Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 April 2026
Revised 18 May 2026
Accepted 16 June 2026

Natural Resources for Human Health 2026, Vol. 6 (3s): 957–975

KEYWORDS:
Arrhythmia, Balanced-Dataset,1D CNN, Accuracy, F1-Score, ROC-AUC

Comparative Analysis of Simplified 1D Convolutional Neural Network Performance on Balanced ECG Datasets for Cardiac Arrhythmia Detection

Madhumita Mishra[1], Ashwinkumar U Motagi[2]

[1,2]School of Computer Science and Engineering, REVA University, Bengaluru, India. Email: madhumita.mishra@reva.edu.in
[1,2]School of Computer Science and Engineering, REVA University, Bengaluru, India. Email: ashwinkumar.um@reva.edu.in
Corresponding-author- madhumita.mishra@reva.edu.in

ABSTRACT: Globally, cardiovascular illnesses continue to be the leading cause of morbidity and mortality. Among them, arrhythmia is a prevalent cardiovascular disorder that affects a significant portion of the population. Timely detection and precise classification of cardiac arrhythmias are essential for actual intervention and patient care. Convolutional Neural Networks (CNNs), a cutting-edge advancement in deep learning, have shown remarkable success in automating the identification of cardiac arrhythmias from electrocardiogram (ECG) data. However, the performance of these models can vary significantly depending on the class distribution of the dataset, with imbalanced datasets presenting a significant challenge. This study presents a comparative analysis of one-dimensional convolutional neural network (1D CNN) performance when applied to multiple balanced ECG datasets. It sheds light on the impact of data distribution on model effectiveness. In the balanced dataset, the 1D CNN achieved improved accuracy and F1-score to detect abnormal ECG patterns.

The comparison was conducted on different batch sizes of 16, 32, 64, and 128. Among batch sizes 16 and 32, the BorderlineSMOTE data balancing on MIT-BIH 1D CNN mode yielded superior performance to other balancing methods based on the accuracy, F1-score and ROC-AUC Score. Similarly, in the case of an increase in batch sizes to 64 and 128, ADASYN oversampling applying on the specified model provided improved results compared to others. The accuracy and F1-score vary from a range of 0.9752 to 0.9804 and 0.8622 to 0.8994 respectively. These findings emphasize the need for careful consideration of medical dataset-balancing techniques while applying deep learning models.

1 INTRODUCTION

Arrhythmias encompass a broad spectrum of irregular heart rhythms, ranging from benign to life-threatening conditions. The typical electrical signal of a healthy heart during one complete heartbeat, as visualized on an electrocardiogram (ECG) is presented in Figure 1[2]. A deviancy from this standard electrical pattern during a heartbeat is considered an arrhythmia. Timely and accurate classification of arrhythmias is vital for guiding appropriate clinical interventions, such as medication adjustments, implantable device therapies, or surgical procedures. Manual analysis is time-consuming and susceptible to errors. Deep learning models, particularly 1D CNNs, offer a promising avenue for automated arrhythmia detection, allowing for faster and more precise diagnoses [1].

In medical datasets, the prevalence of certain diseases or conditions may be naturally low compared to normal conditions. For example, rare diseases or conditions may have fewer instances in the dataset, leading to class imbalance. In the context of the arrhythmia dataset, the majority of the data that have been gathered or made available for the classifications shows that the number of normal instances well exceeds the number of abnormal cases. Highly imbalanced data demonstrates bias in favour of majority classes which is common in medical datasets and in the worst case, it may overlook the minority class completely [3].

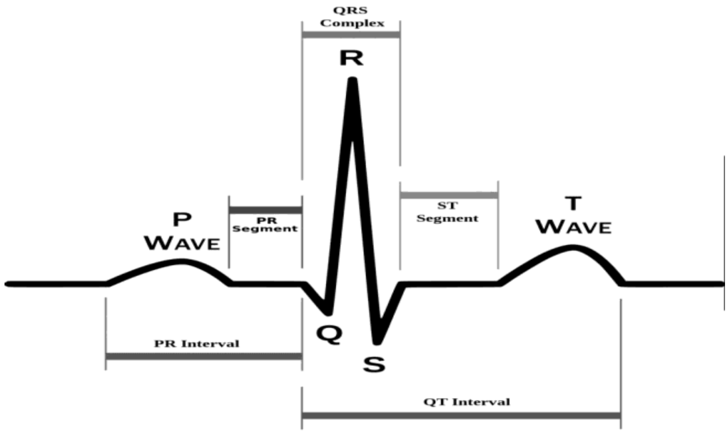


Fig. 1. Normal Heartbeat

Dataset imbalance, a common challenge occurs when certain classes are underrepresented compared to others. Balancing datasets in arrhythmia classification is essential to improve the model performance, generalization, and decision-making in healthcare, ultimately leading to better patient outcomes and advancements in medical research [23] [24]. Sampling is widely used in various fields, including research and medical studies, to collect representative data and make reliable conclusions or predictions about the dataset of interest. Broadly, sampling is divided into two main categories: oversampling and undersampling [25-28]. The purpose of oversampling is to replicate existing minority class samples or create synthetic samples to boost the representation of minority classes. Undersampling aims to decrease the representation of majority classes by eliminating or subsampling instances from the majority classes. Oversampling techniques provide a balanced and representative dataset for training models in medical applications, leading to improved model performance and better clinical outcomes [33]. Combining both oversampling and undersampling, also ensures that the machine learning model is exposed to sufficient samples across all categories, allowing it to learn the underlying patterns and characteristics of the data more effectively.

1D CNNs have proven to be incredibly effective at analysing a variety of sequential data sets, including time-series data like ECG readings. Their ability to analyse recurring structures within a specific region and hierarchical features makes them well-suited for detecting complex temporal relationships within ECG data. By leveraging these characteristics, 1D CNNs have shown promising results in enhancing the accurateness and efficiency of the arrhythmia classification model [12]. It is crucial to assess 1D CNN model performance over a range of oversampling balancing strategies to find the competent method for resolving class imbalance in medical datasets [21]. The current work investigates the performance of 1D CNNs in arrhythmia classification using multiple oversampling and a combination of oversampling and undersampling balancing techniques in ECG datasets.

The widely used MIT-BIH ECG database, employed in this study, primarily comprises normal recordings, while abnormal recordings are notably fewer. This disparity in data distribution is a common challenge in ECG databases, potentially leading to biases in deep learning-based classification frameworks due to the scarcity of abnormal signals [17] [18]. To mitigate this issue, oversampling techniques are employed, which augment minority class instances, enabling machine learning models to better learn from all classes and enhance their performance on imbalanced datasets. The comparison of 1D CNN model performance in this study encompasses various balancing techniques, including SMOTE, ADASYN, Random Oversampling, SMOTEN, BorderlineSMOTE, SMOTETomek, and SMOTE-ENN. While SMOTE is prevalent in existing research works, this study offers a comprehensive analysis by exploring multiple oversampling and hybrid sampling techniques, thereby providing a more nuanced understanding of their effectiveness in improving model performance.

Previous research overlooked a comprehensive comparison of various sampling techniques across different batch sizes. This review investigates the application of models for arrhythmia classification, utilizing various data balancing techniques with different batch sizes. We review the literature to determine if 1D CNNs are useful for classifying arrhythmias, discuss the problems caused by unbalanced datasets, and show how various data balancing strategies affect model performance.

The paper is organized as follows: It highlights recent advancements in both balanced and imbalanced datasets, with a focus on the application of 1D CNNs in ECG analysis. It includes a detailed description of the dataset, a discussion of several oversampling techniques, and an outline of the architecture of our 1D CNN model. Subsequently, we present experimental results and conduct a comprehensive comparative analysis of model performance on balanced datasets. Finally, we conclude with a summary of key insights and provide recommendations for future research directions

2 RELATED WORK

In medical diagnosis, the issue of an imbalanced dataset is highly predominant. When a classifier is given unbalanced data, it will result in unfavourable outcomes, such as testing performance that is significantly worse than training performance [4]. It is challenging to create a large-scale balanced ECG database because of its sensitive information. If trained on an imbalanced ECG database, the learning model will be biased towards the majority type, resulting in inferior recognition performance for the minority type. Synthesizing the ECG data has been the primary method used in recent years to address the issue of imbalanced data. Various techniques are available to address the problem of class imbalance in datasets. These techniques include balancing at the algorithmic and data-level techniques. The advantages of data-level techniques, including their simplicity and independence from specific algorithms, contribute to their widespread adoption. Various data sampling techniques have been proposed to address class imbalance, including Random Oversampling, Random Undersampling, SMOTE (Synthetic Minority Over-sampling Technique) [6], ADASYN (Adaptive Synthetic Sampling) [7] and multiple variants of SMOTE like SMOTEN [8], BorderlineSMOTE [9], SVMSMOTE.

