Active Balancing Mechanism for Imbalanced Medical Data in Deep Learning–Based Classification Models

HONGYI ZHANG and HAOKE ZHANG, Xiamen University of Technology
SANDEEP PIRBHULAL, Chinese Academy of Sciences
WANQING WU, Sun Yat-Sen University
VICTOR HUGO C. DE ALBUQUERQUE, University of Fortaleza

Imbalanced data always has a serious impact on a predictive model, and most under-sampling techniques consume more time and suffer from loss of samples containing critical information during imbalanced data processing, especially in the biomedical field. To solve these problems, we developed an active balancing mechanism (ABM) based on valuable information contained in the biomedical data. ABM adopts the Gaussian naïve Bayes method to estimate the object samples and entropy as a query function to evaluate sample information and only retains valuable samples of the majority class to achieve under-sampling. The Physikalisch Technische Bundesanstalt diagnostic electrocardiogram (ECG) database, including 5,173 normal ECG samples and 26,654 myocardial infarction ECG samples, is applied to verify the validity of ABM. At imbalance rates of 13 and 5, experimental results reveal that ABM takes 7.7 seconds and 13.2 seconds, respectively. Both results are significantly faster than five conventional under-sampling methods. In addition, at the imbalance rate of 13, ABM-based data obtained the highest accuracy of 92.23% and 97.52% using support vector machines and modified convolutional neural networks (MCNNs) with eight layers, respectively. At the imbalance rate of 5, the processed data by ABM also achieved the best accuracy of 92.31% and 98.46% based on support vector machines and MCNNs, respectively. Furthermore, ABM has better performance than two compared methods in $F1$-measure, G-means, and area under the curve. Consequently, ABM could be a useful and effective approach to deal with imbalanced data in general, particularly biomedical myocardial infarction ECG datasets, and the MCNN can also achieve higher performance compared to the state of the art.

CCS Concepts: • **Computing methodologies** → *Machine learning*; • **Applied computing** → *Life and medical sciences*;

1 INTRODUCTION

Today, deep learning has achieved great success in many fields, such as anomaly detection in medical images [12], target monitoring and recognition [1, 32, 48], and feature learning [45]. However, most models are built on balanced data, and their performance will be limited when they face imbalanced data. Additionally, imbalanced data are widely found in the medical field, especially in biomedical data. For instance, myocardial infarction (MI) is one of the most deadly cardiovascular diseases in the world, because the blood supply to the heart is blocked by thrombosis or atherosclerosis, resulting in tissue death. For many reasons, MI electrocardiogram (ECG) data is more difficult to collect than normal ECG data, so it often causes imbalanced distribution of ECG data. Recently, many methods have been proposed, including machine learning and deep learning, for imbalanced MI ECG data classification. Baloglu eg al. [2] applied deep learning to classify MI ECG data. They built an end-to-end convolutional neural network (CNN) model with 10 layers for 12-lead MI ECG data and achieve high accuracy. Han and Shi [9] took the energy entropy based on wavelet transform coefficients and morphological features as feature vectors. In addition, the dimensions of these features would be reduced by principal component analysis (PCA) and then classified using the support vector machine (SVM). Sadhukhan et al. [29] proposed a method based on Fourier transform for MI feature extraction and threshold and logistic regression to classify MI ECG. Although all of these can achieve higher precision to some extent, they do not take into account the effects of data imbalance. In imbalanced data, the predictive model priority increases for the majority class and reduces the classification priority for the minority class [11], and this problem has a serious impact on the reliability of ECG disease predictions. Thus, there is an urgent need to handle the problem of imbalanced biomedical data in classification.

Several researchers have designed different techniques to handle data imbalance problems. These techniques are divided into cost-sensitive learning [4, 16, 27, 43, 44], ensemble methods, and resampling methods. Imbalanced data processing based on cost-sensitive learning is primarily achieved by changing some existing algorithms through cost-based learning, incorporating more punishment into the minority class samples, and small-scale penalties in the majority class samples. These methods have high computational efficiency [8], whereas the actual costs are usually varied and unknown from case to case and must determine the cost factor by following the actual data class [41]. Meanwhile, the classification models may lead to over-fitting during training. Ensemble methods [5, 26, 29, 40] primarily adopt the technology of multiple classifiers to improve the reliability of a single classifier and predict the sample class through the classification results of all classifiers, such as classifier-based voting or averaging all classifier results. Although these techniques will not give up every sample in the imbalanced data, a well-known problem with these ensemble techniques is a long processing time.

Resampling methods are mainly divided into over-sampling [17, 18, 21, 31] and under-sampling [13, 15, 35], and two over-sampling methods are widely used. One is random over-sampling, which requires less time to deal with minority class samples by copying minority class samples into the minority class. This method introduces duplicate samples into the minority class dataset, causing

the classification model to overfit, which severely affects the classification performance [13]. Another one is the synthetic minority over-sampling technique (SMOTE) proposed by Chawla et al. [3], which randomly selects many minority class samples to generate new synthetic samples between these selected samples and their corresponding k-nearest neighbor (kNN) samples. Nevertheless, this approach has some expensive computing time and overlapping samples. The problem of overlapping is due to the failure to consider the majority class samples and may limit the performance of the classification model. In general, the performance of over-sampling methods is worse than that of another resampling technique, namely under-sampling [6].

Under-sampling methods balance data by reducing the majority class data. The RaUS method is the simplest under-sampling method. This approach randomly removes a certain number of samples from the majority class data to achieve balance with minority class samples. This process takes less time, yet there is a risk of losing important data, affecting the classification performance. Various under-sampling approaches have been developed to address this problem. Mani and Zhang [19] developed four under-sampling methods based on the four versions of kNN. In addition, NearMiss-1 presents a method that selects the majority sample with the smallest average distance from the closest three minority samples. NearMiss-2 selects majority class samples when their average distances to the three farthest minority class samples are the smallest. NearMiss-3 selects a given number of the majority class samples nearest to each minority class sample. Most distant methods are opposite NearMiss-1. These four methods exclude the majority class samples based on the distance between samples. They mitigate the risk of reduced model classification performance. However, these methods need more time to calculate the distances between samples.

