KSÜ Mühendislik Bilimleri Dergisi, 29(1), 2026
Araştırma Makalesi

KSU J Eng Sci, 29(1), 2026
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 30.12.2025
Kabul Tarihi : 10.02.2026

Received Date : 30.12.2025
Accepted Date : 10.02.2026

ECG SIGNAL CLASSIFICATION WITH DEEP LEARNING AND MACHINE LEARNING: REPRESENTATION OF QRS COMPLEXES IN TWO-DIMENSIONAL FRAMES AND DATA BALANCING WITH SMOTE

DERİN ÖĞRENME VE MAKİNE ÖĞRENİMİ İLE EKG SİNYALİ SINIFLANDIRMASI: İKİ BOYUTLU ÇERÇEVELERDE QRS KOMPLEKSLERİNİN TEMSİLİ VE SMOTE İLE VERİ DENGELEME

Mehmet Reşat ÖNER[1*] (ORCID: 0000-0003-2084-0181)
İsmail BABAOĞLU[1] (ORCID: 0000-0002-2503-1482)
Gülay TEZEL[1] (ORCID: 0000-0003-1698-0106)

[1] Konya Technical University, Department of Computer Engineering, Konya, Turkey
* Corresponding Author: Mehmet Reşat ÖNER, mroner@ktun.edu.tr

ABSTRACT

This study aims to support reliable ECG signal interpretation by reducing human-dependent variability through computer-aided analysis methods. Machine learning and deep learning methods were employed to examine 2D ECG representations and Synthetic Minority Over-Sampling Technique (SMOTE)-based balancing in ECG classification. Unlike existing ECG classification studies that typically address signal representation and class imbalance separately, this study jointly investigates the interaction between two-dimensional QRS representation and SMOTE-based data balancing within a unified experimental framework, thereby providing a systematic analysis of their combined impact on classification performance. Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and K-Nearest Neighbors (KNN) algorithms were implemented and comparatively analyzed. ECG beats from record 108 of the MIT-BIH Arrhythmia dataset were represented in a vision-based form for classification. To address severe class imbalance, SMOTE was applied only to the training data, and its effect on two-dimensional ECG representations was explicitly examined. Normal and Abnormal heartbeats were classified using a stratified 5-fold cross-validation strategy. Experimental results demonstrated that the CNN model achieved the most successful performance after applying SMOTE, reaching a weighted average F1-score of 99.82% ± 0.002, highlighting the combined effectiveness of two-dimensional QRS representation and data balancing in improving automated ECG classification.
Keywords: deep learning, electrocardiogram, QRS complex, SMOTE

ÖZET

Bu çalışma, bilgisayar destekli analiz yöntemleri aracılığıyla insan kaynaklı değişkenliği azaltarak güvenilir ECG sinyali yorumlamasını desteklemeyi amaçlamaktadır. Makine öğrenmesi ve derin öğrenme yöntemleri, ECG sınıflandırmasında iki boyutlu ECG temsillerini ve Synthetic Minority Over-sampling Technique (SMOTE) tabanlı veri dengelemenin etkisini incelemek amacıyla kullanılmıştır. Mevcut ECG sınıflandırma çalışmalarında genellikle sinyal temsili ve sınıf dengesizliği ayrı ayrı ele alınırken, bu çalışmada iki boyutlu QRS temsili ile SMOTE tabanlı veri dengelemenin etkileşimi tek bir deneysel çerçeve içerisinde birlikte incelenmiş ve bu yaklaşımların sınıflandırma performansı üzerindeki birleşik etkisi sistematik olarak analiz edilmiştir. Yapay Sinir Ağları (ANN), Evrişimsel Sinir Ağları (CNN) ve K-En Yakın Komşu (KNN) algoritmaları uygulanmış ve karşılaştırmalı olarak analiz edilmiştir. MIT-BIH Aritmi veri setinin 108 numaralı kaydından elde edilen ECG atımları, sınıflandırma amacıyla görsel tabanlı bir biçimde temsil edilmiştir. Ciddi sınıf dengesizliğini gidermek amacıyla SMOTE yalnızca eğitim verisine uygulanmış ve iki boyutlu ECG temsilleri üzerindeki etkisi açıkça incelenmiştir. Normal ve Anormal kalp atımları, tabakalı 5 katlı çapraz doğrulama stratejisi kullanılarak sınıflandırılmıştır. Deneysel sonuçlar, SMOTE uygulandıktan sonra CNN modelinin %99,82 ± 0,002 ağırlıklı ortalama F1-skoru ile en başarılı performansı elde ettiğini göstermiş; bu durum, iki boyutlu QRS temsili ile veri dengelemenin birlikte kullanımının otomatik ECG sınıflandırma performansını artırmadaki etkinliğini ortaya koymuştur.
Anahtar Kelimeler: derin öğrenme, elektrokardiyogram, QRS kompleksi, SMOTE

INTRODUCTION

Cardiovascular diseases are among the most significant health problems both globally and in Turkey, constituting the leading cause of worldwide mortality according to the World Health Organization. In Türkiye specifically, it is reported that approximately 57 out of every 1000 individuals suffer from heart disease (Onat et al., 1991). Early diagnosis and timely intervention in cardiac disorders are of critical importance for the course of the disease. At this point, one of the most commonly employed diagnostic methods is the Electrocardiogram (ECG) test. ECG records the electrical activity of the heart, enabling the monitoring of cardiac functions, and is widely used in the diagnosis of various conditions such as arrhythmias. However, the accurate interpretation of ECG data largely depends on the expertise of physicians. The diagnostic process can be both time-consuming and prone to misinterpretations, particularly when conducted by inexperienced or non-specialist practitioners. Such challenges may hinder timely access to appropriate treatment and pose risks for clinical outcomes.