Conventional techniques for classifying arrhythmias include machine learning techniques like Support Vector Machines (SVMs) and Decision Trees, as well as rule-based algorithms. However, these methods often rely on handcrafted features and may not capture complex patterns effectively. Deep learning models, particularly CNNs, have shown promising results in automated arrhythmia classification by automatically learning relevant features from raw ECG signals [10]. To understand the relationship between ECG features and the associated types of arrhythmias, deep learning algorithms need a substantial volume of ECG signals as training data.

Various methods have been proposed to address the challenges posed by imbalanced datasets and to improve classification performance. The issue of class imbalance exists in many different domains, and during the last ten years, numerous strategies to address it have been put forth [11]. Research in deep learning has garnered significant attention, yet only a few studies have focused on understanding the relationship between class imbalance and deep learning methods [3]. Deep learning models, particularly 1D CNNs, have shown promising results in automated arrhythmia classification from ECG signals. The study in [1] introduces a 1-D convolutional ResNet model designed to classify five different types of heartbeats, utilizing data from the publicly available PhysioNet MIT-BIH Arrhythmia database, a common resource for arrhythmia classification. By applying SMOTE to the training dataset, the model achieves an average accuracy of 98.63% in classifying various heartbeat signals. In [12], a CNN approach was utilized by researchers to achieve accuracies of 93.53% and 95.22% with and without noise removal, respectively. Meanwhile in [13], authors employed a 1D convolutional neural network for arrhythmia classification, achieving a peak accuracy of 95%. Furthermore, the study described in [14] used random sampling to analyse a given ECG dataset using a 1-dimensional CNN model, yielding an overall accuracy of 96.72% and F1 measure of 85.08%.

In previous work [15], the authors suggested an eleven-layer deep CNN to automatically classify four arrhythmic classes using 21,709 ECG segments from net A and 8,683 ECG segments from net B. The proposed algorithm achieved an accuracy of only 92.50%. Earlier researchers proposed an approach for accurately classifying ECG arrhythmias using a 2D CNN architecture. In this system, [17,27] ECG heartbeats are transformed into 2D time-amplitude images to serve as input data for the CNN. While this approach achieved higher accuracy, the complexity of the architecture is significantly high.

ECG signals, being inherently one-dimensional time-series data, are best processed by a 1D CNN. This allows the data to be handled directly without requiring transformation into a 2D format, preserving the temporal characteristics of the ECG signals and potentially enhancing performance. Transforming 1D ECG signals into 2D images, such as time-amplitude images for 2D CNNs, can introduce redundant or irrelevant information, complicating the classification process and potentially reducing the model's efficiency and accuracy. Additionally, training and classifying arrhythmia using recurrent neural networks (RNNs) [22,23] can be more challenging due to issues like exploding and vanishing gradients, which are less problematic in CNNs. This stability makes CNNs easier to train and tune.

Consequently, in imbalanced learning issues, multiple other evaluation metrics are established to measure the classifier. This comparative analysis will provide insights into the effectiveness of different data sampling methods and their impact on model performance, ultimately contributing to the development of more reliable and accurate arrhythmia classification algorithms.

3 DATASET

This study utilized the MIT-BIH database [20] for training and testing the proposed 1-D convolutional neural network model. The MIT-BIH Arrhythmia database has become a cornerstone dataset in the field of arrhythmia research. Researchers use the dataset for tasks such as developing algorithms for automatic arrhythmia detection and classification, studying the characteristics and patterns of different types of arrhythmias, and evaluating the performance of arrhythmia detection systems. It contains electrocardiogram (ECG) recordings from multiple subjects, each annotated with beat annotations indicating the presence of various types of arrhythmias. These annotations provide valuable labels for the ECG recordings, allowing researchers to develop and evaluate algorithms for automatic arrhythmia detection and classification.

The database comprises a total of 109446 heartbeats recorded at a sampling frequency of 125 Hz from 44 records. The data helps perform code-based heart assessments and comes from a diverse range of patients including both healthy individuals and those with heart disease.

The five main forms of arrhythmias include Normal (N) Ventricular ectopic (V), Supraventricular ectopic (S), Fusion (F), and Unknown (Q), according to the MIT-BIH Arrhythmia Database as shown in Figure 2. The data distribution of five major types of arrhythmias are highly imbalanced as shown in Figure 3. Tables 1 to 3 present the original size of the data, followed by the numerical results after applying various data balancing techniques.

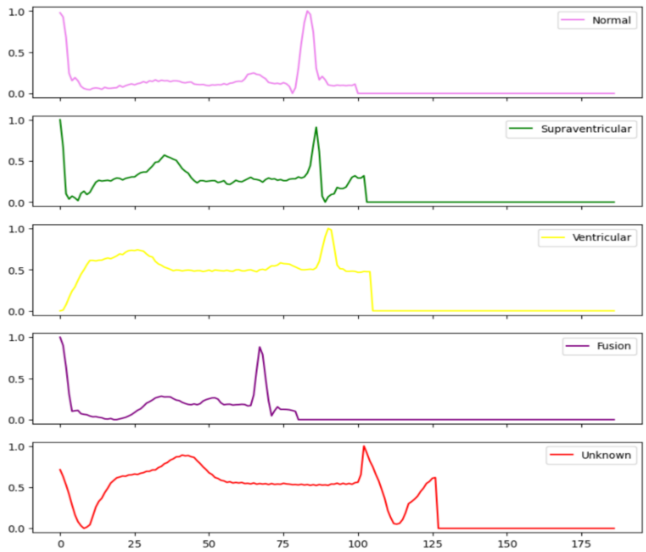


Fig. 2. ECG data belonging to each class label.

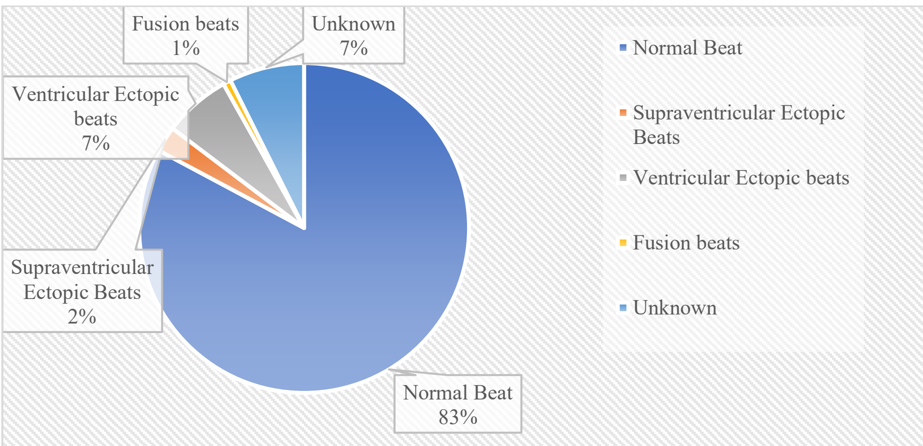


Fig. 3. Distribution of classes in each category in Arrhythmia MIT-BIH Dataset

Table 1. Actual Sample Size of Data

Labels	Sample Tag	Actual Sample
0	Normal Beat	72494
1	Supraventricular Ectopic Beats	2199
2	Ventricular Ectopic beats	5750
3	Fusion beats	663
4	Unknown	6450

Table 2. Sample Size after applying Oversampling Techniques.

Labels	RandomSampler	SMOTE	ADASYN	BorderlineSMOTE	SMOTEN	SVMSMOTE
0	72494	72494	72494	72494	72494	72494
1	72494	72494	72399	72494	72494	72494
2	72494	72494	72180	72494	72494	72494
3	72494	72494	72459	72494	72494	72494
4	72494	72494	72553	72494	72494	72494

Table 3. Sample Size after applying Hybrid Sampling Techniques.

