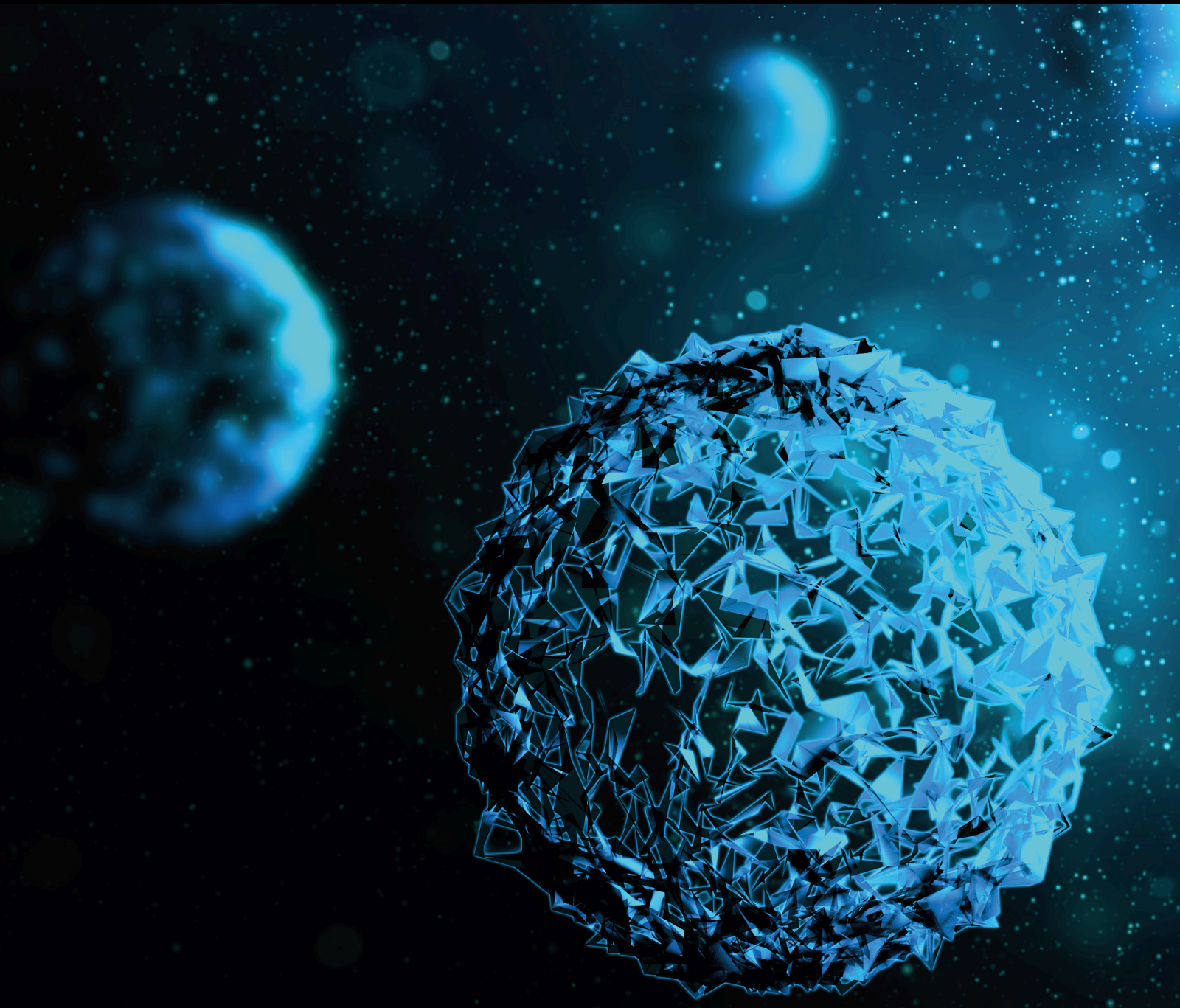


The Emergence of AI Based Health Informatics

Lead Guest Editor: Friso De Boer

Guest Editors: Francis M. Bui, Kawsar Ahmed, and Sami Azam





The Emergence of AI Based Health Informatics

BioMed Research International

The Emergence of AI Based Health Informatics

Lead Guest Editor: Friso De Boer

Guest Editors: Francis M. Bui, Kawsar Ahmed, and Sami Azam



Copyright © 2023 Hindawi Limited. All rights reserved.

This is a special issue published in "BioMed Research International." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Section Editors




Penny A. Asbell, USA
David Bernardo , Spain
Gerald Brandacher, USA
Kim Bridle , Australia
Laura Chronopoulou , Italy
Gerald A. Colvin , USA
Aaron S. Dumont, USA
Pierfrancesco Franco , Italy
Raj P. Kandpal , USA
Fabrizio Montecucco , Italy
Mangesh S. Pednekar , India
Letterio S. Politi , USA
Jinsong Ren , China
William B. Rodgers, USA
Harry W. Schroeder , USA
Andrea Scribante , Italy
Germán Vicente-Rodríguez , Spain
Momiao Xiong , USA
Hui Zhang , China

Academic Editors

Bioinformatics

Contents

An Enhanced Technique of COVID-19 Detection and Classification Using Deep Convolutional Neural Network from Chest X-Ray and CT Images

Md Khairul Islam , Md Mahbubur Rahman , Md Shahin Ali , Md Sipon Miah , and Md Habibur Rahman 

Research Article (18 pages), Article ID 6341259, Volume 2023 (2023)

A Novel Approach for Best Parameters Selection and Feature Engineering to Analyze and Detect Diabetes: Machine Learning Insights

Md Shahin Ali , Md Khairul Islam , A. Arjan Das , D. U. S. Duranta , Mst. Farija Haque , and Md Habibur Rahman 



Research Article (15 pages), Article ID 8583210, Volume 2023 (2023)

HDFCN: A Robust Hybrid Deep Network Based on Feature Concatenation for Cervical Cancer Diagnosis on WSI Pap Smear Slides

Nitin Kumar Chauhan , Krishna Singh , Amit Kumar , and Swapnil Baburav Kolambakar 

Research Article (17 pages), Article ID 4214817, Volume 2023 (2023)

Automated Evaluation of Upper Airway Obstruction Based on Deep Learning

Yunho Jeong, Yeeyeewin Nang , and Zhihe Zhao 

Research Article (7 pages), Article ID 8231425, Volume 2023 (2023)

Research Article

An Enhanced Technique of COVID-19 Detection and Classification Using Deep Convolutional Neural Network from Chest X-Ray and CT Images

Md Khairul Islam ¹, Md Mahbubur Rahman ², Md Shahin Ali ¹, Md Sipon Miah ^{2,3}
and Md Habibur Rahman ⁴

¹Department of Biomedical Engineering, Islamic University, Kushtia 7003, Bangladesh

²Department of Information and Communication Technology, Islamic University, Kushtia 7003, Bangladesh

³Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain

⁴Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh

Correspondence should be addressed to Md Habibur Rahman; habib@iu.ac.bd

Received 22 September 2022; Revised 7 March 2023; Accepted 10 November 2023; Published 11 December 2023

Academic Editor: Nikhil Agrawal

Copyright © 2023 Md Khairul Islam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Coronavirus disease (COVID-19) is an infectious illness that spreads widely over a short period of time and finally causes a pandemic. Unfortunately, the lack of radiologists, improper COVID-19 diagnosing procedures, and insufficient medical supplies have all played roles in these devastating losses of life. Deep learning (DL) could be used to detect and classify COVID-19 for potential image-based diagnosis. **Materials and Methods.** This paper proposes an improved deep convolutional neural network (IDConv-Net) to detect and classify COVID-19 using X-ray and computed tomography (CT) images. Before the training phase, preprocessing methods such as filtering, data normalization, classification variable encoding, and data augmentation were used in conjunction with the proposed IDConv-Net to increase the effectiveness of the detection and classification processes. To extract essential features, deep CNN is then employed. As a result, the suggested model can identify patterns and relationships crucial to the image classification task, resulting in more precise and useful diagnoses. Python and Keras (with TensorFlow as a backend) were used to carry out the experiment. **Results.** The proposed IDConv-Net was tested using chest X-rays and CT images collected from hospitals in Sao Paulo, Brazil, and online databases. After evaluating the model, the proposed IDConv-Net achieved an accuracy of 99.53% and 98.41% in training and testing for CT images and 97.49% and 96.99% in training and testing for X-ray images, respectively. Further, the area under the curve (AUC) value is 0.954 and 0.996 for X-ray and CT images, respectively, indicating the excellent performance of the proposed model. **Conclusion.** The findings of our proposed IDConv-Net model confirm that the model outperformed compared to existing COVID-19 detection and classification models. The IDConv-Net outperforms current state-of-the-art models by 2.25% for X-rays and 2.81% for CT images. Additionally, the IDConv-Net training approach is significantly quicker than the current transfer learning models.

1. Introduction

The worldwide outbreak of the coronavirus disease (COVID-19) is still wreaking havoc on people's lives and health [1]. COVID-19 is a highly infectious disease with limited and less effective treatment options [2]. The transmission of COVID-19 occurs through respiratory droplets released when an infected individual talks, coughs, or

sneezes due to infection with the SARS-CoV-2 virus. The virus can also be spread by contacting the mouth, nose, or eyes after touching a surface or object that has been exposed to the virus [3]. Numerous COVID-19 patients frequently overburden the healthcare systems in many countries. About 347.49 million/5.60 million patients have been diagnosed/died with COVID-19 infection since December 2019. The incidence of illnesses and deaths due to COVID-19 is

increasing each day. According to a report [4] on 12 September 2022, a total of instances of COVID-19 have been reported at 613,958,298 in the world. Among them, 6,516,913 have died, and 592,777,665 have recovered. Furthermore, over 1,075,668 out of 97,095,092 patients have died in America; 684,914 out of 34,574,765 patients have died in Brazil; 528,165 out of 44,500,580 patients have died in India; 29,334 out of 2,014,887 patients have died in Bangladesh, etc. [4].

Generally, late detection of COVID-19 can assault the lungs and harm the tissues of the disease-ridden patient [5]. The lungs and human respiratory system are still particularly susceptible organs where the COVID-19 virus can easily proliferate. Damage results and the air sac is filled with liquid and expelled as an outcome [6, 7]. As a consequence, the patient has trouble breathing with oxygen. So, we want to rapidly and precisely determine the degree of lung injury to survive the patients and reduce fatality [8]. Moreover, early COVID-19 detection can save the patient's life and stop spreading. A significant level of protection should be offered by a parenteral COVID-19 vaccine approach capable of inducing a potent, long-lasting immune response involving neutralizing antibodies and T cells [9, 10]. Different vaccine platforms and strategies have advantages and disadvantages from an immunological perspective. As a result, the COVID-19 vaccine has significantly changed the pandemic's trajectory and reduced the rate of mortality [9, 11].

One of the diagnostic methods used for detecting COVID-19 is real-time reverse transcription polymerase chain reaction (RT-PCR), which is a recommended technique by the WHO for identifying the presence of the virus causing COVID-19 [12]. However, the RT-PCR method takes a few hours to two days to produce test results. Additionally, this technique is difficult, expensive, manual, and unavailable everywhere. The expense and lack of RT-PCR affect many developing and underdeveloped nations [13]. Further, RT-PCR testing needs a laboratory kit; many nations find it difficult to produce or gather during the outbreak [14]. Moreover, the COVID-19 RT-PCR test's reduced sensitivity was noted in several investigations. Many researchers have reported this test's sensitivity to 71% to 98%, which reduces the detection accuracy of COVID-19 cases [15].

Another approach is medical imaging which plays a critical role in COVID-19 detection and management. Specifically, chest X-rays and computed tomography (CT) scans have been used to detect and monitor COVID-19 patients. Medical imaging, such as chest X-rays and CT scans, can be helpful in detecting COVID-19 for several reasons, including the visualization of lung abnormalities, the confirmation of the diagnosis, the severity assessment, and the monitoring of disease progression [16]. Specifically, medical imaging is used by radiologists to verify the COVID-19 diagnosis manually. However, as radiologists must manually diagnose a significant number of COVID-19 patients, it is a laborious, error-prone, and exhausting process that necessitates competent radiologists [17].

Over the years, artificial intelligence (AI) has shown potential in the field of medical imaging. Deep learning (DL) is an effective tool for analyzing medical imaging data because it can automatically identify patterns and features

from large datasets without requiring manual feature engineering [18]. Furthermore, DL has the potential to improve the speed, accuracy, and accessibility of COVID-19 diagnosis, which can help to control the spread of the virus better and improve patient outcomes. There has been a significant amount of research on the use of medical imaging for COVID-19 detection. Several researchers have applied machine learning (ML) and DL methods, such as convolutional neural networks (CNNs), transfer learning (TL), autoencoders, and ensemble, to medical imaging for COVID-19 detection. Further, these ML and DL methods have shown promise in COVID-19 detection using medical imaging [19]. Moreover, deep CNNs have demonstrated potential in COVID-19 detection and classification using medical imaging due to their ability to automatically learn hierarchical features from the input images. Furthermore, CNNs are designed to minimize noise and variation in the input images. Additionally, it allows the network to leverage knowledge learned from a large and diverse dataset, which can improve performance on the target task. However, in some research on COVID-19, lung cancer, monkeypox, brain stroke, etc., detection and classification were performed using CNN with insufficient accuracy [6, 20, 21]. Additionally, in some cases, CNN and transfer learning require a longer training time in the detection and classification.

After considering these issues, we require a compatible deep learning framework that will be able to help consultants and healthcare staff quickly and correctly identify COVID-19 disease from X-ray and CT images [22]. This research is aimed at demonstrating an enhanced deep convolutional neural network-based solution for automatic COVID-19 detection from chest X-rays and CT images. The COVID-19 radiography dataset that is publicly available is limited, and a large dataset is preferred to train deep CNN models. Even after using overfitting mitigation techniques, training DL models on a small dataset can result in overfitting. One of the most critical issues when designing the architecture is limiting the number of trainable parameters to avoid overfitting. An early call-back function can be employed to avoid overfitting. Further, data augmentation can also be used to address problems with small datasets. While developing deep learning models, overcoming the vanishing gradient problem is crucial. Additionally, the problem of accuracy degradation during deeper network training needs to be addressed.

This paper's primary contributions include the following:

- (i) We propose an improved IDConv-Net model that can effectively and precisely differentiate between patients with normal chest conditions and those with COVID-19 by utilizing both chest X-rays and CT images
- (ii) Moreover, the novelty of the IDConv-Net model is that it can detect and classify disease from different modalities of images with higher accuracy. The model training approach time is very quick due to the use of fewer layers
- (iii) Furthermore, to evaluate the performance of IDConv-Net, we run the model using X-ray and CT images

from which the best classification rate is achieved when compared to other existing detection models. In addition, our model provides a lower miss detection which indicates that the model is more reliable in detecting COVID-19 even though the dataset is entirely unknown

- (iv) Lastly, the experimental findings support the notion that the suggested IDConv-Net model outperformed previous state-of-the-art approaches to COVID-19 detection and classification. Additionally, the proposed IDConv-Net model correctly detects COVID-19 in both datasets, such as chest X-rays and CT images

The subsequent sections of this paper are organized as follows: Section 2 provides a literature review of the study, while Section 3 describes the method used and the required materials in detail. Our proposed IDConv-Net model is elaborated on in Section 4, while Section 5 presents the potential training parameters employed in our model. Section 6 of this paper presents the study's findings, while Section 7 provides a discussion of these results. The paper concludes with a summary in Section 8.

2. Related Works

To stop the COVID-19 pandemic from spreading, it is essential to identify the virus quickly and precisely. Chest X-ray and CT images are available in almost all hospitals worldwide and are the most widely used and economically advantageous medical imaging technology for evaluating lung problems [22, 23]. Chest X-ray and CT scans can reliably identify lung injury in COVID-19 patients earlier [9, 24]. It identifies the virus's stage while indicating its presence [16]. However, the lack of distinctive characteristics and the resemblances between lung lesions and other viral diseases make COVID-19 susceptible to misdiagnosis [25]. Considering these, potential AI appliances such as ML and DL can overcome the COVID-19 disease detection errors caused by people from X-ray and CT imaging techniques [26–28]. AI has proven its efficiency and performance in detecting diseases like cancer, tumor, pneumonia, and COVID-19. DL-based approach, such as CNN, plays a key role in processing medical images, particularly in features extracting and classifying [29]. In [30], Bassi and Attux developed a dense CNN to classify COVID-19, pneumonia, and normal from chest X-rays. They proposed a novel approach of output neurons that modifies the twice-transfer learning techniques. They achieved a good performance from their model regarding the classification of COVID-19. However, larger datasets and clinical investigations were required to guarantee accurate generalization. Agrawal and Choudhary [20] suggested a deep CNN for detecting COVID-19, utilizing two datasets of chest X-rays. For image segmentation, they used an encoder-decoder architecture. The CNN structure encoder extracts features and transfers them to the decoder as part of the segmentation procedure for this experiment. The findings demonstrated that, for two datasets used in COVID-19 detection,

the suggested model achieved high accuracy of 94.4% and 95.2%, respectively.

The authors [31] introduced a novel COVID-CXNet utilizing a familiar transfer learning-based CheXNet model. They utilized relevant and meaningful features in the detection of novel coronavirus. A CNN-long short-time memory (CNN-LSTM) model was designed by Purohit et al. [32] to extract features from raw data hierarchically. They employed several COVID-19 chest X-ray datasets to test and investigate the model accomplishment of COVID-19 detection. For the larger dataset, however, the model needs to be trained for longer, which needs to be cut down.

Ayalew et al. [17] presented the DCCNet model for the diagnosis of COVID-19 patients. The authors employed two methods, namely, histograms of oriented gradients (HOG) and CNN, to extract features from input images, and used a support vector machine (SVM) classifier to classify COVID-19. The SVM classifier yielded 99.97% accuracy during training and 99.67% accuracy during testing when combined with CNN and HOG-based features.

Indumathi et al. [33] mentioned an ML algorithm to classify and predict COVID-19-affected zone. From March to July 2020, they used the Virudhunagar district's COVID-19 dataset. They achieved a 98.06% accuracy rate, which was higher than the 95.22% accuracy rate of the C5.0 algorithm.

Salau [34] used an SVM technique to classify and identify COVID-19 using chest CT data. After extracting features from CT scans using a discrete wavelet transform (DWT) technique, the study built a classification model. The findings demonstrated that the suggested model has a high accuracy of 98.2% in COVID-19 detection.

Chaunzwa et al. [35] used a DL framework to detect lung cancer from CT images. In [36], the identification of COVID-19 on CT scans is accomplished using ML methods. However, their investigation only used images that cost 150 for CT scans. Khan et al. [37] highlight promising DL research for understanding radiography pictures and progressing the investigation of constructing specific DL-based assessment methods for unique COVID-19 variations, delta, omicron, and challenges ahead. In [6, 21], to identify COVID-19, the authors implemented an SVM technique. Using 208 test data, they achieved a lower recognition rate. In [38], the authors applied ML techniques to detect COVID-19 automatically using X-ray images to enhance accuracy. Rahimzadeh and Attar [39] considered Xception and ResNet50V2 approaches for COVID-19 identification from X-ray images. In [40], the researchers employed pre-trained transfer learning models, such as ResNetV2, InceptionV3, and ResNet50, for detecting lung disease and COVID-19 using X-ray images. In [40], COVID-19 was identified using just X-ray data by CNN models like Inception-ResNetV2, ResNet50, and InceptionV3, where the models had 98%, 97%, and 87% classification accuracy, respectively. These experiments used a few X-ray data. They may have tested their models' performance on other modalities, such as CT scans.

However, some problems with past research include insufficient detection accuracy for different modalities of images, small datasets, overfitting issues, and using CNN

without first preprocessing images. Further, some works require prolonged training time, which is another drawback. This study used a number of image preprocessing approaches to address these drawbacks. Furthermore, the proposed enhanced model improves detection and classification performance and reduces the training approach time.

3. Materials and Methods

This section elaborates on the methodology provided for identifying COVID-19. Figure 1 depicts the process of the proposed methodology.

3.1. Image Data Acquisition. A dataset is the backbone of research. We used two types of images, 2D CT and X-ray images. For CT images, the settings for the 64-slice scanner were calibrated with the following parameters: collimation of either 128×0.6 mm or 64×0.6 mm, tube voltage of 120 kilovolts (kV), section thickness of five millimeters (mm), slice interval of five millimeters (mm), the pitch of 1.375 mm, and field of view of three hundred fifty by three hundred fifty (mm). In addition, the patient's position was supine; both arms were elevated, and the patient was instructed to hold their breath. The datasets were reconstructed with a wall thickness and increment ranging between 1.5 and 2 mm [41].

We collected 1,252 COVID-19 positive and 1,230 normal images from the SARS-CoV-2 CT-scan dataset which images were gathered from real patients in hospitals from Sao Paulo, Brazil.

Our COVID-19 CT scan image dataset consisted of 7,593 COVID-19 images obtained from 466 patients, as well as 6,893 normal images obtained from 604 patients. Then, merge CT images of both datasets to create a new 2D-CT dataset containing 8,845 COVID-19 and 6,893 normal images, totaling 17,168 images. Similarly, we collected 576 COVID-19 positive and 1,583 normal X-ray images from COVID-19 X-ray dataset and 4273 COVID-19 positive images and 10192 normal X-ray images from the COVID-19 Radiography Database and then merged them to create an enlarged size of the new X-ray dataset. We used a new merged X-ray and 2D-CT image dataset to perform our model better. Overall, the merged X-ray dataset contains 4,192 positive COVID-19 and 11,775 normal images, and the merged 2D-CT dataset consists of 8,845 COVID-19 positive and 8,123 normal CT images. We represent some sample CT and X-ray images in Figures 2 and 3. In Table 1, we highlighted the number of images extracted from the sources dataset. The datasets were partitioned into training, validation, and testing sets, as shown in Table 2. For each dataset, 80% of the images were allocated for training, 10% for validation, and 10% for testing purposes.

3.2. Data Preprocessing. Preprocessing is crucial in transforming raw data into a format appropriate for the ML or DL approach. It primarily enhances the source images by controlling normalization, multicollinearity, scaling, shuffle, and data division [42]. Furthermore, preprocessing methods enhance the image quality, making an experiment more suc-

cessful. Moreover, it is very difficult to handle high dimensions of input data. Sometimes, it may cause overfitting and poor results. For this reason, we downsized the images to 224×224 . We applied the dimensionality reduction technique for less computational time and quick visualization. Before training the model, it is essential to convert string or nonnumeric features into numeric ones. So, we utilized data transformation for data compatibility, which means converting string or non-numeric features into numeric [43–45]. We also use feature engineering, which entails selecting the features that would be helpful in training a model. The normalizing technique was utilized to compare various features on a comparable scale. As a result, we can use higher learning rates or models to converge more quickly for a given learning rate. It also helps to stabilize the gradient descent step.

3.3. Normalization of Data. The significance of data normalization for developing exact prescient models has been analyzed for the different ML algorithms that have recreated a crucial position [46]. The fundamental objective of data normalization is data quality assurance before its application to predictive analytics. Various data normalization techniques can be utilized, including min-max normalization, Z-score normalization, decimal scaling, and median standardization, among others [47, 48]. The prime aims of data normalization are given below:

- (i) This data group makes all entries and attributes appear identical
- (ii) It provides the dataset with relevant information that is more obvious and natural, reducing its size and simplifying its structure so that it is easier to identify, contrast, and retrieve
- (iii) It enhances and simplifies the numerical data without losing the critical characteristics with reduced complexity, leading to easy segmentation

The dataset can be normalized by dividing an image's gray-scale value by 255. However, this study uses Z-score normalization as a normalized technique [49], which is stated as follows:

$$\bar{Z}_k = \frac{Z_k - \bar{S}}{\text{std}(S)}, \quad (1)$$

where \bar{Z}_k indicates the normalized weights of Z-score, Z_k is the weight of S^{th} row and k^{th} column, \bar{S} represents the mean, and $\text{std}(S)$ represents the standard deviation, which can be expressed as

$$\text{std}(S) = \sqrt{\frac{1}{n} \sum_{k=1}^n (Z_k - \bar{S})^2}. \quad (2)$$

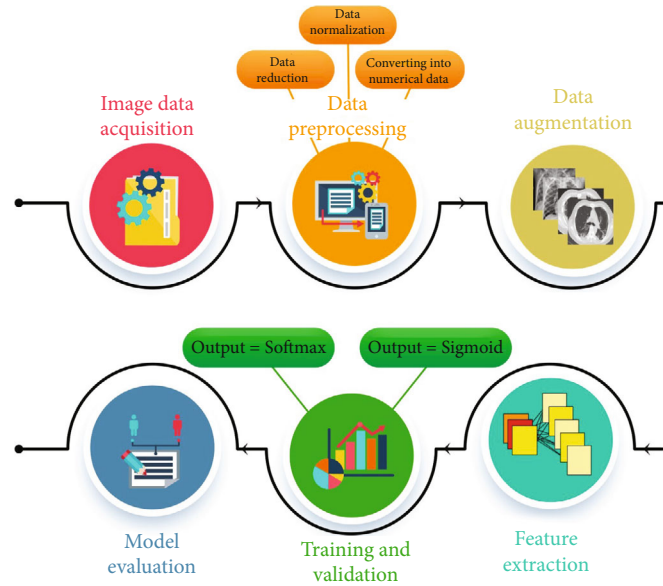


FIGURE 1: Working flow diagram of the proposed methodology.

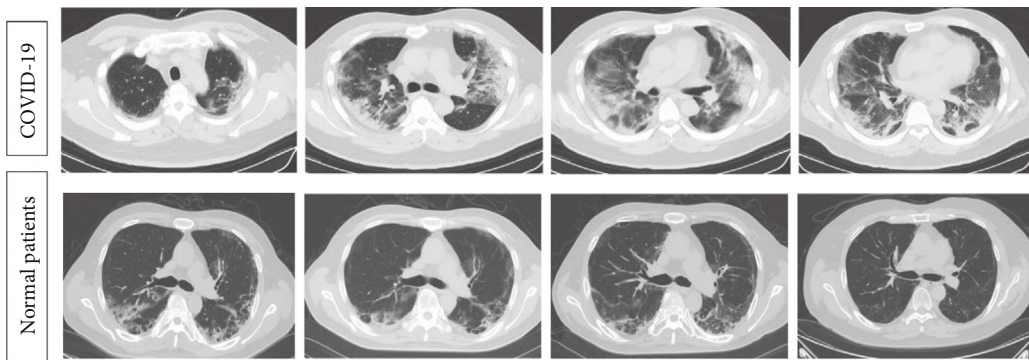


FIGURE 2: CT scan images of COVID-19 and normal patients.

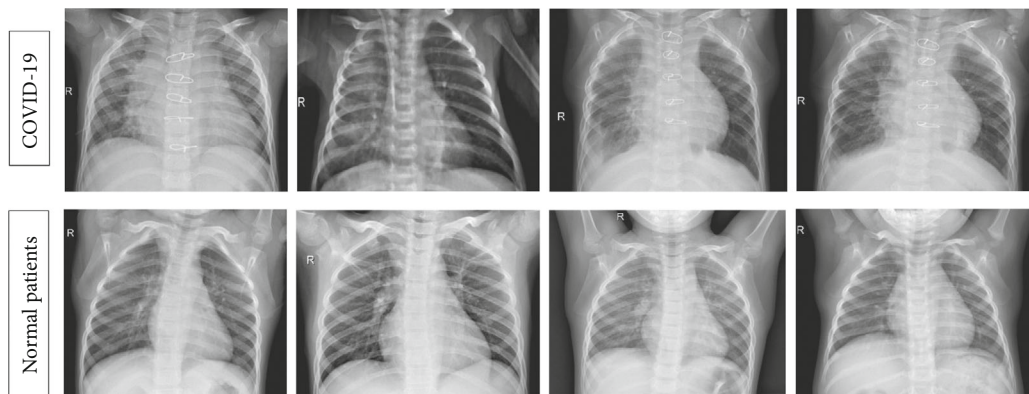


FIGURE 3: X-ray images of COVID-19 and normal patients.

Herein,

$$\bar{S} = \frac{1}{n} \sum_{k=1}^n Z_k. \tag{3}$$

When manipulating data, the values are typically scaled into the [0–1] range, ensuring that the data is stored.

3.4. Working with Numerical Data. The most frequent data types are dealt with in DL methods through numerical

TABLE 1: Details of X-ray and CT datasets before and after merging images.

| Type | Dataset | COVID-19 images | Normal images | Total images |
|-------|--------------------------------------|-----------------|---------------|--------------|
| X-ray | COVID-19 X-ray dataset | 576 | 1,583 | 2,159 |
| | COVID-19 Radiography Database | 3,616 | 10,192 | 13,808 |
| | Merged X-ray dataset | 4,192 | 11,775 | 15,967 |
| CT | SARS-COV-2 CT-scan dataset | 1,252 | 1,230 | 2,482 |
| | Large COVID-19 CT scan slice dataset | 7,593 | 6,893 | 14,686 |
| | Merged CT dataset | 8,845 | 8,123 | 17,168 |

TABLE 2: The datasets used in this study.

| Dataset | Label | Train set | Validation set | Test set | Total |
|---------|----------|-----------|----------------|----------|--------|
| X-ray | COVID-19 | 3354 | 419 | 419 | 15,967 |
| | Normal | 9,419 | 1,178 | 1,178 | |
| CT | COVID-19 | 7,075 | 885 | 885 | 17,167 |
| | Normal | 6499 | 812 | 812 | |

values. For this reason, we had to maintain some procedures to get numerical values for each extracted feature. Furthermore, we apply normalized and standardized operations to get better processing to train models and support various DL networks. In our experiment, we converted the two levels of COVID-19 and non-COVID-19 to 0 and 1 using the LabelEncoder function from the Python standard module [50].

3.5. Data Augmentation. Data augmentation is a powerful and useful technique for improving machine learning models' accuracy and predictive ability by increasing the number of images in a dataset through modified versions of existing training images. Moreover, it reduces the complexity of collecting more images to enlarge the dataset. Data augmentation utilizes techniques such as data wrapping and oversampling to increase the number of images in a dataset. Nevertheless, it may appear to be an overfitting problem in the results [51]. To mitigate this problem, we have applied flipping, rotation shearing, mirroring, zooming, fill mode, and channel shifting using principal component analysis to augment the data [52]. The augmentation parameters that we used to increase the number of images are given below:

Flipping: the image is horizontally and vertically flipped. The flipping operation reconfigures the pixels while preserving the image's attributes. An image's vertical and horizontal position is randomly adjusted by 0.2 degrees.

Rotation: the image is flipped by a number of degrees between 0 and 360. In the model, every flipped image will be different. The rotation range is from -360 to 360 degrees.

Shear: to produce or correct perception angles, the image can be twisted along a particular axis using a shear range of approximately 0.4 degrees.

Zoom: the zoom range of an image in the data augmentation method can be zoomed in or out. This method enlarges the image by zooming in or out randomly and adding pixels around the image. The extent of zoom is around 0.5 degrees.

Fill mode: to fill the empty pixel's values, the default value "nearest" is applied, which replaces the nearest image pixels.

Channel shifting: it randomly shifts channel values to vary the hue by 10.

4. Our Proposed IDConv-Net Model

This section represents the proposed model and working outline. The proposed IDConv-Net model has five convolutional layers and four pooling layers (max pooling), batch normalization, rectified linear unit (ReLU), dense and dropout layers, and sigmoid. The architecture of the IDConv-Net model is shown in Figure 4, which consists of input images, feature extraction, and classification layers. Firstly, the feature extraction layer extracts the critical features from the input images; then, the last part of the model, such as the fully connected layer, performs classification. As a result, the model functions as a feature extractor before acting as a classifier.

4.1. Feature Extraction. Feature extraction is an important step in ML and DL applications, as it can improve the efficiency, accuracy, and interpretability of the subsequent learning algorithms [53]. The feature extraction part of our model consists of five convolutional layers followed by four max-pooling layers through the ReLU activation layer (see Figure 5), while subsequent ReLU layers follow through the batch normalization layer and maximum pooling layer, and finally, the last convolution layer, followed by the flattening and dropout layer, as shown in Figure 4.

Apart from this, the input layer initially receives input images with the size of $224 \times 224 \times 3$ CT or X-ray chest image, where 224×224 is the image's dimension, and 3 is the RGB channel. The convolutional layer is responsible for featuring maps, i.e., feature representation, of the input images [54]. The input image x is convoluted with a set of trainable weights, sometimes referred to as multidimensional filters f_k , and the result is coupled with biases b_k .

