IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Hybrid Model to Predict Arrhythmia in Cancer Patients using Machine Learning

Megha.S
School of Computer Science and Engineering
REVA University
Bengaluru, India

Panchadara Sasidhar
School of Computer Science and Engineering
REVA University
Bengaluru, India.co

Polineni Manoj Kumar
School of Computer Science and Engineering *REVA University* Palakolanu Jaya Avinash Reddy
School of Computer Science and Engineering
REVA University
Bengaluru, India

Konduru Nikhith Raju
School of Computer Science and Engineering
REVA University
Bengaluru, India

Abstract— **Arrhythmia is irregularity in heart beat may be harmless or life threatening. Heart diseases are the important health problem and main cause of the death of the patient. Heart disease early detection and medical therapy can stop patients from passing away suddenly.**

Cancer patients often experience arrhythmia, a condition characterized by an irregular heartbeat. Arrhythmia can lead to serious health complications and even death if not managed properly. Early detection and prediction of arrhythmia can aid in timely medical intervention and improve patient outcomes. In this study, a hybrid model is proposed to predict arrhythmia in cancer patients using clinical and electrocardiogram (ECG) data. The results suggest that the hybrid model has the potential to be an effective tool for early detection and prediction of arrhythmia in cancer patients, thus enabling timely medical intervention and improved patient outcomes.

Keywords— Arrhythmia, Detection, hybrid Model.

I.INTRODUCTION

Arrhythmia, a common cardiovascular complication in cancer patients, can lead to serious health consequences if not detected and managed early. In recent years, there has been growing interest in developing machine learning models to predict arrhythmia in cancer patients. One promising approach is the use of hybrid models that combine multiple machine learning techniques, such as deep learning and feature engineering, to improve the accuracy of arrhythmia prediction.

The hybrid model leverages the strengths of different techniques to provide a more comprehensive understanding of the complex relationship between arrhythmia and cancer. Deep learning techniques can identify patterns in large, complex datasets, while feature engineering can extract meaningful features that are relevant to arrhythmia prediction. By combining these techniques, the hybrid model can generate more accurate predictions, leading to better clinical outcomes for cancer patients.

In this context, the purpose of this project is to develop a hybrid model to predict arrhythmia in cancer patients, using a variety of clinical and genetic data. The model will be trained and validated using a large, diverse dataset of cancer patients, and the performance of the model will be assessed using various evaluation metrics. The ultimate goal is to create a tool that can be used by clinicians to identify cancer patients who are at high risk of developing arrhythmia, enabling earlier intervention and better management of this serious complication.

II. DESIGN FLOW

The design flow of this model will be done with respect to the following objectives:

1.Data Collection and Preprocessing: Collect relevant data for arrhythmia prediction in cancer patients, including clinical data such as record, type, treatment history, and genetic data such as gene expression, genetic mutations, and copy number variations. Preprocess the data to clean and prepare it for analysis, including missing value imputation, normalization, and feature selection.

2. Feature Engineering: Extract relevant features from the preprocessed data using domain knowledge and machine learning techniques, such as principal component analysis (PCA), Linear discriminant analysis (LDA), or clustering. Generate a set of features that can be used as inputs for the model.

3. Model Architecture Selection: Choose a suitable Machine learning architecture for arrhythmia prediction, Determine the number of layers, neurons, and activation functions based on the size and complexity of the dataset.

4. Training and Validation: Split the preprocessed data into training, validation, and testing sets. Train the model using the training set, and validate it using the validation set. Monitor the performance of the model and adjust hyperparameters as necessary to improve accuracy and reduce overfitting.

5. Evaluation and Testing: Evaluate the performance of the model using appropriate metrics such as accuracy. Test the model using the testing set to ensure it is robust and generalizes well to new data.

6. Deployment and Integration: Once the model has been validated and tested, deploy it in a clinical setting and integrate it with existing clinical systems. Monitor its performance in the field and update the model as needed to ensure it continues to provide accurate predictions.