In order to address these limitations, computer-aided diagnostic systems have been increasingly investigated in recent years. The rapid advancements in artificial intelligence have provided substantial advantages in the automated processing and classification of ECG signals. Machine learning and deep learning methods have shown promising results in developing decision-support systems that assist physicians with high levels of accuracy. In the literature, the number of studies applying various algorithms for automatic analysis of ECG signals has been steadily increasing. Nevertheless, challenges such as data imbalance, noise sensitivity of signals, and maintaining stable classification performance remain significant research areas.

For instance, a novel voting rule for interval-valued intuitionistic fuzzy-based classifiers was proposed. Their work focused on imbalanced data classification, which aims to categorize complex datasets with unequal class distributions. Traditional classifiers often perform poorly in such contexts. To address this, a data redistribution approach involving oversampling and noise reduction was applied. Interval-valued fuzzy and intuitionistic fuzzy set-based KNN classifiers were employed, and a novel voting rule was introduced using the Iterative Partitioning Filter (IPF). Comparative experiments with SMOTE and its variants demonstrated the effectiveness of their approach (Zeraatkar and Afsari, 2021).

Similarly, a system for diagnosing COVID-19, normal, or viral pneumonia cases using chest X-ray images was developed. Identifying positive cases from chest radiographs provides a rapid and effective solution for hospitals. While deep learning techniques achieved high performance in classifying COVID-19 images, the limited number of images resulted in class imbalance issues. To overcome this, balanced datasets were used, and six CNN models were trained with transfer learning across three datasets. Weighted Categorical Loss and SMOTE methods were examined to address imbalance. DenseNet201 and VGG-19 achieved the best results, with CheXNet combined with WCL yielding the highest accuracy and performance (Chamseddine et al., 2022).

An experimental study was conducted to investigate the performance of different metaheuristic algorithms for training Artificial Neural Networks (ANNs) in medical data classification tasks. Experiments on 15 benchmark medical datasets compared Levenberg–Marquardt with 13 other metaheuristics. Multiple evaluation metrics and Multi-Criteria Decision Making (MCDM) analyses highlighted the superior performance of the Equilibrium Optimizer algorithm (Si et al., 2022).

Dunstan's infant cry dataset, preprocessed with MFCC and energy features, was utilized to classify five types of infant cries using SVM, MLP, and CNN classifiers. Loss and accuracy metrics were analyzed for MLP and CNN, with performance evaluated through multiple measures. The CNN model demonstrated superior results, achieving the highest accuracy of 92.1% after 10 runs (Abbaskhah et al., 2023).

A hybrid model combining Convolutional Neural Networks (CNN) and Extreme Learning Machine (ELM) was proposed for classifying chest diseases. Their computer-aided system identified 17 different lung diseases from chest X-ray (CXR) images. The CNN-ELM algorithm achieved 90.92% accuracy and 96.93% AUC, with COVID-19 and TB detected with 99.37% and 99.98% accuracy, respectively. The model outperformed state-of-the-art approaches and maintained robustness across multiple datasets (Nahiduzzaman et al., 2023).

SDIF-CNN, a novel visualization-based architecture, was introduced for malware detection and classification. Feature extraction was performed with VGG16, VGG19, ResNet50, and InceptionV3, followed by horizontal concatenation into a single feature map. Irrelevant features were filtered, and six machine learning classifiers were trained. The MLP classifier achieved the highest performance with 98.55% accuracy on the MalImg dataset. The proposed approach also yielded successful results in real-world malware detection (Kumar and Panda, 2023).

A new SMOTE technique enhanced with an additional algorithm was developed to address class imbalance issues. Although SMOTE is widely used, it is prone to generating noisy samples. Their proposed SMOTE-kTLNN method, based on a two-layer nearest neighbor classifier, rebalances data and removes noisy samples via majority voting. Experiments demonstrated that SMOTE-kTLNN consistently outperformed other methods across multiple evaluation metrics (Sun et al., 2024).

As seen in the literature, artificial intelligence techniques, particularly deep learning and machine learning, have proven effective in providing robust solutions across a wide range of applications, including biomedical signal analysis. This potential serves as the core motivation for this study, which aims to improve ECG signal classification under conditions of significant class imbalance. While numerous studies have applied machine learning and deep learning approaches to ECG analysis, two fundamental research gaps remain unaddressed. First, it is still unclear whether the conversion of one-dimensional QRS complexes to two-dimensional representations confers meaningful learning capabilities beyond traditional one-dimensional signal processing, or whether it merely serves as a structural adaptation for convolutional models. Second, although class imbalance is a well-known challenge in ECG datasets, the systematic application and effects of SMOTE on two-dimensional ECG representations have not been comprehensively investigated. To address these shortcomings, this study utilizes ECG beats extracted from the 108th recording of the MIT-BIH Arrhythmia dataset and proposes a visual-based framework by converting one-dimensional QRS complexes into two-dimensional 150x150 grids. Furthermore, the effect of SMOTE-based data equalization on these representations is systematically investigated. This study aims to explicitly distinguish the effects of spatial signal representation and class imbalance management using KNN, ANN, and CNN algorithms within a combined experimental design employing stratified 5-fold cross-validation.