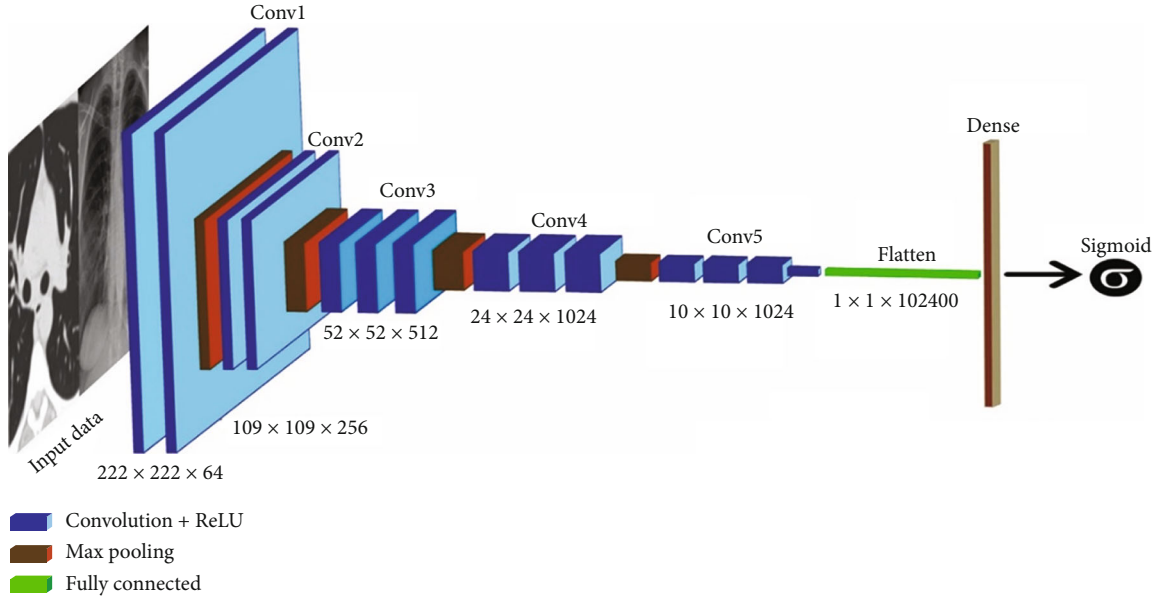


FIGURE 4: The architecture of our recommended IDConv-Net model.

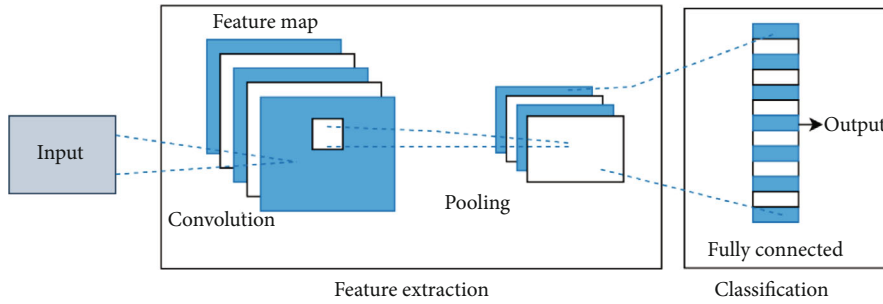


FIGURE 5: The workflow of the feature extractor and classifier.

Assume there are K filters; this layer's k th output can be represented as given in the following equation:

$$V_{p'',q'',n} = \sum_c f_n^c \times x^c(r^{p'',q''}) + b_n, \quad (4)$$

$$= \sum_{p'=1}^M \sum_{q'=1}^N \sum_{c=1}^C f_n^c \times x^c(r^{p'',q''}) + b_n, \quad (5)$$

where $x \in R^{M \times N \times C}$, $f \in R^{M' \times N' \times C}$, $x \in R^{M'' \times N'' \times C}$ and M , N , C are the height, width, and channel of the input, respectively. Further, $N'' = N - N' + 2P/S + 1$, $M'' = M - M' + 2P$ $x^{c(r^{p'',q''})}$ represents the local region of p^{th} row and q^{th} column of input image, where P is the zero padding number and S is the pixels of stride. Further, the batch normalization (BN) layer improves network training and lessens sensitivity to network initialization between the convolutional and activation unit (ReLU) layers [55]. In this paper, the ReLU activation function is applied, mathematically shown in Equation (6), and it only keeps the positive part of the activation.

$$a_{(p,q,n)} = f(z_{(p,q,n)}) = \text{MAX}(z_{(p,q,n)}, 0). \quad (6)$$

Furthermore, the polling layer takes maximum values with a pool size of (2, 2). Consequently, the max-pooling layer pooled the feature maps with the dimension $111 \times 111 \times 64$, followed by the second convolutional layer. Similarly, the second convolutional layer convolved feature maps, followed by a second pooling layer with a similar filter size of 2×2 and a stride of 2. Consequently, the image's dimension will be reduced to $54 \times 54 \times 256$. Table 3 displays the proposed IDConv-Net model, which outlines its constituent layers, including their corresponding output sizes. The IDConv-Net model is comprised of five convolution layers, four activation layers, and four max-pooling layers. The resulting output features are then passed through a flatten layer, a dense layer, a dropout layer, and a sigmoid activation layer.

4.2. Classification. The classification layer is the final layer of the proposed model that produces the network's output in the form of predefined categories or classes. The classification layer follows the feature extraction layer that extracts the high-level features from the image. The output of the feature extraction layer is sent to a flattened layer as the first

TABLE 3: The proposed IDConv-Net model with its layers and output size.

| Layer No | Name | Layer Type | Output size |
|----------|-------|-------------|-----------------------------|
| 1 | Input | — | $224 \times 224 \times 3$ |
| 2 | C1 | conv2d | $222 \times 222 \times 64$ |
| 3 | A1 | Activation | $222 \times 222 \times 64$ |
| 4 | M1 | max_pooling | $111 \times 111 \times 64$ |
| 5 | C2 | conv2d | $109 \times 109 \times 256$ |
| 6 | A2 | Activation | $109 \times 109 \times 256$ |
| 7 | M2 | max_pooling | $54 \times 54 \times 256$ |
| 8 | C3 | conv2d | $52 \times 52 \times 512$ |
| 9 | A3 | Activation | $52 \times 52 \times 512$ |
| 10 | M3 | max_pooling | $26 \times 26 \times 512$ |
| 11 | C4 | conv2d | $24 \times 24 \times 1024$ |
| 12 | A4 | Activation | $24 \times 24 \times 1024$ |
| 13 | M4 | max_pooling | $12 \times 12 \times 1024$ |
| 14 | C5 | conv2d | $10 \times 10 \times 1024$ |
| 15 | F1 | Flatten | $1 \times 1 \times 102400$ |
| 16 | D1 | Dense | $1 \times 1 \times 1024$ |
| 17 | Dr1 | Dropout | $1 \times 1 \times 1024$ |
| 18 | D2 | Dense | $1 \times 1 \times 1$ |
| 19 | A5 | Activation | $1 \times 1 \times 1$ |

step in classification, which converts the data's form into a one-dimensional data vector. In the classification function, a dropout layer is followed by a thick layer with 1024 neurons. A dense layer produces the final output with two neurons and a sigmoid activation function, which identifies the image as belonging to one of the chest diseases: COVID-19 or the normal.

The classification layer in a CNN using the sigmoid activation function can be represented using the following equation:

$$y = \sigma(Wx + b), \quad (7)$$

where x is the output of the previous layer, W is the weight matrix of the classification layer, b is the bias vector, and σ is the sigmoid activation function defined as

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (8)$$

The output of the final layer of the proposed model is passed through the sigmoid function to obtain a value between 0 and 1, which can be interpreted as the probability that the input image belongs to the positive class. The decision boundary can be set to 0.5, so if the output of the sigmoid function is greater than 0.5, the input image is classified as belonging to the positive class. If it is less than or equal to 0.5, the input image is classified as belonging to the negative class. Moreover, a dropout layer is utilized with

a value of 0.3 for the last convolution layer to avoid overfitting between the training and testing performance.

5. Training and Performance

The performance of the training and testing set depends on the experiment setup and performance matrices such as precision, Recall, F -score accuracy, sensitivity, and specificity. Experiment setup and performance matrices were described in this section.

5.1. Experiment Setup. Hyperparameter tuning is an important step in building machine learning models, as it involves selecting the optimal hyperparameters that result in the best performance of the model. To get excellent performance from the model, we repeatedly fine-tuned the model. We used three optimized hyperparameters during our study's training phase: batch size, epochs, and learning rate. Manually tuning that parameter is time-consuming; therefore, we have applied the grid search method to select the best value of the hyperparameter. Table 4 summarizes the initial and optimal parameters found during the experiment. From Table 4, we can infer that the best-optimized batch size is 32, epochs are 50, and the learning rate is 0.001 for both datasets. We performed a grid search method using ML frameworks and libraries, such as scikit-learn in Python to obtain these values.

Adam, also known as adaptive momentum, is used to enhance the performance of our suggested IDConv-Net model because it performs consistently while categorizing binary images [56]. The experiment was conducted using an organization laptop with Windows 10, a Core i7 processor, and 16 GB of RAM. Furthermore, we run the model on a Jupiter laptop and the Google Colab GPU environment with 12 GB of RAM.

The proposed model was developed and fine-tuned using chest X-ray and CT image dataset to get insight into the COVID-19 identification issues. We split our dataset into three sections: training, validation, and testing to evaluate the performance of the IDConv-Net model. To assess the performance, we have used 80% data for the model training up, 10% data for model validation, and the rest 10% for model testing. Table 2 shows the data distribution for the training, validation, and testing sets, respectively, for a better understanding of both datasets.

5.2. Performance Metrics. Performance measures are crucial to assessing the proposed approach. In this study, we measured precision, recall, $F1$ -score, and accuracy using four metrics: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

- (i) *TP.* The experimental result for the COVID-19 patients is accurate. That means the model detects positive results for COVID-19-affected patients
- (ii) *TN.* The experimental result for the Non-COVID-19 patients is accurate. That means the model detects negative results for Non-COVID-19 affected patients

Input: COVID-19 or Non-COVID-19 X-ray or CT images datasets D with resize image dimension (S).
 Featured Vector using IDConv-Net = (F_v)
 1: Initialize $F_v > S_p, p=1$
 2: characteristics extracted from each image $D(p, 1, 570)$.
 3: $F_v(p, 1) = S(x, 1) + F_v(p, 1)$.
 4: F_v = Total characteristics extracted by IDConv-Net.
 5: Assign. H_o = Output of the hidden layer, H_d = Last hidden layer outcome.
 6: $V_t(p, 1) = H_o(p, 1) + H_d(p, 1)$.
 7: F_t = Output of the IDConv-Net through a hidden and FC layer.
 8: Training_feature ($T_{train_feature}$) = [F_v, F_t].
 9: Test_image = $imread(image)$.
 10: Move to: step 1 and 2 to extract essential test features ($T_{test_feature}$) from test set.
 11: Outcome (i) = classify ($T_{train_feature}, T_{test_feature}$).
 12: Output: True for COVID-19 Positive or False for COVID-19 Negative.

ALGORITHM 1: Suggested algorithm for the classification of COVID-19.

TABLE 4: Parameter settings used during this study.

| Parameters | Study X-ray | | Study CT | |
|---------------|-------------------|---------------------|-------------------|---------------------|
| | Initial parameter | Optimize parameters | Initial parameter | Optimize parameters |
| Batch size | 10, 20, 32, 64 | 32 | 10, 20, 32, 64 | 32 |
| Epochs | 30, 40, 50, 60 | 50 | 25, 30, 50, 70 | 50 |
| Learning rate | .001, .01, 0.1 | .001 | .001, .01, 0.1 | .001 |

TABLE 5: Performance comparisons between the IDConv-Net and state-of-the-art models on X-ray images.

| Model | Dataset | Precision | Recall | F1-score | Training accuracy | Testing accuracy |
|---------------------------|------------------------|-----------|--------|----------|-------------------|------------------|
| AlexNet [57] | Lungs X-ray | 69.25 | 90.43 | 81.83 | 69.84 | 67.76 |
| nCOVnet [58] | Chest X-ray | 82.00 | 97.62 | 89.13 | 97.00 | 88.10 |
| Deep CNN [59] | covid-chestxray | 99.17 | 71.76 | 83.27 | 72.78 | 71.90 |
| InceptionResnetV2 [60] | covid-chestxray | 92.11 | 92.38 | 92.07 | 93.83 | 92.18 |
| MobileNetV2 [61] | covid-chestxray | 20.00 | 100 | 33.33 | 62.12 | 60.00 |
| ResNetV2 [61] | covid-chestxray | 40.00 | 100 | 57.14 | 71.89 | 70.00 |
| VGG-16 [62] | COVID-19 and pneumonia | 86.17 | 86.23 | 86.38 | 87.36 | 86.39 |
| AlexNet [40] | Covid_Data | 84.62 | 94.57 | 89.32 | 92.93 | 82.62 |
| DenseNet201 [61] | covid-chestxray | 100 | 83.33 | 90.91 | 92.23 | 90.00 |
| DenseNet121 [63] | Radiography Database | 89.47 | 100 | 94.44 | — | 94.74 |
| Proposed IDConv-Net model | Merged X-ray | 97.14 | 91.87 | 94.43 | 97.49 | 96.99 |

(iii) *FP*. The experimental result is wrong for the Non-COVID-19 patients. That means the model detects positive results for Non-COVID-19 affected patients

(iv) *FN*. The experimental result is wrong for the COVID-19 patients. That means the model detects negative results for COVID-19-affected patients

Each of these performance metrics is used to assess the performance described in the equation below:

Precision (P): it is comprehended as a positive predictive value. It measures the proportion of positive and projected instances out of the total number of cases that are expected to be positive.

TABLE 6: Confusion matrix of the proposed IDCNN model on X-ray images.

| | | Actual value | |
|-----------------|----------|--------------|--------|
| | | COVID-19 | Normal |
| Predicted value | COVID-19 | 407 | 12 |
| | Normal | 36 | 1142 |

$$P = \frac{TP}{TP + FP} \tag{9}$$

Recall (R): it calculates the percentage of individuals who are truly positive out of the total number of individuals who

TABLE 7: Performance comparisons between the IDConv-Net and state-of-the-art models on CT images.

| Model | Dataset | Precision | Recall | F1-score | Training accuracy | Testing accuracy |
|--------------------------------|-------------------------------|-----------|--------|----------|-------------------|------------------|
| Noisy-OR Bayesian [64] | Transverse-section CT | 86.87 | 86.67 | 86.70 | 87.05 | 86.70 |
| Decision fusion [65] | COVID-CT | 88.14 | 88.79 | 86.70 | 89.78 | 88.34 |
| Ensemble [66] | CT (HRCT) | 90.63 | 93.55 | 92.06 | 92.83 | 91.94 |
| DenseNet [67] | HRCT images | 96.00 | 97.00 | 93.00 | 91.36 | 92.00 |
| DenseNet161 [68] | COVID19Net | 85.39 | 77.55 | 81.28 | 84.36 | 82.76 |
| Ensemble (Hard voting) [68] | COVID19Net | 82.80 | 78.57 | 80.63 | 82.88 | 81.77 |
| Ensemble (Soft voting) [68] | COVID19Net | 83.70 | 78.57 | 81.05 | 84.21 | 82.27 |
| VGG16-based DL [69] | Hospital of Tabriz Data | 91.50 | 89.78 | 90.63 | 91.68 | 90.14 |
| 3D-ResNets with attention [63] | Several cooperative hospitals | 86.27 | 92.33 | 85.20 | — | 93.30 |
| U-NET [70] | CT segmentation dataset | 96.38 | 96.04 | 86.10 | 95.93 | 95.60 |
| Proposed IDConv-Net model | Merged 2D-CT | 98.64 | 98.31 | 98.48 | 99.53 | 98.41 |

are either positive or expected to be positive. It is commonly referred to as the actual positive rate.

$$R = \frac{TP}{TP + FN}. \quad (10)$$

F1-score (F1): precision and recall are used to determine *F1-score*. It is significant to the experiment because of indicating the test accuracy.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (11)$$

Accuracy (A): the ratio is the number of right prediction cases separated by the total number of cases.

$$A = \frac{TP + TN}{TP + TN + FP + FN}. \quad (12)$$

The value of all performance metrics ranges from 0 to 1.

6. Results

The study used secondary datasets of two modalities: X-rays and CT images. The experiment utilized 15967 X-ray images, where 14370 images were used for model training and validation, and the remaining 1597 images were used to evaluate the model. Similar to the first experiment, 15471 CT images were used for model training and validation, and the remaining 1697 images were used to evaluate the model. In this experiment, 17168 CT images were used.

Firstly, we ran the experiment five times to optimize the hyperparameters, including node size, batch size, learning rate, and drop rate. The optimized learning rate was 0.0001 and 0.001 for the X-ray and CT studies, respectively, and 0.99 momentum while training the model with the Adam optimizer using the binary cross-entropy loss. In our research, we have used 50 epochs. However, it was completed in 48 epochs for the CT images and 47 epochs for the X-ray images due to the early stopping function, which is accountable for terminating the execution when reaching an optimum result. Moreover, the complete trainable parameters in model train-

TABLE 8: Confusion matrix of the proposed IDCNN model on CT images.

| | Actual value | |
|-----------------|--------------|--------|
| | COVID-19 | Normal |
| Predicted value | COVID-19 | 873 |
| | Normal | 15 |
| | | 12 |
| | | 797 |

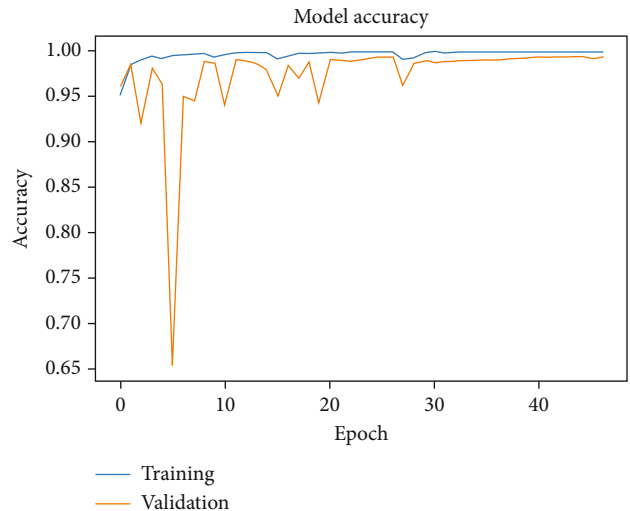


FIGURE 6: Accuracy curve of the proposed IDConv-Net model for X-ray images.

ing were 8,004,481 out of 8,088,129. Furthermore, the sigmoid activation function is used in the final layer since our model works as a binary classifier.

We utilized X-ray and CT images separately to evaluate the performance of the proposed IDConv-Net model. To evaluate the performance of IDConv-Net, firstly, we train the model with X-ray images. We used 80% of the data for model training and 10% for validation. The remaining 10% was used to evaluate the model's performance. After evaluating the IDConv-Net model with X-ray images, we achieved an accuracy of 97.49% and 96.99% for training and testing,

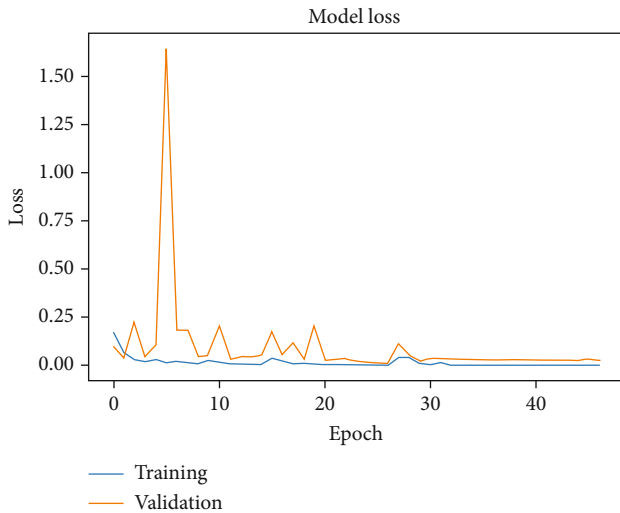


FIGURE 7: Loss curve of the proposed IDConv-Net model for X-ray images.

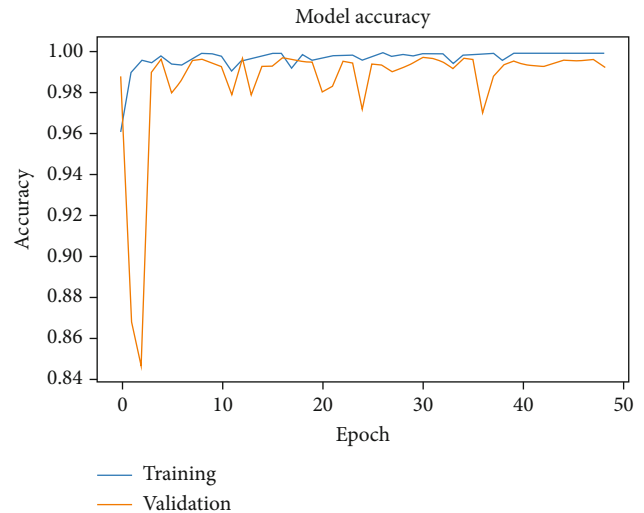


FIGURE 8: Accuracy curve of the proposed IDConv-Net model for CT images.

respectively (see Table 5). Furthermore, the model achieved a precision of 97.14%, recall of 91.87%, and *F1*-score of 94.43% from the X-ray image dataset. From the confusion matrix of the X-ray image dataset (see Table 6), only 12 out of 419 COVID-19 images are misidentified. Furthermore, only 36 out of 1178 normal images are miss identified.

In a different study, we used a CT scan image dataset to train the IDConv-Net model. In this study, 80% of the data were utilized for model training, and 10% were used for model validation. The performance of the model was assessed using the final 10%. We achieved an accuracy of 99.53% and 98.41% for training and testing, respectively, after evaluating the IDConv-Net model on CT images. The findings of IDConv-Net are compared with other state-of-the-art methods shown in Table 7, where the suggested model achieved a precision of 98.64%, recall of 96.31%, *F1*-score 98.48%, training accuracy of 99.53%, and testing accuracy 98.41%. From the confusion matrix of the CT image dataset (see Table 8), only 12 out of 885 COVID-19 images are miss detected, whereas 15 out of 812 normal images are miss detected. The accuracy and confusion matrix proves the model classification reliability even with an entirely new data set.

Finally, we can infer that our proposed model can accurately classify COVID-19 and normal patients from X-ray and CT image datasets. The proposed model obtained a better accuracy with a bit of loss, which is shown in Figures 6 and 7. In another study, the model outperformed the existing models with a little loss on CT images, as shown in Figures 8 and 9.

Moreover, the area under the curve (AUC) summarizes the receiver operating characteristics (ROC) curve demonstrating the classifier's ability to distinguish between classes. The horizontal axis (*X*-axis) represents the false positive rate (FTR), and the vertical axis (*Y*-axis) represents the true positive rate (TPR). The AUC-ROC value is an indicator of the detection performance of the model, with a higher value indicating better performance. The AUC-ROC 0.954 and

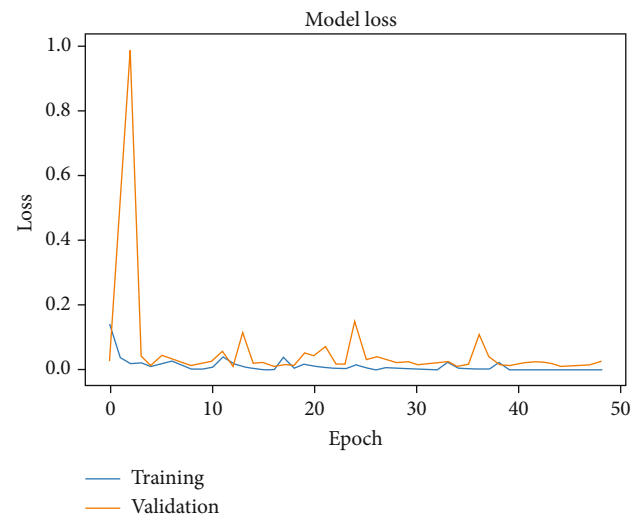


FIGURE 9: Loss curve of the proposed IDConv-Net model for CT images.

0.966 have been achieved simultaneously from our proposed model using X-ray and CT image datasets shown in Figures 10 and 11, respectively. The results of our study indicate that training time for a deep learning model is an important consideration for detecting and classifying COVID-19. Based on the data presented in Table 9, it can be observed that our proposed IDConv-Net model exhibited significantly reduced training times compared to other transfer learning models. Specifically, the training time for the X-ray image dataset was only 31 ± 1 minutes, while the training time for the CT image dataset was 34 ± 1 minutes. These training times were substantially lower than those observed in existing models, which took double as long. Therefore, our IDConv-Net model can be considered a highly efficient and effective approach for image classification tasks. We have also demonstrated the random prediction outcomes of test images using

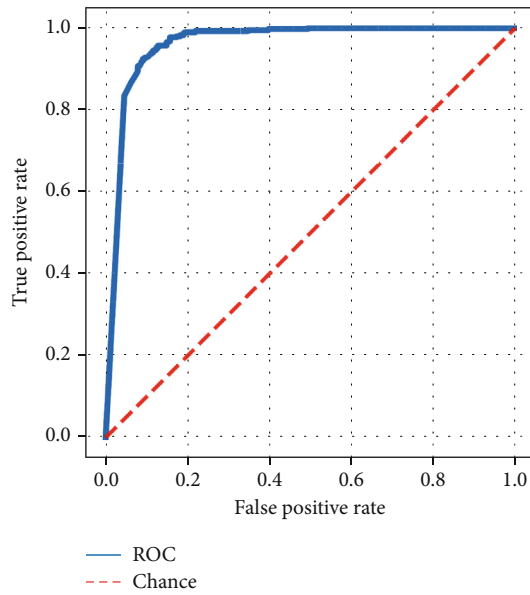


FIGURE 10: AUC-ROC of the proposed IDConv-Net model for X-ray images.

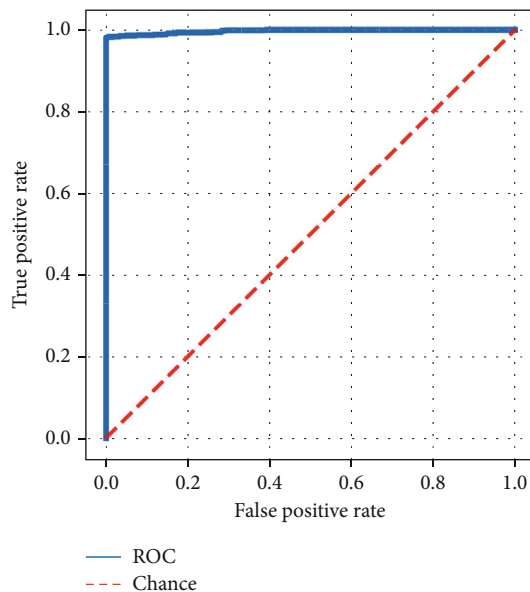


FIGURE 11: AUC-ROC of the proposed IDConv-Net model for CT images.

our suggested IDConv-Net in Figures 12 and 13. In this direction, we evaluated the identification accuracy by comparing the actual and predicted test images with a confidence level.

7. Discussion

In this study, 2D-CT and X-ray images were used. The 2D approach is slice-based, using a single slice image as input to produce a score for each individual. As opposed to this, 3D is a volume-based technique that uses the entire volume (a sequence of slices) as its input to produce a single patient score. However, 2D is still trustworthy for inspecting impor-

tant areas of images and for complex geometry. Moreover, we applied preliminary filtration to all chest images on the train set to control quality and remove incomprehensible slices for processing the chest CT images. Before being approved to train in the IDConv-Net model, two expert physicians graded the diagnosis for the images. A third expert evaluated the evaluation set to ensure no grading errors. Furthermore, we applied the generalization technique to enhance the model performance. The grid-search approach was utilized to identify the optimum hyperparameters. We have chosen the local minima by defining a set of discrete values. The objective function is then evaluated at its grid point by inputting the appropriate parameter values following that. Subsequently, the local minimum can be identified as the lowest objective function value grid point. The minimum and maximum values for each prior were also employed; these parameters were determined empirically according to the characteristics of the images and the number of instances. The parameter settings used in this study are shown in Table 4 with the initial and optimized parameters.

After training, the model can process new data and estimate accurate predictions. In addition, we also used some techniques like data feature extraction preprocessing to get an accurate classification.

We can see the results of the comparison of the proposed model with the state-of-the-art model in Table 5, where we obtained such excellent results due to applying some image preprocessing techniques like noise removal, filtering, data transformation, and feature extraction. Furthermore, our model is quicker because we use fewer layers than the state-of-the-art model. Figure 6 represents the training and validation accuracy of the proposed model during the training. Figure 7 shows the training and validation loss during the model training. From these curves, we can infer that there is no overfitting and exhibits a good model performance. Although we have used 50 epochs, the model terminates execution after 47 epochs for X-ray and 48 epochs for CT images due to the early stopping function.

Similarly, for the CT scan study, Table 7 highlights that our proposed model achieved an excellent performance compared to state-of-the-art models. Moreover, Figure 8 shows the accuracy for the training and validation sets during the model build-up. Similarly, Figure 9 indicates the loss of train and validation sets during the model train. After evaluating the model, we got excellent accuracy and loss curves which indicates the model's good performance.

The study's most significant part was increasing the accuracy level of detection and classification. It is also possible that the goal is to obtain accuracy as close to 100% as possible because, even in a few cases, misdiagnosis is not worth it.

Although similar models of CNN (e.g., AlexNet, nCOV-net, MobileNetV2, and ResNetV2) could detect COVID-19 with insufficient accuracy, moreover, more hidden layers of these models consume more time to yield results. In addition, these models increase the complexity of providing the detection results. Our proposed IDConv-Net model has great significance as a binary classifier. Firstly, it works as a feature extractor, then as a classifier. Moreover, the model

TABLE 9: Time comparison between our proposed IDConv-Net and state-of-the-art models (50 epochs).

| Model | Time of first epoch (seconds) | | Time per remaining epochs (seconds) | | Total time (minutes) | |
|---------------------|-------------------------------|-----|-------------------------------------|---------|----------------------|---------|
| | X-ray | CT | X-ray | CT | X-ray | CT |
| DenseNet201 | 129 | 138 | 103 ± 2 | 113 ± 2 | 86 ± 1 | 95 ± 1 |
| VGG-19 | 296 | 319 | 209 ± 2 | 219 ± 2 | 174 ± 1 | 183 ± 1 |
| ResNet50 | 89 | 103 | 78 ± 2 | 86 ± 2 | 65 ± 1 | 72 ± 1 |
| VGG-16 | 278 | 296 | 193 ± 2 | 207 ± 2 | 161 ± 1 | 173 ± 1 |
| AlexNet | 167 | 187 | 113 ± 2 | 119 ± 2 | 95 ± 1 | 99 ± 1 |
| DenseNet121 | 91 | 106 | 82 ± 2 | 93 ± 2 | 69 ± 1 | 78 ± 1 |
| MobileNetV2 | 79 | 88 | 68 ± 2 | 79 ± 2 | 57 ± 1 | 66 ± 1 |
| InceptionV3 | 132 | 158 | 86 ± 2 | 98 ± 2 | 72 ± 1 | 82 ± 1 |
| Proposed IDConv-Net | 84 | 96 | 37 ± 2 | 44 ± 2 | 31 ± 1 | 34 ± 1 |

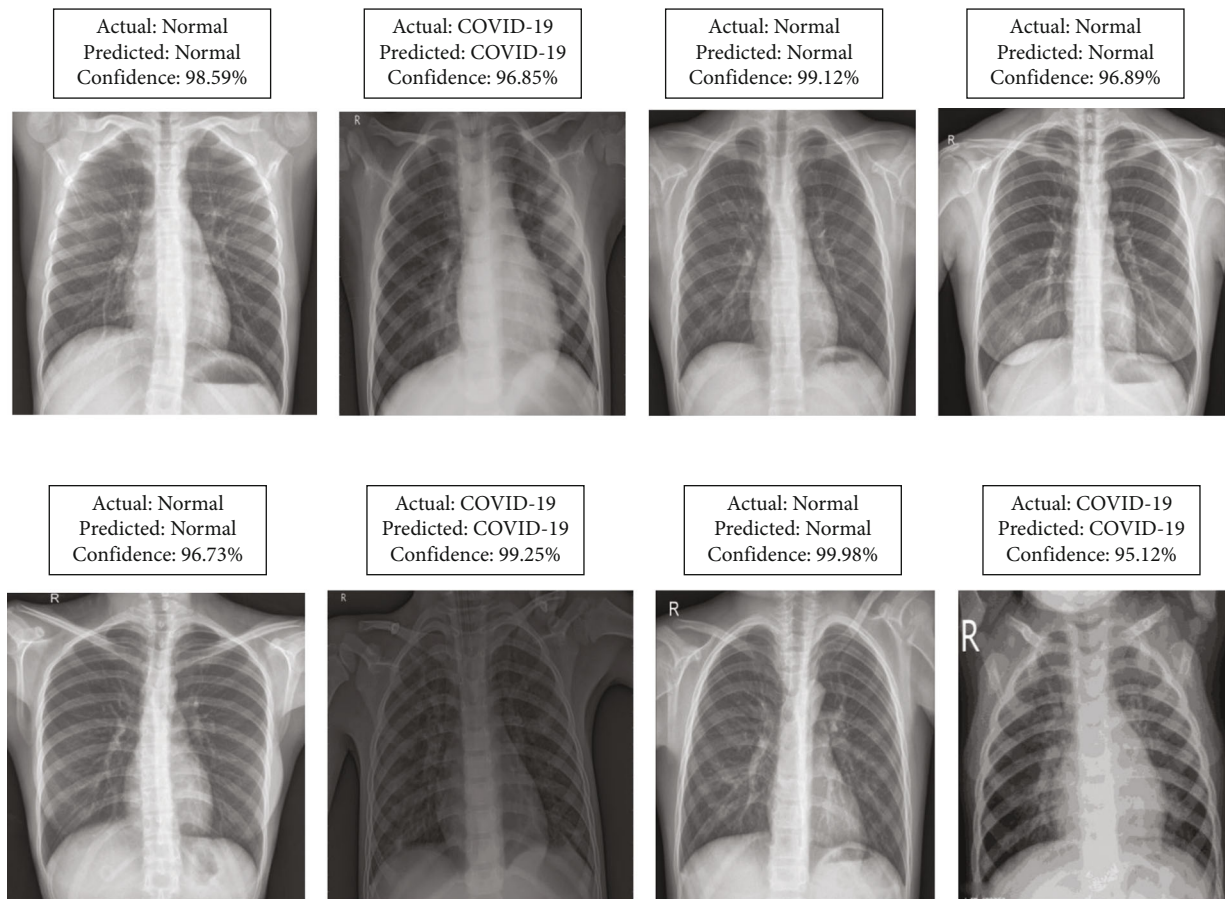


FIGURE 12: Random prediction outcomes using proposed IDConv-Net model from X-ray test images.

is more flexible, less complex, and consumes less time due to fewer hidden layers.

