schlep, "Proxy

indicators for identifying iron deficiency among anemic vegetarians in an area prevalent for thalassemia and hemoglobinopathies," *Acta Haematologica*, vol. 127, no. 4, pp. 250–255, 2012.

[21] P. Dharmani, K. Sehgal, T. Dadu, R. Mankeshwar, A. Shaikh, and S. Khodaiji, "Developing a new index and its comparison with other CBC-based indices for screening of beta thalassemia trait in a tertiary care hospital," *International Journal of Laboratory Hematology*, vol. 35, p. 118, 2013.

[22] S. Pornprasert, A. Panya, M. Punyamung, J. Yanola, and C. Kongpan, "Red cell indices and formulas used in differentiation of β-thalassemia trait from iron deficiency in Thai school children," *Hemoglobin*, vol. 38, no. 4, pp. 258–261, 2014.

[23] N. Sirachainan, P. Iamsirirak, P. Charoenkwan et al., "New mathematical formula for differentiating thalassemia trait and iron deficiency anemia in thalassemia prevalent area: a study in healthy school-age children," *Southeast Asian Journal of Tropical Medicine and Public Health*, vol. 45, no. 1, pp. 174–182, 2014.

[24] E. Bordbar, M. Taghipour, and B. E. Zucconi, "Reliability of different RBC indices and formulas in discriminating between β-thalassemia minor and other causes of microcytic hypochromic anemia," *Mediterranean journal of hematology and infectious diseases*, vol. 7, no. 1, 2014.

[25] J. F. Matos, L. Dusse, K. B. Borges, R. L. de Castro, and W. Coura-Vital, "A new index to discriminate between iron deficiency anemia and thalassemia trait," *Revista Brasileira de Hematologia e Hemoterapia*, vol. 38, no. 3, pp. 214–219, 2016.

[26] A. Janel, L. Roszyk, C. Rapatel, G. Mareynat, M. G. Berger, and A. F. Serre-Sapin, "Proposal of a score combining red blood cell indices for early differentiation of beta-thalassemia minor from iron deficiency anemia," *Hematology*, vol. 16, no. 2, pp. 123–127, 2011.

[27] M. Jahangiri, F. Rahim, and A. S. Malehi, "Diagnostic performance of hematological discrimination indices to discriminate between βeta thalassemia trait and iron deficiency anemia and using cluster analysis: introducing two new indices tested in Iranian population," *Scientific Reports*, vol. 9, no. 1, pp. 1–13, 2019.

[28] A. Vehapoglu, G. Ozgurhan, A. D. Demir et al., "Hematological indices for differential diagnosis of beta thalassemia trait and iron deficiency anemia," *Anemia.*, vol. 2014, pp. 1–7, 2014.

[29] F. Rahim and B. Keikhaei, "Better differential diagnosis of iron deficiencyanemia from beta-thalassemia trait," *Turkish Journal of Hematology*, vol. 26, no. 3, pp. 138–145, 2009.

[30] J. J. Hoffmann, E. Urrechaga, and U. Aguirre, "Discriminant indices for distinguishing thalassemia and iron deficiency in patients with microcytic anemia: a meta-analysis," *Clinical Chemistry and Laboratory Medicine (CCLM).*, vol. 53, no. 12, pp. 1883–1894, 2015.

[31] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2/3, pp. 131–163, 1997.

[32] M. Jahangiri, E. Khodadi, F. Rahim, N. Saki, and A. Saki Malehi, "Decision-tree-based methods for differential diagnosis of β-thalassemia trait from iron deficiency anemia," *Expert Systems*, vol. 34, no. 3, 2017.

[33] M. Maity, T. Mungle, D. Dhane, A. K. Maiti, and C. Chakraborty, "An ensemble rule learning approach for automated morphological classification of erythrocytes," *Journal of Medical Systems*, vol. 41, no. 4, p. 56, 2017.

[34] C. Bellinger, A. Amid, N. Japkowicz, and H. Victor, "Multi-label classification of anemia patients," in *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*, IEEE, 2015.

[35] S. Dogan and I. Turkoglu, "Iron-deficiency anemia detection from hematology parameters by using decision trees," *International Journal of Science & Technology.*, vol. 3, no. 1, pp. 85–92, 2008.

[36] E. H. Elshami and A. M. Alhalees, "Automated diagnosis of thalassemia based on data mining classifiers. The International Conference on Informatics and Applications (ICIA 2012)," in The Society of Digital Information and Wireless Communication, 2012.

[37] W. Wongseree, N. Chaiyaratana, K. Vichittumaros, P. Winichagoon, and S. Fucharoen, "Thalassaemia classification by neural networks and genetic programming," *Information Sciences*, vol. 177, no. 3, pp. 771–786, 2007.

[38] D. Setsirichok, T. Piroonratana, W. Wongseree et al., "Classification of complete blood count and haemoglobin typing data by a C4.5 decision tree, a naive Bayes classifier and a multilayer perceptron for thalassaemia screening," *Biomedical Signal Processing and Control*, vol. 7, no. 2, pp. 202–212, 2012.

[39] E. Urrechaga, U. Aguirre, and S. Izquierdo, "Multivariable discriminant analysis for the differential diagnosis of microcytic anemia," *Anemia.*, vol. 2013, pp. 1–6, 2013.

[40] A. S. Malehi and M. Jahangiri, *Classic and Bayesian Tree-Based Methods*, Enhanced Expert Systems, 2019, Intech Open.

[41] D. G. Denison, B. K. Mallick, and A. F. Smith, "A Bayesian CART algorithm," *Biometrika*, vol. 85, no. 2, pp. 363–377, 1998.

[42] H. A. Chipman, E. I. George, and R. E. McCulloch, "Bayesian CART model search," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 935–948, 1998.

[43] H. Chipman, E. George, and R. McCulloch, "Bayesian treed generalized linear models," *Bayesian statistics.*, vol. 7, pp. 323–349, 2003.

[44] H. A. Chipman, E. I. George, and R. E. McCulloch, "Bayesian treed models," *Machine Learning*, vol. 48, no. 1/3, pp. 299–320, 2002.

[45] Y. Wu, H. Tjelmeland, and M. West, "Bayesian CART: prior specification and posterior simulation," *Journal of Computational and Graphical Statistics*, vol. 16, no. 1, pp. 44–66, 2007.

[46] R. A. O'Leary, J. V. Murray, S. J. Low Choy, and K. L. Mengersen, "Expert elicitation for Bayesian classification trees," *Journal of Applied Probability & Statistics.*, vol. 3, no. 1, pp. 95–106, 2008.

[47] W. Hu, R. A. O'Leary, K. Mengersen, and S. L. Choy, "Bayesian classification and regression trees for predicting incidence of cryptosporidiosis," *PLoS One*, vol. 6, no. 8, article e23903, 2011.

[48] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Chapman and Hall/CRC, New York, 1984.

[49] H. Zhang and B. Singer, *Recursive Partitioning and Applications. Second ed*, P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, and S. Zeger, Eds., Springer, New York, 2010.

[50] W. Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.*, vol. 1, no. 1, pp. 14–23, 2011.

[51] G. De'ath and K. E. Fabricius, "Classification and regression trees: a powerful yet simple technique for ecological data analysis," *Ecology*, vol. 81, no. 11, pp. 3178–3192, 2000.

[52] S. C. Lemon, J. Roy, M. A. Clark, P. D. Friedmann, and W. Rakowski, "Classification and regression tree analysis in public health: methodological review and comparison with logistic regression," *Annals of Behavioral Medicine*, vol. 26, no. 3, pp. 172–181, 2003.

[53] N. Speybroeck, D. Berkvens, A. Mfoukou-Ntsakala et al., "Classification trees versus multinomial models in the analysis of urban farming systems in Central Africa," *Agricultural Systems*, vol. 80, no. 2, pp. 133–149, 2004.

[54] W. W. Moe, H. Chipman, E. I. George, and R. E. McCulloch, "A Bayesian treed model of online purchasing behavior using in-store navigational clickstream," *revising for 2nd review at Journal of Marketing Research*, 2002.

[55] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, CRC Press, 1984.

[56] S. Saha, *Survival Analysis with Bayesian Additive Regression Trees and Its Application*, 2017, https://commons.lib.niu.edu/handle/10843/21175.

[57] H. A. Chipman, E. I. George, and R. E. McCulloch, "BART: Bayesian additive regression trees," *The Annals of Applied Statistics.*, vol. 4, no. 1, pp. 266–298, 2010.

[58] A.-M. Šimundić, "Measures of diagnostic accuracy: basic definitions," *Med Biol Sci.*, vol. 22, no. 4, pp. 61–65, 2008.

[59] L. Donisi, G. Cesarelli, P. Balbi et al., "Positive impact of short-term gait rehabilitation in Parkinson patients: a combined approach based on statistics and machine learning," *Mathematical Biosciences and Engineering*, vol. 18, no. 5, pp. 6995–7009, 2021.

[60] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.

[61] J. B. Gray and G. Fan, "Classification tree analysis using TARGET," *Computational Statistics and Data Analysis*, vol. 52, no. 3, pp. 1362–1372, 2008.

[62] K. Wang, C. A. Phillips, A. M. Saxton, and M. A. Langston, "EntropyExplorer: an R package for computing and comparing differential Shannon entropy, differential coefficient of variation and differential expression," *BMC Research Notes*, vol. 8, no. 1, pp. 1–5, 2015.

[63] Available from: https: //stats. http://stackexchange.com/questions/239973/a-general-measure-of-data-set-imbalance/239982.

[64] A. S. AlAgha, H. Faris, B. H. Hammo, and A.-Z. Ala'M, "Identifying _β_ -thalassemia carriers using a data mining approach: the case of the Gaza Strip, Palestine," *Artificial Intelligence in Medicine*, vol. 88, pp. 70–83, 2018.

[65] J. J. Hoffmann and E. Urrechaga, "Role of RDW in mathematical formulas aiding the differential diagnosis of microcytic anemia," *Scandinavian Journal of Clinical and Laboratory Investigation*, vol. 80, no. 6, pp. 464–469, 2020.

[66] E. Miri-Moghaddam and N. Sargolzaie, "Cut off determination of discrimination indices in differential diagnosis between iron deficiency anemia and β-thalassemia minor," *International journal of hematology-oncology and stem cell research.*, vol. 8, no. 2, pp. 27–32, 2014.

[67] A. Nesa, M. A. Tayab, T. Sultana et al., "RDWI is better discriminant than RDW in differentiation of iron deficiency anaemia and beta thalassaemia trait," *Bangladesh Journal of Child Health.*, vol. 33, no. 3, pp. 100–103, 2010.

[68] C. Beyan, K. Kaptan, and A. Ifran, "Predictive value of discrimination indices in differential diagnosis of iron deficiency anemia and beta-thalassemia trait," *European Journal of Haematology*, vol. 78, no. 6, pp. 524–526, 2007.

[69] M. Ghafouri, S. L. Mostaan, S. Sharifi, G. L. Hosseini, and C. Z. Atar, "Comparison of cell counter indices in differentiation of beta thalassemia minor from iron deficiency anemia," *The Scientific Journal of Iranian Blood Transfusion Organization (KHOON)*, vol. 2, no. 7, pp. 385–389, 2006.

[70] A. Demir, N. Yarali, T. Fisgin, F. Duru, and A. Kara, "Most reliable indices in differentiation between thalassemia trait and iron deficiency anemia," *Pediatrics International*, vol. 44, no. 6, pp. 612–616, 2002.

[71] M. Schoorl, M. Schoorl, J. Linssen et al., "Efficacy of advanced discriminating algorithms for screening on iron-deficiency anemia and β-thalassemia trait: a multicenter evaluation," *American Journal of Clinical Pathology*, vol. 138, no. 2, pp. 300–304, 2012.

[72] N. Tripathi, J. P. Soni, P. K. Sharma, and M. Verma, "Role of haemogram parameters and RBC indices in screening and diagnosis of beta-thalassemia trait in microcytic, hypochromic Indian children," *International Journal of Hematological Disorders.*, vol. 2, no. 2, pp. 43–46, 2015.

[73] I. L. Roth, B. Lachover, G. Koren, C. Levin, L. Zalman, and A. Koren, "Detection of β-thalassemia carriers by red cell parameters obtained from automatic counters using mathematical formulas," *Mediterranean journal of hematology and infectious diseases*, vol. 10, no. 1, 2017.

[74] J. F. Matos, L. M. S. A. Dusse, R. V. B. Stubbert et al., "Comparison of discriminative indices for iron deficiency anemia and β thalassemia trait in a Brazilian population," *Hematology*, vol. 18, no. 3, pp. 169–174, 2013.

[75] H. A. Getta, H. A. Yasseen, and H. M. Said, "Hi & Ha, are new indices in differentiation between iron deficiency anemia and beta-thalassaemia trait," *A Study in Sulaimani City-Kurdistan/Iraq IOSR-JDMS*, vol. 14, no. 7, pp. 67–72, 2015.

[76] T. Jameel, M. Baig, I. Ahmed, M. B. Hussain, and M. bin Doghaim Alkhamaly, "Differentiation of beta thalassemia trait from iron deficiency anemia by hematological indices," *Pakistan journal of medical sciences*, vol. 33, no. 3, pp. 665–669, 2017.

[77] L. Tong, J. Kauer, S. Wachsmann-Hogiu, K. Chu, H. Dou, and Z. J. Smith, "A new red cell index and portable rbc analyzer for screening of iron deficiency and thalassemia minor in a chinese population," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[78] C. Shen, "Evaluation of indices in differentiation between iron deficiency anemia and β-thalassemia trait for Chinese children," *Journal of Pediatric Hematology/Oncology*, vol. 32, no. 6, pp. e218–e222, 2010.

[79] E. Urrechaga and J. J. Hoffmann, "Critical appraisal of discriminant formulas for distinguishing thalassemia from iron deficiency in patients with microcytic anemia," Clinical Chemistry and Laboratory Medicine (CCLM), 2017.

[80] M. Jahangiri, F. Rahim, A. Saki Malehi, S. M. S. Pezeshki, and M. Ebrahimi, "Differential diagnosis of microcytic anemia, thalassemia or iron deficiency anemia: a diagnostic test accuracy meta-analysis," *Modern Medical Laboratory Journal.*, vol. 3, no. 1, pp. 1–14, 2019.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 5288844, 8 pages
https://doi.org/10.1155/2021/5288844

Research Article

Comparing the Prognostic Value of Stress Myocardial Perfusion Imaging by Conventional and Cadmium-Zinc Telluride Single-Photon Emission Computed Tomography through a Machine Learning Approach

Valeria Cantoni,[1] **Roberta Green**,[1] **Carlo Ricciardi**,[2,3] **Roberta Assante**,[1] **Leandro Donisi**,[1] **Emilia Zampella**,[1] **Giuseppe Cesarelli**,[3,4] **Carmela Nappi**,[1] **Vincenzo Sannino**,[2] **Valeria Gaudieri**,[1] **Teresa Mannarino**,[1] **Andrea Genova**,[1] **Giovanni De Simini**,[1] **Alessia Giordano**,[1] **Adriana D'Antonio**,[1] **Wanda Acampa**,[1,5] **Mario Petretta**,[6] **and Alberto Cuocolo**[1]

[1]*Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy*
[2]*Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy*
[3]*Bioengineering Unit, Institute of Care and Scientific Research Maugeri, Telese Terme, Campania, Italy*
[4]*Department of Chemical, Materials and Production Engineering, University of Naples Federico II, Naples, Italy*
[5]*Institute of Biostructure and Bioimaging, National Council of Research, Naples, Italy*
[6]*IRCCS SDN, Naples, Italy*

Correspondence should be addressed to Carlo Ricciardi; carloricciardi.93@gmail.com

Received 8 June 2021; Revised 30 September 2021; Accepted 5 October 2021; Published 16 October 2021

Academic Editor: Rafik Karaman

We compared the prognostic value of myocardial perfusion imaging (MPI) by conventional- (C-) single-photon emission computed tomography (SPECT) and cadmium-zinc-telluride- (CZT-) SPECT in a cohort of patients with suspected or known coronary artery disease (CAD) using machine learning (ML) algorithms. A total of 453 consecutive patients underwent stress MPI by both C-SPECT and CZT-SPECT. The outcome was a composite end point of all-cause death, cardiac death, nonfatal myocardial infarction, or coronary revascularization procedures whichever occurred first. ML analysis performed through the implementation of random forest (RF) and *k*-nearest neighbors (KNN) algorithms proved that CZT-SPECT has greater accuracy than C-SPECT in detecting CAD. For both algorithms, the sensitivity of CZT-SPECT (96% for RF and 60% for KNN) was greater than that of C-SPECT (88% for RF and 53% for KNN). A preliminary univariate analysis was performed through Mann-Whitney tests separately on the features of each camera in order to understand which ones could distinguish patients who will experience an adverse event from those who will not. Then, a machine learning analysis was performed by using Matlab (v. 2019b). Tree, KNN, support vector machine (SVM), Naïve Bayes, and RF were implemented twice: first, the analysis was performed on the as-is dataset; then, since the dataset was imbalanced (patients experiencing an adverse event were lower than the others), the analysis was performed again after balancing the classes through the Synthetic Minority Oversampling Technique. According to KNN and SVM with and without balancing the classes, the accuracy (p value = 0.02 and p value = 0.01) and recall (p value = 0.001 and p value = 0.03) of the CZT-SPECT were greater than those obtained by C-SPECT in a statistically significant way. ML approach showed that although the prognostic value of stress MPI by C-SPECT and CZT-SPECT is comparable, CZT-SPECT seems to have higher accuracy and recall.

1. Introduction

Risk stratification by noninvasive cardiac imaging has become increasingly important to optimize management and outcome in patients with coronary artery disease (CAD) [1]. Previous research indicated that stress single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) has been the most widely used nuclear cardiac imaging technique for the noninvasive assessment of cardiac disease, including the prognosis and choice of the most appropriate treatment strategies for patients with CAD [2]. Conventional- (C-) SPECT systems utilize sodium iodide crystals and parallel-hole collimators. This approach presents some technical limits; for instance, we can mention extended imaging time, low spatial resolution, and large doses of radiopharmaceuticals [3]. Recently, these limitations have been overcome with the introduction of gamma cameras with semiconductor cadmium-zinc-telluride (CZT) allowed to directly convert radiation into electric signals, bringing an improvement in image accuracy and acquisition time [4, 5].

Previous studies showed that CZT-SPECT findings can be used for risk stratification of patients referred to MPI for suspected or known CAD. Lima et al. [6] demonstrated that CZT-SPECT and C-SPECT provide similar prognostic results, with lower prevalence of hard events in patients with normal scan [6]. Yokota et al. [7] showed that the prognostic value of normal stress-only CZT-SPECT is at least comparable and may be even better than that of normal C-SPECT [7].

These biomedical technologies can produce big amount of data and, nowadays, different techniques have been used to obtain as much information as possible from data and signals [8–12]. Introducing machine learning (ML) in the healthcare sector can help clinicians in diagnosis and therapy planning, as well as in management of resources [13, 14]. Several studies have been conducted to test CAD detection using ML algorithms and to predict patient outcome [15–18]. An innovative approach is to use ML models to compare the performance of biomedical technologies, and an evaluation of the performance in terms of diagnostic power has already been reported [19, 20], demonstrating CZT-SPECT has a better ability to detect CAD. To the best of our knowledge, the prognostic value of CZT-SPECT and C-SPECT has not been investigated to date by using ML techniques.

Therefore, the purposes of the present investigation were as follows:

(1) To evaluate the prognostic value of C-SPECT and CZT-SPECT using ML-based approaches in patients with suspected or known CAD

(2) To compare the prognostic performance of these biomedical instrumentations through ML

This use of ML—in this particular case, aimed at compering two biomedical technologies—represents, to authors' best knowledge, one of the first attempts in literature.

2. Materials and Methods

2.1. Patients. Between February 2016 and May 2017, a total of 453 consecutive patients with suspected or known CAD were submitted by referring physicians to stress MPI for assessment of myocardial ischemia. For overall population, clinical history and cardiac risk factors were collected. Patients with a previous history of myocardial infarction, revascularization procedures, or a diagnosed atherosclerotic coronary disease were considered to have known CAD. The review committee of our institution approved the study (Protocol Number 110/17), and all patients gave informed consent.

2.2. Study Protocol. All patients were submitted to stress technetium-99m sestamibi-gated SPECT MPI by physical exercise or dipyridamole stress test, according to the recommendations of the European Association of Nuclear Medicine and European Society of Cardiology [21]. The protocol followed in this paper was the same employed in our previous research [20]. All patients underwent MPI by both C-SPECT and CZT-SPECT systems according to a randomized scheme in 1 : 1 ratio that determined which camera was used for first acquisition. For C-SPECT, a dual-head rotating gamma camera (E.CAM, Siemens Medical Systems, Hoffman Estates, IL, USA) was used. The acquisition time was 20 min for both stress and rest images. For CZT-SPECT (D-SPECT, Spectrum Dynamics, Caesarea, Israel), recordings were obtained using 9 pixilated CZT crystal detector columns mounted vertically spanning a 90 geometry. Scan duration was lower than 10 minutes for stress and lower than 5 minutes for rest imaging.

An automated software program (e-soft, 2.5, QGS/QPS, Cedars- Sinai Medical Center, Los Angeles, CA) was utilized to compute left ventricular (LV) volumes and ejection fraction (EF) and the scores incorporating both the extent and severity of perfusion defects, employing a standard segmentation of the 17 myocardial regions. The extent and grade of the quantitative defect were determined based on sex-specific normal limits while adding the scores of the 17 segments (from 0 for normal to 4 for absent perfusion) of the stress images allowed us to compute the summed stress score (SSS). A poststress LVEF greater than 45% and a SSS lower than 3 were considered normal.

2.3. Follow-Up Data. A follow-up questionnaire was collected by calling all patients by examinators blinded to patient's test results. The outcomes evaluated as endpoints were all-cause death, cardiac death, nonfatal myocardial infarction, or coronary revascularization procedures whichever occurred first. Cardiac death occurred subsequently to acute myocardial infarction, congestive heart failure, and cardiac interventional procedure related. Myocardial infarction was recorded when chest pain or equivalent symptom complex, positive cardiac biomarkers, or typical electrocardiographic changes were reported [22]. The length of follow-up was determined according to the date of the last medical visit.

2.4. Statistical Analysis. Statistical analyses were performed by using IBM SPSS statistics software (v. 26), both to test data distribution and to perform statistical tests. The process was carried out separately on both the parameters of the C-SPECT and the CZT-SPECT. First, the Kolmogorov-Smirnov test was performed to test data normality, in order to understand the type of test to be used (parametric or nonparametric): in particular, normality was tested for all parameters, for both groups, and for both camera types. Subsequently, a two-tailed t-test was performed for parameters with a normal distribution, while Mann-Whitney test was performed for the remaining parameters, and both tests were conducted considering a significance level of 0.05. After the use of ML algorithms, a chi-square test was used in order to compare the performances of different the models, trained with C-SPECT and the CZT-SPECT data, and to understand if there were statistical differences among them. The results are shown and discussed in the "Results" and "Discussion" sections, respectively.

2.5. Machine Learning Algorithms. The ML analysis was performed by using the Classification Learning App, provided by Matlab (v. 2019b), which trains models to classify data using supervised ML. The 10-fold crossvalidation was used to train and test the models; the dataset was divided into 10 groups of data, 9 were used for training the model and one group for testing it; the procedure was repeated 10 times, and the evaluation metrics are computed by averaging all those obtained [23]. The tree-based approach has shown in literature great results not only in the cardiologic context in cases such as diagnosis [24–26], prognosis [27, 28], and comparison of biomedical technologies [19, 20] but also in other medical specialties [29–31]. The classification tree is a simple and effective model consisting of nodes, branches, and leaves: each node has a rule that the data is routed along several branches while the leaves represent the output of the system [32].

Random forests (RF) model is part of the ensemble algorithms and allows to train together a set number of decision trees using the technique of Bootstrap Aggregation; this model turns out to have better accuracy than the single weak learner and reduces the chance of overfitting [33]. K-nearest neighbor (KNN) algorithm is a distance-based method. In fact, an example's membership in a class is determined by proximity to other known class examples. The critical aspect is the choice of the value of k that is the number of neighbors to consider for the decision [34]. Support vector machine (SVM) is a classification model that is based on finding the best surface that allows you to separate the two classes. In particular, the algorithm tries to maximize the margin between classes, the space that separates them, and in this way, bases learning on the most difficult examples, decreasing the influence of outliers [35]. Naïve Bayes (NB) was also employed in this study; it is a well-known algorithm based on the a priori probability theorem [36], thus being a completely different algorithm compared to the previous ones. These algorithms were used to predict an adverse event by using the features of the two cameras, and then, the evaluation metrics were compared through a statistical test for proportions in order to understand which one had the best capacity to detect the adverse event.

The present dataset is unbalanced; indeed, people with adverse events turned out to be much less than those with no events. In the literature, the problems that arise in training ML models using unbalanced data are well known [37, 38]. To deal with this problem, the Synthetic Minority Oversampling Technique (SMOTE) [39] was used; this oversampling technique generates new artificial data of the minority class, on the basis of those already present, allowing to rebalance the dataset. After that, the training phase of the models was repeated. This can be considered fair because it will be employed on both cameras allowing a fair comparison; moreover, the aim of the study is to compare C-SPECT and the CZT-SPECT rather than build the best prognostic model.

To evaluate the performance of the models, several metrics [40] were used: accuracy, sensitivity or recall, specificity, and precision. Furthermore, area under the curve (AUC) receiver-operating characteristic (ROC) was computed because it is a good method to assess model performance [41]. In addition, a feature selection process was performed to understand which parameters resulted more significant in reference to the target variable. We tested 14 features: perfusion parameters as SSS, summed rest score (SRS), summed difference score (SDS), and total perfusion defect (TPD) and functional parameters as systolic wall motion (SWM), systolic wall thickening (SWT), end-diastolic volume (EDV), end-systolic volume (ESV), and EF. In particular, two algorithms were used: Maximum Relevance–Minimum Redundancy (MRMR) that selects the variables with the most relevance to the destination one by calculating the mutual information of the parameters [42] and chi-square independence test [43].

3. Results

3.1. Patient Characteristics and Outcome. The clinical characteristics of patient population are shown in Table 1. The study group comprised 204 (45%) patients with suspected CAD and 249 (55%) with known CAD. The mean follow-up was 2.5 ± 0.5 years. During follow-up, 41 events occurred. The events were cardiac death in 1 patient, nonfatal myocardial infarction in 5, coronary revascularization procedures in 20, and 15 all-cause of death.

3.2. Statistical Analysis. The first step was to evaluate the possible normal distribution of the features between patients with events and patients with no events evaluated by both cameras, applying Kolmogorov-Smirnov test. The test revealed that, among the features of C-SPECT, only stress and rest EF (p value > 0.05) showed a normal distribution for both groups; similarly, no features of CZT-SPECT resulted to have a Gaussian distribution. Therefore, t-test was used only for stress and rest EF by C-SPECT, while Mann-Whitney test was performed for all other parameters, and the results are reported in Table 2.

Table 1: Clinical characteristics of patient population.

Characteristic	
Age (years)	64 ± 10
Male gender, *n* (%)	331 (73)
Body mass index ≥ 30 kg/m^2, *n* (%)	110 (24)
Diabetes, *n* (%)	153 (34)
Dyslipidemia, *n* (%)	333 (74)
Smoking, *n* (%)	196 (43)
Hypertension, *n* (%)	386 (85)
Atypical angina, *n* (%)	162 (36)
Family history of CAD, *n* (%)	231 (51)
Previous myocardial infarction, *n* (%)	148 (33)
Previous revascularization procedures, *n* (%)	173 (38)

Data are presented as mean ± SD or *n* (%) of subjects. CAD: coronary artery disease.

3.3. Machine Learning Analysis. The ML analysis was conducted separately and by using a 10-fold crossvalidation for C-SPECT and CZT-SPECT, both before and after SMOTE application in order to compare camera's performance with and without the augmentation of the dataset. The evaluation metrics regarding the models without SMOTE are reported in Table 3. Among the ML algorithms used for the analysis, RF reached the highest value of accuracy (90.3% and 90.1%, respectively, for C-SPECT and CZT-SPECT) and recall (98.5% and 99.0%, respectively, for C-SPECT and CZT-SPECT), but it presented the lowest value of specificity (7.3% and 0%, respectively, for C-SPECT and CZT-SPECT), showing a low capacity to detect adverse future events. Despite achieving these performances, statistically significant differences between the two cameras were not available, and this was also verified for Tree, SVM, and NB models. KNN model had an accuracy and recall lower than RF for both cameras (accuracy of 74.4% and 80.8%, recall of 78.6% and 87.4%, respectively, for C-SPECT and CZT-SPECT) but higher specificity (ranging from 14.6%, in CZT-SPECT, to 31.7% in C-SPECT). Nevertheless, the accuracy and the capacity to detect the absence of adverse event were statistically significant in favour of the CZT camera (p value = 0.021 for accuracy and p value = 0.001 for recall). These results were influenced by the imbalanced nature of the datasets; indeed, although accuracy and recall were high, they were affected by the bias introduced by the presence of a majority class for subjects with a negative prognosis, as also validated by the low AUCROC values of the models (ranging from 0.53 to 0.60 for C-SPECT and from 0.50 to 0.61 for CZT-SPECT). To overcome this issue, the dataset was balanced by introducing artificial samples of the minority class (patients with future adverse events), generated with SMOTE. The evaluation metrics values are reported in Table 4. The overall performance of classifiers increased significantly with a balanced dataset, especially in terms of specificity and AUCROC. Considering the C-SPECT, RF reached the highest values of accuracy (93.4%), recall (90.3%), and AUCROC (0.99), while SVM and KNN reached higher values of specificity (95.0% and 99.8%, respectively). Regarding CZT camera models performances, SVM classifier reached the highest values of accuracy (94.5%), recall (92.2%), and specificity (96.8%). Moreover, SVM turned out to have statistically significant performances: the accuracy and the recall showed a statistical significance in favor of the CZT-camera (p value = 0.016 for accuracy and p value = 0.028 for recall), while, despite showing in CZT-SPECT a higher capacity to detect adverse events, the specificity of SVM was not found to be statistically significant (p value = 0.279).

