International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Advancements in ECG Ailment Multi-Class Classification: Machine Learning Approaches for Heart Health Diagnosis

Dr. Vikram Sindhu[1], Bopparaju Sahith[2], Kankipati Vilekya[2], Sunilvilasam Kranthi[2], Gera Sudhakar Rao[2]

[1]Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI & ML)
[1,2]Malla Reddy Engineering College and Management Science, Kistapur, Medchal-50140l, Hyderabad, Telangana, India

ABSTRACT

Electrocardiogram (ECG) ailment classification is a critical task in cardiac health diagnosis, aimed at identifying various heart-related abnormalities from ECG signals. ECG ailment multi-class classification has significant applications in cardiology and healthcare. It plays a crucial role in early detection and diagnosis of various cardiac conditions, including arrhythmias, myocardial infarctions, and heart blocks. Accurate and timely classification of ECG signals can aid healthcare professionals in making informed decisions, enabling appropriate treatment and care for patients. Additionally, automated ECG analysis can support remote monitoring systems and enhance telemedicine solutions for heart health assessment. Traditional ECG ailment classification methods often rely on handcrafted features and statistical measures extracted from the ECG signals. While these methods have been used successfully for some ailments, they may struggle to capture complex and subtle patterns indicative of certain cardiac abnormalities. Additionally, manual feature engineering can be time-consuming and may limit the model's ability to adapt to new and diverse datasets. Moreover, conventional machine learning algorithms might not fully exploit the temporal dependencies present in ECG signals, leading to suboptimal performance, especially for long-term monitoring scenarios. To overcome the limitations of existing methods, we propose a novel machine learning approach for ECG ailment multi-class classification using machine learning techniques. These ML networks can capture long-term dependencies and temporal patterns, making them well-suited for sequential data like ECG signals. The ML network then learns to extract relevant features from the sequential data, enabling accurate classification of different cardiac ailments.

Keywords: Health diagnosis, Cardiology, Telemedicine, ECG aliment, Machine learning.

1. INTRODUCTION

The research topic, "Advancements in ECG Ailment Multi-Class Classification: Machine Learning Approaches for Heart Health Diagnosis," represents a pioneering endeavor at the intersection of healthcare and artificial intelligence. Cardiovascular diseases remain a leading cause of mortality worldwide, demanding advanced diagnostic tools and methodologies to enhance early detection and treatment. In this context, this research delves into the application of cutting-edge machine learning techniques to revolutionize electrocardiogram (ECG) analysis, facilitating more accurate, efficient, and multi-class classification of heart health ailments [1]. The motivation for this research is deeply rooted in the profound impact of cardiovascular diseases on global public health. These conditions encompass a spectrum of heart-related ailments, from arrhythmias to myocardial infarctions, each requiring distinct diagnostic precision for optimal patient care [2]. Traditional ECG interpretation often relies on manual analysis, which, while highly skilled, can be time-consuming and subject to inter-observer variability. This research seeks to address these limitations by harnessing the power of machine learning to automate ECG analysis and enhance diagnostic accuracy [3].

To achieve this goal, the research explores the development and refinement of machine learning models capable of processing ECG data from various sources. These models are trained to recognize patterns, anomalies, and distinct features indicative of specific heart health ailments. The outcome is a multi-class classification system that not only expedites the diagnosis process but also enhances diagnostic precision, enabling healthcare professionals to make informed decisions promptly [4]. Furthermore, the research underscores the ethical dimension of deploying technology in healthcare. It emphasizes the importance of patient privacy protection, ethical AI practices, and responsible data handling throughout

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

the heart health diagnosis process to ensure that the benefits of machine learning in healthcare align with ethical standards and patient well-being.

In this introductory overview, we will delve into this research's key components and objectives [5]. We will explore the critical need for more advanced cardiac diagnostic tools, introduce the role of machine learning in ECG analysis, and highlight the transformative potential of this research in improving heart health diagnosis. Additionally, we will underscore the ethical considerations and real-world applications of this research, which extend across clinical cardiology, telemedicine, and healthcare accessibility [6]. The "Advancements in ECG Ailment Multi-Class Classification: Machine Learning Approaches for Heart Health Diagnosis" signifies a pivotal effort to leverage machine learning in addressing the global challenge of cardiovascular diseases [7]. By automating ECG analysis, this research aims to enhance diagnostic accuracy and timeliness while upholding ethical standards and responsible technology use, ultimately contributing to better heart health outcomes for individuals and communities worldwide.