Labels	SMOTENN	SMOTETomek
0	68793	72493
1	72458	72493
2	72425	72494
3	72493	72494
4	72492	72494

4 PROPOSED APPROACH

A generalized workflow for this work is represented in Figure 4. A summarized workflow for processing the MIT-BIH dataset using sampling techniques and a simplified 1D CNN model for arrhythmia classification:

MIT-BIH Dataset: The MIT-BIH Arrhythmia Database is obtained containing ECG signals with annotations for different arrhythmia classes.

Data Preprocessing Stage: The ECG signals is pre-processed by normalizing the data to ensure consistency and reliability.

- Split into Training, Test and Validation Dataset: The pre-processed dataset is split into training, test and validation sets in a ratio of 80:10:10 to evaluate the performance of the model.

- Sampling Techniques Applied: Sampling techniques such as SMOTE, ADASYN, or BorderlineSMOTE are applied to address class imbalance, ensuring that each arrhythmia class is adequately represented in the training dataset.

- Addition of Noise: Gaussian noise is added to the training dataset to enhance the model's robustness and generalization ability.

Training Model: A simplified 1D CNN model is applied on the balanced noise-augmented training dataset to enhance generalization. The model leverages convolutional layers to extract features, followed by pooling layers for dimensionality reduction.
Prediction: The trained model is used to make predictions on the test dataset, classifying ECG signals into five major arrhythmia classes. (N, S, V, F and Q)
Evaluation: The performance of the model is evaluated based on metrics such as accuracy, precision, recall, and F1-score, [19] to assess its ability for accurate arrhythmias classification.

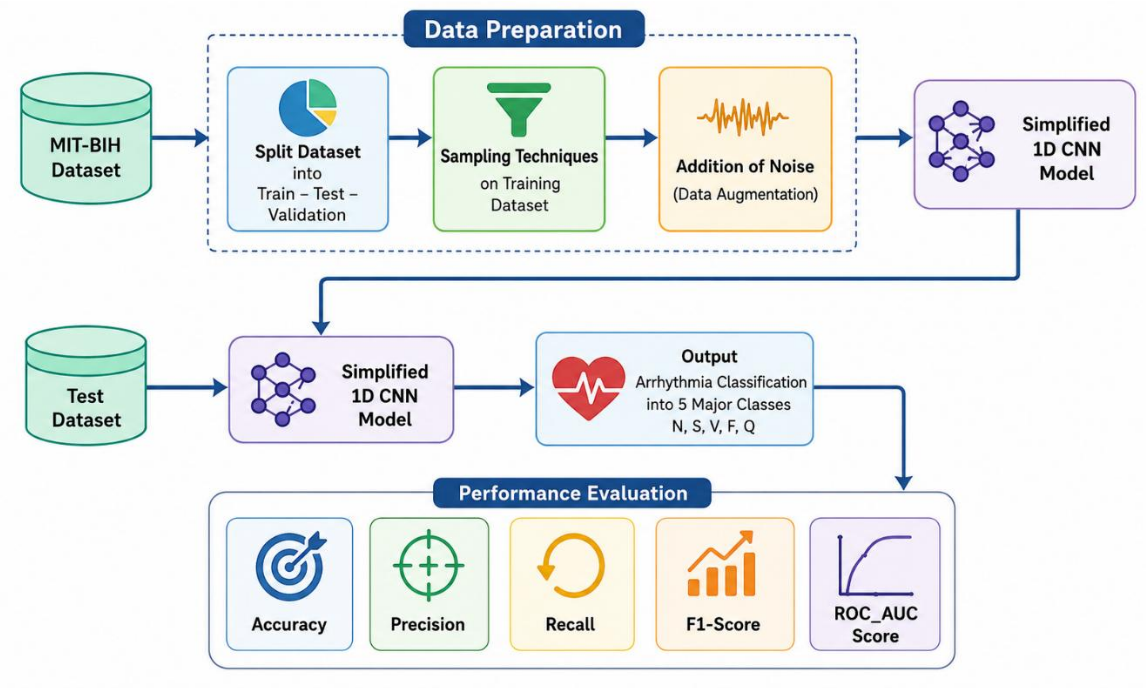


Fig. 4. Summarized workflow for processing MIT-BIH Dataset for Arrhythmia Classification

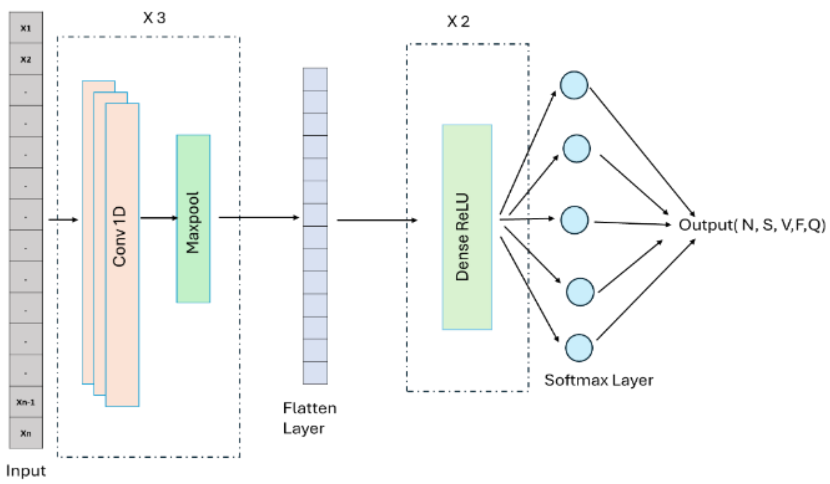


Fig. 5. Simplified 1D CNN Model

Figure 5 depicts the simplified 1D convolutional neural network (CNN) model for arrhythmia classification. The diagram shows layers stacked horizontally, representing the typical order of operations in a CNN. Each layer is represented by a block. The arrows indicate the flow of data from one layer to the next. The first block consists of three pairs of 1D Convolution layers followed by MaxPooling layers. After the MaxPooling layer, the data is flattened into one-dimensional form. This is followed by three Dense layers, with the first two having ReLU activation functions and the last one having a Softmax activation function for multi-class classification. Each layer performs a specific function to transform the input data into a format suitable for arrhythmia classification. Here's a breakdown of the layers:

Input: This layer represents the raw ECG signal input into the model. The ECG signal is a 1D time series that captures the electrical activity of the heart over time.

Conv 1D: This layer applies a 1D convolution operation. Convolution is a mathematical operation that extracts features from the ECG signal by sliding a filter across the signal. The filter learns weights that identify important patterns within the signal.

Maxpool: This layer performs down-sampling, reducing the dimensionality of the data. It extracts the maximum value from a predefined window of elements in the convolutional layer's output, summarizing the activation across that region.

Flatten: This layer performs a flattening operation on the 2D output tensor from the convolutional layer. This transformation converts the spatial data (arranged in rows and columns) into a 1D vector, enabling it to be fed into fully connected layers that process data as a single list of features.

Dense ReLU: This layer represents a densely connected layer with ReLU (Rectified Linear Unit) activation function. The dense layer connects all neurons from the previous layer to every neuron in this layer. ReLU activation introduces non-linearity, allowing the model to learn more complex patterns.

Output: This layer represents the final output of the model. It uses a softmax activation function to generate probabilities for each arrhythmia class. The class with the highest probability is predicted as the ECG's rhythm.

Overall, the simplified 1D CNN model aims to automatically classify ECG signals into different arrhythmia classes by learning distinctive features from the raw ECG data.

5 DATASET BALANCING TECHNIQUE

Multiple data balancing techniques can be employed to mitigate the effects of class imbalance in arrhythmia classification with 1D CNNs. Three main categories exist for addressing class imbalance: data-level methods (e.g., oversampling, undersampling), algorithm-level techniques (e.g., cost-sensitive learning), and hybrid approaches that combine these strategies. The focus of our work is on data level balancing techniques which attempt to reduce the level of imbalance through various sampling methods. By leveraging the strengths of CNN architectures and addressing the challenges of class imbalance through data sampling, we aim to enhance the performance and clinical utility of arrhythmia classification systems.

