

Research Article

Premature Ventricular Contractions' Detection Based on Active Learning

Xianrong Zhang ¹, Muhammad Shafiq ², Guijun Zheng ³, Junping Wan ¹
and Zhe Sun ¹

¹Cyberspace Institute of Advanced Technology (CIAT), Guangzhou University, Guangzhou 510006, China

²School of Computer Science & Technology, Harbin Institute of Technology, Harbin 150001, China

³School of Software Engineering, University of Science and Technology of China, Hefei 230051, China

Correspondence should be addressed to Zhe Sun; sunzhe@gzhu.edu.cn

Received 12 January 2021; Revised 12 February 2021; Accepted 25 February 2021; Published 8 March 2021

Academic Editor: Shaukat Ali

Copyright © 2021 Xianrong Zhang 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.

Premature ventricular contractions (PVCs) are one of the most common cardiovascular diseases with high risk to a large population of patients. It has been shown that supervised learning algorithms can detect PVCs from beat-level ECG data. However, a huge human effort is needed in order to achieve an accurate detection rate. A convolutional autoencoder was trained in this work in an unsupervised fashion to extract features automatically with zero prior specialized knowledge. Random forest was adopted as a supervised algorithm trained on the features generated by the autoencoder. Various active learning selection strategies, uncertainty-based and diversity-based, were studied on top of the random forest. In each iteration of active learning, the training data are updated with newly selected samples and fed into the classifier. The performance on an independent validation set is recorded in each iteration. As a result, among the different uncertainty sampling strategies, the least confidence score shows a better F_1 score of 0.85 than other methods. In between the two diversity-based strategies, the representative clustering sample had the best F_1 score than the k -center-greedy algorithm. By comparing the performance of different active learning methods trained on half of the original data size with the same classifier trained on the full set, the F_1 score of least confidence is still better than the full set. This study demonstrates that active learning could help reduce human annotation effort by achieving the same level of performance as the classifier trained on the fully annotated training data.

1. Introduction

Premature ventricular contractions (PVCs) are one of the most common arrhythmias that occur in a large population of patients [1]. In a more serious case, when PVCs happen with other cardiac risk factors concurrently, it could lead to many extreme situations, like cardiac death or heart attack. Electrocardiograms (ECG) are recognized as the most useful and noninvasive technique for monitoring cardiac activity, within which the monitoring of various arrhythmias are the main tasks [2]. Also, it would be difficult for clinicians to recognize PVCs from ECG if only a short period of ECG is provided since the evaluation of PVCs needs a reference from neighbor ECG segments. Therefore, it is desirable if one approach can automatically detect PVCs from just one beat of ECG.

Applying traditional machine learning algorithms on detection of PVCs is one mainstream category, in which feature engineering is the most important process in order to get better performance. ECG morphological based features have been proved to be useful. Geddes [3] applied the QRS complexity length interval gap between the two R peaks, and the first-order signal derivative of the signal as the manually selected features, and then fed them into a tree-based classification algorithm. However, to follow the same approach, in addition to the annotation of PVCs, the PQRS annotation for each beat of the ECG is also needed, which causes much more work for human clinicians. In work [4], the author came up with a combination of features from a different aspect. Besides the temporal features, sparse signal decomposition is also adopted for each segment of ECG. At the same time, human-design rules were

also proposed as an additional filter for features. The results show that PVCs can be detected in a robust way compared to just applying any signal aspect of the feature.

Within all the work above, domain knowledge is required to find the most plausible features. In recent years, the deep learning algorithms have become the focus of the research community. Due to their powerful classification capabilities and end-to-end methods, deep learning does not require tedious feature engineering to process the data. Researchers only need to feed the raw data into a deep neural network, and the network can extract the most important features and achieve a better performance than traditional algorithms. In [5], they used a convolutional autoencoder to extract features and then fed the extracted feature list into the random forest for the classification task. The results achieved over 90% accuracy over the whole patient cohort.

Despite traditional machine learning or deep neural network, the success of one classifier highly depends on the massive and accurate annotated training set, which requires tremendous human effort.

Active learning has been brought into much research as a novel idea due to its ability to achieve similar or better performance using only half of the original data [6–8], where the core idea behind active learning is to seek out the most informative data samples to annotate. The general process of active learning starts with an initial well-labeled training set and a data pool with no annotation. Classifiers are then trained on a baseline set of training and each data sample from the data pool will be evaluated by a pre-trained classifier. The output probability for each sample in the data pool will be used as the input for the active learning selection strategy. Finally, selected samples from the data pool will be transmitted to annotators to obtain the exact label and then appended to the training set for the second iteration of the training process. It is evident throughout the whole workflow that the selection strategy is the most important part of the active learning process [9].

Some of the main contributions of this study are listed as follows:

- (1) Overall framework: An active learning framework is proposed to detect premature ventricular beats, which reduced the workload and the cost of manual labeling data. The advantages of artificial intelligence are creatively applied to biomedicine to help clinicians improve the accuracy of PVCs detection.
- (2) Feature engineering: Convinced autoencoders are designed to automatically extract features from data without the prior knowledge of medical experts, providing novel insights for feature engineering of physiological data. After the extracted features are input into the classifier, the results show that the traditional machine learning methods have advantages. Using convolutional autoencoders to extract features is more convenient and fast than incorporating human annotation efforts.
- (3) Data distribution: Initial training data distributions were investigated, and in most cases, random sampling of initial training data may not be sufficient to represent the entire data set. We propose an alternative approach to initial training data and test the impact of each approach.
- (4) Selection strategy: In active learning, the selection strategy is crucial. By comparing the selection strategies, we put forward the selection strategies suitable for PVC testing.

This paper is organized as follows. Additional related work is reported in Section 2. Methods and the entire design of the work are reported in Section 3. A comprehensive illustration of the work is given in Section 4. Finally, the discussion section of this paper is in Section 5 and the conclusion is presented in Section 6.

2. Related Work

2.1. Traditional Machine Learning. Active learning has attracted a lot of spotlight in the machine learning community. Different selection strategies have proven to be useful in many tasks, which were designed from different perspectives. When considering uncertainty, the uncertainty is calculated based on output probability from the initiation classifier. When it comes to diversity, many geometry-based approaches are proposed. The core idea behind them is to calculate the distance between samples in the unlabeled pool and then select the data points that represent the entire distribution of the data pool. Many state-of-art methods have been proposed. In [10], the author took the selection strategy as the k -cover problem, which is the solution to find the best k data point out of the whole data pool. While the k -cover problem is NP-hard and has been proven, the authors tried a k -greedy algorithm to simulate the k -cover problem. Reference [11] proposed a novel way by building an auxiliary model to estimate the loss for each input, by which those samples with higher loss will be selected for annotation in each iteration. These works examine the data distribution of active learning, but these advanced active learning techniques have not been explored in physiological research.

2.2. Traditional Selection Strategies. Traditional selection strategies for active learning in ECG data have been pursued recently. Reference [12] applied active learning as an effective approach for finding the most relevant signals with motion artifacts in order to accurately classify the human activity, in which only 16% of the original training data is used. In [13], active learning is used mainly to produce more generalizable training data among the patient cohort rather than reducing the human annotation effort. The authors adopted a global recurrent neural network that captures the time order of the input signal. The selection strategy is based on a combination of entropy index, model output, and Premature-or-Escape-Flag index, which is temporal information learned from the embedding layer of the model. At the same time, the comparison between different selection strategies [14–17] has not been investigated in the field of physiological measurement research. In the biomedical area, active learning was adopted as the strategy combined with SVM trying to discriminate an ECG-based classification task

[18]. In [19], a novel selection strategy called AIFT was created. The results show that the proposed method can help improve the performance of classifying three biomedical image classification tasks, with less human effort involved. These works had contributed to the development of active learning, and selection strategies suitable for PVCs have not been studied.