4. Discussion

To our knowledge, this is the first study using ML approach to compare the prognostic value of two technologies used in clinical routine practice (C-SPECT and CZT-SPECT) in patients with suspected or known CAD. Indeed, the ML analysis did not aim to create the best model to predict adverse events because, probably, it would not have been possible considering the highly unbalanced nature of the dataset. The aim was to test the feasibility of the cameras in predicting adverse future events in order to understand which could be the one with the better performance.

Although a similar evaluation has already been performed, ML techniques have never been used. Lima et al. [6] compared the prognostic value of MPI using an ultrafast protocol with low radiation in CZT-SPECT and a C-SPECT in different groups of patients. They concluded that the new protocol of MPI in CZT-SPECT showed similar prognostic results to those obtained in dedicated cardiac Na-I SPECT camera, with lower prevalence of hard events in patients with normal scan. Similarly, Yokota et al. [7] compared the prognosis of patients with normal stress-only at both CZT-SPECT and C-SPECT. They showed that the prognostic value of normal stress-only CZT-SPECT is at least comparable and may be even better than that of normal stress-only C-SPECT. In a recent study, Liu et al. [44] showed that ultra-low dose thallium perfusion imaging using CZT-SPECT provides good prognostic results, with a more severe prognosis in patients with abnormal MPI.

However, ML has been recently employed for the comparison of biomedical technologies. In previous studies using ML techniques to compare the diagnostic performance of C-SPECT and CZT-SPECT, we highlighted how algorithms trained with CZT-SPECT data achieved better accuracy, recall, and specificity than C-SPECT [19, 20]. Concerning the ML models, it has been observed that they generally present a high accuracy and recall. In particular, accuracy (p value = 0.021) and recall (p value = 0.001) were statistically significant for CZT-SPECT through the KNN algorithm. This result would demonstrate that CZT-SPECT has better performance to detect the absence of adverse event. To enhance the results obtained on the unbalanced dataset, a process of rebalancing the dataset was applied using SMOTE and repeating all the ML analyses. As expected, the performance of all models improved significantly for both cameras after rebalancing. However, SVM showed marked differences in all metrics values: accuracy, recall, and specificity had higher values in CZT-SPECT than C-

Table 2: Univariate statistical analysis of all the parameters of C-SPECT and the CZT-SPECT.

Parameters	C-SPECT			CZT-SPECT		
	Patients with no event	Patients with event	p value	Patients with no event	Patients with event	p value
SSS	9.90 ± 8.10	15.10 ± 11.50	0.053	9.30 ± 7.70	14.30 ± 11.70	<0.001***
SRS	6.80 ± 7.90	11.40 ± 11.60	0.163	5.10 ± 7.10	9.30 ± 11.10	0.240
SDS	3.10 ± 3.20	3.00 ± 2.50	0.841	4.10 ± 3.10	4.50 ± 2.80	0.310
TPD	13.10 ± 11.70	20.40 ± 16.70	0.043*	13.10 ± 11.80	20.20 ± 17.30	<0.001***
Stress SWM	14.90 ± 12.20	18.80 ± 15.00	0.007*	10.70 ± 12.10	14.70 ± 14.10	0.018*
Stress SWT	8.90 ± 9.20	11.70 ± 10.40	0.002*	6.70 ± 8.50	9.10 ± 9.40	0.020*
Stress EDV	92.90 ± 37.80	105.00 ± 51.50	0.031*	106.10 ± 42.40	121.10 ± 57.40	0.044*
Stress ESV	48.10 ± 32.30	60.40 ± 44.80	0.006*	54.70 ± 36.50	71.10 ± 51.50	0.008**
Stress EF	52.30 ± 14.20	48.40 ± 15.50	<0.001*	51.30 ± 11.80	46.80 ± 13.10	0.005**
Rest SWM	15.40 ± 12.70	21.50 ± 15.00	0.048*	10.20 ± 12.20	15.40 ± 13.00	0.070
Rest SWT	9.40 ± 9.40	12.65 ± 10.30	0.128	6.10 ± 8.30	9.80 ± 13.30	0.110
Rest EDV	91.86 ± 41.05	99.77 ± 41.48	0.493	106.30 ± 45.40	114.10 ± 52.50	0.790
Rest ESV	48.10 ± 35.70	57.60 ± 37.20	0.283	55.30 ± 41.50	65.50 ± 43.90	0.420
Rest EF	51.70 ± 13.80	46.60 ± 14.90	0.098	50.80 ± 12.00	46.90 ± 14.10	0.160

Statistically significant at: *0.05, **0.001, ***<0.001. Abbreviations. EDV: end-diastolic volume; EF: ejection fraction; ESV: end-systolic volume; SDS: summed difference score; SRS: summed rest score; SSS: summed stress score; SWM: wall motion; SWT: wall thickening; TPD: total perfusion defect.

Table 3: Machine learning analysis and statistical comparison through chi square test for proportions on the original dataset.

		Accuracy (%)		Error (%)	Recall (%)	Specificity (%)
Tree	C-SPECT	87.4		12.6	94.4	17.1
	CZT-SPECT	89.0		11.0	97.1	7.32
	p value		0.471		0.057	0.177
KNN	C-SPECT	74.4		25.6	78.6	31.7
	CZT-SPECT	80.8		19.2	87.4	14.6
	p value		**0.021**		**0.001**	0.067
SVM	C-SPECT	85.9		14.1	92.2	21.6
	CZT-SPECT	86.5		13.5	92.6	21.6
	p value		0.773		0.597	1.000
NB	C-SPECT	83.4		16.6	89.1	26.8
	CZT-SPECT	84.1		15.9	90.1	24.4
	p value		0.787		0.649	0.800
RF	C-SPECT	90.3		9.7	98.5	7.3
	CZT-SPECT	90.1		9.9	99.0	0.0
	p value		0.591		0.525	0.078

Abbreviations: KNN: K nearest neighbor; SVM: support vector machine; NB: Naïve Bayes; RF: random forests.

SPECT. In particular, accuracy and recall were statistically significant in favour of CZT-SPECT (accuracy p value = 0.016, recall p value = 0.028). Therefore, even considering the balanced data, CZT-SPECT proved to achieve a better accuracy and ability in predicting the absence of adverse event. It is likely that patients affected by an adverse event had a particular pattern of input variables which have allowed instance-based algorithms (KNN and SVM) to capture the outcome better than tree-based and probability-based algorithms. As regards the computational costs and the runtime of our models, there was no specific problem because all the models followed a simple workflow without applying heavy preprocessing algorithms (such as backward or forward feature selection methods). Indeed, all the models required less than a minute to be run.

The novel CZT technique provides patients with several advantages, as lower radiation dose and imaging time. Moreover, the higher energy and intrinsic spatial resolution of CZT detectors lead to lower artifacts and need for rest imaging, with a consequent reduction in radiopharmaceutical dosage which enables nuclear MPI to be more cost-effective [45].

Table 4: Machine learning analysis and statistical comparison through chi square test for proportions after SMOTE implementation.

		Accuracy (%)	Error (%)	Recall (%)	Specificity (%)
Tree	C-SPECT	88.1	11.9	86.2	90.1
	CZT-SPECT	88.1	11.9	86.9	89.3
	p value	1.000		0.760	0.731
KNN	C-SPECT	91.9	8.1	83.9	99.8
	CZT-SPECT	91.6	8.4	84.7	98.5
	p value	0.858		0.774	0.058
SVM	C-SPECT	91,5	8.5	87,6	95.0
	CZT-SPECT	94.5	5.5	92.2	96.8
	p value	**0.016**		**0.028**	0.279
NB	C-SPECT	59.3	40.7	86.7	32.0
	CZT-SPECT	59.0	41.0	87.6	30.3
	p value	0.880		0.677	0.599
RF	C-SPECT	93.4	6.6	90.3	94.4
	CZT-SPECT	93.0	7.0	91.0	94.9
	p value	0.637		0.720	0.757

Abbreviations. KNN: K nearest neighbor; SVM: support vector machine; NB: Naïve Bayes; RF: random forests.

4.1. Limitations and Future Developments. This study has some limitations that need to be considered. The dataset was strongly imbalanced, to the detriment of patients who present adverse events. It influenced the learning process of the models, introducing biases into evaluation metrics. SMOTE technique has been applied to balance the dataset and overcome these issues. However, the samples introduced were artificial, which represented another limitation. Nevertheless, the aim of the paper was not to evaluate the performance of the models in order to create a tool for clinical support, but to compare the performance of two technologies; therefore, the limitation introduced by the oversampling process is attenuated. Regarding future developments, it would be necessary to try to balance the dataset with original data rather than with artificial samples in order to increase the reliability of the evaluation metrics.

5. Conclusions

The novelty introduced in this study was the use of supervised learning techniques to compare the prognostic value of C-SPECT and CZT-SPECT. The results obtained showed that although the prognostic value of the two systems is comparable; CZT-SPECT seems to have higher accuracy and recall.

Data Availability

The dataset used to support the findings of this study have not been made available because of the privacy policy.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Conflicts of Interest

The authors declare that they have no conflict of interests.

References

[1] L. J. Shaw and A. E. Iskandrian, "Prognostic value of gated myocardial perfusion SPECT," *Journal of Nuclear Cardiology*, vol. 11, no. 2, pp. 171–185, 2004.

[2] W. Acampa, M. Petretta, L. Evangelista et al., "Stress cardiac single-photon emission computed tomographic imaging late after coronary artery bypass surgery for risk stratification and estimation of time to cardiac events," *Journal of Thoracic and Cardiovascular Surgery*, vol. 136, no. 1, pp. 46–51, 2008.

[3] H. O. Anger, "Scintillation camera with multichannel collimators," *Journal of Nuclear Cardiology*, vol. 5, pp. 515–531, 1964.

[4] T. Sharir, S. Ben-Haim, K. Merzon et al., "High-speed myocardial perfusion imaging: initial clinical comparison with conventional dual detector anger camera imaging," *JACC: Cardiovascular Imaging*, vol. 1, no. 2, pp. 156–163, 2008.

[5] D. S. Berman, X. Kang, B. Tamarappoo et al., "Stress thallium-201/rest technetium-99m sequential dual isotope high-speed myocardial perfusion imaging," *JACC Journals of the American College of Cardiology*, vol. 2, no. 3, pp. 273–282, 2009.

[6] R. Lima, T. Peclat, T. Soares, C. Ferreira, A. C. Souza, and G. Camargo, "Comparison of the prognostic value of myocardial perfusion imaging using a CZT-SPECT camera with a conventional anger camera," *Journal of Nuclear Cardiology*, vol. 24, no. 1, pp. 245–251, 2017.

[7] S. Yokota, M. Mouden, J. P. Ottervanger et al., "Prognostic value of normal stress-only myocardial perfusion imaging: a comparison between conventional and CZT-based SPECT," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 43, no. 2, pp. 296–301, 2016.

[8] P. Bifulco, M. Cesarelli, L. Loffredo, M. Sansone, and M. Bracale, "Eye movement baseline oscillation and variability of eye position during foveation in congenital nystagmus," *Documenta Ophthalmologica*, vol. 107, no. 2, pp. 131–136, 2003.

[9] M. Romano, F. Clemente, G. D'Addio, A. M. Ponsiglione, G. Improta, and M. Cesarelli, "Symbolic dynamic and frequency analysis in foetal monitoring," in *Proceedings of the IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2014Lisbon, Portugal.

[10] M. Cesarelli, A. Fratini, P. Bifulco, A. La Gatta, M. Romano, and G. Pasquariello, "Analysis and modelling of muscles motion during whole body vibration," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, no. 1, 2009.

[11] V. Onesto, L. Cancedda, M. L. Coluccio et al., "Nano-topography enhances communication in neural cells networks," *Scientific Reports*, vol. 7, no. 1, 2017.

[12] M. Recenti, C. Ricciardi, K. Edmunds, M. K. Gislason, and P. Gargiulo, "Machine learning predictive system based upon radiodensitometric distributions from mid-thigh CT images," *European Journal of Translational Myology*, vol. 30, no. 1, pp. 121–124, 2020.

[13] H. C. Koh and G. Tan, "Data mining applications in healthcare," *Journal of Health Informatics & Management*, vol. 19, pp. 64–72, 2005.

[14] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: review of the state-of-the-art and opportunities for healthcare," *Artificial Intelligence in Medicine*, vol. 104, p. 101822, 2020.

[15] A. Lin, M. Kolossváry, M. Motwani et al., "Artificial intelligence in cardiovascular imaging for risk stratification in coronary artery disease," *Radiology: Cardiothoracic Imaging*, vol. 3, no. 1, 2021.

[16] A. Akella and S. Akella, "Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution," *Future Science OA*, vol. 7, no. 6, 2021.

[17] M. Abdar, W. Ksiazek, R. Archarya, R. Tan, V. Makarenkov, and P. Plawiak, "A new machine learning technique for an accurate diagnosis of coronary artery disease," *Computer Methods and Programs in Biomedicine*, vol. 179, p. 104992, 2019.

[18] J. Betancur, Y. Otaki, M. Motwani et al., "Prognostic value of combined clinical and myocardial perfusion imaging data using machine learning," *JACC Journals of the American College of Cardiology*, vol. 11, no. 7, pp. 1000–1009, 2018.

[19] T. Mannarino, R. Assante, C. Ricciardi et al., "Head-to-head comparison of diagnostic accuracy of stress-only myocardial perfusion imaging with conventional and cadmium-zinc telluride single-photon emission computed tomography in women with suspected coronary artery disease," *Journal of Nuclear Cardiology*, vol. 20, 2021.

[20] V. Cantoni, R. Green, C. Ricciardi et al., "A machine learning-based approach to directly compare the diagnostic accuracy of myocardial perfusion imaging by conventional and cadmium-zinc telluride SPECT," *Journal of Nuclear Cardiology*, vol. 18, 2020.

[21] Task Force Members, G. Montalescot, U. Sechtem et al., "2013 ESC guidelines on the management of stable coronary artery disease," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 38, pp. 2949–3003, 2013.

[22] K. Thygesen, J. S. Alpert, A. S. Jaffe, M. L. Simoons, B. R. Chaitman, and H. D. White, "Third universal definition of myocardial infarction," *Circulation*, vol. 126, no. 16, pp. 2020–2035, 2012.

[23] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Ijcai*, vol. 14, no. 2pp. 1137–1145, IJCAI, 1995.

[24] C. Ricciardi, V. Cantoni, G. Improta et al., "Application of data mining in a cohort of Italian subjects undergoing myocardial perfusion imaging at an academic medical center," *Computer Methods and Programs in Biomedicine*, vol. 189, p. 105343, 2020.

[25] C. Wang, Y. Zhao, B. Jin et al., "Development and validation of a predictive model for coronary artery disease using machine learning," *Frontiers in Cardiovascular Medicine*, vol. 8, 2021.

[26] C. Ricciardi, A. S. Valente, K. Edmund et al., "Linear discriminant analysis and principal component analysis to predict coronary artery disease," *Health Informatics Journal*, vol. 26, no. 3, pp. 2181–2192, 2020.

[27] C. Ricciardi, V. Cantoni, R. Green, G. Improta, and M. Cesarelli, "Is it possible to predict cardiac death?," in *Mediterranean Conference on Medical and Biological Engineering and Computing*, pp. 847–854, Springer, 2019.

[28] C. Ricciardi, K. J. Edmunds, M. Recenti et al., "Assessing cardiovascular risks from a mid-thigh CT image: a tree-based machine learning approach using radiodensitometric distributions," *Scientific Reports*, vol. 10, no. 1, 2020.

[29] A. Stanzione, C. Ricciardi, R. Cuocolo et al., "MRI radiomics for the prediction of Fuhrman grade in clear cell renal cell carcinoma: a machine learning exploratory study," *Journal of Digital Imaging*, vol. 33, no. 4, pp. 879–887, 2020.

[30] C. Ricciardi, R. Cuocolo, G. Cesarelli et al., "Distinguishing functional from non-functional pituitary macroadenomas with a machine learning analysis," *Paper presented at the IFMBE Proceedings*, vol. 76, pp. 1822–1829, 2020.

[31] D. Scrutinio, C. Ricciardi, L. Donisi et al., "Machine learning to predict mortality after rehabilitation among patients with severe stroke," *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.

[32] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone, *Classification and Regression Trees*, IJCAI, 1983.

[33] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[34] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, 2018.

[35] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[36] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3no. 22, pp. 41–46, 2001.

[37] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial," *Special Interest Group on Knowledge Discovery in Data*, vol. 6, no. 1, pp. 1–6, 2004.

[38] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: a review," *GESTS international Transactions on Computer Science and Engineering*, vol. 30, 2006.

[39] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[40] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal*

of Data Mining & Knowledge Management Process, vol. 5, no. 2, pp. 01–11, 2015.

[41] A. P. Brandley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[42] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

[43] A. Satorra and P. M. Bentler, "A scaled difference chi-square test statistic for moment structure analysis," *Psychometrika*, vol. 66, no. 4, pp. 507–514, 2001.

[44] L. Liu, F. A. Abdu, G. Yin et al., "Prognostic value of myocardial perfusion imaging with D-SPECT camera in patients with ischemia and no obstructive coronary artery disease (INOCA)," *Journal of Nuclear Cardiology*, 2020.

[45] W. Acampa, R. R. Buechel, and A. Gimelli, "Low dose in nuclear cardiology: state of the art in the era of new cadmium-zinc-telluride cameras," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 17, no. 6, pp. 591–595, 2016.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 2213194, 15 pages
https://doi.org/10.1155/2021/2213194

Research Article

A Hybrid Method to Predict Postoperative Survival of Lung Cancer Using Improved SMOTE and Adaptive SVM

Jiang Shen,[1] Jiachao Wu,[1] Man Xu,[2] Dan Gan,[3] Bang An,[1] and Fusheng Liu[1]

[1]*College of Management and Economics, Tianjin University, Tianjin 300072, China*
[2]*Business School, Nankai University, Tianjin 300071, China*
[3]*School of Economics and Management, Hebei University of Technology, Tianjin 300071, China*

Correspondence should be addressed to Jiachao Wu; hhtaizhen@163.com

Received 19 April 2021; Revised 9 June 2021; Accepted 21 August 2021; Published 11 September 2021

Academic Editor: Mario Cesarelli

Predicting postoperative survival of lung cancer patients (LCPs) is an important problem of medical decision-making. However, the imbalanced distribution of patient survival in the dataset increases the difficulty of prediction. Although the synthetic minority oversampling technique (SMOTE) can be used to deal with imbalanced data, it cannot identify data noise. On the other hand, many studies use a support vector machine (SVM) combined with resampling technology to deal with imbalanced data. However, most studies require manual setting of SVM parameters, which makes it difficult to obtain the best performance. In this paper, a hybrid improved SMOTE and adaptive SVM method is proposed for imbalance data to predict the postoperative survival of LCPs. The proposed method is divided into two stages: in the first stage, the cross-validated committees filter (CVCF) is used to remove noise samples to improve the performance of SMOTE. In the second stage, we propose an adaptive SVM, which uses fuzzy self-tuning particle swarm optimization (FPSO) to optimize the parameters of SVM. Compared with other advanced algorithms, our proposed method obtains the best performance with 95.11% accuracy, 95.10% *G*-mean, 95.02% F1, and 95.10% area under the curve (AUC) for predicting postoperative survival of LCPs.

1. Introduction

Lung cancer (LC) is the deadliest cancer in the world. More than 85% of lung cancer patients are diagnosed with non-small-cell LC [1]. Surgical resection is the standard and most effective treatment for LC stage I, stage II, and nonsmall cell stage III A [1]. A major problem of the clinical decision on LC operation is to select candidates for surgery based on the patient's short-term and long-term risks and benefits, where survival time is one of the most important measures. Accurately predicting a patient's survival after surgery can help doctors make better treatment decisions. At the same time, it can help patients better understand their conditions to have good psychological expectations and financial preparation.

In recent years, more and more data-driven methods have been used to predict the postoperative survival of LCPs. In terms of statistical methods, Kaplan–Meier curves, multivariable logistic regression, and Cox regression are the three most widely used statistical methods to predict survival or complications for LCPs [2]. However, taking into account the shortcomings of traditional statistical methods and the incompleteness of medical data, data mining and machine learning techniques are introduced in recent years. Mangat and Vig [3] proposed an association rule algorithm based on a dynamic particle swarm optimizer, and the classification accuracy is 82.18%. Saber Iraji [4] compared the accuracy of adaptive fuzzy neural networks, extreme learning machine, and neural networks for predicting the 1-year postoperative survival of LCPs. The results show that sensitivity (90.05%) and specificity (81.57%) of an extreme learning machine are the highest, respectively. Tomczak et al. [5] used the boosted support vector machine (SVM) algorithm to predict the postoperative survival of LCPs. This algorithm combines the advantages of ensemble learning and cost-sensitive SVM, and the *G*-mean can reach

65.73%. As can be seen from the previous research, most of them ignore the impact of imbalanced data distribution, which may reduce the performance of classifiers.

Class imbalance refers to the phenomenon in which one class of data in a dataset is much larger than the others [6]. Standard machine learning classifiers are effective for balanced data, but they are not good for imbalanced data. Specifically, with the progress of medical technology, the number of long-term survivors after surgery for LCPs is much larger than that of short-term deaths. This will lead to higher prediction accuracy for survivors (majority class) and poorer recognition for deceases (minority class). Therefore, it is necessary to propose a method that has good classification performance for both survivors and deceased ones for predicting postoperative survival of LCPs.

During the past decades, the imbalanced data classification problem has widely become a matter of concern and has been intensively researched. The existing papers on imbalanced data processing methods have two main research directions: data level and algorithm level [7]. The data-level processing methods create a balanced class distribution by resampling the input data. Algorithm-level processing methods mainly involve two aspects: ensemble learning and cost-sensitive learning. Among these imbalanced data processing methods, the synthetic minority oversampling technique (SMOTE) is one of the most widely used methods, as it is relatively simple and effective [8]. However, it is likely to be unsatisfactory or even counterproductive if SMOTE is used alone, which is because its blind oversampling ignores the distribution of samples, such as the existence of noise [9, 10]. To solve this problem, many approaches are proposed to improve SMOTE. Ramentol et al. [11] combined rough set theory with SMOTE and proposed the SMOTE-RSB algorithm. SMOTE-RSB first uses SMOTE for oversampling and then removes noise and outliers in the dataset based on rough set theory. SSMNFOS [12] is a hybrid method based on stochastic sensitivity measurement (SSM) noise filtering and oversampling, which can improve the robustness of the oversampling method with respect to noise samples. The CURE-SMOTE [13] uses CURE (clustering using representatives) to cluster minority samples for removing noise and outliers and then uses SMOTE to insert artificial synthetic samples between representative samples and central samples to balance the dataset. However, most of these methods need to set the noise threshold through prior parameters, which increases the risk of misidentification of noise. In addition, some researchers consider ensemble filtering methods, which have been proven to be generally more efficient than single filters [14]. In this paper, we propose to use the cross-validated committees filter (CVCF) to detect and remove noise before applying SMOTE and record this method as CVCF-SMOTE. CVCF is an ensemble-based filter, which can reduce the risk of error in the threshold setting of prior parameters [15].

In addition, SVM as one of the most advanced classifiers has not been well used to predict postoperative survival of LC. In the previous research, SVM has been widely used in statistical classification and regression analysis due to its excellent performance [16]. Considering the limitations of SVM on imbalanced data, some studies combine resampling technology and SVM to deal with imbalanced data. D'Addabbo and Maglietta [17] proposed a method combining parallel selective sampling and SVM (PSS-SVM) to process imbalanced big data. Experimental results show that the performance of PSS-SVM is better than that of SVM and RUSBoost classifiers. Huang et al. [18] designed an undersampling technique based on clustering and combined it with optimized SVM to deal with imbalanced data. The classification performance of SVM is improved by the linear combination of SVM based on a mixed kernel. Fan et al. [19] proposed a hybrid technology combining principal component analysis (PCA), SMOTE, and SVM to diagnose chiller fault. Experimental results prove that this hybrid technology can improve the overall performance of chiller fault diagnosis.

However, these studies usually require a manual setting of SVM parameters, which may lead to failure to obtain the best experimental results. The standard SVM has a limitation that its performance depends on the selection of initial parameters. Some studies optimize the parameters of SVM through evolutionary calculations which have achieved good results. In these optimization algorithms, the particle swarm optimization- (PSO-) optimized SVM has been widely used with promising results due to its simplicity and fast convergence [20]. With the development of PSO technology, some improved PSO algorithms are used to optimize SVM. Wei et al. [21] proposed a binary PSO-optimized SVM method for feature selection, which overcomes the problem of premature convergence and obtained high-quality features. A switching delayed particle swarm optimization- (SDPSO-) optimized SVM is proposed to diagnose Alzheimer's disease [22]. Experimental results show that the proposed method outperforms several other variants of SVM and has obtained excellent classification accuracy. However, these methods often require parameter settings for PSO or improved PSO, such as particle size and inertial weight. In general, getting the best settings is complicated and time-consuming. If the PSO parameters are set improperly, it will even reduce the performance of the SVM.

In recent years, many new metaheuristics techniques have been proposed, such as Monarch Butterfly Optimization (MBO) [23], slime mould algorithm [24], Moth Search (MS) [25], Hunger Games Search (HGS) [26], and Harris Hawks Optimizer (HHO) [27]. However, most of these methods require users to tune parameters to achieve satisfactory performance. Fuzzy self-tuning PSO (FPSO) is a kind of setting-free adaptive PSO proposed in recent years [28]. The advantage of FPSO is that every particle is adaptively adjusted during the optimization process without any PSO expertise and parameter settings. Moreover, experimental results show that FPSO is better than several previous competitors in convergence speed and finding optimal solution aspects. Based on the above considerations, the FPSO algorithm is exploited to optimize the parameters of SVM, which leads to a novel FPSO-SVM classification algorithm.

Based on the improved SMOTE and FPSO-SVM, we propose a two-stage hybrid method to improve the performance

of the postoperative survival prediction of LCPs. In the first stage, CVCF is used to remove noise samples to improve the performance of SMOTE. Then, SMOTE is adopted to handle the imbalanced nature of the dataset. In the second stage, we apply FPSO-SVM to predict the postoperative survival of LCPs. The experimental results show that the proposed hybrid method outperforms other comparative state-of-the-art algorithms. This hybrid method can effectively improve the accuracy of survival prediction after LC surgery and provide reliable medical decision-making support for doctors and patients. Our contributions are summarized as follows:

(i) A novel hybrid method that combines improved SMOTE with adaptive SVM is proposed for predicting postoperative survival of LCPs

(ii) We apply CVCF to clean up data noise to improve the performance of SMOTE

(iii) FPSO is used to optimize the parameters of SVM and achieve an adaptive SVM

(iv) The proposed hybrid method not only performs higher predictive accuracy than other compared algorithms for predicting postoperative survival of LCPs but also has better *G*-mean, F1, and area under the curve (AUC)

The rest of this paper is as follows: Section 2 shows the materials and methods. The experiment design, performance metrics, and experimental results are described in Section 3. A brief summary is described in Section 4.

2. Materials and Methods

2.1. Data Description. In this paper, the thoracic surgery dataset in Zięba et al. [5], is selected to predict the postoperative survival of LCPs. Data were collected from the Wroclaw Thoracic Surgery Center. These patients underwent lung resection for primary LC from 2007 to 2011. It contains 470 samples with an imbalance rate of 5.71. There are 400 patients who survived more than one year and 70 patients who survived less than one year in this dataset. Table 1 shows the features of the dataset. These features were selected from 36 preoperative predictors by the information gain method and were used to predict the postoperative survival expectancy. Our task is to predict whether the survival time in patients after surgery was greater than one year.

Table 1: Feature details of the thoracic surgery dataset.

Feature ID	Description	Type of attribute
1	Size of the original tumor, from OC11 (smallest) to OC14 (largest)	Nominal
2	Diagnosis (specific combination of ICD-10 codes for primary and secondary as well multiple tumors if any)	Nominal
3	Forced vital capacity	Numeric
4	Pain (presurgery)	Binary
5	Age at surgery	Numeric
6	Performance status	Nominal
7	Weakness (presurgery)	Binary
8	Dyspnoea (presurgery)	Binary
9	Cough (presurgery)	Binary
10	Haemoptysis (presurgery)	Binary
11	Peripheral arterial diseases	Binary
12	MI up to 6 months	Binary
13	Asthma	Binary
14	Volume that has been exhaled at the end of the first second of forced expiration	Numeric
15	Smoking	Binary
16	Type 2 diabetes mellitus	Binary
17	1-year survival period (true value if died)	Binary

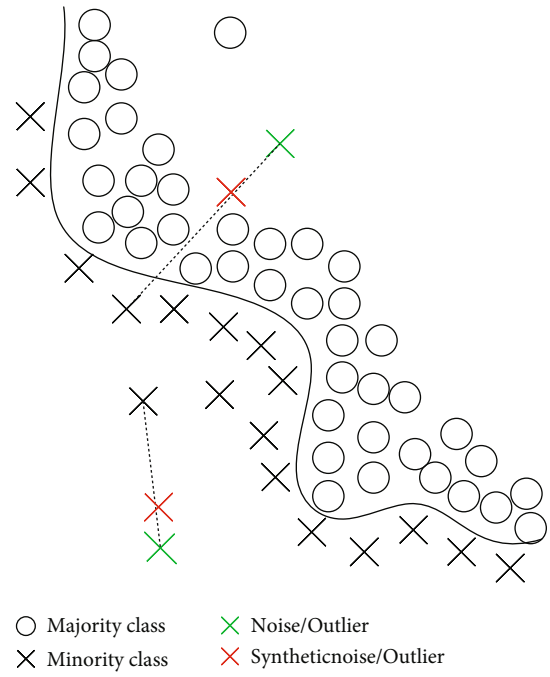


Figure 1: Using SMOTE alone may indiscriminately aggravate the noise.