MATERIALS AND METHOD

Proposed Method

In this study, three classifiers were employed to categorize arrhythmias labeled as Normal and Abnormal from ECG signals obtained from the MIT-BIH Arrhythmia dataset, and their performances were compared. KNN, ANN, and CNN algorithms were selected as classifiers, and two different experimental settings were considered; classification using the original imbalanced data and classification after balancing the dataset using the SMOTE algorithm. This study demonstrates both the effectiveness of the SMOTE algorithm in handling class imbalance and the classification performance of two different paradigms, namely machine learning and deep learning.

Experiments were conducted using ECG beats extracted from record 108 of the MIT-BIH Arrhythmia dataset, which was selected due to its noticeably imbalanced distribution of Normal and Abnormal heartbeats, enabling a controlled evaluation of SMOTE-based data balancing.

To enable the application of the CNN model, one-dimensional ECG signals were transformed into two-dimensional representations by framing QRS complexes into 150x150 matrices, thereby allowing the use of two-dimensional convolutional layers. Specifically, each QRS complex was extracted using a fixed window length of 150 samples (approximately 0.42 seconds at a sampling frequency of 360 Hz). The resulting one-dimensional segment was then expanded into a square matrix of size 150x150 by replicating the signal along one dimension, producing a two-dimensional representation suitable for convolutional processing. Although this transformation is obtained by replicating the one-dimensional segment along a single axis without introducing new physiological information, it is intentionally designed to provide a structured spatial arrangement. This approach allows the model to leverage sophisticated spatial feature maps and shared-weight filtering mechanisms that are extensively optimized in the two-dimensional deep learning ecosystem. By mapping the temporal morphology of the QRS complex into a two-dimensional grid, convolutional kernels can effectively capture subtle shape variations, local transitions, and edge-like patterns across multiple receptive fields. This spatial organization facilitates a more nuanced extraction of discriminative morphological characteristics compared to standard one-dimensional processing. Therefore, the proposed representation does not artificially inflate performance through data duplication; rather, it provides a more robust framework for the CNN to exploit its inherent ability to learn complex patterns in a spatially organized form.

An example of this transformation, illustrating how a one-dimensional QRS segment is embedded into a 150x150 two-dimensional frame, is visually presented in Figure 1.

Figure 1. Two-dimensional 150x150 Frame Representation of a One-dimensional QRS Segment.

In this study, QRS complexes were identified using expert-tagged R-peak locations provided in conjunction with the MIT-BIH Arrhythmia Database. Specifically, the wfdb library was used to read the tagging files, and the R-peak locations provided by cardiologists were used directly, rather than implementing an additional automated QRS detection algorithm. This approach ensures consistency with established benchmarks and eliminates potential variability due to detector-specific errors. Each QRS complex was segmented using a fixed-length window of 150 samples centered around the tagged R-peak. This window length was chosen to accurately capture QRS morphology while maintaining a consistent segment size across all beats. Since the MIT-BIH signals were prerecorded clinical ECG signals with sufficient signal quality, no additional filtering was applied. Furthermore, raw amplitude values were preserved to avoid altering morphological features critical for arrhythmia classification.

Since the SMOTE algorithm operates on vector-based feature representations, each 150x150 two-dimensional ECG frame was temporarily reshaped into a one-dimensional vector of length 22500. SMOTE was applied exclusively to the training data to generate synthetic samples for the minority class using KNN interpolation in the feature space. After the oversampling process, the generated synthetic vectors were reshaped back into their original 150x150 two-dimensional form. This procedure ensures that SMOTE contributes only to data balancing while preserving the spatial structure of the ECG representations used for CNN-based classification.

Since the KNN algorithm and ANN model cannot directly process two-dimensional data, the two-dimensional ECG representations were reshaped into one-dimensional feature vectors of 22500 dimensions before being fed into these models. Figure 2 illustrates the proposed methodology. Section (1) of Figure 2 presents the details of the classification performed with imbalanced data, while Section (2) visualizes the classification process after applying the SMOTE technique for data balancing.

Dataset

In this study, the publicly available MIT-BIH Arrhythmia dataset was utilized. The MIT-BIH Arrhythmia dataset consists of two-channel ambulatory ECG recordings obtained from 47 individuals examined between 1975 and 1979 by the BIH Arrhythmia Laboratory. The ECG recordings were collected from a mixed population of hospitalized patients at Beth Israel Hospital in Boston (approximately 60%) and outpatients (approximately 40%). The records were digitized at a sampling rate of 360 samples per second per channel with 11-bit resolution over a 10 mV range. Each recording was independently annotated by two or more cardiologists.

Although the MIT-BIH Arrhythmia dataset includes ECG recordings from multiple subjects and a total of 16 annotated heartbeat classes, this study focuses on ECG beats extracted exclusively from record 108. This record was selected due to its noticeably imbalanced distribution of Normal and Abnormal heartbeats, which enables a controlled evaluation of SMOTE-based data balancing techniques. By focusing on a single record, the proposed methodology can be systematically analyzed without the confounding effects of inter-patient variability.