2.3. Deep Learning Algorithms. Deep learning algorithms are also being used to create active learning classifiers for ECG-based classification tasks [20, 21]. In [22], a global recurrent neural network was adopted for the purpose of ECG beat classification. Morphological and temporal features for ECG beats were investigated, and active learning was utilized to select the most representative samples for training. As for the work [23], a convolutional neural network was applied for the wearable ECG classification. Breaking-ties and modified breaking-ties algorithms were used with active learning simultaneously to improve model performance. ECG abnormalities with the convolutional neural network were studied in [24]. Despite the noisy part, additional six arrhythmia events within beat ECG were also detected. Active learning was also planned in the procedure in order to deal with the unseen pattern inside the original training data. Manually set decision rules are used for PVCs detection in [25]. By identifying statistical and rhythm rules, the PVC beats could be detected with high accuracy. Electrocardiograms (ECG) data are a powerful tool for reflecting cardiovascular events. Different arrhythmias can be automatically detected through machine learning algorithms. In the most recent work [26], the authors proposed a deep learning method that can detect PVCs without any human annotation effort by localizing the PVC beats via deep learning algorithms.

However, active learning for PVCs detection is not well explored. The above methods are not particularly good for PVC detection. Firstly, there is a lack of research on initialization training data. Secondly, the above work is still lacking in the extraction of data features. Moreover, most of the work requires manual annotation of data, and sufficient sample features cannot be extracted from small samples. How to improve the accuracy with fewer data remains to be studied. Finally, there was less research on active learning selection strategies for those works. In particular, there are few studies on the classification of active learning in biomedicine.

3. Materials and Methods

Active learning algorithms are applied in PVCs for detection in this paper. The algorithm model comprehensively considers data initialization, data feature extraction, and sampling strategies. In order to improve the performance of the original classifier, this paper uses k -means++ for initialization. In order to extract features better, this paper designs a convolutional autoencoder. To fully study the impact of sampling strategies on active learning, uncertain sampling and diversity sampling were also studied.

3.1. Overview. Compared to related work, we hope to learn from small samples, to reduce the workload of manual annotation. In addition, in order to improve training accuracy, we proposed an initialization method for training data. To fully examine the data characteristics, we propose a convolutional autoencoder. Finally, we comprehensively examined selection strategies to improve the effectiveness of active learning. The overall flow of the framework is as follows:

- (1) Initial training data: we propose a training data initialization method that is more suitable for PVC detection, and the method is k -means++.
- (2) Feature engineering: in this framework, we design a convolutional autoencoder for feature engineering. Self-convolutional encoders can learn the characteristics of data by themselves to solve the problem of low efficiency of manual annotation data sets.
- (3) Data pool selection strategy: in this work, two main aspects of the selection strategy, uncertainty and diversity, were well explored in the task of PVCs detection on beat-level ECG data. Uncertainty sampling is discussed as follows: ① Least confidence sampling. ② Margin sampling. ③ Shannon entropy sampling. Diversity sampling is discussed as follows: ① K -center-greedy sampling. ② Representative cluster sampling.

The process can be summarized as follows.

A convolutional autoencoder was trained to extract the features automatically. These data-driven features are then fed into a random forest classifier. During each iteration of active learning, the same algorithm is trained on the updated training data. The overall framework is shown in Figure 1.

We first downloaded PVCs detection data sets from the website, and these data sets have not been annotated by experts. In the initial phase, we performed random initialization and initialization of the K -means++ algorithm. This was done to study the impact of data initialization on the effect of active learning. After the initialization algorithm is complete, the active learning algorithm can select a portion of the representative data. This small part of the initialized data is integrated into the initial subset. We hope that the initial subset of the initialization algorithm can represent the overall data distribution. A selected initial subset was manually marked by oracle. After manually marking the initial subset, we constructed a convolutional autoencoder to extract features from the initial subset. In addition, these features are trained using a random forest classification algorithm. Before the initial data subset is trained, the weights of the convolutional autoencoder and the random forest classifier are randomly distributed.

Because we assume that the meaning of the initial data subset is that a portion of the data represents the total data, after training on the initial data subset, our convolutional autoencoder and classifier may fully learn the characteristics of the data set. In the next iteration of the algorithm, we need to reassign a representative part of the data set to the oracle for annotation. Therefore, in an iterative process, we first

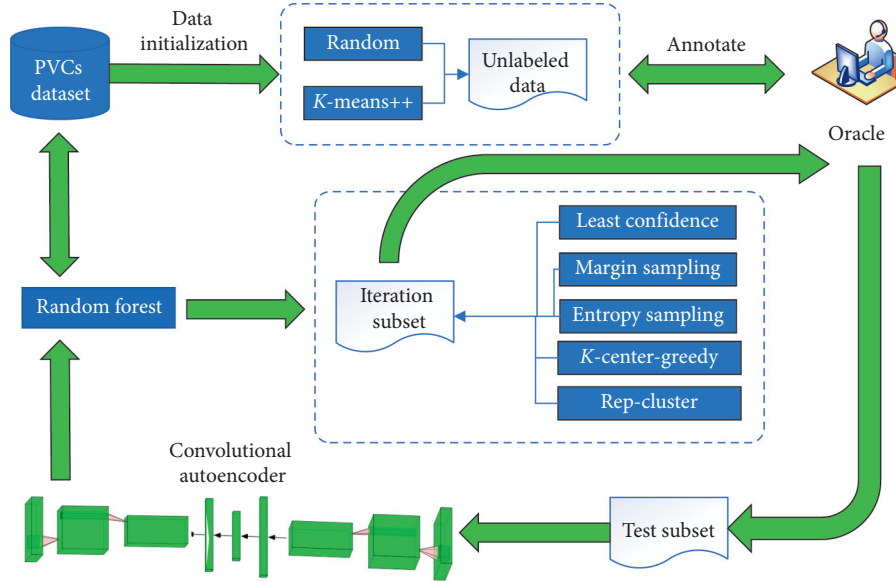


FIGURE 1: The overview of the workflow.

input the remaining unlabeled data into a convolutional autoencoder and classifier for predictive classification. We selected the data after classification using the uncertainty sampling strategy and the diversity sampling strategy. We call this data set an iterative subset. We resubmit data that are difficult to classify to Oracle for labeling. After multiple iterations, until the labeled data reaches half of the data pool, the iteration ends.

This paper presents a model from three aspects: training data initialization, iterative process, and selection strategy classifier.

3.2. Training Data Initialization. In most of the work, the researcher tended to select the initial training set randomly. At the same time, random selection might lead to the skewness of model parameters due to a lack of the big picture of all the training data. In this paper, we also propose a data initiation approach, k -means++, to investigate whether it is important to consider initiating the training set by design versus selecting randomly. The ultimate goal of k -means++ is to find the best subset that covers the whole big picture of the original training data. A detailed description of the significance of the results is presented in Table 1.

After data initialization is completed, feature engineering is required for the data. Feature selection directly affects the performance of the algorithm [27, 28]. Common deep learning algorithms often require a large number of classification tags, which need manual labeling. Manual labeling of data is a huge workload, which is also a disadvantage of deep learning. In order to reduce manual labeling data and fully extract features of small samples, we designed a convolutional autoencoder. An autoencoder is one of the most popular unsupervised neural network methods that take the input data also as output data. The fundamental architecture of the autoencoder is a three-layer neural network that includes an input layer, hidden layer, and

TABLE 1: The algorithm of k -means++.