In medical imaging, evaluating ML and DL models used for COVID-19 detection commonly employs the AUC-ROC performance metric, in addition to accuracy. The ROC curve is a graphical representation of the sensitivity and specificity of a binary classifier, and the AUC-ROC quantifies the model's ability to distinguish between positive and negative

classes in a binary classification task. Figures 10 and 11 represent the AUC-ROC of the proposed model for X-ray and CT images. The proposed model obtained 0.954 and 0.966 AUC-ROC for X-ray and CT images, respectively. A high AUC-ROC value indicates that the model has good discrimination between COVID-19 positive and negative cases, with high sensitivity and specificity. However, its use should be combined with other performance metrics and clinical

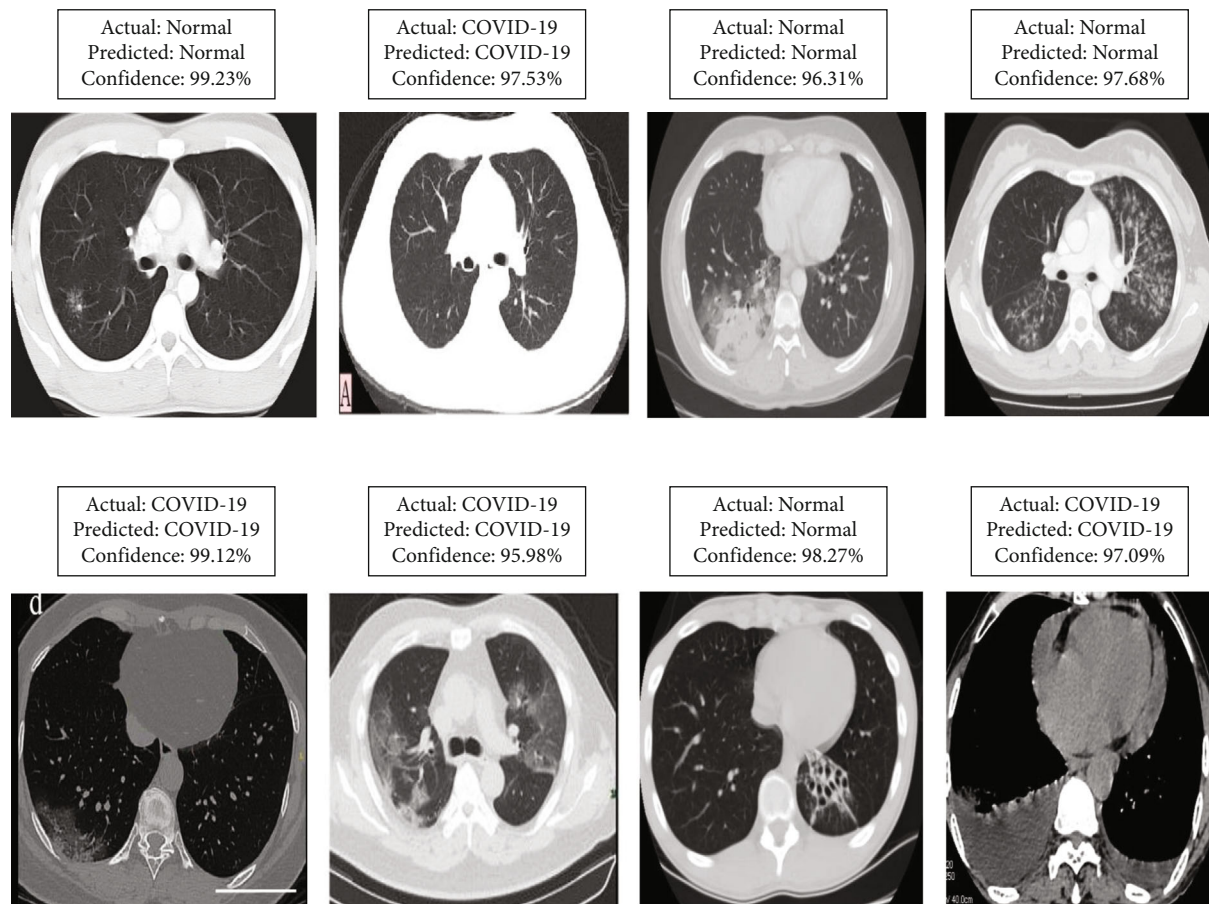


FIGURE 13: Random prediction outcomes using proposed IDConv-Net model from CT test images.

validation to ensure that models are effective and safe for clinical practice.

Overall, the proposed IDConv-Net provides effective results individually on the X-ray and CT images. Finally, according to Tables 5 and 7, our suggested IDConv-Net model achieved the best accuracy for the X-ray and CT image datasets, respectively. Moreover, to avoid overfitting, we used dropout with a value of 0.3 in the last convolution layer of the proposed model. Furthermore, we used an early stopping function during the training of our proposed model to ensure that the model is not overfitted. Thus, the model is good and reliable for detecting COVID-19 in an unknown dataset. We also performed a qualitative analysis where the proposed IDConv-Net achieved a high prediction outcomes rate with a confidence level ranging from 95 to 99+ on the testing set, indicating that it can accurately classify COVID-19 using both X-ray and CT images. Figures 12 and 13 illustrate the actual and predicted outcomes with a confidence level of identification for X-ray and CT images, respectively. These results suggest that a deep CNN model can be an effective tool for COVID-19 diagnosis and potentially assist healthcare professionals in detecting and treating the virus. Moreover, the results of our study demonstrate that our proposed model can detect and classify COVID-19 in a relatively short time frame. As shown in Table 9, our proposed model achieved comparable outcomes to a transfer

learning model while requiring less training time across different image modalities. The reduced training time of our proposed model can be attributed to several factors, including the use of fewer layers in the model architecture and the implementation of enhanced preprocessing techniques. By using a more streamlined model architecture, we were able to reduce the computational demands of the training process while still achieving high levels of performance. Therefore, we can infer that the proposed model works appropriately for both datasets and acquires better accuracy than state-of-the-art detection and classification models. Additionally, DL models' predictions could have been understood and interpreted with the use of a collection of tools and frameworks called explainable AI (XAI). Furthermore, XAI develops a set of ML techniques that produce more understandable models while preserving high performance (prediction accuracy) and enabling human users to comprehend, properly trust, and manage the new breed of AI partners. Another solution to prevent COVID-19 is wearing a face mask and practicing regular hand washing. These are two important measures that effectively reduce the spread of COVID-19. Low-cost sensor-based hand washing techniques can contribute to reducing the spread of COVID-19. However, these measures are most effective when combined with other prevention strategies, such as social distancing and avoiding large gatherings [71, 72].

The advantage of the study is the proposed model consists of fewer layers than other detection models. As a result, it reduces complexity and training time due to a lower layer than other models. Another advantage of the model is that it can detect and classify both data types with higher accuracy. The most vital advantage of the model is that it does not contain overfitting in both datasets' training and testing results. In addition, the following advantages of the model can increase its accuracy if we use balanced datasets. In contrast, the drawbacks of the study are that the model yields less accuracy for X-ray images than CT images due to poor resolution and bony structure of chest scan. However, it can be overcome using high-resolution X-ray images. Another drawback of the model might reduce accuracy if we use imbalanced datasets. The other drawback is that some slices among hundreds of pieces do not contain disease features. These slices are taken from the chest scan's superior/upper, middle, or inferior/lower part. As a result, the model sometimes provides a minor misclassification for COVID-19.

8. Conclusion and Recommendation

COVID-19 poses a severe threat to all living things in the world. A new variant of COVID-19 (e.g., Omicron) will be dangerous and deadly if it mutates with delta or another lethal variant and then spreads quickly worldwide. As a result, early detection of COVID-19 can protect against its spread by isolating affected people. For this purpose, our proposed IDConv-Net can compensate by detecting and classifying COVID-19 at an early stage. Our proposed IDConv-Net model achieves a training accuracy of 99.53% and a testing accuracy of 98.41% for CT images. On the other hand, the IDConv-Net model also achieves a training accuracy of 97.49% and a testing accuracy of 96.99% for X-ray images. Furthermore, our suggested IDConv-Net model outperforms previous COVID-19 detection and classification models that are currently available. Additionally, our proposed model requires less training time than existing models to detect and classify COVID-19.

Overall, while the proposed model has shown great promise in medical imaging applications, several challenges still need to be addressed to make them more effective and practical for use in real-world settings. The model is considered black-box, meaning it can be difficult to understand how they make their predictions. In the future, we plan to use Grad-CAM and XAI to make the model more comprehensive and user-friendly for disease diagnosis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

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

Authors' Contributions

Md Khairul Islam was responsible for the conceptualization, analysis, data curation, methodology, software, validation, writing—original draft preparation, reviewing, and editing. Md Mahbubur Rahman was responsible for the supervision and writing—reviewing and editing. Md Shahin Ali was responsible for the data curation, software, and writing—reviewing and editing. Md Sipon Miah was responsible for the supervision and writing—reviewing and editing. Md Habibur Rahman was responsible for the supervision and writing—reviewing and editing.

Acknowledgments

We would like to acknowledge the support provided by the Bio-Imaging Research Lab, Department of Biomedical Engineering, Islamic University, Kushtia 7003, Bangladesh, in carrying out our research successfully.

References

- [1] L. Wang, Z. Q. Lin, 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, article 19549, 2020.
- [2] S. Liang, H. Liu, Y. Gu et al., "Fast automated detection of COVID-19 from medical images using convolutional neural networks," *Communications Biology*, vol. 4, no. 1, p. 35, 2021.
- [3] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos, Solitons & Fractals*, vol. 134, article 109761, 2020.
- [4] Worldometer, *COVID Live-Coronavirus Statistics*, Worldmeter, 2022.
- [5] M. Ahsan, M. A. Based, J. Haider, and M. Kowalski, "COVID-19 detection from chest X-ray images using feature fusion and deep learning," *Sensors*, vol. 21, no. 4, p. 1480, 2021.
- [6] M. M. Hossain, M. Asadullah, A. Rahaman et al., "Prediction on domestic violence in Bangladesh during the COVID-19 outbreak using machine learning methods," *Applied System Innovation*, vol. 4, no. 4, p. 77, 2021.
- [7] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases," *Computers in Biology and Medicine*, vol. 132, article 104348, 2021.
- [8] M. Ghaderzadeh and F. Asadi, "Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review," *Journal of Healthcare Engineering*, vol. 2021, Article ID 6677314, 10 pages, 2021.
- [9] D. Gursoy, A. S. Can, N. Williams, and Y. Ekinici, "Evolving impacts of COVID-19 vaccination intentions on travel intentions," *The Service Industries Journal*, vol. 41, no. 11-12, pp. 719-733, 2021.
- [10] M. Jeyanathan, S. Afkhami, F. Smaill, M. S. Miller, B. D. Lichty, and Z. Xing, "Immunological considerations for COVID-19 vaccine strategies," *Nature Reviews Immunology*, vol. 20, no. 10, pp. 615-632, 2020.
- [11] O. J. Watson, G. Barnsley, J. Toor, A. B. Hogan, P. Winskill, and A. C. Ghani, "Global impact of the first year of COVID-19 vaccination: a mathematical modelling study," *The Lancet Infectious Diseases*, vol. 22, no. 9, pp. 1293-1302, 2022.

- [12] V. Shah, R. Keniya, A. Shridharani, M. Punjabi, J. Shah, and N. Mehendale, "Diagnosis of COVID-19 using CT scan images and deep learning techniques," *Emergency Radiology*, vol. 28, no. 3, pp. 497–505, 2021.
- [13] O. Vandenberg, D. Martiny, O. Rochas, A. van Belkum, and Z. Kozlakidis, "Considerations for diagnostic COVID-19 tests," *Nature Reviews Microbiology*, vol. 19, no. 3, pp. 171–183, 2021.
- [14] L. J. Carter, L. V. Garner, J. W. Smoot et al., *Assay techniques and test development for COVID-19 diagnosis*, ACS Publications, 2020.
- [15] J. Watson, P. F. Whiting, and J. E. Brush, "Interpreting a COVID-19 test result," *BMJ*, vol. 369, 2020.
- [16] D. Kollias, A. Arsenos, and S. Kollias, *Ai-mia: COVID-19 detection and severity analysis through medical imaging*, Springer, 2023.
- [17] A. M. Ayalew, A. O. Salau, B. T. Abeje, and B. Enyew, "Detection and classification of COVID-19 disease from X-ray images using convolutional neural networks and histogram of oriented gradients," *Biomedical Signal Processing and Control*, vol. 74, article 103530, 2022.
- [18] W. Zhao, W. Jiang, and X. Qiu, "Deep learning for COVID-19 detection based on CT images," *Scientific Reports*, vol. 11, no. 1, article 14353, 2021.
- [19] S. Singh, P. Sapra, A. Garg, and D. K. Vishwakarma, "CNN based COVID-aid: COVID 19 detection using chest X-ray," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1791–1797, Erode, India, 2021.
- [20] T. Agrawal and P. Choudhary, "FocusCovid: automated COVID-19 detection using deep learning with chest X-ray images," *Evolving Systems*, vol. 13, no. 4, pp. 519–533, 2022.
- [21] M. Singh, S. Bansal, S. Ahuja, R. K. Dubey, B. K. Panigrahi, and N. Dey, "Transfer learning-based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data," *Medical & Biological Engineering & Computing*, vol. 59, no. 4, pp. 825–839, 2021.
- [22] M. S. Alam, M. M. Rahman, M. A. Hossain et al., "Automatic human brain tumor detection in MRI image using template-based K means and improved fuzzy C means clustering algorithm," *Big Data and Cognitive Computing*, vol. 3, no. 2, p. 27, 2019.
- [23] A. Sakdinawat, *X-rays and extreme ultraviolet radiation-principles and applications*, Cambridge University Press, 2017.
- [24] P. C. Barman, M. S. Miah, B. C. Singh, and M. T. Khatun, "MRI image segmentation using level set method and implement an medical diagnosis system," *Computer science & Engineering: An international journal (CSEIJ)*, vol. 1, no. 5, pp. 1–10, 2011.
- [25] Y. Li and L. Xia, "Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management," *American Journal of Roentgenology*, vol. 214, no. 6, pp. 1280–1286, 2020.
- [26] M. K. Islam, M. S. Ali, M. M. Ali et al., "Melanoma skin lesions classification using deep convolutional neural network with transfer learning," in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, pp. 48–53, Riyadh, Saudi Arabia, 2021.
- [27] M. K. Islam, M. S. Ali, A. A. Das, D. Duranta, and M. Alam, "Human brain tumor detection using k-means segmentation and improved support vector machine," *International Journal of Scientific Engineering Research*, vol. 11, no. 6, p. 6, 2020.
- [28] L. Li, L. Qin, Z. Xu et al., "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, 2020.
- [29] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, article 100036, 2021.
- [30] P. R. Bassi and R. Attux, "A deep convolutional neural network for COVID-19 detection using chest X-rays," *Research on Biomedical Engineering*, vol. 38, pp. 139–148, 2022.
- [31] A. Haghaniifar, M. M. Majdabadi, Y. Choi, S. Deivalakshmi, and S. Ko, "COVID-CXNet: detecting COVID-19 in frontal chest X-ray images using deep learning," *Multimedia Tools and Applications*, vol. 81, no. 21, pp. 30615–30645, 2022.
- [32] K. Purohit, A. Kesarwani, D. Ranjan Kisku, and M. Dalui, "COVID-19 Detection on Chest X-Ray and CT Scan Images Using Multi-image Augmented Deep Learning Model," in *Proceedings of the Seventh International Conference on Mathematics and Computing. Advances in Intelligent Systems and Computing*, D. Giri, K. K. Raymond Choo, S. Ponnusamy, W. Meng, S. Akleyek, and S. Prasad Maity, Eds., vol. 1412, Springer, Singapore, 2022.
- [33] N. Indumathi, M. Shanmuga Eswari, A. O. Salau, R. Ramalakshmi, and R. Revathy, "Prediction of COVID-19 Outbreak with Current Substantiation Using Machine Learning Algorithms," in *Intelligent Interactive Multimedia Systems for e-Healthcare Applications*, A. K. Tyagi, A. Abraham, and A. Kaklauskas, Eds., Springer, Singapore, 2022.
- [34] A. O. Salau, "Detection of corona virus disease using a novel machine learning approach," in *2021 International Conference on Decision Aid Sciences and Application (DASA)*, pp. 587–590, Sakheer, Bahrain, 2021.
- [35] T. L. Chaunzwa, A. Hosny, Y. Xu et al., "Deep learning classification of lung cancer histology using CT images," *Scientific Reports*, vol. 11, no. 1, p. 5471, 2021.
- [36] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (COVID-19) classification using ct images by machine learning methods," 2020, <http://arxiv.org/abs/2003.09424>.
- [37] A. Khan, S. H. Khan, M. Saif, A. Batool, A. Sohail, and M. Waleed Khan, "A survey of deep learning techniques for the analysis of COVID-19 and their usability for detecting omicron," *Journal of Experimental & Theoretical Artificial Intelligence*, pp. 1–43, 2023.
- [38] J. Rasheed, A. A. Hameed, C. Djeddi, A. Jamil, and F. al-Turjman, "A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 13, no. 1, pp. 103–117, 2021.
- [39] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics in Medicine Unlocked*, vol. 19, article 100360, 2020.
- [40] 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.
- [41] W. Ning, S. Lei, J. Yang et al., *iCTCF: An Integrative Resource of Chest Computed Tomography Images and Clinical Features*

- of Patients with COVID-19 Pneumonia, Research Square, 2020.
- [42] R. Garg, S. Maheshwari, and A. Shukla, "Decision support system for detection and classification of skin cancer using CNN," 2019, <http://arxiv.org/abs/1912.03798>.
- [43] M. K. Islam, M. S. Ali, M. S. Miah, M. M. Rahman, M. S. Alam, and M. A. Hossain, "Brain tumor detection in MR image using superpixels, principal component analysis and template based K-means clustering algorithm," *Machine Learning with Applications*, vol. 5, article 100044, 2021.
- [44] M. M. Hossain, M. S. Ali, M. M. Ahmed et al., "Cardiovascular disease identification using a hybrid CNN-LSTM model with explainable AI," *Informatics in Medicine Unlocked*, vol. 42, article 101370, 2023.
- [45] O. Singh, A. K. Singh, A. K. Agrawal, and H. Zhou, "SecDH: security of COVID-19 images based on data hiding with PCA," *Computer Communications*, vol. 191, pp. 368–377, 2022.
- [46] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, article 105524, 2020.
- [47] M. S. Ali, M. K. Islam, J. Haque, A. A. Das, D. S. Duranta, and M. A. Islam, "Alzheimer's disease detection using m-random forest algorithm with optimum features extraction," in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, pp. 1–6, Riyadh, Saudi Arabia, 2021.
- [48] S. Nayak, B. B. Misra, and H. S. Behera, "Impact of data normalization on stock index forecasting," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 6, no. 2014, pp. 257–269, 2014.
- [49] S. Patro and K. K. Sahu, "Normalization: a preprocessing stage," 2015, <http://arxiv.org/abs/1503.06462>.
- [50] M. I. Hasan, M. S. Ali, M. H. Rahman, and M. K. Islam, "Automated detection and characterization of colon cancer with deep convolutional neural networks," *Journal of Healthcare Engineering*, vol. 2022, Article ID 5269913, 12 pages, 2022.
- [51] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [52] A. Ech-Cherif, M. Misbhauddin, and M. Ech-Cherif, "Deep neural network based mobile dermoscopy application for triaging skin cancer detection," in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–6, Riyadh, Saudi Arabia, 2019.
- [53] A. O. Salau and S. Jain, "Feature extraction: a survey of the types, techniques, applications," in *2019 International Conference on Signal Processing and Communication (ICSC)*, pp. 158–164, Noida, India, 2019.
- [54] G. Danala, S. K. Maryada, H. Pham, W. Islam, M. Jones, and B. Zheng, "Comparison of performance in breast lesions classification using radiomics and deep transfer learning: an assessment study," in *Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment*, (Vol. 12035, pp. 215–222)SPIE.
- [55] W. Islam, M. Jones, R. Faiz, N. Sadeghipour, Y. Qiu, and B. Zheng, "Improving performance of breast lesion classification using a ResNet50 model optimized with a novel attention mechanism," *Tomography*, vol. 8, no. 5, pp. 2411–2425, 2022.
- [56] A. A. Hameed, B. Karlik, and M. S. Salman, "Back-propagation algorithm with variable adaptive momentum," *Knowledge-Based Systems*, vol. 114, pp. 79–87, 2016.
- [57] M. Umer, I. Ashraf, S. Ullah, A. Mehmood, and G. S. Choi, "COVINet: a convolutional neural network approach for predicting COVID-19 from chest X-ray images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 1, pp. 535–547, 2022.
- [58] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, and V. Singh, "Application of deep learning for fast detection of COVID-19 in X-rays using nCOVnet," *Chaos, Solitons & Fractals*, vol. 138, article 109944, 2020.
- [59] M. Z. Che Azemin, R. Hassan, M. I. Mohd Tamrin, and M. A. Md Ali, "COVID-19 deep learning prediction model using publicly available radiologist- adjudicated chest X-ray images as training data: preliminary findings," *International Journal of Biomedical Imaging*, vol. 2020, Article ID 8828855, 7 pages, 2020.
- [60] K. El Asnaoui and Y. Chawki, "Using X-ray images and deep learning for automated detection of coronavirus disease," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 10, pp. 3615–3626, 2021.
- [61] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidxnet: a framework of deep learning classifiers to diagnose COVID-19 in X-ray images," 2020, <http://arxiv.org/abs/2003.11055>.
- [62] J. Civit-Masot, F. Luna-Perejón, M. Domínguez Morales, and A. Civit, "Deep learning system for COVID-19 diagnosis aid using X-ray pulmonary images," *Applied Sciences*, vol. 10, no. 13, p. 4640, 2020.
- [63] H. M. Shyni and E. Chitra, "A comparative study of X-ray and CT images in COVID-19 detection using image processing and deep learning techniques," *Computer Methods and Programs in Biomedicine Update*, vol. 2, article 100054, 2022.
- [64] X. Xu, X. Jiang, C. Ma et al., "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.
- [65] A. K. Mishra, S. K. das, P. Roy, and S. Bandyopadhyay, "Identifying COVID19 from chest CT images: a deep convolutional neural networks based approach," *Journal of Healthcare Engineering*, vol. 2020, Article ID 8843664, 7 pages, 2020.
- [66] A. Abbasian Ardakani, U. R. Acharya, S. Habibollahi, and A. Mohammadi, "COVIDiag: a clinical CAD system to diagnose COVID-19 pneumonia based on CT findings," *European Radiology*, vol. 31, no. 1, pp. 121–130, 2021.
- [67] S. Yang, L. Jiang, Z. Cao et al., "Deep learning for detecting corona virus disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study," *Annals of Translational Medicine*, vol. 8, no. 7, 2020.
- [68] J. F. H. Santa Cruz, "An ensemble approach for multi-stage transfer learning models for COVID-19 detection from chest CT scans," *Intelligence-Based Medicine*, vol. 5, article 100027, 2021.
- [69] A. K. Abdar, S. M. Sadjadi, H. Soltanian-Zadeh, A. Bashirgonbadi, and M. Naghibi, "Automatic detection of coronavirus (COVID-19) from chest CT images using VGG16-based deep-learning," in *2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME)*, pp. 212–216, Tehran, Iran, 2020.
- [70] A. Saood and I. Hatem, "COVID-19 lung CT image segmentation using deep learning methods: U-net versus SegNet," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–10, 2021.

- [71] T. Natnael, Y. Alemnew, G. Berihun et al., "Facemask wearing to prevent COVID-19 transmission and associated factors among taxi drivers in Dessie City and Kombolcha Town, Ethiopia," *PloS One*, vol. 16, no. 3, article e0247954, 2021.
- [72] A. G. Yadessa and A. O. Salau, "Low cost sensor based hand washing solution for COVID-19 prevention," in *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pp. 93–97, Zallaq, Bahrain, 2021.

Research Article

A Novel Approach for Best Parameters Selection and Feature Engineering to Analyze and Detect Diabetes: Machine Learning Insights

Md Shahin Ali ¹, Md Khairul Islam ¹, A. Arjan Das ¹, D. U. S. Duranta ¹,
Mst. Farija Haque ¹ and Md Habibur Rahman ^{2,3}

¹Department of Biomedical Engineering, Islamic University, Kushtia 7003, Bangladesh

²Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh

³Center for Advanced Bioinformatics and Artificial Intelligent Research, Islamic University, Kushtia 7003, Bangladesh

Correspondence should be addressed to Md Habibur Rahman; habib@iu.ac.bd

Received 18 October 2022; Revised 26 December 2022; Accepted 13 April 2023; Published 4 May 2023

Academic Editor: Francis M. Bui

Copyright © 2023 Md Shahin Ali et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Humans are familiar with “diabetes,” a chronic metabolic disease that causes resistance to insulin in the human body, and about 425 million cases worldwide. Diabetes is a hazard to human health since it can gradually cause significant damage to the heart, blood vessels, eyes, kidneys, and nerves. As a result, it is critical to recognize diabetes early on to minimize its negative consequences. Over the years, artificial intelligence (AI) technology and data mining methods are playing a crucial role in detecting diabetic patients. Considering this opportunity, we present a fine-tuned random forest algorithm with the best parameters (RFWBP) that is used with the RF algorithm and feature engineering to detect diabetes patients at an early stage. We have employed several data processing techniques (e.g., normalization, conversion into numerical data) to raw data during the preprocessing phase. After that, we further applied some data mining techniques, adding related characteristics to the primary dataset. Finally, we train the proposed RFWBP and conventional methods like the AdaBoost algorithm, support vector machine, logistic regression, naive Bayes, multilayer perceptron, and a regular random forest with the dataset. Furthermore, we also utilized 5-fold cross-validation to enhance the performance of the RFWBP classifier. The proposed RFWBP achieved an accuracy of 95.83% and 90.68% with and without 5-fold cross-validation, respectively. Moreover, the proposed RFWBP is compared with conventional machine learning methods to evaluate the performance. The experimental results confirm that the proposed RFWBP outperformed conventional machine learning methods.

1. Introduction

1.1. Background. Diabetes mellitus (DM) is the most common long-lasting noncommunicable public health concern that causes serious health complications, e.g., kidney disease, cardiovascular disease, and lower-limb amputations that increase morbidity and reduce lifespan [1, 2]. A high blood sugar level is responsible for DM, leading to human metabolic disorders. Insulin is a type of hormone released from the pancreas into the bloodstream. Insulin helps glucose enter the body cell from the bloodstream and balance the sugar level. When the pancreas fails to secrete enough insulin, sugar also fails to enter the body cell; subsequently, the

sugar level increases and causes diabetes. Diabetes is influenced by various factors such as height, weight, genetic factors, and insulin, but the most important thing is remembering the sugar concentration [3]. It is the cause of many fatal diseases like cardiovascular disease, nerve damage, kidney damage, and depression. They are sorts of type 1 (during childhood), type 2 (at any age), and gestational (pregnant women) [4]. The latest prediction shows that the disease burden of DM had a global prevalence of 425 million people with diabetes in 2017, which is estimated to rise to 629 million by 2045 due to the majority of obesity, physical inactivity, poor diets, sedentary lifestyle, and also genetics [5]. Most of these numerical increments will face in

developing countries [6]. According to the world health organization (WHO), more than 77% of patients have reached severe cases due to DM over more than 20 years [7]. Diabetes and its complications affect individuals physically, financially, and socially. According to the report, 1.2 million people die yearly from an untreated health condition. Diabetes-related risk factors, such as cardiovascular and other disorders, resulted in about 2.2 million deaths. Generally, the DM diagnosis process is time-consuming and complex because of the physician's manipulation. On the other hand, the physician only focuses on the present patient report. But computational detection using a machine learning (ML) algorithm compares the current report with many other factors, which gives a more accurate result. In addition, DM is a life-quality-reducing disease that can lead to more severe issues in the human body. For this reason, it is challenging and essential to diagnose and identify diabetes at the primary phase early. Early diagnosis is a procedure for detecting a disease or disorder in patients in the early stages. It enables people to make important decisions about their care, support, and financial and legal matters. Furthermore, it helps them get crucial information, counsel, and support as they face new problems. However, detecting diabetes early on becomes more challenging due to the uncertainty of the parameters of different physical, environmental, and family backgrounds. Besides, the value of the parameters varies from person to person.

ML is an artificial intelligence- (AI-) based application that automatically builds an analytical model which can be learned from data, identifying the patterns and determining with minimal latency. It can learn something and overcome the deficiencies from experience as humans do [7, 8]. ML-based algorithms are essential for investigating this issue and developing a more accurate CAD scheme for predicting not only the survival rate but also other factors for diabetes in the current era, as they dominate the various tasks of computer vision and the medical industry, including radiology [8, 9]. It is used in the medical field to detect fatal diseases. Also, it assists in streamlining hospital administrative processes, mapping and treating infectious diseases, and personalizing medical treatments [10, 11]. Moreover, ML is also applicable to biological data to extract knowledge by taking the help of feature engineering techniques and diagnosing human-threatening diseases like DM [10]. To accomplish this analysis, random forest is employed with its best parameters. Random forest is a supervised learning method that can be utilized for data classification and prediction. Nonetheless, it is primarily employed to overcome classification issues. The random forest algorithm constructs decision trees from sample data, generates predictions from each one, and then conducts a vote to identify the optimal option. This ensemble technique is preferred to an individual decision tree since it averages the findings to reduce overfitting [11].

Feature engineering converts raw data into features that may be used to construct a prediction model with ML or statistical modeling. It aims to optimize ML models' performance by preparing an input dataset that best fits the algorithm. In addition, k-fold cross-validation is a frequently employed approach for testing the performance of an ML

model. It involves randomly partitioning the data into a collection of folds, where each fold is utilized as a test set in turn while the remaining folds are used as training data. This procedure is done K times, with each fold serving as the test set exactly once. The performance indicator is then averaged over K iterations.