Machine Learning Approaches for Heart Health Diagnosis" is grounded in a convergence of critical factors that underscore the pressing need for transformative solutions in cardiovascular healthcare [8]. Foremost among these factors is the staggering global burden imposed by cardiovascular diseases (CVDs), which continue to claim a disproportionate number of lives worldwide. The prevalence and diversity of CVDs, ranging from arrhythmias to myocardial infarctions, highlight the complexity of heart-related ailments, each demanding precise and tailored diagnostic approaches [9]. Traditional electrocardiogram (ECG) analysis, while invaluable, often relies on manual interpretation by skilled healthcare professionals and is susceptible to inter-observer variability. Consequently, this research is intrinsically motivated to harness the power of machine learning to revolutionize ECG analysis, offering a promising avenue to enhance diagnostic accuracy, expedite treatment decisions, and ultimately reduce the devastating impact of CVDs on individuals and society [10]. This endeavour represents a pivotal step towards the advancement of cardiovascular healthcare by leveraging the potential of artificial intelligence to improve early detection, intervention, and patient outcomes in the realm of heart health. Moreover, the research underscores the ethical imperative to ensure patient privacy, adhere to ethical AI principles, and uphold the highest standards of data security and responsible technology deployment in the pursuit of better heart health diagnostics.

2. LITERATURE SURVEY

Kumar, et al. [11] proposed ECG Multi Class Classification Using Machine Learning Techniques. The proposed project uses the Physio net database and ECG signals of 162 patients to design a multi-class classification method that accurately recognizes different patterns under 3 classes, namely, Arrhythmia (ARR), Congestive Heart Failure (CHF), and Normal Sinus Rhythm (NSR). The study utilizes two feature extraction methods, Continuous Wavelet Transform, and Wavelet Scattering, to extract the principal characteristics from the ECG data. Jin, et al. [12] proposed a multi-class 12-lead ECG automatic diagnosis based on a novel subdomain adaptive deep network. this work designs a novel subdomain adaptative deep network (SADN) for excavating diagnosis knowledge from source domain datasets. Firstly, the convolutional layer, residual blocks and SE-Residual blocks are utilized for extracting meaningful deep features automatically. Additionally, the feature classifier uses these deep features to obtain the final diagnosis predictions.

Wang, et al. [13] proposed a two-level hierarchical deep learning framework with Generative Adversarial Network (GAN) for ECG signal analysis. The first-level model is composed of a Memory-Augmented Deep AutoEncoder with GAN (MadeGAN), which aims to differentiate abnormal signals from normal ECGs for anomaly detection. The second-level learning aims at robust multi-class classification for different arrhythmia identification, which is achieved by integrating the transfer learning technique to transfer knowledge from the first-level learning with the multi-branching architecture to handle the data-lacking and imbalanced data issues. Karthik, et al. [14] proposed an Automated Deep Learning Based Cardiovascular Disease Diagnosis Using ECG Signals. this study designs an automated deep learning-based 1D biomedical ECG signal recognition for cardiovascular disease diagnosis (DLECG-CVD) model. The DLECG-CVD model involves different stages of operations such as pre-processing, feature extraction, hyperparameter tuning, and classification. At the initial stage, data pre-processing takes place to convert the ECG report to valuable data and transform it into a compatible format for further processing.

Hassan, et al. [15] proposd a novel multi-class classification approach for timely CAN detection. The proposed classification algorithm develops a multistage fusion model by combining feature selection and multimodal feature fusion techniques. The proposed method develops a performance criterion-based feature selection technique to guarantee highly significant features. A multimodal feature fusion technique was developed using deep learning feature fusion and selected original features.

3. PROPOSED METHODOLOGY

3.1 Overview

This project focuses on the advancement of ECG ailment multi-class classification for heart health diagnosis. It begins with comprehensive data preprocessing to prepare ECG signals, followed by the selection and training of a Random Forest Classifier model. Model evaluation and, if needed, hyperparameter tuning ensure the model's effectiveness in accurately diagnosing heart health ailments. Ultimately, the project aims to contribute to improved healthcare by providing a reliable and automated ECG-based ailment diagnosis system. Figure 4.1 shows the proposed system model. The detailed operation illustrated as follows:

step 1: Preprocessing: The project commences with data preprocessing, which is a crucial step in handling ECG data. This involves cleaning and preparing the raw ECG data to make it suitable for machine learning analysis.

