International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Impact of Feature Extraction Combined with Data Sampling Methods on Heartbeat Categorization

Sampath Kini K[1], Karthik Pai B H[2]
[1]Computer Science And Engineering
NMAM Institute Of Technology (Deemed To Be University), Karkala, India
sampath@nitte.edu.in
[2]Information Science And Engineering
NMAM Institute Of Technology (Deemed To Be University), Karkala, India
karthikpai@nitte.edu.in

Abstract— Dealing with class-imbalanced datasets in data analytics poses challenges, especially when faced with high-dimensional data. In order to handle this issue, researchers often utilize preprocessed methods like feature selection. Feature selection attempts to create a more informative and condensed feature set, while data sampling helps alleviate class imbalance. In our study, aim is to explore the effectiveness of data sampling preprocessed techniques combined with feature extraction using a dataset on ECG Heartbeat. We evaluate ensemble classifiers: Decision Tree; Random Forests (RF), Gradient-Boosted Trees (GBT) for feature extraction. In terms of data sampling, we assess the effectiveness of two methods: Random Under sampling (RUS) and Synthetic Minority Oversampling (SMOTE). The performance of this feature extraction is measured using the sensitivity and the specificity, two important metrics used for accuracy. Our findings depict that the combination of the RUS and GBT method yields the highest performance for ECG Heartbeat detection.

Keywords—ECG Heartbeat, Random Under sampling, SMOTE, Feature Extraction, Classification, Data Analytics

I. INTRODUCTION

Monitoring a patient's health status at the early stage can assist doctors to treat a patient's disease properly [1]. An electrocardiogram (ECG) serves as a comprehensive portrayal of the electrical actions by the heart, which is captured on the outer surface of the human's body. Its significance lies in its crucial role in the heart diseases diagnosis [2]. It serves as a reliable tool for monitoring the functionality of the cardiovascular system. The simple and non-invasive feature of ECG signals make them extensively employed in the detection of heart diseases. Millions of individuals are impacted by irregular heartbeats, which can be life-threatening in specific instances. Consequently, there is a significant need for precise and cost-effective methods to diagnose arrhythmic heartbeats [3]. Numerous studies have devised approaches for classifying arrhythmias by leveraging automatic analysis and diagnostic systems that make use of ECG signals. In analysis and diagnosis of cardiac diseases, the effectiveness of feature extraction and beat classification holds paramount importance. Over the past few years, there has been a proliferation of techniques proposed for the classification of ECG signals, leading to notable progress and positive outcomes [4].The effectiveness of ECG pattern classification is greatly influenced by both the feature extraction process and class imbalance. These will have a crucial role in determining the discriminative power of the extracted features from the ECG signal Class imbalance, which refers to an uneven distribution of classes in a dataset, can present difficulties for machine learning algorithms. When the majority class is overrepresented, it can lead to biased classification performance and make it more challenging to accurately identify the minority class. To address this concern, several approaches can be utilized, such as data-level techniques like RUS [5] and SMOTE [6]. The objective of these methods is to restore balance to the dataset by either decreasing instances of the major class or generating synthetic samples for the minor class. In addition, algorithm-level strategies often employ cost-sensitive approaches to deal with the influence of class imbalance [7]. However, for the purposes of this study, the main emphasis will be on data-level techniques specifically aimed at addressing class imbalance. Random Under sampling (RUS) is a technique that targets to remove the instances from the major class in the dataset until a predefined ratio between the major and minor class is attained. In contrast, SMOTE serves as a technique for increasing the presence of the minority class. It accomplishes this by generating artificial instances for the minority class, utilizing existing instances that are closely related to each other. The procedure of feature extraction begins with an initial dataset containing original attributes and strives to produce fresh attributes that convey valuable information while avoiding redundancy. This extraction procedure has the potential to enhance overall model performance, interpretation, and classification task outcomes. The motivation for our research stems from the increasing occurrence of heart attacks and the effective application of machine learning methods in detecting them. Our research specifically focuses on studying the use of RUS and SMOTE as data sampling techniques, which, have not been previously examined in the context of imbalanced

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

heartbeat data as per our knowledge. The ECG heart attack dataset utilized in our study is sourced from the PhysioBank ATM Dataset [8].Ensemble classifiers have demonstrated remarkable performance in numerous investigations [9]. Hence, we utilize two ensemble learners, namely Random Forest [10], GBT classifier, to evaluate the classification performance. The assessment is based on sensitivity and specificity scores.