5.1 OVERSAMPLING TECHNIQUE

Addressing class imbalance is crucial in deep learning tasks where one class significantly outnumbers the other(s), potentially leading to biased models. The goal of oversampling approaches is to artificially increase the number of cases in the minority class in order to mitigate class imbalance. The generalized equation for all oversampling methods is as mentioned in equation 1 where Nminority and Nmajority represents the number of instances in the minority and majority classes, respectively. IR denotes the imbalance ratio, calculated as:

$$IR = \frac{N\text{majority}}{N\text{minority}} \tag{1}$$

The prevalently used Oversampling techniques are as follows:

Random Sampler: To balance the dataset, RandomSampler randomly selects instances from the majority class without replacement [29,30]. This effectively matches the number of instances in the majority and minority classes. The resulting dataset has an equal number of instances for each class, addressing the imbalance issue. In brief, RandomSampler doesn't involve complex equations, but rather a straightforward process of randomly sampling instances from the majority class to achieve class balance.

ADASYN (Adaptive Synthetic Sampling): The method that creates synthetic samples for the minority class adaptively depending on their density distribution in order to address class imbalance [7,31]. For each minority instance $\boldsymbol{x_i}$, calculate its density $\boldsymbol{D_i}$ as the inverse of the average distance to its $\boldsymbol{k}$ nearest minority class neighbours:

$$D_i = \frac{1}{k}\sum_{x_j \in neighbors(x_i)} \frac{1}{dist(x_i, x_j)} \quad (2)$$

where k is the number of nearest neighbours, neighbours(x_i) is the set of k nearest neighbours of x_i, and **dist(**x_i, x_j**)** is the distance between instances x_i and x_j. The importance score is computed for each minority instance based on its density and the imbalance ratio (IR) as calculated in equation 1. The importance score W_i for each instance x_i is computed as:

$$W_i = \frac{D_i}{\sum_{i=1}^{N_{minority}} D_i} \times \frac{1-IR}{N_{minority}} \quad (3)$$

where N_{minority} is the number of minority class instances.

For each minority instance x_i, a number of synthetic samples proportional to its importance score W_i is generated. The synthetic samples are generated by interpolating between x_i and its k nearest minority class neighbours. The original minority class instances are combined with the synthetic samples to form a new, balanced dataset.

SMOTE (Synthetic Minority Over-sampling Technique): Generates synthetic samples by interpolating between existing minority class instances and their nearest neighbours in feature space [6,25]. As shown in equation 1 the imbalance ratio is computed. A minority class instance x_i is randomly selected from the dataset. The k-nearest neighbours of the selected minority instance x_i from the minority class instances are obtained mostly using Euclidean distance or other distance metric. For each selected minority instance x_i randomly selects one of its k-nearest neighbours x_j. Then, generate a synthetic sample x_{new} by interpolating between x_i and x_j in the feature space.

$$x_{\text{new}} = x_i + \lambda \times (x_j - x_i) \quad (4)$$

where $0<\lambda<1$is a random value chosen uniformly. The above phases are repeated till the desired number of synthetic samples is generated or a predetermined criterion is met. SMOTE addresses class imbalance by creating synthetic data for the minority class. It does this by taking a random minority class point and its nearest neighbours (also from the minority class) and then creating a new data point somewhere in between them.

BorderlineSMOTE: The methodology builds upon SMOTE by strategically generating synthetic samples for the minority class. These samples are positioned close to the decision boundary, the critical area where the minority and majority classes are most difficult to distinguish. It first identifies the instances in the minority class that are near to the decision boundary and considers them as borderline instances Randomly a borderline minority class instance x_i is selected from the dataset. Similar to SMOTE the k-nearest neighbours of the selected minority instance x_i from both the minority and majority class instances are found using Euclidean distance. The borderline score for the selected instance x_i based on the proportion of its neighbours that belong to the majority class: is computed as mentioned in the equation:

$$Bordeline\ Score = \frac{\text{Number of majority neighbors}}{k} \quad (5)$$

For each selected borderline minority instance x_i, if the borderline score is above a certain threshold (e.g., 0.5), then a synthetic sample x_{new} is generated by interpolating between x_i and one of its minority class neighbours. The phases are repeated till the number of synthetic samples is generated or a predetermined criteria is met. It uses a similar approach to SMOTE but selects only borderline instances and generates synthetic samples based on their proximity to the decision boundary [32]. The imbalance ratio is computed as mentioned earlier.

SMOTEN:

It extends over SMOTE to handle categorical features by introducing a modified approach for generating synthetic samples [34]. For categorical features, instead of linear interpolation, SMOTEN selects a random value from the set of possible values for each categorical feature.

Given a minority instance x_i with both numerical and categorical features, SMOTEN selects k nearest neighbours from the minority class to generating synthetic samples.

For each selected neighbour x_j, a synthetic sample $\textbf{new}_{\textbf{x}}$ is generated by combining the numerical features using linear interpolation and randomly selecting values for categorical features.

For numerical features, SMOTEN follows the same equation as SMOTE:

$$new_x = x_i + \lambda \times (x_j - x_i) \quad (6)$$

Where $0 < \lambda < 1$ is a random value chosen uniformly.

SVMSMOTE (Support Vector Machine Synthetic Minority Over-Sampling Technique):

An oversampling method that employs Support Vector Machines (SVMs) to generate synthetic samples for the minority class. The basic idea is to use SVMs to identify informative minority class instances (support vectors) and generate synthetic samples around them [35]. Synthetic samples are generated around the identified support vectors, similar to the SMOTE algorithm. The process of generating synthetic samples follows the same equation as SMOTE as mentioned in equation 6. The key difference lies in the selection of informative minority instances (support vectors) using SVMs.

5.2 HYBRID SAMPLING TECHNIQUE

SMOTEENN (SMOTE combined with Edited Nearest Neighbours)

A fusion sampling technique that combines the oversampling capabilities of SMOTE with the under-sampling approach of Edited Nearest Neighbours (ENN). To address class imbalance, SMOTEENN begins by applying SMOTE, which generates synthetic data points for the minority class [36]. It then applies ENN to the combined dataset (original minority class instances + synthetic samples) to remove noisy instances. For each instance x_i, ENN calculates the majority class label among its k nearest neighbours. If x_i has a minority class label but most of its neighbours belong to the majority class, x_i is considered noise and removed. The result is a balanced dataset with reduced noise.

SMOTETomek

A hybrid sampling technique that combines the oversampling capabilities of SMOTE with the under-sampling approach of Tomek links. SMOTETomek first applies SMOTE to oversample the minority class and generate synthetic samples. It then identifies Tomek links in the combined dataset (original minority class instances + synthetic samples) and removes the instances involved in these links [37]. Tomek linkages are pairs of instances that are closest to each other but are in separate classes— a majority class instance and a minority class instance. Removing instances involved in Tomek links can help improve the separation between classes and remove noisy samples. SMOTETomek offers a comprehensive approach that addresses both class imbalance and noisy instances in the dataset.

6 RESULTS AND DISCUSSION

In this work we have applied 6 Oversampling techniques and 2 hybrid sampling techniques to balance the dataset. Tables 4 to 11 display the results obtained from running the 1D CNN model on multiple sampled datasets, using different batch sizes (16, 32, 64, 128) for 24 epochs with a learning rate of 0.01.