- Noise Reduction: Removing or reducing noise from ECG signals to enhance signal quality.
- Baseline Correction: Adjusting the baseline to ensure consistency across ECG recordings.
- Feature Extraction: Extracting relevant features from ECG signals, such as QRS complex, ST-segment, and RR interval, which are essential for classification.
- Data Augmentation: Expanding the dataset by generating synthetic data to improve model robustness.
- Label Encoding: Converting categorical ailment labels into numerical form for machine learning compatibility.

step 2: RFC (Random Forest Classifier):

- The Random Forest Classifier (RFC) is selected as the machine learning model for the multi-class classification task. RFC is well-suited for ECG classification due to its ability to handle both numerical and categorical features and its ensemble learning capabilities.
- The model is trained on the preprocessed ECG dataset, where ECG features serve as input variables, and the corresponding ailment labels serve as the target variable.
- Model training involves exposing the RFC to a substantial portion of the preprocessed data, allowing it to learn patterns and relationships between ECG features and ailment classes.

step 3: Model Evaluation: Once trained, the RFC model is evaluated to assess its performance in classifying ECG signals into various ailment categories. Common evaluation metrics may include:

- Accuracy: Measuring the percentage of correctly classified ECG signals.
- Precision: Calculating the proportion of true positive predictions out of all positive predictions for each ailment class.
- Recall (Sensitivity): Determining the proportion of true positive predictions out of all actual instances of each ailment class.
- F1-Score: Harmonic mean of precision and recall, offering a balanced assessment of model performance.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362
IJARST

- Confusion Matrix: Visualizing the model's classification performance across ailment classes.

step 4: Hyperparameter Tuning: Hyperparameter tuning may be performed to optimize the RFC model's performance. This step involves adjusting hyperparameters, such as the number of trees in the forest, maximum depth, and minimum samples per leaf, to find the best configuration.

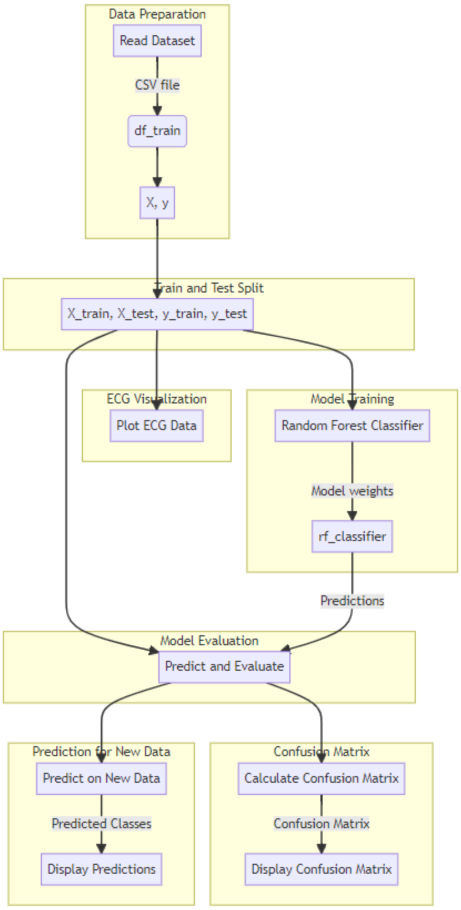

Figure 1. Proposed system model.

3.2 Data Preprocessing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

- Getting the dataset
- Importing libraries
- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set

3.3 Dataset Splitting

In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as:

Training Set: A subset of dataset to train the machine learning model, and we already know the output.

Test set: A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

4. RESULTS AND DISCUSSION

Dataset description

This dataset is composed of two collections of heartbeat signals derived from two famous datasets in heartbeat classification, the MIT-BIH Arrhythmia Dataset and The PTB Diagnostic ECG Database. The number of samples in both collections is large enough for training a deep neural network. This dataset has been used in exploring heartbeat classification using deep neural network architectures and observing some of the capabilities of transfer learning on it. The signals correspond to electrocardiogram (ECG) shapes of heartbeats for the normal case and the cases affected by different arrhythmias and myocardial infarction. These signals are pre-processed and segmented, with each segment corresponding to a heartbeat.