The section labeled "Background and Related work" provides a comprehensive review of pertinent literature, while the "Methodology" section discusses the pre-processing of data and the tasks associated with classification. The "Results and Discussion" section presents and examines our findings, while the "Conclusion" section captures brief summary of the key findings in our experiments and proposes avenues for the future study.

II. BACKGROUND AND RELATED WORK

Numerous works and studies have been conducted on ECG classification, employing both traditional and advanced techniques. In a study conducted in and it was observed that Indonesia has a high mortality rate due to cardiovascular diseases [11]. To address this problem, a tele-ECG system was developed to enable early detection and monitoring of heart diseases, specifically addressing the challenges associated with processing large-scale data. This system utilized decision tree (DT) and random forest (RF) algorithms for classifying ECG data, marking the first implementation of big data analytics in heartbeats classification. The achieved accuracy rates were 97.14% using decision trees and 98.92% using random forest. In a different investigation mentioned in [12], neural networks and dimensionality reduction methods were utilized. This approach's effectiveness was assessed by applying it to the Massachusetts Institute of Technology arrhythmia database. The results demonstrated an accuracy of 96.97% when tested on a limited dataset comprising 18 ECG records, each with duration of 30 minutes.

In a research conducted by Alarsan and Younes J [13], they focused on extracting various types of heartbeats for classification purposes. Their approach involved employing distinct models for each heartbeat type (e.g., normal heartbeats, type 1 heartbeats, etc.) and using classification methods. They utilized neural networks and SVM, achieving high accuracy rates, exceeding 90%. This study highlighted the potential for developing a swift classifier for heartbeat categorization. In another investigation [14], a Deep Neural Network (DNN) was harnessed for deep learning in heartbeat classification. The author compared these results with multiple studies and achieved an impressive classification accuracy of 99%. However, this particular classification was limited to two types (Normal and Abnormal), and the dataset encompassed nearly 85,000 records. Additionally, another study explored multiple heartbeat types, achieving a 93.4% accuracy using a Convolutional Neural Network (CNN) for ECG beat type classification [15].It's essential to acknowledge that the highest accuracy scores in these previous studies are notably lower when compared to the superior results obtained in our research. Furthermore, none of the mentioned studies in this section integrated data sampling and feature extraction concurrently. In our study, we propose a novel method for classifying ECG signals by combining multiple widely-used classifiers with data sampling techniques. This method is thoroughly assessed using the MIT BIH Arrhythmia Database, and the classification performance is measured on a collection of 51 ECG records with varying durations. Noteworthy aspects of this study include: The tested records comprised 205,146 obtained from 51 patients. The inclusion of heartbeat types, encompassing normal and abnormal beats, in both the training and testing sets. The adoption of the data sampling and utilization of machine learning methods for the classification task.

III. METHODOLOGY

The following subsections will describe the experimental study conducted for the entirety of this work.

A. Heartbeat Data Set Preparation

To carry out this study, it was imperative to make use of a database comprising digital ECG records. This facilitated the computational examination of patients with various medical conditions. Consequently, our choice was to employ the well-known Massachusetts Institute of Technology (MIT) arrhythmia database, specifically the MIT-BIH Arrhythmia Dataset. To acquire the essential annotations for the records following the desired setup, we accessed the physio-net ATM Bank. Every record was extracted in its entirety, reaching its full duration (e.g., record 100 covers 1805 seconds). The finalized dataset encompasses data from 51 individuals, leading to a cumulative count of 205,146 rows encompassing all patients. Every row is composed of 16 distinct feature columns. Each record's data is organized into three primary files, recognizable by their unique extensions: .atr, .hea, and .dat. Within these files, the .atr file holds the annotations, the .dat file contains the digitized signals, and the .hea file functions as the header file. The recordings underwent digitization at a rate of 360 samples per second per channel, featuring an 11-bit resolution and a 10 mV range. To guarantee accuracy, every record underwent separate annotation by two or more cardiologists. Any disparities were addressed and resolved to attain computer-readable reference annotations for every heartbeat contained in the database. The annotations were acquired from the Physionet website and stored as text files. Many of the features examined in the research were generated through the Discrete Wavelet Transform (DWT) of the continuous ECG signal. Finally, the resultant dataset was preserved in CSV format. Alongside the ECG signal, the annotations also include information about beat localization and beat class. For classification purposes, two primary classes were identified: Normal and Abnormal. Within this research, we assess the effectiveness of classifiers in diverse scenarios, examining the impact of two data sampling methods: RUS (Random Under Sampling) and SMOTE (Synthetic Minority Over Sampling Technique).For the RUS technique, we employ five different ratios (1:1, 1:5, 1:10, 1:20, and 1:50) to