In a recent study, Yen and Lee [42] proposed a cluster-based under-sampling method. The method divides data into n clusters, calculates the ratio of majority to minority samples in each cluster, and deletes a certain number of majority category samples based on a sample ratio. Similarly, D'Addabbo and Maglietta [7] proposed a parallel selective sampling method. It divides the majority class samples into N subsets, with each subset and the minority class samples forming a dataset. Then, it applies Tomek links (TL) to reduce the majority class samples and repeats this step several times until the between classes are balanced. In the entire process, these N subsets are performed at the same time. Tasi [13] introduces a cluster-based and instance selection method, which uses clustering analysis to group similar samples in the majority class into "subclasses," whereas the instance selection filters out unrepresentative data samples from each subclass to achieve data balance. Although these methods are quicker, they might remove some important sample data. Therefore, it is worthwhile to study an alternative method that quickly deals with imbalanced MI ECG data and improves the reliability of the data classification.

Some different models based on active learning (AL) have also been used for addressing imbalanced data, which does not include MI ECG data. For instance, Zhang et al. [46] proposed an online asymmetric AL algorithm that uses asymmetric query strategies to process imbalanced data. However, this method tends to query for more negative data, which may result in poor training of positive data. Zhang et al. [47] adapted to maximize the sum of weighted sensitivity and specificity or minimize the weighted cost of misclassification to handle imbalanced data, but this method ignored the overall accuracy. Therefore, a new under-sampling method based on AL called the *active balancing mechanism* (ABM) is proposed for imbalanced ECG data in this article. Our main contributions are as follows:

- Considering the time consumption and sample importance, this work proposes an efficient ABM for processing imbalanced ECG data.
- An imbalanced ECG data processing method based on Gaussian naïve Bayes (GNB) and entropy is proposed for high efficiency while handling imbalanced ECG data.

- An eight-layer modified convolutional neural network (MCNN) with low time consumption and high performance of imbalanced ECG classification is constructed. A global average pooling layer is used to reduce training parameters and increase the training rate.

The rest of the article is organized as follows. Section 2 proposes ABM and describes it in detail. Section 3 presents the comparison of ABM with several conventional methods for examining performance. Section 4 concludes the article.

2 METHODOLOGY

Recent advancements in mobile and wearable devices have allowed researchers to passively collect real-time data on individuals without disrupting their daily lives [23, 24, 36–39]. On the one hand, this data can help researchers inform individuals of their risk profiles and enable clinicians and patients to make more informed care decisions. On the other hand, to improve the availability of a wearable system for long-term and real-time monitoring, it is important to propose a novel methodology for decreasing the influences of the imbalanced distribution of samples on the accuracy and reliability of a classifier, because imbalanced data make classifiers have high accuracies on majority class samples and low accuracies on minority samples. More importantly, the minority class samples are often abnormal disease data that are essential to disease prediction; once the prediction is wrong, it may lead to serious results. Therefore, imbalanced data must be processed before the application and need to be handled in a short time for real-time monitoring.

The well-designed methodology for imbalanced ECG data processing includes three parts: signal pre-processing derived from a wearable system, proposed ABM implementation, and disease predication (or decision). The signal pre-processing stage aims to remove noise interference and divide the ECG data into a single heartbeat sample based on the location of the R-peak. In the ABM stage, m samples are randomly selected from the majority class data, and N samples of the minority class constitute the initial dataset for training the GNB estimator. Then, the remaining M-m majority class samples are estimated by the trained estimator, and the entropy values of these samples are calculated from their estimates. Afterward, k samples with the largest entropy values are selected and added to the training set to retrain the estimator. The second and third steps are repeated K times until the two types of samples in the training set are equal. The last stage is application of the balance data using deep networks, such as disease predication. The overall pre-processing phase with data imbalance and performance evaluation is illustrated in Figure 1.

2.1 Dataset

In our work, the Physikalisch Technische Bundesanstalt (PTB) diagnostic ECG database was considered as an experimental dataset. This database contains 148 MI subjects and 52 normal patients between the ages of 17 and 87 years, with an average age of 55.5 years for men and 61.6 years for women, and each signal sampling frequency is 1,000 Hz. In addition, lead II ECG data are usually used for ECG classification; therefore, in our study, lead II was also used as experimental data.

In the pre-processing stage, we applied a wavelet transform for processing the raw ECG signal. The MI and normal ECG records are segmented according to the position of the R-peaks detected by the Pan Tompkins algorithm [22]. Each ECG sample consists of 220 interval points located at the left of the R-peak and 380 interval points to the right. There are 5,173 normal ECG samples and 26,654 MI ECG samples.

2.2 Proposed ABM

This section balances MI ECG data and normal ECG by adopting ABM to obtain balanced data between MI and normal ECG samples. A visualization process of imbalanced data processing using
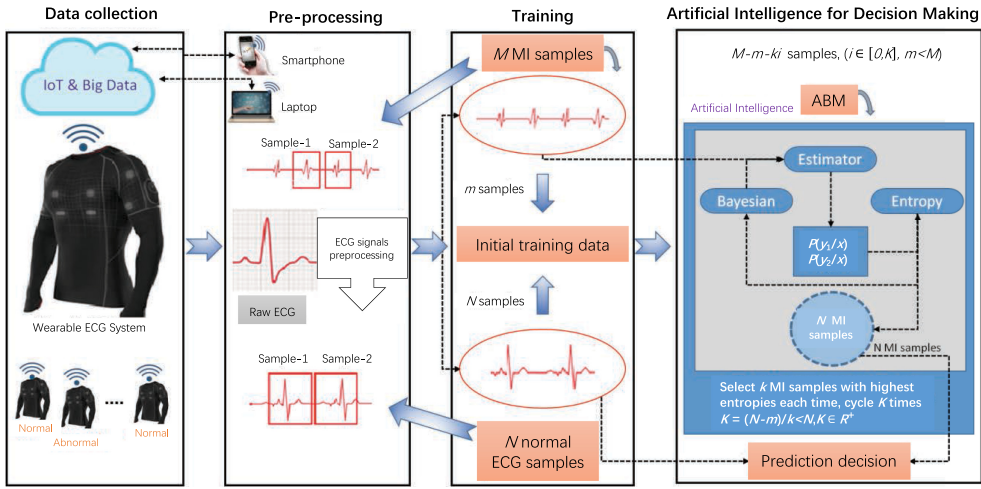
Fig. 1. The whole process of imbalanced MI data processing.
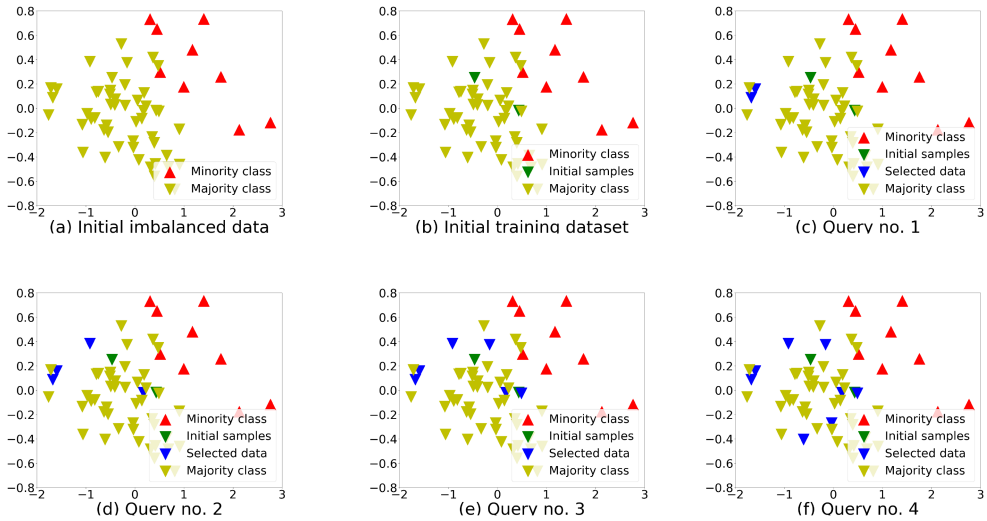
Fig. 2. Instance of an ABM-based under-sampling majority class. (a) Two classes of initial data. (b) Initial training dataset was used to train a model to measure the uncertainty of the yellow samples. (c–f): The process of obtaining the balance data based on ABM.