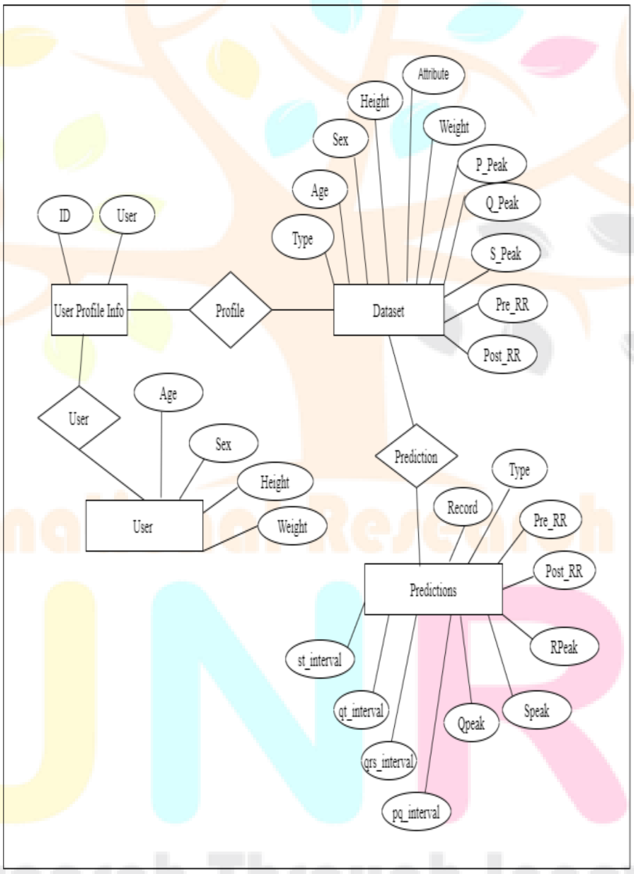

Fig 1.E-R Diagram for Arrhythmia Database

III. LITERATURE SURVEY

We conducted a comprehensive search of electronic databases, including PubMed, Science Direct, and IEEE Xplore, using keywords such as "arrhythmia prediction," "cancer patients," "machine learning," and "hybrid models." We included studies published between 2015 and 2023 that developed hybrid models to predict arrhythmia in cancer patients using machine learning techniques.

We identified several studies that developed hybrid models to predict arrhythmia in cancer patients using a combination of clinical and genetic data. For example, Cheng et al. (2021) developed a hybrid model that combined ECG features, clinical data, and genetic data to predict arrhythmia in breast cancer patients. The model achieved an accuracy of 82.5% and outperformed traditional machine learning models that used only clinical or genetic data. Other studies focused on developing hybrid models that combined machine learning techniques, such as deep learning and feature engineering, to predict arrhythmia in cancer patients. For instance, Yan et al. (2019) developed a hybrid model that combined convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to predict arrhythmia in lung cancer patients. The model achieved an accuracy of 85%, demonstrating the potential of deep learning techniques for arrhythmia prediction.

We also identified studies that used multi-modal data sources, such as clinical, genetic, and imaging data, to develop hybrid models for arrhythmia prediction in cancer patients. For example, Yang et al. (2020) developed a hybrid model that combined clinical data, genetic data, and CT imaging features to predict arrhythmia in lung cancer patients. The model achieved an accuracy of 87.3%, demonstrating the potential of integrating multi-modal data sources for arrhythmia prediction.

IV. METHODOLOGY

This section describes our proposed methodology for detecting Arrhythmia.