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

obtain datasets with down-sampled minority instances. These ratios indicate the proportion of minority to majority instances after the down sampling process. By considering this range of ratios, we aim to carry out the validation of our study. To implement the RUS and SMOTE algorithms, we utilize the imbalance edlearn Python library as depicted in "Fig. 1" and "Fig. 2" respectively. Additionally, we incorporate the PCA (Principal Component Analysis) algorithm with Scikit-Learn , implemented. The learners utilized in this investigation encompass Random Forests as well as Gradient-Boosted Trees (GBT), which constitutes an ensemble of Decision Trees techniques.

```python
import pandas as pd
from imblearn.over_sampling import SMOTE

# Load the dataset
data = pd.read_csv('heart_beat.csv')

# Separate the features and the target variable
X = data.drop('target', axis=1)
y = data['target']

# Apply SMOTE to balance the classes
smote = SMOTE()
X_resampled, y_resampled = smote.fit_resample(X, y)

# Print the class distribution before and after applying SMOTE
print("Before SMOTE:")
print(y.value_counts())

print("After SMOTE:")
print(pd.Series(y_resampled).value_counts())
```

Fig. 1. Python code snippet of SMOTE sampling.

```python
import pandas as pd
from imblearn.under_sampling import RandomUnderSampler

# Load the dataset
data = pd.read_csv('heart_beat.csv')

# Separate the features and the target variable
X = data.drop('target', axis=1)
y = data['target']

# Apply RUS to balance the classes
rus = RandomUnderSampler()
X_resampled, y_resampled = rus.fit_resample(X, y)

# Print the class distribution before and after applying RUS
print("Before RUS:")
print(y.value_counts())

print("After RUS:")
print(pd.Series(y_resampled).value_counts())
```

Fig. 2. Python code snippet of RUS sampling.

B. Machine learning-based classification of heartbeats

In this study, we opted to utilize the MIT-BIH Arrhythmia and MIT-BIH Supraventricular Arrhythmia datasets as our selected databases. "Fig. 3" illustrates the different stages of dataset processing aimed at developing a heartbeat classification model. To ensure compatibility with software programs, the acquired features were saved in a csv file along with their respective column data types. The file contains values with a combination of integer and double types, which is necessary information for proper csv file interpretation. The process of obtaining the final model consisted of multiple stages, as illustrated in Fig. 3.We utilized frame sampling techniques to gather the necessary data. To facilitate the data processing using Python ML libraries, we established a schema for the columns in the CSV files. The steps for producing the final models involved several stages, as depicted in "Fig. 4".The steps for producing the final accuracy involved several stages, as depicted in "Fig. 5".