2.2. Data Preprocessing

2.2.1. CVCF for Noise Cleaning. Although SMOTE is one of the most widely used methods for imbalanced data processing, it has some drawbacks in dealing with data noise. A major concern is that SMOTE may exacerbate the presence of noise in the data, as shown in Figure 1. Given the good performance of CVCF, we consider using it to improve SMOTE.

The CVCF algorithm is a well-known representative of an ensemble-based noise filter [29]. It induces multiple single classifiers by means of cross-validation. Afterward, samples mislabeled by all classifiers (or most classifiers) will be marked as noise and removed from the dataset. Choosing an appropriate base classifier is a key operation to ensure the excellent performance of CVCF. In this paper, we choose the

C4.5 algorithm as the base classifier of CVCF because it has better robustness to noise data and suitability for ensemble learning [30, 31].

C4.5 is an improved version of the ID3 algorithm [32]. It improves ID3 by handing numeric attributes and missing values and by introducing pruning. In addition, essentially different from the ID3, the information gain ratio is used to select split attributes in C4.5, which can be denoted by

$$\text{InfoGainRatio}(S, A) = \frac{\text{InfoGain}(S, A)}{\text{SpiltInfo}(S, A)}, \tag{1}$$

where $\text{InfoGainRatio}(S, A)$ represents the information gain ratio of attribute A in dataset S. $\text{InfoGain}(S, A)$ is the information gain of dataset S after splitting through attribute A and can be denoted by

$$\text{InfoGain}(S, A) = \text{Info}(S) - \text{Info}(S, A), \tag{2}$$

where $\text{Info}(S)$ is the entropy of dataset S. $\text{Info}(S, A)$ is the conditional entropy about attribute A. $\text{SpiltInfo}(S, A)$ denotes the splitting information of attribute A and is expressed by

$$\text{SpiltInfo}(S, A) = -\sum_{i=1}^{m} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}, \tag{3}$$

where $|S|$ represents the number of samples of dataset S. $|S_i|$ indicates the number of samples of subset i after the original dataset is divided into m subsets according to the attribute value of A.

2.2.2. SMOTE to Balance Data. The core idea of SMOTE is to insert artificial samples of similar values into the minority class, thereby improving the imbalanced distribution of classes. More specifically, the sampling ratio is set firstly, and then, the k nearest neighbors of each minority sample are found. Finally, according to equation (4), one of the neighbors is randomly selected to generate a synthetic sample that is put back into the dataset until the sampling number reaches the set ratio. The synthesized new sample is calculated as follows:

$$\mathbf{X_{new}} = \mathbf{X} + \partial(\mathbf{X_i}\text{-}\mathbf{X}), \quad \partial \in (0, 1), \tag{4}$$

where $\boldsymbol{X}_{\mathbf{new}}$ represents a new synthetic sample, $\mathbf{X}$ is the feature vector for each sample in the minority class, and $\mathbf{X_i}$ is the i-th nearest neighbor of sample $\mathbf{X}$. ∂ is a random number between 0 and 1.

2.3. The Proposed FPSO-Optimized SVM (FPSO-SVM)

2.3.1. SVM. SVM is a supervised learning classifier based on statistical theory and structural risk optimization [33]. SVM is not prone to overfitting and can handle high-dimensional data well. The principle of SVM is to map the original data to a high-dimensional space to discover a hyperplane that maximizes the margin determined by the support vectors. Suppose there is a dataset $D = \{(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), \cdots, (\mathbf{x_n}, y_n)\}$. The optimal hyperplane of dataset D can be expressed as

$$\mathbf{a}^T\mathbf{x} + b = 0, \tag{5}$$

where $\mathbf{a}^T$ is the weight vector and b represents the bias.

For nonlinear problems, the above-mentioned optimal hyperplane can be transformed into

$$\begin{cases} \min\limits_{\mathbf{a},b} & \frac{1}{2}\mathbf{a}^T\mathbf{a}\text{-}C\sum\limits_{i=1}^{n}\zeta_i, \\ \text{s.t.} & y_i\left(\mathbf{a}^T \cdot x_i + b\right) \geq 1 - \zeta_i, \quad \zeta_i \geq 0\, i = 1, 2, \cdots, n, \end{cases} \tag{6}$$

where C is the penalty factor and ζ_i is the slack variable. The above constrained objective function can satisfy the KKT condition by introducing the Lagrange formulation. The original objective function is transformed into

$$\begin{cases} \min\ \frac{1}{2}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} y_i y_j \beta_i \beta_j \left(\mathbf{x}_i \cdot \mathbf{x}_j\right) - \sum\limits_{i=1}^{n}\beta_i, \\ \text{s.t.}\ \sum\limits_{i=1}^{n}\beta_i y_i = 0,\ 0 \leq \beta_i \leq C, \quad i, j = 1, 2, \cdots, n, \end{cases} \tag{7}$$

where β is a Lagrangian multiplier. According to the previous experimental experience, a larger value of C means a larger separation interval and a greater generalization risk. Conversely, when the value of C is too small, it is easy to have an underfitting problem.

Finally, the decision function is shown in

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n}\beta_i^* y_i K < \mathbf{x}_i \cdot \mathbf{x}_j > + b^*\right), \tag{8}$$

Table 2: Confusion matrix.

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Table 3: Defuzzification of w, c_{soc}, c_{cog}, η, and λ.

Output	Level		
	Low	Medium	High
w	0.3	0.5	1.0
c_{soc}	1.0	2.0	3.0
c_{cog}	0.1	1.5	3.0
λ	0.0	0.001	0.01
η	0.1	0.15	0.2

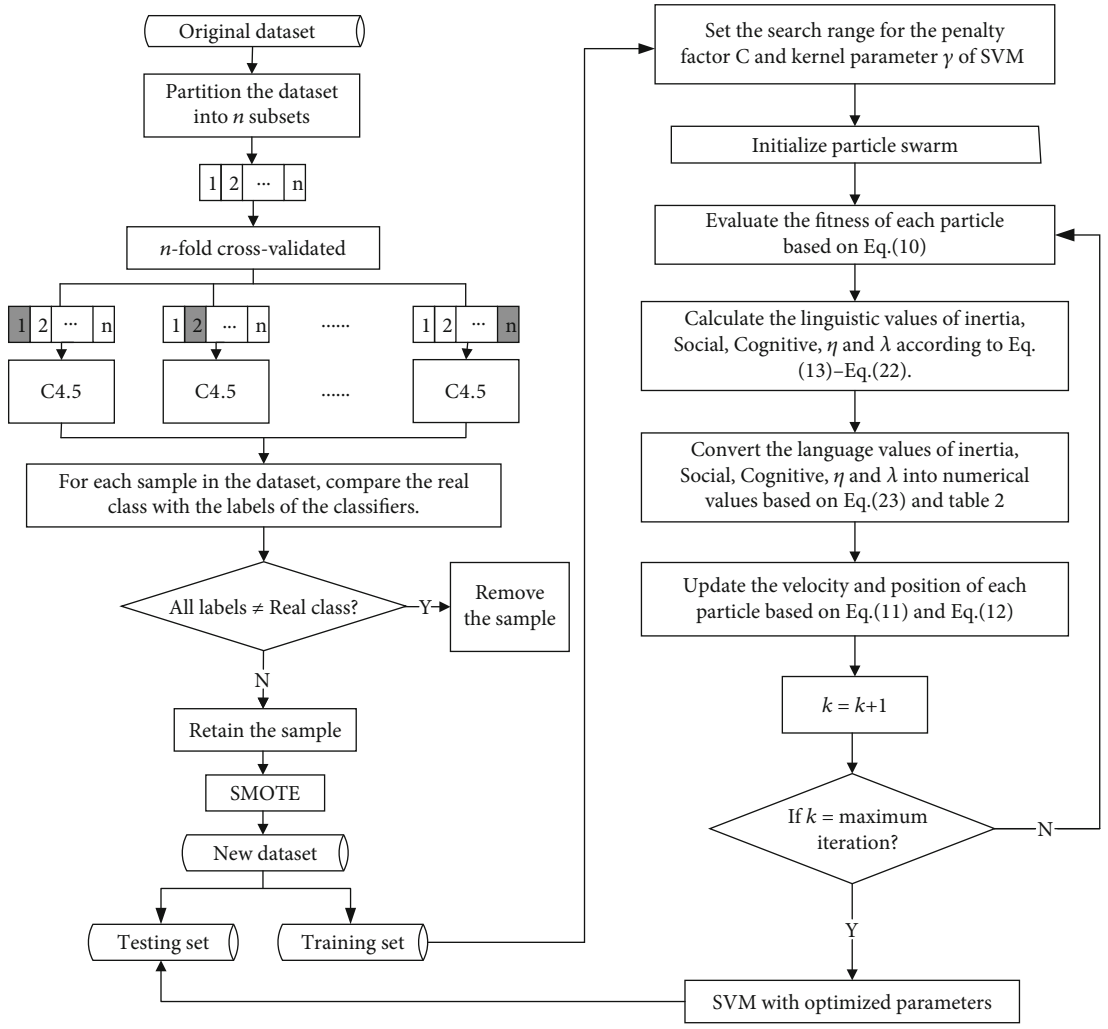


Figure 2: Flowchart of the proposed hybrid method for predicting postoperative survival of LCPs.

where β_i^* and b^* are the optimal Lagrangian multiplier and optimal value of b, respectively, and sgn $(\cdot)$ represents a symbolic function. $K < \mathbf{x}_i \cdot \mathbf{x}_j >$ is a kernel function. Usually, the radial basis function (RBF) kernel function is selected for SVM, which can be expressed as

$$K < \mathbf{x}_i \cdot \mathbf{x}_j > = \exp\left(-\gamma \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right), \tag{9}$$

where γ is the kernel parameter. The classification performance of SVM depends heavily on the setting of penalty factor C and kernel parameter γ. Therefore, parameter setting is a key step in applying SVM.

2.3.2. FPSO-SVM Model. In order to make SVM have better classification performance, we use FPSO to optimize the penalty factor C and kernel parameter γ of SVM, called FPSO-SVM. The classification accuracy is taken as the fitness function of FPSO, which is defined as

$$\text{Fitness} = \text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{10}$$

where TP, TN, FP, and FN represent four different classification results which are shown in Table 2.

FPSO is a fully adaptive version of PSO, which calculates the inertia weight, learning factor, and velocity independently for each particle based on fuzzy logic. The outstanding advantages of FPSO are that it does not require any prior knowledge about PSO and its optimization performance and convergence speed are better than those of PSO.

In FPSO, first, the number of particle swarms is set to $N = 10 + 2\sqrt{M}$ based on the heuristic [34, 35]. Here, M is the dimension of the optimization problem. In this paper, since there are two SVM parameters that need to be optimized, $M = 2$ and $N = 12$ (round down). After initializing the particles, we need to update them according to the position and velocity of the particles. Let $\mathbf{x}_i^k$ and $\boldsymbol{v}_i^k$ be the velocity and position of the i-th particle at the k-th iteration, respectively. At the $(k+1)$-th iteration, the velocity $\boldsymbol{v}_i^{k+1}$ and position $\mathbf{x}_i^{k+1}$ of the i-th particle can be defined as

$$\begin{aligned}\mathbf{v}_i^{k+1} = {} & w_i^k \cdot \mathbf{v}_i^k + c_{\text{soc}_i}^k \cdot \mathbf{r}_1 \cdot \left(\mathbf{x}_i^k - \mathbf{g}^k\right) \\ & + c_{\text{cog}_i}^k \cdot \mathbf{r}_2 \cdot \left(\mathbf{x}_i^k \text{-} \mathbf{b}_i^k\right), \quad i = 1, 2 \cdots, 12,\end{aligned} \tag{11}$$

Table 4: Accuracy comparison for different algorithms with different preprocessing methods.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	**0.8440**	**0.7149**	**0.6385**	0.7378	**0.8679**	**0.9511**
PSO-SVM	**0.8440**	0.6570	0.6217	0.6776	0.7267	0.8643
SVM	**0.8440**	0.5294	0.5561	0.4781	0.5493	0.5204
RF	0.8369	**0.7149**	0.6023	**0.7388**	0.8430	0.8869
GBDT	0.8156	0.7059	0.5864	0.7025	0.8213	0.9276
KNN	0.8227	0.6561	0.5833	0.6910	0.7905	0.9005
AdaBoost	0.7943	0.6652	0.5615	0.6458	0.7674	0.9095

Table 5: *G*-mean comparison for different algorithms with different preprocessing methods.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	0	0.6942	**0.6148**	0.7203	**0.8625**	**0.9510**
PSO-SVM	0	0.5832	0.5628	0.6150	0.6567	0.8501
SVM	0	0	0	0.1537	0.1015	0.1659
RF	0	**0.7092**	0.6017	**0.7385**	0.8404	0.8868
GBDT	**0.2938**	0.6901	0.5835	0.7024	0.8154	0.9274
KNN	0	0.6572	0.5819	0.6874	0.7919	0.9000
AdaBoost	0.2059	0.6550	0.5552	0.6464	0.7597	0.9096

Table 6: F1 comparison for different algorithms with different preprocessing methods.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	0	0.6612	0.5549	0.7059	**0.8482**	**0.9502**
PSO-SVM	0	0.5089	0.4995	0.5600	0.6022	0.8336
SVM	0	0	0	0.2823	0.0605	0.0536
RF	0	**0.6834**	**0.5713**	**0.7458**	0.8241	0.8889
GBDT	**0.1333**	0.6524	0.5470	0.7025	0.7950	0.9292
KNN	0	0.6545	0.5473	0.7094	0.7760	0.9035
AdaBoost	0.0645	0.6186	0.5101	0.6425	0.7323	0.9099

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1}, \tag{12}$$

where w_i^k is the inertia weight of particle i at the k-th iteration and $c_{\mathrm{soc}_i}^k$ and $c_{\mathrm{cog}_i}^k$ are social and cognitive factors of particle i at the k-th iteration, respectively. In FPSO, unlike conventional PSO, the values of w_i^k, $c_{\mathrm{soc}_i}^k$, and $c_{\mathrm{cog}_i}^k$ are not fixed but are calculated separately for different particles at each iteration. $\mathbf{r}_1$ and $\mathbf{r}_2$ are two random vectors, respectively. $\mathbf{b}_i^k$ and $\mathbf{g}^k$ are the position of the i-th particle and the best global position in the swarm at the k-th iteration.

The maximum velocity ($v_{\max_m}$) and minimum velocity ($v_{\min_m}$) of all particles in the m-th dimension are defined as

$$v_{\max_m} = \eta \cdot \left(b_{\max_m} - b_{\min_m}\right), \quad \eta \in (0, 1]. \tag{13}$$

$$v_{\min_m} = \lambda \cdot \left(b_{\max_m} - b_{\min_m}\right), \quad \lambda \in (0, 1], \tag{14}$$

where $b_{\max_m}$ and $b_{\min_m}$ represent upper and lower bounds of the m-th dimension for the optimization problem, respectively. η and λ ($\eta > \lambda$) are two coefficients determined by linguistic variables, in order to clamp $v_{\max_m}$ and $v_{\min_m}$ of each particle.

In order to get the w, c_{soc}, c_{cog}, η, and λ values of each particle in each iteration, two concepts are introduced: the distance between each particle and the global optimal particle and the fitness increment of each particle relative to the previous iteration.

The distance between any two particles in the k-th iteration is expressed as

$$\begin{aligned}\delta\left(x_i^k, x_j^k\right) &= \left\| x_i^k - x_j^k \right\|_2 \\ &= \sqrt{\sum_{m=1}^{2} \left(x_{i,m}^k - x_{j,m}^k\right)^2}, \quad i, j = 1, 2, \cdots, 12.\end{aligned} \tag{15}$$

Table 7: AUC comparison for different algorithms with different preprocessing methods.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	0.5000	**0.7265**	**0.6268**	**0.7400**	**0.8639**	**0.9510**
PSO-SVM	0.5000	0.6426	0.6069	0.6754	0.7094	0.8631
SVM	0.5000	0.5000	0.5000	0.4993	0.5059	0.5138
RF	0.4958	0.7115	0.6038	0.7397	0.8411	0.8873
GBDT	**0.5202**	0.6993	0.5857	0.7052	0.8171	0.9281
KNN	0.4874	0.6581	0.5842	0.6919	0.7927	0.9010
AdaBoost	0.4891	0.6603	0.5582	0.6483	0.7621	0.9097

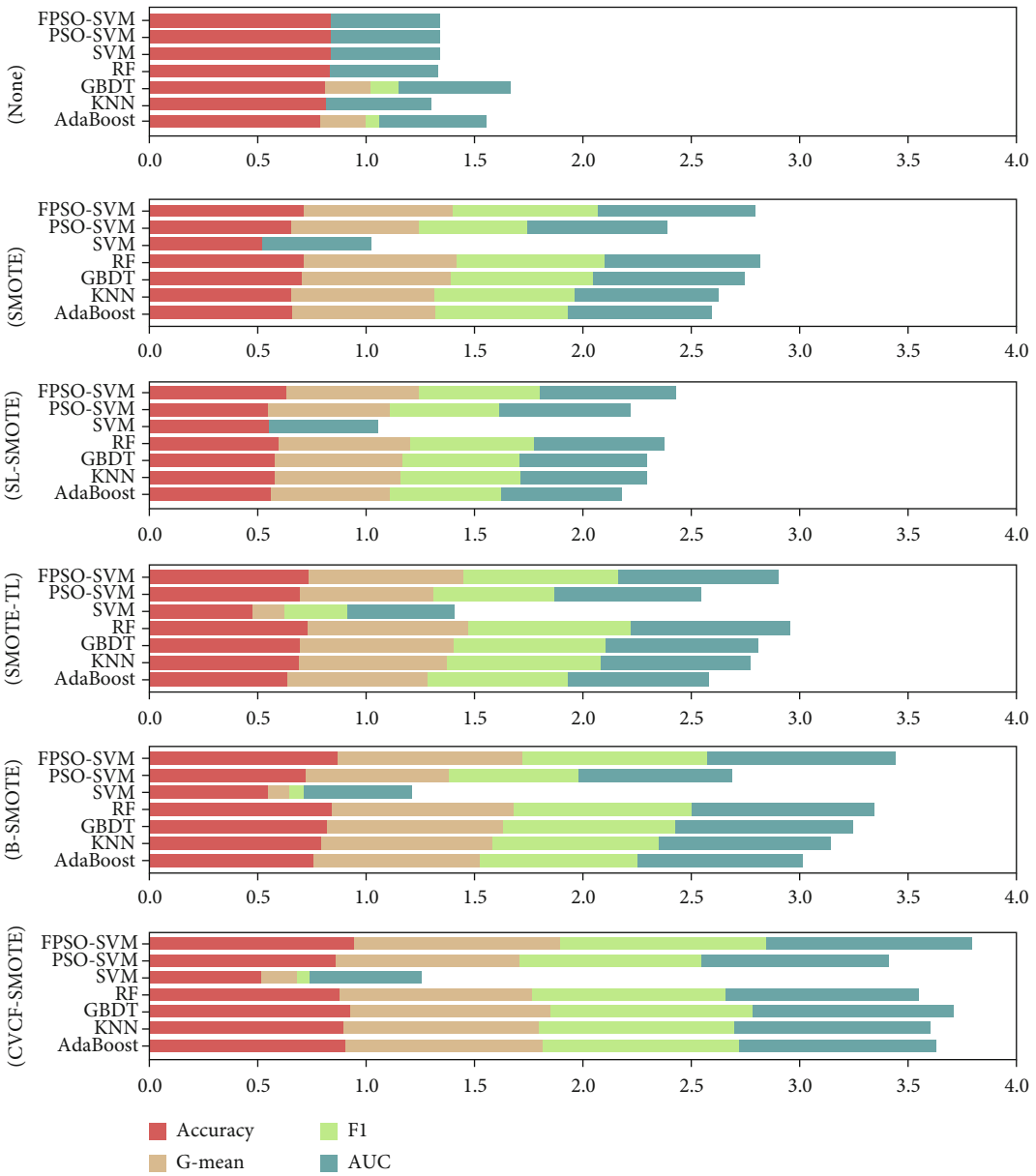


Figure 3: Stacked histograms of accuracy, *G*-mean, F1, and AUC for different algorithms under different preprocessing methods.

Table 8: Paired t-test results of CVCF-SMOTE+FPSO-SVM and the best performance under different preprocessing methods in terms of accuracy, F1, G-mean, and AUC on the thoracic surgery dataset. For CVCF-SMOTE, the p value is the statistic of the best result and the second best result.

Methods	Accuracy	F1	G-mean	AUC
NONE	11.034 (0.000)	25.502 (0.000)	21.102 (0.000)	27.01 (0.000)
SMOTE	14.348 (0.000)	16.01 (0.000)	10.261 (0.000)	12.469 (0.000)
SL-SMOTE	29.947 (0.000)	25.764 (0.000)	30.349 (0.000)	31.255 (0.000)
SMOTE-TL	29.815 (0.000)	30.281 (0.000)	22.248 (0.000)	26.895 (0.000)
B-SMOTE	6.541 (0.000)	5.176 (0.001)	5.297 (0.000)	5.997 (0.000)
CVCF-SMOTE	5.237 (0.001)	4.994 (0.001)	4.67 (0.001)	4.719 (0.001)

The function ϕ represents the normalized fitness increment of particle i for the previous iteration, which is calculated as

$$\phi\left(x_i^{k+1}, x_i^k\right) = \frac{\delta\left(x_i^{k+1}, x_i^k\right)}{\delta_{\max}} \cdot \frac{\min\left\{f\left(x_i^{k+1}\right), f_{\text{wor}}\right\} - \min\left\{f\left(x_i^k\right), f_{\text{wor}}\right\}}{|f_{\text{wor}}|}, \tag{16}$$

where $\delta_{\max}$ is the diagonal length of the rectangle formed by the search space. f_{wor} is the worst fitness value.

The linguistic variable of function δ is defined as Same, Near, and Far, which is used to measure the distance from a particle to the global best particle. The trapezoid membership function of Same is defined as

$$\delta = \begin{cases} 1, & \text{if } 0 \le \delta < \delta_1, \\ \dfrac{\delta_2 - \delta}{\delta_2 - \delta_1}, & \text{if } \delta_1 \le \delta < \delta_2, \\ 0, & \text{if } \delta_2 \le \delta \le \delta_{\max}. \end{cases} \tag{17}$$

The triangle membership function of Near is defined as

$$\delta = \begin{cases} 0, & \text{if } 0 \le \delta < \delta_1, \\ \dfrac{\delta - \delta_1}{\delta_2 - \delta_1}, & \text{if } \delta_1 \le \delta < \delta_2, \\ \dfrac{\delta_3 - \delta}{\delta_3 - \delta_2}, & \text{if } \delta_2 \le \delta < \delta_3, \\ 0, & \text{if } \delta_3 \le \delta \le \delta_{\max}. \end{cases} \tag{18}$$

The trapezoid membership function of Far is defined as

$$\delta = \begin{cases} 0, & \text{if } 0 \le \delta < \delta_2, \\ \dfrac{\delta - \delta_2}{\delta_3 - \delta_2}, & \text{if } \delta_2 \le \delta < \delta_3, \\ 1, & \text{if } \delta_3 \le \delta \le \delta_{\max}, \end{cases} \tag{19}$$

where $\delta_1 = 0.2 \cdot \delta_{\max}$, $\delta_2 = 0.4 \cdot \delta_{\max}$, and $\delta_3 = 0.6 \cdot \delta_{\max}$.

The linguistic variable of function ϕ is defined as Better, Same, and Worse, which is used to measure the improvement of a particle's fitness value for the previous iteration. The trapezoid membership function of Better can be obtained by

$$\phi = \begin{cases} 1, & \text{if } \phi = -1, \\ -\phi, & \text{if } -1 < \phi < 0, \\ 0, & \text{if } 0 \le \phi \le 1. \end{cases} \tag{20}$$

The triangle membership function of Same is expressed as follows:

$$\phi = 1 - |\phi|. \tag{21}$$

The triangle membership function of Worse is as follows:

$$\phi = \begin{cases} 0, & \text{if } -1 \le \phi < 0, \\ \phi, & \text{if } 0 \le \phi < 1, \\ 1, & \text{if } \phi = 1. \end{cases} \tag{22}$$

According to the preset fuzzy rules, w, c_{soc}, c_{cog}, η, and λ have three levels including Low, Medium, and High [28]. Table 3 shows the defuzzification values of w, c_{soc}, c_{cog}, η, and λ, which are calculated by the Sugeno inference method [36]. It is defined as follows:

$$\text{output} = \frac{\sum_{r=1}^{R} \rho_r z_r}{\sum_{r=1}^{R} \rho_r}, \quad r = 1, 2 \cdots R, \tag{23}$$

where R represents the number of rules. ρ_r and z_r are the membership degree of the input variable and output value of the r-th rule, respectively.

Table 9: Comparative results with previous studies based on accuracy.

Authors	Methods	Accuracy
Mangat and Vig [3]	DA-AC	82.18%
Elyan and Gaber [46]	RFGA	84.67%
Li et al. [47]	STDPNF	85.32%
Muthukumar and Krishnan [48]	IFSSs	88%
Saber Iraji [4]	ELM (wave kernel)	88.79%
Our work	CVCF-SMOTE+FPSO-SVM	95.11%

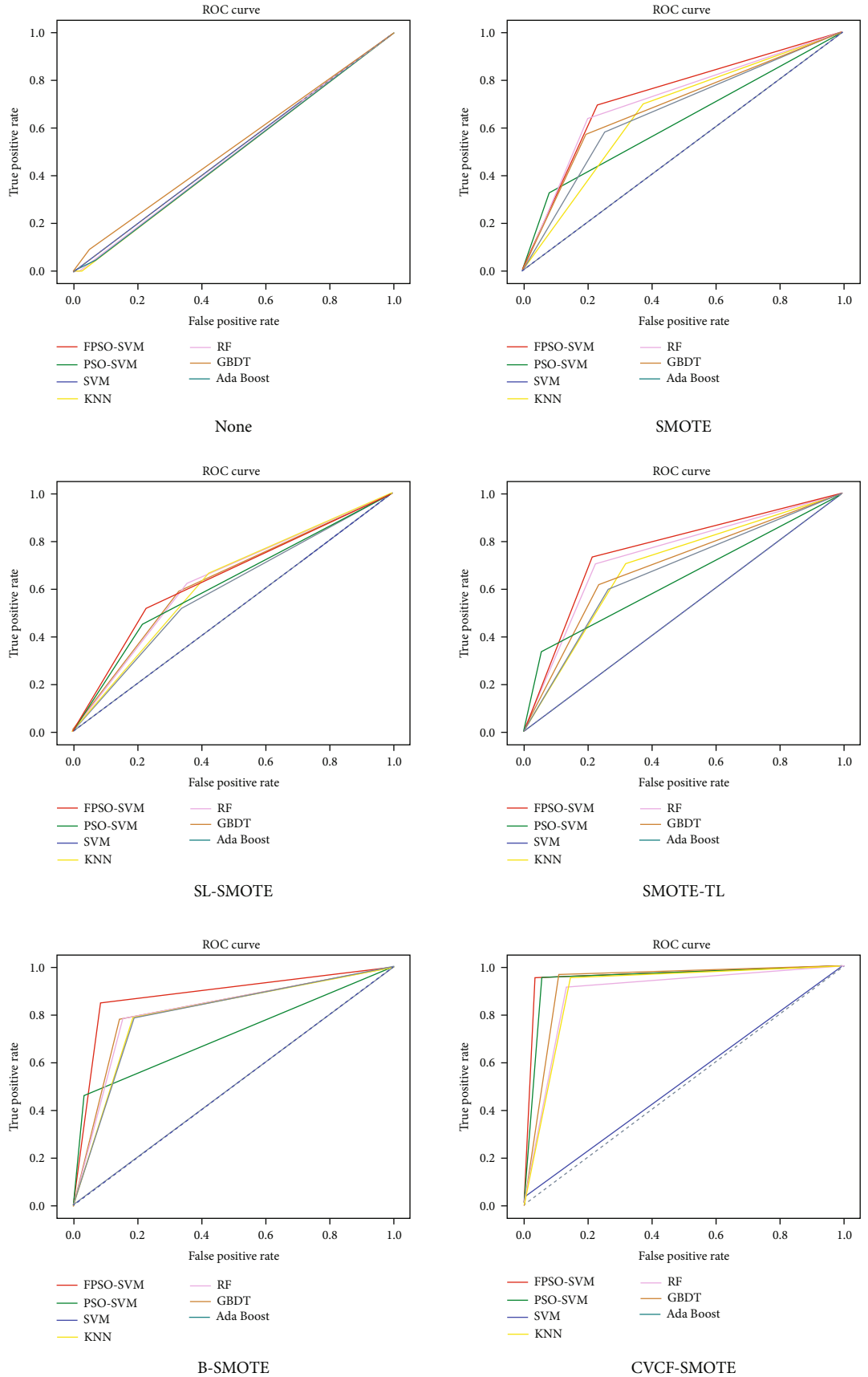
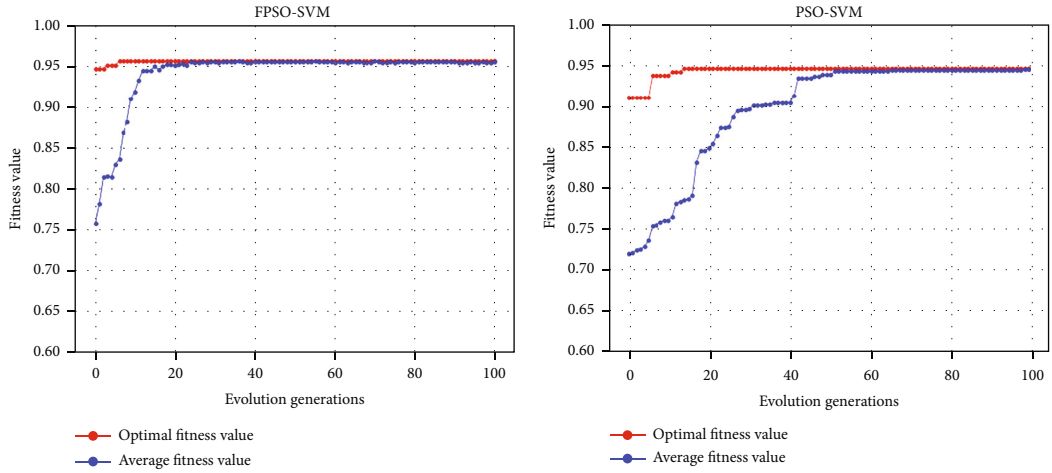


Figure 4: ROC curve comparison of different algorithms under different preprocessing methods.