ABM is shown in Figure 2. Figure 2(a) represents an imbalanced dataset containing two class samples: red denotes the minority class samples, and yellow denotes the majority class. During the process of imbalanced data processing, we first randomly obtain a certain number of red and green samples from the two class samples as the initial training data, respectively, as shown in Figure 2(b). An estimator is then trained to calculate the probability of the remaining majority class samples. After that, the entropy of each sample is calculated; two majority class samples with high entropy values are selected and added to the training set to update training data, as shown in Figure 2(c). The second and third steps are repeated one time to get to a new training dataset, which is shown in Figure 2(d); after the second repeating, we obtain another training dataset, as shown in

Figure 2(e); the third repeating achieves a balanced dataset, which is shown in Figure 2(f). At the same time, the main purpose of this process is to select samples from the entire sample space, not just high-entropy samples, such as boundary samples. Figure 2 shows that when the two types of samples are almost balanced, the selected samples are mainly at the boundary.

To build an ABM model, there is an urgent need for an estimator object and a query-based strategic function. In our study, we compared GNB to some generative and discriminative approaches, including linear discriminant analysis, decision tree, random forest, and quadratic discriminant analysis, among others. Linear and quadratic discriminative analyses take more time than GNB, and the classification accuracy of balanced data achieved by these two methods is lower than that of GNB. Decision tree and random forest take even a longer time than GNB consumption. Thus, GNB is known as an estimator, whereas entropy is considered a query strategy function.

Moreover, GNB is one of the naïve Bayes (NB) models including a set of supervised learning algorithms based on Bayes' theorem. It is also a generation model and can converge to a real model more quickly than other methods, such as decision trees and random forest. Furthermore, GNB presumes that the data features are continuous, which is different from NB, since NB assumes that the features of the samples remain independent of each other, so when the sample features are continuous, NB is not suitable. GNB uses a Gaussian density estimate and assumes that the conditional distribution satisfies the Gaussian distribution, checking that the sample features follow the Gaussian distribution. Therefore, GNB is applied to process ECG data in our study. To learn about the posterior probability of sample data, the conditional probability must be carefully interpreted. The GNB model is carefully interpreted as shown in Equation (1):

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - u_y)^2}{2\sigma_y^2}}, \tag{1}$$

where the parameters u_y and σ_y^2 represent the mean and variance, respectively. They are estimated using the maximum likelihood method with $P(x_i|y)$, which is the conditional probability. In this section, $P(x_i|y_j)$ indicates the probability of x_i belonging to the y_j class. All MI ECG samples are calculated with GNB without initially trained MI ECG samples. This is the first step to balance the two classes of data.

After calculating the posterior probability of the samples, we need to obtain the sample information for MI ECG beats. Entropy is used to predict uncertainty in a sample. It is the mathematical expectation of the self-information of random variables. When the data produces low probability values, it indicates that the event itself contains more uncertainty, namely the low probability event contains more information compared to the data with high probability. The amount of information for each event becomes a random variable. The Shannon entropy of probabilities p_x is defined in Equation (2):

$$H(x) = -\sum_k p_x \log p_x, \tag{2}$$

where p_x represents the different probability values that the sample can have regarding K classes. In addition, H is the sample entropy, which is greater than the respective values. Higher entropy indicates samples with more information to predict the training entities. Here, p_x represents the probability of the x samples in association with the normal or MI ECG beats, whereas $H(x)$ represents the entropy of the x MI samples.

After obtaining sample entropies, we choose some MI ECG beats with higher H and add these with high-entropy samples to prepare the training data. Repeating all three steps to balance the datasets is defined next.

ALGORITHM 1: Balanced data based on ABM

Dataset: $D_{maj} = \{x_i\}_1^M$, $i \in [1, M]$, represents the majority class data; $D_{min} = \{x_j\}_1^N$, $j \in [1, N]$, indicates the minority class samples. M is the number of the majority class samples, N is the number of the minority class samples.

Input: $D_{Train} = D_{min} + D_{training}$, represents the initial training data; $D_{training} = \{x_a\}_1^m$, $a \in [1, m]$ and $m < M$, is the initial training data selected from the majority class samples; $D_{test} = D_{maj} - D_{training}$, indicates the test set.

Output: $D_{balance}$ represents balance data.

Imbalanced data processing:
Iteration parameter setting: $K = N - m/k$, k is the number of samples per update.
Using D_{Train} to train Bayes model and calculate sample entropy of D_{test}
For each $i \in [1, K]$
    Select k highest entropy samples from D_{test},
    Update D_{Train} with these k samples,
    Retrain Bayes model based on D_{Train} and calculate sample entropy of D_{test}.
End
$D_{balance} = D_{Train}$.
Balance data: $D_{balance}$.

Suppose that $D = \{x_i, y_i\}_{i=1}^T$ is a set of ECG beats, $\{x_i\}_{i=1}^N$ represents N ECG samples, and $\{y_i\}_{i=1}^N, y_i \in (1, k)$ are the relevant labels of N ECG beats. In addition, $D_{Mj} = \{x_i, y_i = 0\}_{i=1}^M$ indicates the MI ECG sample dataset, with M MI ECG beats, $y_i = 0$ reveals the class label, $D_{Mi} = \{x_i, y_i = 1\}$, n is the normal ECG beats, and $y_i = 1$ is its label. Meanwhile, $T = M + n$. The implementation procedure of the proposed ABM is described in the pseudocode of Algorithm 1.