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

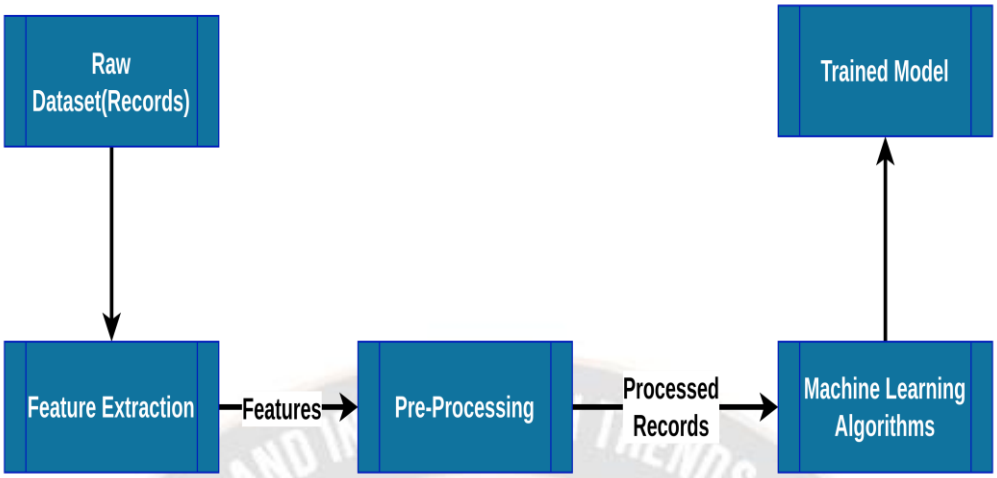

Fig. 3. Different stages of dataset processing.

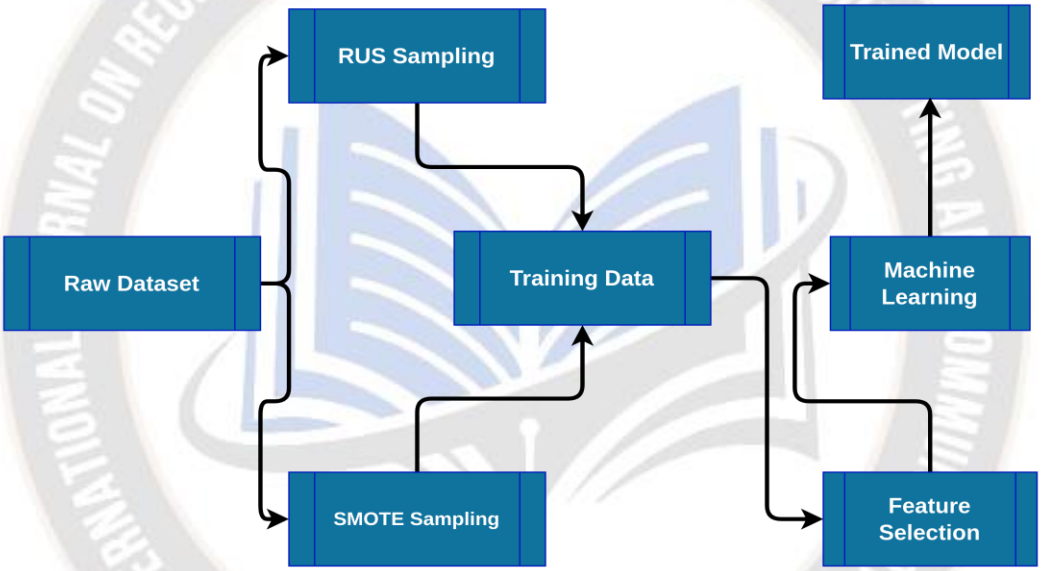

Fig. 4. Process of obtaining final models.