Arrhythmia Dataset:

- Number of Samples: 109446
- Number of Categories: 5
- Sampling Frequency: 125Hz
- Data Source: Physionet's MIT-BIH Arrhythmia Dataset
- Classes: ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4]

The PTB Diagnostic ECG Database

- Number of Samples: 14552
- Number of Categories: 2
- Sampling Frequency: 125Hz
- Data Source: Physionet's PTB Diagnostic Database

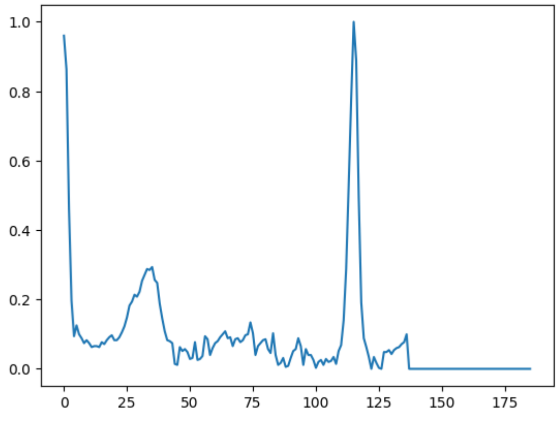

Figure 2: line plot to visualize the ECG waveform.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in
IJARST