Algorithm K -means ++
Input: Training data $Z = (x_i, y_i), i = 1, \dots, N$, Number of selected centroids K .
initiation centroids = [one random selected points]
For j from 1 to $K - 1$
Distance $\leftarrow \{ \}$
For t from 1 to length (centroids)
compute distances of "point" in Z to centroids and store the point P with minimum distance
End
centroids = centroids + P
End

output layer. The restrictions for the architecture are as follows: (1) The whole architecture is symmetric, being connected by the hidden layer. (2) The number of perceptrons of input and output are identical. (3) The number of perceptrons of the hidden layer must be smaller than the number of perceptrons of the input layer. By the "learning myself" philosophy, as shown in equation (1), the autoencoder managed to get a usable representation of the original input in a lower dimension from the hidden layer, and minimize the distance between the input x and output z as follows:

$$J(x, z) = \|x - z\|^2. \quad (1)$$

However, the ability of the autoencoder to handle complex real-world data is highly restricted if only the traditional three-layer architecture is adopted. The PVCs data set is sequence information and the potential connection between the data before and after cannot be ignored. Traditional autoencoders use ordinary neural networks to train data, which causes data stacking, and the context or spatial information of the data is also ignored. Therefore, it is

critical to design an autoencoder that can consider contextual information. As shown in the figure, the convolutional autoencoder improves the traditional autoencoder. Following the convolutional layer, linear rectification layer, and pooling layer, the convolutional autoencoder retains the spatial information of the PVCs data set. At the same time, more variants of autoencoders were proposed [29, 30]. Among these, convolutional autoencoders are widely used. Instead of importing raw data, the input data first go through several layers of convolutional kernels for feature extraction. Then, the extracted feature list is fed into a fully connected hidden layer. As shown in Figure 2, in our study, the input ECG recordings will go through a batch of one-dimensional convolutional layers, and then the length of the latent layer is set at 25. After that, the convolutional autoencoder makes the difference between the input and output as small as possible.

In this paper, each beat was extracted with the length of 250 data points, in which 89 samples were before the position of each R peak annotated by humans, and 160 after the position of each R peak annotated. In the model aspect, the dimension of latent space in the hidden layer is selected at 25, which has been proven to be sufficient to represent the raw 250-length signal [5]. We use the mean squared error as our loss function and Adam as the optimizer for the training process, and a more detailed description of the model is described in Table 2.

After the hyperparameters are set, the CAE is trained for 50 epochs at a batch size of 200. The final weight is selected in the epoch with the least loss during training, and this weight is used for inference purposes only except in the experiment in which a different approach of updating the weight of CAE is applied. In Table 1, the output size column indicates the number of samples.

3.3. Iterative Process and Selection Strategy. After the data are initialized with K -means++, we also design a convolutional autoencoder to extract features from the data. We will now discuss the iterative process of active learning for each round and the selection strategy for the iterative process.

As illustrated in Figure 3, active learning begins with initial training data to obtain the intermediate model. Then for each iteration, a batch of unlabeled data is selected according to the output probability of the intermediate model across the entire data pool, as illustrated in equation.

$$\min_{s^1: |s^1| \leq b} E_{x, y \sim P_z} [l(x, y; A_{s^0, s^1})], \quad (2)$$

where s^0 and s^1 are denoted as labeled data sets and unlabeled data candidates, respectively, and b represents the selected samples during each iteration. The math behind equation (2) is trying to select b samples from s^1 which results in a minimal loss based on the current intermediate model.

Newly selected samples will be appended to the training set and fed into the next iteration. An independent validation set is used to evaluate the performance of the model in each iteration. The entire active learning

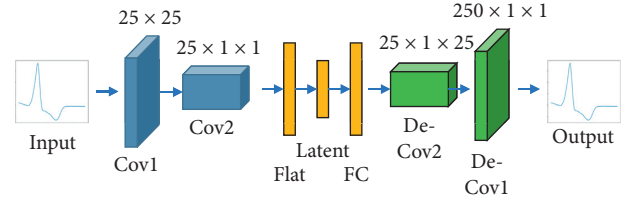


FIGURE 2: The architecture of the convolutional autoencoder.

TABLE 2: A detailed description of the architecture of the convolutional autoencoder model.

Layer number	Layer name	Kernel size	Output size
0	Input		(-1, 250)
1	Reshape		(-1, 250, 1)
2	Convolution1D	20	(-1, 25, 25)
3	Reshape		(-1, 25, 25, 1)
4	Convolution2D	(1, 20)	(-1, 25, 1, 1)
5	Convolution2D transpose	(1, 20)	(-1, 25, 1, 25)
6	Convolution2D transpose	(20, 1)	(-1, 250, 1, 1)
7	Reshape/output		(-1, 250)

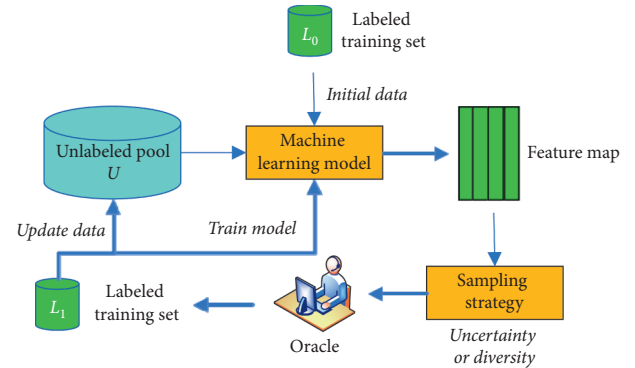


FIGURE 3: General workflow of active learning.

process stops when the performance on the validation set is satisfied.

In each iteration of active learning, the selection strategy determines the quality of the results of active learning. Therefore, we discuss two types of sampling strategies, namely, uncertainty sampling and diversity sampling. Specific uncertainty sampling is discussed as follows: ① Least confidence sampling. ② Margin sampling. ③ Shannon entropy sampling. Diversity sampling is discussed as follows: ① K -center-greedy sampling. ② Representative cluster sampling. The discussion of different selection strategies enables to design a framework that is more suitable for the detection of PVCs.

As pool-based sampling strategies are the most widely used, and when there is only one unlabeled sample in the sample pool, it is equivalent to a stream-based sampling

strategy. Therefore, this article mainly studies this type of sampling strategy, especially uncertainty sampling strategies and differential sampling strategies.

A sampling strategy based on uncertainty is the most widely applicable type of sampling strategy. This sampling strategy selects the classifier to mark the samples whose predicted value of p is closest to 0.5. The sampling strategy is not only suitable for most classifiers; it effectively reduces the workload of human experts and greatly improves classifier accuracy and generalization abilities.

The sampling strategy selects one or a batch of samples in each iteration. We certainly hope that the information provided by the sample inquired is comprehensive, and the information provided by each sample is not repeated or redundant, that is, there are certain differences between the samples. In the case of extracting a single sample with the largest amount of information in each iteration and adding it to the training set, the model is retrained in each iteration, so that the acquired knowledge is used in the evaluation of sample uncertainty and can effectively avoid data redundancy. But if you query a batch of samples for each iteration, you should find ways to ensure the diversity of the samples and avoid data redundancy. This is the differential selection strategy.