2.3 Classification of the Proposed Method

We applied SVM and MCNN classifiers to evaluate the performance of the ABM, RaUS, and NearMiss-1 methods. The SVMs are commonly used methods, especially in binary classification, whereas the MCNN is an improved model of the CNN. Moreover, the MCNN needs to train fewer model parameters.

2.3.1 Support Vector Machines. The SVM is a binary classification model with the largest interval of heterogeneous support vectors defined in the feature space. Its key purpose from the learning cycle viewpoint is to maximize the spacing of heterogeneous support vectors. The SVM can be expressed as a problem-solving convex quadratic programming and contains multiple learning models. We merely focused on the linear SVMs that were used as the classifier.

2.3.2 Modified Convolutional Neural Network. The MCNN is a modified model based on CNN. The global average pool layer replaces the fully connected layers in smaller training parameters. The MCNN comprises three convolutional layers, three max-pooling layers, a global average pool layer, and a classification layer. The three-layer convolution layer is in the first layer, and the third and fifth layers function as feature extractors. The filter sizes of the three convolutional layers are 3, 5, and 5, respectively. The activation functions for all convolutional layers are set to Relu, and the strides of all convolutional layers are set to 1. Max-pooling layers are used to reduce computation for MCNNs. Max-pooling layers are located on the second, fourth, and sixth layers, respectively. All pool sizes and strides are set to 2. The seventh layer is the global average pooling layer, and the classification layer is located on the eighth layer. The Softmax function is selected as the classification function of the last layer, as shown in Table 1. The hyper-parameters of the model are set as follows: momentum was set to 0.5, the regularization parameter was 0.2, and the

Table 1. Detailed Summary of the MCNN Structure

Layers	Kernel Size	Output Shape	Stride
Convolution layer	120	(482, 3)	1
Max-pooling layer	2	(241,3)	2
Convolution layer	22	(220, 5)	1
Max-pooling layer	2	(110, 5)	2
Convolution layer	9	(102, 5)	1
Max-pooling layer	2	(51,5)	2
Global average pooling layer	—	2	—
Classification layer	—	2	—

learning rate was set to 2×10^{-3}. Meanwhile, the optimizer method is Adam at 150 epochs in our experimental setup.

2.4 Data Evaluation Parameters

During the data evaluation stage, we first compared the time consumption of ABM to six existing under-sampling methods. Second, the MCNN and SVM were adopted to classify the balanced data method, taking less time. To evaluate the classification results, four other prediction performance approaches with accuracy, G-mean [30], area under the curve (AUC) [25], and Macro-$F1$ [20] were adopted.

2.4.1 Accuracy. Given a dataset, accuracy represents the ratio of the correct classification data to the total data as given in Equation (3):

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{3}$$

where TP represents the correct number of positive class estimations, FN is the number of negative class prediction errors, TN is the correct number of negative class predictions, and FP reveals the number of positive class prediction errors.

2.4.2 G-mean. The G-mean evaluates the classification results of two class samples at the same time. If one class of samples has a high accuracy and the other has a low accuracy, a low G-mean will be achieved as shown in Equation (4):

$$G_{mean} = \sqrt{\frac{TP}{TP + FN} + \frac{TN}{TP + FP}}. \tag{4}$$

2.4.3 Macro-F1. Macro-$F1$ is the performance metric that is widely used for analyzing the imbalanced sample performance as given in Equation (5):

$$F1_{macro} = \frac{2 \times P_{macro} \times R_{macro}}{P_{macro} + R_{macro}}, \tag{5}$$

$$P_{macro} = \frac{1}{n} \sum_{i=1}^{n} P_i, \tag{6}$$

$$R_{macro} = \frac{1}{n} \sum_{i=1}^{n} R_i, \tag{7}$$

where $P_i = \frac{TP_i}{TP_i + FP_i}$ and $R_i = \frac{TP_i}{TP_i + FN_i}$ in each test, and n represents the test times.

2.4.4   Area Under the Receiver Operating Curve (AUC). The abbreviation AUC stands for the area under the receiver operating curve (ROC) and is one of the performance indicators, whereas the classification model is evaluated based on the classification results of positive and negative samples.

3   EXPERIMENTAL RESULTS AND DISCUSSION

In this study, experimental results are deployed in Python with an Integrated Development Environment known as PyCharm and trained our model on a workstation with two Intel Core i7-6700s, 3.40-GHz processors, and 16 GB of RAM. We examined our approach using the PTB diagnostic ECG database. The database contains a 12-lead ECG for 148 MI subjects and 52 normal subjects. Each ECG record is digitized at 1,000 samples per second. However, we merely adopt the lead II ECG datasets using the proposed work. Each ECG sample contains 601 sampling points. Moreover, our results are tested by 5,173 normal ECG samples and 26,654 MI ECG samples. In the meantime, the two class samples are adopted with different patterns. The different data samples are collected to continue the experimental imbalance ratios to verify the performance of the proposed ABM. Experimental datasets have imbalance ratios of 13 and 5, respectively.

3.1   Experimental Setup

In our research work, ABM is compared to other conventional under-sampling methods with distinct imbalances. The samples' imbalance ratios between MI ECG beats and normal ECG beats are 5 and 13, respectively. Each experiment was repeated four times, and the average of the four experiments is taken for improving the reliability. The SVM and MCNN classifiers are adopted to examine the under-sampling approaches. During the evaluation phase, the ECG samples are categorized into 10 equal parts to obtain more reliable performance. Moreover, 90% of the ECG samples are used as training data, whereas the remaining 10% are considered to verify the reliability of the under-sampling methods.

To obtain a sample set with an imbalance ratio of 13, we randomly selected 2,000 samples from 5,173 normal ECG samples and used 26,654 MI ECG data to construct the imbalanced dataset. Afterward, 2,000 normal ECG samples and 200 MI samples were selected randomly from a 26,654-ECG dataset to construct the initial training data. Finally, we repeated this nine times; each time, we obtained 200 samples with high entropy and low confidence metrics for the MI samples from the 26,454-MI dataset, and a total of 1,800 MI ECG samples were added to the training dataset to balance the datasets between different classes.

To construct the sample data with an imbalance ratio of 5, all normal ECG samples and 26,654 MI ECG samples were used to build the imbalanced data, and 1,173 MI ECG samples and all normal ECG samples to build initial training data were selected. Thereafter, the proposed ABM method was repeated 10 times, and 400 samples with high entropy were achieved each time to form 4,000 MI ECG data. All of these samples were added to the initial training dataset to achieve a data balance between classes.