(a) (b)

Figure 5: Fitness curves of FPSO-SVM (a) and PSO-SVM (b) with CVCF-SMOTE.

Then, update the position of each particle based on the obtained values of w, c_{soc}, c_{cog}, η, and λ. Finally, recalculate the fitness of each particle, that is, accuracy of the SVM corresponding to each particle. Repeat the above process until the maximum number of iterations is reached and output SVM with the optimal parameters.

The time complexity of FPOS-SVM consists of two parts: FPSO and SVM. In FPSO, the velocity and position of each particle are calculated in each iteration. Therefore, the computational complexity of FPSO is determined by the number of iterations, the particle swarm size, and the dimensionality of each particle. Thus, FPSO requires $O(TNm)$ time complexity, where T is the number of iterations of FPSO, N is the particle swarm size of FPSO, and m is the dimensionality of the optimization problem. For SVM, the optimal hyperplane is obtained by computing the distance between the support vector and the decision boundary. Then, the time complexity required for SVM is $O(dn_{sv})$, where d is the input vector dimension and n_{sv} is the number of support vectors. In FPSO-SVM, the number of SVM computations depends on the particle swarm size and the number of iterations of FPSO. Therefore, the time complexity of FPSO-SVM is $O(TNm + TNdn_{sv})$.

2.4. Specific Steps of the Proposed Hybrid Method for Predicting Postoperative Survival of LCPs. Based on improved SMOTE and FPSO-SVM, we propose a two-stage hybrid method to improve the performance of the postoperative survival prediction of LCPs. In the first stage, CVCF is used to remove noise samples to improve the performance of SMOTE. Then, apply SMOTE to balance data. In the second stage, FPSO-SVM is adopted to predict postoperative survival of LCPs. Figure 2 shows the flowchart of the proposed hybrid method. The specific steps of the hybrid method are presented as follows:

(1) Set CVCF to n-fold cross-validation. Then, the original dataset is divided into n subsets

Table 10: Details of Haberman and appendicitis datasets.

Datasets	Case number	Attribute number	Class distribution
Haberman	306	3	225/81
Appendicitis	106	7	85/21

(2) Take a different subset from the n subsets each time as the testing set and the remaining $n-1$ subsets as the training set. Therefore, a total of n different C4.5 classifiers are trained. Then, all the trained C4.5 classifiers will vote for each sample in the dataset. In this way, each sample has a real class label and n labels marked by C4.5

(3) For each sample, determine whether all (or most) labels marked with C4.5 are different from the real one. If all (or most) of them are different from the real class label, the sample will be treated as noise and removed from the dataset. On the contrary, the sample is retained. Finally, all the retained samples make up a cleaned dataset

(4) Oversample from the cleaned dataset with SMOTE until the class distribution of the dataset is balanced

(5) After data preprocessing with CVCF-SMOTE, the new dataset is divided into a training set and a testing set

(6) Set the search range for the penalty factor C and kernel parameter γ. Initialize particle swarm

(7) Evaluate the fitness of each particle based on equation (10). Calculate the linguistic values of Inertia, Social, Cognitive, η, and λ according to equations (13)-(22)

(8) Convert the language values of Inertia, Social, Cognitive, η, and λ into numerical values based

Table 11: Accuracy comparison for different algorithms with different preprocessing methods on the Haberman dataset.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	**0.7402**	**0.6890**	0.6386	**0.7396**	**0.7795**	**0.8205**
PSO-SVM	0.7098	0.6435	**0.6504**	0.6538	0.6831	0.7205
SVM	0.7196	0.6291	0.6409	0.6423	0.6772	0.7165
RF	0.6989	0.6795	0.6142	0.7315	0.7559	0.7772
GBDT	0.6837	0.6606	0.6299	0.7252	0.7465	0.7764
KNN	0.7174	0.6630	0.6417	0.7000	0.7449	0.7992
AdaBoost	0.7163	0.6402	0.6331	0.6117	0.6819	0.7559

Table 12: AUC comparison for different algorithms with different preprocessing methods on the Haberman dataset.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	0.5274	0.6813	0.6288	**0.7310**	**0.7748**	**0.8206**
PSO-SVM	0.5012	0.6131	0.6325	0.6669	0.6518	0.7121
SVM	0.5077	0.6096	0.6246	0.6598	0.6566	0.7035
RF	0.5731	**0.6815**	0.6132	0.7283	0.7588	0.7784
GBDT	0.5492	0.6607	0.6274	0.7226	0.7475	0.7765
KNN	0.5737	0.6649	**0.6418**	0.6997	0.7433	0.8009
AdaBoost	**0.5809**	0.6359	0.6293	0.6118	0.6779	0.7549

on equation (23) and Table 3. Update the velocity and position of each particle based on equations (11) and (12)

(9) Determine whether the maximum number of iterations has been reached. If it is reached, the optimized SVM is output. Otherwise, return to steps (7) and (8)

(10) Apply the optimized SVM on the testing set

3. Experiments and Results

3.1. Experiment Design. To evaluate our proposed hybrid method, we compare it with several state-of-the-art algorithms including PSO-optimized SVM (PSO-SVM), SVM, k-nearest neighbor (KNN) [37], random forest (RF) [38], gradient boosting decision tree (GBDT) [39], and AdaBoost [40]. In addition, we consider six preprocessing approaches, including CVCF-SMOTE, Borderline-SMOTE (B-SMOTE) [41], Safe-Level-SMOTE (SL-SMOTE) [42], SMOTE-TL [43], SMOTE, and no preprocessing (marked as NONE), to explore the performance of our proposed CVCF-SMOTE method. B-SMOTE, SL-SMOTE, and SMOTE-TL are three representative SMOTE extensions, which can handle imbalanced data with noise. In addition, in order to better evaluate the effectiveness of the proposed hybrid method, we tested its performance on two other imbalanced data. The value range of penalty factor C and kernel parameter γ is set to $[0, 30]$, and the maximum number of iterations is set to 30. All of these algorithms are programmed in the Python programming language, except for CVCF-SMOTE which is run in the KEEL software [44]. To eliminate randomness, experiments are repeated 10 times and the average performance is shown in this study.

Table 13: Paired t-test results of CVCF-SMOTE+FPSO-SVM and the best performance under different preprocessing methods in terms of accuracy and AUC on the Haberman dataset.

Methods	Accuracy	AUC
NONE	6.603 (0.000)	18.744 (0.000)
SMOTE	6.555 (0.000)	10.315 (0.000)
SL-SMOTE	15.959 (0.000)	15.806 (0.000)
SMOTE-TL	4.506 (0.001)	3.539 (0.006)
B-SMOTE	2.601 (0.029)	2.83 (0.02)
CVCF-SMOTE	4.669 (0.001)	4.392 (0.002)

3.2. Performance Metrics. In this section, we introduce the selected widely used imbalanced data classification performance metrics, including accuracy (defined by equation (10)), G-mean, F1, and AUC. They can be calculated according to the confusion matrix in Table 2.

$$G\text{-mean} = \sqrt{\frac{\text{TP}}{\text{TP} + \text{FN}} \times \frac{\text{TN}}{\text{TN} + \text{FP}}}, \tag{24}$$

$$\text{F1} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}, \tag{25}$$

where precision $= \text{TP}/(\text{TP} + \text{FP})$ and recall $= \text{TP}/(\text{TP} + \text{FN})$. Precision can be regarded as a measure of the exactness of a classifier, while recall can be regarded as a measure of the completeness of a classifier.

Table 14: Accuracy comparison for different algorithms with different preprocessing methods on the appendicitis dataset.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	**0.8688**	**0.8792**	**0.8208**	**0.9381**	**0.9167**	**0.9511**
PSO-SVM	0.8625	0.8713	0.7620	0.8104	0.8714	0.9277
SVM	0.8469	0.7979	0.7854	0.8310	0.8813	0.9021
RF	0.8438	0.8438	0.7271	0.8714	0.9083	0.9106
GBDT	0.8188	0.8479	0.7146	0.8690	0.8917	0.9085
KNN	0.8500	0.7708	0.7354	0.8476	0.8708	0.8957
AdaBoost	0.8031	0.8396	0.7458	0.8690	0.8896	0.9106

Table 15: AUC comparison for different algorithms with different preprocessing methods on the appendicitis dataset.

Algorithms	NONE	SMOTE	SL-SMOTE	SMOTE-TL	B-SMOTE	CVCF-SMOTE
FPSO-SVM	0.6878	**0.8807**	**0.8167**	**0.9411**	**0.9135**	**0.9512**
PSO-SVM	0.5893	0.7602	0.7708	0.9311	0.8917	0.9239
SVM	0.6674	0.7966	0.7832	0.8423	0.8788	0.8982
RF	**0.6930**	0.8475	0.7324	0.8755	0.9064	0.9070
GBDT	0.6460	0.8539	0.7207	0.8713	0.8909	0.9092
KNN	0.6885	0.7736	0.7374	0.8499	0.8676	0.8954
AdaBoost	0.6352	0.8461	0.7492	0.8685	0.8888	0.9102

AUC is defined as the area under the ROC curve and the coordinate axis. AUC is very suitable for the evaluation of imbalanced data classifiers because it is not sensitive to imbalanced distribution and error classification costs, and it can achieve the balance between true positive and false positive [45].

3.3. Result and Discussion. Tables 4–7 demonstrate the accuracy, *G*-mean, F1, and AUC values of different algorithms under different preprocessing methods for predicting postoperative survival of LCPs, respectively. The best experimental results of different preprocessing methods are marked in bold. We can see from Tables 4–7 that the proposed CVCF-SMOTE+FPSO-SVM model obtains the best performance among all methods with 95.11% accuracy, 95.10% *G*-mean, 95.02% F1, and 95.10% AUC. This shows that our proposed hybrid method can balance the classification accuracy of the minority class and the majority class while ensuring overall accuracy. That is, the proposed CVCF-SMOTE+FPSO-SVM method has a higher recognition rate for patients who survived after LC surgery for both longer than 1 year and less than 1 year.

In addition, it is easy to see from Tables 5–7 that the *G*-mean, F1, and AUC performances of different classifiers for the original dataset without preprocessing are extremely poor. However, it can be found from Table 4 that the classification accuracy of all the classifiers for the original dataset is higher than the accuracy after SMOTE preprocessing. This indicates susceptibility to imbalanced data; although the classifiers perform well in the majority class, it performs very poorly in the minority class. That is to say, these classifiers fail to balance the classification accuracy of LCPs whose survival time after surgery is longer than 1 year and less than 1 year.

Table 16: Paired *t*-test results of CVCF-SMOTE+FPSO-SVM and the best performance under different preprocessing methods in terms of accuracy and AUC on the appendicitis dataset.

Methods	Accuracy	AUC
NONE	6.591 (0.000)	15.628 (0.000)
SMOTE	4.562 (0.001)	5.176 (0.001)
B-SMOTE	3.024 (0.014)	3.373 (0.008)
SL-SMOTE	6.227 (0.000)	7.009 (0.000)
SMOTE-TL	1.089 (0.304)	0.785 (0.453)
CVCF-SMOTE	2.764 (0.022)	2.787 (0.21)

For the performance after preprocessing with SMOTE, we found that the *G*-mean, F1, and AUC values of most classifiers (except SVM) are higher than those of the original dataset. However, as can be seen from Table 4, the accuracy of all classifiers with SMOTE is lower than that of the original dataset. This shows that although SMOTE can balance precision and recall, it leads to a decrease in accuracy. For the three SMOTE extensions SL-SMOTE, SMOTE-TL, and B-SMOTE, we find that B-SMOTE has the most competitive performance. B-SMOTE+FPSO-SVM obtained the experimental results second only to CVCF-SMOTE+FPSO-SVM.

Figure 3 shows the stacked histograms of accuracy, *G*-mean, F1, and AUC for different algorithms under different preprocessing methods. It can be seen from Figure 3 that our proposed CVCF-SMOTE+FPSO-SVM has the best performance in predicting postoperative survival of LCPs. The main reasons behind the experimental results are as follows:

Table 17: Running time (in second) by CVCF-SMOTE+FPSO-SVM and state-of-the-art algorithms.

Datasets	Algorithms		
Thoracic surgery	CVCF-SMOTE+GBDT	CVCF-SMOTE+PSO-SVM	CVCF-SMOTE+FPSO-SVM
	31.2	53.6	43.5
Haberman	CVCF-SMOTE+KNN	CVCF-SMOTE+PSO-SVM	CVCF-SMOTE+FPSO-SVM
	18.8	27.5	24.5
Appendicitis	SMOTE-TL+FPSO-SVM	CVCF-SMOTE+PSO-SVM	CVCF-SMOTE+FPSO-SVM
	13.8	22.2	17.3

first, CVCF identifies and removes noise to improve the data quality so that blind oversampling can be reduced when applying SMOTE. Second, FPSO-SVM can search the optimal parameters of SVM adaptively, which improves the classification accuracy of SVM.

In order to further test the difference between CVCF-SMOTE+FPSO-SVM and other combination methods, a paired t-test was conducted among CVCF-SMOTE+FPSO-SVM and the best results under different preprocessing methods. A p value less than 0.05 is considered to be statistically significant in the experiment. From Table 8, it can be seen that CVCF-SMOTE+FPSO-SVM achieves significantly better results than the best results under different preprocessing methods in terms of the accuracy, F1, G-mean, and AUC at the prescribed statistical significance level of 5%.

We also compare the accuracy of our proposed model with previous studies as shown in Table 9. We can see from Table 9 that the accuracy of the CVCF-SMOTE+FPSO-SVM model is higher than that of other methods of the previous literature. Finally, we compare the ROC curves of different algorithms under different preprocessing methods, as shown in Figure 4. The greater the AUC value, the better the classifier performance. It can be seen that the AUC of our proposed CVCF-SMOTE+FPSO-SVM is the largest, which means that our proposed model is outperforming other comparison methods for predicting postoperative survival of LCPs.

In order to further prove that the performance of our proposed FPSO-SVM is superior to that of PSO-SVM, we draw the fitness curves of these two algorithms. Figures 5(a) and 5(b) show fitness curves of FPSO-SVM and PSO-SVM with CVCF-SMOTE preprocessing. As can be seen from (Figures 5(a) and 5(b)), we can clearly see that compared with PSO-SVM, FPSO-SVM not only has a higher fitting degree but also a faster convergence speed. This shows that our proposed FPSO-SVM algorithm can identify the optimal solution in the search space faster and more accurately than PSO-SVM.

3.4. Works on Other Datasets. To show the generalization ability of our proposed method, we apply CVCF-SMOTE +FPSO-SVM to the other two imbalanced datasets collected from KEEL (https://sci2s.ugr.es/keel/) [44]. Table 10 shows the details of the two selected datasets.

Tables 11 and 12 show the accuracy and AUC of different algorithms in different preprocessing methods on the Haberman dataset. It can be seen from Tables 11 and 12 that under different preprocessing methods, accuracy and AUC of CVCF-SMOTE+FPSO-SVM are higher than those of the comparison classifiers. As shown in Table 13, the results of the paired t-test also show that CVCF-SMOTE+FPSO-SVM is significantly better than the best experimental results under different preprocessing methods on the Haberman dataset. For the appendicitis dataset, it can be seen from Tables 14 and 15 that CVCF-SMOTE+FPSO-SVM also obtains the highest accuracy and AUC value compared to other preprocessing methods and classifier combinations. As can be seen from Table 16, for the appendicitis dataset, CVCF-SMOTE+FPSO-SVM achieves significantly better results than the best performance under NONE, SMOTE, SL-SMOTE, and B-SMOTE. However, it is not a significant difference for the best performance under SMOTE-TL.

From the experimental results, we see that CVCF-SMOTE+FPSO-SVM outperforms the compared algorithms for both the thoracic surgery dataset and the other two imbalanced datasets. On the one hand, it is because CVCF-improved SMOTE is well adapted to different datasets. On the other hand, FPSO-SVM automatically adjusts the optimal parameters according to different datasets, thus improving the generalization ability of the SVM.

3.5. Running Time Analysis. We compared the running time of CVCF-SMOTE+FPSO-SVM with the algorithms with the highest accuracy among all the compared methods. For the three datasets thoracic surgery, Haberman, and appendicitis, the algorithms with the highest accuracy among the compared methods are CVCF-SMOTE+GBDT, CVCF-SMOTE+KNN, and SMOTE-TL+FPSO-SVM, respectively. In addition, in order to compare the running time of FPSO-SVM with that of PSO-SVM, CVCF-SMOTE+PSO-SVM is also involved in the comparison. The comparison results are shown in Table 17. It can be seen from Table 17 that the running time for CVCF-SMOTE+FPSO-SVM is less than that of CVCF-SMOTE+PSO-SVM for the three datasets. However, the running time of CVCF-SMOTE+FPSO-SVM is slower than that of CVCF-SMOTE+GBDT, CVCF-SMOTE+KNN, and SMOTE-TL+FPSO-SVM for the thoracic surgery, Haberman, and appendicitis datasets, respectively. Considering the higher classification performance of our proposed method, it can still be considered superior to other algorithms.

4. Conclusion

In this work, we proposed a hybrid improved SMOTE and adaptive SVM method to predict the postoperative survival

of LCPs. In our proposed hybrid model, CVCF is adopted to clear the data noise to improve the performance of SMOTE. Then, we use FPSO-optimized SVM to estimate whether the postoperative survival of LCPs is greater than one year. Experimental results show that our proposed CVCF-SMOTE+FPSO-SVM hybrid method obtains the best accuracy, *G*-mean, F1, and AUC as compared to other compared algorithms for postoperative survival prediction of LCPs.

Our proposed hybrid method can provide valuable medical decision-making support for LCPs and doctors. Considering the excellent classification performance for the other two imbalanced datasets, in the future, we will try to apply the proposed method to other problems based on imbalanced data, such as disease diagnosis and financial fraud detection. There are two limitations that need to be pointed out: one is that we only consider the 1-year survival after lung cancer surgery. In future studies, we will try to predict survival at other time points, such as survival 3 or 5 years after lung cancer surgery. The other is that the value range of the parameters of SVM in FPSO-SVM needs to be set manually, which may require some experience or experimental attempts. Designing a setting-free SVM is our future research direction.

Data Availability

The dataset for this study can be obtained from the UCI machine learning database (http://archive.ics.uci.edu/ml/datasets/Thoracic+Surgery+Data).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (71971123).

References

[1] J. A. Rotman, A. J. Plodkowski, S. A. Hayes et al., "Postoperative complications after thoracic surgery for lung cancer," *Clinical Imaging*, vol. 39, no. 5, pp. 735–749, 2015.

[2] C. A. Osuoha, K. E. Callahan, C. P. Ponce, and P. S. Pinheiro, "Disparities in lung cancer survival and receipt of surgical treatment," *Lung Cancer*, vol. 122, pp. 54–59, 2018.

[3] V. Mangat and R. Vig, "Novel associative classifier based on dynamic adaptive PSO: application to determining candidates for thoracic surgery," *Expert Systems with Applications*, vol. 41, no. 18, pp. 8234–8244, 2014.

[4] M. S. Iraji, "Prediction of post-operative survival expectancy in thoracic lung cancer surgery with soft computing," *Journal of Applied Biomedicine*, vol. 15, no. 2, pp. 151–159, 2017.

[5] M. Zięba, J. M. Tomczak, M. Lubicz, and J. Świątek, "Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients," *Applied Soft Computing*, vol. 14, pp. 99–108, 2014.

[6] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.

[7] C.-F. Tsai, W.-C. Lin, Y.-H. Hu, and G.-T. Yao, "Under-sampling class imbalanced datasets by combining clustering analysis and instance selection," *Information Sciences*, vol. 477, pp. 47–54, 2019.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[9] J. A. Sáez, J. Luengo, J. Stefanowski, and F. Herrera, "SMOTE-IPF: addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering," *Information Sciences*, vol. 291, pp. 184–203, 2015.

[10] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE," *Information Sciences*, vol. 465, pp. 1–20, 2018.

[11] E. Ramentol, Y. Caballero, R. Bello, and F. Herrera, "SMOTE-RSB: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory," *Knowledge and Information Systems*, vol. 33, no. 2, pp. 245–265, 2011.

[12] J. Zhang and W. W. Ng, "Stochastic sensitivity measure-based noise filtering and oversampling method for imbalanced classification problems," in *In 2018 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 403–408, IEEE, 2018.

[13] L. Ma and S. Fan, "CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests," *BMC Bioinformatics*, vol. 18, no. 1, pp. 169–169, 2017.

[14] J. Luengo, S.-O. Shim, S. Alshomrani, A. Altalhi, and F. Herrera, "CNC-NOS: class noise cleaning by ensemble filtering and noise scoring," *Knowledge-Based Systems*, vol. 140, pp. 27–49, 2018.

[15] D. O. Afanasyev and E. A. Fedorova, "On the impact of outlier filtering on the electricity price forecasting accuracy," *Applied Energy*, vol. 236, pp. 196–210, 2019.

[16] Z. Tao, L. Huiling, W. Wenwen, and Y. Xia, "GA-SVM based feature selection and parameter optimization in hospitalization expense modeling," *Applied Soft Computing*, vol. 75, pp. 323–332, 2018.

[17] A. D'Addabbo and R. Maglietta, "Parallel selective sampling method for imbalanced and large data classification," *Pattern Recognition Letters*, vol. 62, pp. 61–67, 2015.

[18] B. Huang et al., "Imbalanced data classification algorithm based on clustering and SVM," Journal of Circuits, Systems and Computers, 2020.

[19] Y. Fan, X. Cui, H. Han, and H. Lu, "Chiller fault diagnosis with field sensors using the technology of imbalanced data," *Applied Thermal Engineering*, vol. 159, no. 10, p. 113933, 2019.

[20] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Applied Soft Computing*, vol. 43, pp. 117–130, 2016.

[21] J. Wei, R. Zhang, Z. Yu et al., "A BPSO-SVM algorithm based on memory renewal and enhanced mutation mechanisms for

feature selection," *Applied Soft Computing*, vol. 58, pp. 176–192, 2017.

[22] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, and Y. Li, "A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 320, pp. 195–202, 2018.

[23] G. G. Wang, S. Deb, and Z. Cui, "Monarch butterfly optimization," *Neural Computing and Applications*, vol. 31, 2015.

[24] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: a new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020.

[25] G.-G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, no. 2, pp. 151–164, 2018.

[26] Y. Yang, H. Chen, A. A. Heidari, and A. H. Gandomi, "Hunger games search: visions, conception, implementation, deep analysis, perspectives, and towards performance shifts," *Expert Systems with Applications*, vol. 177, p. 114864, 2021.

[27] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

[28] M. S. Nobile, P. Cazzaniga, D. Besozzi, R. Colombo, G. Mauri, and G. Pasi, "Fuzzy self-tuning PSO: a settings-free algorithm for global optimization," *Swarm and Evolutionary Computation*, vol. 39, pp. 70–85, 2018.

[29] S. Verbaeten and A. Van Assche, "Ensemble methods for noise elimination in classification problems," in *In international workshop on multiple classifier systems*, pp. 317–325, Springer, Berlin, Heidelberg, 2003.

[30] S.-J. Lee, Z. Xu, T. Li, and Y. Yang, "A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making," *Journal of Biomedical Informatics*, vol. 78, pp. 144–155, 2017.

[31] L. P. F. Garcia, J. Lehmann, A. C. P. L. F. de Carvalho, and A. C. Lorena, "New label noise injection methods for the evaluation of noise filters," *Knowledge Based Systems*, vol. 163, pp. 693–704, 2019.

[32] J. R. Quinlan, "Improved use of continuous attributes in C4.5," *Journal of Artificial Intelligence Research*, vol. 4, no. 1, pp. 77–90, 1996.

[33] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[34] N. Hansen, R. Ros, N. Mauny, M. Schoenauer, and A. Auger, "Impacts of invariance in search: when CMA-ES and PSO face ill-conditioned and non-separable problems," *Applied Soft Computing*, vol. 11, no. 8, pp. 5755–5769, 2011.

[35] M. S. Nobile, G. Pasi, P. Cazzaniga, D. Besozzi, R. Colombo, and G. Mauri, "Proactive particles in swarm optimization: a self-tuning algorithm based on fuzzy logic," in *In 2015 IEEE international conference on fuzzy systems (FUZZ-IEEE)*, pp. 1–8, IEEE, 2015.

[36] M. Sugeno, *Industrial Applications of Fuzzy Control*, Elsevier Science Inc., 1985.

[37] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[38] T. K. Ho, "Random decision forests," *In Proceedings of 3rd international conference on document analysis and recognition*, vol. 1, 1995, pp. 278–282, IEEE, 1995.

[39] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, 2001.

[40] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and Computation*, vol. 121, no. 2, pp. 256–285, 1995.

[41] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in *In international conference on intelligent computing*, pp. 878–887, Springer, Berlin, Heidelberg, 2005.

[42] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-SMOTE: safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *In Pacific-Asia conference on knowledge discovery and data mining*, pp. 475–482, Springer, Berlin, Heidelberg, 2009.

[43] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

[44] J. Alcala-fdez, "KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple Valued Logic & Soft Computing*, vol. 17, no. 2-3, pp. 255–287, 2011.

[45] D. Veganzones and E. Severin, "An investigation of bankruptcy prediction in imbalanced datasets," *Decision Support Systems*, vol. 112, pp. 111–124, 2018.

[46] E. Elyan and M. M. Gaber, "A genetic algorithm approach to optimising random forests applied to class engineered data," *Information Sciences*, vol. 384, pp. 220–234, 2017.

[47] J. Li, Q. Zhu, and Q. Wu, "A self-training method based on density peaks and an extended parameter-free local noise filter for k nearest neighbor," *Knowledge-Based Systems*, vol. 184, p. 104895, 2019.

[48] P. Muthukumar and G. S. S. Krishnan, "A similarity measure of intuitionistic fuzzy soft sets and its application in medical diagnosis," *Applied Soft Computing*, vol. 41, pp. 148–156, 2016.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 8924293, 8 pages
https://doi.org/10.1155/2021/8924293

Research Article

Research on Key Technologies of Personalized Intervention for Chronic Diseases Based on Case-Based Reasoning

Lin Zhang[1,2] **and Ping Qi**[1]

[1]*Institute of Service Computing, Tongling University, Tongling, Anhui 244061, China*
[2]*Institute of Robotics Engineering, Anhui Sanlian University, Hefei, Anhui 230000, China*

Correspondence should be addressed to Ping Qi; qiping929@tlu.edu.cn

Received 10 April 2021; Revised 15 July 2021; Accepted 2 August 2021; Published 13 August 2021

Academic Editor: Giovanni D Addio

In recent years, with the acceleration of industrialization, urbanization, and aging process, the number of patients with chronic diseases in the world is increasing year by year. In China, the number of chronic diseases has increased tenfold in 10 years. The percentage of the disease burden in the whole society accounts for 79.4%. Chronic diseases have become the top killer for Chinese people's health. However, for chronic diseases, prevention is more important than treatment. It is the best way to keep healthy. Therefore, health intervention is the key to prevent chronic diseases. Especially now, with the spread of COVID-19 pandemic, reducing the times of hospital check-ups and treatments for chronic patients is practically significant for releasing the stress on medical staffs and decreasing the rate of transmission and infection of COVID-19. In this paper, case-based reasoning (CBR) technology is used to assist personalized intervention for chronic diseases, and the key technologies of personalized intervention for chronic diseases based on case-based reasoning are proposed. The case organization, case retrieval, and case retention techniques of CBR technology in chronic disease personalized intervention are designed, and the calculation of interclass dispersion is added to the distribution of feature words, which is used to describe the distribution of feature attributes in different categories of cases. It provides an effective method for the establishment of personalized intervention model for chronic disease.

1. Introduction

In recent years, with the acceleration of industrialization, urbanization, and aging process, the number of chronic diseases in China has exploded, which has increased tenfold during the 10 years. There are nearly 300 million people with chronic diseases, 350 million overweight and obese people, 200 million people with hypertension, 100 million with hyperlipidemia, and 92.4 million with diabetes. The death rate of chronic disease has risen to 86.6% of the total death rate of Chinese residents. The percentage of the disease burden in the whole society accounts for 79.4%. In the next 10 years, 80 million Chinese people will die of chronic diseases. Chronic disease has become China's top one killer, and huge medical expenses will also be the heavy burden for individuals, families, and society.

Common chronic diseases mainly include cardiovascular and cerebrovascular diseases, metabolic diseases, and pulmonary diseases, such as hypertension, diabetes, and coronary heart disease. These chronic diseases are characterized by long course of disease, many complications, and long treatment, which have a serious impact on the health and normal life of patients [1]. In fact, for chronic diseases, prevention is better than treatment. Prevention is the best way to keep healthy. As traditional Chinese medicine says, "three parts cure, seven parts raise." People cannot live forever, but people can gradually enhance the physical fitness and improve the ability of rehabilitation and antiaging through good living habits and later recuperation, so as to achieve the purpose of prolonging life and to improve the quality of life. Therefore, health intervention is the key to prevent and cure chronic diseases. However, health intervention has a high

requirement for specialization, and it is difficult for ordinary residents to carry out their own health intervention. Therefore, case-based reasoning technology can be used to assist the personalized intervention of chronic diseases.