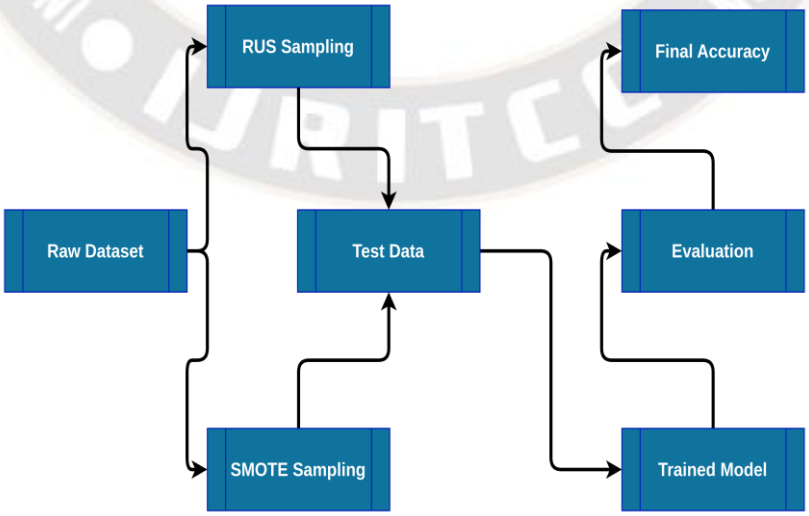

Fig. 5. Process of obtaining final accuracy.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Table I. provides metadata of the column types in the dataset. Excluding the last column, which represents the label, all other columns contain double values. The label column, on the other hand, is of integer type. Data mining Techniques used to diagnose heart related illness has become one of the interesting fields for the researchers in recent years[16].In machine learning, irrespective of the particular algorithm selected, there are certain components that are commonly found in most models [17].

TABLE I. METADATA OF THE COLUMN TYPES IN THE DATASET

Field Number	Type	Field Number	Type
Field 1	Float64	Field 10	Float64
Field 2	Float64	Field 11	Float64
Field 3	Float64	Field 12	Float64
Field 4	Float64	Field 13	Float64
Field 5	Float64	Field 14	Float64
Field 6	Float64	Field 15	Float64
Field 7	Float64	Field 16	Float64
Field 8	Float64	Field 17	Int64
Field 9	Float64		

C. Data Modeling Process

The dataset is divided into two subsets, one intended for constructing the training model and the other for assessing its performance. The authors employ techniques such as Random under Sampling and SMOTE to ensure balanced representation. In this study, null values are addressed through the data sampling process. The column representing the heartbeat type is labeled using two different classes,

namely "Normal" and "Abnormal." Relevant columns from the dataset are selected as features for the model. The authors make use of either the Gradient-Boosted Trees algorithm (GBT), incorporating parameters like MaxDepth and MaxIter, or the Random Forest algorithm (RF), involving parameters such as MaxDepth and NumTrees. After selecting the algorithm, the model is trained with the training dataset, resulting in a prepared model for testing.

To study the performance of each of the trained models, evaluators are utilized to calculate accuracy. In this particular study, we have constructed and evaluated both the Gradient-Boosted Trees model (GBT) and the Random Forest model (RF).

D. Gradient-Boosted Trees Classification Model

The authors employ techniques such as Random under Sampling and SMOTE to ensure balanced representation. Gradient-Boosted Trees (GBT) is a machine learning technique primarily employed for addressing classification and regression tasks. It builds a predictive model by amalgamating numerous weaker prediction models. The concept of gradient boosting has its roots in the understanding that boosting can be seen as an optimization procedure associated with a particular cost function [18]. Fig 5 illustrates the GBT model in question. The process of training this GBT model entailed exploring various combinations of Max Iteration and Max Depth values, as detailed in Table II.

TABLE II. MAX ITERATIONS AND MAX DEPTH VALUES OF GBT MODEL

Depth: Max	Iterations: Max	GBT Model
18	8	M1
8	8	M2
6	8	M3
6	6	M4
6	16	M5
8	10	M6
8	10	M7

E. Random Forest Model (RF)

Random Forests, also known as random decision forests, are a widely employed ensemble learning method employed in tasks like classification, regression, and similar endeavors. This strategy entails constructing numerous decision trees during the training stage and then making predictions by

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

considering the mode for classification tasks or the average prediction for regression tasks among these individual trees [19]. The formation of the RF decision tree model relies significantly on key parameters [20], which are vital for its performance. Throughout the training phase, we established this model using different Max Depth and Number of Trees values. We fine-tuned these parameters, and the resulting combinations are detailed in Table III.

TABLE III. PARAMETER VALUES OF MAX DEPTH AND NO. OF TREES OF RF MODEL

Depth: Max	#Trees	RF Model
16	16	M1
6	16	M2
8	16	M3
16	8	M4
6	8	M5
8	8	M6
18	18	M7

IV. RESULTS OF RESEARCH

To validate this study, the model's accuracy was evaluated using both GBT and RF algorithms with the BinaryClassificationEvaluator. Furthermore, We have evaluated the impact of RUS and SMOTE data sampling techniques on classifier accuracy.

The definitions of SE (Sensitivity) and SP (Specificity) are as follows: Sensitivity (SE) represents the ability of a clinical test to correctly identify individuals with the disease, while Specificity (SP) refers to the ability of a clinical test to accurately recognize individuals without the disease. CC (Classification Correctness) denotes the percentage of instances correctly classified.