1.2. Motivation. Our research investigated the various aspects of diabetes that helped us identify it early. We used feature engineering techniques to train the algorithm, which helped to provide the best output. In our study, we used a random forest with its best parameters to improve the performance of diabetes identification by employing the tuning parameters and applying the grid search approach. The parameters of RF are tweaked to create a superior classifier that is more robust and precise. GridSearchCV holds all the best parameters to obtain such a type of classifier.

Using the best parameters of a random forest classifier can provide several benefits and could be our study's novelty and motivating factors. These motivating factors are as follows.

1.2.1. Improved Performance. By tuning the parameters of a random forest classifier, we can often achieve better performance in terms of accuracy, precision, recall, and other evaluation metrics. This is a key motivation for using the best parameters, as they can improve the classifier's effectiveness [12].

1.2.2. Reduced Overfitting. Overfitting occurs when a model is too complex and has too many parameters, leading to a poor generalization of new data. Using the best parameters of a random forest classifier can reduce the risk of overfitting and improve the model's ability to generalize to unseen data [13].

1.2.3. Increased Efficiency. Some parameters of a random forest classifier, such as the number of trees and the maximum depth of each tree, can impact the computational efficiency of the model. Using the best parameters, we can optimize the computational efficiency of the classifier and potentially reduce the time and resources required for training and prediction [14].

1.2.4. Enhanced Interpretability. The parameters of a random forest classifier can also affect the interpretability of the model. For example, using the best parameters may result in a simpler and more easily interpretable model, which can also be a motivating factor for using the best parameters in our paper [15].

In summary, using the best parameters of a random forest classifier can lead to improved performance, reduced overfitting, increased efficiency, and enhanced interpretability.

1.3. Contributions. The following is a summary of our paper's primary contribution:

- (i) We present a random forest algorithm with its optimal parameters (RFBWP) that more effectively diagnoses diabetes in early-stage patients

- (ii) Our proposed RFWBP achieves much better accuracy when compared with other existing ML algorithms within a short time
- (iii) We use feature engineering techniques to extract the features from the raw data, taking some preprocessing strategies that help get better performances
- (iv) We use the k-fold cross-validation technique with the best parameters for the proposed RFWBP algorithm as well as some ML algorithms like decision tree (DT), random forest (RF), support vector machine (SVM), AdaBoost, and linear regression (LR) which determines the detection diabetes, giving the reader better insight regarding the classification approach

The remainder of this work is structured as follows: the literature review discussion is represented by Section 2. Section 3 contains materials and methodologies. Section 4 elaborates on the suggested RFWBP technique. The outcome and discussion are mentioned in Section 5, whereas the conclusion and future recommendations are summarized in Section 6.

2. Literature Review

Diabetes is a chronic and significant health problem that leads to many complications in the human body. Many researchers investigated diabetes using ML techniques to extract features for predicting and identifying diabetes. Sisodia S and Sisodi D [16] proposed predictive analysis models based on DT, SVM, and naive Bayes (NB) algorithms. They got 76.30% as the highest accuracy from NB, which could be improved using a large dataset with some fruitful preprocessing steps. In [17], Alehegn used several ML algorithms, including logistic regression, NB, and SVM, to evaluate the method with 10fold cross-validation [18]. They showed that SVM obtained the best performance and accuracy of 84%. However, the accuracy needed to be increased for the prediction of DM. Perveen et al. [19] looked into the effectiveness of AdaBoost and bagging ensemble ML algorithms in classifying DM patients based on diabetic risk factors utilizing the J48 decision tree as a baseline. The experiment results indicate that AdaBoost surpasses bagging and a J48 decision tree regarding efficiency. Shakeel et al. [4] proposed a cloud-based framework to diagnose DM using k-means clustering, where they compared their work with the other two clustering methods and found better results than the other two. But their framework gives the predicted outcome for only a specific group of affected people. Vijayan and Anjali [3] have taken another ML approach implementing SVM, k-nearest neighbors (KNN), and a decision tree. They found the highest 80.72% accuracy using the AdaBoost algorithm with a decision stamp as a base classifier. Barakat et al. [20] used an SVM classifier to detect DM with good accuracy. Moreover, they used an additional explanation module to make SVM more effective, which helped get better performances. A survey has been done by Shivakumar [21] on data mining technologies for diabetic prediction. After analyzing essential research papers, they found some relation among the

diseases like wheezing, edema, oral disease, female pregnancy, and age with having a person diabetes. Choudhury and Gupta [22] surveyed various ML techniques using a dataset (PIMA Indian diabetes dataset) to analyze different models. Finally, they found the best 77.61% of accuracy at LR. SVM and KNN also worked well on that dataset. Sumangali [23] made a model by combining RF and classification and regression tree (CART), which gave them an excellent performance. They also found that a combined classifier model is much more effective than a single classifier model. Experimental work has been done by Chowdhary et al. [24] on diabetes retinopathy detection using ensemble ML algorithms. They found that their model outperformed other existing ML algorithms. Zou et al. [25] tried to detect DM with ML algorithms such as decision trees, random forests, and neural networks. They also used 5-fold cross-validation to examine their model precisely. To reduce the dimensionality, principal component analysis (PCA) and minimum redundancy with maximum relevance have been used and finally found the maximum 80.84% accuracy from the random forest classifier. In [26], Rahman et al. used LR based on p value and odds ratio to predict risk factors for diabetes disease. They proposed a combined LR-based feature selection and RF-based classifier model, which gives better results than other models. Saxena et al. [27] proposed a method using KNN, which acquired an accuracy of 70%, where it should be improved considering a larger dataset. In [28], there is a proposed method based on an NB classifier with good accuracy of 77.01%. In addition, Perveen et al. [19] applied the AdaBoost classifier, offering better performance in detecting DM. However, the work could be more impactful using a large dataset with some preprocessing steps. In [29], Nai-arun and Moungrmai used an algorithm to classify the risk of DM. The authors used DT, ANN, LR, and NB ML classification methods to achieve the outcomes. Additionally, bagging and boosting techniques are utilized to increase the consistency of the constructed model. According to the test results, the RF algorithm performed best against all the algorithms used. However, all the associated parameters are needed to fit the model perfectly. Also, they could have increased the performance by using the best parameter of the algorithms.

In prior research, the authors employed traditional statistical machine-learning methods to identify diabetes in tabular data. Their investigation on a few small datasets utilized a black-box-like algorithm that obtained 70-85% accuracy based on their experiment. However, our research employed the random forest technique with its optimal parameters. When using the optimal parameters for an ML algorithm, we are effectively fine-tuning the model to perform optimally on a certain dataset. This can result in enhanced accuracy, precision, and recall, as well as shorter training and inference times, regardless of the dataset size or complexity.

3. Materials and Methods

The materials and methods section elaborates the working procedures from first to last, which helps understand the

method well-handled. Here, we describe the steps that help to analyze our research study in Figure 1. We use several ML techniques to identify whether the patient has diabetes or not.

3.1. Dataset. A dataset aggregates some necessary data to help the model perform better. It is fed to the ML algorithm to ensure how accurately the algorithm is interpreted [30]. In our research paper, we used a dataset of different features based on health information to diagnose whether the patient has diabetes. We collected the dataset from Kaggle [31], the world's largest data science community, with various tools and services to assist in achieving data science objectives. The dataset named "Pima Indians Diabetes Database" contains some health condition features like pregnancies, glucose, blood pressure, age, skin thickness, insulin, BMI, and diabetes pedigree function from the patients, shown in Table 1. The dataset was manipulated by the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to determine the probability that an individual has diabetes based on the specific diagnostic metrics provided in the information. Multiple restrictions governed the inclusion of these occurrences from a more extensive database. There are 768 female patients aged 21-81 years old. The patients' average age and standard deviation are 33 and 11.76, respectively. Moreover, the descriptive statistics of the dataset are shown in Table 2, explicitly. The numbers of diabetes and no diabetes are 268 and 500, respectively, as shown in Figure 2. Besides this, some parameters have been added (see Table 3) by applying the feature engineering technique further to make this model more precise.

3.2. Data Preprocessing. Preprocessing transforms raw data into form machines, and computers can interpret and evaluate it. Text, photos, video, and other information about the real world are jumbled. In addition to including errors and inconsistencies, it is typically inadequate and needs a consistent style. Since computers prefer to work with clean data, they interpret it as 1s and 0s. It is estimated that data preparation accounts for 60% of all effort and time utilized in the data mining process [29–31]. We have utilized some preprocessing strategies in this work, such as data normalization, transformation, outlier identification, feature engineering, and feature selection, detailed in the following subsections.

3.3. Data Normalization. Normalization of data is a preprocessing approach that entails scaling or altering the data to ensure that every attribute contributes equally. The term normalization refers to the process of arranging data into multiple related tables in order to eliminate data redundancy. The performance of ML algorithms depends on the data quality used to build a comprehensive statistical model for the categorization problem. Recent research has highlighted the importance of data normalization for improving data quality and, subsequently, the performance of ML algorithms [32].

It entails data discretization, removal of outliers and noise, data integration from diverse sources, incomplete data

handling, and data transformation to comparable dynamic ranges [33, 34]. The researchers give various options for rescaling or transforming the data using these metrics, such as z-score normalization, min-max normalization, max normalization, decimal scaling normalization, and MaxAbsScaler. Our experiment used the MaxAbsScaler normalization technique, which performed better on this dataset. MinMaxScaler was also applied to this dataset. The performance of the two scaling procedures is nearly identical because all of the data are positive. MaxAbsScaler scales and transforms each feature in the dataset by its most outstanding absolute value [32, 35, 36]. This estimator scales and encodes each component independently, resulting in a potential overall mass of 1.0 for each training set feature. It neither moves nor centers the data. Thus, there is no reduction in sparsity [37, 38]. Mathematically,

$$X_{\text{scaler}} = \frac{X_{\text{std}}}{(\max - \min) + \min}. \quad (1)$$

3.4. Outlier Detection. Data quality is essential to ensure the robust result of high-dimensional datasets. Outlier is a solution for providing the data quality of datasets. The conventional technique of outlier detection excludes the distribution's tails and ignores the data generation process of a particular dataset [39]. But outlier detection in ML brings a new dimension to ensuring data quality in a dataset. Outliers are data points significantly different from other data points present in given data sets [40]. Generally, we apply outlier detection on training data to eradicate outlier pollution of train data. They have various applications of outliers in multiple sectors like military service for enemy activity identification, deception identification, medical and public health data, industrial damage identification, and image processing [41]. The datasets contain features like patient age, blood group, height, and weight in the medical sector. One of the most critical tasks in the statistical analysis of time series data is detecting outliers or typical data structures, as outliers can significantly impact the study's outcome [42]. The numeric outlier technique is employed in this study to identify data mistakes that can then be removed. As an outlier detector, Tukey's fencing is utilized in this study [43]. It is the simplest nonparametric outlier identification method in a one-dimensional feature space. In this case, the interquartile range (IQR) is used to calculate outliers, and hereinafter, scale (k) ranges from 1.5 for regular and 3 for extreme outliers. The first and third quartiles are determined for Q1 and Q3. An outlier is a data point x_i that is outside the interquartile limit. Mathematically,

$$x_i > Q3 + k(\text{IQR}); x_i < Q1 - k(\text{IQR}), \quad (2)$$

where $\text{IQR} = Q3 - Q1$ and $k \geq 0$.

3.5. Feature Engineering. Feature engineering is a significant step before building a precise model. Finding all the necessary features in a compatible format while working with an ML algorithm [44] is crucial. Without these essential features, the algorithm does not perform properly, and the result also goes down. The term feature engineering presents

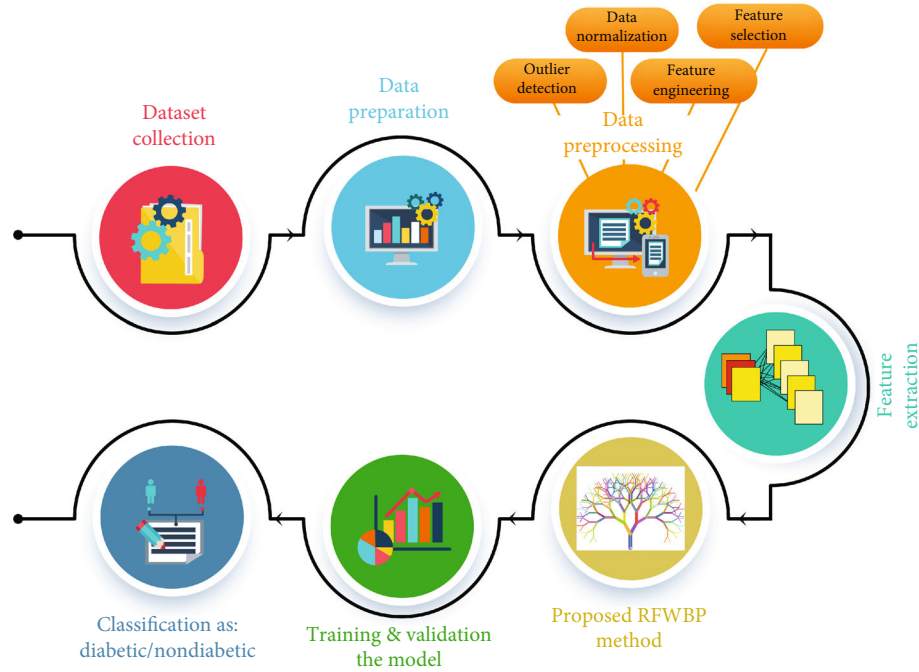


FIGURE 1: The following steps of our proposed methodology.

TABLE 1: The detailed information of the dataset (5 instances) before applying the feature engineering technique.

| Pregnancies | Glucose | Blood pressure | Skin thickness | Insulin | BMI | Diabetes pedigree function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|----------|----------------------------|-----|---------|
| 6 | 148 | 72 | 35 | 0 | 33.60000 | 0.62700 | 50 | 1 |
| 1 | 85 | 66 | 29 | 0 | 26.60000 | 0.35100 | 31 | 0 |
| 8 | 183 | 64 | 0 | 0 | 23.30000 | 0.67200 | 32 | 1 |
| 1 | 89 | 66 | 23 | 94 | 28.10000 | 0.16700 | 21 | 0 |
| 0 | 137 | 40 | 35 | 168 | 43.10000 | 2.28800 | 33 | 1 |

TABLE 2: Descriptive statistics of the dataset.

| Variable | Distinct | Min | Max | Zeros | Mean | STD | Variance | Skewness | Missing |
|----------------------------|----------|-------|------|-------|---------|---------|-----------|----------|---------|
| Pregnancies | 17 | 0 | 17 | 111 | 3.845 | 3.369 | 11.354 | 0.902 | 0 |
| Glucose | 136 | 0 | 199 | 5 | 120.895 | 31.973 | 1022.248 | 0.174 | 0 |
| Blood pressure | 47 | 0 | 122 | 35 | 69.105 | 19.356 | 374.647 | -1.844 | 0 |
| Skin thickness | 51 | 0 | 99 | 227 | 20.536 | 15.952 | 254.473 | 0.109 | 0 |
| Insulin | 186 | 0 | 846 | 374 | 79.799 | 115.244 | 13281.180 | 2.272 | 0 |
| BMI | 248 | 0 | 67.1 | 11 | 31.993 | 7.884 | 62.159 | -0.428 | 0 |
| Diabetes pedigree function | 517 | 0.078 | 2.42 | 0 | 0.472 | 0.331 | 0.109 | 1.919 | 0 |
| Age | 52 | 21 | 81 | 0 | 33.241 | 11.760 | 138.303 | 1.129 | 0 |
| Sex | | | | | Female | | | | |
| Sample size | | | | | 768 | | | | |

similar activities like improving the existing features (see Figure 3) and adding some new features [45]. It is all about feeding the model and making it more fruitful. Some practical steps of feature engineering include feature generation, feature extraction, feature transformation, feature selection, and feature analysis and evaluation. It is also a method for

transforming unprocessed data into features that better address the core problem with ML models, resulting in increased model accuracy on previously unknown data [46]. Our research study has added exclusive features based on existing features of raw data labeled as BMI category, glucose category, blood category, skin-thickness category, and

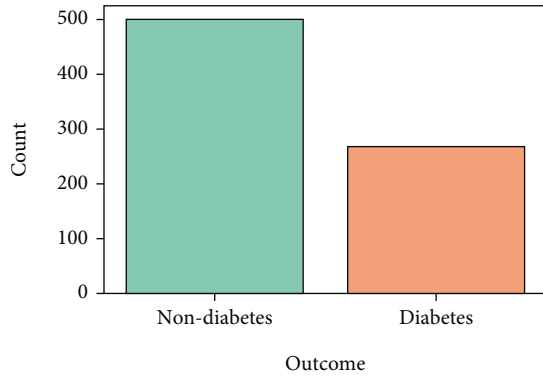


FIGURE 2: The number of instances of the target (outcome) column.

insulin category to get the best performances from our method. Table 4 shows the added features based on the value ranges of raw data.

Moreover, one-hot encoding may also use in feature engineering that encodes the categorical variable to a numeric form for the better prediction skill of an ML algorithm [47]. In ML, the dataset contains many categories of data. Some algorithms can work with the categorized values, but most need help. Labeled data is a big problem for them, so the data must be converted into numeric. To make the data more acceptable, we rebuild the encoding dimension of the main network package in the data collection by applying individual-heat coding to generate two-dimensional data. We used a one-hot encoding technique to add a binary variable for each unique categorical value. It deals with only 1 and 0. The actual values are assigned to 1, and the remaining variables are considered false and assigned to 0 [48].

3.6. Feature Selection. Feature selection automatically or manually selects those features, contributing much to predicting the results from a model [31]. It is classified into three groups based on filters, wrappers, and embeds used for statistical measures between the input variables. The wrapper feature selection method uses an induction learning algorithm to evaluate the feature subset. It measures the performance based on categorizing the rate gained from the testing set. The embedded process uses a particular supervised and unsupervised ML algorithm to incorporate sense about the specific form of the class. The filter method shows complete independence between the learning machine and raw data, which is relatively robust against overfitting [49]. They can be filtered to select the relevant features, reducing the noise effects from the overall raw data [50]. It has been discussed in ML and the data mining field to find the best k features and avoid generalization errors in the generalization errors [51]. In addition, Figure 4 depicts the histograms of each feature in our experimented dataset, which is the quickest way to understand the distribution of each attribute in the dataset.

Our research used the filter-based feature selection method (see Figure 5) to select the best features from our raw data that provide good identification performances. Filter techniques assess the quality of data subsets by looking at

just the intrinsic data features in which a single data or a group of data is generally compared to a class label [52]. Rather than cross-validation performance, filter approaches focus on the inherent qualities of features as assessed by univariate statistics. It states that if a feature is valid, it can be independent of the input data but not of the class labels, i.e., a feature that does not affect the class labels can be ignored [50]. It selects the features based on various statistical correlations with outcome variables of any ML algorithm independently. Here, the correlation is a subjective matter for the continuous variables, whose value varies from -1 to +1. It must reduce multicollinearity before training the model. Moreover, the Pearson correlation among the input features is shown in Figure 6. Mathematically,

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum((x_i - \bar{x}))^2 \sum(y_i - \bar{y})^2}}. \quad (3)$$

Here, r is the correlation coefficient; x_i is the x -variable values in a sample; \bar{x} is the mean of the x -value variables; y_i is the y -variable values in a sample; \bar{y} is the mean of the y -value variables [53].

3.7. Feature Extraction. Feature extraction is the process of selecting the essential and relevant data by separating all data into some groups [51, 52]. While working with a large dataset, collecting all the necessary information or reducing the loss of relevant data is crucial. Feature extraction helps manage the critical information out of the massive raw dataset reducing the data loss rate. A large dataset causes many problems. It requires a lot of memory, computation power also goes slow, causing overfitting to training samples, and the most important one is that it also lowers the model's performance [54]. To overcome these, feature extraction derives all the nonredundant values from the initially measured dataset. It is similar to dimensionality reduction, increasing the algorithm speed [53, 55, 56]. It is critical for future data analysis; whether it is model acknowledgment, denoising, data abbreviation, or imagination, the data must be represented in a way that makes resolution easier [55, 57]. The extraction of features begins with the collection of quantitative information. It generates derived values (features) that are intended to be valuable and nonredundant, facilitating the learning and adaptation procedures and, in some instances, leading to superior human interpretations by using several feature extraction techniques such as principal component analysis (PCA), random projection algorithm (RPA), and Isomap to recognize unnecessary features and reduce ineffective and redundant ones [56]. Tables 1 and 3 represent the final set of attributes employed in the analysis of this study. The precise characteristics were derived from the combined form of these records used for further assessment. Some equations for feature extraction are as follows: the essential concept is that a linear, causal, stable, time-invariant system with impulse response can provide a random sequence as an output $h(n)$ and a white noise sequence

TABLE 3: The added features which were taken after applying the feature engineering technique.

| New_BMI_cat | New_glucose_cat | New_blood_cat | New_skin thickness_cat | New_insulin_cat |
|--------------|-----------------|---------------|------------------------|-----------------|
| Obese | Prediabetes | Normal | 0 | Abnormal |
| Slightly_fat | Normal | Normal | 0 | Normal |
| Normal | Prediabetes | Normal | 0 | Abnormal |
| Slightly_fat | Normal | Normal | 0 | Normal |
| Obese | Normal | Normal | 0 | Abnormal |

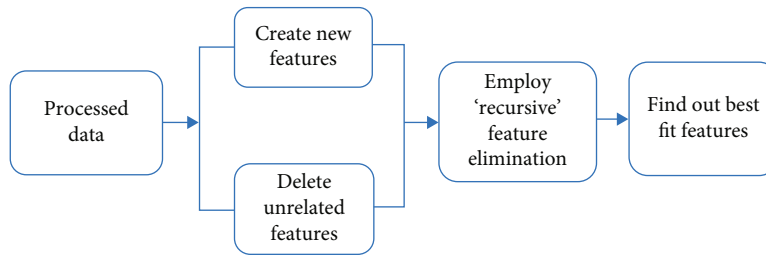


FIGURE 3: The working principle of feature engineering.

TABLE 4: The added categorical features based on the value ranges.

| New_BMI_cat | | New_glucose_cat | | New_blood_cat | | New_skin thickness_cat | | New_insulin_cat | |
|-------------|--------------|-----------------|---------------|----------------------|----------------------|------------------------|----------------------|-----------------|----------------|
| BMI range | BMI label | Glucose range | Glucose label | Blood pressure range | Blood pressure label | Skin thickness range | Skin thickness label | Insulin range | Insulin labels |
| 0-18.4 | Weakness | 0-139 | Normal | 0 - 79 | Normal | 1-18 | Normal | 0-16 | Normal |
| 18.4-25.0 | Normal | 139-200 | Prediabetes | 79-90 | Hypertension_S1 | 19-88 | Abnormal | 17-166 | Abnormal |
| 25.0-30.0 | slightly_fat | — | — | 90-123 | Hypertension_S2 | — | — | — | — |
| 30.0-70.0 | Obese | — | — | — | — | — | — | — | — |

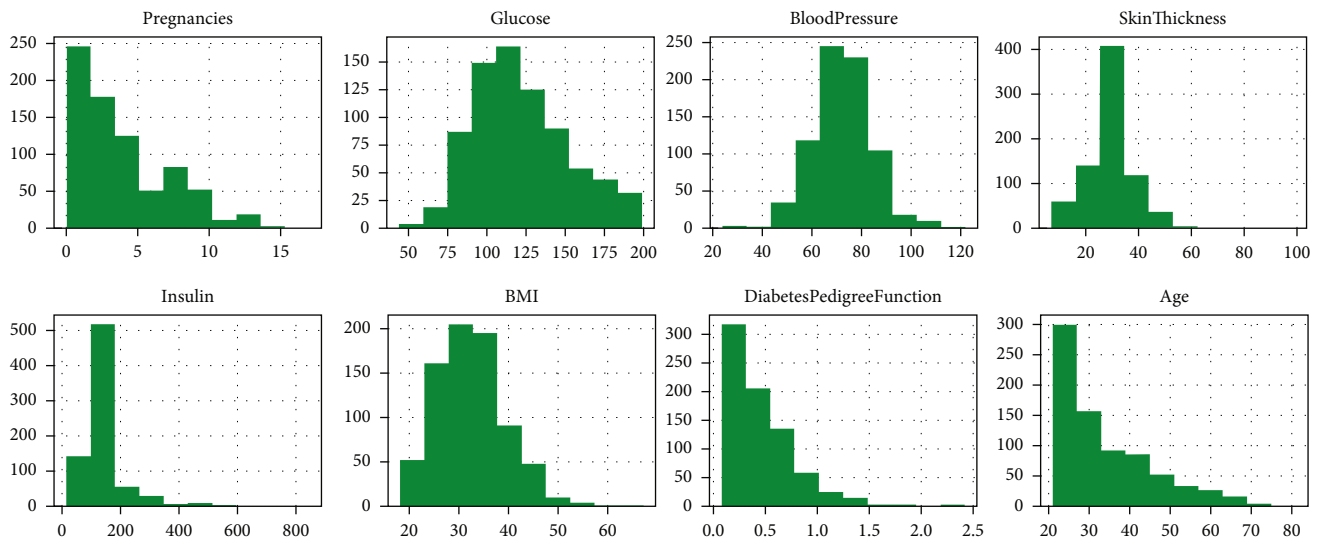


FIGURE 4: Histogram for each feature of our dataset.

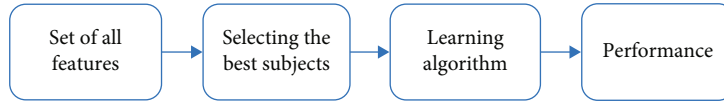


FIGURE 5: Working principle of filter method.

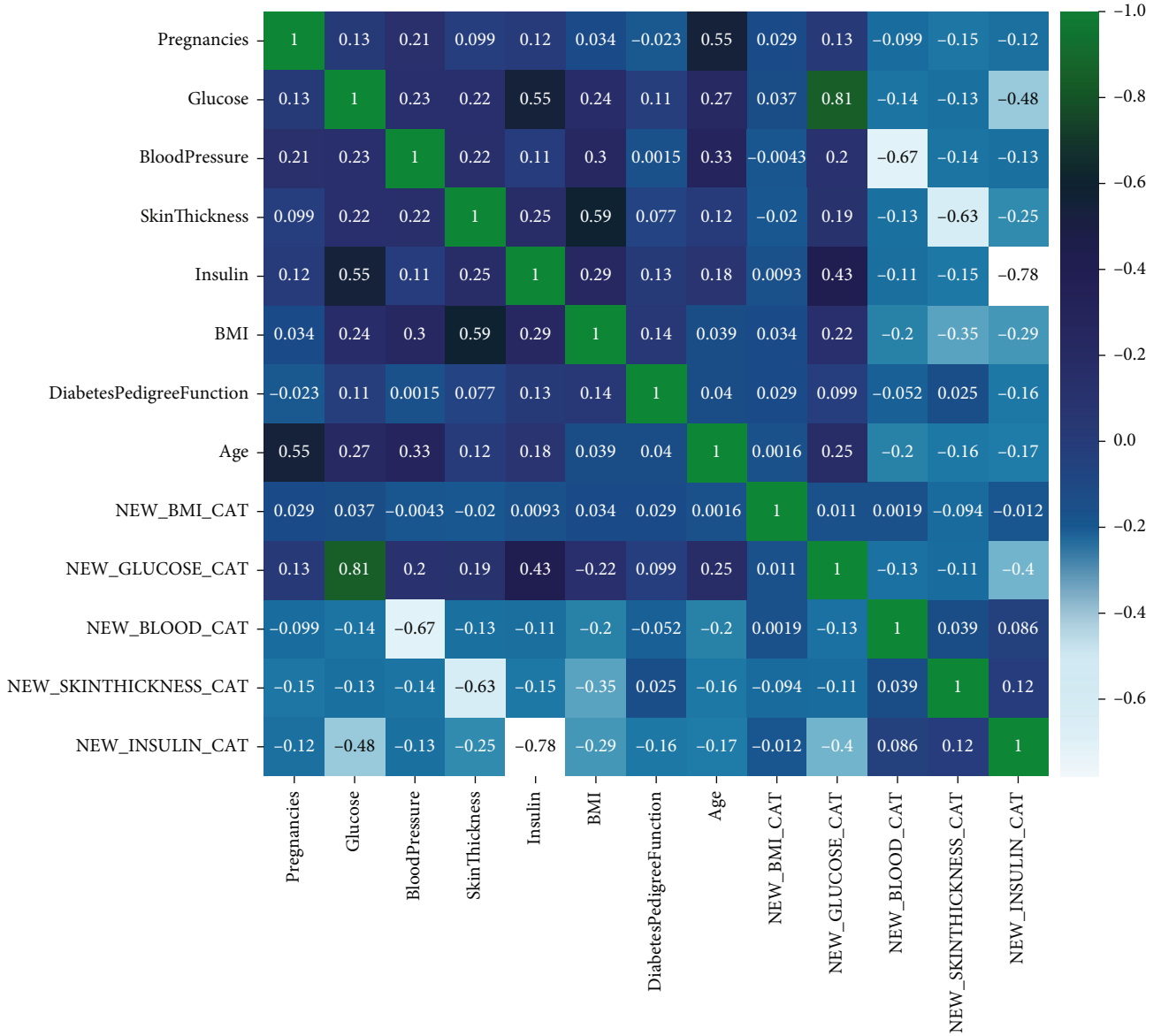


FIGURE 6: Correlation among all input features.

as input. Let $I(n)$ be a stationary random sequence with $R(k)$ autocorrelation.

$$I(n) = \sum_{k=0}^{\infty} h(k)\mu(n-k). \tag{5}$$

$$R(k) = E[I(n)I(n-k)]. \tag{4}$$

The process is known as an autoregressive (AR) process and is developed recursively.

We get equivalently for I and (n) when (n) represents a white noise sequence (n) .

$$I(n) = \sum_{k=1}^p a(k)I(n-k) + \mu(n). \tag{6}$$

It can be seen right away that $I(n)$ is a linear combination of preceding random sequences. $I(n, k)$ values plus an additive constant (n). The AR model's order is denoted by p . With $k = 1, 2 \dots$, the correlation coefficients $a(k), p$ are the AR model's parameters, and at the same time, whenever the sequence's predictor parameters (n). To put it another way, they reflect the weighting terms of previous sampled values $I(n1), \dots, I(np)$ and serve as a predictor of the actual value $I(n)$.

$$I(n) = \sum_{k=1}^p a(k)I(n-k) = a^T I(n-1). \quad (7)$$

With $I^T(n-1) = [I(n-1), \dots, I(n-p)]$ and the prediction error $\mu(n)$, $a^T = [a(1), a(2), \dots, a(p)]$ is an unknown parameter vector.

$$E[\mu^2(n)] = E[(I(n) - \hat{I}(n))^2] = E[(I(n) - a^T I(n-1))^2]. \quad (8)$$

The unknown parameters can be deduced from the data.

$$E[I(n-1)a^T(n-1)]a = E[I(n)I(n-1)]. \quad (9)$$

In a matrix, notation is equivalent to

$$R^T a = R, \quad (10)$$

with $r \equiv [r(1), \dots, r(p)]^T$. The prediction error $\sigma\mu^2$ is calculated based on

$$\sigma^2\mu = E[\mu^2(n)] = R(0) - a(k)R(k). \quad (11)$$

This attribute is very desirable according to the Levinson-Durbin algorithm.

4. Proposed RFWBP Method

Like the RF method, our proposed RFWBP method is a supervised learning technique used for classification and regression problems but primarily used in classification problems. It is blended with the RF algorithm and feature engineering. It selects the best parameters from the total number of parameters and uses them to predict and classify the problem.

RF with a single tree is a simple decision tree that tends to overfit. The proposed RFWBP algorithm is developed of multiple trees based on the premise that a forest with more trees is more adaptable while reducing model variance. It makes the decision trees on data samples and gives a prediction for each tree to select the solutions by means and voting shown in Figure 7. RFWBP uses exclusive features based on existing raw data features labeled as BMI, glucose, blood, skin thickness, and insulin to get the best performance. The best parameters used in our proposed RFWBP are in Table 5.

The training dataset's cross-validation accuracy and the significance of every element as the performance parameter are measured using the RF algorithm. Fast trees or the basic units of an RF algorithm are distinct and can create collaterally. After that, we choose the best subset by observing the maximum aggregate of the average score and median score with minimum standard deviation (SD). To best prevent the overfitting problem, the k-fold cross-validation technique ensures stable performance.

All the procedures are given below.