Case-based reasoning (CBR) is written in the book Dynamic Memory, which is written by Roger Schank from Yale University in 1982. It is an important knowledge-based problem solving and learning method emerging in the field of artificial intelligence. It can be used to solve the problem that nonprofessionals are difficult to obtain and to express professional knowledge. CBR solves the existing problem through the reuse or modification of the solution of the most similar case by building a rich case base and looks for the most similar cases in the case base. In the problem to solve mechanism, CBR uses the case-based reasoning strategy and imitates the cognitive way of analogy in human decision-making process to solve the unstructured and knowledge poor domain problems effectively [2–11].

The process of case reasoning usually includes four steps: case representation, case retrieval, case reuse and modification, and case evaluation and learning. Among them, case retrieval is the key step of case reasoning. Only by finding similar cases through case retrieval can it be better for case-based reasoning. At present, case retrieval techniques used commonly include nearest neighbor retrieval, knowledge-guided retrieval, inductive reasoning retrieval, neural network retrieval, classification retrieval, rough set retrieval, and fuzzy retrieval. However, this paper does not use common case retrieval methods. Instead, it is based on the characteristics of common chronic disease cases, draws on the concept of TF-IDF (term frequency–inverse document frequency), combines the calculation method of information entropy, and then determines the weight of the case attributes through the calculation of the interclass dispersion distribution to solve the problem of different attribute weights. In addition, the paper finally compares the relative similarity of cases through the simple theorem of cosines, which greatly improves the efficiency of case similarity retrieval.

2. Related Research

Through the CBR research for many years, the author has designed a children's common diseases diagnosis method based on case-based reasoning and the elderly health assessment method based on case-based reasoning and has applied for successfully key project of Anhui province natural science foundation of the higher institutions, The Children's Common Diseases Diagnosis Method Based on Case-based Reasoning Research, and the supported project of excellent young talents in colleges and universities in Education Department of Anhui province, The Study of the Elderly Health Assessment Method Based on Case-based Reasoning. In the process of project research, the author not only puts the designed algorithm into practice and develop children's common diseases diagnosis model software to get access to the software copyright (see Annex 1 for the copyright certificate) but also standardizes the algorithm to make it be applied to other fields of case-based reasoning, successfully applies standardized algorithm to urban traffic guidance, and successfully develops the urban road traffic congestion channel decision support system software to get access to the software copyright (see Annex 1 for the copyright certificate) [11].

In the preliminary research results, either the diagnosis of common diseases in children, or the health assessment of the elderly, or the decision-making of urban traffic congestion, the application fields are relatively narrow. Although the software designed can use the concept of TF-IDF and the calculation method of information entropy to build a case model, and determine the similarity of unknown cases, the descriptions of the distribution of different characteristics in different cases are not too ideal. The results are usually based on known case diagnosis or artificial intervention, directly according to the known diagnostic results of similar cases, without human intervention. Therefore, the intelligent ability needs to be improved.

In order to solve the problem of the generality of the case-based reasoning method and the distribution description of characteristic attributes to improve the intelligence of the algorithm application process. The research groups have established the health big data through the questionnaire survey of urban residents' lifestyle and health status and have proposed the general case-based reasoning method to add interclass dispersion calculation through the analysis of the original model and the continuous testing and improvement of the software. This method is not only applicable to most fields of case-based reasoning but also describes the distribution of feature words among different classes, which solves the problem that IDF overamplifies the function of rare words. The authors apply this approach to personalize interventions for chronic diseases. Through the questionnaire survey of residents' lifestyle and health status, the case base of the case-based reasoning model has been established. Through the search of similar cases, the probability of chronic diseases caused by residents' lifestyle is calculated, and suggestions for reasonable adjustment of residents' lifestyle are given based on the diagnosis and treatment protocol of known patients.

3. A Framework of Key Technology Models for Personalized Interventions for Chronic Diseases Based on Case-Based Reasoning

Through the questionnaire survey of the lifestyle and health status of patients with chronic diseases, as well as the diagnosis and treatment protocol of patients with chronic diseases, the case database is established. Through the similarity retrieval of the unknown cases, several cases whose similarity meets the ranking requirements, or several cases whose similarity meets the threshold, are found out. Then, through the analysis of the chronic disease diagnosis and treatment protocol of similar cases, the diagnosis and treatment protocol of new cases can be obtained, so as to provide the diagnosis and treatment service for the chronic disease patients or to provide reasonable preventive measures for the potential chronic disease patients, to reduce the number of chronic patients hospitalized for examination and treatment. With

the prevalence of COVID-19, this has practical implications for reducing the stress on medical staffs at this particular time and the rate of transmission and infection of COVID-19. The key technology model framework of personalized intervention for chronic diseases based on case-based reasoning is shown as follows in Figure 1.

In the key technology model of personalized intervention for chronic disease based on case-based reasoning, the first cases collected need to have specific diagnosis and treatment protocols or preventive measures. Then, they need to have standardized descriptions. Different eigenvectors are used to describe the different attributes of the case state and treatment protocol. By retrieving case status one by one, several matching cases with the highest similarity with the new case are extracted from the case base. Then, the availability of the new case is calculated through the utilization rate of a diagnosis and treatment protocol of the most similar case to recommend the diagnosis and treatment protocol of the new case.

4. Case-Based Reasoning for Individualized Intervention of Chronic Diseases

The method of personalized intervention for chronic diseases based on case-based reasoning mainly includes four key technologies: standardized representation of case knowledge, case similarity retrieval, case reuse, and case personalized intervention.

4.1. Case Knowledge Standardized Representation. Before using CBR, the data should be cleaned and collated first. The data structure should be standardized. Various medical institutions have a large amount of medical data. However, due to local and temporal differences, many data are not only scattered but also have differences in storage structure, description of illness and diagnosis scheme, and attribute characteristics, so it is difficult to compare a lot of data on the same platform.

Here, we use Boolean eigenvectors to represent case knowledge. Since data is not all structured data, and different fields have different emphases on data requirements, so we first set up a Boolean attribute statistical diagram, which means that all evaluation indicators are structured and all attributes are broken down into Boolean options.

Take the questionnaire of lifestyle and health status of urban residents as an example, the sex can be divided into male and female, so the attribute "sex" can be made. The attribute option 1 represents male, and the attribute option 0 represents female. Age is continuous numerical data, which can be divided into several optional Boolean options such as "Child," "Teenager," "Youth," "Middle age," and "Old age" according to age. The daily sleep time has "less than 6 hours," "6-7 hours," "7-8 hours," and "more than 8 hours" options, so it is divided into "Daily sleep (less than 6 hours)," "Daily sleep (6-7 hours)," "Daily sleep (7-8 hours)," and "Daily sleep (more than 8 hours)" several Boolean options. Then, all the options are made into Boolean eigenvectors, and the attribute statistics of the evaluation indicators are obtained based on this, as shown in the following Diagram 1.

According to the attribute statistics table (Table 1), the original case library can be converted into a Boolean case library. Assuming that the original case library is shown in the following Table 2, the corresponding Boolean case library is shown in the following Table 3.

Through the transformation of the case base, we found that the case attributes would increase. Many optional attributes are divided into several normalized Boolean attributes, which are decomposable from the same attribute. In each case, only one of the Boolean attributes can be selected. However, the converted Boolean case base can make the cases into vectors, which is helpful for the contrast of similar cases. Realizing the structure of data is more helpful for data process. Even if different regions and institutions have different descriptions of the cases, the standardized conversion of the cases can become a structured case.

Assuming that the attributes of the original case database are decomposed into n Boolean attributes in the attribute statistics table, each Boolean case after transformation can be represented by an n-dimensional feature vector X, $X = (x_1, x_2 \cdots, x_n)$. In this vector, if the Boolean attribute does not appear, $x_i = 0$, otherwise, $x_i = 1$.

We can easily find that the weight of each Boolean attribute should be different in a case of eigenvector representation. The fewer times a Boolean attribute has a value of 1 in all cases, the more typical this attribute is in case evaluation, so its weight should be greater when carrying out case similarity retrieval. On the contrary, if a Boolean attribute has a value of 1 in a large number of cases, that is to say, it is difficult to judge the actual situation of the case through this attribute, and then its weight in the process of case similarity retrieval should be small. Therefore, it is not reasonable to set the weight of all the attributes that appear in the case to 1. This is similar to the inverse document frequency (IDF) of information theory.

IDF, simply to say, is that if a keyword w appears in N pages, the greater the N is, the smaller the weight of w is, vice versa [12].

Combining with the calculation method of information entropy, namely, the calculation method of information needed to express the uncertainty of information, we can get the formula for calculating the weight of the attribute of the case: $w_i = \log_2(D/D_i)$, where D is the total number of cases in the case base, and D_i is the number of times that the value of attribute i is 1 in all cases in the case base.

It is assumed that there are 1000 cases in the case base, among which 489 cases have a "sex" attribute value of 1. There are 489 males among 1000 cases, so the weight of "sex" attribute is $\log_2(1000/489) \approx 1.03$. Similarly, if the number of times that attribute i and attribute j value 1 in the case are 200 and 50, respectively, that is, $D = 1000$, $D_i = 200$, and $D_j = 50$, then, the weight of attribute I is $\log_2(D/D_i) = \log_2(1000/200) \approx 2.32$, while the weight of attribute j is $\log_2(D/D_j) = \log_2(1000/50) \approx 4.32$. By parity of reasoning, we can get the weight of all the evaluation indicators, so we get the attribute statistics table with weights which are shown in the following Table 4:

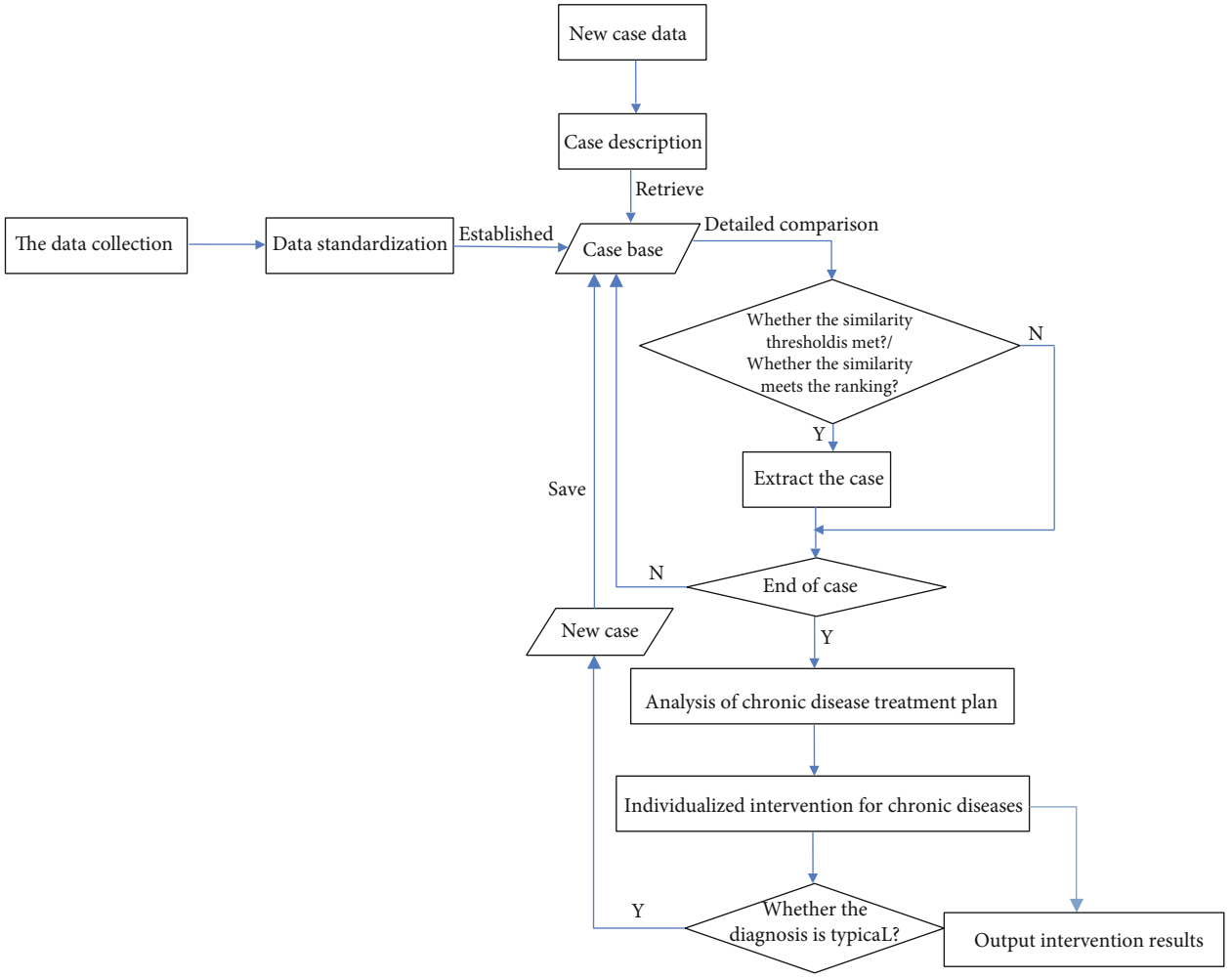


Figure 1: The key technology model framework of personalized intervention for chronic diseases based on case-based reasoning.

Table 1: Attribute statistics.

Attribute ID	Attribute content	Attribute description
1	Sex	Male : 1, female : 0
2	Child	Under the age of 12
3	Teenager	Age between 12 and 18
……	……	……
i	Daily sleep (less than 6 hours)	/
……	……	……
j	Eat fruit per week (more than 1000 g)	/
……	……	……
k	Does anyone in the immediate family have diabetes	Yes : 1, no : 0
……	……	……

Table 2: Original case library.

ID	Name	Sex	Age	……	Daily sleep	……	Eat fruit per week	……	Does anyone in the immediate family have diabetes
1	Zhang San	Male	65	……	Less than 6 hours	……	250 g-1000 g	……	No
……	……	……	……	……	……	……	……	……	……

Table 3: Original case library.

ID	Name	Sex	Child	……	Old age	Daily sleep (less than 6 hours)	……	Eat fruit per week (250 g-1000 g)	……	Does anyone in the immediate family have diabetes
1	Zhang San	1	0	……	1	1	……	1	……	0
……	……	……	……	……		……	……	……	……	……

Table 4: Attribute statistics with weights.

Attribute ID	Attribute content	Attribute description	The weight
1	Sex	Male : 1, female : 0	1.03
2	Child	Under the age of 12	0.86
3	Teenager	Age between 12 and 18	
……	……	……	……
i	Daily sleep (less than 6 hours)	/	2.32.
……	……	……	……
j	Eat fruit per week (more than 1000 g)	/	4.32
……	……	……	……
k	Does anyone in the immediate family have diabetes	Yes : 1, no : 0	0.32
……	……	……	……

Thus, the weight vector of the Boolean attribute can be obtained as follows:

$$W = (w_1, w_2, \cdots\cdots, w_i, \cdots\cdots)^T = (1.03{,}0.86, \cdots\cdots, 2.32, \cdots\cdots)^T. \tag{1}$$

And the weighted vector of each case i in the Boolean case base is deduced as follows:

$$\begin{aligned} wi &= X \times W = (x_1, x_2, \cdots\cdots, x_n) \times (w_1, w_2, \cdots\cdots, w_n)^T \\ &= (1, 0, \cdots\cdots, 1, \cdots\cdots) \times (1.03{,}0.86, \cdots\cdots, 2.32, \cdots\cdots)^T \\ &= (1.03{,}0, \cdots\cdots, 2.32, \cdots\cdots). \end{aligned} \tag{2}$$

Through the calculation method of IDF and information entropy, the case weighted vector obtained shows a good application effect in the allocation of case eigenvalue weight. However, the original intention of introducing IDF is to suppress the negative impact of the meaningless high-frequency attribute in the case. In addition, when the ratio between the total number of cases and the attribute with the value of 1 is large, the role of the low-frequency attribute is highlighted. However, here is a question which should be discussed: Common attributes are not necessarily meaningless. On the contrary, some patients with chronic diseases will have some inherent habits, or physical health indicators will have some inherent changes. These habits and changes often indicate that people with these habits or changes will suffer from a chronic disease precursor. In the same way, the occasional presence of low-frequency attributes will be treated as high-weight keywords, which will overamplify the importance of these attributes. Moreover, due to the differences of climate, environment, region, living habits, age, sex, and other factors, different categories of people in different regions will lead to the difference in the prevalence of different chronic diseases. In view of these deficiencies, the frequency of occurrence of the ith attribute in different classes will directly affect whether this attribute can become the characteristic attribute of the case. Therefore, an item can be added between the original cases to represent the distribution of feature attributes among different classes, that is to say, the interclass dispersion of feature attribute distribution.

The so-called interclass dispersion is the description of the distribution of characteristics attributed in different categories of cases. The characteristic attributes centrally distributed in a certain type of case often have a strong ability to distinguish categories. It is assumed that all cases can be divided into n categories, and $f(i)$ represents the frequency of occurrence of feature attribute i in a certain category of cases, while $\overline{f(i)}$ represents the average frequency of occurrence of feature I in all types of cases.

$$\overline{f(i)} = \frac{1}{n}\sum_{k=1}^{n} f_k(i). \tag{3}$$

The overall interclass dispersion is

$$D(i) = \frac{\sqrt{1/n - 1\sum_{k=1}^{n}\left(f_k(i) - \overline{f(i)}\right)^2}}{\overline{f(i)}}. \tag{4}$$

Substitute (3) into (4) to get:

$$D(i) = \frac{\sqrt{1/n - 1\sum_{k=1}^{n}(f_k(i) - 1/n\sum_{k=1}^{n}f_k(i))^2}}{\overline{1/n\sum_{k=1}^{n}f_k(i)}}. \quad (5)$$

Combine the main idea of weight calculation before, if the feature attribute in Formula (5) only appears in a certain type of case, it has the strongest classification ability, so $D(i)$ is 1. If the frequency of the feature attribute appearing in each category of cases is equal, it is considered that the feature does not have the classification ability. Therefore, $D(i)$ is 0, and the feature is useless and can be discarded. Thus, the value of $D(i)$ is between [0,1]. After considered the dispersion between classes, the weight calculation is as follows:

$$w_i = \left(\log_2 \frac{D}{D_i}\right) * \left(\frac{\sqrt{1/n - 1\sum_{k=1}^{n}(f_k(i) - 1/n\sum_{k=1}^{n}f_k(i))^2}}{\overline{1/n\sum_{k=1}^{n}f_k(i)}}\right). \quad (6)$$

Although the discreteness between classes is considered here, if the distribution of attributes with two features is basically similar in the same class case, we still cannot accurately judge the distribution of the two fault features. Therefore, we define the information entropy within the same kind of cases, so as to reflect the distribution of feature attributes within the same kind of cases. If the distribution of some feature attribute i in a similar case is more uniform, the information entropy in this kind of case is larger, and the feature attribute i can more easily reflect the feature information of this kind of case. The calculation formula of the information entropy of a case within the class is

$$E(t, C_k) = -\sum_{j}^{D} \frac{Nd_j}{NC_k} \lg \frac{Nd_j}{NC_k}, \quad (7)$$

wherein Nd represents the frequency of occurrence of the jth value (0 or 1) of feature attribute i in class CK cases, and NC_k represents the total frequency of occurrence of feature attribute I in class C_k cases.

Finally, based on the interclass dispersion and intraclass information entropy, a relatively accurate calculation method to determine the weight of feature attributes is obtained for the calculation of case class differentiation:

$$w_i = \left(\log_2 \frac{D}{D_i}\right) * \left(\frac{\sqrt{1/n - 1\sum_{k=1}^{n}(f_k(i) - 1/n\sum_{k=1}^{n}f_k(i))^2}}{\overline{1/n\sum_{k=1}^{n}f_k(i)}}\right) * \left(-\sum_{j}^{D} \frac{Nd_j}{NC_k} \lg \frac{Nd_j}{NC_k}\right). \quad (8)$$

According to Formula (8), the improved weight algorithm can be used to select the feature attributes, to calculate the weight of each feature attribute, and then to select the N cases with the largest weight as the feature vectors of CBR.

4.2. Case Similarity Retrieval. Case similarity retrieval is the core of CBR, which aims to retrieve as few approximate similar cases as possible from a large number of cases, as the reference to the solution of the current problem. Common case search strategies include template search strategy, literature search strategy, inductive index strategy, knowledge guide strategy, and nearest neighbor strategy. In this paper, the nearest neighbor strategy is used for case retrieval, but the calculation of similarity is determined by the law of cosines instead of Euclidean distance.

In the knowledge representation of the case, since we have established an attribute eigenvector for each case, we can calculate the size of the angle between two eigenvectors by using the cosine theorem. Since the weights of all indicators are positive, the cosine value between the two eigenvectors is between 0 and 1. The closer the cosine value between two eigenvectors is to 1, the smaller the angle between the two vectors is. It means that the closer the two eigenvectors are to each other. On the contrary, the closer the cosine value between the eigenvectors is to 0, the greater the angle between the two eigenvectors is. It means that the two eigenvectors represent less correlation between the cases.

We know that the cosine of $\triangle ABC$ is $\cos A = b^2 + c^2 - a^2/2bc$.

At this point, if b and c are regarded as two vectors starting from A, the above formula can be equivalent to $\cos A = \langle b, c\rangle / |b| \cdot |c|$, where $<b, c>$ said vector inner product, and $|b|$ and $|c|$ has said the length of the vector.

Suppose the eigenvectors of the Boolean attributes of case X are $(x_1, x_2 \cdots, x_n)$, where xi is 0 or 1, and the attribute weight vector $Y = (y_1, y_2 \cdots, y_n)^T$, then, its weighted eigenvector is $(x_1, x_2 \cdots \cdots, x_n) \cdots (y_1, y_2, \cdots, y_n)^T = (x_1y_1, x_2y_2, \cdots, x_ny_n)$.

Therefore, if we assume that the weighted eigenvectors of two cases A and B are $(a_1, a_2 \cdots \cdots, a_n)$ and $(b_1, b_2, \cdots, b_n)$, then, the cosine of the angle between them is $\cos\theta = a_1b_1 + a_2b_2 + \cdots \cdots + a_nb_n / \sqrt{a_1^2 + a_2^2 + \cdots \cdots + a_n^2}\sqrt{b_1^2 + b_2^2 + \cdots \cdots + b_n^2}$.

The smaller $\cos\theta$ value of the two vectors is, the smaller the approximation degree of the case will be. On the contrary, the larger $\cos\theta$ value is, the closer the two cases will be. When $\cos\theta = 1$, the two vectors will completely overlap, that is to say, the attribute indexes of the two cases will be exactly the same.

Therefore, we use the vector angle calculated by the law of cosines to express the similarity of two vectors. For example, if the result of two vectors calculated by the law of cosines is 0.5, then we reckon that the similarity of the two vectors is 50%. Although the nonlinear cosine function is not very accurate to calculate the similarity of the cases, but here, we do not need to calculate the accurate similarity between the cases to be evaluated and each case in the case library, but to know the relative similarity between the cases to be evaluated and the cases in the case library. That is to say, we only need to know which cases in the case library are more similar

Table 5: Diagnosis and treatment of similar cases.

Case ID	Similarity	Diagnosis and treatment protocol 1	Diagnosis and treatment protocol 2	Diagnosis and treatment protocol 3	……	Diagnosis and treatment protocol n
798	98.62%	1	0	1	……	1
1103	96.98%	1	1	1	……	0
6	95.33%	1	0	0	……	1
235	93.75%	1	0	0	……	1
……	……	……	……	……	……	……
39	89.99%	0	0	1	……	1
1295	88.73%	1	1	1	……	0

to the case to be evaluated. Therefore, using the law of cosines to evaluate similarity is simple, which can obtain a good result of corresponding approximation judgment.

4.3. Case Reuse and Case Personalized Intervention. Through the Boolean attribute feature vector expression of the above cases and the case similarity retrieval method calculated by the law of cosines, as well as the method of setting a threshold or setting the number of similar cases, a certain number of cases that are most similar to the current case can be obtained, such as setting search for cases where the similarity is over 90%, or search for the top 50 cases with similarity, etc. By obtaining chronic disease diagnosis and treatment plans of similar cases, we can obtain personalized intervention methods for the diagnosis and treatment of new chronic disease patients.

In the process of case similarity retrieval, if we can find cases with a similarity of 100%, we will find exactly the same cases. Then, we can directly reuse the diagnosis and treatment scheme of the case, otherwise.

First of all, we standardize the diagnosis and treatment protocols of all chronic disease cases in the case base and convert the diagnosis and treatment protocols of all cases into Boolean options after comprehensive conversion. This transformation is consistent with the standardized conversion method of cases in the process of case similarity retrieval. When a certain diagnosis and treatment scheme is adopted in a case, it means that the Boolean option value of the scheme is 1; otherwise, it is 0.

After the standardization of diagnosis and treatment schemes, the personalized intervention of diagnosis and treatment schemes in unknown cases are carried out according to the similarity of similar cases CR_i and the application degree of a diagnosis and treatment scheme CT_i in all selected cases. Then, the optional rate of diagnosis and treatment schemes in article jth of unknown cases is

$$\mathrm{New}\left(\mathrm{CT}_j\right) = \frac{\sum_{i=1}^{n}\left(\mathrm{CR}_i * \mathrm{CT}_j\right)}{\sum_{i=1}^{n}\mathrm{CR}_i} * 100\%. \tag{9}$$

Suppose, in the case base, N optional Boolean diagnosis and treatment protocols can be obtained after the comprehensive and decomposed treatment plans of all cases. Through case search, we find the top 50 cases are the most similar to the current unknown cases. The similarity between similar cases and new cases, as well as the diagnosis and treatment protocol of similar cases, is shown in Table 5.

Then, the probability of the new case adopting the diagnosis and treatment protocol 1 is

$$\frac{98.62\% * 1 + 96.98\% * 1 + 95.33\% * 1 + \cdots\cdots + 89.99\% * 0 + 88.73\% * 1}{98.62\% + 96.98\% + 95.33\% + \cdots\cdots + 89.99\% + 88.73\%} * 100\% = 97.30\%. \tag{10}$$

The diagnosis and treatment protocol of the new case can be given after the adoption rate of all the diagnosis and treatment protocols of the new case has been calculated, and the threshold value of the case adoption rate has been given through manual intervention.

For example, after manual intervention, the adoption rate of diagnosis and treatment protocol in new cases is more than 95%, and these plans can be regarded as the necessary treatment plan. The adoption rate of diagnosis and treatment protocol in new cases is between 75% and 95%, which can be regarded as the optional treatment plan. The adoption rate of diagnosis and treatment protocol in new cases is between 60% and 75%, as reference treatment plan.

In the process of personalized case intervention, in addition to providing case auxiliary diagnosis and treatment information, it can also be used to expand the case base. In the process of case similarity retrieval, if the similarity between the new case and the cases in the case base is lower than a certain threshold (for example, the similarity is lower than 95%), the auxiliary diagnosis and treatment scheme of the new case will be added to the case base as a case after manual intervention.

5. Conclusion

This paper puts forward the method of personalized intervention for chronic disease based on case-based reasoning and gives several key techniques in the process of intervention. This algorithm model can be used in the prevention of chronic diseases and also in the auxiliary diagnosis and treatment of chronic diseases. The main idea is to prevent or treat unknown cases through the judgment of case similarity and the diagnosis and treatment scheme of similar cases. In people's daily life, diseases are inevitable. In addition, different medical staff may give different results in the process of disease diagnosis. At this point, diagnosis and treatment experience is particularly important. Patients are more inclined to the diagnosis and treatment plan given by the medical staff with rich diagnosis and treatment experience. We are not saying that experience is always right, but in the case of ambiguity, the experience will be an important reference. The algorithm proposed in this paper is to integrate the experience of different medical institutions and medical staff and then to be applied. Therefore, the algorithm proposed in this paper can not only be used for personalized intervention for chronic diseases but also for personalized intervention for other diseases, even used in other fields. The premise is that the corresponding accurate case base can be established.

The accuracy of the algorithm proposed in this paper depends on the construction of the case base. The richer the cases in the case base are and the more accurate the diagnosis and treatment scheme in the case base is, the higher the feasibility of the auxiliary diagnosis and treatment scheme finally obtained by the algorithm will be. Of course, there are some problems with the algorithm itself:

Second, when the eigenvector is used to represent knowledge, many attributes in the Boolean case base are decomposed from the same attribute in the original case base, which leads to the fact that the eigenvector used is usually a sparse vector. In addition, the thresholds mentioned in case reuse and personalized intervention techniques need to be set by professionals. The manual intervention of professionals is necessary when new cases are added to the case base, which will undoubtedly increase the degree of manual intervention. Therefore, in practical application, how to simplify the existing algorithm by sparse vector algorithm on the basis of ensuring its effectiveness, and how to reduce the degree of manual intervention to improve its working efficiency as far as possible are the directions of future research.

Finally, the effectiveness of the algorithm in the application process is related to the size of the case base. However, with the continuous expansion of the case base, case similarity retrieval will become more and more complex. Therefore, how to improve the efficiency of the algorithm is also one of the future directions.

Data Availability

The experiment data supporting this experiment analysis are from previously reported studies, which have been cited, and are also included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by key research and development program of Anhui province, under Grant no. 202004a05020010; key program in the youth elite support plan in universities of Anhui province, under Grant no. gxyqZD2020043; and key natural science project of science foundation of Anhui Sanlian University, under Grant no. KJZD2021005.