In the original MIT-BIH annotation scheme, one class is labeled as Normal, while the remaining heartbeat types correspond to various arrhythmias. In this study, all non-Normal heartbeat classes were grouped under the Abnormal label, resulting in a binary classification task distinguishing between Normal and Abnormal heartbeats. As a result of the fixed-length windowing procedure, a total of 1458 (1391 Normal and 67 Abnormal) QRS segments that satisfied the 150-sample window criteria were retained and used in the experiments.

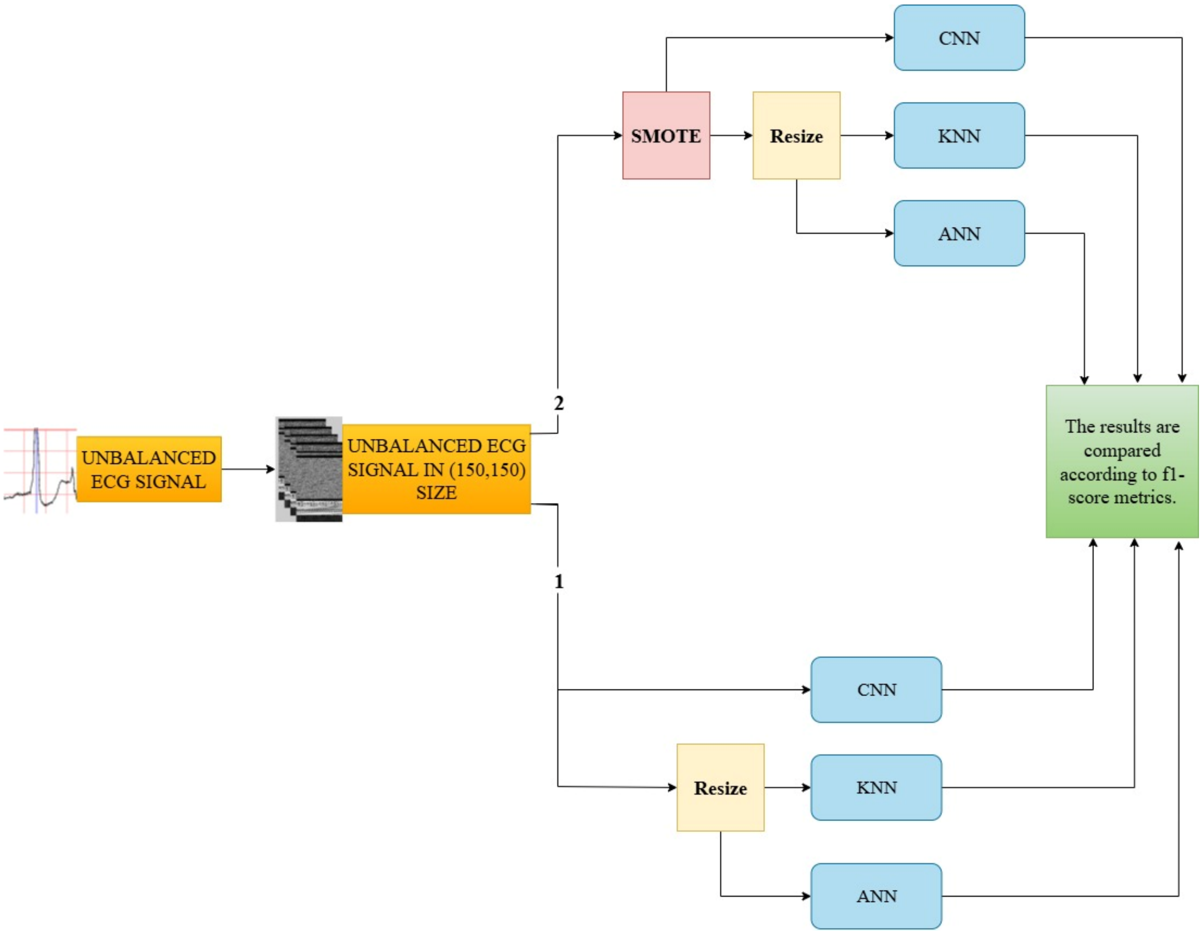

Figure 2. Proposed Method

Deep Learning and Machine Learning

Machine learning is a discipline of artificial intelligence that enables computers to acquire the ability to learn from data and experiences. It aims to make predictions about future data based on knowledge accumulated from past data using statistical and mathematical methods (Peng, 2015). Machine learning has a wide range of applications, including image and speech recognition, natural language processing, financial analysis, healthcare systems, and cybersecurity. Well-known machine learning algorithms include Decision Trees, Support Vector Machines, K-Means, and KNN.

Deep learning is a subfield of machine learning based on artificial neural networks, developed to learn more abstract and complex representations of data. Its multilayered network architecture allows it to decompose data into increasingly abstract and intricate features for learning. Unlike traditional machine learning, deep learning does not require manual rule input; instead, it learns through relationships derived from abstract representations of data (Goodfellow et al., 2016; Özcan, 2014). Deep learning finds extensive applications in speech recognition and synthesis, autonomous vehicles, healthcare, and entertainment. Well-known deep learning algorithms include CNN, Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM).

K-Nearest Neighbors (KNN)

KNN is a non-parametric machine learning method widely used for classification due to its simplicity and discriminative capability. However, KNN has two major disadvantages: Low computational efficiency and dependence on the selection of the k value. Being a lazy learning method, its applicability is limited for large datasets. Furthermore, the classification performance largely depends on the chosen k value. Although some studies have aimed to reduce the dependency on k (S. Zhang et al., 2018). The most effective approach to date remains conducting multiple experiments with different k values to determine the optimal choice (Guo et al., 2003).