By using the least confidence strategy, in each iteration, the learner will select samples which the intermediate model is most unconfident about, as shown in equation (3), in which X_{new} represents all the data from the unlabeled pool. For example, in a binary classification task class A and class B, there are two different unlabeled data samples s_1 and s_2 . The intermediate model predicts samples s_1 with label A at a probability of 0.9 and samples s_2 with label A at a probability of 0.5. The least confidence strategy will select s_2 and transport it to the annotators for its actual label.

$$\phi_{\text{LC}}(x) = 1 - P_{\theta}(y^* | x_{\text{new}}). \quad (3)$$

Although the least confidence has been proven to be useful, it still has disadvantages when the model is only unconfident in one class, which will lead to a data imbalance problem. At the same time, the parameters in the intermediate model will skew toward one class. Margin sampling is capable of solving this problem. Instead of only focusing on the probability of one class, margin sampling also calculates the difference of probabilities between the first possible label (y_{first} in equation (4)) and the second possible label (y_{second} in equation (4)), as shown in equation (4). The sample with the least difference means the model is also confused as to which label this sample truly belongs to. Those samples will be selected under this strategy.

$$\text{Sample}_{\text{selected}} = \operatorname{argmin}_x (P(y_{\text{first}} | x_{\text{new}}) - P(y_{\text{second}} | x_{\text{new}})). \quad (4)$$

Moving the margin sampling one step further, Shannon entropy allows us to consider the probabilities from all the possible classes in a classification task. In the field of information theory, entropy is a popular measurement of the randomness of a system. In the study of active learning, for each iteration, the Shannon entropy is calculated over all of

the predicted label probabilities, as shown in equation (5). The higher the entropy value, the more uncertainty there will be. Under entropy sampling, samples with the highest entropy value will be selected for the annotation process.

$$\text{Sample}_{\text{selected}} = \operatorname{argmax}_x - \sum_i P(y_i | x_{\text{new}}) \log P(y_i | x_{\text{new}}). \quad (5)$$

The main idea behind K -center-greedy is to select K points, which can represent the whole distribution of the unlabeled data pool. The K -center-greedy method starts with initiating the centroid with one randomly selected data point. For each iteration, the centroid is updated by adding the data point with the longest distance to the original centroid. The distance between one point and the centroid is calculated based on its distance from the nearest center point. Formally, we can denote existing pool s^1 , labeled set s^0 , and a budget b for each iteration, then the idea of K -center-greedy can be defined as follows :

$$\min_{s^1: |s^1| \leq b} \max_i \min_{j \in s^1 \cup s^0} \Delta(x_i, x_j). \quad (6)$$

Representative cluster sampling improves the K -center-greedy by selecting new points from the margin of the classes instead of from the whole data set. K -center-greedy has been proven to be successful, but its performance can be contaminated if there are too many outliers in the data. By only selecting new samples from the margin of each class, this situation can be greatly alleviated.

3.4. Random Forest. A good classifier can greatly improve the effectiveness of active learning. Therefore, a classifier suitable for PVCs data is the core component of the entire algorithm.

Random forest classifier is more commonly used for medical data because of its remarkable ability to resist training overfitting, which makes it a perfect choice for downstream active learning approaches. Since the output probability of the classifier is one major criterion for selection strategies, bini impurity is used as the criteria for splitting the nodes when the tree is building in this work.

The overall algorithm flow of this paper is shown in Table 3:

4. Results and Discussion

In this paper, we aimed to investigate the influence of active learning in terms of PVCs detection in three aspects: Initiation training data creation, different selection strategies, and different choices for updating the weights for CAE. This section presents the results of those experiments.

4.1. Experiment Data. To evaluate the proposed approach, we adopted the MIT-BIH arrhythmia database [26]. In this database, 48 30-minute two-channel ECG recordings were collected from 47 patients from 1975 to 1979. Those recordings were sampled at 360 Hz. In the annotation aspect,

TABLE 3: Overall algorithm flow of PVC detection.

Algorithm ALPVCsD

Input: Unlabeled data set $Z = (x_i, y_i), i = 1, \dots, N$,
Number of initial selected samples: K ,
Number of samples selected during iteration: i .

Output: Labeled data set $L = (a_j, b_j), j = 1, \dots, M$ ($0 < M < N$),
Trained classifier model: random forest

Initialize: centroids = [one random selected points]
 $Z_len \leftarrow \text{length}(Z)$
 $L_len \leftarrow \text{length}(L)$
Feature extraction from data by self convolution encoder

For j **from** 1 **to** $K - 1$
Distance $\leftarrow \{ \}$
For t **from** 1 **to** length (centroids)
compute distance of “point” in Z to centroids and store the point P with minimum distance
End centroids = centroids + P

End
 $L \leftarrow L \cup \text{centroids}$ (annotated the K data by Oracal)
 $Z \leftarrow Z \setminus \text{centroids}$
 $Z_len \leftarrow Z_len - K$
 $L_len \leftarrow L_len + K$

Repeat:
Training random forest with centroids.
If the number of unlabeled data $< 1/2 N$:
Break

For t **in** Z
calculate the minimum confidence using equation (3) and store it in Q
End
 $P_data \leftarrow \text{Sort } Q$ and select the smallest 50 data
 $L \leftarrow L \cup P_data$
 $Z \leftarrow Z \setminus P_data$ train random forest use the 50 labeled data
 $Z_len \leftarrow Z_len - 50$
 $L_len \leftarrow L_len + 50$

End

both the QRS complexes are annotated automatically first and then reviewed by a human expert, and beat-level arrhythmia types are also annotated by human experts into ten categories, in which normal beats and PVCs are the two populated classes. In this paper, we used the first 20 recordings, indexed from 100 to 125, as the training set and the remaining 24 records, indexed from 200–234 as the testing set. Figure 4(a) represents a normal ECG, and Figure 4(b) represents an ECG of PVCs beat. Ventricular premature beats occur before the sinus node impulse reaches the ventricle, at any part of the ventricle or ectopic rhythm point of the ventricular septum, and an electric pulse is sent out in advance, causing ventricular depolarization, called ventricular premature contraction, or PVC for short. The summary is: heart attacks resulting from untimely impulses originating in the ventricles are the most common arrhythmia.

The data preprocessing procedure follows the same method as in [5]. Each recording is split into beat-levels according to the position of the labeled R peak. The segment is then constructed using 89 data points before the R peak and 160 data points after. In this way, the input shape for the downstream classifier is 250, in which the position of the R peak is located at 90. A standard normalization is then applied on each input of length 250 to obtain the new segment with all data points valued between 0 and 1.

4.2. Evaluation Index. The problems studied in this paper belong to classification problems. The common evaluation indexes of classification problems include accuracy P , recall R , and F_1 score. The confusion matrix is needed to calculate the above indexes, and the confusion matrix [31] is shown in Table 4:

Accuracy is the score of the classifier correctly predicted in all samples [32]. F_1 value is a comprehensive consideration of precision and recall, as shown in equation (7). It can reflect the classification performance more comprehensively, and so it is the main evaluation index to measure the experimental effect in this paper.

$$\begin{aligned}
 \text{accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\
 \text{precision} &= \frac{TP}{TP + FP}, \\
 \text{recall} &= \frac{TP}{TP + FN}, \\
 F_1 &= \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}.
 \end{aligned} \tag{7}$$

In the metrics presented above, although the accuracy rate can judge the overall correct situation, it cannot be used as a good indicator to measure the result when the sample is