References

[1] C. Ping, "Discussion on the intervention effect of individualized community chronic disease management," *Medicine and health care*, vol. 12, no. 2, p. 284, 2018.

[2] L. Ji-qiong, L. Xing-guo, G. Dong-xiao, and F. Shuai, "Case based reasoning ISP knowledge reuse method," *Computer Engineering*, vol. 36, pp. 36–39, 2010.

[3] L. I. Li, F. A. Xiao-zhong, Q. Quan, and L. I. Xiao-ming, "Ontology- based question expansion for question similarity calculation," *Journal of Beijing Institute of Technology*, vol. 20, no. 2, pp. 244–248, 2011.

[4] H. Min and L.-h. Shen, "Case-based resoning based on FCM and neural network," *Control and Decision*, vol. 27, pp. 1421–1424, 2012.

[5] M. F. Abdelwahed, A. E. Mohamed, and M. A. Saleh, "Machine learning; findings from Helwan University broaden understanding of machine learning (solving the motion planning problem using learning experience through case-based reasoning and machine learning algorithms)," *Journal of Engineering*, vol. 36, no. 4, pp. 251–256, 2020.

[6] A. Zia and R. Boumans, *Designing Participatory Decision Support Systems: Towards Meta-Decision Making Analytics in Then Generation of Ecological Economics*, Edward Elgar Publishing, 2020.

[7] M. Benamina, B. Atmani, S. Benbelkacem, and A. Mansoul, *Fuzzy Adaptation of Surveillance Plans of Patients with Diabetes*, Springer International Publishing, 2019.

[8] H. Zhang, *Research on Case-Based Reasoning for Urban Road Traffic Congestion Safety Decision Support Technology*, Anhui University of Science and Technology, 2018.

[9] Z. Lin and D. Zhang, "A novel diagnosis method for paediatric common disease using case-based reasoning," *International Journal of Simulation Systems, Science & Technology*, vol. 12, no. 17, pp. 37.1–37.5, 2016.

[10] N. Yuguang, K. Junjie, L. Fengqiang, G. Weichun, and Z. Guiping, "Science-automation science; researchers at North China Electric Power University detail findings in automation science (case-based reasoning based on grey-relational theory for the optimization of boiler combustion systems)," *Energy Weekly News*, vol. 352, no. 5, pp. 374–378, 2020.

[11] Z. Lin, "Research on case reasoning method based on TF-IDF," *International Journal of System Assurance Engineering and Management*, vol. 12, pp. 608–615, 2021.

[12] W. Jun, *The Beauty of Mathematics*, People's Posts and Telecommunications Press, 2012.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 7252280, 9 pages
https://doi.org/10.1155/2021/7252280

Research Article

Clinical Feature-Based Machine Learning Model for 1-Year Mortality Risk Prediction of ST-Segment Elevation Myocardial Infarction in Patients with Hyperuricemia: A Retrospective Study

Zhixun Bai,[1,2,3] **Jing Lu**,[4] **Ting Li**,[3] **Yi Ma**,[3] **Zhijiang Liu**,[3] **Ranzun Zhao**,[1,3] **Zhenglong Wang**,[3] **and Bei Shi**[1,3]

[1]*Program of Artificial Intelligence in Medicine, College of Medicine, Soochow University, Suzhou 215123, China*
[2]*Department of Internal Medicine, The Second Affiliated Hospital of Zunyi Medical University, Zunyi 563000, China*
[3]*Department of Cardiology, Affiliated Hospital of Zunyi Medical University, Zunyi, China*
[4]*Department of Pathology, Zunyi Medical and Pharmaceutical College, Zunyi 563006, China*

Correspondence should be addressed to Bei Shi; shibei2147@163.com

Received 14 April 2021; Accepted 16 June 2021; Published 5 July 2021

Academic Editor: Giovanni D Addio

Accurate risk assessment of high-risk patients is essential in clinical practice. However, there is no practical method to predict or monitor the prognosis of patients with ST-segment elevation myocardial infarction (STEMI) complicated by hyperuricemia. We aimed to evaluate the performance of different machine learning models for the prediction of 1-year mortality in STEMI patients with hyperuricemia. We compared five machine learning models (logistic regression, *k*-nearest neighbor, CatBoost, random forest, and XGBoost) with the traditional global (GRACE) risk score for acute coronary event registrations. We registered patients aged >18 years diagnosed with STEMI and hyperuricemia at the Affiliated Hospital of Zunyi Medical University between January 2016 and January 2020. Overall, 656 patients were enrolled (average age, 62.5 ± 13.6 years; 83.6%, male). All patients underwent emergency percutaneous coronary intervention. We evaluated the performance of five machine learning classifiers and the GRACE risk model in predicting 1-year mortality. The area under the curve (AUC) of the six models, including the GRACE risk model, ranged from 0.75 to 0.88. Among all the models, CatBoost had the highest predictive accuracy (0.89), AUC (0.87), precision (0.84), and F1 value (0.44). After hybrid sampling technique optimization, CatBoost had the highest accuracy (0.96), AUC (0.99), precision (0.95), and F1 value (0.97). Machine learning algorithms, especially the CatBoost model, can accurately predict the mortality associated with STEMI complicated by hyperuricemia after a 1-year follow-up.

1. Introduction

The most common cardiovascular diseases currently include hypertension, heart failure, coronary atherosclerosis, and myocardial infarction (MI); there is widespread interest in these conditions, as they are associated with high morbidity and mortality. In recent years, the incidence and death rate associated with MI have increased in China. The incidence of MI, though not strongly associated with the regions in China, has been found to increase with age [1]. Research has shown that MI typically starts to develop in young and middle-aged people. Therefore, the prevention, detection, and treatment of MI have become an area of interest among medical experts and scholars. In recent years, uric acid (UA) has been increasingly recognized as a well-known cardiovascular risk factor, along with hypertension, diabetes, chronic kidney disease (CKD), and obesity [2–7]. Although it is unclear whether UA is an independent predictor of cardiovascular disease, recent retrospective studies have demonstrated that hyperuricemia is an independent predictor of short- and long-term mortality in patients with AMI [8–10]. Machine learning is a multidisciplinary field involving artificial intelligence, computational complexity theory, probability and statistics, cybernetics, information theory,

philosophy, physiology, neurobiology, and other disciplines that can be characterized by system self-improvement. Machine learning was developed from the research method based on neuron models and function approximation theory; rule learning and decision tree learning were then incorporated based on symbolic calculus [11]. Furthermore, machine learning plays an essential role in clinical practice and cardiology. Each machine learning algorithm has its advantages in different fields. Previous studies have found that machine learning has good predictive power in predicting intrahospital mortality and short-term prognosis in acute MI. However, imbalanced data distribution and quality of deaths and survivors, that may lead to misclassification, are great challenges in machine learning. If the model evaluation places excessive emphasis on the area under the curve (AUC) index, it may ignore the weakness of truly predicting actual deaths. At present, there has been no research for developing a more comprehensive machine learning prediction model for the prognosis of ST-segment elevation myocardial infarction (STEMI) patients with hyperuricemia. Therefore, in this study, we evaluated multiple performance indicators for predicting 1-year mortality in STEMI patients with hyperuricemia, by using different machine learning models including logistic regression (LR), *k*-nearest neighbor (KNN), CatBoost, random forest (RF), and XGBoost. We then compared these models with the traditional GRACE risk score. To improve the prediction accuracy of imbalanced learning, we used SMOTEENN, a hybrid sampling algorithm of synthetic minority oversampling technique (SMOTE), and edited nearest neighbor (ENN) algorithms to oversample the minority class by creating synthetic samples.

2. Materials and Methods

2.1. Patients. This investigation followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines for cohort studies [12]. We enrolled consecutive patients aged >18 years diagnosed with STEMI at the Affiliated Hospital of Zunyi Medical University between January 2016 to January 2020 (Figure 1). The inclusion criteria were as follows: (1) increase or occurrence of ischemic chest discomfort at rest; (2) elevation of ST – segment ≥ 0.1 mV; (3) elevation of ST-segment in two consecutive leads; (4) elevated cardiac troponin I ($\geq 0.03\,\mu$g/L) or cardiac troponin T levels (≥42 ng/L); (5) diagnosed with hyperuricemia on admission; (6) no history of recent nephrotoxic drug intake; and (7) receipt of emergency percutaneous coronary intervention (PCI) treatment. The use of drugs was based on the treatment standards recommended by the published guidelines. Research approval was obtained from the Ethics Committee of the Affiliated Hospital of Zunyi Medical University (approval No. KLL [2020]0144). The need for written informed consent was waived owing to the retrospective nature of the study.

2.2. Outcomes. The primary outcome was defined as cardiac and sudden deaths during the 1-year clinical follow-up after discharge. Patients who had died during hospital admission were excluded from the analysis; the follow-up period ended in January 2021. All eligible patients enrolled in this study were followed up through telephone interviews or outpatient visits.

2.3. Candidate Predictors. Data on demographic characteristics, disease, electrocardiographic findings, laboratory parameters on admission, and in-hospital events were obtained from the patient's medical records. Data on baseline characteristics, demographics (age and gender), risk factors (hypertension, diabetes, current smoking, family history), nonweekday admission (NWDS), delay (defined as patient FMC > 12 hours), medical history (previous stroke, previous CKD), and electrocardiography (ECG) findings (inferior, anterior, right ventricular, and other) were all obtained from our electronic database. Hyperuricemia was defined by serum UA levels of >7 mg/dL (417 mmol/L) in men and >6 mg/dL (357 mmol/L) in women at admission. The patient data collected included demographic information, baseline characteristics at admission, diagnosis and treatment during hospitalization, diseased vessel identified during procedure, diagnosis at discharge and drug treatment, and comorbidities, such as hypertension, diabetes, and renal disease; in total, 41 characteristics were analyzed. Based on TRIPOD reporting guidelines, the rule of thumb for sample size is to have at least 10 outcome events per variable (EPV).

2.4. Data Collection. In our data source, all attributes that can be subdivided are categorized into independent classes, and each class generates a new attribute. The new attribute is encoded with the one-hot encoding rule. The data were susceptible to incorrect notation by the researcher; data cleansing and editing, consisting of removing typographical errors, and reviewing data quality in data reporting, were performed by a second researcher to avoid a flawed model training process. Assessment of predictors in our study has been performed without knowledge of the participant's outcome. A single investigator assessed all demographic information and clinical data and was blinded to the outcome of mortality. Additionally, a different researcher assessed the plausibility of the results regarding the outcome of mortality.

2.5. Missing Values. Complete case data were collected from the electronic health records (EHRs) and analyzed; all variables can be queried in the EHRs. Some patients were excluded as they refused to undergo the candidate predictor laboratory test or failed to comply with 1-year follow-up.

2.6. Statistical Analysis. Continuous variables are presented as the mean ± standard deviation, and classified variables are indicated by counts and percentages. Differences in baseline characteristics between groups were analyzed using the independent sample *t*-test. The Mann–Whitney *U* test was used for continuous variables, and the chi-square test or Fisher's exact test was used for categorical variables. The previously described GRACE risk score was used to analyze mortality, and it was calculated according to the published formula [13]. Five machine learning classifiers (LR, KNN, CatBoost, RF, and XGBoost) and the ensemble model were used as the supervised machine learning methods to predict survival status after 1-year follow-up. In order to solve the

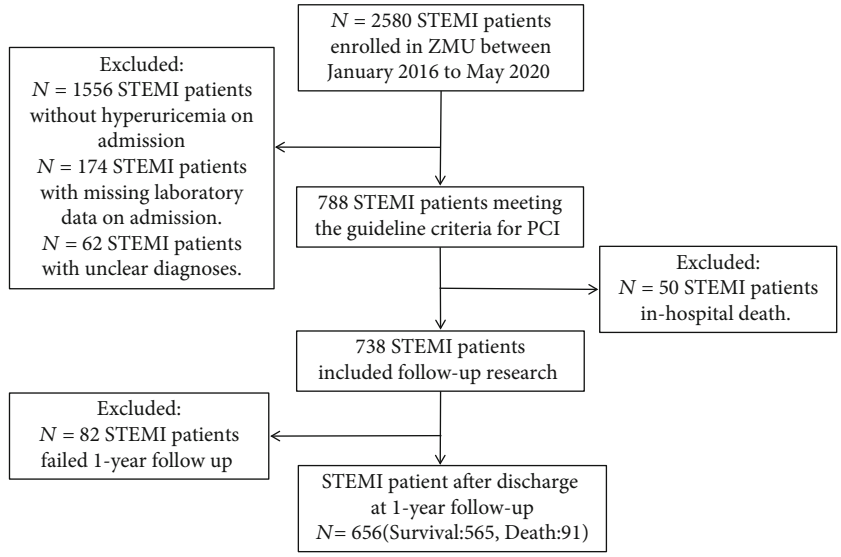


Figure 1: A flow diagram showing the study process.

problem of imbalanced data classification owing to medical diagnosis, we used SMOTEENN, a hybrid sampling algorithm of SMOTE and ENN algorithms; this helped to oversample the minority (death cases) class by creating synthetic samples, followed by cleaning the mislabeled instances. Supervised learning aims to establish a concise model of outcome type distribution (called label in machine learning), based on predictor parameters [14]. All models were validated by 10-fold crossvalidation. In feature engineering, all classification features were transformed by one-hot encoding, and the missing values were provided by the missForest method. Compared with the traditional chain multifilling method, this method results in significant performance improvement [14–16]. The following indicators were used to define model performance: AUC, recall, precision, and F1 value. Python (version 3.7, https://www.python.org/) was used for all statistical analyses.

3. Results and Discussion

Between January 2016 and January 2020, a total of 738 STEMI patients registered in the database met the inclusion criteria. After excluding those who were lost to follow-up ($n = 82$), 656 patients were enrolled in this study. The patients' average age was 62.5 years (±13.6 years), and 83.6% were male. All patients underwent emergency PCI. The median follow-up duration was 25 months, and 91 patients died within 1 year of admission, resulting in a mortality rate of 13.8%. Table 1 summarizes the differences in demographic information, admission baseline characteristics, and diseased vessels between the patients who survived and those who died. Considering the imbalance of classification data among samples (death cases : survival cases = 91 : 565), five machine learning algorithms (logistic regression, KNN, RF, XGBoost, and CatBoost) were developed to predict the 1-year mortality rate with all available features. RF (accuracy = 0.89, AUC = 0.88) and CatBoost (accuracy = 0.89, AUC = 0.87) provided similar AUC values in our study, and the predicted performance was higher than that of the traditional GRACE score. As a traditional risk assessment tool, GRACE (accuracy = 0.84, AUC = 0.80) also showed good discriminatory ability in our study (Table 2). The RF classifier outperformed the other models in terms of the AUC crossvalidation results (Figure 2). This study used SMOTEENN to further optimize the models; thus, the performance of all machine learning models was improved significantly (Table 2, Figure 3). After using SMOTEENN to generate more minority class samples, the CatBoost model (accuracy = 0.96, AUC = 0.99, recall = 0.98, precision = 0.95 , F1 value = 0.97) demonstrated the highest performance (Figure 4). We investigated the possibility of combining different models to improve performance. In particular, we tried several ensembles and combination methods, including training of the above classifiers and combining their predictions to check whether combination is better than any single classifier (Table 3). The CatBoost was separately integrated with Bagging and Boosting. Further, when the prediction probability of each model was used as the combination rule through the combination of LR, KNN, and XGBoost models after 10-fold crossvalidation, the performance of some models partially improved (recall from 0.33 to 0.53; F1 value from 0.44 to 0.58) compared with that of a single model. This shows that different models can be regarded as partially complementary. When the other abovementioned models were included in the integration method according to different combinations, very similar results were obtained.

Owing to the recent widespread development of chest pain centers in China, 70.8% of patients with acute STEMI were admitted to the hospital within 12 hours of onset and received prompt reperfusion treatment. Hospital mortality rates have therefore decreased significantly. Timely and effective revascularization treatment is key for the reduction of mortality and improved prognosis following AMI. The rescue system based on chest pain centers has played an essential role in improving the timeliness of revascularization in AMI patients and in reducing mortality.

Table 1: Comparison of characteristics of patients with and without mortality in the cohort.

Variables	Total ($n = 656$)	Survival ($n = 565$)	Death ($n = 91$)	P value
Demographic characteristics				
Sex, n (%)				0.008
Female	107 (16)	83 (15)	24 (26)	
Male	549 (84)	482 (85)	67 (74)	
Age, y	64.00 (52, 74)	63.00 (51, 73)	70.00 (59, 78)	<0.001
Smoking, n (%)	453 (69)	396 (70)	57 (63)	0.192
Weekend on admission, n (%)	248 (38)	205 (36)	43 (47)	0.059
Delay, n (%)	167 (25)	133 (24)	34 (37)	0.007
Vascular risk factors				
Hypertension, n (%)	380 (58)	326 (58)	54 (59)	0.857
Diabetes mellitus, n (%)	121 (18)	98 (17)	23 (25)	0.096
Prior-stroke, n (%)	35 (5)	30 (5)	5 (5)	1
CKD, n (%)	152 (23)	122 (22)	30 (33)	0.024
Clinical data				
HR, beats/min	80 (72, 92)	80.00 (72, 91)	85 (73, 106)	0.003
SBP, mmHg	124 (108, 140)	127 (110, 143)	111 (92, 129)	<0.001
DBP, mmHg	80 (68, 91)	80 (70, 92)	74 (58, 85)	<0.001
Shock_index	0.65 (0.55, 0.77)	0.64 (0.54, 0.75)	0.75 (0.61, 1.04)	<0.001
Electrocardiographic data				
Inferior, n (%)	300 (46)	263 (47)	37 (41)	0.351
Anterior, n (%)	322 (49)	276 (49)	46 (51)	0.851
Other, n (%)	21 (3)	16 (3)	5 (5)	0.194
Right ventricular, n (%)	7 (1)	6 (1)	1 (1)	1
Laboratory examinations on admission				
WBC, $^*10^9$/L	11.27 (8.60, 14.19)	10.97 (8.34, 13.57)	13.92 (10.56, 19.51)	<0.001
Neutrophil count, $^*10^9$/L	8.85 (6.34, 11.83)	8.46 (6.11, 11.19)	11.40 (8.43, 16.26)	<0.001
NLR	6.65 (3.89, 10.77)	6.25 (3.78, 9.85)	9.74 (5.89, 14.93)	<0.001
PLR	149.03 (104.31, 224.60)	148.96 (107.43, 220.27)	151.40 (81.96, 250.07)	0.518
MLR	0.54 (0.37, 0.82)	0.51 (0.36, 0.76)	0.75 (0.41, 1.12)	<0.001
SIRI	4.51 (2.63, 8.44)	4.19 (2.47, 7.39)	8.41 (4.38, 15.27)	<0.001
SII	1285.43 (746.84, 2247.28)	1233.05 (735.30, 2139.17)	1923.99 (894.50, 2898.80)	0.003
HB, g/L	139.00 (123.00, 154.00)	140.00 (124.00, 155.00)	128.00 (115.00, 147.00)	0.001
RBC, $^*10^{12}$/L	4.54 (3.98, 5.01)	4.58 (4.05, 5.02)	4.21 (3.71, 4.88)	0.006
PLT, $^*10^9$/L	205.00 (161.00, 249.25)	207.00 (164.00, 250.00)	196.00 (138.00, 246.50)	0.113
ALT, U/L	33.00 (23.00, 56.00)	32.00 (22.25, 51.75)	56.00 (30.00, 193.00)	<0.001
AST, U/L	72.00 (36.50, 169.5)	67.00 (35.00, 143.00)	225.00 (73.50, 456.00)	<0.001
GGT, U/L	44.00 (27.00, 75.00)	43.00 (27.00, 72.75)	61.00 (29.00, 104.00)	0.007
BUN, mmol/L	6.72 (5.25, 9.37)	6.38 (5.09, 8.50)	10.33 (7.45, 13.15)	<0.001
Creatinine, umol/L	101.00 (82.00, 128.00)	98.00 (81.00, 119.00)	134.00 (109.00, 174.50)	<0.001
Uric acid, umol/L	484.00 (449.00, 542.00)	481.00 (447.00, 535.00)	523.00 (461.00, 637.00)	<0.001
Cystatin C, mg/L	1.22 (0.97, 1.58)	1.17 (0.95, 1.49)	1.65 (1.32, 2.18)	<0.001
CK, U/L	507.00 (186.00, 1368.75)	463.50 (172.00, 1322.50)	745.00 (303.25, 2012.50)	0.002
CKMB, U/L	52.00 (25.00, 127.00)	48.00 (24.00, 117.25)	86.00 (33.00, 190.00)	<0.001
LDH, U/L	375.00 (266.25, 639.75)	350.50 (255.25, 556.75)	695.50 (407.75, 1229.75)	<0.001
α-HBDH, U/L	259.00 (173.00, 475.00)	240.00 (165.00, 427.50)	490.50 (273.75, 773.00)	<0.001
CTnT, ng/L	1014.00 (213.50, 3480.00)	786.95 (185.97, 3069.00)	3077.00 (1133.00, 6711.00)	<0.001
BNP, pg/mL	1022.50 (255.15, 3860.75)	884.90 (204.85, 2713.00)	5349.00 (2058.00, 15267.00)	<0.001
Glucose, mmol/L	6.66 (5.56, 8.66)	6.52 (5.44, 8.19)	8.47 (6.41, 11.60)	<0.001

Table 1: Continued.

Variables	Total ($n = 656$)	Survival ($n = 565$)	Death ($n = 91$)	P value
Myoglobin, ng/mL	341.10 (104.50, 910.40)	308.95 (95.96, 820.28)	615.00 (203.50, 2251.00)	<0.001
Diseased vessel identified during procedure				
LM, *n* (%)	13 (2)	13 (2)	0 (0)	0.233
LAD, *n* (%)	213 (33)	185 (33)	28 (31)	0.836
LCX, *n* (%)	70 (11)	62 (11)	8 (9)	0.674
RCA, *n* (%)	157 (24)	134 (24)	23 (26)	0.819
Risk assessment				
GRACE, score	125.00 (102.00, 154.00)	121.00 (101.00, 146.00)	178.00 (140.00, 206.50)	<0.001

Values are expressed as medians with interquartile ranges for continuous data. Other values are presented as numbers and percentages. Shock index: ratio of HR to SBP; SIRI: systemic inflammatory response index; SII: systemic inflammatory reaction index; PLR: ratio of platelets to lymphocytes; NLR: the ratio of neutrophils to lymphocytes; MLR: ratio of monocytes to lymphocytes; OHCA: out-of-hospital cardiac arrest; GRACE: Global Registry of Acute Coronary Events score; α-HBDH: α-hydroxybutyrate dehydrogenase; BNP: B-type natriuretic peptides.

Table 2: Comparison of validation results of machine learning models.

Models	Accuracy	AUC	Recall	Precision	F1 value
CatBoost	0.89	0.87	0.33	0.78	0.44
RF	0.89	**0.88**	0.26	**0.82**	0.38
XGBoost	**0.90**	0.83	**0.41**	0.81	**0.51**
LR	0.89	0.82	0.38	0.63	0.46
KNN	0.88	0.75	0.21	0.61	0.31
Model with oversampling (SMOTEENN)					
CatBoost	0.96	0.99	0.98	0.95	0.97
RF	0.95	0.99	0.98	0.94	0.96
XGBoost	0.94	0.98	0.98	0.92	0.95
LR	0.91	0.95	0.92	0.92	0.92
KNN	0.92	0.96	0.98	0.88	0.93
Tradition risk score model					
GRACE score	0.84	0.80	0.46	0.59	0.51

AUC and F1 score: the higher, the better. XGBoost: Extreme Gradient Boosting; RF: random forest; LR: logistic regression; KNN: K-nearest neighbors.

Previous studies have confirmed that baseline renal dysfunction and acute kidney injury are strong predictors of in-hospital and long-term adverse cardiovascular outcomes after STEMI complicated by cardiogenic shock [17]. STEMI-related mortality is considerably higher in those who have had unsuccessful invasive procedures or those with diabetes, chronic kidney failure, or high serum lactate or glucose levels [17, 18].

UA is the final product of purine metabolism and is metabolized by xanthine oxidase. Hyperuricemia can lead to gout and nephrolithiasis; it has also been implicated as an indicator for diseases, such as the metabolic syndrome, diabetes mellitus, cardiovascular disease, and chronic renal disease. Previous studies have suggested that hyperuricemia with STEMI is associated with a poor prognosis and a high incidence of death and major adverse cardiovascular events (MACEs) [19]. Although the pathophysiological mechanisms of adverse reactions to hyperuricemia have not been fully elucidated, it appears to be multifactorial. In the light of the experimental evidence, hyperuricemia was linked to a variety of proatherogenic processes, including increased oxidative stress, inhibition of endothelial nitric oxide, activation of the renin-angiotensin system, and increase in the microvascular damage via endothelial dysfunction and vascular smooth muscle cell proliferation [20–23].

There is currently no effective evaluation method to predict the long-term prognosis of these patients. GRACE risk scores can be used to estimate follow-up results after acute coronary syndrome. Although Asian populations were not included during the development of the model, the use of GRACE revealed a good discriminatory accuracy in predicting both short-term and long-term MACEs in Asian patients with MI [24]. Our cohort had a median follow-up duration of 25 months, similar to those of previously published studies (accuracy = 0.84, AUC = 0.8). However, the statistical methods in these traditional assessment tools include the Cox proportional hazard regression model. Researchers make presumptions and employ subjective feature selection before model fitting, potentially leading to loss of information [15]. As we enter the era of precision medicine, the demand for risk assessment tools has gained importance. In cases where the research goal is to generate a model that

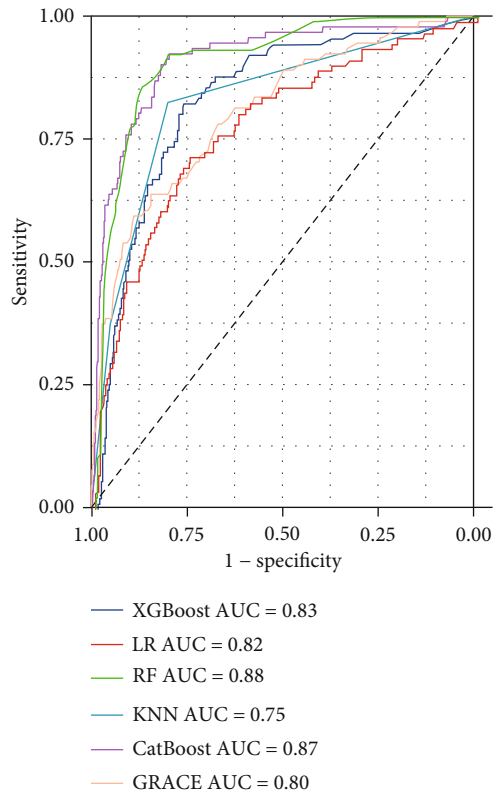


Figure 2: ROC analysis result of five classifiers and GRACE for the prediction of 1-year mortality with all available features.

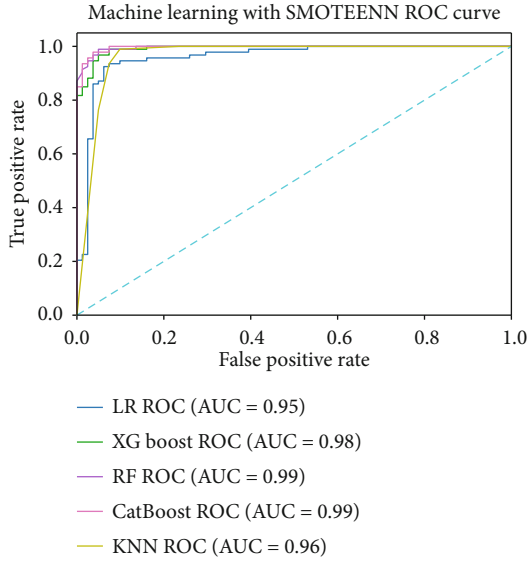


Figure 3: ROC analysis result of five classifiers with SMOTEENN.

can predict the results most accurately, machine learning algorithms may be more advantageous compared to traditional regression methods. First, machine learning methods can compute multiple related predictions, nonlinear relationships, and the internal interaction between predictors and end events in large datasets. Second, as a critical component of the TRIPOD original declaration report [25], the model should be verified after establishment. In cases where the machine learning method is used, model performance is more robust after external verification. In this study, we compared several standard machine learning methods and performed 10-fold crossinternal verification of the dataset in the absence of external data to ensure model robustness. However, in the traditional regression model, internal validation is not necessary, because one (ideally) posits an analytic model before fitting it to the data [15]. Considering the different effects of each machine learning method in solving medical professional problems, this study compares the efficiency and robustness of various machine learning methods with that of the traditional risk score to obtain more cautious results.

In previous studies, machine learning methods showed a better ability to predict short-term mortality after STEMI, while XGBoost showed better predictive ability than other machine learning models in patients with anterior wall STEMI [14]. Gradient boosted tree (GBT) methods, such as XGBoost, RF, and CatBoost, provided similar AUC values in our study. However, after model optimization, the CatBoost model showed more accurate prediction ability. The CatBoost algorithm, which was released in 2017, LightGBM, and XGBoost are the three mainstream machine learning methods for GBT. The CatBoost algorithm is a GBT framework based on an asymmetrical decision tree (oblivious trees) algorithm, with only a few parameters; it supports class variables and has high accuracy. It mainly addresses the issue of dealing with category features efficiently and reasonably.

Furthermore, to improve the algorithm's accuracy and generalization ability, a new method was proposed to account for gradient deviation (gradient bias) and prediction partial (prediction shift) problems. As a new algorithm released in 2017, this method can account for category features in clinical practice and can effectively prevent overfitting; its high training accuracy has provoked widespread interest. Our study also demonstrated the high accuracy of the model. Interestingly, under the premise of the imbalance of clinical samples, the machine learning method with the oversampling technique SMOTEENN could significantly improve performance. SMOTEENN is a hybrid sampling technique of SMOTE and ENN algorithms, that is often employed to oversample the minority class by creating synthetic samples, followed by cleaning of mislabeled instances [26]. It is essential to be aware of the dramatic effects of these synthetic sampling techniques on machine learning models.