6.1. OVERSAMPLING MODELS RESULTS

Table 4. Result obtained for Batch Size-16

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
RandomSampler	0.9930	0.9781	0.8974	0.8600	0.8783	0.9230
ADASYN	0.9936	0.9766	0.8953	0.8534	0.8731	0.9181
SMOTE	0.9924	0.9775	0.8902	0.8702	0.8800	0.9268
BorderlineSMOTE	0.9930	0.9785	0.9084	0.8703	0.8889	0.9311
SMOTEN	0.9924	0.9784	0.9142	0.8510	0.8814	0.9174

SVMSMOTE	0.9931	0.9757	0.9083	0.8348	0.8699	0.9064

Table 5. Result obtained for Batch Size-32

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
RandomSampler	0.9942	0.9778	0.9161	0.8555	0.8847	0.9197
ADASYN	0.9941	0.9772	0.8744	0.8677	0.8710	0.9216
SMOTE	0.9944	0.9783	0.8964	0.8650	0.8804	0.9266
BorderlineSMOTE	0.9944	0.9788	0.9101	0.8699	0.8895	0.9375
SMOTEN	0.9941	0.9783	0.9092	0.8596	0.8837	0.9222
SVMSMOTE	0.9939	0.9752	0.8607	0.8650	0.8628	0.9335

Table 6. Result obtained for Batch Size-64

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
RandomSampler	0.9935	0.9788	0.9192	0.8560	0.8864	0.9199
ADASYN	0.9940	0.9804	0.9201	0.8685	0.8931	0.9265
SMOTE	0.9946	0.9787	0.9189	0.8519	0.8841	0.9170
BorderlineSMOTE	0.9945	0.9789	0.8968	0.8670	0.8816	0.9261
SMOTEN	0.9976	0.9774	0.8936	0.8590	0.8759	0.9214
SVMSMOTE	0.9946	0.9779	0.9322	0.8470	0.8877	0.9145

Table 7. Result obtained for Batch Size-128

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
RandomSampler	0.9923	0.9780	0.8903	0.8751	0.8826	0.9294
ADASYN	0.9928	0.9804	0.9201	0.8798	0.8994	0.9384
SMOTE	0.9928	0.9768	0.8981	0.8730	0.8853	0.9289
BorderlineSMOTE	0.9935	0.9763	0.8788	0.8590	0.8687	0.9324

SMOTEN	0.9928	0.9774	0.8936	0.8685	0.8808	0.9255
SVMSMOTE	0.9930	0.9773	0.8831	0.8728	0.8778	0.9288

6.2. HYBRID SAMPLING TECHNIQUES

Table 8. Result obtained for Batch Size 16

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
SMOTE+ENN	0.9935	0.9789	0.9290	0.8457	0.8812	0.9140
SMOTETomek	0.9936	0.9791	0.9019	0.8728	0.8869	0.9286

Table 9. Result obtained for Batch Size 32

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
SMOTE+ENN	0.9943	0.9771	0.8853	0.8649	0.8749	0.9292
SMOTETomek	0.9945	0.9777	0.9118	0.8689	0.8898	0.9293

Table 10. Result obtained for Batch Size 64

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
SMOTE+ENN	0.9938	0.9781	0.8989	0.8620	0.8801	0.9214
SMOTETomek	0.9939	0.9784	0.9015	0.8635	0.8820	0.9238

Table 11. Result obtained for Batch Size 128

MODELS	Training Accuracy	Test Accuracy	Precision Score	Recall	F1-Score	ROC-AUC
SMOTE+ENN	0.9934	0.9788	0.9262	0.8495	0.8862	0.9164
SMOTETomek	0.9931	0.9781	0.9072	0.8709	0.8887	0.9214

It is important to carefully evaluate the performance metrics of each sampling technique and consider the specific requirements of the classification task when selecting the most suitable technique. Based on the provided table showcasing the performance metrics for various sampling techniques, we can draw several inferences:

Consistent High Performance Across Techniques: All sampling techniques demonstrate high training and test accuracies, ranging from approximately **97.52% to 98.04%** for test accuracy and from **99.23% to 99.76%** for training accuracy. This indicates that the models trained on the resampled datasets perform well in terms of correctly classifying instances.

Variability in F1-Score, Precision and Recall: There is variability noticed in F1-score, precision and recall metrics across different sampling techniques. For instance, across all table SVMSMOTE shows the highest precision score (93.22%) whereas with a low recall value (84.70%). However, it is observed ADASYN has the highest recall (87.98%) with a batch size of 128 compared to other techniques. The F1-scores range from approximately 86.28% to 89.94%, indicating a balance between precision and recall.

ROC-AUC Performance: The ROC-AUC values, which denote the area under the receiver operating characteristic curve, remain relatively consistent across various sampling techniques, ranging from approximately **90.64% to 93.84%.** This suggests that the models have good discrimination capabilities in distinguishing between positive and negative classes.

While there are slight variations in performance metrics across different sampling techniques, all techniques appear to effectively address class imbalance and produce models with robust performance. The choice of sampling technique may depend on specific objectives, preferences, and trade-offs between different performance metrics.

In medical diagnostic scenarios, such as screening for life-threatening conditions like arrhythmia detection, maximizing recall (sensitivity) is often prioritized. Recall measures the percentage of relevant cases identified. While high precision is also important to minimize false positives and unnecessary treatments, balancing precision and recall is crucial in medical diagnostics. Consequently, the F1-score, which considers both metrics, is taken into account.

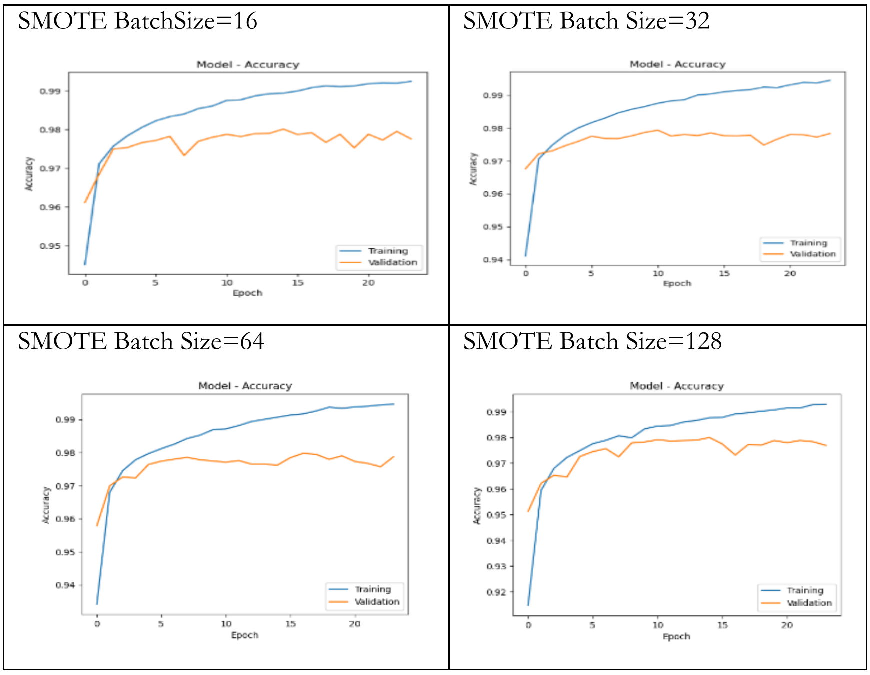
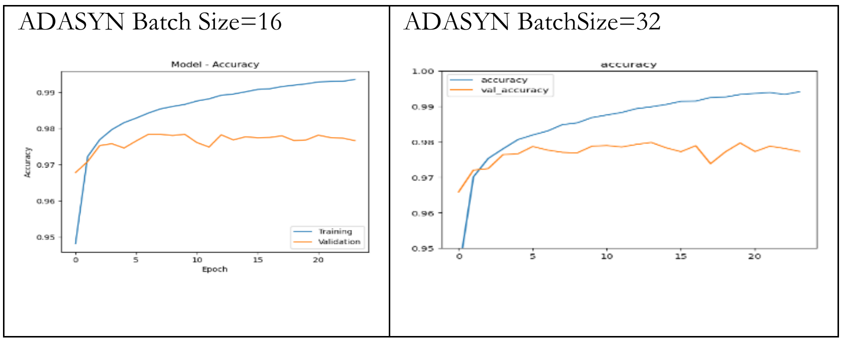
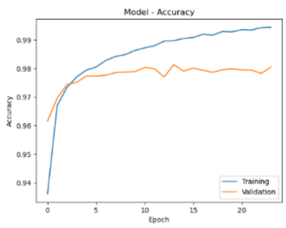


Fig. 6. Figure 6. Comparison of accuracy achieved during training and validation phases on applying SMOTE balancing technique to Simplified1D CNN model.



Natural Resources for
Human Health
eISSN: 2583-1194

ADASYN Batch Size=64



ADASYN BatchSize=128

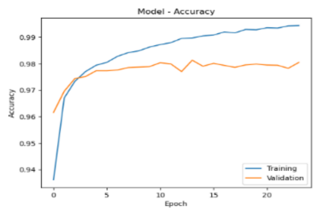
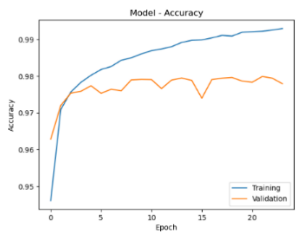
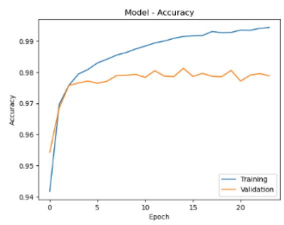


Fig. 7. Comparison of accuracy achieved during training and validation phases on applying ADASYN balancing technique to Simplified1D CNN model.