Artificial Neural Network (ANN)

ANN is an information-processing system inspired by biological neural networks. These networks consist of a series of interconnected units called neurons, which process input data and generate outputs. ANN typically comprises three main layers: The input layer, the hidden layers, and the output layer. The input layer receives raw data and passes it to the hidden layers. Hidden layers process this information, transforming it into more abstract and meaningful representations. Each hidden layer receives outputs from the previous layer, processes them through weights and activation functions, and forwards the results to the subsequent layer. Finally, the output layer produces the network's predictions or classifications. ANN learns and optimizes through forward and backward propagation processes across these layers. This architecture provides strong modeling capabilities for complex datasets and large-scale problems (Wu and Feng, 2018).

Convolutional Neural Network (CNN)

CNN is a subfield of neural network-based machine learning, particularly effective for processing two-dimensional data such as images and videos. CNNs utilize three main types of layers to process data and extract features: Convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply various filters to the input data to generate feature maps, extracting hierarchical features ranging from low-level patterns such as edges and corners to more complex structures. Pooling layers reduce the dimensions of feature maps, thereby lowering computational costs. Fully connected layers transform the incoming features into a flat vector to produce system outputs for tasks such as classification or regression. Through forward and backward propagation across these layers, CNN can achieve high accuracy learning and generalization on large datasets (Tuncer et al., 2021).

Synthetic Minority Over-Sampling Technique (SMOTE)

SMOTE is a widely used method developed to address imbalanced classification problems. This technique is employed to augment the number of samples in the minority class. Initially, for each minority class sample, denoted as x, the k nearest neighbors are identified, typically with k=5. Among these neighbors, those belonging to the minority class are selected. New synthetic samples are then generated by randomly interpolating between x and its selected neighbors at a certain rate. This process effectively increases the number of minority class samples, thereby balancing the dataset (Chawla et al., 2002).

Evaluation Metrics

The method responsible for producing the results of a classification task is referred to as a classifier. Interpreting and comparing the relationship between the predicted class labels generated by a classifier and the true class values is scientifically valuable and necessary. To perform such evaluations, a comparison framework must be established. This framework is represented by the confusion matrix. The number of rows and columns in the confusion matrix corresponds to the number of classes included in the study (Ruuska et al., 2018).

Confusion matrices consist of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. To evaluate the classification performance of a model, several metrics are calculated using these values. The commonly used metrics include Accuracy, Precision, Recall, and F1-Score. The formulas for these metrics are provided in Equations 1, 2, 3, and 4, respectively (Gonzalez-Huitron et al., 2021).

$$accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{1}$$

$$\mathrm{precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{2}$$

$$recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{3}$$

$$f1 - score = \frac{2*\mathrm{precision}*\mathrm{recall}}{\mathrm{precision+recall}} \tag{4}$$

In this study, model performance was evaluated using stratified 5-fold cross-validation. For each fold, the evaluation metrics were computed independently, and the final performance values were reported as the mean and standard deviation across the five folds. Given the imbalanced nature of the dataset, the F1-score was considered the primary evaluation metric, as it provides a balanced measure by jointly considering precision and recall.

EXPERIMENTAL RESULTS

In this study, the classification of Normal and Abnormal heartbeats was conducted using an imbalanced dataset derived from record number 108 of the MIT-BIH Arrhythmia database, comprising a total of 1458 ECG beats (1391 Normal and 67 Abnormal). To compare the performances of deep learning and machine learning approaches and to evaluate the impact of the SMOTE algorithm on classification success, CNN, ANN, and KNN models were employed. To ensure a fair and transparent comparison, the ANN and CNN models were trained under the same strategy with 100 epochs and a batch size of 32. The ANN model was structured with fully connected layers containing 128, 64, and 32 neurons, respectively, utilizing the ReLU activation function and dropout layers to prevent overfitting. The CNN model was designed to process two-dimensional 150x150 QRS representations. The architecture consists of two convolutional stages: the first layer utilizes 32 filters of size (3,3) followed by a (2,2) max-pooling and a 0.25 dropout rate, while the second layer employs 64 filters of size (3,3) followed by identical pooling and dropout configurations. The features are then flattened and fed into a 128-neuron dense layer with 0.5 dropout. For both neural networks, the Adam optimizer was used, and a Sigmoid activation function was implemented in the final output layer to address the binary classification problem. In the KNN model, a neighbor count of k=40 was utilized, as preliminary experiments indicated that this value provided a more balanced performance and demonstrated robustness against noise. All experiments were conducted using stratified 5-fold cross-validation to maintain the sample ratios between classes. Training and testing procedures were executed independently for each fold, and the results were combined to generate aggregated confusion matrices. Model performances were reported using the mean and standard deviation of the metrics calculated for each fold. Due to the imbalanced nature of the dataset, the weighted average F1-score, which considers precision and recall together, was taken as the primary performance metric for evaluation. Comparing the experiments conducted before and after the application of the SMOTE algorithm, it was observed that SMOTE enhanced the classification performance of all models by strengthening the representation of minority class samples. In the SMOTE algorithm, a KNN approach was adopted for each minority class sample, with the number of nearest neighbors set to k=5, as commonly used in the literature. Synthetic samples were generated via linear interpolation between a minority class sample and a randomly selected sample from its k nearest neighbors. In this process, the number of minority class samples was increased by generating a new synthetic instance at a random point within the feature space between the two samples. To prevent data leakage, SMOTE was applied exclusively to the training data, and synthetic samples were strictly excluded from the test set. Thus, the generalizability of the model on real-world data was preserved. Since two-dimensional 150x150 QRS frames cannot be processed directly by the SMOTE algorithm, each frame was first converted into a single 22500-dimensional feature vector, and the SMOTE process was applied within this high-dimensional feature space. After the minority class was balanced with the majority class, the resulting synthetic samples were reshaped back to 150x150 dimensions for the CNN model to remain compatible with the convolutional layers. This approach enabled leveraging both the vector-based sampling capabilities of SMOTE and the effective learning of two-dimensional spatial information by CNN. The comparative test results of the conducted experiments are presented in Table 1.