Finally, to analyze the performance of ABM, we applied the MCNN and SVM methods to classify these balanced datasets. At the same time, the proposed ABM and the two less-time-consuming methods in the six under-sampling methods were observed in terms of classification accuracy, G-mean, AUC, and Macro-*F*1 to verify the effectiveness of our proposed method.

3.2   Comparative Analysis

In our work, we first compared the time consumption of ABM to six other under-sampling techniques: (1) TL, (2) cluster centroids, (3) condensed nearest neighbors, (4) RaUS, (5) NearMiss-1, and (6) ALLKNN, as shown in Table 2. The results proved that ABM consumed more time than the

Table 2. Time Consumption Comparison of Seven Methods at Different Imbalance Ratios

Imbalance Ratio	ABM (s)	TL [33] (s)	CONN [10] (s)	RaUS [14] (s)	NearMiss-1 [19] (s)	ALLKNN [34] (s)
13	7.7	87.5	122.1	1.3	32.9	290.2
5	13.2	112.3	183.9	2.2	53.7	360.1

CONN, Condensed nearest neighbors.

Table 3. Average Accuracy, G-mean, and Macro-*F*1 at Imbalance Ratio 13

Classification Methods	Under-Sampling Methods	Accuracy (%)	F1_macro	G_mean
SVM	ABM	**92.23**	**0.934**	**0.922**
	RaUS	89.38	0.885	0.884
	NearMiss-1	90.05	0.912	0.902
MCNN	ABM	**97.252**	0.977	**0.987**
	RaUS	95.13	0.922	0.916
	NearMiss-1	97.30	**0.981**	0.975

Table 4. Average Accuracy, G-mean, and Macro-*F*1 at Imbalance Ratio 5

Classification Methods	Under-Sampling Methods	Accuracy (%)	F1_macro	G_mean
SVM	ABM	**92.31**	**0.927**	**0.942**
	RaUS	90.20	0.893	0.911
	NearMiss-1	91.64	0.917	0.920
MCNN	ABM	**98.46**	**0.964**	**0.984**
	RaUS	95.92	0.923	0.928
	NearMiss-1	97.81	0.956	0.974

RaUS method but far less than the other five under-sampling methods. Since time loss is an important factor in practical applications, we only compared ABM to the NearMiss-1 and RaUS methods in our work, as both approaches took less time. Meanwhile, we evaluated the performance of the three under-sampling methods by considering four major aspects: accuracy, G-mean, Macro-*F*1, and AUC.

Tables 3 and 4 depict the comparative analysis of the average accuracy, average G-mean, and average Macro-*F*1 with different imbalance ratios for the ABM, NearMiss-1, and RaUS methods. Moreover, Table 3 examines the accuracy, G-mean, and Macro-*F*1 at the imbalance ratio of 13. The accuracy, G-mean, and Macro-*F*1 at the imbalance ratio of 5 are mentioned in Table 4. Moreover, Table 3 presents the samples with the imbalance ratio of 13, and convincing experimental results are highlighted with underlined bold. The results show the average accuracy of the data classification based on SVMs and MCNNs, and our approach obtained the average accuracies of 92.23% and 97.52%, respectively, which are better than the accuracies of the RaUS and NearMiss-1 methods. Meanwhile, the proposed ABM had an average Macro-*F*1 of 0.934 and 0.977, respectively, whereas 0.934 was the highest value compared to RaUS and NearMiss-1 in the SVM, but 0.977 was lower than NearMiss-1 in the MCNN. The average G-mean of the SVM and MCNN based on our method are 0.922 and 0.987, respectively. Both of them were higher than the RaUS and NearMiss-1 methods.

Table 4 illustrates the imbalance ratio of 5 with the average accuracy of the data classification based on the SVM and MCNN methods and ABM at average accuracy of 92.31% and 98.46%,
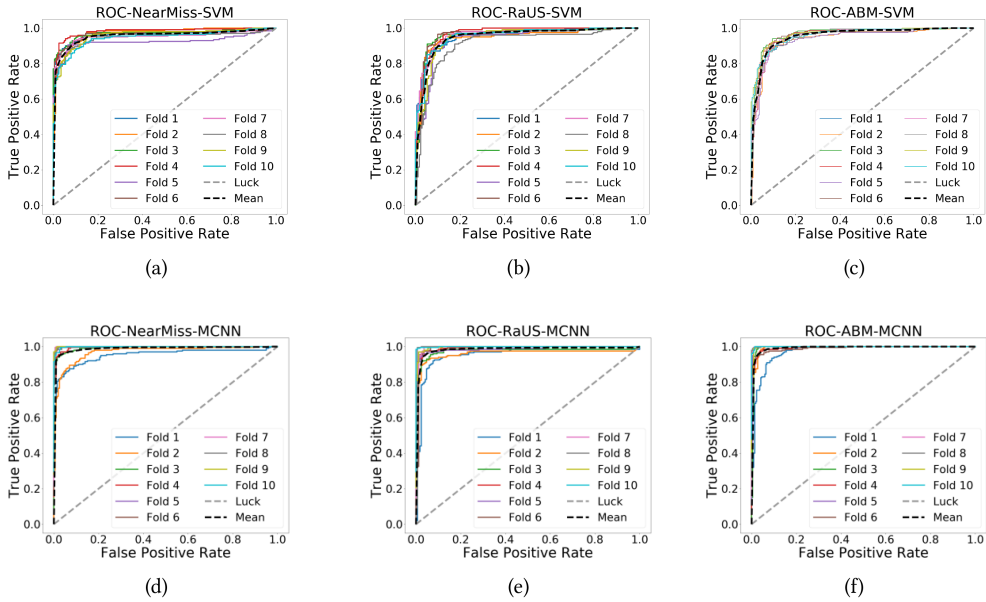
Fig. 3. Ten-fold cross-validation ROC and AUC based on the ABM, NearMiss-1, and RaUS methods at an imbalance ratio of 13. Different color curves represent the ROC at each cross validation, and the mean represents the average ROC of 10 verifications. (a) Results of NearMiss-1 based on SVM. (b) Results of RaUS based on SVM. (c) Results of ABM based on SVM. (d) Results of NearMiss-1 based on MCNN. (e) Results of RaUS based on MCNN. (f) Results of ABM based on MCNN.

respectively, which outperforms the RaUS and NearMiss-1 methods. The proposed method shows the average Macro-F1 of 0.927 and 0.964, which are higher than RaUS and NearMiss-1. The average G-mean of 0.942 and 0.984 are also adopted by ABM. In summary, these results demonstrat that ABM performs better than the RaUS and NearMiss-1 methods in terms of accuracy and G-mean for the imbalanced MI ECG data and normal ECG data.