Our research has several limitations. Owing to the retrospective design of this study, the process of patient data collection may have been accompanied by a risk of bias. Further, this was a single-center study, including only Chinese patients. Under the premise of the imbalance of clinical samples, the machine learning method based on clinical data alone could not obtain a higher AUC value; even the oversampling technique could not significantly improve performance. Second, although the machine model based on hybrid sampling technology has achieved excellent performance in this study, the samples in hybrid sampling technology are computer-generated samples and not real patients; thus, making a more accurate assessment of the prognosis

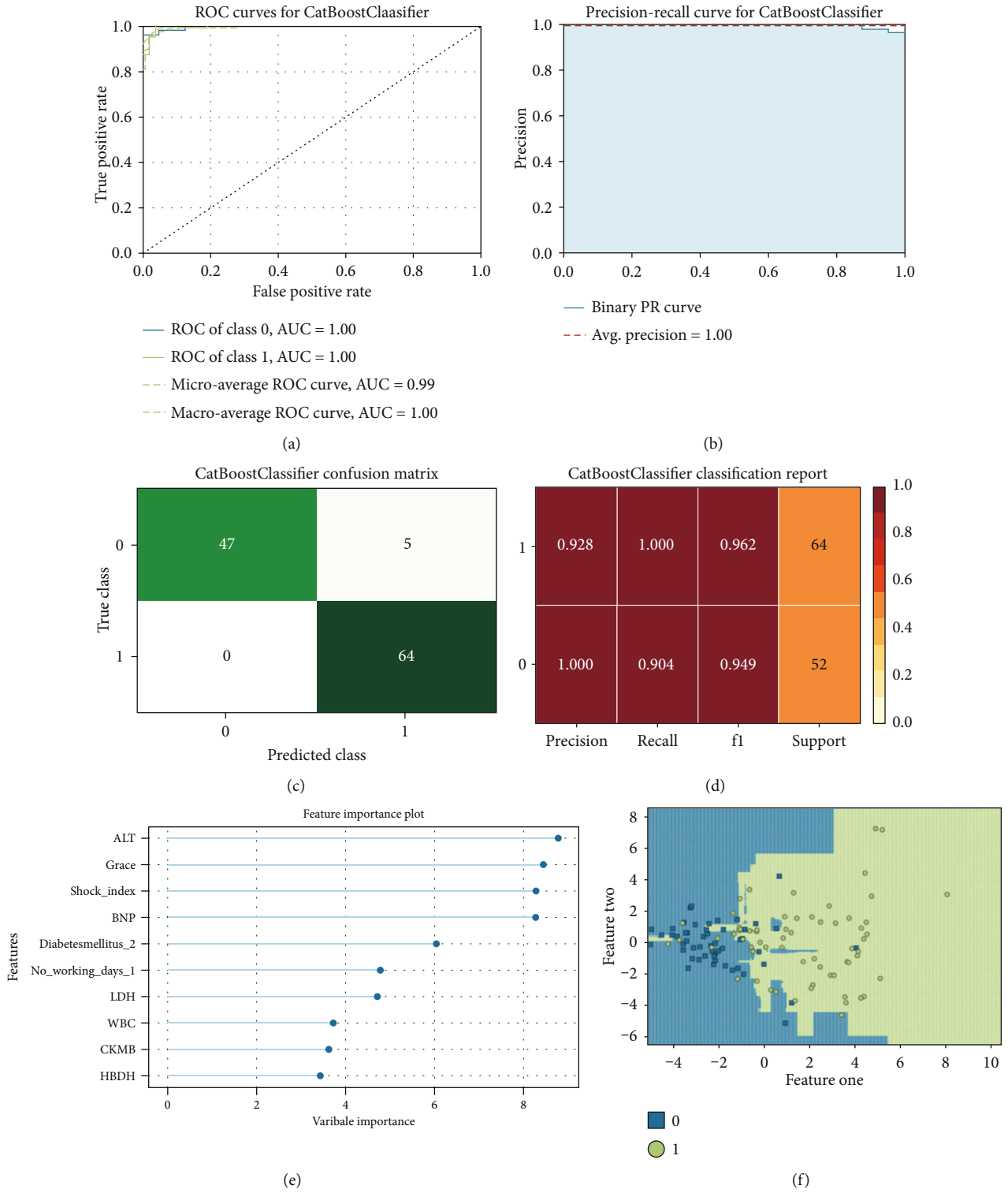


Figure 4: CatBoost model performance visualization. (a) ROC curve. (b) Precision-recall curve. (c) Confusion matrix. (d) Classification report. (e) Feature importance. (f) Decision boundary.

Table 3: Ensemble of machine learning models.

Ensemble	Accuracy	AUC	Recall	Precision	F1 value
RF+CatBoost+XGBoost	0.90	0.87	0.36	0.72	0.47
XGBoost+LR+KNN	0.90	0.84	0.33	0.72	0.43
RF+LR+KNN	0.89	0.86	0.30	0.71	0.39
RF+XGBoost+LR+KNN	0.90	0.86	0.37	0.73	0.46
All	0.90	0.86	0.37	0.73	0.47

of STEMI patients with hyperuricemia using big clinical data requires further analysis using a more extensive dataset. Despite the abovementioned limitations, our study also has some strengths. The results provide an effective and robust method for predicting 1-year mortality in patients with STEMI complicated by hyperuricemia, through the crossvalidation of machine learning models. Further study requires the combination of social factors, environmental parameters, and phenotypic information (such as genome or proteomics data) in MI for prognostic prediction.

4. Conclusion

In conclusion, the predictive ability of machine learning methods is significantly higher than that of the traditional statistical scoring model. The machine learning model will be helpful for the prediction and early detection of MACEs in patients with STEMI complicated by hyperuricemia. In addition, in cases of clinically unbalanced samples, the oversampling technology can significantly improve model performance and ability; however, it is essential to be aware of the dramatic effects of the synthetic sampling techniques on models. There is still uncharted territory in clinical medicine, and methods for accurately predicting the occurrence of some diseases or adverse events will remain the enduring focus of clinical research. Although machine learning presently appears to have good predictive effect, further reasonable and scientific verification is required.

Data Availability

The datasets are available from the corresponding author upon reasonable request.

Ethical Approval

The study was locally approved by the Ethics Committee of Affiliated Hospital of Zunyi Medical University (approval no. KLL[2020]0144).

Consent

The need to obtain written informed consent from the patients was waived because of the study's retrospective nature.

Conflicts of Interest

The authors have no disclosures to make with respect to this manuscript.

Authors' Contributions

Ranzun Zhao, Zhijiang Liu, Zhenglong Wang, and Yi Ma are responsible for the acquisition, analysis, or interpretation of data; Zhixun Bai for drafting of the manuscript; Bei Shi for the critical revision of the manuscript for important intellectual content; Zhixun Bai, Jing Lu, and Ting Li for the statistical analysis; Zhixun Bai, Jing Lu, and Ting Li served as co-first authors; and Zhixun Bai, Jing Lu, and Ting Li contributed equally to this work.

References

[1] J. C. Kwong, K. L. Schwartz, M. A. Campitelli et al., "Acute myocardial infarction after laboratory-confirmed influenza infection," *The New England Journal of Medicine*, vol. 378, no. 4, pp. 345–353, 2018.

[2] A. Srivastava, A. D. Kaze, C. J. McMullan, T. Isakova, and S. S. Waikar, "Uric acid and the risks of kidney failure and death in individuals with CKD," *American Journal of Kidney Diseases*, vol. 71, no. 3, pp. 362–370, 2018.

[3] L. G. Sanchez-Lozada, B. Rodriguez-Iturbe, E. E. Kelley et al., "Uric acid and hypertension: an update with recommendations," *American Journal of Hypertension*, vol. 33, no. 7, pp. 583–594, 2020.

[4] G. Ndrepepa, "Uric acid and cardiovascular disease," *Clinica Chimica Acta*, vol. 484, pp. 150–163, 2018.

[5] B. Lacey, W. G. Herrington, D. Preiss, S. Lewington, and J. Armitage, "The role of emerging risk factors in cardiovascular outcomes," *Current Atherosclerosis Reports*, vol. 19, no. 6, p. 28, 2017.

[6] A. E. Berezin, "Is serum uric acid a pretty accurate prognostic predictor of ST elevated acute coronary syndrome?," *International Journal of Cardiology*, vol. 254, p. 49, 2018.

[7] F PM, "2016 European guidelines on cardiovascular disease prevention in clinical practice : the Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)," *International Journal of Behavioral Medicine*, vol. 24, no. 3, pp. 321–419, 2017.

[8] W. Guo, The RESCIND Group, D. Yang et al., "Hyperuricemia and long-term mortality in patients with acute myocardial infarction undergoing percutaneous coronary intervention," *Annals of Translational Medicine*, vol. 7, no. 22, p. 636, 2019.

[9] M. Magnoni, M. Berteotti, F. Ceriotti et al., "Serum uric acid on admission predicts in-hospital mortality in patients with acute coronary syndrome," *International Journal of Cardiology*, vol. 240, pp. 25–29, 2017.

[10] M. Tscharre, R. Herman, M. Rohla et al., "Uric acid is associated with long-term adverse cardiovascular outcomes in

patients with acute coronary syndrome undergoing percutaneous coronary intervention," *Atherosclerosis*, vol. 270, pp. 173–179, 2018.

[11] M. Chegini, J. Bernard, P. Berger, A. Sourin, K. Andrews, and T. Schreck, "Interactive labelling of a multivariate dataset for supervised machine learning using linked visualisations, clustering, and active learning," *Visual Informatics*, vol. 3, no. 1, pp. 9–17, 2019.

[12] G. S. Collins, J. B. Reitsma, D. G. Altman, K. G. Moons, and TRIPOD Group, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)," *Circulation*, vol. 131, no. 2, pp. 211–219, 2015.

[13] C. B. Granger, R. J. Goldberg, O. Dabbous et al., "Predictors of hospital mortality in the global registry of acute coronary events," *Archives of Internal Medicine*, vol. 163, no. 19, pp. 2345–2353, 2003.

[14] Y. M. Li, L. C. Jiang, J. J. He, K. Y. Jia, Y. Peng, and M. Chen, "Machine learning to predict the 1-year mortality rate after acute anterior myocardial infarction in Chinese patients," *Therapeutics and Clinical Risk Management*, vol. Volume 16, pp. 1–6, 2020.

[15] B. A. Goldstein, A. M. Navar, and R. E. Carter, "Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges," *European Heart Journal*, vol. 38, no. 23, pp. 1805–1814, 2017.

[16] D. J. Stekhoven and P. Buhlmann, "MissForest–non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012.

[17] M. I. Hayiroglu, E. Bozbeyoglu, O. Yildirimturk, A. I. Tekkesin, and S. Pehlivanoglu, "Effect of acute kidney injury on long-term mortality in patients with ST-segment elevation myocardial infarction complicated by cardiogenic shock who underwent primary percutaneous coronary intervention in a high-volume tertiary center," *Türk Kardiyoloji Derneği Arşivi*, vol. 48, no. 1, pp. 1–9, 2020.

[18] M. I. Hayiroglu, Y. Canga, O. Yildirimturk et al., "Clinical characteristics and outcomes of acute coronary syndrome patients with intra-aortic balloon pump inserted in intensive cardiac care unit of a tertiary clinic," *Türk Kardiyoloji Derneği Arşivi*, vol. 46, no. 1, pp. 10–17, 2018.

[19] W. Guo, Y. Liu, J. Y. Chen et al., "Hyperuricemia is an independent predictor of contrast-induced acute kidney injury and mortality in patients undergoing percutaneous coronary intervention," *Angiology*, vol. 66, no. 8, pp. 721–726, 2015.

[20] F. Barkas, M. Elisaf, E. Liberopoulos, R. Kalaitzidis, and G. Liamis, "Uric acid and incident chronic kidney disease in dyslipidemic individuals," *Current Medical Research and Opinion*, vol. 34, no. 7, pp. 1193–1199, 2018.

[21] A. Testa, F. Mallamaci, B. Spoto et al., "Association of a polymorphism in a gene encoding a urate transporter with CKD progression," *Clinical Journal of the American Society of Nephrology*, vol. 9, no. 6, pp. 1059–1065, 2014.

[22] K. Hughes, T. Flynn, J. de Zoysa, N. Dalbeth, and T. R. Merriman, "Mendelian randomization analysis associates increased serum urate, due to genetic variation in uric acid transporters, with improved renal function," *Kidney International*, vol. 85, no. 2, pp. 344–351, 2014.

[23] the RESCIND group, W. Guo, F. Song et al., "The relationship between hyperuricemia and contrast-induced acute kidney injury undergoing primary percutaneous coronary intervention: secondary analysis protocol for the ATTEMPT RESCIND-1 study," *Trials*, vol. 21, no. 1, p. 567, 2020.

[24] Y. H. Chen, S. S. Huang, and S. J. Lin, "TIMI and GRACE risk scores predict both short-term and long-term outcomes in Chinese patients with acute myocardial infarction," *Acta Cardiologica Sinica*, vol. 34, no. 1, pp. 4–12, 2018.

[25] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement," *BMJ*, vol. 350, no. jan07 4, 2014.

[26] G. Idakwo, S. Thangapandian, J. Luttrell et al., "Structure-activity relationship-based chemical classification of highly imbalanced Tox 21 datasets," *Journal of Cheminformatics*, vol. 12, no. 1, p. 66, 2020.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 6691177, 10 pages
https://doi.org/10.1155/2021/6691177

Research Article

Automated Atrial Fibrillation Detection Based on Feature Fusion Using Discriminant Canonical Correlation Analysis

Jingjing Shi[1], **Chao Chen**[1], **Hui Liu**[1], **Yinglong Wang**[1], **Minglei Shu**[1] and **Qing Zhu**[2]

[1]*Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Sciences), China*
[2]*Qilu Hospital of Shandong University, China*

Correspondence should be addressed to Qing Zhu; 198862000790@email.sdu.edu.cn

Received 21 December 2020; Revised 5 March 2021; Accepted 26 March 2021; Published 9 April 2021

Academic Editor: Mario Cesarelli

Atrial fibrillation (AF) is one of the most common cardiovascular diseases, with a high disability rate and mortality rate. The early detection and treatment of atrial fibrillation have great clinical significance. In this paper, a multiple feature fusion is proposed to screen out AF recordings from single lead short electrocardiogram (ECG) recordings. The proposed method uses discriminant canonical correlation analysis (DCCA) feature fusion. It fully takes intraclass correlation and interclass correlation into consideration and solves the problem of computation and information redundancy with simple series or parallel feature fusion. The DCCA integrates traditional features extracted by expert knowledge and deep learning features extracted by the residual network and gated recurrent unit network to improve the low accuracy of a single feature. Based on the Cardiology Challenge 2017 dataset, the experiments are designed to verify the effectiveness of the proposed algorithm. In the experiments, the F1 index can reach 88%. The accuracy, sensitivity, and specificity are 91.7%, 90.4%, and 93.2%, respectively.

1. Introduction

Atrial fibrillation (AF) is the most common persistent cardiovascular disease, which can easily lead to strokes, hemiplegia, and other diseases, seriously threatening patients' health; thus, timely diagnosis and treatment are necessary. However, owing to the shortage of medical resources and the single model of doctor diagnosis, it becomes urgent to improve automatic detection technology. Automatic detection of cardiac rhythm is a meaningful and important issue in different age groups, including adults [1] and fetuses [2]. Computational techniques and deep learning methods detecting various types of arrhythmia have been widely developed to analyse ECG signals and are strong candidates to help clinical advances by providing a better understanding of medical challenges [3, 4]. With the development of medicine, people have gained more understanding of the physiological mechanism of atrial fibrillation, but further research is still needed [5]. Physiologically, the occurrence of atrial fibrillation is due to irregular atrial contraction, which is reflected in the electrocardiogram: P waves disappear, irregular fibrillation waves (f waves) of different sizes and shapes appear [6, 7], and there is a severe irregularity of the RR interval.

The detection of atrial fibrillation signals is mainly divided into four parts, including data preprocessing, feature extraction, feature selection, and classification. Among them, feature extraction directly affects the accuracy and efficiency of atrial fibrillation signal classification. Commonly used feature extraction in the literature usually falls into two categories, traditional feature extraction and feature extraction based on deep learning methods. Traditional feature extraction methods are generally divided into three categories. The first is to extract the statistical characteristics of ECG signals, that is, use the statistical data to summarize a series of ECG data. Typical statistics include mean, maximum, minimum, variance, skewness, kurtosis, count, and percentage.

Kaya et al. [8] calculated the statistical and time characteristics of a heartbeat, such as skewness, kurtosis, standard, deviation, and average, and they used the best feature reduction and classification methods, the highest classification accuracy, sensitivity, and specificity rates of 99.30%, 98.84%, and 98.40%, respectively. Athif et al. [9] extracted statistical and morphological features and then used a support vector machine classifier to classify records into three categories: "normal," "AF," and "other." The algorithm has a sensitivity of 77.5%, a specificity of 97.9%, and an accuracy of 96.1% in the "Computing in Cardiology Challenge 2017" database. The second is signal processing, which is to transform the ECG data from the time domain to the frequency domain or other domains through discrete Fourier transform, discrete wavelet transform, and other methods. Yin et al. [10] proposed a multidomain ECG feature extraction method. The RR intervals were extracted as time domain feature. The fifth-order approximate coefficients of wavelet decomposition are used to represent the frequency domain features. In addition, the sample entropy values of six wavelet coefficients are used as nonlinear characteristics. These three features were fed to a classifier for automated diagnosis. The average accuracy of the SVM classifier in the MIT-BIH arrhythmia database was 99.70%. The third is to directly extract the time domain or morphological features of ECG signals, including RR interval, QRS wave width, and PR interval. Dash et al. [11] used a statistical method to evaluate the complexity, randomness, and variability of the RR interval. Verification by the MIT-BIH atrial fibrillation database shows the sensitivity is 94.4%, and the specificity is 95.1%. Zabihi et al. [12] adopted time-frequency, phase space, tuples, and other characteristics in multiple fields and used a random forest classifier for feature selection. F1 was 82.6% on the PhysioNet Challenge 2017 atrial fibrillation competition database. Deep learning feature extraction and classification include convolutional neural work (CNN) [13, 14] and long and short memory networks (LSTM) [15, 16] as well as their variants [17, 18]. Warrick and Homsi [19] combined convolutional neural networks and long short-term memory networks (LSTM) and used pooling, step size, and normalization techniques to improve its accuracy. The network predicts a classification every 18 and then selects the final prediction for classification. The total F1 on the PhysioNet Challenge 2017 dataset is 80%.

With the rapid development of deep learning, the advantages of feature-level fusion have become more and more obvious. In recent years, some researchers have used feature fusion for ECG signal detection. Smoleń [20] first used a sequential Recurrent Neural Network (RNN) classifier to get the probabilities for each class and then combined the probabilities with hand-designed features. Finally, F1 is 79% in PhysioNet Challenge 2017 (CinC 2017). Chu et al. [21] proposed a new method for arrhythmia classification based on multilead ECG signals; the core of the design is to fuse two types of deep learning features with some common traditional features and then use a support vector machine (SVM) classifier to classify the feature vectors, and according to the AAMI standard, the accuracy on the 12-lead INCAET dataset is 88.565%. Ghiasi et al. [22] proposed two different classification methods, of which the first is a feature-based method, and the second adopts a deep neural network. Finally, they used the decision table to combine the output results of the two methods and divided all records into three categories. The proposed method is evaluated using a scoring function from the 2017 PhysioNet/CinC Challenge and achieved an overall score of 80% and 71% on the training dataset and hidden test dataset.

This paper presents a robust method capable of detecting AF from single short ECG lead recording. Here are the four main contributions of this paper: (1) novel combination of deep learning and the traditional features; (2) proposed an improved residual network and gated recurrent unit network, which extracted deep learning features in spatial and time series; (3) performing ECG feature fusion used discriminant canonical correlation analysis; and (4) achieving superior classification results compared to the above-cited method of the same database [23–27].

The structure of this paper is as follows: Section 2 introduces the feature extraction method, Section 3 presents the feature fusion method, Section 4 the performance metrics, Section 5 the experimental results and analysis, and Section 6 the summary.

2. Feature Extraction

This section mainly introduces deep learning feature extraction methods and traditional feature extraction methods based on expert knowledge.

2.1. Dataset. This article uses a large dataset released by the PhysioNet/CinC Challenge in 2017, which contains 8528 single-lead ECG records [28]. Each ECG record in the dataset is collected from an individual. Compared to most of the researches based on the relatively simple dataset, such dataset is of higher research significance. These records are collected by AliveCor equipment. The dataset consists of single-lead ECGs of 8528 subjects of different lengths (about 23,878 heartbeats). The categories include normal rhythm, atrial fibrillation rhythm, other rhythms, and noise. The data duration is 9-60 s. Table 1 shows the details of the database.

2.2. Data Preprocessing

2.2.1. Denoising and Padding. The Butterworth band-pass filter is used to denoise the original ECG. The frequency response of the Butterworth filter is maximally flat (i.e., has no ripples) in the passband and rolls off towards zero in the stopband [29]. The attenuation of the first-order filter is 6 dB per octave, and the attenuation rate of the sixth-order Butterworth filter is 36 dB per octave. Since the frequency range of the ECG signal is mainly concentrated in 0.5 Hz~45 Hz, the blocking frequency is set to 45 Hz here, and the frequency signal output above 45 Hz will be attenuated. Because the convolutional neural network requires the input data to have the same size, but the length of the electrical signal in the center of the dataset is 9 seconds to 61 seconds, the ECG signal should be padded with zeros to adapt to the model.

Table 1: The PhysioNet 2017 dataset.

Type	Recording	Average time length (s)
Normal	5076	31.9
AF	758	31.6
Other rhythm	2415	34.1
Noisy	279	27.1

2.2.2. Sample Balancing. Due to the uneven number of samples in the database, the number of normal rhythms and other rhythm samples is large, namely, 5076 and 2415, respectively, while the number of atrial fibrillation rhythms and noise samples is small, 758 and 279, respectively, which easily affect the performance of model training and overfitting occurs. In this paper, class_weight is used to balance the sample and it provides weights for each output class. The weight of normal and other signals is very small, while the weight of atrial fibrillation and noise signal is much bigger. The class_weight method uses balance, and its weight calculation method: n_samples/(n_classes ∗ np.bincount(y)), where n_classes = 4, np.bincount(y) is the total number of samples for a certain class, and n_sample is the total number of samples, which is 8528. After calculation, the weight of normal ECG recording is 0.42, the weight of the atrial fibrillation signal is 2.81, the other weights are 0.88, and the weight of noise is 7.64.

2.3. Deep Learning Feature Extraction. This paper adopts residual network and gated recurrent unit for deep learning network feature extraction, which can not only reduce the depth of the network and effectively prevent overfitting but also extract the timing characteristics of the signal while extracting their spatial characteristics. The specific network structure is shown in Figure 1.

To deal with the degradation of neural networks, the method of establishing identity mapping with residual structure simplifies the multilayer network into a shallower network. According to the characteristics of the residual network, a one-dimensional residual network suitable for processing atrial fibrillation signals is designed. The residual network consists of six residual convolution blocks. In the first two residual blocks, the filter is 16. The residual ConvBlock is composed of four convolution blocks and a one-dimensional average pooling layer. Each convolution block contains a one-dimensional convolution with a step length of 1, a batch normalization, a linear unit with leakage correction, and a spatial random loss. The active layer is finally followed by a one-dimensional average pooling layer, the commonly used batch normalization (BN), LeakyRelu, and SpatialDropout. The spatial random activation function prevents overfitting, which is more conducive to promoting independence between feature maps than dropout. The number of filters in every two residual blocks is doubled, and the convolution step length in each convolution block is 1. The data obtained through the residual network is input into the gated recurrent unit network, and the number of neurons is set to 32; finally, the output of the last hidden layer is extracted as the deep learning feature.

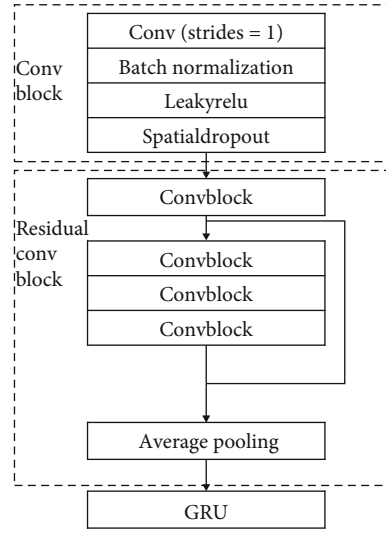


Figure 1: Deep learning feature extraction uses ResNet (residual network) and GRU (gated recurrent unit).

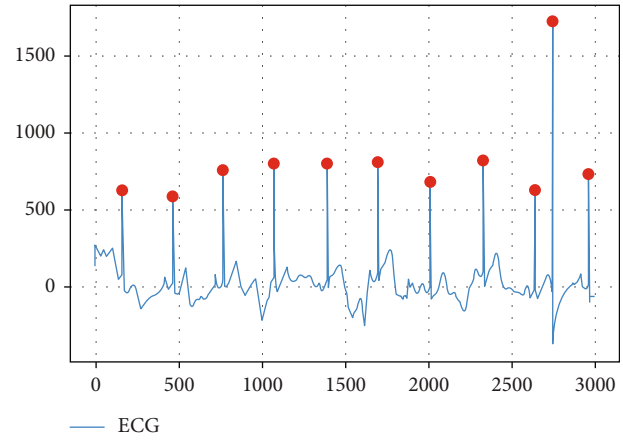


Figure 2: ECG detection algorithm detects QRS.

2.4. Traditional Feature Extraction. In fact, the ECG signal is used as input to extract relevant statistical features. First, the multilead differential electrocardiogram summation absolute value and adaptive threshold real-time detection algorithm [30] are used to detect QRS points. Taking A0003 in the dataset as an example, the corresponding waveform and the marked R wave are shown in Figure 2.

After the R wave is detected, the RR interval is calculated based on the R wave, and the RR interval is calculated as follows:

$$\mathrm{RRI} = \frac{R_{\mathrm{peaks}}(n+1) - R_{\mathrm{peaks}}(n)}{f_s}. \tag{1}$$

$R_{\mathrm{peaks}}(n)$ is the position of the nth R peak in the sample, and f_s is the sample rate. According to the RR interval and the traditional features of the ECG signal computed by QRS wave, these features are outputs as a feature vector. The RR interval and P wave are shown in Figure 3 [31].

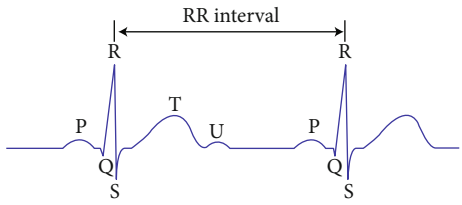


Figure 3: RR intervals and P waves [32].

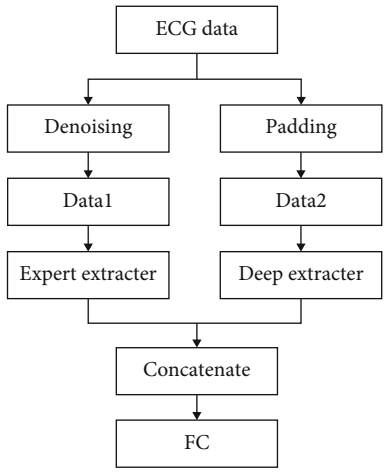


Figure 4: The structure of the proposed simple feature fusion.

2.4.1. RR Interval Feature. The statistical characteristics of RR intervals include standard deviation and variance, maximum RR interval, minimum RR interval, average RR interval, pNN50 (the proportion of the number of RR intervals in the ECG sequence whose RR interval difference is greater than 50 ms in all RR intervals), RMSSD (root mean square of the difference between the RR intervals), SDSD (standard deviation of the difference between the RR intervals), and the mean, variance, skewness, and kurtosis of each of the RR intervals divided into six segments.

2.4.2. P Wave Feature. The statistical characteristics of the P wave include the mean, variance, skewness, kurtosis, sample entropy, and sample entropy coefficient, and the P wave is divided into the average value, variance, and skewness of each of the six segments.

2.4.3. Signal Procession Feature. In order to extract the features of the ECG signal more comprehensively, we also extract the signal features based on the medical field and the frequency domain. These features first transform ECG data from time domain into frequency domain; then, frequency-related features are extracted. In the presented paper, the periodogram power spectral density (PSD) and energy spectral density are calculated. PSD is calculated using Fast Fourier Transform (FFT). After the transformation, energy within a specific range (band) is obtained. The chosen bands are between 5 frequencies: 0.1, 6, 12, 20, and 30 Hz. Another four features compute the variation based on QRS [1], compute the sample entropy (SampleEn) [2], compute the coefficient of variation and density histograms (CDF) [3], compute the thresholding on the median absolute deviation (MAD) [4], and compute the heart rate variability (variability).

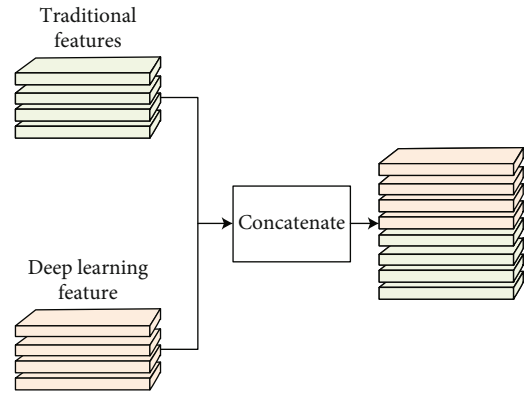


Figure 5: The specific process of simple feature fusion.