Step 1. Using a parallel random forest (PRF) classifier, train the dataset and then measure and sort the median of the variables by their importance through 20 trials

Step 2. Select and add every feature containing the highest variables' importance and train the dataset by PRF with k-fold cross-validation

Step 3. Compute the score for every feature's F_i where $i = 1 \dots n$ (n expresses the number of features in the executing loop)

Step 4. Choose the best features' subsets by selecting the rules described below

Step 5. Repeat the steps until it arrives at the expected criteria

In Step 2, we train the classifier using PRF with k-fold cross-validation. In the j^{th} cross-validation, a set of $(F_i, A_j^{\text{learn}}, A_j^{\text{validation}})$ is obtained, representing the feature importance, learning accuracy, and validation accuracy, respectively. In Step 3, the score criterion is calculated using the above data. Step 3 takes the data from Steps 1 and 2 to create a score criterion used in Step 4. The following formula is used to compute the score of the feature i_{th} :

$$F_i^{\text{score}} = \sum_{j=1}^n F_{ij} \times (A_j^{\text{learn}} + A_j^{\text{validation}}). \quad (12)$$

The best features will be selected using the following rules in the next stage, the primary step of our algorithm: the best average + median score and the lowest standard deviation (SD).

- (i) *Rule 1.* Choose attributes that have the highest median score
- (ii) *Rule 2.* Choose features that have the highest average score
- (iii) *Rule 3.* Look for features that have the lowest SD

The best accuracy and lowest SD are obtained using these guidelines. As a result, the best selection of features tends to minimize the number of output features to the least possible. The RF importance of the component is determined using ML algorithms. We discover the subset of features with negligible characteristics while still accomplishing the problem's goal based on the estimated relevance value. We have

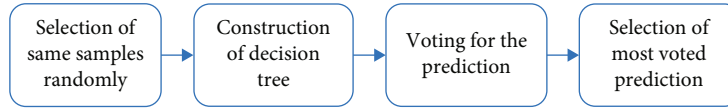


FIGURE 7: Working principle of random forest.

TABLE 5: Random forest best parameters after the tuning parameter using GridSearchCV.

| Tuning parameter | Best parameter | Parameter function |
|--|------------------------|--|
| “n_estimators”: [10, 17, 25, 33, 41, 48, 56, 64, 72, 80] | “n_estimators”: 33 | Number of trees that builds before taking the maximum voting |
| “Max_features”: (“auto” and “sqrt”) | “Max_features”: ‘auto’ | Number of features to consider when looking for the best split |
| “Max_depth”: [2, 4] | “Max_depth”: 4 | The depth of each tree in the forest |
| “Min_samples_split”: [2, 5] | “Min_samples_split”: 5 | Minimum number of samples required to split an internal node |
| “Min_samples_leaf”: [1, 2] | “Min_samples_leaf”: 2 | Minimum number of samples required to be at a leaf node |
| “Bootstrap”: (true, false) | “Bootstrap”: true | Involves random sampling of a dataset with replacement |

Input: Final dataset after all the preprocessing steps and feature engineering techniques.

1. Use a parallel random forest classifier and sort out the median of variables through 20 trials.
2. Train the classifier using PRF with k-fold cross validation and in the j^{th} cross-validation, a set of F_i , A_j^{learn} , and $A_j^{\text{validation}}$ is obtained.
3. Calculate the F_i score in every feature.
4. $i = 1 \dots n$ (n is the number of features).
5. $F_i^{\text{score}} = \sum_{j=1}^n F_{ij} \times (A_j^{\text{learn}} + A_j^{\text{validation}})$. (13)
6. Choose subsets containing the best features following the rules.
 - (i) Choose attributes that have the highest median score
 - (ii) Choose features that have the highest average score
 - (iii) Look for the features that have the lowest SD
7. Repeat the above steps until it arrives at the expected criteria.
8. Training with the RFWBP algorithm
9. Evaluate the RFWBP model
10. **Output: result** = nondiabetic and diabetic.

ALGORITHM 1: Proposed RFWBP method for diabetes identification [53].

implemented our model using rf_RandomGrid for searching the trees randomly by tuning the parameters that increase the model’s generalizability. Evaluating metrics are used for the conversion from the grid, and random combinations of hyperparameters are considered in every iteration in this search pattern which helps the model to show accurate performance. Furthermore, Algorithm 1 describes the overall Diabetes identification process using the proposed RFWBP method [58].

5. Result and Discussion

Our research study aims to identify diabetic patients based on diabetes risk factors like age, glucose level, blood sugar concentration, pregnancies, BMI, and skin thickness. We evaluated our study on a dataset from Kaggle [31]. The study is implemented using Jupyter Notebook and Google Colab.

In this research, we utilized the RFWBP algorithm to achieve the best results when comparing our technique to different ML algorithms, including DT, RF, SVM, NB, and AdaBoost. We also evaluated the findings using a 5-fold cross-validation and an alternative (without cross-validation) based on precision, recall, $F1$ score, and accuracy.

Precision measures the number of positive class predictions that have a place with the positive class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (13)$$

Recall evaluates the quantity of positive class prediction made out of all sure models in the dataset.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (14)$$

TABLE 6: The performance comparison of our proposed RFWBP method over other existing classifiers taking 70% as training with 30% as testing data.

| Model | Precision | Recall | F1 score | Accuracy |
|---------------------------|-----------|--------|----------|--------------|
| Decision tree | 92.90 | 91.72 | 92.31 | 88.61 |
| Support vector machine | 91.61 | 89.87 | 90.73 | 87.45 |
| AdaBoost | 94.85 | 88.02 | 91.30 | 87.87 |
| Naive Bayes | 85.81 | 90.48 | 88.08 | 84.42 |
| Logistic regression | 90.83 | 91.67 | 91.24 | 87.66 |
| Random forest | 90.32 | 90.32 | 90.32 | 87.01 |
| Gradient boosting machine | 88.99 | 90.65 | 89.81 | 85.71 |
| CatBoost | 89.91 | 91.59 | 90.74 | 86.84 |
| Multi-layer perceptron | 91.74 | 92.59 | 92.17 | 88.82 |
| Proposed RFWBP | 94.13 | 91.73 | 92.81 | 90.32 |

TABLE 7: Results comparison of different ML algorithms against our proposed RFWBP method using a 5-fold cross-validation technique (with a 95% confidence interval).

| Algorithm | 1st fold CV | 2nd fold CV | 3rd fold CV | 4th fold CV | 5th fold CV | Mean CV (accuracy) |
|---------------------------|-------------|-------------|-------------|-------------|-------------|--------------------|
| Decision tree | 87.01 | 87.66 | 88.31 | 88.24 | 87.58 | 87.76 |
| Support vector machine | 87.66 | 85.72 | 87.01 | 86.28 | 80.40 | 85.42 |
| AdaBoost | 87.66 | 84.42 | 87.01 | 88.24 | 87.58 | 86.98 |
| Naive Bayes | 85.07 | 79.87 | 86.36 | 86.28 | 81.05 | 83.73 |
| Logistic regression | 86.36 | 84.21 | 86.48 | 86.53 | 88.33 | 86.38 |
| Random forest | 87.66 | 84.41 | 87.01 | 89.54 | 88.24 | 87.37 |
| Gradient boosting machine | 88.32 | 87.37 | 88.89 | 91.21 | 89.68 | 89.01 |
| CatBoost | 92.19 | 88.21 | 94.56 | 91.12 | 93.78 | 91.97 |
| Multi-layer perceptron | 91.51 | 91.35 | 95.21 | 92.35 | 94.87 | 93.01 |
| Proposed RFWBP | 95.67 | 95.55 | 95.99 | 96.58 | 95.35 | 95.83 |

TABLE 8: The performance comparison of our proposed RFWBP method over other existing algorithms taking 80% as training with 20% as testing data.

| Model | Precision | Recall | F1 score | Accuracy |
|---------------------------|-----------|--------|----------|--------------|
| Decision tree | 88.99 | 91.51 | 90.23 | 86.36 |
| Support vector machine | 89.91 | 90.74 | 90.32 | 86.36 |
| Logistic regression | 90.83 | 91.67 | 91.24 | 87.66 |
| AdaBoost | 90.83 | 88.39 | 89.59 | 85.06 |
| Naive Bayes | 83.49 | 90.10 | 86.67 | 81.82 |
| Random forest | 92.66 | 90.99 | 91.82 | 88.31 |
| Gradient boosting machine | 88.07 | 91.43 | 89.72 | 85.53 |
| CatBoost | 88.99 | 93.27 | 91.08 | 87.50 |
| Multi-layer perceptron | 91.74 | 93.46 | 92.59 | 89.47 |
| Proposed RFWBP | 92.38 | 94.21 | 93.29 | 90.68 |

The *F1* score, also known as the harmonic mean, attempts to achieve a compromise between precision and recall. It accepts false negatives and false positives for calculation and operates well on an asymmetrical dataset.

$$F1 = \frac{2TP}{(2TP + FP + FN)}. \quad (15)$$

The total quantity of correctly predicted data points from the entire dataset is known as accuracy.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}. \quad (16)$$

Throughout the experiment, the performances were

divided into three distinct segments. Taking 70% for training with 30% for testing, 80% for training with 20% for testing, and a 5-fold cross-validation technique on the entire dataset. These are as follows:

in Table 6, 70% of the total dataset was utilized for training and 30% for testing. Comparing our proposed RFWBP method (without cross-validation) to existing ML algorithms, we obtained the lowest values from DT and the best deals from RF with the best parameters based on precision, recall, $F1$ score, and accuracy. Using 5-fold cross-validation, we compared the performance of our proposed method to that of various existing ML algorithms in Table 7. Our proposed classifier attained a maximum accuracy of 95.83% with confidence intervals of 95%. A few ML algorithms performed well; however, NB's cross-validation technique yielded the worst results. Table 8 contains a performance comparison between our proposed model (without cross-validation) and other existing ML algorithms using 80% of the whole dataset as training data and 20% as testing data based on precision, recall, $F1$ score, and accuracy. In addition, our suggested model achieved a maximum accuracy of 90.68 percent compared to existing ML methods.

In Table 9, we compared the results of our proposed RFWBP approach to the existing related work that they obtained from their research. It implies that our proposed method employing the best RF parameters provided the best results.

Figure 8 shows the graphical representation of the performance of our proposed classifier. It plots the true positive rate on the y -axis against the false positive rate on the x -axis at different classification thresholds. In creating a ROC curve, the classifier is first trained on a dataset and then tested on a separate dataset. The true positive and false positive rates for each classification threshold are calculated and plotted on the ROC curve. The resulting curve shows the trade-off between the true positive rate and the false positive rate for the classifier. Observing the curve, we find that the area under the curve (AUC) is 0.92, while we used 5-fold cross-validation on our proposed RFWBP classifier. A ROC curve can also be used to compare different classifiers' performance and identify the optimal classification threshold for a given classifier. Furthermore, the mean squared error (MSE) determined by the Python function is 0.0117.

Regarding activity versus better performance, gathering more data and feature engineering pays off the most. Still, once we have saturated all databases, it is time to move on to model hyper-parameter tuning. Random forest parameters are often used to boost the model's prediction power or make it easier to train. Hyperparameters are best compared to the settings of an algorithm that can be tweaked to improve performance.

In our study, we used the best parameters of the random forest algorithm instead of default parameters. Hence, we got the best performances that the random forest algorithm shows with its default parameters. Our dataset needs to have some best features from which we may get the best performances. We evaluated the dataset with the default parameter of the random forest, but it shows fewer performances than the random forest with the best parameters. A detailed dis-

TABLE 9: Performance comparison of our study on the same dataset against the existing related works.

| Authors | Algorithm | Accuracy (%) | Year |
|--------------------------|-----------|--------------|------|
| Saxena et al. [27] | KNN | 70.00 | 2014 |
| Rani and Jyothi [28] | NB | 77.01 | 2016 |
| Choudhury and Gupta [22] | LR | 77.61 | 2019 |
| Vijayan and Anjali [3] | AdaBoost | 80.72 | 2015 |
| Zou et al. [25] | RF | 80.84 | 2018 |
| Faruque [18] | SVM | 84.00 | 2019 |
| Khanam and Foo [59] | ANN | 88.60 | 2021 |
| Proposed method | RFWBP | 95.83 | 2023 |

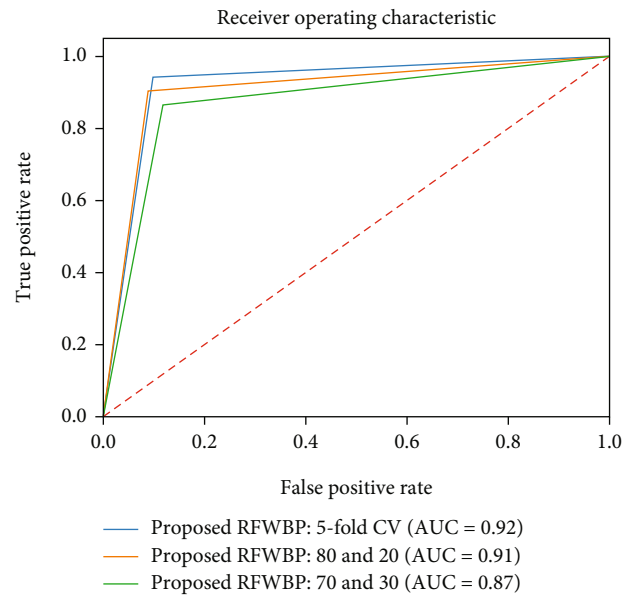


FIGURE 8: Receiver operating characteristics (ROC) curve with the area under the curve (AUC).

cussion of our proposed method is conducted based on numerical performance and visual results. In this study, we got a reasonable identification rate with the considerable help of data processing techniques described in the preprocessing section. After several times of fine-tuning, we got the best results using RF with its best parameters as a classifier. In our study, the RFWBP model was exclusively applied to data collected from Pima Indians. We will examine the performance of our suggested approach on other large datasets in the near future. Besides, k -fold cross-validation reduces the variation of the performance estimate by averaging the performance of multiple test sets. In addition, it enables us to use all the data for training and testing in order to obtain a more accurate estimate of the model's performance, which is essential when data is scarce. Furthermore, as it requires fewer iterations than other validation techniques, such as leave-one-out cross-validation, it is more computationally efficient. The performance estimate can be used to determine the hyperparameters that result in the best model performance.

The proposed method can be deployed in a computer-aided diagnosis system that will help effectively to identify diabetic patients at the early stage. In addition, the early identification of diabetes growth in humans, especially those without admittance to doctors, can significantly encourage them to get the treatment and enrich the survival possibility.

6. Conclusion and Future Work

In this study, we propose a method using the random forest algorithm with its best parameters to assemble a comprehensive data set, including diabetic and nondiabetic patients, to figure out the issue of inaccurate-accurate conclusions in diabetes identification. A medical diagnosis requires lots of information on the patient's physical condition. The motive for using these parameters was the same. It can detect the abnormality and identify the diabetic patient quickly in a short time. We have shown how to identify diabetes in two ways in our study. Finally, we got 95.83% of the highest accuracy using 5-fold cross-validation and 90.68% accuracy without k-fold cross-validation. Experimental results implied better accuracy, and the mentioned procedure has identical to other diabetes detection algorithms. When applied clinically, our proposed method can be used to detect diabetes quite accurately and precisely. Additionally, it will aid any organization's ability to diagnose many diabetes patients. However, it has some risk factors, such as incorrect blood glucose and insulin information, which reduces the ability to diagnose diabetes. The number of samples in our study is modest, and the results may need to be more generalizable to other groups or contexts due to the sample. The results of this study might not apply to real-world situations due to its artificial character or controlled conditions. In the future, we will extend our analysis by maximizing the number of subjects and features of both balanced and imbalanced datasets, which could provide detailed insights into the aspects that allow our model to identify diabetes patients more precisely.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

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

Authors' Contributions

Shahin Ali contributed to the conceptualization, methodology, software, visualization, writing—original draft preparation, and reviewing and editing. Khairul Islam contributed to the supervision, and writing—reviewing and editing. A Arjan Das contributed to the data curation, validation, and writing. D U S Duranta contributed to the data curation and writing. Farija Haque contributed to the data curation and writing. Md Habibur Rahman was responsible for supervision, and writing—reviewing and editing.

Acknowledgments

This work was supported by the Department of Biomedical Engineering (BME), Islamic University, Kushtia 7003, Bangladesh.

References

- [1] S. Dutta, B. C. S. Manideep, M. Basha, R. D. Caytiles, and N. C. S. N. Iyengar, "Classification of diabetic retinopathy images by using deep learning models," *International Journal of Grid and Distributed Computing*, vol. 11, no. 1, pp. 99–106, 2018.
- [2] L. Math and R. Fatima, *Adaptive Machine Learning Classification for Diabetic Retinopathy*, 2020.
- [3] V. V. Vijayan and C. Anjali, "Prediction and diagnosis of diabetes mellitus—A machine learning approach," in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, pp. 122–127, Trivandrum, India, 2015.
- [4] P. M. Shakeel, S. Baskar, V. R. S. Dhulipala, and M. M. Jaber, "Cloud based framework for diagnosis of diabetes mellitus using K - means clustering," *Health information science and systems*, vol. 6, no. 1, pp. 16–17, 2018.
- [5] N. G. Forouhi and N. J. Wareham, "Epidemiology of diabetes," *Medicine (Baltimore)*, vol. 38, no. 11, pp. 602–606, 2010.
- [6] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes," *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004.
- [7] P. S. Kumar, R. U. Deepak, A. Sathar, V. Sahasranamam, and R. R. Kumar, "Automated detection system for diabetic retinopathy using two field fundus photography," *Procedia computer science*, vol. 93, pp. 486–494, 2016.
- [8] Y. Wang, C. Wang, A. Sensoy, S. Yao, and F. Cheng, "Can investors' informed trading predict cryptocurrency returns? Evidence from machine learning," *Research in International Business and Finance*, vol. 62, article 101683, 2022.
- [9] M. M. Ahsan, T. E. Alam, T. Trafalis, and P. Huebner, "Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients," *Symmetry*, vol. 12, no. 9, p. 1526, 2020.
- [10] B. B. Gupta, A. Gaurav, E. C. Marin, and W. Alhalabi, "Novel graph-based machine learning technique to secure smart vehicles in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2022.
- [11] A. A. Das and D. S. Duranta, "Alzheimer ' S Disease Detection Using M-Random Forest Algorithm with Optimum Features Extraction," in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, Riyadh, Saudi Arabia, 2021.
- [12] S. Ali, S. Miah, J. Haque, M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, article 100036, 2021.
- [13] J. Mistry and B. Inden, "An approach to sign language translation using the intel realsense camera," in *2018 10th Computer Science and Electronic Engineering (CEECE)*, Colchester, UK, 2019.
- [14] G. T. Reddy, M. P. K. Reddy, K. Lakshmana et al., "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020.

- [15] K. Islam, S. Ali, S. Miah, M. Rahman, M. S. Alam, and M. A. Hossain, "Brain tumor detection in MR image using superpixels, principal component analysis and template based K-means clustering algorithm," *Machine Learning with Applications*, vol. 5, article 100044, 2021.
- [16] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia computer science*, vol. 132, pp. 1578–1585, 2018.
- [17] M. Alehegn, "Analysis and prediction of diabetes mellitus using machine learning algorithm," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 9, pp. 871–878, 2018.
- [18] F. Faruque, "Performance analysis of machine learning techniques to predict diabetes mellitus," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–4, Cox'sBazar, Bangladesh, 2019.
- [19] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance analysis of data mining classification techniques to predict diabetes," *Procedia Computer Science*, vol. 82, pp. 115–121, 2016.
- [20] N. H. Barakat, A. P. Bradley, S. Member, and M. N. H. Barakat, "Intelligible support vector machines for diagnosis of diabetes mellitus," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 4, pp. 1114–1120, 2010.
- [21] B. L. Shivakumar, "A Survey on Data-Mining Technologies for Prediction and Diagnosis of Diabetes," in *2014 International Conference on Intelligent Computing Applications*, Coimbatore, India, 2014.
- [22] A. Choudhury and D. Gupta, "A survey on medical diagnosis of diabetes using machine learning," in *Recent Developments in Machine Learning and Data Analytics: IC3 2018*, Springer Singapore, 2019.
- [23] K. Sumangali, "A Classifier Based Approach for Early Detection of Diabetes Mellitus," in *2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pp. 389–392, Kumaracoil, India, 2016.
- [24] C. L. Chowdhary, S. Bhattacharya, S. Hakak, and R. Kaluri, "An ensemble based machine learning model for diabetic retinopathy classification," in *2020 international conference on emerging trends in information technology and engineering (ic-ETITE)*, pp. 1–6, Vellore, India, 2020.
- [25] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in genetics*, vol. 9, p. 155, 2018.
- [26] J. Rahman, B. Ahammed, and M. Abedin, "Classification and prediction of diabetes disease using machine learning paradigm," *Health information science and systems*, vol. 8, no. 1, pp. 1–14, 2020.
- [27] K. Saxena, Z. Khan, and S. Singh, "Diagnosis of diabetes mellitus using K nearest neighbor algorithm," *International Journal of Computer Science Trends and Technology (IJCSST)*, vol. 2, no. 4, pp. 36–43, 2014.
- [28] A. S. Rani and S. Jyothi, "Performance analysis of classification algorithms under different datasets," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2016April 2023, <https://ieeexplore.ieee.org/abstract/document/7724534>.
- [29] N. Nai-arun and R. Mounngmai, "Comparison of classifiers for the risk of diabetes prediction," *Procedia Computer Science*, vol. 69, pp. 132–142, 2015.
- [30] M. M. Ahsan, M. R. Uddin, M. S. Ali et al., "Deep transfer learning approaches for monkeypox disease diagnosis," *Expert Systems with Applications*, vol. 216, article 119483, 2023.
- [31] Kaggle, "Pima Indians diabetes database," 2023, April 2023, <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>.
- [32] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, article 105524, 2020.
- [33] A. A. AlJarullah, "Decision tree discovery for the diagnosis of type II diabetes," in *2011 International conference on innovations in information technology*, pp. 303–307, Abu Dhabi, United Arab Emirates, 2011.
- [34] I. Hasan, S. Ali, H. Rahman, and K. Islam, "Automated detection and characterization of colon Cancer with deep convolutional neural networks," *Journal of Healthcare Engineering*, vol. 2022, Article ID 5269913, 12 pages, 2022.
- [35] K. Islam, S. Ali, A. A. Das, D. U. S. Duranta, and M. S. Alam, "Human brain tumor detection using k-means segmentation and improved support vector machine," *International Journal of Scientific Engineering Research*, vol. 11, 2020.
- [36] G. Dougherty, "Classification," in *Pattern recognition and classification: an introduction*, pp. 9–26, Springer Science & Business Media, 2013.
- [37] "sklearn.preprocessing.MaxAbsScaler — scikit-learn 1.2.2 documentation," April 2023, <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MaxAbsScaler.html>.
- [38] I. M. Pires, F. Hussain, N. M. Garcia, and P. Lameski, "Homogeneous data normalization and deep learning: a case study in human activity classification," *Future Internet*, vol. 12, no. 11, p. 194, 2020.
- [39] H. Paulheim and R. Meusel, "A decomposition of the outlier detection problem into a set of supervised learning problems," *Machine Learning*, vol. 100, no. 2-3, pp. 509–531, 2015.
- [40] L. Tian, Y. Fan, L. Li, and N. Mousseau, "Identifying flow defects in amorphous alloys using machine learning outlier detection methods," *Scripta Materialia*, vol. 186, pp. 185–189, 2020.
- [41] K. Singh and M. Cantt, "Outlier detection: applications and techniques," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 1, pp. 307–323, 2012.
- [42] G. K. Vishwakarma, C. Paul, and A. M. Elsawah, "An algorithm for outlier detection in a time series model using backpropagation neural network," *Journal of King Saud University-Science*, vol. 32, no. 8, pp. 3328–3336, 2020.
- [43] W. P. Zijlstra, L. A. Van Der Ark, and K. Sijtsma, "Outlier detection in test and questionnaire data," *Multivariate Behavioral Research*, vol. 42, no. 3, pp. 531–555, 2007.
- [44] J. Heaton, "An empirical analysis of feature engineering for predictive modeling," in *SoutheastCon 2016*, Norfolk, VA, USA, 2016.
- [45] M. F. Uddin, J. Lee, S. Rizvi, and S. Hamada, "proposing enhanced feature engineering and a selection model for machine learning processes," *Applied Sciences*, vol. 8, no. 4, p. 646, 2018.
- [46] G. Dong and H. Liu, *Feature Engineering for Machine Learning and Data Analytics*, CRC Press, 2020.
- [47] P. Rodríguez, M. A. Bautista, J. González, and S. Escalera, "Beyond one-hot encoding: lower dimensional target embedding," *Image and Vision Computing*, vol. 75, pp. 21–31, 2018.

- [48] J. Li, Y. Si, T. Xu, and S. Jiang, "Deep convolutional neural network based ECG Classification system using information fusion and one-hot encoding techniques," *Mathematical problems in engineering*, vol. 2018, Article ID 7354081, 10 pages, 2018.
- [49] J. Hua, W. D. Tembe, and E. R. Dougherty, "Performance of feature-selection methods in the classification of high- dimension data," *Pattern Recognition*, vol. 42, no. 3, pp. 409–424, 2009.
- [50] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [51] J. R. Vergara and P. A. Este, "A Review of Feature Selection Methods Based on Mutual Information," *Neural computing and applications*, vol. 24, no. 1, pp. 175–186, 2014.
- [52] F. Herrera, "A review of microarray datasets and applied feature selection methods," *Information sciences*, vol. 282, pp. 111–135, 2014.
- [53] "Correlation coefficient: simple definition, formula, easy calculation steps," April 2023, <https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/>.
- [54] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 science and information conference*, pp. 372–378, London, UK, 2014.
- [55] E. L. Hall, R. P. Kruger, S. J. Dwyer, D. L. Hall, R. W. McLaren, and G. S. Lodwick, "A survey of preprocessing and feature extraction techniques for radiographic images," *IEEE Transactions on Computers*, vol. C-20, no. 9, pp. 1032–1044, 1971.
- [56] W. Islam, G. Danala, H. Pham, and B. Zheng, "Improving the performance of computer-aided classification of breast lesions using a new feature fusion method," *Medical Imaging 2022: Computer-Aided Diagnosis*, vol. 12033, no. 4, pp. 98–105, 2022.
- [57] M. K. Islam, M. S. Ali, M. M. Ali et al., "Melanoma skin lesions classification using deep convolutional neural network with transfer learning," in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, pp. 48–53, Riyadh, Saudi Arabia, 2021.
- [58] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and naive Bayes," *The Journal of Supercomputing*, vol. 77, no. 5, pp. 5198–5219, 2021.
- [59] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021.

Research Article

HDFCN: A Robust Hybrid Deep Network Based on Feature Concatenation for Cervical Cancer Diagnosis on WSI Pap Smear Slides

Nitin Kumar Chauhan ^{1,2}, Krishna Singh ³, Amit Kumar ^{2,4}
and Swapnil Baburav Kolambakar ⁵

¹USIC&T, Guru Gobind Singh Indraprastha University, New Delhi 110078, India

²Department of ECE, Indore Institute of Science & Technology, Indore 453331, India

³DSEU Okhla Campus-I, Formerly G. B. Pant Engineering College, New Delhi 110020, India

⁴Department of Electronics Engineering, Indian Institute of Technology (BHU), Varanasi 221005, India

⁵Bluecrest university college, Monrovia, Liberia

Correspondence should be addressed to Swapnil Baburav Kolambakar; k.swapnil@bluecrest.edu.lr

Received 1 November 2022; Revised 6 January 2023; Accepted 18 March 2023; Published 17 April 2023

Academic Editor: Sami Azam

Copyright © 2023 Nitin Kumar Chauhan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cervical cancer is a critical imperilment to a female's health due to its malignancy and fatality rate. The disease can be thoroughly cured by locating and treating the infected tissues in the preliminary phase. The traditional practice for screening cervical cancer is the examination of cervix tissues using the Papanicolaou (Pap) test. Manual inspection of pap smears involves false-negative outcomes due to human error even in the presence of the infected sample. Automated computer vision diagnosis revamps this obstacle and plays a substantial role in screening abnormal tissues affected due to cervical cancer. Here, in this paper, we propose a hybrid deep feature concatenated network (HDFCN) following two-step data augmentation to detect cervical cancer for binary and multiclass classification on the Pap smear images. This network carries out the classification of malignant samples for whole slide images (WSI) of the openly accessible SIPaKMeD database by utilizing the concatenation of features extracted from the fine-tuning of the deep learning (DL) models, namely, VGG-16, ResNet-152, and DenseNet-169, pretrained on the ImageNet dataset. The performance outcomes of the proposed model are compared with the individual performances of the aforementioned DL networks using transfer learning (TL). Our proposed model achieved an accuracy of 97.45% and 99.29% for 5-class and 2-class classifications, respectively. Additionally, the experiment is performed to classify liquid-based cytology (LBC) WSI data containing pap smear images.

1. Introduction

Cancer is coerced by the obsolete and irregular evolution of cells in the human body. This deformity can infiltrate the nearby cells in that tissue together with the other tissues and may disperse into more body organs. Cervical cancer arises due to the contagion of the human papillomavirus (HPV) which causes an anomaly in the cervix through which the lower portion of the uterus and vagina connect. Cervical cancer was the fourth most genre of cancer in statistics of new indices and fatalities following breast, colorectum,

and lung cancer in 2020 [1]. The scarcity of screening and therapeutic systems consequences a high mortality rate in low and middle-income economies. The preliminary traits of cervical cancer comprise an erratic feminine cycle, postintercourse vagary bleeding, strong vaginal stink with discharge, inexplicable and relentless pelvic, intestinal, or back agony, exhaustion, and diminution in weight [2, 3].

An adequate diagnosis can be procured by using the potential preliminary investigation of cervical lesions for the contraction of the mortality rate by cause of cervical cancer. The most trusted and well-known approach for the

detection of precancerous cells and cervical lesions is a Pap smear test [4]. The conventional and prevailing method for scrutinizing anomalies of peculiar cells of pap slides using a microscope by clinical experts is a quite sophisticated, tedious, and slow proceeding that requires decent knowledge and familiarization.

Artificial intelligence (AI) in computer vision is in growing vogue amid the researchers to go beyond the bounds of manual screening of pap smear data. Even though conventional machine learning (ML) models have reduced computing intricacy, they still require incorporating the extraction of features manually [5]. DL comes out as one of the fine automated end-to-end solutions to the multitudinous challenges in biomedical image processing [6–10]. To produce a decision support system with high efficacy and potential, a deep model requires plenty of medical image data for training. TL addresses this issue, by pretraining the model over a large amount of data followed by utilizing it with a limited image dataset of the specific problem [11].

This paper proposes a novel approach that utilizes concatenated features for the classification of pap smear WSI images. Here, we used TL for the automatic retrieval of key features using fine-tuned deep models. Three fine-tuned DL models VGG-16, ResNet-152, and DenseNet-169 are used for feature extraction. These models use pretrained weights on the ImageNet dataset. The features retrieved from these individual models are concatenated and used for the prediction of test samples. A fully connected network (FCN) is used for classification by training it on concatenated features. The structural outline of the proposed framework is shown in Figure 1.

Commonly, the cells are first segmented from the WSI slides, and then the segmented cell images are used for the prediction of cancer. However, cell segmentation is a complex process in itself, and no algorithm performs equally well for cell segmentation from test slides obtained in distinct medical procedures and test conditions. Here, in this paper, we propose a model to analyze the multicell cervix images to detect the various grades of abnormality. The detection of abnormal samples from this proceeding enables other downstream investigations by segmentation of cells from abnormal WSI images.

The contribution of this research work is listed below:

- (1) A brief insight into current research in cervical cancer diagnosis, the use of intensive two-step augmentation, and the proposed HDFCN model to detect WSI pap-smear cervical images
- (2) The proposed method uses optimally extracted deep features to get concatenated into hybrid features that are used with FCN for test image classification
- (3) This proposed model gets validated on two WSI cervical cancer data—(1) SIPaKMeD data for 5-class and 2-class classifications and (2) LBC data classification
- (4) McNemar’s statistical tool indicating p value is used to illustrate the demarcation of efficacy among the

proposed classifier and the individual deep classifier utilized in the proposed hybrid model.

The remainder of the paper is articulated as follows: Section 2 contains the recent work related to cervical cancer diagnosis using different DL models. Section 3 describes a detailed description of the proposed method and material utilized. Section 4 depicts the result of the proposed model and discussion on various aspects of its performance. Lastly, Section 5 has some brief conclusions and the futuristic possibilities of the work.

2. Related Work

With the advancement of computer-aided diagnostic (CAD) tools, the use of DL has been motivated in the classification of cervix malignancies. Papers [12–16] give a detailed survey of the latest DL applications in cervigram histopathological image classification. The most renowned publicly available databases of cervical cancer are the Herlev [17], ISBI [18], Risk factor-based dataset [19], Cervix93 [20], SIPaKMeD [21], LBC dataset [22], etc.