The performance results of AUC are summarized in Figures 3 and 4 and Tables 5 and 6. The highly convincing results are represented in underlined bold. The 10-fold cross validation of data processed for the ABM, RaUS, and NearMiss-1 methods are shown in Figures 3 and 4. In Table 5 and Figure 3, the sample imbalance ratio is 13, and the three methods obtained the mean AUC based on the MCNN, which are 0.9832, 0.9756, and 0.9763, respectively. In the SVM, the ABM, RaUS, and NearMiss-1 methods achieved a mean for the AUC of 0.9645, 0.9480, and 0.9540, accordingly. Our method also performs better than the conventional results. From the curve distribution of Figure 3, the fluctuations between the 10 ROC curves obtained based on ABM are small, but the fluctuation range of the other two methods (RaUS and NearMiss-1) is large; the main reason is that the distribution of selected samples based on entropy are more stable. In other words, the data obtained by the ABM method contains less variance and higher classification stability.

In Table 6 and Figure 4, the imbalance ratio is 5, and the three methods achieve the mean AUC based on the MCNN, which are 0.9911, 0.9706, and 0.9852, respectively. The ABM method shows better results, and the average AUC values for the three conventional methods following SVM classification are 0.9694, 0.9554, and 0.9546, respectively. In addition, as shown in Figure 4, when the imbalance ratio is 5, the curve distribution differences of the three methods are reduced, and all of the ROC curve distributions are stable, which means that when the sample imbalance ratio is small, the samples selected by the three methods have little effect on the results. Considering that
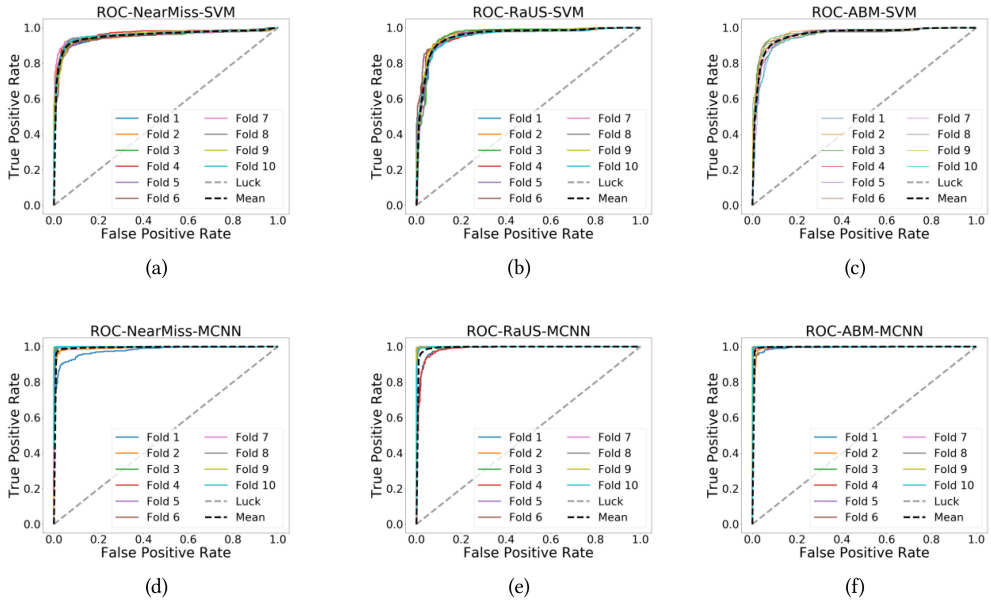
Fig. 4. Ten-fold cross-validation ROC and AUC based on the ABM, NearMiss-1, and RaUS methods at an imbalance ratio of 5. Different color curves represent the ROC at each cross validation, and the mean represents the average ROC of 10 verifications. (a) Results of NearMiss-1 based on SVM. (b) Results of RaUS based on SVM. (c) Results of ABM based on SVM. (d) Results of NearMiss-1 based on MCNN. (e) Results of RaUS based on MCNN. (f) Results of ABM based on MCNN.

Table 5. Average AUC of 10-Fold Cross Validation After Data
Processed by Three Under-Sampling Methods (Sample
Imbalance Ratio Is 13)

Classification Methods	Under-Sampling Methods		
	ABM	NearMiss-1	RaUS
SVM	**0.9645**	0.9540	0.9480
MCNN	**0.9832**	0.9763	0.9756

Table 6. Average AUC of 10-Fold Cross Validation After Data
Processed by Three Under-Sampling Methods (Sample
Imbalance Ratio Is 5)

Classification Methods	Under-Sampling Methods		
	ABM	NearMiss-1	RaUS
SVM	**0.9694**	0.9546	0.9554
MCNN	**0.9911**	0.9852	0.9706

in imbalanced data processing the data distribution based on our method is more concentrated, the total sample attributes are not much different. Figures 3 and 4 and Tables 5 and 6 demonstrate that our proposed method has better performance than the RaUS and NearMiss-1 methods. In comparison, the accuracy, G-mean, Macro-F1, and mean AUC of the proposed ABM surpasses the other conventional methods.

4 CONCLUSION AND FUTURE WORK

In this article, a novel effective under-sampling mechanism called *ABM* is proposed to address the imbalanced MI dataset. Moreover, the experiment results show that our method based on GNB and entropy has high efficiency to deal with imbalanced data, and it consumes less time than the other five conventional under-sampling methods but it takes slightly longer than RaUS. In other words, the proposed ABM outperforms RaUS and NearMiss-1 in PTB ECG data. Meanwhile, the constructed eight-layer MCNN can be trained with less time and has high performance for imbalanced ECG classification; all of them verify the effectiveness of our proposed method.

However, the proposed ABM still has some challenges. For example, the number of minority samples should not be too small to achieve ABM. In addition, ABM can only validate a single two class of datasets. In the future, we will modify the proposed ABM to verify more multi-class datasets (i.e., real clinical ECG data).