Table 1. Test Results

Algorithms	**Weighted average F1 score**		**Accuracy**	
	Before applying SMOTE	**After applying SMOTE**	**Before applying SMOTE**	**After applying SMOTE**
KNN	%93.16±0.002	%95.29±0.008	%95.40±0.001	%95.20±0.008
ANN	%93.83±0.012	%96.55±0.009	%95.75±0.006	%96.55±0.009
CNN	%96.17±0.009	%99.82±0.002	%96.91±0.006	%99.82±0.002

Upon examination of Table 1, it is evident that the application of the SMOTE algorithm provided a marked improvement in the performance of all classifiers. Specifically, when evaluated in terms of the weighted average F1-score, significant increases were achieved across the KNN, ANN, and CNN models following the implementation of SMOTE. This observation demonstrates that a more balanced representation of minority class samples directly enhances classification success. In terms of the accuracy metric, a consistent after-SMOTE increase was particularly observed in the deep learning-based models.

The CNN model, which exhibited successful performance compared to other approaches both before and after data balancing, reached its highest and most stable classification success with the application of SMOTE. The fact that the model's weighted average F1-score reached a high value of 99.82% with a very low standard deviation (± 0.002) serves as an indicator of reliable and consistent results across different data splits. The ANN model significantly improved its performance (particularly regarding the weighted average F1-score metric) after SMOTE experiments, producing results close to the CNN model and emerging as a strong alternative. Although the KNN model enhanced its performance after SMOTE, it remained behind the deep learning-based methods in terms of absolute success. These findings reveal that deep learning approaches are more effective than traditional machine learning methods, especially when operating on two-dimensional ECG representations.

Overall, the obtained results confirm that SMOTE data balancing techniques significantly improve classification performance in imbalanced ECG datasets, and that the CNN model provides more successful outcomes under both balanced and imbalanced data conditions. The aggregated confusion matrices for the CNN model, which delivered the most successful classification performance both before and after the data balancing process, are presented in Figure 3 and Figure 4, respectively.

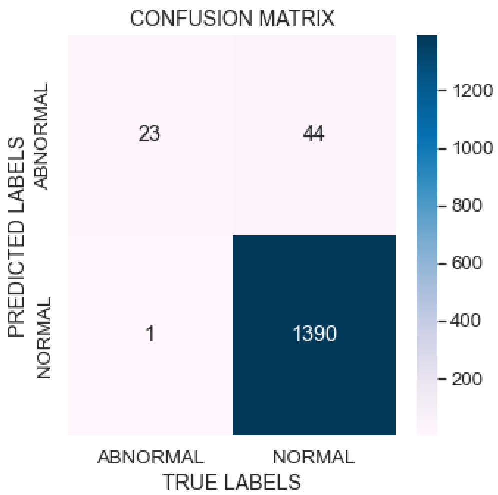

Figure 3. Aggregated Confusion Matrix of the CNN Model Before SMOTE

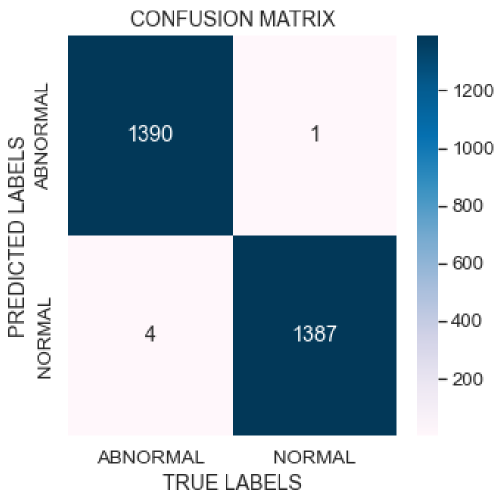

Figure 4. Aggregated Confusion Matrix of the CNN Model After SMOTE

As illustrated in Figures 3 and 4, the CNN model exhibits a clear improvement in classification performance following the application of the SMOTE algorithm. Figure 3 presents the aggregated confusion matrix obtained from experiments conducted on the original imbalanced dataset, where the majority of misclassifications are associated with the minority (Abnormal) class. Although the model successfully identifies Normal heartbeats with a very low false-alarm rate, a visible portion of Abnormal samples is erroneously classified as Normal, reflecting the adverse impact of class imbalance.