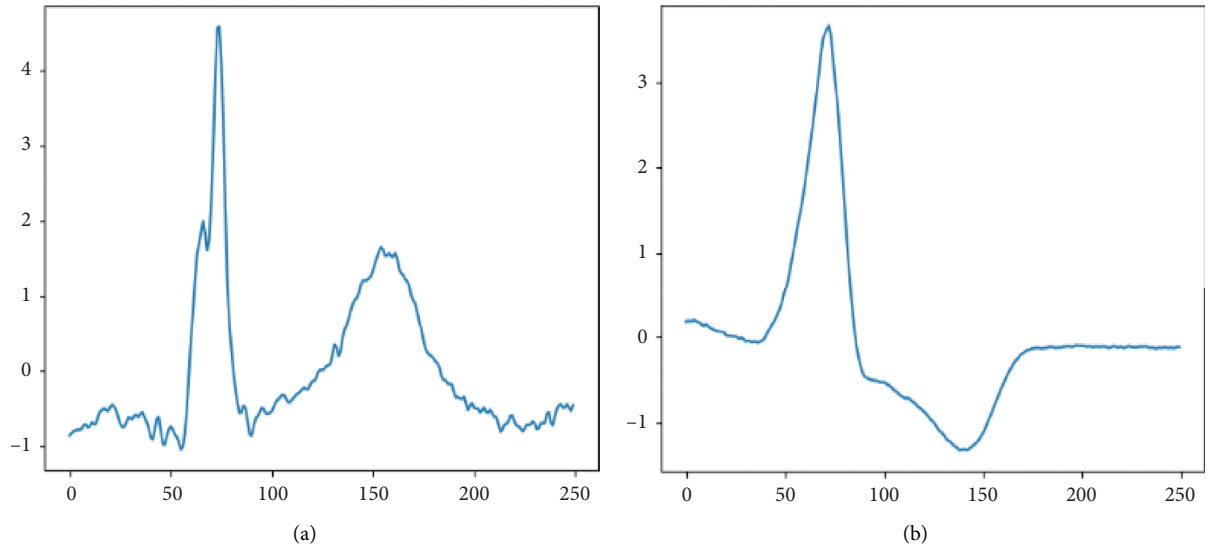


FIGURE 4: G Data preview of normal sinus rhythm and PVCs within one beat. (a) Normal beat. (b) PVC beat.

TABLE 4: Confusion matrix of the classification problem.

	Test positive	Test negative
Condition positive	True positive (TP)	False negative (FN)
Condition negative	False positive (FP)	True negative (TN)

not balanced. Precision and accuracy may seem similar, but they are two entirely different concepts. Precision represents the precision of the prediction in the positive sample results, but the accuracy rate represents the overall correctness of the prediction, including positive samples and negative samples. The recall rate is for our original sample, and it indicates how many positive examples in the sample are predicted correctly. F_1 scores are a harmonic measure of accuracy and recall.

4.3. Experiment Setup. As the first step in active learning, preparing the initial training set is crucial. In most of the work, the initial training set is selected randomly. One reason is that they ignored the importance of the training data at the start point. Another realistic reason is that, in a real-world situation, the chance of selecting the initial training data is not always available in many tasks.

The benchmark classifier maintained in the learning module must have a certain classification accuracy. Consequently, the benchmark classifier must be trained initially before active learning. The key to solving the problem is how to construct a high-performance initial training sample set. Overall, the initial training set selected at random is not representative, and the initial training set composed of representative samples is a prerequisite for training a high-precision benchmark classifier, and it can also speed up the active learning process more effectively.

A method based on clustering or distance similarity measurements is a common method for selecting

representative examples. K -medoids form an initial training set; hierarchical clustering by example selection, K -means, and other measures have accelerated the process of active learning to varying degrees. The classification surface of the benchmark classifier is not far from the real classification surface from the beginning, avoiding the situation wherein the classification surface stays in the wrong direction for a long time. However, K -medoids construct the initial training set and hierarchical clustering sample selection is more suitable for image processing. The K -means method requires a manual setting of K , which leads to inaccurate algorithms.

As described in the method section, in this paper, we aimed to compare the difference between random initiation and selection using k -means++.

Active learning starts with training the classifier on initial training data. As with most work, the initial training data are generated by random selection. In the paper, we introduced an additional method called k -means++. The k -means++ algorithm can select the first K to initiate data points, which can represent the distribution of the whole data set, as described in the method section. The results are shown in Figure 5, where we observe that there is a difference between random and k -means++. The initial training data size, which is 3000 in our study, is not large enough, which makes the difference between the two distributions not as huge as we expected. Another rational reason could be that only the least confidence was incorporated, which may introduce bias into this experiment. For the next steps of active learning in this paper, k -means++ is adapted to initiate the training data. However, in general, the accuracy, recall rate, and F_1 score of data initialization with the K -means++ method are better than random initialization. As a processing feature, in our convolutional autoencoder, the dimensionality of the hidden space in the hidden layer is assumed to be 25, which was shown to be sufficient to represent the original signal of 250 length [5, 33]. We have already introduced it in the section on convolutional autoencoders.

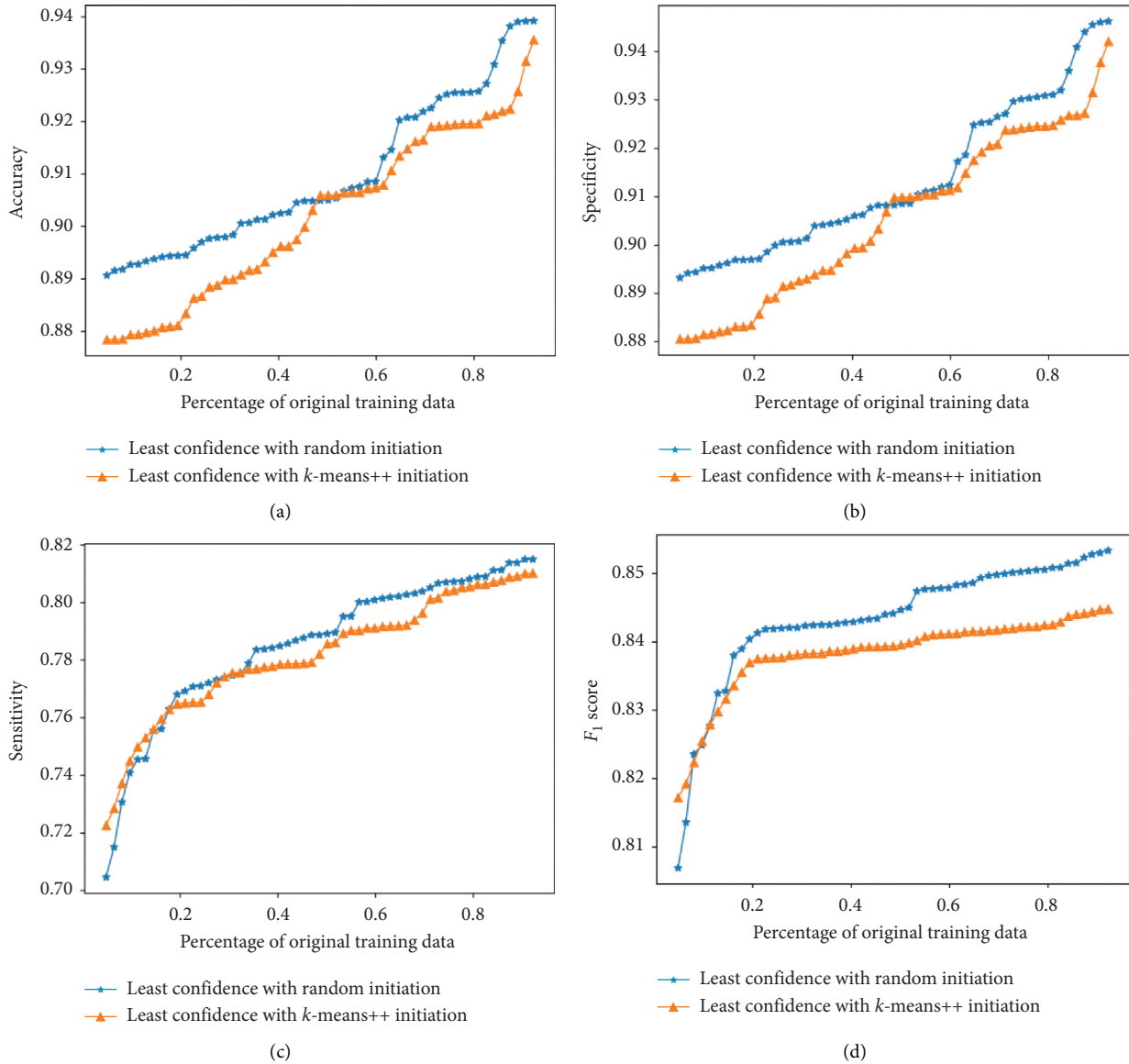


FIGURE 5: Performance comparison of different data initialization methods.