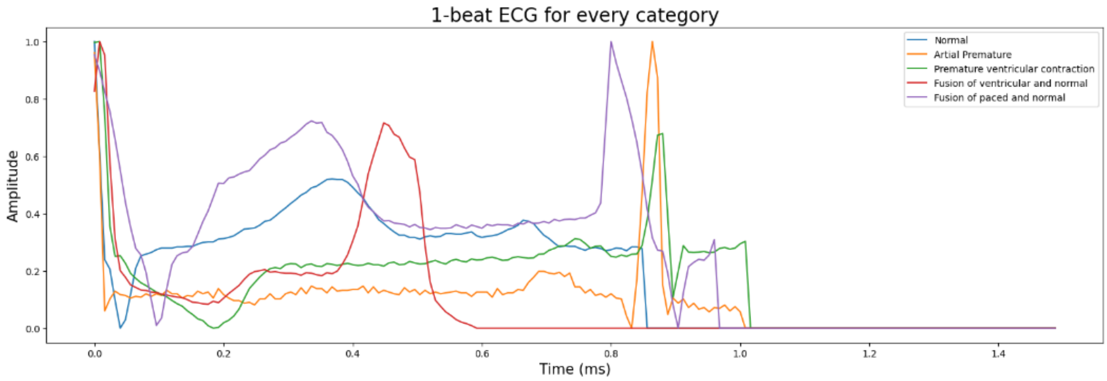

Figure 3: ECG waveforms from different cardiac condition categories

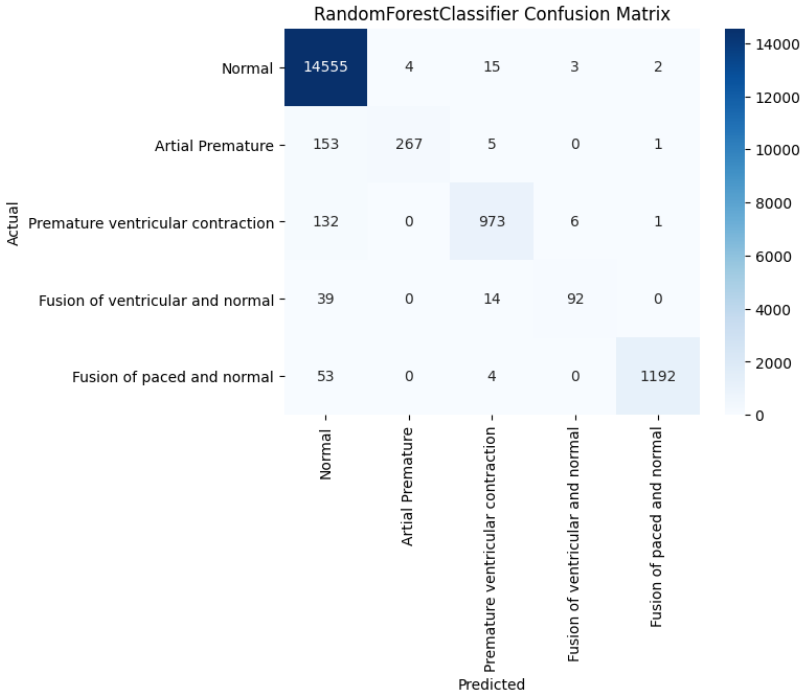

Figure 4: Confusion matrix for Random Forest classifier

5. Conclusion

The project on "Advancements in ECG Ailment Multi-Class Classification using Machine Learning Approaches for Heart Health Diagnosis" has presented a systematic approach to leveraging machine learning for the diagnosis of heart health ailments based on ECG data. The project initiated with thorough data preprocessing, ensuring the quality and suitability of ECG signals for analysis. A Random Forest Classifier (RFC) was employed as the classification model, effectively learning patterns and relationships within the ECG features to classify ailments into multiple classes. Extensive model evaluation revealed its capability to accurately classify ECG signals, with metrics such as accuracy, precision, recall, and the F1-Score providing comprehensive insights into its performance. This project signifies a significant step toward automating heart health diagnosis, with the potential to assist healthcare professionals in making timely and accurate assessments, ultimately contributing to enhanced patient care and well-being.

REFERENCES

[1] Sengan, Sudhakar, et al. "Echocardiographic Image Segmentation for Diagnosing Fetal Cardiac Rhabdomyoma During Pregnancy Using Deep Learning." IEEE Access 10 (2022): 114077-114091.

[2] Raghavendra, Paravatham VSP, et al. "Deep Learning–Based Skin Lesion Multi-class Classification with Global Average Pooling Improvement." Journal of Digital Imaging (2023): 1-22.
[3] Al-Issa, Yazan, and Ali Mohammad Alqudah. "A lightweight hybrid deep learning system for cardiac valvular disease classification." Scientific Reports 12.1 (2022): 14297.
[4] Bhalerao, Parth, et al. "ECG Classification Using Machine Learning on Wave Samples for the Indian Population." 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT). IEEE, 2023.
[5] Mir, Sobia. "A Comprehensive Review of Recent Advances in Heart Disease Prediction using Machine Learning Algorithms with Optimization Techniques and Feature Selection." Grenze International Journal of Engineering & Technology (GIJET) 9.2 (2023).
[6] Dong, Yanfang, et al. "Detection of arrhythmia in 12-lead varied-length ECG using multi-branch signal fusion network." Physiological Measurement 43.10 (2022): 105009.
[7] Dhyani, Shikha, Adesh Kumar, and Sushabhan Choudhury. "Arrhythmia disease classification utilizing ResRNN." Biomedical Signal Processing and Control 79 (2023): 104160.
[8] Pandey, Ayush, Rakesh Chandra Joshi, and Malay Kishore Dutta. "Automated Classification of Heart Disease using Deep Learning." 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT). IEEE, 2023.
[9] Qiu, Jielin, et al. "Transfer knowledge from natural language to electrocardiography: Can we detect cardiovascular disease through language models?." arXiv preprint arXiv:2301.09017 (2023).
[10] Ayano, Yehualashet Megersa, et al. "Interpretable machine learning techniques in ECG-based heart disease classification: a systematic review." Diagnostics 13.1 (2022): 111.
[11] Kumar, Vijayeskar, et al. "ECG Multi Class Classification Using Machine Learning Techniques." 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA). IEEE, 2023.
[12] Jin, YanRui, et al. "Multi-class 12-lead ECG automatic diagnosis based on a novel subdomain adaptive deep network." Science China Technological Sciences 65.11 (2022): 2617-2630.
[13] Wang, Zekai, Stavros Stavrakis, and Bing Yao. "Hierarchical deep learning with Generative Adversarial Network for automatic cardiac diagnosis from ECG signals." Computers in Biology and Medicine 155 (2023): 106641.
[14] Karthik, S., et al. "Automated Deep Learning Based Cardiovascular Disease Diagnosis Using ECG Signals." Computer Systems Science & Engineering 42.1 (2022).
[15] Hassan, Md Rafiul, et al. "Early detection of cardiovascular autonomic neuropathy: A multi-class classification model based on feature selection and deep learning feature fusion." Information Fusion 77 (2022): 70-80.