To calculate SE and SP, the following definitions are used: TP (True positive) represents cases where the patient has the disease and the test result is positive. FP (False positive) corresponds to situations where the patient doesn't have the disease, but the test result is positive. TN (True negative) pertains to instances where the patient is disease-free, and the test result is negative. FN (False negative) signifies cases where the patient has the disease, but the test result is negative.

The formulae for calculating SE, SP, and CC are captured in Table IV. The result summary of the GBT and RF algorithms combined with RUS and SMOTE sampling are captured in Tables V, VI, VII and VIII.

TABLE IV. EQUATIONS FOR SE, SP, AND CC

Target	Formula
SE	TP / (TP + FN)
SP	TN / (TN + FP)
CC	(TP + TN) / (TP + TN + FP + FN)

TABLE V. GBT RESULTS WITH RUS SAMPLING

#Model	Accuracy %	SP%	SE%	CC%
GBT M1	98.2	97	98	98
GBT M2	94.6	94	92	95
GBT M3	86.21	87	87	86
GBT M4	83.65	83	84	84
GBT M5	87.10	88	87	87
GBT M6	95.32	95	95	95
GBT M7	96.76	97	96	97

TABLE VI. GBT RESULTS WITH SMOTE SAMPLING

#Model	Accuracy%	SP%	SE%	CC%
M1	96.22	96	95	96
M2	92.62	92	90	92
M3	85.21	85	83	85
M4	83.25	82	84	83
M5	86.10	87	85	86
M6	91.32	91	91	91
M7	92.76	92	90	92

TABLE VII. RF RESULTS WITH RUS SAMPLING

Model	Accuracy	SP%	SE%	CC%
RF M1	97.2	97	96	97
RF M2	92.3	92	90	92
RF M3	82.21	83	81	82

Model	Accuracy	SP%	SE%	CC%
RF M4	82.27	81	78	82
RF M5	82.10	80	82	82
RF M6	90.32	90	89	90
RF M7	91.76	91	91	92

TABLE VIII. RF RESULTS WITH SMOTE SAMPLING

Model	Accuracy	SP%	SE%	CC%
RF M1	91.2	91	92	90

Model	Accuracy	SP%	SE%	CC%
RF M2	89.6	90	87	89
RF M3	87.57	85	84	88
RF M4	85.35	83	84	85
RF M5	83.17	82	85	83
RF M6	89.32	91	88	89
RF M7	89.77	91	90	90

V. CONCLUSION

In this research, we employed a cardiac dataset to investigate how data sampling and feature extraction influence ensemble classifiers. More precisely, we assessed the effects of two data sampling methods, RUS and SMOTE, as well as two feature extraction techniques, RF and GBT, on the classifiers' performance. The results demonstrated that combining the RUS data sampling method with the GBT feature extraction technique produced the most favorable results. In conclusion; this research has effectively validated the efficiency of machine learning classification algorithms in classifying heartbeats. It has successfully incorporated data sampling techniques like RUS and SMOTE to address class imbalance and improve accuracy, which are crucial aspects of classification. The developed GBT and RF models demonstrate the capability to accurately classify different types of ECG heartbeats, making them suitable for integration into ECG systems to enhance diagnostic efficiency and reliability. The proposed model shows promise as a valuable tool for aiding cardiologists in interpreting ECG heartbeat signals and gaining deeper insights from them. Future work could focus on real-time implementation and continuous training to further enhance accuracy.

REFERENCES

[1] Tiwari K, Ansari S, Kushwaha G. Earlier Heart Disease Prediction System Using Machine Learning. InComputer Vision and Robotics: Proceedings of CVR 2022 2023 Apr 28 (pp. 191-204). Singapore: Springer Nature Singapore.