As the most important part of active learning, different selection strategies are well investigated in this experiment. As described in the method section, the three most classical strategies based on uncertainty sampling have been tested in this experiment.

Uncertainty sampling is one important aspect in the design of selection strategies. In this paper, we investigated three classic uncertainty sampling methods: least confidence sampling, margin sampling, and entropy sampling. Besides, random sampling was also included as a control approach. The performance of each method is reported in Figure 6. One thing worth noting is that there is an imbalance problem in both training and testing data, in which normal sinus rhythm has a higher prevalence. Instead of being measured by accuracy, the F_1 score is more appropriate in this situation. We can observe that the least confidence and margin sampling have similar performance across all

metrics. In terms of F_1 score and sensitivity, the least confidence and margin sampling have the best performance than random sampling and entropy sampling. This phenomenon demonstrates that selecting a more informative sample can help achieve a better performance than random selection. Among all the four selection strategies, entropy sampling has the least performance and is even worse than random selection. The most plausible reason is that entropy sampling is more vulnerable to multiclass classification. From a theoretical perspective, the least confidence and margin sampling only focus on the classes with the best or the second-best prediction probability. However, the entropy sampling takes care of all the possible classes, which is a huge imbalance in our data set. This problem can be further alleviated by narrowing down the task from multiclass classification to binary because these three strategies were proven to be theoretically equal in the binary tasks.

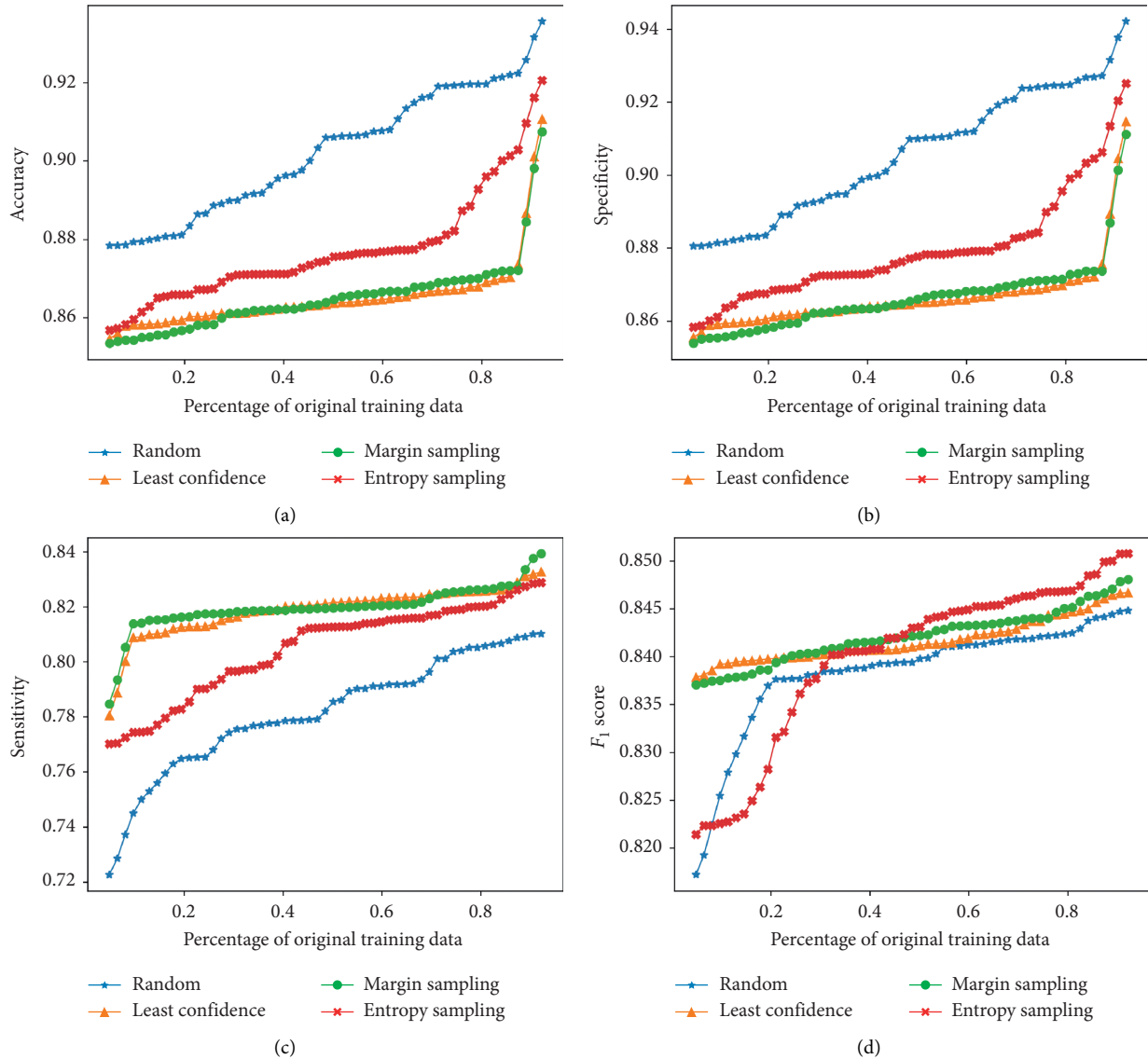


FIGURE 6: Performance comparison between different uncertainty selection strategies.

In addition to the uncertainty sampling, two geometry-based selection strategies were also investigated in this paper, k -center-greedy and representative cluster sampling.

Another important aspect of the active learning selection strategy is diversity. We studied two different approaches, k -center-greedy and its advanced version, representative clustering sampling, as described in the method section. The results are reported in Figure 7. We can observe that the representative clustering sampling outperforms k -center-greedy in all the metrics, which is expected. However, both methods are sampling K points that can represent the whole original distribution. The k -center-greedy selects points from all the data sets, which makes it more vulnerable to the outlier points. However, the representative clustering sampling would calculate the margin of each class and only select samples near the margins, by which the outlier problem could be heavily alleviated.

Uncertainty and diversity are two important aspects that active learning methods are trying to capture. In this paper, we studied five different selection strategies that cover both uncertainty and diversity. In Table 5, we list the performance of each strategy at the level of half of the original data being trained.