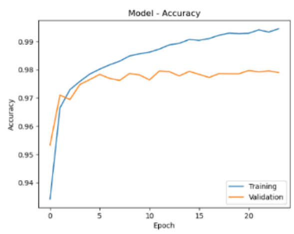
BorderlineSMOTE Batch Size=16



BorderlineSMOTE BatchSize=32



BorderlineSMOTE Batch Size=64



BorderlineSMOTEBatchSize=128

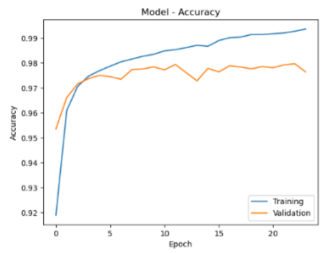
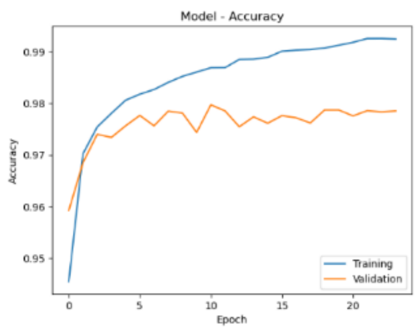
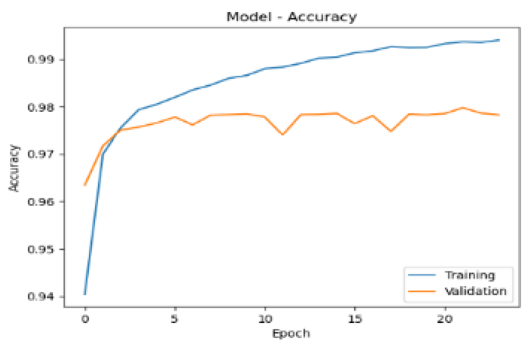


Fig. 8. Comparison of accuracy achieved during training and validation phases on applying BorderlineSMOTE balancing technique to Simplified1D CNN model.

SMOTEN Batch Size=16



SMOTEN BatchSize=32



Natural Resources for
Human Health
eISSN: 2583-1194

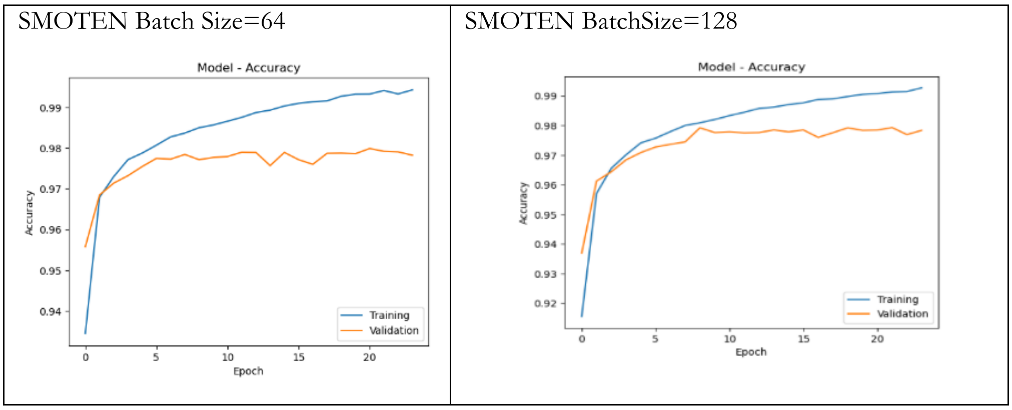


Fig. 9. Comparison of accuracy achieved during training and validation phases on applying SMOTETomek balancing technique to Simplified1D CNN model.

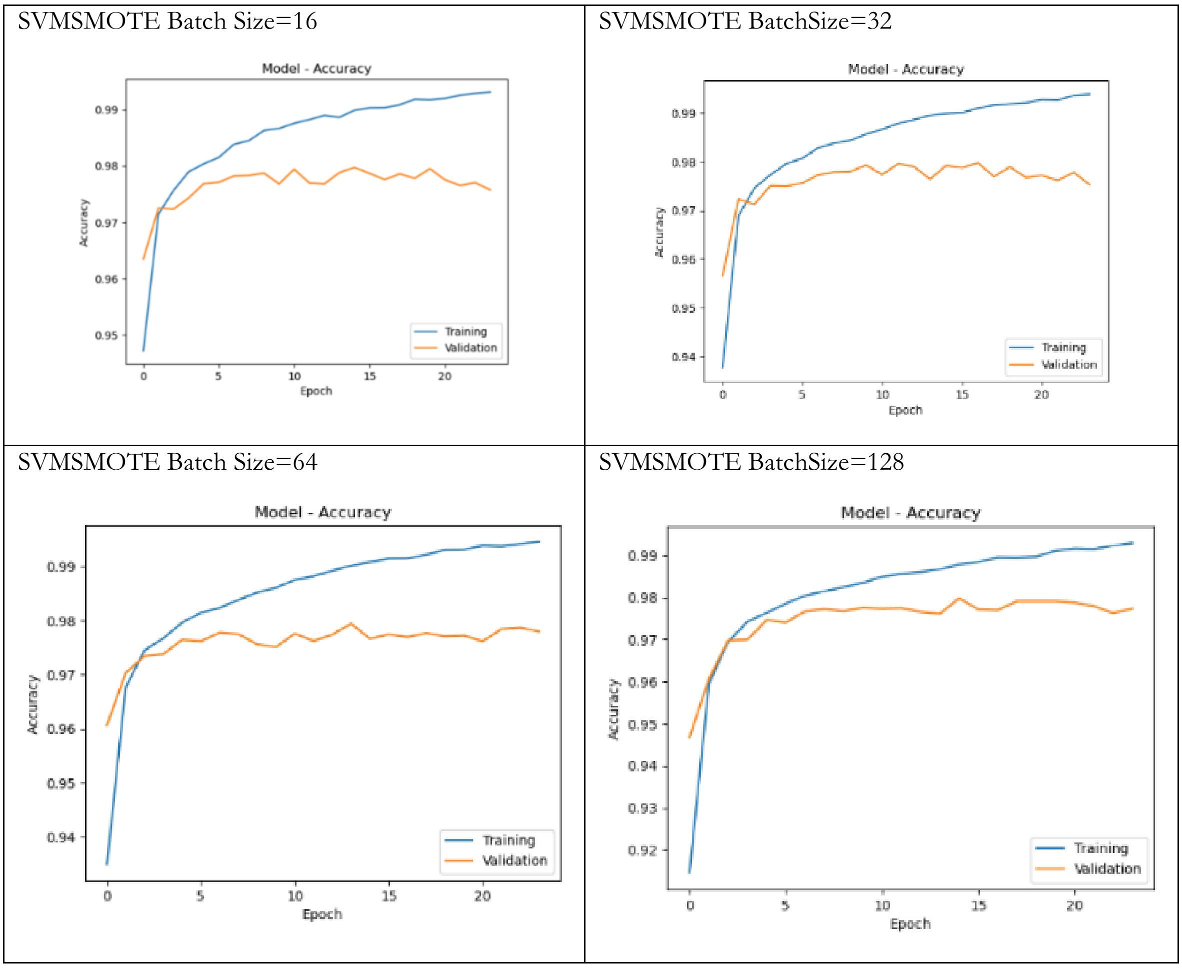


Fig. 10. Comparison of accuracy achieved during training and validation phases on applying SMOTETomek balancing technique to Simplified1D CNN model