Upon the implementation of SMOTE, the aggregated confusion matrix shown in Figure 4 reveals a significant increase in the recognition of the Abnormal class. While the number of correctly classified Abnormal samples increases, the misclassification of minority class instances decreases substantially. Simultaneously, the model maintains its successful performance on the Normal class by preserving a high number of true negative predictions and a very low false positive rate.

These observations confirm that the SMOTE algorithm effectively alleviates the class imbalance problem by enriching the representation of minority class samples during training. When integrated with the CNN architecture, SMOTE leads to a more balanced decision boundary, enhancing overall classification performance without compromising the accuracy of the majority class. Consequently, the proposed SMOTE-enhanced CNN model demonstrates a more reliable and robust performance in detecting Abnormal heartbeats, which is particularly critical for clinical decision support systems. Table 2 shows a comparison of this study with similar studies.

Table 2. Studies Using Similar Materials in Literature

	Summary of the Study	Dataset	Convolution Layer	Results
(Chen et al., 2021)	The proposed approach uses Borderline-SMOTE to balance one-dimensional EEG signal data. Binary classification of valence and arousal is performed using the CNN model.	DEAP	1D	%97.76 (accuracy)
(P. Zhang et al., 2022)	The proposed approach Employs a 1D-MobileNet model to classify epileptic seizure data and proposes a new algorithm, BNNSMOTE, to address the imbalanced data problem.	CHB-MIT	1D	%91.90 (f1 score)
(Ullah et al., 2022)	The proposed approach Combines SMOTE and Random Undersampling methods to overcome data imbalance. Classifies four different class labels using the DenseNet model.	MIT-BIH and INCART	1D	%98.91 (f1 score)
Proposed Method	In this study, after adding a second dimension to ECG signals, SMOTE was applied for balancing, and the signals were classified as Normal and Abnormal.	MIT-BIH	2D	%99.82 (f1 score)

As shown in Table 2, introducing a second dimension to the ECG signals in the proposed method enables a richer data representation. Furthermore, it enhances the model’s flexibility and its capacity to learn complex patterns. By incorporating this second dimension, the use of two-dimensional convolutional layers in the CNN model becomes feasible, which has played a significant role in achieving the reported successful outcomes. Addressing data imbalance through SMOTE substantially improves classification performance, particularly in imbalanced datasets. This technique, working in harmony with the employed dataset, has proven more effective than algorithms such as Borderline-SMOTE or BNNSMOTE, contributing to the achievement of weighted average F1-score values as high as 99.82%.

CONCLUSION AND DISCUSSION

In this study, KNN, ANN, and CNN algorithms were investigated to classify Normal and Abnormal arrhythmias using the MIT-BIH arrhythmia dataset, and the effectiveness of the SMOTE algorithm in addressing class imbalance was evaluated. Experimental results revealed a comparison between the performances of deep learning and machine learning techniques, proving the superiority of deep learning methods. An examination of the existing literature indicates that the combination of one-dimensional signals and the application of the SMOTE technique for data balancing has not been previously explored. Furthermore, while SMOTE is generally not preferred during the data balancing process in two-dimensional studies, one-dimensional convolutional layers are commonly utilized for one-dimensional data.

This study highlights the success of the SMOTE algorithm on two-dimensional data and demonstrates that converting signals into two dimensions enables the development of CNN models equipped with two-dimensional convolutional layers, which exhibit more effective performance. Consequently, the study introduces an innovative perspective to literature and validates its efficacy. Analysis of the experimental results indicates that, as expected, the SMOTE algorithm improved the weighted average F1-scores across all KNN, ANN, and CNN algorithms. The findings confirm that SMOTE performs highly effective data balancing. Following the application, the CNN algorithm exhibited the best performance, achieving a weighted average F1-score and an accuracy rate of 99.82%.

The CNN model demonstrated its superiority over other models by exhibiting successful classification performance on both imbalanced and balanced datasets. Meanwhile, the ANN model also yielded successful results, achieving a weighted average F1-score and accuracy rate of 96.55%, thereby suggesting its potential as a viable alternative to the CNN. These results emphasize that deep learning techniques are more effective than traditional machine learning methods for the dataset at hand.

Overall, this study demonstrates that deep learning approaches capable of preserving spatial information are more effective than traditional machine learning methods in ECG signal classification. Additionally, class imbalance correction via SMOTE not only improves classification accuracy but also contributes to more reliable and consistent detection of abnormal arrhythmias. While the proposed framework focuses on binary Normal/Abnormal classification, many aspects remain open for future research. A natural extension of this study involves multiclass arrhythmia classification to distinguish different types of abnormal rhythms. Furthermore, validation of the proposed approach on larger and independent ECG datasets and the adoption of patient-independent experimental setups will provide a more comprehensive assessment of its generalizability. Future studies could also explore alternative data imbalance correction strategies beyond SMOTE and examine their interaction with two-dimensional ECG representations. Such extensions are expected to further enhance the success and applicability of the proposed framework.

Artificial Intelligence Contribution Statement

Artificial intelligence was utilized exclusively for language editing purposes. The study design, methodology, data analysis, results, figures, and manuscript structure were fully developed and prepared by the authors without any Artificial intelligence involvement in content generation.