As discussed before, the ultimate goal of active learning is trying to reduce human annotation effort by keeping a similar performance at the same time. In this paper, we compared all the applied strategies trained on half the size of the original data set with the performance of the same classifier trained on the full data set. The results are displayed in Table 5. The best F_1 score is achieved with the least confidence, which is even better than training on the full data set, which demonstrates that active learning can help reduce human annotation without compromising on the performance. In terms of sensitivity, the classifier trained on the full set has the

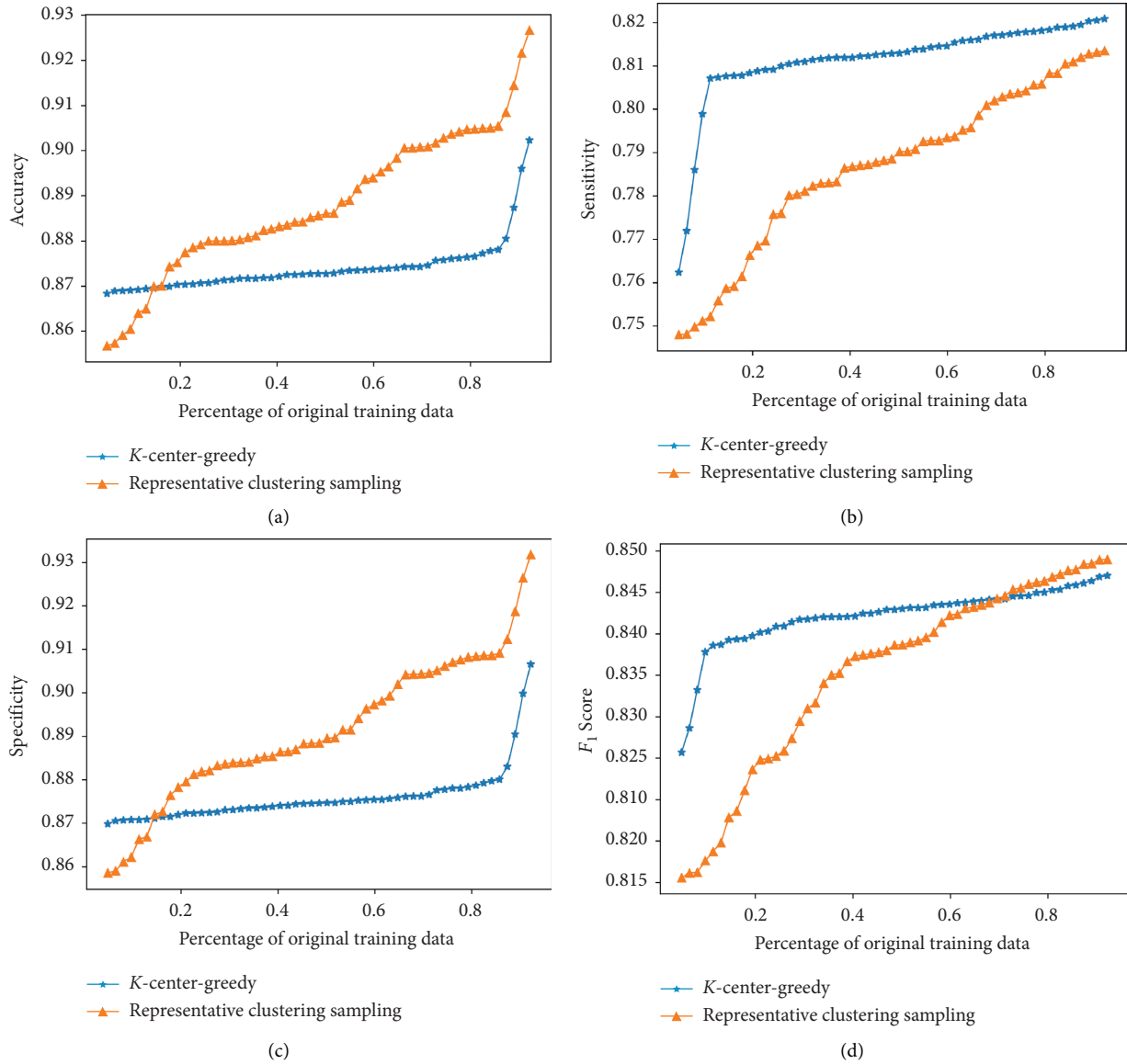


FIGURE 7: Performance comparison between different diversity selection strategies.

TABLE 5: Each strategy was trained to perform at the level of half of the original data.

	Random	Least confidence	Margin sampling	Entropy sampling	<i>K</i> -center-greedy	Rep-cluster	Train on the full data set
Accuracy	0.9068	0.8621	0.8485	0.8738	0.8648	0.8852	0.9385
Sensitivity	0.786	0.8470	0.8232	0.77267	0.81279	0.786	0.7704
Specificity	0.9106	0.8626	0.8493	0.8769	0.8663	0.8883	0.9562
F ₁ score	0.8237	0.8547	0.8360	0.8215	0.8387	0.8341	0.8533

lowest sensitivity value. One possible reason is that there are too many positive samples in the data set and too few negative samples. This reason is also proven by the highest specificity achieved by the full set training. This phenomenon indicates that the classifier is better at detecting normal sinus rhythms than PVCs.

The fundamental idea behind active learning is to find out the most informative samples from the original training data, which can help the classifier achieve similar

performance with much less human effort involved. In the study, as the comparison factor, we first train our model on the whole training data set to be the reference for the downstream experiment. Following the same strategy described above, we can get similar results as [5]. The results are shown in Tables 6 and 7. Table 6 shows the representation of the two methods in the whole data set, and Table 7 shows the confounding matrix of the two methods on different data sets. As shown in Table 6, our accuracy and

TABLE 6: The performance of our method in the whole training data.

		Accuracy	PVCs sensitivity	PVCs specificity	PVCs F_1 score
Full database	Work at [5]	98.43	85.64	98.90	0.9179
	Our approach	96.49	81.50	97.56	0.8881
Test set	Work at [5]	87.80	86.65	88.09	0.8736
	Our approach	93.85	77.04	95.62	0.8533

TABLE 7: Confusion matrix of two methods on different data sets.

	Full data set				Test data set			
	Work at [5]		Our approach		Work at [5]		Our approach	
	True normal	True PVC	True normal	True PVC	True normal	True PVC	True normal	True PVC
Detected normal	50483	299	93734	1266	50542	810	49943	1266
Detected PVC	559	3334	2283	5595	501	2823	2283	4249

specificity are both higher than the existing algorithms, and the F_1 score is not far from the existing algorithm. The experimental results show that our algorithm can indeed play a significant role in PVCs detection of small samples.

5. Discussion

Since active learning engages with a large number of unlabeled samples and a small number of labeled samples, its advantages are evident in the advantages of traditional supervised learning. This paper conducts research on MIT's ECG data set. Research findings show that active learning techniques can effectively reduce the number of high-quality training samples required to build a classifier. On the basis of not affecting the generalization performance of the classifier, it can effectively reduce the burden of human experts.

Nonetheless, the results of this research have certain limitations. There are still many problems to be solved in the design of sampling strategies, algorithm theory, and practical applications. First, from the perspective of the algorithm architecture of this article, for new tasks how to select new instances and label instances and which features are selected to be labeled by human experts, in order for a highly versatile selection strategy.

These problems are worthy of further study, and secondly, whether the classifier proposed in this article can be replaced by a deep learning algorithm remains to be further studied. An in-depth study of feature selection [34] and classification algorithms will effectively improve the accuracy of recognition. Consequently, improvements to feature algorithms are also our next research focus.

6. Conclusions

In the medical research area, massive human annotation efforts are necessary in order to achieve higher detection performance from supervised machine learning algorithms. Active learning is a promising technique that utilizes the output of a trivial classifier to select the most informative

samples for the request of annotation. By active learning, similar performance could be achieved with much less human annotation.

In this work, two main aspects of selection strategy, uncertainty and diversity, were well explored in the task of PVCs detection on beat-level ECG data. One convolutional autoencoder was trained to extract the features automatically. These data-driven features are then fed into a random forest classifier. During each iteration of active learning, the same algorithm is trained on the updated training data. It can be seen from the experimental results that the F_1 value of the least confidence sampling algorithm is better than other algorithms. In addition, by comparing different active learning methods trained on half of the original data size with the same classifier trained on the full set, the performance of least confidence is still better than the full set one, which demonstrates that active learning works perfectly for the task.