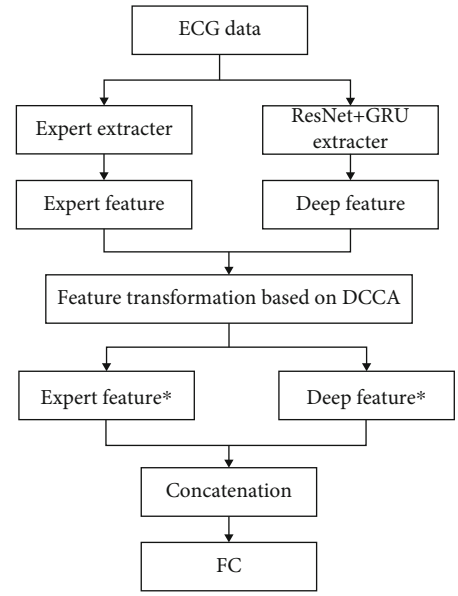


Figure 6: The structure of the proposed DCCA feature fusion.

3. Feature Fusion

3.1. Feature Fusion Based on Feature Concatenation. Based on expert knowledge, this model performs time domain and frequency domain feature extraction on the denoised ECG signal to obtain feature vectors. It uses a convolution residual network and gated recurrent unit to form a deep learning network, and input data filled ECG signal deep learning network to obtain deep feature vectors.

The two feature vectors obtained are fused into one feature vector in series and input into the classifier composed of the fully connected layers to classify ECG signals, as shown in Figures 4 and 5. This method is simple and but highly applicable. Compared with single feature extraction and classification [33], this method has improved accuracy [32]. However, since the method of fusion features is simple and

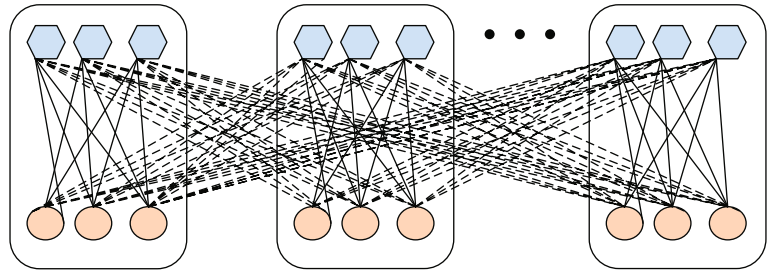


Figure 7: A graphical representation of the relationship between sample characteristics. Among them, hexagon and circle represent each feature, solid line represents the correlation within the class, and dashed line represents the correlation between classes.

rough, there are problems of redundancy and a large amount of calculation [34].

3.2. Feature Fusion Based on DCCA. In view of the shortcomings of the above-mentioned concatenation method, this section uses discriminant canonical correlation analysis (DCCA) [35] for feature fusion. DCCA is an improvement in canonical correlation analysis (CCA) [36]. The CCA feature fusion process does not consider the class structure. The DCCA method can not only optimize the correlation among the four types of samples but also minimize the correlation among the features of different types of samples. The proposed DCCA feature fusion method is shown in Figure 6.

In this paper, the discriminant canonical correlation analysis (DCCA) method is used for deep learning feature and traditional feature fusion, the preprocessed ECG signals are extracted separately to obtain two feature vectors, and then the DCCA method is used for feature fusion. The specific implementation is divided into four steps as follows:

(1) Find a set of projection direction w_x and w_y to achieve the maximum correlation among the features of samples of the same type and the minimum correlation among the features of different types of samples. Mathematically, DCCA is to maximize the correlation coefficient. The formula is as follows:

$$J_{\mathrm{d}}(w_x, w_y) = \frac{w_x^T \tilde{S}_{xy} w_y}{\sqrt{w_x^T S_{xx} w_x w_y^T S_{yy} w_y}}, \tag{2}$$

where $\tilde{S}_{xy} = S_w - \eta S_b$ (adjustable parameter > 0), S_w is the intraclass correlation matrix, S_b is the interclass correlation matrix, adjustable parameters η measure the relativity of the intraclass correlation and the interclass correlation of the sample characteristics, and the definitions of intraclass correlation and interclass correlation are shown in Figure 7

(2) Calculate the intraclass correlation matrix S_w and the interclass correlation matrix S_b, and set the processed sample set as

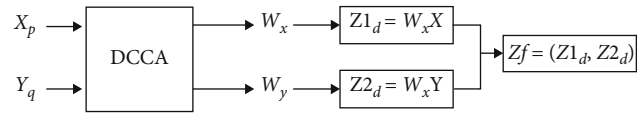


Figure 8: Block diagram for realizing canonical correlation analysis.

$$X = \left[x_1^{(1)}, \cdots x_{n1}^{(1)}, \cdots, x_1^{(c)}, x_{nc}^{(1)}\right] \in R^{p \times n},$$
$$Y = \left[y_1^{(1)}, \cdots, y_{n1}^{(1)}, \cdots, y_1^{(c)}, y_{nc}^{(1)}\right] \in R^{q \times n}. \tag{3}$$

Then, the intraclass correlation matrix and the interclass correlation matrix are, respectively, shown as

$$S_w = \sum_{i=1}^{c} \sum_{k=1}^{n_i} \sum_{l=1}^{n_i} x_k^{(i)} y_l^{(i)T} = XDY^T, \tag{4}$$

where D is a block diagonal matrix, which is also a positive semidefinite matrix. The difference between the interclass correlation matrix and the intraclass correlation matrix is just a negative sign [37]

(3) Solve the eigenvalues and eigenvectors. The optimization problem of DCCA can be transformed into

$$\max w_x^T S_w w_y \text{ s.t.} w_x^T S_{xx} w_x = w_y^T S_{yy} w_y = 1. \tag{5}$$

Use the Lagrangian multiplier method to solve the above optimization problem turning the above problem into a problem of finding characteristic roots and characteristic vectors.

$$S_w S_{yy}^{-1} (S_w)^T w_x = \lambda^2 S_{xx} w_x,$$
$$(S_w)^T S_{xx}^{-1} S_w w_y = \lambda^2 S_{yy} w_y. \tag{6}$$

The eigenvector $\{w_x, w_y\}_1^d$ corresponds to the first d generalized eigenvalues, and the $\lambda_1 \geq \lambda_2 \geq \lambda_d$

(4) For each pair of samples (x, y), fusion is performed according to the tandem method. The block diagram of feature fusion using the DCCA algorithm is shown in Figure 8

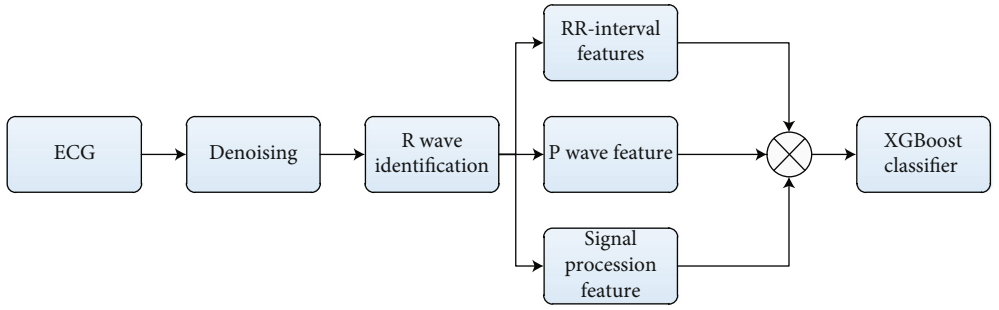


Figure 9: Block diagram of AF by traditional feature experimental pipeline.

4. Performance Metrics

In order to optimize the atrial fibrillation detection model, a large number of experiments are carried out using a single-lead ECG dataset. The experiment in this article is to train on a server equipped with Tesla V100-SXM2 GPU and Ubuntu 16.04 operating system, and its dynamic memory of the computer is 32480MiB.

In this paper, normal F1 score, atrial fibrillation F1 score, other F1 score, and the average value of three categories of F1 score are four metrics for evaluating the classification performance of the experiments. The definition of these four metrics can be defined as

$$F_{1a} = \frac{2 \times A_a}{\sum A + \sum a}, \tag{7}$$

where A is the total number of signals identified as atrial fibrillation by the algorithm, A_a is the number of signals correctly classified as atrial fibrillation by the algorithm, and a is the total number of atrial fibrillation signals.

$$F_{1n} = \frac{2 \times N_n}{\sum N + \sum n}, \tag{8}$$

where N is the total number of normal signals recognized by the algorithm, N_n is the number of correct signals classified as normal by the algorithm, and n is the total number of normal signals.

$$F_{1o} = \frac{2 \times O_o}{\sum O + \sum o}, \tag{9}$$

where O is the total number of signals identified by the algorithm as "other," O_o is the correct number of signals classified by the algorithm as "other," and o is the total number of "other" signals.

$$F_{1p} = \frac{2 \times P_p}{\sum P + \sum p}, \tag{10}$$

where P is the total number of noise signals recognized by the algorithm, P_p is the correct number of noise signals classified by the algorithm, and p is the total number of noise signals.

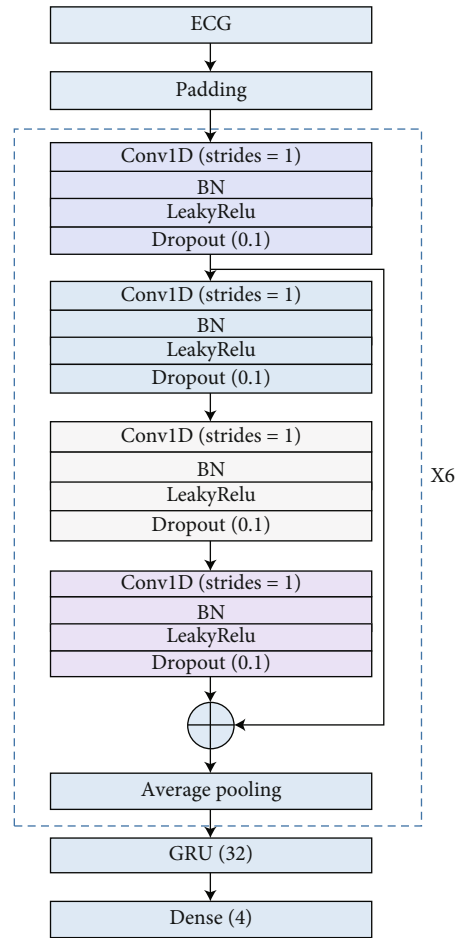


Figure 10: Block diagram of AF by deep learning feature experimental pipeline.

$$F_{\text{overall}} = \frac{(F_{1n} + F_{1a} + F_{1o})}{3}. \tag{11}$$

Because the noise signals are too small and unbalanced, the result of the entire dataset is unstable, and the first three types of signals are selected as the final F1 index. Even so, the F1 score of noise will also affect the other three types. In addition to F1, we also use true positive (TP), true negative (TN), false positive (FP), and false negative (FN) to calculate

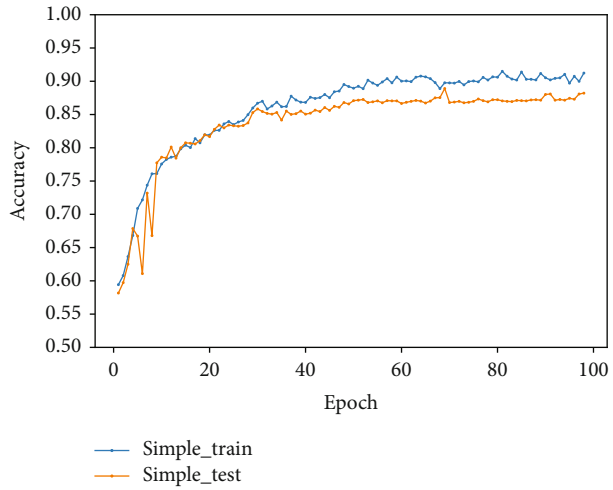


Figure 11: The accuracy diagram of series feature fusion.

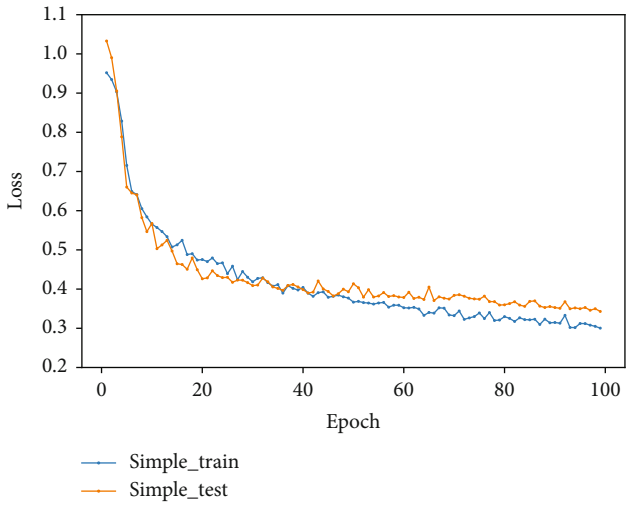


Figure 12: The loss diagram of series feature fusion.

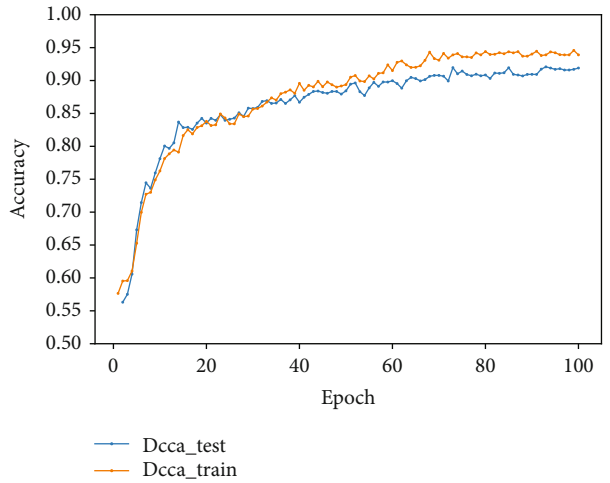


Figure 13: The accuracy diagram of DCCA feature fusion.

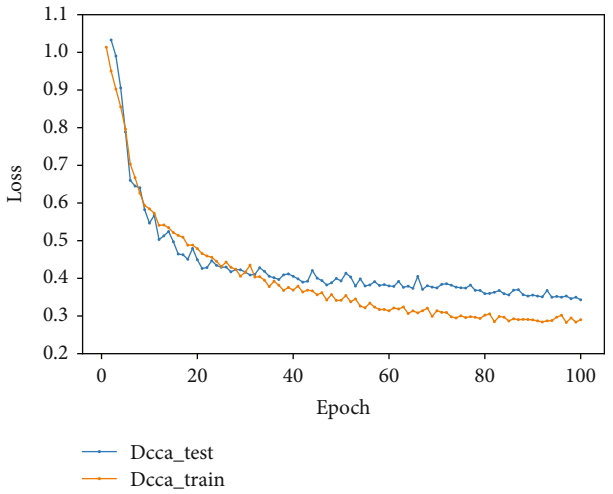


Figure 14: The loss diagram of DCCA feature fusion.

accuracy (Acc), specificity (Spe), and sensitivity (Sen). The calculation formula is as follows:

$$\begin{aligned} \text{Acc} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Spe} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\ \text{Sen} &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned} \tag{12}$$

5. Results

Four experiments are used to verify the feasibility and efficiency of the proposed feature fusion model. The first three experiments are comparative experiments.

5.1. Experiments Based on Single Feature

5.1.1. Experiments Based on Traditional Feature. In this experiment, after the ECG signal is denoised, its statistical features and frequency domain features are extracted manually based on expert knowledge, and finally, the XGBoost (Extreme Gradient Boosting) classifier is used for classification. The experimental block diagram based on traditional feature extraction and classification is shown in Figure 9.

The XGBoost parameters are tuned using random grid search cross-validation, and the optimal parameters are selected. The minimum leaf node weight is set to 20, the maximum depth of the tree is set to 11, the subsample is set to 0.8, the colsample_bytree is set to 0.9, the learning rate is 0.2, and the maximum depth of the tree is 11.

The minimum loss function is reduced to 1, the softmax objective function is used for classification, and the final F1 is 75%.

5.1.2. Experiments Based on Deep Learning Feature. In this experiment, the ECG signal is detected based on the model of residual network and gated recurrent unit. The experimental block diagram of using deep learning feature extraction to classify atrial fibrillation is shown in Figure 10.

Table 2: The result of the different model.

Model	F_{1n}	F_{1a}	F_{1o}	F_{overall}	Acc	Spe	Sen
Expert features	87%	73%	65%	75%	79%	82%	72%
Resnet+GRU	91%	81%	77%	83%	86%	85%	84%
Simple fusion	92%	83%	80%	85%	88%	89%	86%
Proposed	93%	88%	84%	88%	92%	93%	90%

Table 3: Comparison of previous studies of ECG based on the PhysioNet/CinC challenge 2017 public dataset.

Method	F_{1n}	F_{1a}	F_{1o}	F_{overall}	Acc	Spe	Sen
Convolutional recurrent neural network [23]	92.4%	81.4%	80.9%	84.9%	87.5%	94.6%	82.9%
Decision tree ensemble [24]	88.9%	79.1%	70.2%	79.4%	——	——	——
16-layer 1D residual convolutional network [25]	90.0%	82.0%	75.0%	82.0%	80.2%	——	——
2D convolutional network with LSTM layer [26]	88.8%	76.4%	72.6%	79.2%	82.3%	——	——
1DCNN containing residual blocks and recurrent layers [27]	91.9%	85.8%	81.6%	86.4%	——	——	——
Proposed in this paper	93.1%	88.3%	84.0%	88.3%	91.7%	93.2%	90.4%%

Firstly, padding the original ECG data. Since the central electrical data of the database varies from 9 s to 61 s and the convolutional network requires equal length input, the ECG data is padded the same length. This paper uses the maximum length of the ECG signal. The sampling rate is 300 Hz, and the calculated maximum length is 18286. Each ECG data is inputted into the residual network. The residual network includes six residual convolution blocks, and each of them consists of a convolution block, a residual block, and a one-dimensional average pooling layer. Each convolutional block includes four parts: a one-dimensional convolution layer with a step size of 1, a batch normalization layer, a linear unit with leakage correction, and a spatial random inactivation layer. After the residual network, data is inputted to the gated recurrent unit for training. The number of neurons in the gated recurrent unit is 32. Finally, it is output through the fully connected layer. F1 ended up at 83%.

5.2. Experiments Based on Feature Concatenation Fusion. In this experiment, the features are simply spliced and fused and input to the fully connected layer for classification.

The feature vectors based on expert knowledge and the feature vectors extracted by the residual network and gated recurrent unit are spliced in series to obtain the fused features and input to the fully connected layer for classification. The specific process is as follows: firstly, add a flatten layer to make the traditional feature vector one-dimensional; then, use the deep learning model for training, the output of the last hidden layer of the recurrent unit as the deep learning feature vectors; finally, use the concatenation method to integrate the two feature vectors into one, and add a fully connected layer for classification. The value of F1 is 85%, and the accuracy and loss diagrams are shown in Figures 11 and 12.

5.3. Experiments Based on DCCA Feature Fusion. In this experiment, the feature vectors extracted by the traditional feature extraction method based on expert knowledge and the deep learning feature vectors extracted using the gated recurrent unit and residual network are fused with discriminant canonical correlation analysis and then input to the fully connected layer for feature classification. The final accuracy on the verification set is 91.7, and F1 is 88%. The accuracy and loss diagrams are shown in Figures 13 and 14. From Table 2, it can be seen that the DCCA-based fusion method is better than the concatenation fusion method. Compared with simple concatenation fusion, the DCCA method considers the correlation among samples and the category information of the sample, which contains less redundant information than the series fusion method.

As can be seen from Table 2 and Figure 12, that compared to using single feature, the method of feature fusion for AF signal detection can obtain better classification accuracy. Compared with single feature extraction, the F1 score is increased by 2% when using simple feature fusion, and compared with the simple feature fusion method, the F1 score is increased by 3% when using DCCA feature fusion.

5.4. Experimental Comparative Analysis. In order to verify the effectiveness of the proposed method, comparisons are also performed with previous studies. Table 3 lists some of the published ECG signal detection research results based on the same dataset, which includes traditional feature extraction, machine learning based on expert knowledge, and deep learning-based methods. It can be seen from Table 3 that the use of a single method requires complex preprocessing, and the final F1 value is 79.4%, which is not ideal [24]. The signal detection model using the expert knowledge feature extraction algorithm has better interpretability. On the other hand, deep neural networks are used to autonomously learn features from ECG records. The conventional method is very easy to learn. Xiong et al. [25] proposed a 16-layer deep convolutional neural network for the automatic classification of ECG signal, the final F1 is 82.0%, and the accuracy is 80.2%. The feature fusion method based on discriminative canonical correlation analysis proposed in this paper can fuse the advantages of the two and achieve a more ideal result. The F1 value is 88%. The accuracy, sensitivity,

and specificity are 91.7%, 90.4%, and 93.2%, respectively, conducive to more accurate ECG signal detection. It is foreseeable that with the further accumulation of datasets, the feature fusion model can achieve more powerful classification capabilities.

6. Conclusion

This paper proposes a classification method for atrial fibrillation signals based on the feature fusion of discriminant canonical correlation analysis. This method can not only extract the deep learning features of ECG signals but also fuse the traditional features of ECG signal samples. With DCCA, the maximum and minimum correlations among classes of different sample types are considered, and the recognition results are better than that of series feature fusion as well as the use of deep learning or traditional features alone. This method has been verified on the public short single-lead ECG dataset of the 2017 PhysioNet/CinC Challenge, with a verification accuracy of 91.7%, a sensitivity of 90.4%, and a specificity of 93.2%. The database used in this article itself has the problem of large differences among various categories, which shows that the fusion method in this article improves the overall accuracy while taking into account other measurement standards, and steadily improves the classification performance of ECG signals. However, this paper only considers the comprehensive and complementary representation of ECG features through feature-level fusion and does not consider the fusion of decision-making layers, such as neural network algorithms, hidden Markov models, and combinations of multiple classifiers. In future researches, the classification model and feature fusion method will be further improved. On the basis of DCCA feature fusion technology, core-based DCCA will be introduced. At the same time, more cutting-edge classifiers will be selected for classification and recognition, which will be more effective to improve recognition results.

Data Availability

The datasets used during the present study are available from the corresponding author upon reasonable request or can be downloaded from https://www.physionet.org/content/challenge-2017/1.0.0/.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Q.Z. and C.C. performed the conceptualization; J.S. contributed to the methodology; J.S. and C.C. helped in the validation; Q.Z., H.L., and M.S. performed the formal analysis; J.S. did the investigation; Q.Z., H.L., and M.S. helped in finding resources; J.S. wrote and prepared the original draft; C.C. and Q.Z. wrote, reviewed, and edited the manuscript; Q.Z. did the supervision, project administration, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

References

[1] S. A. Shufni and M. Y. Mashor, "ECG signals classification based on discrete wavelet transform, time domain and frequency domain features," in *2015 2nd international conference on biomedical engineering (ICoBE)*, pp. 1–6, Penang, Malaysia, 2015.

[2] M. Romano, P. Bifulco, A. M. Ponsiglione, G. D. Gargiulo, F. Amato, and M. Cesarelli, "Evaluation of floatingline and foetal heart rate variability," *Biomedical Signal Processing and Control*, vol. 39, pp. 185–196, 2018.

[3] A. Lyon, A. Mincholé, J. P. Martínez, P. Laguna, and B. Rodriguez, "Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances," *Journal of the Royal Society Interface*, vol. 15, no. 138, p. 20170821, 2018.

[4] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Systems with Applications: X*, vol. 7, article 100033, 2020.

[5] O. Berenfeld and J. Jalife, "Complex fractionated atrial electrograms: is this the beast to tame in atrial fibrillation?," *Circulation: Arrhythmia and Electrophysiology*, vol. 4, no. 4, pp. 426–428, 2011.

[6] S. A. Guidera and J. S. Steinberg, "The signal-averaged p wave duration: a rapid and noninvasive marker of risk of atrial fibrillation," *Journal of the American College of Cardiology*, vol. 21, no. 7, pp. 1645–1651, 1993.

[7] S. Mehta, N. Lingayat, and S. Sanghvi, "Detection and delineation of p and t waves in 12-lead electrocardiograms," *Expert Systems*, vol. 26, no. 1, pp. 125–143, 2009.

[8] Y. Kaya, H. Pehlivan, and M. E. Tenekeci, *Effective ECG Beat Classification Using Higher Order Statistic Features and Genetic Feature Selection*, Biomedical Research, India, 2017.

[9] M. Athif, P. C. Yasawardene, and C. Daluwatte, "Detecting atrial fibrillation from short single lead ECGs using statistical and morphological features," *Physiological Measurement*, vol. 39, no. 6, article 064002, 2018.

[10] L. Yin, F. Chen, Q. Zhang, and X. Ma, "Arrhythmia classification based on multi-domain feature extraction," *Journal of Physics: Conference Series*, vol. 1237, no. 2, article 022062, 2019.

[11] S. Dash, K. Chon, S. Lu, and E. Raeder, "Automatic real time detection of atrial fibrillation," *Annals of Biomedical Engineering*, vol. 37, no. 9, pp. 1701–1709, 2009.

[12] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier," in *2017 Computing in Cardiology Conference (CinC)*, Rennes, France, 2017.

[13] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

[14] M. Zubair, J. Kim, and C. Yoon, "An automated ECG beat classification system using convolutional neural networks," in *2016 6th International Conference on IT Convergence and Security (ICITCS)*, pp. 1–5, Prague, Czech Republic, 2016.

[15] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, and H. Yang, "Patient-specific ECG classification based on recurrent neural networks and clustering technique," in *IASTED Int Conf Biomedical Engineering*, Innsbruck, Austria, 2017.

[16] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita, and U. R. Acharya, "Automated detection of atrial fibrillation using long short-term memory network with RR interval signals," *Computers in Biology and Medicine*, vol. 102, pp. 327–335, 2018.

[17] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, vol. 115, pp. 465–473, 2019.

[18] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1744–1753, 2018.

[19] P. Warrick and M. N. Homsi, "Cardiac arrhythmia detection from ECG combining convolutional and long short-term memory networks," in *2017 computing in cardiology (Cin C)*, pp. 1–4, Rennes, France, 2017.

[20] D. Smoleń, "Atrial fibrillation detection using boosting and stacking ensemble," in *2017 computing in cardiology (Cin C)*, pp. 1–4, Rennes, France, 2017.

[21] J. Chu, H. Wang, and L. U. Wei, "A novel two-lead arrhythmia classification system based on cnn and lstm," *Journal of Mechanics in Medicine and Biology*, vol. 19, no. 3, article 1950004, 2019.

[22] S. Ghiasi, M. Abdollahpur, N. Madani, K. Kiyani, and A. Ghaffari, "Atrial fibrillation detection using feature based algorithm and deep convolutional neural network," in *2017 computing in cardiology (Cin C)*, pp. 1–4, Rennes, France, 2017.

[23] J. Van Zaen, O. Chételat, M. Lemay, E. Calvo, and R. Delgado-Gonzalo, "Classification of cardiac arrhythmias from single lead ECG with a convolutional recurrent neural network," in *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies*, pp. 33–41, Prague, Czech Republic, 2019.

[24] R. Muhammed, M. Whitaker Bradley, and V. Anderson David, "AF detection from ECG recordings using feature selection, sparse coding, and ensemble learning," *Physiological Measurement*, vol. 39, no. 12, article 124007, 2018.

[25] X. Zhaohan, K. Stiles Martin, and Z. Jichao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," in *2017 Computing in Cardiology (Cin C)*, pp. 1–4, Rennes, France, 2017.

[26] Z. Martin, P. Dmytro, and T. Michael, "Convolutional recurrent neural networks for electrocardiogram classification," in *2017 Computing in Cardiology (Cin C)*, pp. 1–4, Rennes, France, 2017.

[27] Z. Xiong, M. P. Nash, E. Cheng, V. V. Fedorov, M. K. Stiles, and J. Zhao, "ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network," *Physiological Measurement*, vol. 39, no. 9, article 094006, 2018.

[28] A. L. Goldberger, L. A. Amaral, L. Glass et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[29] G. Bianchi and R. Sorrentino, *Electronic Filter Simulation & Design*, McGraw Hill Professional, 2007.

[30] I. I. Christov, "Real time electrocardiogram QRS detection using combined adaptive threshold," *BioMedical Engineering OnLine*, vol. 3, no. 1, p. 28, 2004.

[31] J. M. Bote, J. Recas, F. Rincon, D. Atienza, and R. Hermida, "A modular low-complexity ECG delineation algorithm for real-time embedded systems," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 429–441, 2018.

[32] C. Huang, Y. Jin, Q. Wang, L. Zhao, and C. Zou, "Multi-modal emotion recognition based on speech signal and ECG signal," *Journal of Southeast University (Natural Science Edition)*, vol. 40, no. 5, pp. 895–900, 2010.

[33] L. Fengjuan, *Research on Face Recognition Method Based on Canonical Correlation Analysis*, Nanjing University of Science and Technology, Jiangsu, 2009.

[34] H. P. Martínez and G. N. Yannakakis, "Deep multimodal fusion: combining discrete events and continuous signals," in *Proceedings of the 16th International conference on multimodal interaction*, pp. 34–41, Bogazici University, Istanbul, Turkey, 2014.

[35] Q.-S. Sun and P.-A. Heng, "Face recognition based on generalized canonical correlation analysis," in *Advances in Intelligent Computing*, Springer, Berlin Heidelberg, 2005.

[36] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.

[37] S. Hoch, F. Althoff, G. McGlaun, and G. Rigoll, "Bimodal fusion of emotional data in an automotive environment," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. ii/1085–ii/1088, Philadelphia, PA, USA, 2005.