In [23], 502 pathological images were collected at Xinjiang Medical University which were used to generate two groups of image datasets. The first group of 3012 images was generated by resizing and cropping, and the other group contains 108,432 images obtained using augmentation functions like rotation, flipping, and image enhancement on the original set of image data. A convolutional neural network (CNN) was used to perform 3-class classification with 93.33% and 89.48% accuracy for the first and second groups of images, respectively. In another work [24], a shallow CNN classifier was used for the classification of 684 positive and negative area handcrafted patches of 15×15 pixels retrieved from cervigram slides obtained by VIA of 102 women patients.

In this study [25], binary classification was performed using a stacked autoencoder followed by a softmax layer on the UCI dataset with 668 samples. The stack autoencoder and softmax layer were used for dimension reduction of raw data and classification, respectively, with 97.8% accuracy. In [26], TL used fine-tuning of the Inception-V3 model for the classification of 307 histopathology images collected by AQP, HIF, and VEGF staining at Shengjing Hospital of China Medical University. Original data got enhanced by 256 multiples, resulting in a total of 78,592 images using augmentation operations—rotation and flipping, provided a mean accuracy of 77.3% with the classifier model.

The authors [27] used four CNN classifiers, namely, AlexNet, DenseNet-121, ResNet-50, and GoogLeNet, to classify the Herlev dataset by combining the morphology and appearance-based features to provide topmost accuracy of 94.5%, 71.3%, and 64.5% for 2 class, 4 class, and 7 class, respectively, with the GoogLeNet classifier. In research [28], an approach using Mask R-CNN was used for the segmentation and classification of the Herlev database, using ResNet-10 as a key pillar. During segmentation, the proposed network was pretrained on the COCO dataset and provided a precision of 0.92 ± 0.06 , recall of 0.91 ± 0.05 , and ZSI of 0.91 ± 0.04 . The VGG-like model was used for

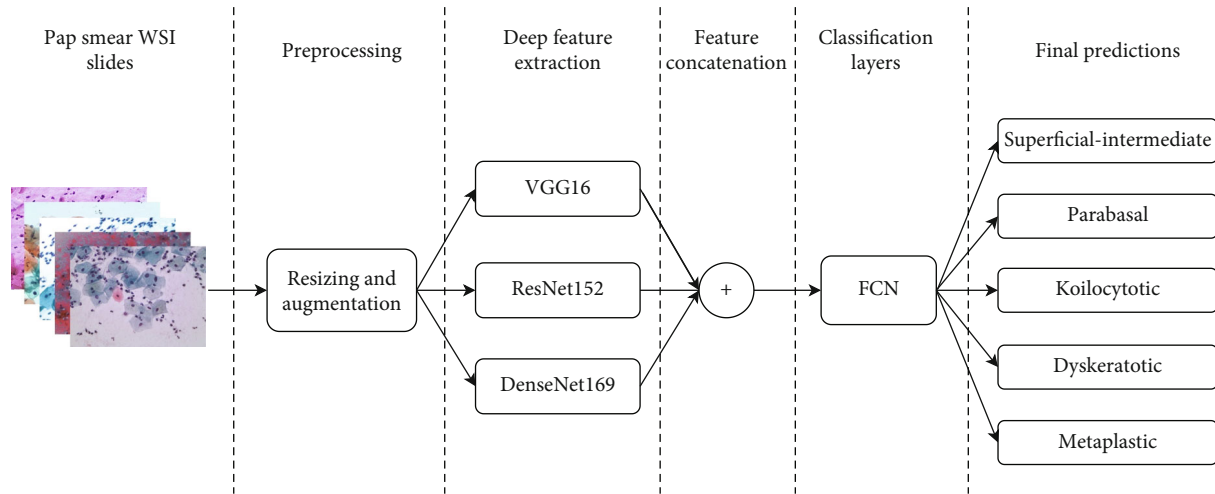


FIGURE 1: Structural outline of the proposed framework.

TABLE 1: Pipeline of augmentation functions.

| S.no. | Aug_function | Augmentations |
|-------|-----------------|---|
| 1 | Aug | Rotate (-45, 45), scale {"x": (0.8, 1.2), "y": (0.8, 1.2)}, translation {"x": (-0.15, 0.15), "y": (-0.15, 0.15)}, shear (-2, 2), h_flip (1.0), v_flip (1.0) |
| 2 | Clahc | clip_limit (1, 10), tile_grid_size (3, 21), gamma_contrast (0.5, 2.0), channel_clahc |
| 3 | Edge | edge_detect (alpha = (0, 0.5), directed_edge_detect (alpha = (0, 0.5), direction = (0.0, 1.0)) |
| 4 | Sharp | Sharpen (alpha = (0.2, 0.8), lightness = (0.75, 1.5)) |
| 5 | Canny | Canny (alpha = (0.5, 0.8), sobel_kernel_size = (3, 7)) |
| 6 | Red | red_channel (add ((10, 100), rotate = (0, 45)) |
| 7 | Green | green_channel (add ((10, 100), rotate = (0, 45)) |
| 8 | Blue | blue_channel (add ((10, 100), rotate = (0, 45)) |
| 9 | Noise | Blur (sigma = (0, 1.22), gauss_noise (scale = 0.111*255), laplace_noise (scale = (0, 0.111*255)) |
| 10 | Color | channel_shuffle (1.0), grayscale (1.0), hue_n_saturation (0.5, 1.5), add_hue_saturation (-50, 50), kmeans_color (n.colors = (4, 16) |
| 11 | Flip | histogram_equalization, v_flip (1.0), h_flip (1.0) |
| 12 | contrast_n_shit | Contrast (LinearContrast (0.75, 1.5)), brightness (0.35, 1.65), brightness_channel ((0.5, 1.5), per_channel = 0.75) |

the classification of segmented cells with an accuracy of 98.1% for 2-class and 95.9% for 7 class. The work [29] introduced an AI classifier by modification of layers in the ResNet-50 model which classifies the colposcopy database of 310 images having 213 HSIL and 97 LSIL cases with an accuracy of 82.3%.

The experiment [30] proposed a pipeline of CNN-based feature extraction and classification networks founded on the identification of the cervix region of interest (ROI) and trained on two datasets: Intel & MobileODT Dataset and NCI Guanacaste Project Dataset. The developed model has a lightweight and faster framework and is found to be quite useful in mobile application development for low-income nations for the diagnosis of cervigram images. The work [31] presented ensemble learning and a CNN model and achieved an accuracy of 90.4% and 91.6%, respectively, for the two classifiers on the Herlev dataset with preprocessing methods. In [32], hybrid TL was implemented with pre-

TABLE 2: Structure of augmented SIPaKMeD data.

| Class | Train | Validation | Test |
|-------|-------|------------|------|
| SI | 975 | 25 | 26 |
| P | 832 | 22 | 22 |
| K | 1846 | 48 | 48 |
| M | 2106 | 54 | 55 |
| D | 1729 | 45 | 45 |
| Total | 7488 | 194 | 196 |

trained AlexNet and VGG-16 models on 1644 cervix cell images collected at the National Institute of Health (NIH) and National Cancer Institute (NCI). This analysis provided a result of 91.46% accuracy for 2-class classification.

In the study [33], a deep network ColpoNet was introduced to detect colposcopy images collected at the National Cancer Institute (NCI) and achieved an accuracy of 81.35%,

TABLE 3: Structure of augmented LBC data.

| Class | Train | Validation | Test |
|-------|-------|------------|------|
| NILM | 4758 | 123 | 123 |
| LSIL | 871 | 23 | 23 |
| HSIL | 1261 | 33 | 33 |
| SCC | 572 | 15 | 15 |
| Total | 7462 | 194 | 194 |

providing better efficiency than GoogleNet, AlexNet, ResNet-50, VGG-16, and LeNet. In the research [34], a deep framework containing seven networks, namely, ResNet-50, 5 incidents of ResNet-101, and a graph convolution network with edge traits (E-GCN), was introduced. The initial six networks were used for feature interpretation, and GCN classified the cervix images based on fused node and edge features. The proposed model evaluated a database with 7,688 colposcopic images and obtained an accuracy of 78.33% for positive and negative instance classification.

The work [35] implemented two classifiers TL-based VGG-19 and an ensemble model on colposcopy (CYENET) which were used to classify the colposcopy image data with 5,679 images collected by Intel & MobileODT and available at Kaggle. The proposed CYENET classifier gave a result of 92.3% with a 19% improvement in classification accuracy to VGG-19 (TL). In another study [36], the ResNet-50 model-based deep CNN (DCNN) model is used to perform classification besides three ML classifiers XGB, SVM, and RF, on cervicography data comprising 4,119 images with positive and negative incidences. Linear regression was applied to select 10 attributes out of 300; those were fed to the ResNet-50 classifier which performed better than others with accuracy and AUC of about 90.65% and 0.97, respectively.

In this experiment [37], CytoBrain was introduced which comprises a compact VGG network to identify the cervical cell from WSI slides. This model showed faster and more precision with an accuracy of about 88.3% on huge data containing 198,952 cervix cell images. In [38], a deep residual network was developed that assesses the performance with distinct activation functions like ReLU, PReLU, and Leaky-ReLU with accuracies of 98.3%, 100%, and 99.2%, respectively. In the work [39], a deep CNN-based capsule module (CNN-CapsNet) was implemented by utilizing a few residual blocks which classified 8-stages in magnetic resonance (MR) images of the TCGA-CESC database with an accuracy of 90.28%.

The study [40] deployed a TL-based exemplar pyramid model utilizing DarkNet19 or DarkNet53 for the retrieval of 21,000 attributes in which 1000 utmost-informative were selected by neighborhood component analysis (NCA). SVM used these elected attributes for the classification of the LBC and SIPaKMeD databases and provided accuracies of 99.47% and 98.26%, respectively. In research [41], a novel deep architecture, HLDnet, was evolved that utilizes a faster RCNN network to detect HSIL+ cervigram images by dual-channel detection (acetic acid and Lugol's iodine cervigram). It achieved an accuracy of 0.86 for 400 training and validation and 200 tests, better than single-channel detection

(either acetic acid or Lugol's iodine cervigram). There is more related work available in clinical practices [42–46] thanks to advances in computer vision methods. Some of these are observed as competent in performing the same or even better as the pathologists on medical data.

3. Material and Methods

3.1. Experimental Data. This work utilizes a publicly accessible SIPaKMeD database containing 966 WSI pap smear images and 4,049 images of handcrafted cropped cells [21]. An optical magnifying device (OLYMPUS BX53F) with a camera having a charge-coupled device (CCD) sensor (Lumenera's INFINITY-1) has been used to capture these pictures. The dataset is categorized into 5 classes by clinical professionals. The classes "superficial-intermediate (SI)" and "parabasal (P)" refer to "normal," images sorted as "koilocytotic (K)" and "dyskeratotic (D)" indicate "abnormal," and the remaining "metaplastic (M)" belongs to have "benign" cells. The experiment is performed on WSI slides and grouped into 5 class and 2 class (normal and abnormal).

Furthermore, the proposed framework is evaluated using liquid-based cytology (LBC) data available online at Mendeley data [22]. Based on the Bethesda system, the collection includes 963 WSI LBC high-resolution images organized into four sets of classes: "no squamous intraepithelial lesion (NILM)," "low-grade squamous intraepithelial lesion (LSIL)," "high-grade squamous intraepithelial lesion (HSIL)," and "squamous cell carcinoma (SCC)." The "NILM" indicates a "normal" grade, while the "LSIL," "HSIL," and "SCC" refer to "abnormal."

3.2. Preprocessing

3.2.1. Resizing and Division. The cervigram WSI images of both databases are of high resolution with 2048×1536 pixels. The images are resized to 224×224 pixels to reduce computation costs and make them fit into DL models. The resized data is get divided into a train, validation, and test data in the ratio of 3 : 1 : 1.

3.2.2. Data Augmentation. DL models require a sufficient amount of data to train them efficaciously. Two-step data augmentation is utilized on the resized training data to increase the amount of data to be learned. Firstly, training data is augmented using a heavy augmentation pipeline consisting of numerous augmentation strategies such as affine transformations, perspective transformations, contrast changes, Gaussian noise, dropout of regions, hue/saturation changes, cropping/padding, and blurring. This augmentation pipeline, shown in Table 1, has 12 sets of augmentation functions that generated 12 augmented images for one training sample. A detailed description of these augmentations can be found in the ImgAug library [47].

Now, training data becomes 13 times more multiple than before. The structure of train, validation, and test data after resizing and augmenting SIPaKMeD and LBC data is given in Tables 2 and 3, respectively. Further, real-time data augmentation is performed using the "ImageDataGenerator" function of the Keras library that performs a random transformation on

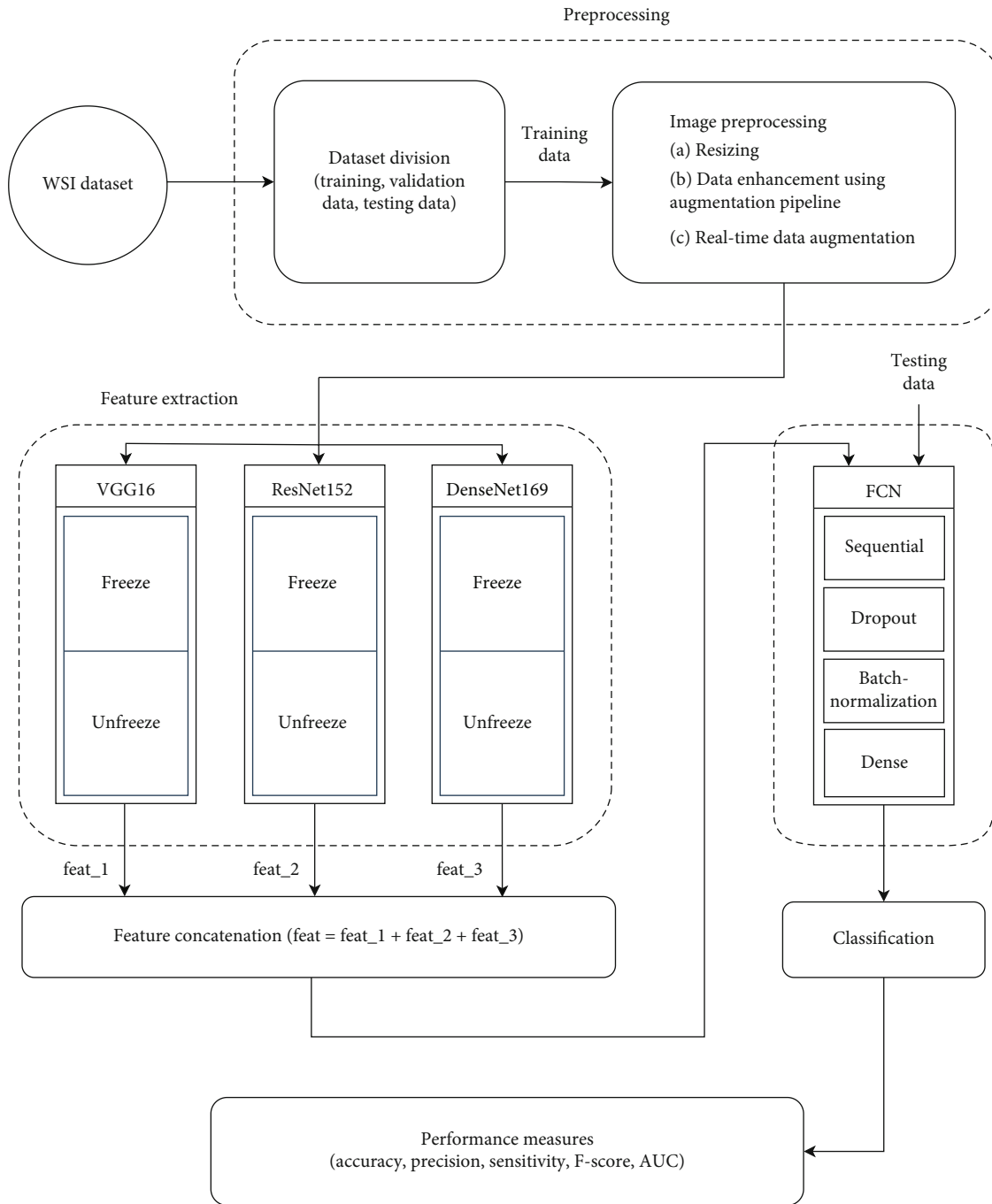


FIGURE 2: The overall architecture of the proposed method.

images so the model can be trained on distinct images on each epoch. The different arguments of the function are set to be: `featurewise_center = false`, `rotation_range = 25`, `fill_mode = nearest`, `zoom_range = 0.2`, `width_shift_range = 0.1`, `height_shift_range = 0.1`, `horizontal_flip = true`, `vertical_flip = true`, `brightness_range = (0.5, 1.5)`, and `channel_shift_range = 20`.

3.3. Methods

3.3.1. *Feature Extraction Using Transfer Learning.* The overall architecture of the proposed method is shown in Figure 2.

The preprocessed training data is being used for feature extraction using the fine-tuning of DL models. The DL models are certain subtypes of ML structures having more complex architecture based on a neural network with lesser human intercede [48]. These models are extensively utilized in extricating high-level features, providing progressive execution over the conventional approach, and expanding interpretability conjointly with the understanding and handling of biological information. Commonly used DL frameworks are CNNs, recurrent neural networks (RNNs), and recursive neural networks (RvNNs). Among these, CNN is

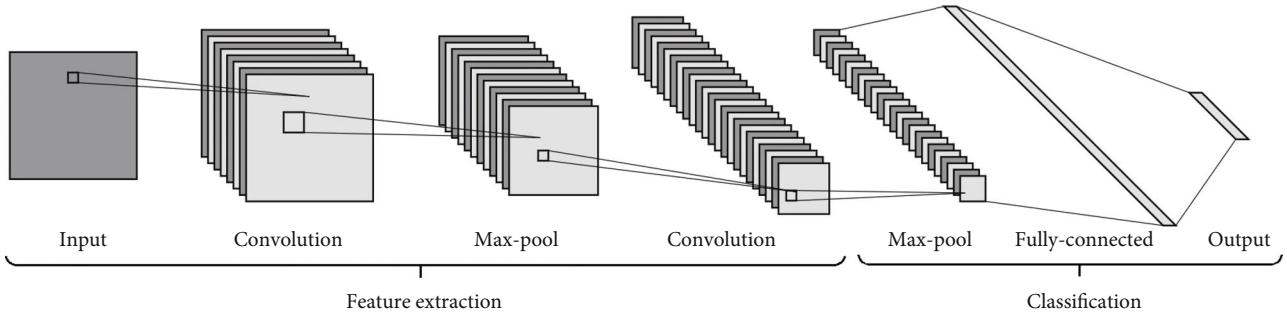


FIGURE 3: CNN model.

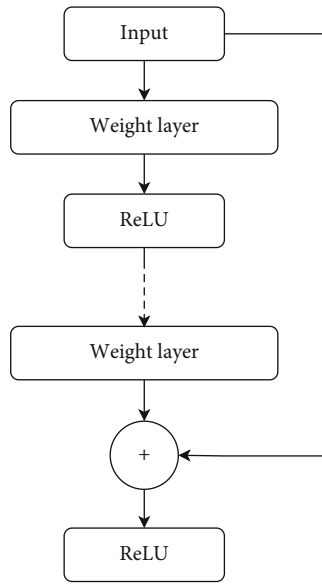


FIGURE 4: The residual block structure of ResNet.

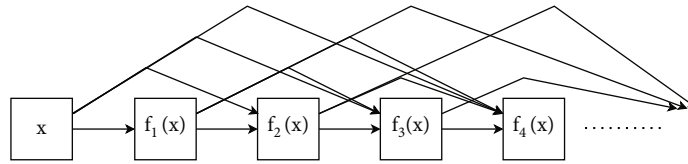


FIGURE 5: The densely connected structure of DenseNet.

the most prominent structure of DL having interconnected networks of neurons drafted in a combination of convolutional layers, pooling layers, and fully connected (FC) layers. Feature learning and classification are the prime utilization of the CNN model, as shown in Figure 3 [49].

The convolution layers extract a variety of visual features, including edges, objects, and textures by performing convolution of input and the kernel filter. The preactivation output z^l of the convolution layer is defined as

$$z^l = v^{l-1} * W^l \quad (1)$$

where v^{l-1} is the activation output of the previous layer, $*$ is the convolutional operator, and W is the weights. Subsequently, pooling layers include various operations like global

average pooling, L-2 normalization, and max pooling to retain relevant features from the convolution layer output. Some of the pooling operators are given

$$\begin{aligned} \text{average - pool } v_{xy}^l &= \frac{1}{s^2} \sum_{i,j} h_{(x+i)(y+j)}^{l-1} \\ \text{max - pool } v_{xy}^l &= \max_{i,j} h_{(x+i)(y+j)}^{l-1} \end{aligned} \quad (2)$$

The multidimensional feature map retrieved from these layers is used to convert it into a vector at the classification stage using the fully connected network (FCN). Here, in this work, three prevalent CNN-based DL methods VGG-16, ResNet-152, and DenseNet-169 are utilized for this purpose.

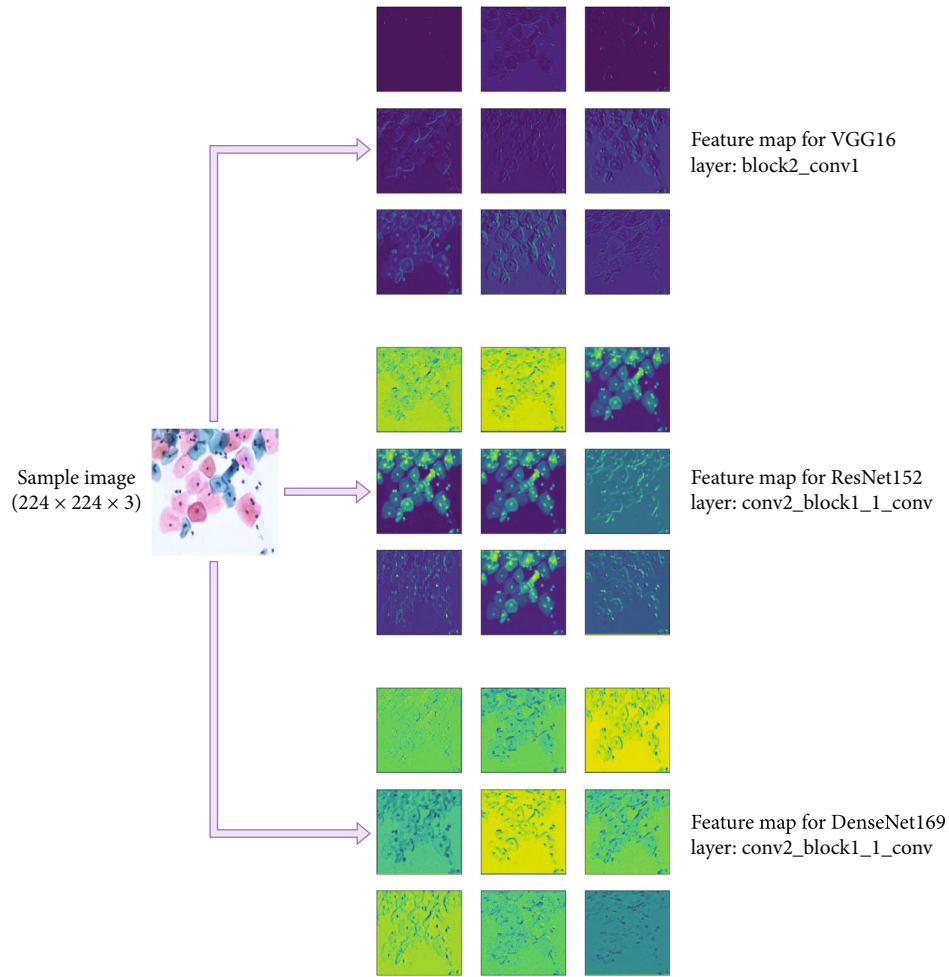


FIGURE 6: Visualization of feature maps for VGG-16, ResNet-152, and DenseNet-169.

VGG-16. The Visual Geometry Group (VGG) was introduced with the concept of inserting a pile of small filters of size 3×3 instead of filters with sizes 5×5 and 11×11 in the preceding deep networks. Additionally, the convolution filter of size 1×1 with rectified linear unit (ReLU) activation function is inserted in between convolutional layers for complexity regulation and linear transformation. VGG-16 holds 16 layers, comprising convolution layers of kernel size 3×3 with padding of 1, and max pool layers of size 2×2 with a stride of 2, followed by three FC layers. This model has an input size of $224 \times 224 \times 3$ with approx. 138 million computation parameters [50].

ResNet-152. The residual network (ResNet) came with a more deep network by skipping connections to resolve the vanishing gradient issue. The substitute route connection made by the residual structure shown in Figure 4 permits the gradient to flow without fading, resulting in enhanced performance of the network. The residual behavior of ResNet for the l^{th} stage with x input and output H of the activation function can be formulated as

$$x_l = H_l(x_{l-1}) + x_{l-1}. \quad (3)$$

The ResNet-152 model has 152 layers, more than 8 times deeper than VGG networks, nevertheless having a lesser computational complexity with approximately 60.4 million parameters [51].

DenseNet-169. This method is also used to resolve the issue of the vanishing gradient followed by the ResNet model. The contrast between ResNet and DenseNet is that ResNet utilizes an additive approach to associate all preceding feature maps, whereas DenseNet concatenates all the previous layers as shown in Figure 5. The value of variable x in DenseNet can be a map with the progressive complex group of functions as

$$x \longrightarrow [x, f_1(x), f_2([x, f_1(x)]) + f_3([x, f_1(x), f_2([x, f_1(x))]) \dots \dots]. \quad (4)$$

The DenseNet-169 model comprises 169 layers, which is larger than the other members of the DenseNet group and has less computational complexity than the VGG and ResNet model with approximately 14.3 million parameters [52]. Figure 6 gives a visualization of feature maps for a sample image of SIPaKMeD data through a specific layer of the aforementioned DL models.

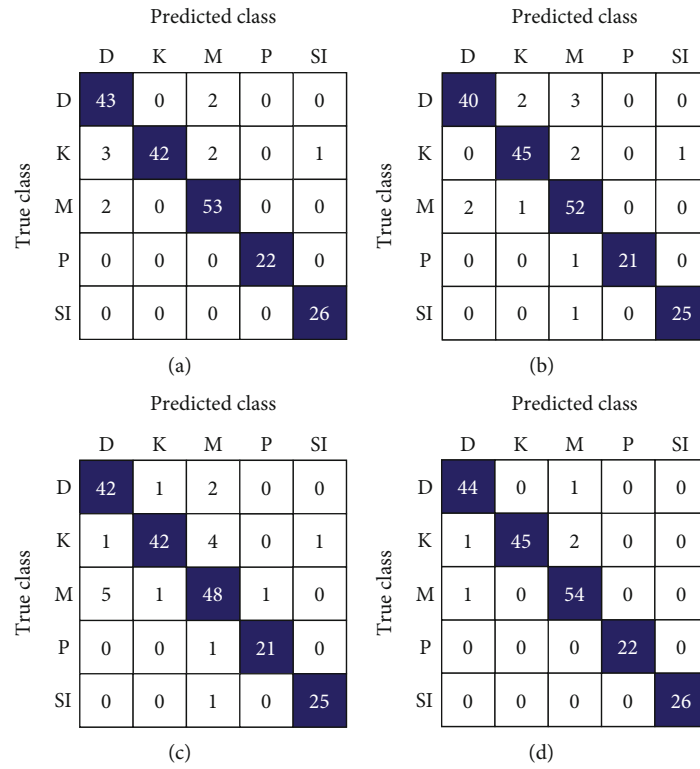


FIGURE 7: Confusion matrix for SIPaKMeD WSI 5-class classification using (a) VGG-16, (b) ResNet-152, (c) DenseNet-169, and (d) proposed model.

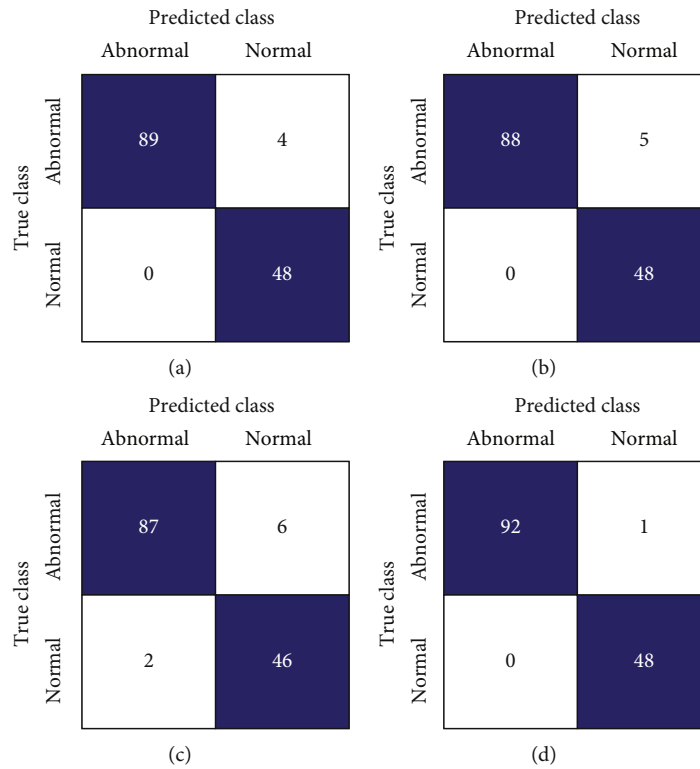


FIGURE 8: Confusion matrix for SIPaKMeD WSI 2-class classification using (a) VGG-16, (b) ResNet-152, (c) DenseNet-169, and (d) proposed model.

TABLE 4: Performance metrics for different fine-tuned classifiers and proposed model on SIPaKMeD WSI data.

| Data | Models | Accuracy (%) | Precision (%) | Recall (%) | F-score (%) |
|----------------------|----------------|--------------|---------------|--------------|--------------|
| SIPaKMeD WSI 5-class | VGG-16 | 94.89 | 95.88 | 95.77 | 95.83 |
| | ResNet-152 | 93.37 | 93.76 | 94.66 | 94.21 |
| | DenseNet-169 | 90.82 | 91.94 | 92.06 | 92 |
| | Proposed model | 97.45 | 97.94 | 98.08 | 98.01 |
| SIPaKMeD WSI 2-class | VGG-16 | 97.16 | 95.7 | 100 | 97.8 |
| | ResNet-152 | 96.45 | 94.62 | 100 | 97.24 |
| | DenseNet-169 | 94.33 | 93.55 | 97.75 | 95.6 |
| | Proposed model | 99.29 | 98.92 | 100 | 99.46 |

Best results are shown in bold.

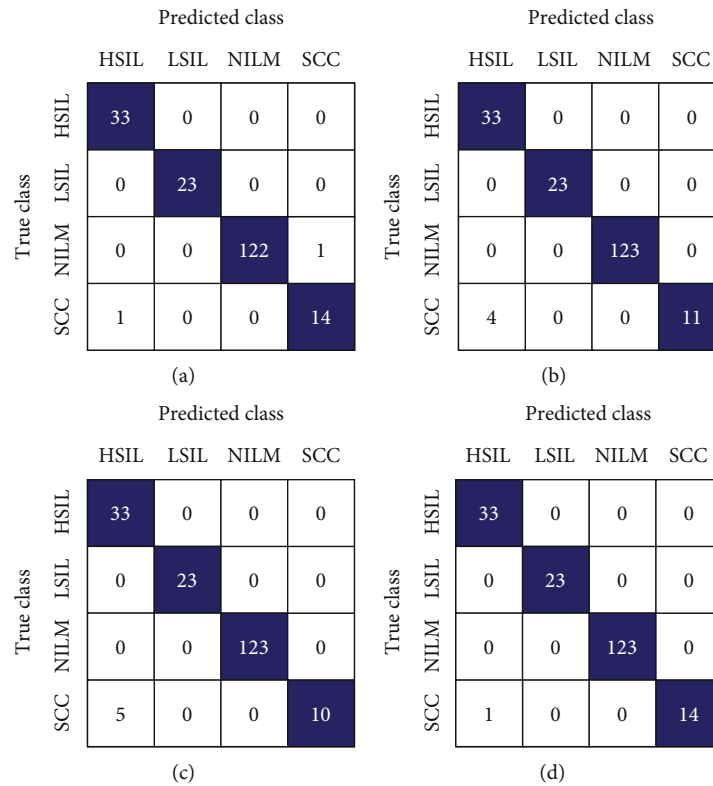


FIGURE 9: Confusion matrix for LBC WSI data classification using (a) VGG-16, (b) ResNet-152, (c) DenseNet-169, and (d) proposed model.