Overall, the experimental effect and the sensitivity of our method are higher than the existing algorithms, demonstrating the superiority of active learning in PVC-detecting techniques. In F_1 , active learning and existing methods can fundamentally improve the efficiency of relevant work. We will continue to improve our feature engineering in the coming period, deeply studying the distribution of data as we design our self-encoder, and working to ensure that the impact of active learning exceeds that of the existing work.

Data Availability

The data sets used in this paper to produce the experimental results are publicly available. ECG recordings can be downloaded from <https://www.physionet.org/content/mitdb/1.0.0/>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant no. 62002077), in part by the China Postdoctoral Science Foundation (Grant no. 2020M682657), and in part by Guangdong Basic and Applied Basic Research Foundation (Grant no. 2020A1515110385).

References

- [1] C. L. Stanfield and W. J. Germann, *Principles of Human Physiology*, Pearson Benjamin Cummings, London, UK, 2008.
- [2] H. J. L. Marriott, G. S. Wagner, and D. G. Strauss, *Marriott's Practical Electrocardiography*, Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2014.
- [3] J. S. Geddes and H. R. Warner, "A PVC detection program," *Computers and Biomedical Research*, vol. 4, no. 5, p. 493, 1971.
- [4] M. S. Manikandan, B. Ramkumar, P. S. Deshpande, and T. Choudhary, "Robust detection of premature ventricular contractions using sparse signal decomposition and temporal features," *Healthcare Technology Letters*, vol. 2, no. 6, pp. 141–148, 2015.
- [5] M. Gordon and C. Williams, "PVC detection using a convolutional autoencoder and random forest classifier," in *Proceedings of the Pacific Symposium on Biocomputing (PSB)*, pp. 42–53, Fairmont Orchid, Hawaii, January 2019.
- [6] P. Bachman, A. Sordoni, and A. Trischler, "Learning algorithms for active learning," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 477–486, Sydney, Australia, August 2017.
- [7] B. Settles, *Active Learning Literature Survey*, University of Wisconsin-Madison Department of Computer Sciences, Madison, WI, USA, 2009.
- [8] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian active learning with image data," in *Proceedings of the International Conference on Machine Learning (PMLR)*, Sydney, Australia, August 2017.
- [9] S. Sinha, S. Ebrahimi, and T. Darrell, "Variational adversarial active learning," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5972–5981, Seoul, Korea, October 2019.
- [10] L. Yang, "Suggestive annotation: a deep active learning framework for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Quebec City, QC, Canada, September 2017.
- [11] D. Yoo and I. S. Kweon, "Learning loss for active learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 93–102, Long Beach, CA, USA, May 2019.
- [12] V. Xia, N. Jaques, S. Taylor, S. Fedor, and R. Picard, "Active learning for electrodermal activity classification," in *Proceedings of the 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–6, IEEE, Philadelphia, PA, USA, December 2015.
- [13] G. Wang, C. Zhang, Y. Liu et al., "A global and updatable ECG beat classification system based on recurrent neural networks and active learning," *Information Sciences*, vol. 501, 2019.
- [14] W. H. Beluch, T. Genewein, A. Nurnberger, and J. M. Kohler, *The Power of Ensembles for Active Learning in Image Classification*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (VBPR)*, pp. 9368–9377, Salt Lake City, UT, USA, June 2018.
- [15] H. Ranganathan, H. Venkateswara, S. Chakraborty, and S. Panchanathan, "Deep active learning for image classification," in *Proceedings of the 2017 IEEE International Conference on Image Processing*, Beijing, China, September 2017.
- [16] G. Yang, Q. Luo, Y. Yang, and Y. Zhuang, "Deep Learning and Machine Learning for Object Detection in Remote Sensing Images," in *Proceedings of Signal and Information Processing, Networking and Computers. (ICSINC 2017)*, pp. 249–256, Chongqing, China, December 2017.
- [17] S. Priya, S. Singh, S. Kumar Dandapat, K. Ghosh, and J. Chandra, "Identifying Infrastructure Damage during Earthquake Using Deep Active Learning," in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '19)* pp. 551–552, New York, NY, USA, August 2019.
- [18] S. Sun, P. Zhong, H. Xiao, and R. Wang, "An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1074–1088, 2015.
- [19] P. H. Gosselin and M. Cord, "Active learning methods for interactive image retrieval," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1200–1211, 2008.
- [20] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM-based auto-encoder model for ECG arrhythmias classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1232–1240, 2019.
- [21] M. Jangra, S. K. Dhull, and K. K. Singh, "ECG arrhythmia classification using modified visual geometry group network (mVGGNet)," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 3, pp. 3151–3165, 2020.
- [22] G. Wang, C. Zhang, Y. Liu et al., "A global and updatable ECG beat classification system based on recurrent neural networks and active learning," *Information Sciences*, vol. 501, pp. 523–542, 2019.
- [23] Y. Xia and Y. Xie, "A novel wearable electrocardiogram classification system using convolutional neural networks and active learning," *IEEE Access*, vol. 7, pp. 7989–8001, 2019.
- [24] J. Pestana, D. Belo, and H. Gamboa, "Detection of abnormalities in electrocardiogram (ECG) using deep Learning," in *Proceedings of 13th International Conference on Bio-inspired Systems and Signal Processing (BIOSSTEC 2020)*, pp. 236–243, Valletta, Malta, February 2020.
- [25] D. L. Lustgarten, G. Rajagopal, J. Reiland, J. Koehler, and S. Sarkar, "Premature ventricular contraction detection for long-term monitoring in an implantable cardiac monitor," *Pacing and Clinical Electrophysiology*, vol. 43, no. 5, pp. 462–470, 2020.
- [26] P. Novotna, T. Vicar, M. Ronzhina, J. Hejc, and J. Kolarova, "Deep-Learning Premature Contraction Localization in 12-Lead ECG from Whole Signal Annotations," in *Proceedings of the 2020 Computing in Cardiology Conference*, pp. 1–4, Rimini, Italy, September 2020.
- [27] M. Shafiq, Z. Tian, M. Bashir, and X. Du, "Corrauc: a malicious Bot-IoT traffic detection method in iot network using machine learning techniques," *IEEE Internet of Things Journal*, vol. 99, p. 1, 2020.
- [28] M. Shafiq, Z. Tian, Y. Sun, X. Du, and M. Guizani, "Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city," *Future Generation Computer Systems*, vol. 107, pp. 443–442, 2020.
- [29] G. Dizaji, Kamran, A. Herandi, C. Deng, W. Cai, and H. Huang, "Deep clustering via joint convolutional

- autoencoder embedding and relative entropy minimization,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5736–5745, Venice, Italy, October 2017.
- [30] S. Pan, “Adversarially regularized graph autoencoder for graph embedding,” in *Proceedings of the IJCAI International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, July 2018.
- [31] Y.-X. Li, C. L. Tan, X. Ding, and C. Liu, “Contextual post-processing based on the confusion matrix in offline handwritten Chinese script recognition,” *Pattern Recognition*, vol. 37, no. 9, pp. 1901–1912, 2004.
- [32] T. Xia and X. Chen, “A discrete hidden Markov model for SMS spam detection,” *Applied Sciences*, vol. 10, no. 14, p. 5011, 2020.
- [33] B. M. Mathunjwa, Y.-T. Lin, C.-H. Lin, M. F. Abbod, and J.-S. Shieh, “ECG arrhythmia classification by using a recurrence plot and convolutional neural network,” *Biomedical Signal Processing and Control*, vol. 64, Article ID 102262, 2021.
- [34] M. Shafiq, “Data mining and machine learning method for sustainable cities traffic classification,” *Sustainable Cities and Society*, vol. 60, Article ID 102177, 2020.