REFERENCES

[1] Kashif Ahmad, Mohamed Lamine Mekhalfi, Nicola Conci, Farid Melgani, and Francesco De Natale. 2018. Ensemble of deep models for event recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications* 14, 2 (2018), 51. DOI : https://doi.org/10.1145/3199668

[2] Ulas Baran Baloglu, Muhammed Talo, Ozal Yildirim, Ru San Tan, and U. Rajendra Acharya. 2019. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognition Letters* 122 (2019), 23–30. DOI : https://doi.org/10.1016/j.patrec.2019.02.016

[3] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16 (2002), 321–357. DOI : https://doi.org/10.1613/jair.953

[4] Jie Chen, ZhongCheng Wu, and Jun Zhang. 2019. Driving safety risk prediction using cost-sensitive with nonnegativity-constrained autoencoders based on imbalanced naturalistic driving data. *IEEE Transactions on Intelligent Transportation Systems* 20, 12 (2019), 4450–4465. DOI : https://doi.org/10.1109/TITS.2018.2886280

[5] Noel C. F. Codella, Q.-B. Nguyen, Sharath Pankanti, D. A. Gutman, Brian Helba, A. C. Halpern, and John R. Smith. 2017. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development* 61, 4–5 (2017), Article 5, 15 pages. DOI : https://doi.org/10.1147/JRD.2017.2708299

[6] Chris Drummond and Robert C. Holte. 2003. C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. In *Proceedings of the Workshop on Learning from Imbalanced Datasets II*, Vol. 11. 1–8.

[7] Annarita D'Addabbo and Rosalia Maglietta. 2015. Parallel selective sampling method for imbalanced and large data classification. *Pattern Recognition Letters* 62 (2015), 61–67. DOI : https://doi.org/10.1016/j.patrec.2015.05.008

[8] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. 2017. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications* 73 (2017), 220–239. DOI : https://doi.org/10.1016/j.eswa.2016.12.035

[9] Chuang Han and Li Shi. 2019. Automated interpretable detection of myocardial infarction fusing energy entropy and morphological features. *Computer Methods and Programs in Biomedicine* 175 (2019), 9–23. DOI : https://doi.org/10.1016/j.cmpb.2019.03.012

[10] Hart. 2003. The condensed nearest neighbor rule. *IEEE Transactions on Information Theory* 14, 3 (2003), 515–516. DOI : https://doi.org/10.1109/TIT.1968.1054155

[11] Haibo He and Edwardo A. Garcia. 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* 9 (2009), 1263–1284. DOI : https://doi.org/10.1109/TKDE.2008.239

[12] M. Shamim Hossain, Syed Umar Amin, Mansour Alsulaiman, and Ghulam Muhammad. 2019. Applying deep learning for epilepsy seizure detection and brain mapping visualization. *ACM Transactions on Multimedia Computing, Communications, and Applications* 15, 1s (2019), 1–17. DOI : https://doi.org/10.1145/3241056

[13] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. 2016. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5375–5384. DOI : https://doi.org/10.1109/CVPR.2016.580

[14] Nathalie Japkowicz. 2000. The class imbalance problem: Significance and strategies. In *Proceedings of the International Conference on Artificial Intelligence.*

[15] Qi Kang, Lei Shi, MengChu Zhou, XueSong Wang, QiDi Wu, and Zhi Wei. 2017. A distance-based weighted under-sampling scheme for support vector machines and its application to imbalanced classification. *IEEE Transactions on Neural Networks and Learning Systems* 29, 9 (2017), 4152–4165. DOI : https://doi.org/10.1109/TNNLS.2017.2755595

[16] Fenglian Li, Xueying Zhang, Xiqian Zhang, Chunlei Du, Yue Xu, and Yu-Chu Tian. 2018. Cost-sensitive and hybrid-attribute measure multi-decision tree over imbalanced data sets. *Information Sciences* 422 (2018), 242–256. DOI : https://doi.org/10.1016/j.ins.2017.09.013

[17] Jinyan Li, Lian-Sheng Liu, Simon Fong, Raymond K. Wong, Sabah Mohammed, Jinan Fiaidhi, Yunsick Sung, and Kelvin K. L. Wong. 2017. Adaptive swarm balancing algorithms for rare-event prediction in imbalanced healthcare data. *Computerized Medical Imaging and Graphics* 12, 7 (2017), 1–25. DOI : https://doi.org/10.1371/journal.pone.0180830

[18] Chien-Liang Liu and Po-Yen Hsieh. 2019. Model-based synthetic sampling for imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*. Epub ahead of print. March 18, 2019. DOI : https://doi.org/10.1109/TKDE.2019.2905559

[19] Inderjeet Mani and I. Zhang. 2003. kNN approach to unbalanced data distributions: A case study involving information extraction. In *Proceedings of the Workshop on Learning from Imbalanced Datasets*, Vol. 126. 42–48.

[20] Antonio Maratea, Alfredo Petrosino, and Mario Manzo. 2014. Adjusted *F*-measure and kernel scaling for imbalanced data learning. *Information Sciences* 257 (2014), 331–341. DOI : https://doi.org/10.1016/j.ins.2013.04.016

[21] Iman Nekooeimehr and Susana K. Lai-Yuen. 2016. Adaptive semi-unsupervised weighted oversampling (A-SUWO) for imbalanced datasets. *Expert Systems with Applications* 46 (2016), 405–416. DOI : https://doi.org/10.1016/j.eswa.2015.10.031

[22] Jiapu Pan and Willis J. Tompkins. 1985. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering* 32, 3 (1985), 230–236. DOI : https://doi.org/10.1109/TBME.1985.325532

[23] Sandeep Pirbhulal, Heye Zhang, Md E. Alahi, Hemant Ghayvat, Subhas Mukhopadhyay, Yuan-Ting Zhang, and Wanqing Wu. 2017. A novel secure IoT-based smart home automation system using a wireless sensor network. *Sensors* 17, 1 (2017), 69. DOI : https://doi.org/10.3390/s17030606

[24] Sandeep Pirbhulal, Heye Zhang, Wanqing Wu, Subhas Chandra Mukhopadhyay, and Yuan-Ting Zhang. 2018. Heartbeats based biometric random binary sequences generation to secure wireless body sensor networks. *IEEE Transactions on Biomedical Engineering* 65, 12 (2018), 2751–2759. DOI : https://doi.org/10.1109/TBME.2018.2815155

[25] Foster Provost, Tom Fawcett, and Ron Kohavi. 1997. The case against accuracy estimation for comparing induction algorithms. In *Proceedings of the 15th International Conference on Machine Learning*. 445–453.