Performing image classification utilizing the DL approach requires plenty of computations that need the support of large data and long-running time of graphic processing unit (GPU) amid training of the model, which results in outlays of considerable computational resources and less productivity. TL accelerates the training process by utilizing the pre-trained model on a contemporary problem, which may result in more efficacy and overall accuracy. TL is substantially used as a design method in ML models to train with the small dataset by utilizing parameters of the pre-trained model trained on one of some publicly available large datasets [53].

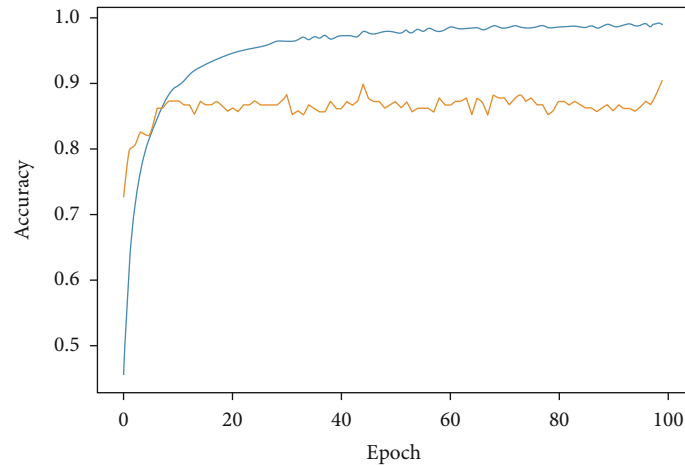
This paper utilizes the TL approach with the three DL models VGG-16, ResNet-152, and DenseNet-169 using pre-trained weights on ImageNet datasets. The ImageNet dataset accommodates 1000 categories of objects with 1,281,167,

50,000, and 100,000 samples of training, validation, and testing instances, respectively. These models are fine-tuned by keeping some of the lower layers to be frozen and the higher layers kept unfrozen as shown in Figure 2. In this experiment, the lower layers of models VGG-16, ResNet-152, and DenseNet-169, up to layers “block4_pool,” “conv4_block29_out,” and “conv4_block1_0_relu,” respectively, are kept frozen, and the higher layers beyond this are being kept unfreeze. The weight parameters obtained by pretraining these models on the ImageNet dataset are utilized for fine-tuning the models by utilizing our training data of the pap-smear dataset. To extract the array of features some of the additional layers like global max pooling, batch normalization, dropout, and dense layers have been added to all three models which result in a bunch of 1024 features set from each model. All three models are trained on training

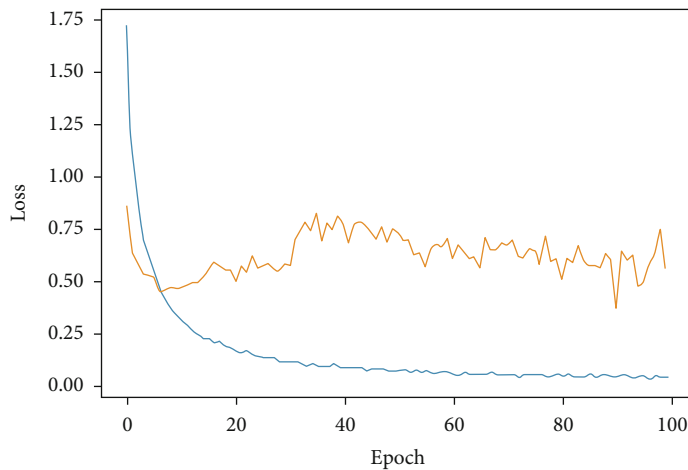
TABLE 5: Performance metrics for different fine-tuned classifiers and proposed model on LBC data.

| Data | Models | Accuracy (%) | Precision (%) | Recall (%) | F-score (%) |
|--------------|----------------|--------------|---------------|--------------|--------------|
| LBC WSI data | VGG-16 | 98.97 | 98.13 | 97.59 | 97.86 |
| | ResNet-152 | 97.94 | 93.33 | 97.29 | 95.27 |
| | DenseNet-169 | 97.42 | 91.67 | 96.71 | 94.12 |
| | Proposed model | 99.49 | 98.33 | 99.26 | 98.79 |

Best results are shown in bold.



(a)

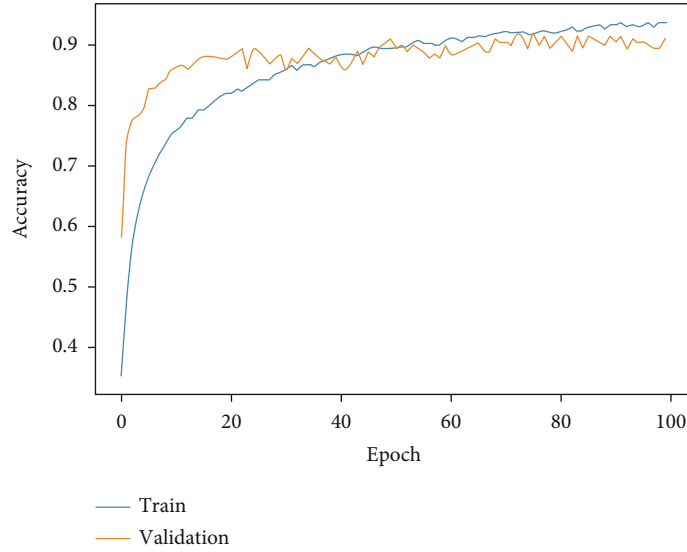


(b)

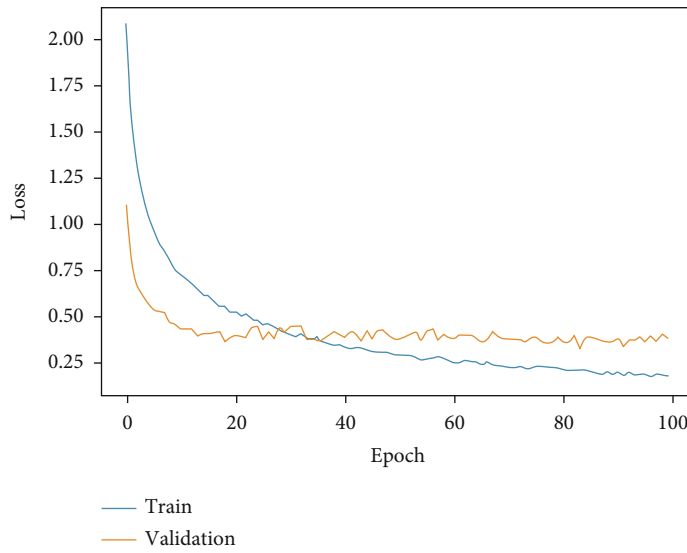
FIGURE 10: Fine-tuned VGG-16. (a) Accuracy curve. (b) Loss curve for SIPaKMeD 5-class classification.

data with a learning rate of 10^{-4} for 100 epochs, batch size of 32, and Adam optimizer. The batch size is set to 1 for validation and test data. The feature combinations “feat_1,” “feat_2,” and “feat_3” from each model are extracted in terms of weight files and saved as H5 extension files.

3.3.2. Feature Concatenation. Each of the feature combinations extracted from the three fine-tuned models i.e. ‘feat_1’, ‘feat_2’, and ‘feat_3’ has an array of 1024 features. These feature combinations are concatenated to produce hybrid features. Let there be two feature vectors v and w of



(a)



(b)

FIGURE 11: Fine-tuned ResNet-152. (a) Accuracy curve. (b) Loss curve for SIPaKMeD 5-class classification.

dimensions $n \times 1$ and $m \times 1$, respectively, their concatenated feature vector having dimension $k \times 1$. This concatenation of two feature vectors can be represented as:

$$z = v \oplus w \text{ where } v \in R^n \text{ and } w \in R^m \tag{5}$$

$$z = [v_1, v_2, \dots, v_n]^T \oplus [w_1, w_2, \dots, w_m]^T \tag{6}$$

$$z = [v_1, \dots, v_n, w_1, \dots, w_m]^T \in R^{n+m} \tag{7}$$

$$f(z): R^{n+m} \longrightarrow R^k \tag{8}$$

Here \oplus is called the concatenation operator. Each of the feature combinations ‘feat_1’, ‘feat_2’, and ‘feat_3’ has a vector of size 1024×1 , and the feature vector ‘feat’ produced by concatenation of these feature combinations is having a size of 3072×1 .

3.3.3. *Fully Connected Network (FCN)*. Lastly, a segment of the FCN network has been implemented to perform the classification of cervical cancer test images. This network contains a sequential model with dropout, batch normalization, and dense (softmax activation function) layers. The input dimension of the FCN network is set to 3072 to feed the concatenated feature set “feat.” The output node of dense layers is set equal to the number of classes in the dataset. The output prediction “ a_k ” for any test cervix image of class “ k ” is related as

$$a_k \in \sum_{i=1, \dots, k, \dots, n} a_i. \tag{9}$$

Here, n is the total number of classes in the cervical dataset. The experiment is performed by setting the learning rate

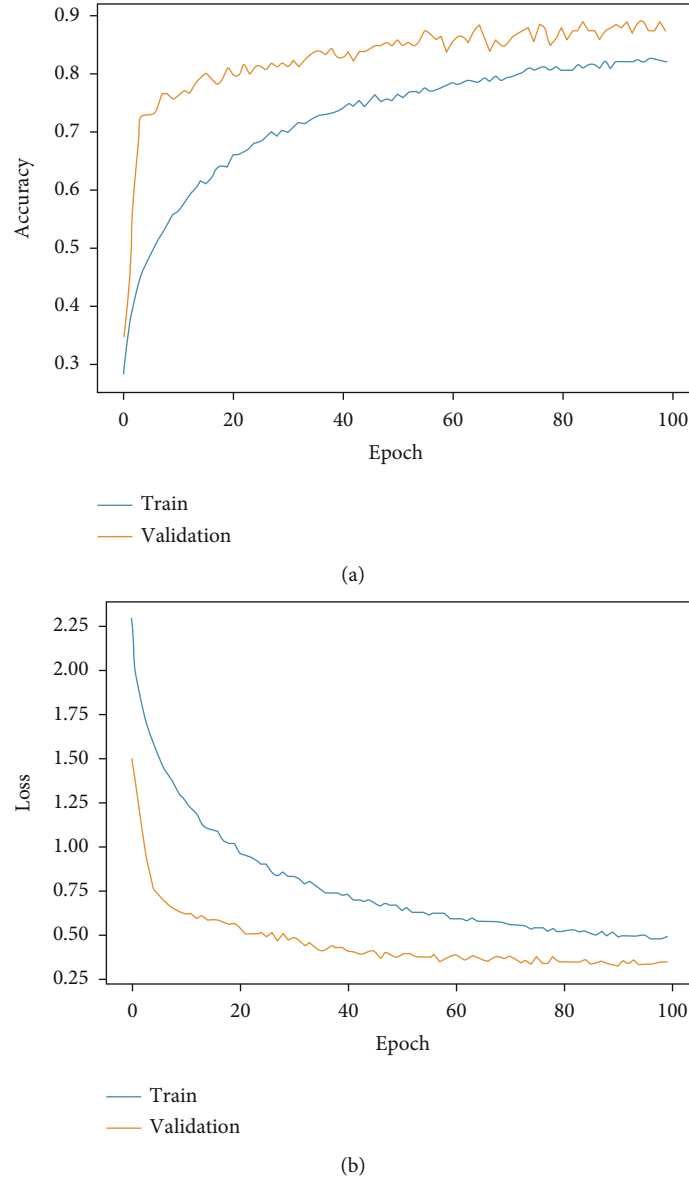


FIGURE 12: Fine-tuned DenseNet-169. (a) Accuracy curve. (b) Loss curve for SIPaKMeD 5-class classification.

to 10^{-4} for 300 epochs and a batch size of 1 with an Adam optimizer for testing of the data.

4. Results and Discussion

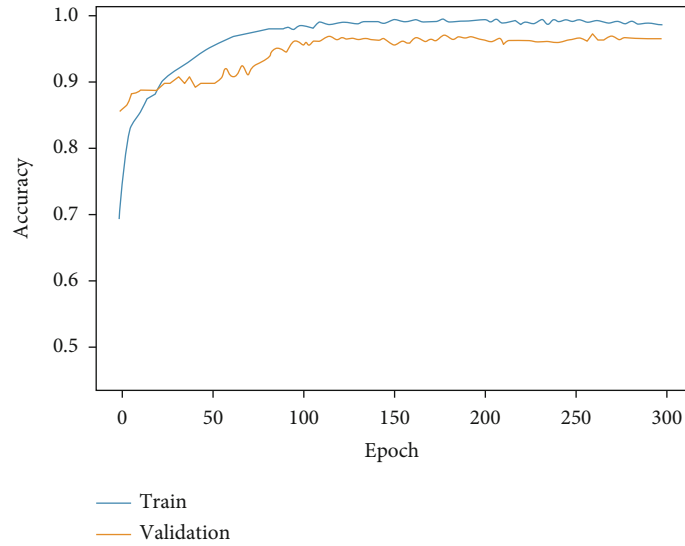
4.1. Implementation Environment. The proposed model is developed on a system with an operating system of Windows 10 having a graphic card from Nvidia Tesla V100 and 16GB of GPU RAM. The experiment is performed on the Python 3.7 programming environment with requisite libraries like Scikit-learn, Tensorflow, Keras, Cuda, CuDNN, etc.

4.2. Performance Measures. The selection of the optimum classifier for best performance is achieved by choosing competent performance metrics. Accuracy, precision, recall, and F-score are some of the indices used in this research to assess

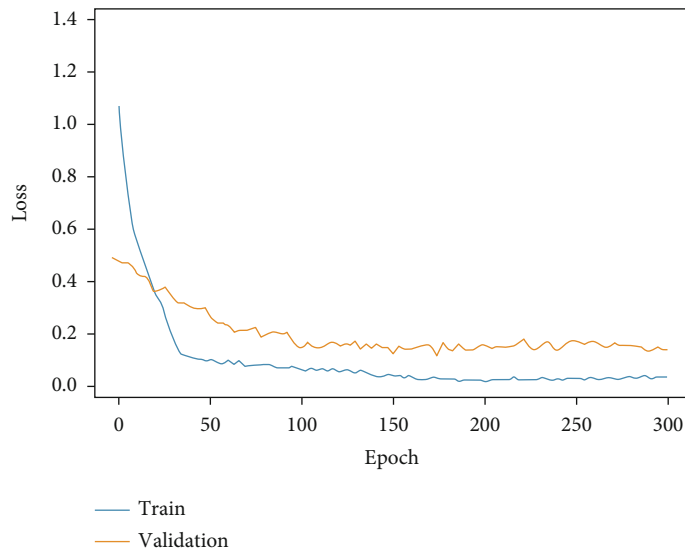
the performance of the model [54]. Accuracy (ACC) in equation (10) gives the proportion of correctly predicted cases to the total number of cases. Precision (PRE) in equation (11) represents the fraction of correctly estimated positive patterns to all positively predicted patterns. Sensitivity (SEN) in equation (12) shows the rate of genuinely anticipated positive to all correctly classified instances. F-score (FS) in equation (13) formulates the harmonic average of precision (PRE) and sensitivity (SEN).

$$ACC = \frac{TRPOS + TRNEG}{TRPOS + TRNEG + FALPOS + FALNEG}, \quad (10)$$

$$PRE = \frac{TRPOS}{TRPOS + FALPOS}, \quad (11)$$



(a)



(b)

FIGURE 13: Proposed model. (a) Accuracy curve. (b) Loss curve for SIPaKMeD 5-class classification.

$$SEN = \frac{TRPOS}{TRPOS + FALNEG}, \quad (12)$$

$$FS = \frac{2 * PRE * SEN}{PRE + SEN}, \quad (13)$$

where TRPOS is the true positive, TRNEG is the true negative, FALPOS is the false positive, and FALNEG is the false negative.

4.3. Results. This section shows the outcomes of the experiment performed on WSI SIPaKMeD cervical cancer data. Here, the implementation of the proposed model has also been analyzed on another cervical cancer database, i.e., LBC data. The performance is evaluated for individual fine-tuned models such as VGG-16, ResNet-152, and DenseNet-169 classifiers as well as for the proposed model. Figures 7 and 8 show the confusion matrix obtained for

TABLE 6: Comparison of classification accuracy for the proposed model with previous methods on SIPaKMeD WSI data.

| Methods | Classification accuracy (%) | |
|---------------------------|-----------------------------|--------------|
| | 5 class | 2 class |
| Ensemble [56] | 94.09 | 98.27 |
| CNN+PCA [57] | 96.37 | — |
| Fuzzy-based ensemble [58] | 95.43 | 98.55 |
| RCAN-DenseNet-121 [59] | 91.09 | — |
| ResNet-50 [60] | 91 | — |
| Wavelet+CNN+RF [61] | 97.01% | — |
| Proposed model | 97.45 | 99.29 |

Best results are shown in bold.

TABLE 7: Comparison of classification accuracy for the proposed model with previous methods on LBC WSI data.

| Methods | Classification accuracy (%) |
|--------------------------------|-----------------------------|
| Fuzzy-based ensemble [58] | 99.23 |
| CNN [62] | 96.89 |
| T2T-ViT+Transfer learning [63] | 98.79 |
| Exemplar Pyramid+NCA+SVM [40] | 99.47 |
| Proposed model | 99.49 |

Best results are shown in bold.

the 5-class and 2-class classifications of WSI SIPaKMeD data for the different classifiers. Both the figures describe that the confusion matrix obtained for the proposed model predicts the least incorrect instances.

Table 4 depicts the results obtained from different classifiers and our proposed model on SIPaKMeD WSI data for classification into 5 class and 2 class. Here is the performance of the individual fine-tuned models of VGG-16, ResNet-152, and DenseNet-169 are confronted with our proposed HDFCN model. Our proposed model achieves better results than the other classifiers with an accuracy of 97.45% and 99.29% for 5-class and 2-class classifications, respectively. For 5-class classification, our model predicts the pap-smear images with a precision score of 97.94%, a recall value of 98.08%, and an F-score of 98.01%. This model outperforms the other classifiers with the precision, recall, and F-score values of 98.92%, 100%, and 99.46%, respectively, in 2 class classification. Furthermore, the proposed model is assessed for LBC WSI pap-smear data. The confusion matrix and the performance indices obtained for the baseline fine-tuned classifiers and the proposed model are shown in Figure 9 and Table 5, respectively. The accuracy-loss curve for 5-class classification of SIPaKMeD pap smear data is shown in Figures 10–13. Observation shows that the proposed model achieves excellent results in classifying this data with the accuracy, precision, recall, and F-score value of 99.49%, 98.33%, 99.26%, and 98.79%, respectively.

4.4. Discussion

4.4.1. Comparative Analysis. Table 6 represents the comparison of the proposed model with the existing state-of-the-art methods for 5-class and 2-class classifications of SIPaKMeD WSI data. The comparison review of the performance of the proposed model for LBC WSI data is given in Table 7. These comparisons convincingly conclude the preciseness of the framework in cervical cancer classification.

4.4.2. McNemar’s Test. Here, we performed McNemar’s statistical non-parametric test on the baseline classifiers and our proposed model [55]. Table 8 shows the p value calculated through McNemar’s test between the individual baseline classifiers (VGG-16, ResNet-152, and DenseNet-169) and the proposed model. This test signifies the fact that the individual baseline classifiers and the proposed model are dissimilar in extracting features if the p value computed is less than 5% or 0.05. As this p value is calculated below

TABLE 8: McNemar’s test result (p value).

| McNemar’s test for the proposed model and the baseline classifier as | p value | |
|--|-------------------|--------------|
| | SIPaKMeD WSI data | LBC WSI data |
| VGG-16 | 0.0007 | 0.0196 |
| ResNet-152 | 0.0008 | 0.0012 |
| DenseNet-169 | 0.0106 | 0.0073 |

$p < 0.5$ conclude false null hypothesis.

0.05 for all the cases on both the datasets utilized, the null hypothesis is false, and the proposed classifier is dissimilar to the other DL classifier with enhanced performance. This scrutiny proves the reliability and authenticity of the framework we proposed in the classification of cervical cancer WSI pap-smear data.

5. Conclusion and Future Work

The escalating use of computer vision models in the early-stage detection of cervical cancer motivates us to propose this hybrid framework. The proposed work utilizes two-step data augmentation to increase the amount of training data. The proposed HDFCN model utilizes the hybrid features obtained from the concatenation of features extracted from the fine-tuned models of three prevalent DL algorithms: VGG-16, ResNet-152, and DenseNet-169. These hybrid features are used for the classification of cervical cancer WSI pap-smear data. The proposed model is evaluated on SIPaKMeD data and gives an accuracy of 97.45% for 5-class classification and 99.29% for 2-class classification. Moreover, the experiment performed on LBC WSI data provides 99.49% accuracy. The precise recognition of infected WSI images enables experts to perform a more in-depth analysis of cells within the images. The futuristic approach to this method involves the utilization of more optimal feature selection algorithms, progressive resizing, and advanced ensemble methods to further improve model performance and computation cost-cutting.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., “Global cancer statistics 2021: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] T. Sarenac and M. Mikov, “Cervical cancer, different treatments and importance of bile acids as therapeutic agents

- in this disease,” *Frontiers in Pharmacology*, vol. 10, p. 484, 2019.
- [3] B. Sudha, N. S. Kumar, and S. Sumathi, “Absence of knowledge and awareness about cervical cancer among educated women: a need for education about cervical cancer,” *Indian Journal of Gynecologic Oncology*, vol. 20, no. 1, pp. 1–8, 2022.
 - [4] P. L. Sachan, M. Singh, M. L. Patel, and R. Sachan, “A study on cervical cancer screening using pap smear test and clinical correlation,” *Asia-Pacific Journal of Oncology Nursing*, vol. 5, no. 3, pp. 337–341, 2018.
 - [5] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, “A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images,” *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 15–22, 2018.
 - [6] J. Ma, Y. Song, X. Tian, Y. Hua, R. Zhang, and J. Wu, “Survey on deep learning for pulmonary medical imaging,” *Frontiers in Medicine*, vol. 14, no. 4, pp. 450–469, 2020.
 - [7] J. Wang, H. Zhu, S. H. Wang, and Y. D. Zhang, “A review of deep learning on medical image analysis,” *Mobile Networks and Applications*, vol. 26, no. 1, pp. 351–380, 2021.
 - [8] J. Manhas, R. K. Gupta, and P. P. Roy, “A review on automated cancer detection in medical images using machine learning and deep learning based computational techniques: challenges and opportunities,” *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2893–2933, 2022.
 - [9] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, “A review on deep learning in medical image analysis,” *International Journal of Multimedia Information Retrieval*, vol. 11, no. 1, pp. 19–38, 2022.
 - [10] R. Yousef, G. Gupta, N. Yousef, and M. Khari, “A holistic overview of deep learning approach in medical imaging,” *Multimedia Systems*, vol. 28, no. 3, pp. 881–914, 2022.
 - [11] F. Zhuang, Z. Qi, K. Duan et al., “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
 - [12] O. E. Aina, S. A. Adeshina, and A. M. Aibinu, “Deep learning for image-based cervical cancer detection and diagnosis - a survey,” in *2019 15th international conference on electronics, computer and computation, ICECCO*, pp. 1–7, Abuja, Nigeria, 2019.
 - [13] C. Li, H. Chen, X. Li et al., “A review for cervical histopathology image analysis using machine vision approaches,” *Artificial Intelligence Review*, vol. 53, no. 7, pp. 4821–4862, 2020.
 - [14] M. M. Rahaman, C. Li, X. Wu et al., “A survey for cervical cytopathology image analysis using deep learning,” *IEEE Access*, vol. 8, pp. 61687–61710, 2020.
 - [15] H. Kaur, R. Sharma, and L. Kaur, “Automated cervical cancer image analysis using deep learning techniques from pap-smear images: a literature review,” in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pp. 1–7, Noida, India, 2021.
 - [16] Y. Kumar, S. Gupta, R. Singla, and Y. C. Hu, “A systematic review of artificial intelligence techniques in cancer prediction and diagnosis,” *Archives of Computational Methods in Engineering*, vol. 29, pp. 2043–2070, 2022.
 - [17] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, *Pap-smear benchmark data for pattern classification*, Nature inspired smart information systems (NiSIS 2005), 2005.
 - [18] Z. Lu, G. Carneiro, A. P. Bradley et al., “Evaluation of three algorithms for the segmentation of overlapping cervical cells,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, pp. 441–450, 2017.
 - [19] K. Fernandes, J. S. Cardoso, and J. Fernandes, “Transfer learning with partial observability applied to cervical cancer screening,” in *Iberian Conference on Pattern Recognition and Image Analysis*, L. Alexandre, J. Salvador Sánchez, and J. Rodrigues, Eds., vol. 10255 of Lecture Notes in Computer Science, , pp. 243–250, Springer, Cham, 2017.
 - [20] H. A. Phoulady and P. R. Mouton, “A new cervical cytology dataset for nucleus detection and image classification (Cervix93) and methods for cervical nucleus detection,” 2018, <https://arxiv.org/abs/1811.09651>.
 - [21] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, “SIPAKMED: a new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3144–3148, Athens, Greece, 2018.
 - [22] E. Hussain, L. B. Mahanta, H. Borah, and C. R. Das, “Liquid based-cytology pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions,” *Data in Brief*, vol. 30, article 105589, 2020.
 - [23] M. Wu, C. Yan, H. Liu, Q. Liu, and Y. Yin, “Automatic classification of cervical cancer from cytological images by using convolutional neural network,” *Bioscience Reports*, vol. 38, no. 6, 2018.
 - [24] V. Kudva, K. Prasad, and S. Guruvare, “Automation of detection of cervical cancer using convolutional neural networks,” *Biomedical Engineering*, vol. 46, no. 2, pp. 135–145, 2018.
 - [25] K. Adem, S. Kiliçarslan, and O. Cömert, “Classification and diagnosis of cervical cancer with stacked autoencoder and softmax classification,” *Expert Systems with Applications*, vol. 115, pp. 557–564, 2019.
 - [26] C. Li, D. Xue, X. Zhou et al., “Transfer learning based classification of cervical cancer immunohistochemistry images,” in *Proceedings of the third international symposium on image computing and digital medicine*, pp. 102–106, New York, NY, USA, 2019.
 - [27] H. Lin, Y. Hu, S. Chen, J. Yao, and L. Zhang, “Fine-grained classification of cervical cells using morphological and appearance based convolutional neural networks,” *IEEE Access*, vol. 7, pp. 71541–71549, 2019.
 - [28] K. H. S. Allehaibi, L. E. Nugroho, L. Lazuardi, A. S. Prabuwono, and T. Mantoro, “Segmentation and classification of cervical cells using deep learning,” *IEEE Access*, vol. 7, pp. 116925–116941, 2019.
 - [29] Y. Miyagi, K. Takehara, and T. Miyake, “Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images,” *Molecular and Clinical Oncology*, vol. 11, no. 6, pp. 583–589, 2019.
 - [30] Z. Alyafeai and L. Ghouti, “A fully-automated deep learning pipeline for cervical cancer classification,” *Expert Systems with Applications*, vol. 141, article 112951, 2020.
 - [31] M. Kuko and M. Pourhomayoun, “Single and clustered cervical cell classification with ensemble and deep learning methods,” *Information Systems Frontiers*, vol. 22, no. 5, pp. 1039–1051, 2020.
 - [32] V. Kudva, K. Prasad, and S. Guruvare, “Hybrid transfer learning for classification of uterine cervix images for cervical

- cancer screening,” *Journal of Digital Imaging*, vol. 33, no. 3, pp. 619–631, 2020.
- [33] S. K. Saini, V. Bansal, R. Kaur, and M. Juneja, “ColpoNet for automated cervical cancer screening using colposcopy images,” *Machine Vision and Applications*, vol. 31, no. 3, pp. 1–15, 2020.
- [34] Y. Li, J. Chen, P. Xue et al., “Computer-aided cervical cancer diagnosis using time-lapsed colposcopic images,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3403–3415, 2020.
- [35] V. Chandran, M. G. Sumithra, A. Karthick et al., “Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images,” *BioMed Research International*, vol. 2021, Article ID 5584004, 15 pages, 2021.
- [36] Y. R. Park, Y. J. Kim, W. Ju, K. Nam, S. Kim, and K. G. Kim, “Comparison of machine and deep learning for the classification of cervical cancer based on cervicography images,” *Scientific Reports*, vol. 11, no. 1, p. 16143, 2021.
- [37] H. Chen, J. Liu, Q. M. Wen et al., “CytoBrain: cervical cancer screening system based on deep learning technology,” *Journal of Computer Science and Technology*, vol. 36, no. 2, pp. 347–360, 2021.
- [38] K. M. A. Adweb, N. Cavus, and B. Sekeroglu, “Cervical cancer diagnosis using very deep networks over different activation functions,” *IEEE Access*, vol. 9, pp. 46612–46625, 2021.
- [39] A. Cibi and R. J. Rose, “Classification of stages in cervical cancer MRI by customized CNN and transfer learning,” *Cognitive Neurodynamics*, 2022.
- [40] O. Yaman and T. Tuncer, “Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images,” *Biomedical Signal Processing and Control*, vol. 73, article 103428, 2022.
- [41] L. Yan, H. Song, Y. Guo et al., “HLDnet: novel deep learning based artificial intelligence tool fuses acetic acid and Lugol’s iodine cervicograms for accurate pre-cancer screening,” *Biomedical Signal Processing and Control*, vol. 71, article 103163, 2022.
- [42] S. Fekri-Ershad, “Pap smear classification using combination of global significant value, texture statistical features and time series features,” *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 31121–31136, 2019.
- [43] F. Shafiei and S. F. Ershad, “Detection of lung cancer tumor in CT scan images using novel combination of super pixel and active contour algorithms,” *Traitement du Signal*, vol. 37, no. 6, pp. 1029–1035, 2020.
- [44] A. Ghoneim, G. Muhammad, and M. S. Hossain, “Cervical cancer classification using convolutional neural networks and extreme learning machines,” *Future Generation Computer Systems*, vol. 102, pp. 643–649, 2020.
- [45] S. Fekri-Ershad and S. Ramakrishnan, “Cervical cancer diagnosis based on modified uniform local ternary patterns and feed forward multilayer network optimized by genetic algorithm,” *Computers in Biology and Medicine*, vol. 144, article 105392, 2022.
- [46] S. Kumar, S. K. Gupta, V. Kumar, M. Kumar, M. K. Chaube, and N. S. Naik, “Ensemble multimodal deep learning for early diagnosis and accurate classification of COVID-19,” *Computers and Electrical Engineering*, vol. 103, article 108396, 2022.
- [47] A. B. Jung, K. Wada, J. Crall et al., *Imgaug*, GitHub, San Francisco, CA, USA, 2020.
- [48] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021.
- [49] L. Alzubaidi, J. Zhang, A. J. Humaidi et al., “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, no. 1, p. 53, 2021.
- [50] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [52] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Honolulu, HI, USA, 2017.
- [53] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, “Transfer learning for medical image classification: a literature review,” *BMC Medical Imaging*, vol. 22, no. 1, p. 69, 2022.
- [54] 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. 1–11, 2015.
- [55] N. Japkowicz and M. Shah, “Performance evaluation in machine learning,” in *Machine Learning in Radiation Oncology*, pp. 41–56, Springer, Cham, 2015.
- [56] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. Myo Aung, “Computer-assisted screening for cervical cancer using digital image processing of pap smear images,” *Applied Sciences*, vol. 10, no. 5, p. 1800, 2020.
- [57] G. V. K. Kiran and G. Meghana Reddy, “Automatic Classification of Whole Slide Pap Smear Images Using CNN With PCA Based Feature Interpretation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1074–1079, Long Beach, CA, USA, 2019.
- [58] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, and R. Sarkar, “A fuzzy rank-based ensemble of CNN models for classification of cervical cytology,” *Scientific Reports*, vol. 11, no. 1, p. 14538, 2021.
- [59] R. Wang, M. A. Armin, S. Denman, L. Petersson, and D. Ahmedt-Aristizabal, “Towards interpretable attention networks for cervical cancer analysis,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3613–3616, Mexico, 2021.
- [60] R. Hardatin, M. N. Muslimah, and S. N. Hasanah, “Perbandingan Arsitektur ResNet50 dan ResNet101 dalam Klasifikasi Kanker Serviks pada Citra Pap Smear,” *Faktor Exacta*, vol. 14, no. 3, pp. 160–167, 2021.
- [61] S. Karim Kutty and G. Chandrasekhara Menon, “Enhancing convolutional neural network model with spectral features for the identification of cervical dysplasia,” *International Journal of Imaging Systems and Technology*, vol. 32, no. 6, pp. 1916–1927, 2022.