REFERENCES

Abbaskhah, A., Sedighi, H., & Marvi, H. (2023). Infant cry classification by MFCC feature extraction with MLP and CNN structures. *Biomedical Signal Processing and Control,* 86, 105261. doi:https://doi.org/10.1016/j.bspc.2023.105261

Chamseddine, E., Mansouri, N., Soui, M., & Abed, M. (2022). Handling class imbalance in COVID-19 chest X-ray images classification: Using SMOTE and weighted loss. *Applied Soft Computing,* 129, 109588. doi:https://doi.org/10.1016/j.asoc.2022.109588

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research,* 16, 321-357. doi:https://doi.org/10.1613/jair.953

Chen, Y., Chang, R., & Guo, J. (2021). Effects of data augmentation method borderline-SMOTE on emotion recognition of EEG signals based on convolutional neural network. *IEEE Access,* 9, 47491-47502. doi:https://doi.org/10.1109/access.2021.3068316

Gonzalez-Huitron, V., León-Borges, J. A., Rodriguez-Mata, A., Amabilis-Sosa, L. E., Ramírez-Pereda, B., & Rodriguez, H. (2021). Disease detection in tomato leaves via CNN with lightweight architectures implemented in Raspberry Pi 4. *Computers and Electronics in Agriculture,* 181, 105951. doi:https://doi.org/10.1016/j.compag.2020.105951

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning.* (Vol. 1): MIT press Cambridge. doi: https://doi.org/10.1007/s10710-017-9314-z.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). *KNN model-based approach in classification.* Paper presented at the OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". doi: https://doi.org/10.1007/978-3-540-39964-3_62

Kumar, S., & Panda, K. (2023). SDIF-CNN: Stacking deep image features using fine-tuned convolution neural network models for real-world malware detection and classification. *Applied Soft Computing,* 146, 110676. doi:https://doi.org/10.1016/j.asoc.2023.110676

Nahiduzzaman, M., Goni, M. O. F., Hassan, R., Islam, M. R., Syfullah, M. K., Shahriar, S. M., . . . Kowalski, M. (2023). Parallel CNN-ELM: A multiclass classification of chest X-ray images to identify seventeen lung diseases including COVID-19. *Expert Systems with Applications*, 229, 120528. doi:https://doi.org/10.1016/j.eswa.2023.120528

Onat, A., Şurdumavcı, G., Şenocak, M., Örnek, E., Gözükara, Y., Karaaslan, Y., . . . Özcan, R. (1991). Türkiye'de erişkinlerde kalp hastalığı ve risk faktörleri sıklığı taraması: 3. Kalp hastalıkları prevalansı. *Türk Kardiyoloji Derneği Arşivi,* 19(1), 26-33.

Özcan, H. (2014). Çok düşük çözünürlüklü yüz imgelerinde derin öğrenme uygulamaları. *Deniz Harp Okulu/Deniz Bilimleri ve Mühendisliği Enstitüsü, İstanbul, Erişim adresi: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp*.

Peng, H. (2015). *Air quality prediction by machine learning methods.* University of British Columbia. doi: https://doi.org/10.14288/1.0166787,

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., & Mononen, J. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behavioural processes,* 148, 56-62. doi:https://doi.org/10.1016/j.beproc.2018.01.004

Si, T., Bagchi, J., & Miranda, P. B. (2022). Artificial neural network training using metaheuristics for medical data classification: an experimental study. *Expert Systems with Applications,* 193, 116423. doi:https://doi.org/10.1016/j.eswa.2021.116423

Sun, P., Wang, Z., Jia, L., & Xu, Z. (2024). SMOTE-kTLNN: A hybrid re-sampling method based on SMOTE and a two-layer nearest neighbor classifier. *Expert Systems with Applications,* 238, 121848. doi:https://doi.org/10.1016/j.eswa.2023.121848

Tuncer, S. A., Ayyıldız, H., Kalaycı, M., & Tuncer, T. (2021). Scat-NET: COVID-19 diagnosis with a CNN model using scattergram images. *Computers in Biology and Medicine,* 135, 104579. doi:https://doi.org/10.1016/j.compbiomed.2021.104579

Ullah, H., Bin Heyat, M. B., Akhtar, F., Sumbul, Muaad, A. Y., Islam, M. S., . . . Lin, Y. (2022). An End-to-End Cardiac Arrhythmia Recognition Method with an Effective DenseNet Model on Imbalanced Datasets Using ECG Signal. *Computational Intelligence and Neuroscience,* 2022(1), 9475162. doi:https://doi.org/10.1155/2022/9475162

Wu, Y.-c., & Feng, J.-w. (2018). Development and application of artificial neural network. *Wireless Personal Communications,* 102(2), 1645-1656. doi:https://doi.org/10.1007/s11277-017-5224-x

Zeraatkar, S., & Afsari, F. (2021). Interval–valued fuzzy and intuitionistic fuzzy–KNN for imbalanced data classification. *Expert Systems with Applications,* 184, 115510. doi:https://doi.org/10.1016/j.eswa.2021.115510

Zhang, P., Zhang, X., & Liu, A. (2022). Effects of Data Augmentation with the BNNSMOTE Algorithm in Seizure Detection Using 1D-MobileNet. *Journal of Healthcare Engineering,* 2022(1), 4114178. doi:https://doi.org/10.1155/2022/4114178

Zhang, S., Cheng, D., Deng, Z., Zong, M., & Deng, X. (2018). A novel kNN algorithm with data-driven k parameter computation. *Pattern Recognition Letters,* 109, 44-54. doi:https://doi.org/10.1016/j.patrec.2017.09.036