[2] Li H, Yuan D, Ma X, Cui D, Cao L.,"Genetic algorithm for the optimization of features and neural networks in ECG signals classification", Sci Rep;7:41011,(2017), DOI: https://doi.org/10.1038/srep41011

[3] Kachuee M, Fazeli S, Sarrafzadeh M.,"ECG heartbeat classification: a deep transferable representation", In: 2018 IEEE international conference on healthcare informatics (ICHI),(2018), New York: IEEE; p. 443–4. DOI: https://doi.org/10.1109/ICHI.2018.00092

[4] Sannino G, De Pietro G,"A deep learning approach for ECG-based heart rate classification for arrhythmia detection", Future Gener Comput Syst,(2018),86:446–55. DOI: https://doi.org/10.1016/j.future.2018.03.057

[5] Liu B, Tsoumakas G," Dealing with class imbalance in classifier chains via random undersampling", Knowl-Based Syst,(2020), 192:105292. https://doi.org/10.1016/j.knosys.2019.105292

[6] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP., "Smote: synthetic minority over-sampling technique." J Artif Intell ,Res, (2002),16:321–57. https://doi.org/10.1613/jair.953

[7] Thai-Nghe N, Gantner Z, Schmidt-Thieme L,"Cost-sensitive learning methods for imbalanced data", In: The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE. pp. 1–8,(2010), DOI: 10.1109/IJCNN. 5596486

[8] https://archive.physionet.org/cgi-bin/atm/ATM, (May 2023)

[9] Alarsan, F.I., Younes, M. "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms",J Big Data (2019) 6,81,. https://doi.org/10.1186/ s40537-019-0244-x

[10] Patel HH, Prajapati P,"Study and analysis of decision tree based classification algorithms", Int J Computer Sci Eng, (2018),6(10):74–8.

[11] Ma'sum M.A, Jatmiko W, Suhartanto H.,"Enhanced tele ECG system using hadoop framework to deal with big data processing", In: 2016 international workshop on Big Data and information security (IWBIS),(2026),New York: IEEE; pp. 121–6. DOI: 10.1109/IWBIS. 2016.7872900

[12] Dalvi RdF, Zago GT, Andreão RV.,"Heartbeat classification system based on neural networks and dimensionality reduction",(2016) Res Biomed Eng.;32(4):318–26.

[13] Alarsan, F.I., Younes, M. "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms",J Big Data (2019) 6,81,. https://doi.org/10.1186/ s40537-019-0244-x

[14] Celesti F, Celesti A, Carnevale L, Galletta A, Campo S, Romano A, Bramanti P, Villari M.,"Big data analytics in genomics: the point on deep learning solutions", In: 2017 IEEE symposium on computers and communications (ISCC). New York: IEEE; pp. 306–9. DOI: 10.1109/ISCC.2017.8024547

[15] Kachuee M, Fazeli S, Sarrafzadeh M.,"ECG heartbeat classification: a deep transferable representation", In: 2018 IEEE international conference on healthcare informatics (ICHI),(2018), New York: IEEE; p. 443–4. DOI: https://doi.org/10.1109/ICHI.2018.00092

[16] Safa Mejhoudi, Rachid Latif, Amine Saddik, Wissam Jenkal and Abdelhafid El Ouardi, "Speeding up an

Adaptive Filter based ECG Signal Pre-processing on Embedded Architectures" International Journal of Advanced Computer Science and Applications(IJACSA), 12(5), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120544

[17] Alzubi J, Nayyar A, Kumar, "A..Machine learning from theory to algorithms: an overview", In: Journal of physics: conference series, vol. 1142, (2018) Bristol: IOP Publishing; P. 012012. DOI:10.1088/17426596/1142/1/012012

[18] Hatwell, Julian, Mohamed Medhat Gaber, and R. Muhammad Atif Azad, "gbt-HIPS: Explaining the Classifications of Gradient Boosted Tree Ensembles",Applied Sciences 11, (2021),no. 6: 2511. https://doi.org/10.3390/app11062511

[19] Woś M, Drop B, Kiczek B. Predicting and Detecting Coronary Heart Disease in Patients Using Machine Learning Method. In International Work-Conference on Bioinformatics and Biomedical Engineering 2023 Jun 29 (pp. 367-377). Cham: Springer Nature Switzerland.

[20] Barandiaran I,"The random subspace method for constructing decision forests", IEEE Trans Pattern Anal Mach Intell. ;20(8):122,(1998), DOI: 10.1109/34.709601