- [62] N. K. Chauhan and K. Singh, "Impact of variation in number of channels in CNN classification model for cervical cancer detection," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pp. 1–6, Noida, India, 2021.
- [63] C. Zhao, R. Shuai, L. Ma, W. Liu, and M. Wu, "Improving cervical cancer classification with imbalanced datasets combining taming transformers with T2T-ViT," *Multimedia Tools and Applications*, vol. 81, no. 17, pp. 24265–24300, 2022.

Research Article

Automated Evaluation of Upper Airway Obstruction Based on Deep Learning

Yunho Jeong, Yeeyewin Nang , and Zhihe Zhao 

State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases, Department of Orthodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, Sichuan 610041, China

Correspondence should be addressed to Zhihe Zhao; zhzhao@scu.edu.cn

Received 28 July 2022; Revised 31 October 2022; Accepted 25 January 2023; Published 18 February 2023

Academic Editor: Kawsar Ahmed

Copyright © 2023 Yunho Jeong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objectives. This study is aimed at developing a screening tool that could evaluate the upper airway obstruction on lateral cephalograms based on deep learning. **Methods.** We developed a novel and practical convolutional neural network model to automatically evaluate upper airway obstruction based on ResNet backbone using the lateral cephalogram. A total of 1219 X-ray images were collected for model training and testing. **Results.** In comparison with VGG16, our model showed a better performance with sensitivity of 0.86, specificity of 0.89, PPV of 0.90, NPV of 0.85, and F1-score of 0.88, respectively. The heat maps of cephalograms showed a deeper understanding of features learned by deep learning model. **Conclusion.** This study demonstrated that deep learning could learn effective features from cephalograms and automated evaluate upper airway obstruction according to X-ray images. **Clinical Relevance.** A novel and practical deep convolutional neural network model has been established to relieve dentists' workload of screening and improve accuracy in upper airway obstruction.

1. Introduction

Upper airway obstruction can result in reduction of breathing or impediment of gas exchange, and it is usually associated with sleep-disordered breathing (SDB) [1, 2]. The cause of upper airway obstruction includes polyps, environmental irritants, allergic rhinitis, and adenotonsillar hypertrophy [3, 4]. Increasing evidence has shown an association between dentofacial anomalies and obstruction in upper airway. Lopatiene et al. analyzed examination results including dental casts and radiographs of 49 children with respiratory obstruction, and they found significant link between nasal resistance and increase overjet, open bite, and maxillary crowding [5]. Children with upper airway obstruction may manifest as mouth breathing, which can result in narrow maxilla, mandibular skeletal retrognathism, increased lower facial height, and high palate [6]. Most children with this type of malocclusion and craniofacial deformity present to dental clinics complaining of occlusal disorder or dissatisfaction with their profile. Lateral cephalogram was a useful and common tool for dentists to evaluate the severity of upper

airway obstruction. Although there are multiple other tools applied to assess upper airway obstruction, including computed tomography (CT), fluoroscopy, magnetic resonance imaging (MRI), and fiberoptic pharyngoscopy; lateral cephalometry is still an appealing approach for screening upper airway obstruction in dental clinics as it is a cheap and easily available technique with less radiation and certain diagnostic value [7–9]. Cephalometric analysis based on McNamara method is a classical measurement for airway analysis [10]. However, the landmark-label process is time- and energy-consuming even for a senior orthodontist. Besides, this experience-dependent technique is difficult to master for young dentists and dentists who have not investigated cephalometric measurements such as endodontists or prosthodontists, which may lead to missed or delayed diagnosis. Thus, it would be useful for orthodontists to develop an automated evaluation method to improve efficiency in upper airway obstruction using lateral cephalograms.

In the past several years, automated methods based on deep learning have achieved excellent results in diagnosis, segmentation, detection tasks, and so on [11–13]. For

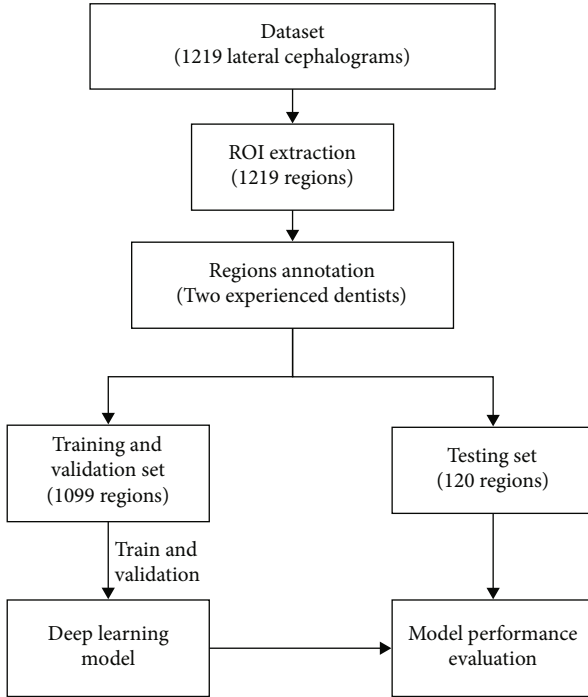


FIGURE 1: Study flowchart.

TABLE 1: Clinical and demographic characteristics of study cohorts.

| Characteristic | Training set (N = 1099) | Testing set (N = 120) |
|---------------------|-------------------------|-----------------------|
| Median age (range) | 12.15 (7-18) | 12.02 (7-18) |
| Sex | | |
| Male | 545 | 54 |
| Female | 554 | 66 |
| Clinical evaluation | | |
| Obstructive | 550 | 60 |
| Nonobstructive | 549 | 60 |

example, Mahdi et al. proposed a residual network-based faster R-CNN model to recognize teeth and evaluate both positional relationship and confidence score of the candidates. This model showed an F1-score of 0.982 which indicated that deep learning was useful and reliable for dental assistance [14]. Similarly, deep learning was also used for identifying the brand and model of a dental implant from a radiograph with a sensitivity of 93.5% and a specificity of 94.2% [15]. In orthodontics, accurate skeletal classification can assist orthodontists in making treatment plans. Yu et al. developed a deep learning model for skeletal classification solely from the lateral cephalogram. In that work, deep learning learned from X-ray imaging labeled by human experts and exhibited superior performance with >90% sensitivity, specificity, and accuracy for vertical and sagittal skeletal diagnosis [16].

In this paper, we developed a novel deep convolutional neural network (DCNN) based on ResNet backbone for automated evaluation of upper airway obstruction. Here,

our study was novel with 4 main contributions. Firstly, this is the first research to evaluate upper airway obstruction based on deep learning using lateral cephalograms with high sensitivity and specificity. Secondly, the heat maps of cephalograms showed a deeper understanding of features learned by deep learning model. This visualization provided interpretable information in upper airway obstruction based on deep learning. Finally, our model is lite and practical and can be deployed in fundamental clinics with less memory and computational overhead.

2. Materials and Methods

2.1. Dataset. A study flowchart of our study was presented in Figure 1. Cephalometric radiographs were retrospectively examined for 1783 cohorts who had initially visited our hospital between March and September 2019. We first excluded the low-quality images (324 images) and eliminated images (240 images) where the anatomic structure (soft palate, tongue, or pharyngeal wall) was difficult to recognize. Finally, a total of 1219 X-ray images of cohorts with lateral X-ray examination were obtained, of which the numbers of upper airway obstruction and nonobstruction were 610 and 609, respectively. Cephalometric radiographs were taken from X-ray machine Morita $\times 550$ (Tube energy 80 kV, Tube current 10 mA; Morita, Kyoto, Japan). The distance between X-ray plate and X-ray machine was 180 cm, and the resolution of X-ray images is 1752×1537 ($22.9 \text{ cm} \times 20.1 \text{ cm}$). We cropped the airway region from the center to the bottom of original X-ray images with a resolution of 1000×500 . The airway regions of X-ray images were randomly divided into 2 groups: a training set (1099 images) and a testing set (120 images). Demographic data are shown in Table 1.

The McNamara method was considered a classic cephalometric analysis for evaluation of upper airway dimension [10]. Linear measurements were performed using Image J software (Rasband software, W.S., Image J, National Institutes of Health, Bethesda, MD, <http://rsb.info.nih.gov/ij/>). Many studies have applied McNamara method for assessment of upper airway, which can be divided into nasopharynx (upper pharynx) and oropharynx (lower pharynx) [17–19]. The upper pharyngeal width (UPW) was measured linearly from a point on the posterior wall of the soft palate to the posterior pharyngeal wall where there was the greatest closure of the airway. The measurement of the lower pharyngeal width (LPW) was a linear distance from an intersection point of the posterior border of the tongue and the lower border of mandible to the closest point on the posterior pharyngeal wall [17] (Figure 2). In our research, X-ray images of patients with mixed dentition were manually labelled as “obstruction” ($UPW < 12 \text{ mm}$ or $LPW < 10 \text{ mm}$) or “nonobstruction” ($UPW \geq 12 \text{ mm}$ and $LPW \geq 10 \text{ mm}$), and the images of patients with permanent dentition were manually labelled as “obstruction” ($UPW < 17.4 \text{ mm}$ or $LPW < 10 \text{ mm}$) or “nonobstruction” ($UPW \geq 17.4 \text{ mm}$ and $LPW \geq 10 \text{ mm}$). All the measurements were carried out by the two experienced dentists blinded by each other. A third senior orthodontic specialist with 30 years of experience was consulted in cases of disagreement. If the three experts

still could not get an agreement, the confusing image would be excluded.

2.2. Deep Learning Model. ResNet-18 is a famous model and achieved excellent results in the field of image classification [20]. In this model, the backbone structure consists of two convolutional layers with skip connection. This backbone has been proven that it has a strong capacity of feature extraction in the image classification task. Hence, we chose ResNet-18 as the backbone to develop a lite and practical model for automated evaluation of upper airway obstruction. Regarding our model, the kernel size of two convolutional layers is 3×3 and its stride is 1. To overcome overparameterization problem, we not only reduce the size and the number of kernels but also reduce the number of backbone blocks. The kernel number is 16 in the first backbone and the second backbone. Max pooling with the size of 2×2 is deployed between two backbone blocks. The features of airway region are extracted in the first extra convolutional layer with 5×5 kernel and 1 stride. And the feature maps produced by first layer are given to backbone structure. At the end of backbone structure, fully-connected layer with softmax switched feature maps into the probability of obstruction or nonobstruction. The rectified linear unit (ReLU) is included in every convolutional layer. The architecture of model is shown in Figure 3.

2.3. Statistical Analysis and Evaluation Criteria. To better measure the performance of the model, we used evaluation metrics of sensitivity (SEN), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV) and F1-score [21–24]. And the metrics equations are calculated as follows:

$$\begin{aligned}
 \text{SEN} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
 \text{SPEC} &= \frac{\text{TN}}{\text{FP} + \text{TN}}, \\
 \text{PPV} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
 \text{NPV} &= \frac{\text{TN}}{\text{FN} + \text{TN}}, \\
 \text{F1 score} &= \frac{2 * \text{PPV} * \text{SEN}}{\text{PPV} + \text{SEN}},
 \end{aligned} \tag{1}$$

where the TP, FP, TN, and FN indicated true positive, false negative, true negative, and false negative, respectively. Positive/negative means that the model predicts that the X-ray image is obstructive/nonobstructive, and true/false means that the prediction is right/wrong. In these metrics, the F1-score is the most overall metric which indicates the harmonic mean of PPV and SEN. The highest possible value of F1-score is 1, indicating perfect PPV and SEN, and the lowest possible value is 0, if either PPV or SEN is zero. All value of metrics is ranged from 0 to 1.

For deeper understanding of the feature in X-ray images, the heat map with class activation mapping (CAM) was generated according to the method proposed by Zhou et al. [25].

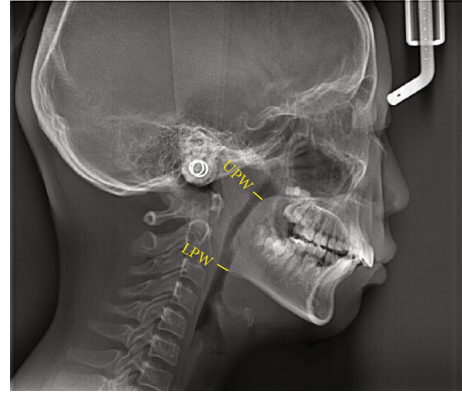


FIGURE 2: Illustration of cephalometric measurements for upper airway. Upper pharyngeal width (UPW): linear distance from a point on the posterior wall of the soft palate (the anterior half part) to the posterior pharyngeal wall where there was the greatest closure of the airway. Lower pharyngeal width (LPW): measured from an intersection point of the posterior border of the tongue and the lower border of mandible to the closest point on the posterior pharyngeal wall.

This map visually highlights the cephalogram region that is most informative in evaluation of upper airway obstruction.

3. Results

The study was developed using 1099 X-ray images for training and 120 X-ray images for testing. All experiments were performed in Python 3.6 and TensorFlow 1.9 on a single NVIDIA RTX 2080Ti [26]. We randomly selected 100 images from training set as a validation set to observe training situation and obtain the highest performance. In the training phase, we used a learning rate of 0.001 in the Adam optimizer and used the “Cross-Entropy” loss function with the batch size of 50. After 30 epochs, automatic evaluation of upper airway obstruction was performed using the testing dataset. In many dental applications, VGG16 is a popular model for classification and diagnosis [27, 28]. Hence, we also carried out experimental comparison between VGG16 and DCNN. In the testing set, DCNN model showed 0.86 sensitivity, 0.89 specificity, 0.90 PPV, 0.85 NPV, and 0.88 F1-score, respectively. DCNN showed higher performance than VGG16 in our study (Table 2).

Figure 4 shows the heat maps created with class activation mapping. This is an indication of a well-trained model that effectively uses the information in the cephalogram. According to the heat maps, the upper airway area was activated when model received a sample with airway obstruction. This activated area revealed that model taught itself according to human annotated conclusion without extra orientation. The processing speed of DCNN was about 5 s for analyzing 120 lateral cephalograms with a single NVIDIA RTX 2080Ti graphic processing unit.

4. Discussion

Upper airway obstruction, as a hot topic studied by dentists and otolaryngologists, showed an intimate association with

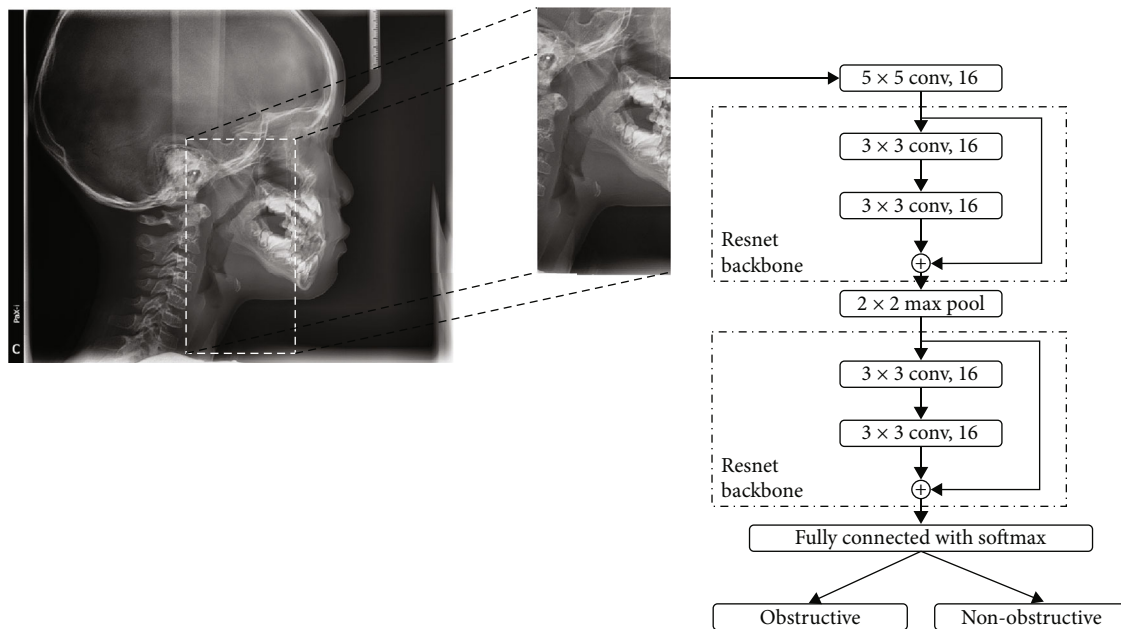


FIGURE 3: Preprocessing and model architecture.

TABLE 2: The performance of deep learning model.

| Operator | SEN | SPEC | PPV | NPV | F1-score |
|--------------|------|------|------|------|----------|
| VGG16 | 0.84 | 0.86 | 0.86 | 0.84 | 0.85 |
| ResNet-18 | 0.83 | 0.86 | 0.87 | 0.82 | 0.85 |
| EfficientNet | 0.84 | 0.88 | 0.88 | 0.83 | 0.86 |
| MobileNet v2 | 0.85 | 0.88 | 0.88 | 0.85 | 0.87 |
| DCNN (ours) | 0.86 | 0.89 | 0.90 | 0.85 | 0.88 |

malocclusion and development of craniofacial complex, and it was also the main etiological factor of obstructive sleep apnea syndrome (OSAS) in children [29]. OSAS may lead to problems which were harmful to children, such as inattention, poor learning, failure to thrive, or even pulmonary hypertension [30]. However, missed or delayed diagnosis was common as signs and symptoms of children were not clear and the experience-dependent diagnosis method was difficult to master [31]. For children who were in an early stage of mental and physical development, airway patency and sleep quality were significant. So, developing a timely and accurate screen system for upper airway obstruction was advantageous.

As a two-dimensional analysis method, lateral cephalometric images have long been discussed about its reliability in assessing pharyngeal volumes. A study consisting of 36 prepubertal children ranging from 4.9–9.8 years old compared the validity of upper airway using MRI and cephalometric measurements, and researchers found that cephalometric measurements showed significant correlations with MRI measurements. The authors concluded that the cephalometric radiograph was a useful screening tool when evaluating nasopharyngeal or retropalatal airway size [9]. Besides, cephalometric analysis was also investigated as a useful tool to evaluate OSAS patients [32]. In this paper,

we developed a novel and practical DCNN model to automatically evaluate upper airway obstruction using the lateral cephalogram. Our data demonstrated that deep learning method was able to evaluate upper airway obstruction with high accuracy and improve screening efficiency in dental clinics.

To the best of our knowledge, so far, there is only one research that is similar to our research, which applied artificial intelligence technology to detect patients with severe obstructive sleep apnea based on cephalometric radiographs [33]. However, it only focused on the oropharynx but not on the nasopharynx. Some research reported automatic segmentation of the airway space with convolutional neural network on CBCT images [34, 35]. We must admit that CBCT offers information on cross-sectional areas, volume, and 3D form that cannot be determined by cephalometric images. However, many studies have confirmed the screening value of cephalometric images [9, 32], which possesses lots of advantages, including lower cost, and less radiation dose. Besides, cephalometric images are more widely used by dentists, especially in developing countries and clinics which cannot afford CBCT machines.

In our study, a DCNN model based on ResNet backbone was applied for automated evaluation of upper airway obstruction. Our experimental results revealed that a simplified model can overcome overparameterization problems to some extent. In dental applications, the size of dental image dataset is always smaller than natural image datasets generally, since the acquisition of data requires the authorization of the patient. Additionally, professional knowledge is required in the work of data annotation. Thus, it is difficult to collect enough high-quality dental samples for training and testing. Under this condition, a typical model like VGG16 with a large number of parameters is suffered from insufficient data so that it can easily overfit the dataset. To avoid the defects mentioned above, we used ResNet as the

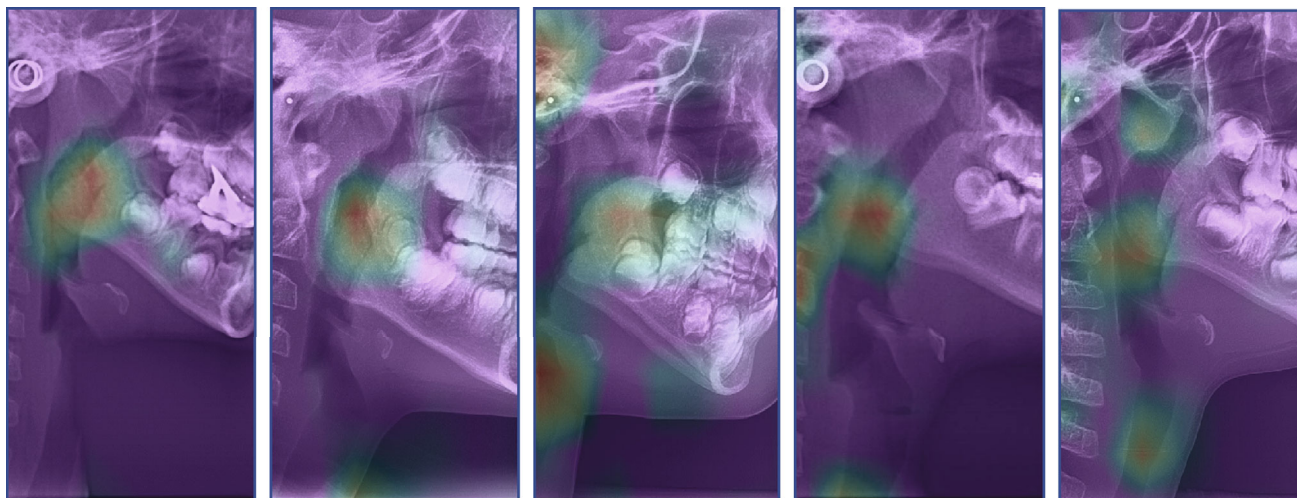


FIGURE 4: Class activation maps. The red area shows key features discovered by deep learning.

backbone to construct a simplified DCNN model, which showed better performance than VGG16.

Although labeling landmarks in the lateral cephalogram were important for orthodontic diagnosis [36–38], errors in landmark identification method are widespread, necessitating time-consuming manual correction. Hence, we applied deep learning method to directly classify X-ray images rather than identification key point methods. Our model showed good performance in both sensitivity and specificity. Moreover, our model did not require extra steps of feature extraction for training or prediction. The heat maps (Figure 4) also confirmed that the well-trained model can discover abnormal area in X-ray images by itself without extra processing.

Nevertheless, our study presented several limitations. First, although 1219 was a large number within the realm of dental research, it was far less than the application requirement of deep learning. Second, our 1219 X-ray images were produced by Morita $\times 550$ at one resolution. Our model may not be robust at other resolutions, which should be addressed through appropriate expansion of the training set with images at other resolutions. Last but not the least, major improvements for the sensitivity and specificity of our research may be achieved in the future by increasing sample size, applying advanced architectures, optimal training strategies, and data generation.

5. Conclusions

This study presents a deep learning model that can automatically detect upper airway obstruction with higher accuracy and more time-efficiency, which would reduce the burden on dentists in clinical work. A simplified DCNN model based on ResNet backbone structure showed good performance for automatic evaluation of upper airway obstruction based on the lateral cephalogram. However, deep learning is not completely accurate in the detection of upper airway obstruction. To avoid false negative diagnosis, regular follow-ups and reevaluations are required if necessary.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to data security but are available from the corresponding author on reasonable request.

Consent

Written informed consent was not required for this study because all the included patients were collected retrospectively. Exemption of informed consent will not affect the rights and health of included patients. The application for free informed consent has been approved by the Institutional Review Board.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was supported by the Sichuan Science and Technology Program (grant numbers. 2022ZDZX0031), Angalign Scientific Research Fund (grant number. SDTS21-3-21), and Public Service Platform of Chengdu (grant number. 2021-0166-1-2).

References

- [1] V. Katyal, Y. Pamula, A. J. Martin, C. N. Daynes, J. D. Kennedy, and W. J. Sampson, "Craniofacial and upper airway morphology in pediatric sleep-disordered breathing: systematic review and meta-analysis," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 143, no. 1, pp. 20–30.e3, 2013.
- [2] C. Flores-Mir, M. Korayem, G. Heo, M. Witmans, M. P. Major, and P. W. Major, "Craniofacial morphological characteristics in children with obstructive sleep apnea syndrome: a

- systematic review and meta-analysis," *Journal of the American Dental Association (1939)*, vol. 144, no. 3, pp. 269–277, 2013.
- [3] C. B. Preston, "Chronic nasal obstruction and malocclusion," *The Journal of the Dental Association of South Africa*, vol. 36, no. 11, pp. 759–763, 1981.
- [4] M. Abe, A. Mitani, A. T. Yao, L. Zong, K. Hoshi, and S. Yanagimoto, "Awareness of malocclusion is closely associated with allergic rhinitis, asthma, and arrhythmia in late adolescents," *Healthcare*, vol. 8, no. 3, p. 209, 2020.
- [5] K. Lopatiene and A. Babarskas, "Malocclusion and upper airway obstruction," *Medicina*, vol. 38, no. 3, pp. 277–283, 2002.
- [6] A. Stellzig-Eisenhauer and P. Meyer-Marcotty, "Interaction between otorhinolaryngology and orthodontics: correlation between the nasopharyngeal airway and the craniofacial complex," *GMS Current Topics in Otorhinolaryngology, Head and Neck Surgery*, vol. 9, article Doc04, 2010.
- [7] C. M. Quinlan, H. Otero, and I. E. Tapia, "Upper airway visualization in pediatric obstructive sleep apnea," *Paediatric Respiratory Reviews*, vol. 32, pp. 48–54, 2019.
- [8] N. Samman, H. Mohammadi, and J. Xia, "Cephalometric norms for the upper airway in a healthy Hong Kong Chinese population," *Hong Kong Medical Journal*, vol. 9, no. 1, pp. 25–30, 2003.
- [9] K. Pirila-Parkkinen, H. Lopponen, P. Nieminen, U. Tolonen, E. Paakko, and P. Pirttiniemi, "Validity of upper airway assessment in children a clinical, cephalometric, and MRI study," *Angle Orthodontist*, vol. 81, no. 3, pp. 433–439, 2011.
- [10] J. A. McNamara, "A method of cephalometric evaluation," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 86, no. 6, pp. 449–469, 1984.
- [11] F. Casalegno, T. Newton, R. Daher et al., "Caries detection with near-infrared transillumination using deep learning," *Journal of Dental Research*, vol. 98, no. 11, pp. 1227–1233, 2019.
- [12] Y. Zhao, P. C. Li, C. Q. Gao et al., "TSASNet: tooth segmentation on dental panoramic X-ray images by two-stage attention segmentation network," *Knowledge-Based Systems*, vol. 206, article 106338, 2020.
- [13] M. G. Endres, F. Hillen, M. Salloumis et al., "Development of a deep learning algorithm for periapical disease detection in dental radiographs," *Diagnostics*, vol. 10, no. 6, p. 430, 2020.
- [14] F. P. Mahdi, K. Motoki, and S. Kobashi, "Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs," *Scientific Reports*, vol. 10, no. 1, p. 12, 2020.
- [15] M. H. Said, M. K. Le Roux, J. H. Catherine, and R. Lan, "Development of an artificial intelligence model to identify a dental implant from a radiograph," *The International Journal of Oral & Maxillofacial Implants*, vol. 35, no. 6, pp. 1077–1082, 2020.
- [16] H. J. Yu, S. R. Cho, M. J. Kim, W. H. Kim, J. W. Kim, and J. Choi, "Automated skeletal classification with lateral cephalometry based on artificial intelligence," *Journal of Dental Research*, vol. 99, no. 3, pp. 249–256, 2020.
- [17] N. N. Silva, R. H. Lacerda, A. W. Silva, and T. B. Ramos, "Assessment of upper airways measurements in patients with mandibular skeletal class II malocclusion," *Dental Press Journal of Orthodontics*, vol. 20, no. 5, pp. 86–93, 2015.
- [18] Z. M. Baka and M. Fidanboy, "Pharyngeal airway, hyoid bone, and soft palate changes after class II treatment with Twin-block and Forsus appliances during the postpeak growth period," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 159, no. 2, pp. 148–157, 2021.
- [19] N. Fareen, M. K. Alam, M. F. Khamis, and N. Mokhtar, "Treatment effects of two different appliances on pharyngeal airway space in mixed dentition Malay children," *International Journal of Pediatric Otorhinolaryngology*, vol. 125, pp. 159–163, 2019.
- [20] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 Ieee Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [21] M. M. Ali, K. Ahmed, F. M. Bui et al., "Machine learning-based statistical analysis for early stage detection of cervical cancer," *Computers in Biology and Medicine*, vol. 139, article 104985, 2021.
- [22] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, article 104672, 2021.
- [23] A. A. Hemu, R. B. Mim, M. M. Ali, M. Nayer, K. Ahmed, and F. M. Bui, "Identification of significant risk factors and impact for ASD prediction among children using machine learning approach," in *Paper presented at: 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Bhilai, India, 2022.
- [24] M. Z. H. Ontor, M. M. Ali, S. S. Hossain, M. Nayer, K. Ahmed, and F. M. Bui, "YOLO_CC: deep learning based approach for early stage detection of cervical cancer from cervix images using YOLOv5s model," in *Paper presented at: 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Bhilai, India, 2022.
- [25] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 Ieee Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929, Las Vegas, NV, USA, 2016.
- [26] M. Abadi, P. Barham, J. Chen et al., "TensorFlow: a system for large-scale machine learning," in *Proceedings of Osd'16: 12th Usenix Symposium on Operating Systems Design and Implementation*, pp. 265–283, Savannah, GA, USA, 2016.
- [27] S. Matsuda, T. Miyamoto, H. Yoshimura, and T. Hasegawa, "Personal identification with orthopantomography using simple convolutional neural networks: a preliminary study," *Scientific Reports*, vol. 10, no. 1, p. 7, 2020.
- [28] S. Sukegawa, K. Yoshii, T. Hara et al., "Deep neural networks for dental implant system classification," *Biomolecules*, vol. 10, no. 7, p. 984, 2020.
- [29] R. K. Garg, A. M. Afifi, C. B. Garland, R. Sanchez, and D. L. Mount, "Pediatric obstructive sleep apnea," *Plastic and Reconstructive Surgery*, vol. 140, no. 5, pp. 987–997, 2017.
- [30] J. L. Carroll, "Obstructive sleep-disordered breathing in children: new controversies, new directions," *Clinics in Chest Medicine*, vol. 24, no. 2, pp. 261–282, 2003.
- [31] C. L. Rosen, L. D'Andrea, and G. G. Haddad, "Adult criteria for obstructive sleep apnea do not identify children with serious obstruction," *The American Review of Respiratory Disease*, vol. 146, no. 5_part_1, pp. 1231–1234, 1992.
- [32] B. Deberryborowiecki, A. Kukwa, and R. H. I. Blanks, "Cephalometric analysis for diagnosis and treatment of obstructive sleep apnea," *The Laryngoscope*, vol. 98, no. 2, pp. 226–234, 1988.

- [33] S. Tsuiki, T. Nagaoka, T. Fukuda et al., “Machine learning for image-based detection of patients with obstructive sleep apnea: an exploratory study,” *Sleep & Breathing*, vol. 25, pp. 2297–2305, 2021.
- [34] S. Shujaat, O. Jazil, H. Willems et al., “Automatic segmentation of the pharyngeal airway space with convolutional neural network,” *Journal of Dentistry*, vol. 111, article 103705, 2021.
- [35] J. Park, J. Hwang, J. Ryu et al., “Deep learning based airway segmentation using key point prediction,” *Applied Sciences*, vol. 11, no. 8, p. 3501, 2021.
- [36] S. Nishimoto, Y. Sotsuka, K. Kawai, H. Ishise, and M. Kakibuchi, “Personal computer-based cephalometric landmark detection with deep learning, using cephalograms on the internet,” *The Journal of Craniofacial Surgery*, vol. 30, no. 1, pp. 91–95, 2019.
- [37] G. Dot, F. Rafflenbeul, M. Arbotto, L. Gajny, P. Rouch, and T. Schouman, “Accuracy and reliability of automatic three-dimensional cephalometric landmarking,” *International Journal of Oral and Maxillofacial Surgery*, vol. 49, no. 10, pp. 1367–1378, 2020.
- [38] J. H. Lee, H. J. Yu, M. J. Kim, J. W. Kim, and J. Choi, “Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks,” *BMC Oral Health*, vol. 20, no. 1, p. 270, 2020.