eISSN: 2583-1194

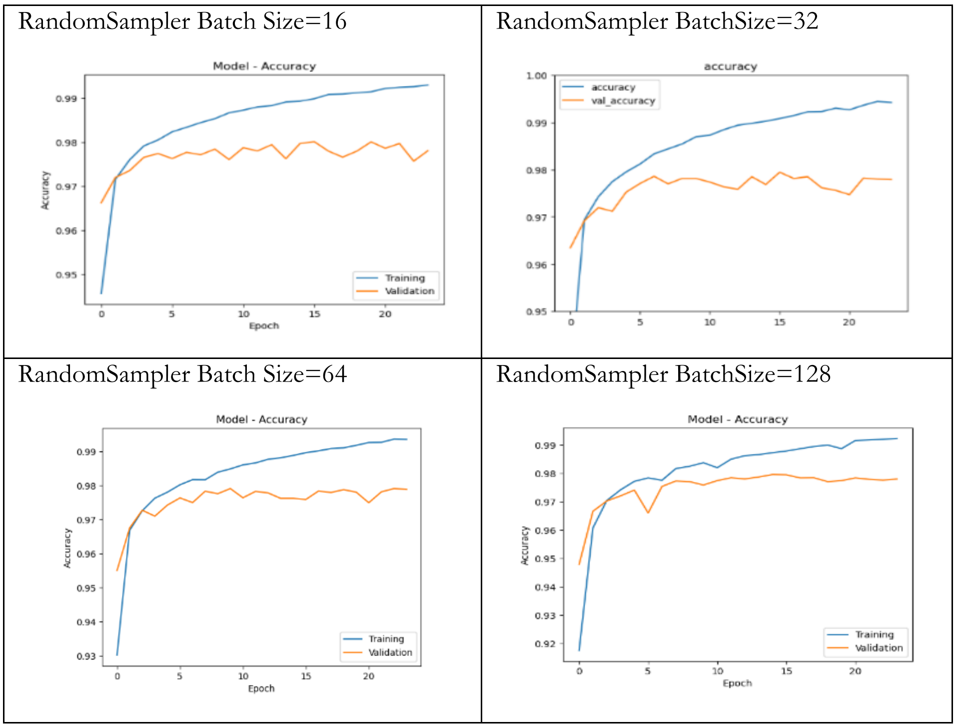


Fig. 11. Comparison of accuracy achieved during training and validation phases on applying RandomSampler balancing technique to Simplified1D CNN model

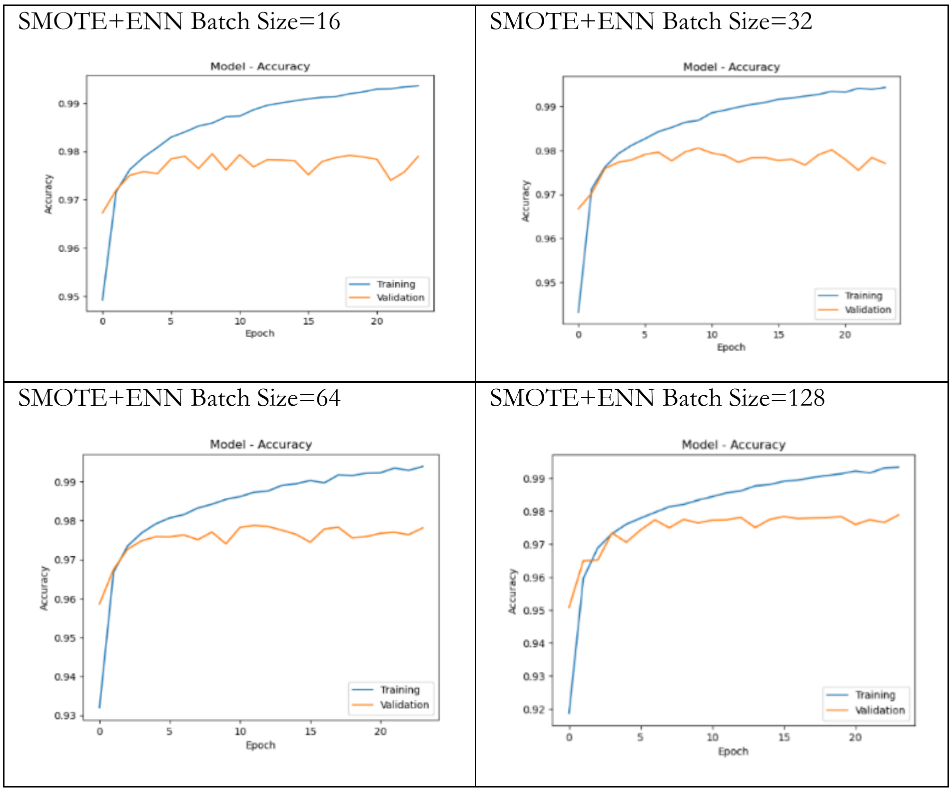


Fig. 12. Comparison of accuracy achieved during training and validation phases on applying SMOTE+ENN balancing technique to Simplified1D CNN model

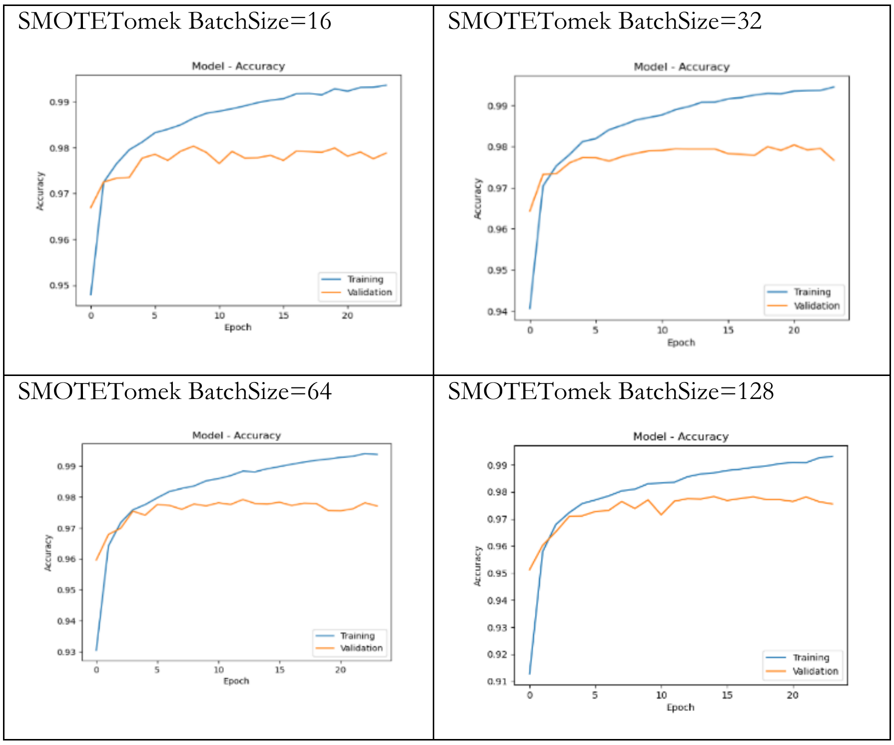


Fig. 13. Comparison of accuracy achieved during training and validation phases on applying SMOTETomek balancing technique to Simplified1D CNN model

Figures 6 to 13 represent the graph of the accuracy of the convolutional neural network (CNN), trained model on the MIT-BIH dataset. The x-axis is the epoch, and the y-axis is the accuracy. There are two lines on the graph, one labelled "Training" and one labelled "Validation." The training accuracy starts at around 0.91 and increases to around 0.99 for 24 epochs. The validation accuracy starts at around 0.92 and increases to around 0.98 for 24 epochs. Batch size is an important hyperparameter to consider when training a machine learning model. A hyperparameter is a setting that is external to the model itself and that can be adjusted to improve the model's performance. The batch size is the number of samples that are processed by the model at a time during training. By generating synthetic samples for the minority class, the balancing technique ensures that the model is exposed to a greater number of examples from the minority class. This has helped the model to learn to better classify samples from the minority class.

7 CONCLUSION

The current work has demonstrated the effectiveness of integrating data sampling techniques with Simplified 1D CNN models for arrhythmia classification, leading to enhanced accuracy and robustness of the model. Addressing class imbalance is essential for developing reliable and robust arrhythmia classification systems. Various data sampling techniques, including Random Oversampling, SMOTE, and ADASYN, have been applied to mitigate the effects of class imbalance. By balancing the class distribution, these techniques improve the model performance by ensuring equal consideration of all classes during training. While oversampling techniques such as SMOTE and ADASYN focus on generating synthetic samples for minority classes, hybrid approaches that combine oversampling and undersampling strategies have shown promising results in achieving a balanced representation of all classes while minimizing information loss. This approach can enable proactive interventions and preventive measures to mitigate the progression of cardiac disorders.