[26] Joseph D. Prusa, Taghi M. Khoshgoftaar, and Naeem Seliya. 2016. Enhancing ensemble learners with data sampling on high-dimensional imbalanced tweet sentiment data. In *Proceedings of the 29th International Flairs Conference.*

[27] Bharat Richhariya and Muhammad Tanveer. 2018. A robust fuzzy least squares twin support vector machine for class imbalance learning. *Applied Soft Computing* 71 (2018), 418–432. DOI : https://doi.org/10.1016/j.asoc.2018.07.003

[28] Deboleena Sadhukhan, Saurabh Pal, and Madhuchhanda Mitra. 2018. Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data. *IEEE Transactions on Instrumentation and Measurement* 67, 10 (2018), 2303–2313. DOI : https://doi.org/10.1109/TIM.2018.2816458

[29] Roghayeh Soleymani, Eric Granger, and Giorgio Fumera. 2018. Progressive boosting for class imbalance and its application to face re-identification. *Expert Systems with Applications* 101 (2018), 271–291. DOI : https://doi.org/10.1016/j.eswa.2018.01.023

[30] Yanmin Sun, Mohamed S. Kamel, and Yang Wang. 2006. Boosting for learning multiple classes with imbalanced class distribution. In *Proceedings of the 6th International Conference on Data Mining (ICDM'06)*. IEEE, Los Alamitos, CA, 592–602. DOI : https://doi.org/10.1109/ICDM.2006.29

[31] Seba Susan and Amitesh Kumar. 2019. SSOMaj-SMOTE-SSOMin: Three-step intelligent pruning of majority and minority samples for learning from imbalanced datasets. *Applied Soft Computing* 78 (2019), 141–149. DOI : https://doi.org/10.1016/j.asoc.2019.02.028

[32] Yonglong Tian, Guang-He Lee, Hao He, Chen-Yu Hsu, and Dina Katabi. 2018. RF-based fall monitoring using convolutional neural networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24. DOI : https://doi.org/10.1145/3264947

[33] Ivan Tomek. 1976. Two modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics* 6 (1976), 769–772. DOI : https://doi.org/10.1109/TSMC.1976.4309452

[34] Ivan Tomek. 2007. An experiment with the edited nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-6, 6 (2007), 448–452. DOI : https://doi.org/10.1109/TSMC.1976.4309523

[35] Chih-Fong Tsai, Wei-Chao Lin, Ya-Han Hu, and Guan-Ting Yao. 2019. Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. *Information Sciences* 477 (2019), 47–54. DOI : https://doi.org/10.1016/j.ins.2018.10.029

[36] Yuandou Wang, Hang Liu, Wanbo Zheng, Yunni Xia, Yawen Li, Peng Chen, Kunyin Guo, and Hong Xie. 2019. Multi-objective workflow scheduling with deep-Q-network-based multi-agent reinforcement learning. *IEEE Access* 7 (2019), 39974–39982. DOI : https://doi.org/10.1109/ACCESS.2019.2902846

[37] Wanqing Wu, Sandeep Pirbhulal, Arun Kumar Sangaiah, Subhas Chandra Mukhopadhyay, and Guanglin Li. 2018. Optimization of signal quality over comfortability of textile electrodes for ECG monitoring in fog computing based medical applications. *Future Generation Computer Systems* 86 (2018), 515–526. DOI : https://doi.org/10.1016/j.future.2018.04.024

[38] Wanqing Wu, Sandeep Pirbhulal, Heye Zhang, and Subhas Chandra Mukhopadhyay. 2018. Quantitative assessment for self-tracking of acute stress based on triangulation principle in a wearable sensor system. *IEEE Journal of Biomedical and Health Informatics* 23, 2 (2018), 703–713. DOI : https://doi.org/10.1109/jbhi.2018.2832069

[39] Wanqing Wu, Heye Zhang, Sandeep Pirbhulal, Subhas Chandra Mukhopadhyay, and Yuan-Ting Zhang. 2015. Assessment of biofeedback training for emotion management through wearable textile physiological monitoring system. *IEEE Sensors Journal* 15, 12 (2015), 7087–7095. DOI : https://doi.org/10.1109/JSEN.2015.2470638

[40] Yawen Xiao, Jun Wu, Zongli Lin, and Xiaodong Zhao. 2018. A deep learning-based multi-model ensemble method for cancer prediction. *Computer Methods and Programs in Biomedicine* 153 (2018), 1–9. DOI : https://doi.org/10.1016/j.cmpb.2017.09.005

[41] Yilin Yan, Min Chen, Mei-Ling Shyu, and Shu-Ching Chen. 2015. Deep learning for imbalanced multimedia data classification. In *Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM'15)*. IEEE, Los Alamito, CA, 483–488. DOI : https://doi.org/10.1109/ISM.2015.126

[42] Show-Jane Yen and Yue-Shi Lee. 2009. Cluster-based under-sampling approaches for imbalanced data distributions. *Expert Systems with Applications* 36, 3 (2009), 5718–5727. DOI : https://doi.org/10.1016/j.eswa.2008.06.108

[43] Hualong Yu, Sun Changyin, Xibei Yang, Shang Zheng, and Haitao Zou. 2019. Fuzzy support vector machine with relative density information for classifying imbalanced data. *IEEE Transactions on Fuzzy Systems* 27, 12 (2019), 2353–2367. DOI : https://doi.org/10.1109/TFUZZ.2019.2898371

[44] Chong Zhang, Kay Chen Tan, Haizhou Li, and Geok Soon Hong. 2019. A cost-sensitive deep belief network for imbalanced classification. *IEEE Transactions on Neural Networks and Learning Systems* 30, 1 (2019), 109–122. DOI : https://doi.org/10.1109/TNNLS.2018.2832648

[45] Qingchen Zhang, Laurence T. Yang, Zhikui Chen, and Peng Li. 2018. Dependable deep computation model for feature learning on big data in cyber-physical systems. *ACM Transactions on Cyber-Physical Systems* 3, 1 (2018), 11. DOI : https://doi.org/10.1145/3110218

[46] Xiaoxuan Zhang, Tianbao Yang, and Padmini Srinivasan. 2016. Online asymmetric active learning with imbalanced data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 2055–2064. DOI : https://doi.org/10.1145/2939672.2939854

[47] Yifan Zhang, Peilin Zhao, Jiezhang Cao, Wenye Ma, Junzhou Huang, Qingyao Wu, and Mingkui Tan. 2018. Online adaptive asymmetric active learning for budgeted imbalanced data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 2768–2777. DOI : https://doi.org/10.1145/3219819.3219948

[48] Zhedong Zheng, Liang Zheng, and Yi Yang. 2016. A discriminatively learned CNN embedding for person re-identification. *ACM Transactions on Multimedia Computing, Communications, and Applications* 14, 1 (2016), 1–20. DOI : https://doi.org/10.1145/3159171