Overall analysis of crucial parameters (recall and F1-score) indicates that Simplified 1D CNN with BorderlineSMOTE performs better with batch sizes of 16 and 32, while ADASYN shows higher performance with batch sizes of 64 and 128, making it the superior oversampling technique in those cases. In the hybrid sampling technique, **SMOTETomek** performance is better across all the different batch sizes. The accuracy and ROC-AUC score vary from a range of 0.9752 to 0.9804 and 90.64% to 93.84% respectively. The ROC AUC score measures a classifier's ability to distinguish between positive and negative classes. A higher ROC AUC score above 90% indicates superior performance of the classifier model. This study adds to the existing body of knowledge by thoroughly examining the performance of 1D CNNs in arrhythmia classification using the MIT-BIH dataset, with a focus on the critical factor of dataset balance. The insights gained from this research will not only inform the design of more effective arrhythmia detection models but also contribute to the broader understanding of the impact of dataset characteristics on the performance of deep learning algorithms in healthcare applications.

Simplified 1D CNN models, combined with balanced data sampling techniques, have the potential to enhance the clinical utility of arrhythmia classification systems. By accurately identifying various types of arrhythmias from ECG signals, these models can assist health specialists in making timely and informed outcomes regarding healthcare services and treatment strategies. Despite its simplicity, the model achieved competitive performance compared to more complex architectures, making it suitable for resource-constrained environments such as mobile devices or wearable health monitors. In conclusion, the comparative analysis of a simplified 1D CNN-based model for arrhythmia classification using multiple balanced data sampling techniques provides valuable insights into the future development of accurate and reliable diagnostic tools for cardiac arrhythmias.

REFERENCES

[1] F. Khan et al., "ECG classification using 1-D convolutional deep residual neural network," PLoS ONE, vol. 18, no. 4, Art. no. e0284791, 2023.

[2] Alim and M. K. Islam, "Application of machine learning on ECG signal classification using morphological features," in Proc. IEEE Region 10 Symp. (TENSYMP), 2020.

[3] K. T. Rajgopal, A. F. J. Alshammari, K. Rashmi, R. Puneeth, G. Sumit, and Y. Ambika Naik, "An automated healthcare framework for single-lead ECG arrhythmia classification using lightweight CNN-BiLSTM and temporal attention," Natural Resources for Human Health (NRFHH), vol. 6, no. 1s, pp. 255–266, 2026.

[4] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," BMC Med. Inform. Decis. Mak., vol. 11, pp. 1–13, 2011.

[5] V. Werner de Vargas et al., "Imbalanced data preprocessing techniques for machine learning: A systematic mapping study," Knowl. Inf. Syst., vol. 65, no. 1, pp. 31–57, 2023.

[6] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," BMC Bioinformatics, vol. 14, pp. 1–16, 2013.

[7] H. He et al., "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.), 2008.

[8] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," arXiv preprint arXiv:1707.01836, 2017.

[9] J. He, J. Rong, L. Sun, H. Wang, Y. Zhang, and J. Ma, "A framework for cardiac arrhythmia detection from IoT-based ECGs," World Wide Web, vol. 23, pp. 2835–2850, 2020.

[10] D. Nankani and R. D. Baruah, "Ventricular arrhythmia classification and interpretation using residual neural network with guided backpropagation," in Proc. IEEE Region 10 Annu. Int. Conf. (TENCON), Auckland, New Zealand, Dec. 7–10, 2021, pp. 574–579.

[11] H S. Mahawar, "Recent advances in heart disease prediction from ECG signals: A survey of machine learning, deep learning, and explainable AI approaches," Int. J. Comput. Intell. Eng. (IJCIE), vol. 1, no. 2, pp. 42–48, 2026.

[12] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," ACM Comput. Surv., vol. 52, no. 4, pp. 1–36, 2019.

[13] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," Inf. Sci., vol. 415, pp. 190–198, 2017.

[14] J. Ferretti, V. Randazzo, G. Cirrincione, and E. Pasero, "1-D convolutional neural network for ECG arrhythmia classification," in Progresses in Artificial Intelligence and Neural Systems, 2021, pp. 269–279.

[15] M. Mishra, T. L. Sharath Kumar, and U. M. Ashwinkumar, "A novel approach to visualize arrhythmia classification using 1D CNN," in Proc. Int. Conf. Cognitive Comput. Cyber Phys. Syst., Cham, Switzerland: Springer Nature, 2023.

[16] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," Inf. Sci., vol. 405, pp. 81–90, 2017.

[17] P. K. Sekharamantry, P. M. Santhoshi, S. Gupta, A. M., K. L. Divya, and S. P. M., "CAF-Net: A convolution and attention fusion network with RR interval context for five-class arrhythmia detection from single-lead ECG," *Natural Resources for Human Health (NRFHH)*, vol. 6, no. 2s, pp. 452–465, 2026.

[18] D. Li, J. Zhang, Q. Zhang, and X. Wei, "Classification of ECG signals based on 1D convolution neural network," in Proc. IEEE 19th Int. Conf. e-Health Netw., Appl. Services (Healthcom), 2017, pp. 1–6.

[19] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, et al., "A deep convolutional neural network model to classify heartbeats," Comput. Biol. Med., vol. 89, pp. 389–396, 2017.

[20] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45–50, 2001.

[21] Ra M. Mishra and U. M. Ashwinkumar, "Time–frequency deep learning for high-accuracy arrhythmia classification: A wavelet-scalogram squeeze-and-excitation residual network," Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 18, no. 6s, pp. 639–648, 2026.

[22] S. Shraddha et al., "Classification of ECG arrhythmia using recurrent neural networks," Procedia Comput. Sci., vol. 132, pp. 1290–1297, 2018.

[23] S. Gupta, U. N. Ranjitha, J. G. S. Banu, K. L. Divya, T. R. Kulkarni, and S. Mannikeri, "Adaptive multi-band dynamic functional-connectivity graph attention network for Parkinson's disease classification from resting-state EEG," Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 18, no. 11s, pp. 840–858, 2026..

[24] Y. Xiang, J. Luo, T. Zhu, S. Wang, X. Xiang, and J. Meng, "ECG-based heartbeat classification using two-level convolutional neural network and RR interval difference," IEICE Trans. Inf. Syst., vol. 101, no. 4, pp. 1189–1198, 2018.

[25] T. Wongvorachan, S. He, and O. Bulut, "A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining," Information, vol. 14, no. 1, Art. no. 54, 2023.

[26] S. K. Pandey and R. R. Janghel, "Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE," Australas. Phys. Eng. Sci. Med., vol. 42, no. 4, pp. 1129–1139, 2019.

[27] E. Izci, M. A. Ozdemir, M. Degirmenci, and A. Akan, "Cardiac arrhythmia detection from 2D ECG images by using deep learning technique," in Proc. Medical Technologies Congress (TIPTEKNO), 2019, pp. 1–4.

[28] Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, pp. 65–69, 2019.

[29] M. G. Kendall and B. Babington Smith, "Randomness and random sampling numbers," J. Roy. Stat. Soc., vol. 101, no. 1, pp. 147–166, 1938.

[30] S. Noor, O. Tajik, and J. Golzar, "Simple random sampling," Int. J. Educ. Lang. Stud., vol. 1, no. 2, pp. 78–82, 2022.

[31] Gosain and S. Sardana, "Handling class imbalance problem using oversampling techniques: A review," in Proc. Int. Conf. Adv. Comput., Commun. Informatics (ICACCI), 2017.

[32] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in Proc. Int. Conf. Intelligent Computing, Berlin, Germany: Springer, 2005.

[33] N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

[34] X. Ge et al., "An empirical study of class rebalancing methods for actionable warning identification," IEEE Trans. Rel., 2023.

[35] Y. Tang et al., "SVMs modeling for highly imbalanced classification," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 39, no. 1, pp. 281–288, 2008.

[36] M. N. Nishat et al., "A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset," Scientific Programming, vol. 2022, pp. 1–17, 2022.

[37] Z. H. E. Wang et al., "SMOTETomek-based resampling for personality recognition," IEEE Access, vol. 7, pp. 129678–129689, 2019.