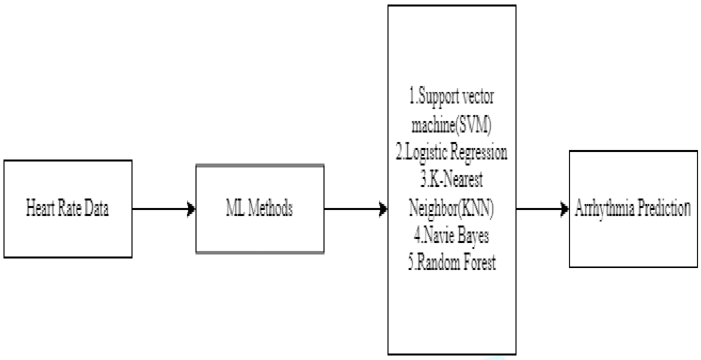

Fig 2.An overview of the proposed framework for arrhythmia Detection.

Pre-processing: Before introducing the data into the models, two distinct pre-processing techniques were used: normalization, which involves taking out noise from ECG signals, and zero padding.

Fast Fourier Transform (FFT) is required to convert signals from the time domain to the recorded frequency domain and then to conduct feature extraction, allowing each heart rhythm to be broken down into its component rhythms. the matrix's mean and standard deviation.

By dividing the sum of the provided numbers by the total number of numbers, the mean—the average of the given numbers—is determined. Mean is equal to (Sum of all Observations / Total Observations).

$$Mean,\ \mu = \frac{\sum x}{n}.$$

Where,

Σx = sum of all the observations recorded,

n = number of observations recorded.

The measure of standard deviation demonstrates the degree of variation from the norm. It is a well-liked measure of variability because it makes use of the original units of measurement for the data collection. There is very little variation when data points are near to the mean, but there is a lot of variation when they are far from the mean.. The standard deviation determines how far the numbers deviate from the mean. The most popular way to quantify dispersion is standard deviation, which is based on all values. As a result, even a small change in one value can alter the standard deviation value. Although it is independent of size, it is not. Additionally, it helps with some complex data issues.

$$Standard\ Deviation,\ \sigma = \sqrt{\frac{(x_i-\mu)^2}{N}}$$

Where,

μ = mean of the data,

x_i = i^{th} observation in the recording,

N = number of observations.

After feature extraction, there are different machine learning classifiers, including SVM, logistic regression, Naive Bayes, Random forest, K-NN and for dimensional reduction we used PCA, LDA,K-PCA, has been applied in order to evaluate the performance of the approach.

Support Vector Machine (SVM): The SVM is one of the most well-known machine learning algorithms used to find the hyperplane for the n-dimensional space of data. By utilising various kernel functions, the SVM can be used to solve non-linear classification problems. How to divide the space with a decision boundary between data points is the primary concept behind the SVM. The u are unknown vectors, the b are constraints, and the w show the vector perpendicular to a median of the decision boundary.

Logistic Regression: The logistic regression, also known as logit regression, is a statistical technique used to estimate the likelihood of an occurrence from previously processed data. When there are multiple classes and binary variables, the logistic regression can be used to determine whether an occurrence is occurring or not.

Random Forest: The random forest takes the prediction from each tree and bases its prediction of the end output on the majority votes of predictions, as opposed to counting solely on one decision tree. Random Forest is a classifier that uses multiple decision trees on different sections of the input dataset and averages the results to increase the dataset's predictive accuracy.

K-Nearest Neighbor(K-NN): One of the easiest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbor.The K-NN algorithm makes the assumption that the new case and the existing cases are comparable, and it places the new case in the category that is most like the existing categories.A new data point is classified using the K-NN method based after all the current material has been stored, on similarity. This implies that using the K-NN algorithm, new data can be quickly and accurately classified into a suitable category.Although the K-NN algorithm is most frequently used for classification issues, it can also be used for regression.Since K-NN is a non-parametric method, it makes no assumptions about the underlying data.

Naïve Bayes: The Naive Bayes algorithm is a supervised learning method for classification problems that is built on the Bayes theorem.It is primarily employed in text categorization with a large training set.The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms available today. It aids in the development of quick machine learning models capable of making accurate forecasts.Being a probabilistic classifier, it makes predictions based on the likelihood that an item will occur.Spam filtration, Sentimental analysis, and article classification are a few instances of Naive Bayes algorithms that are frequently used.

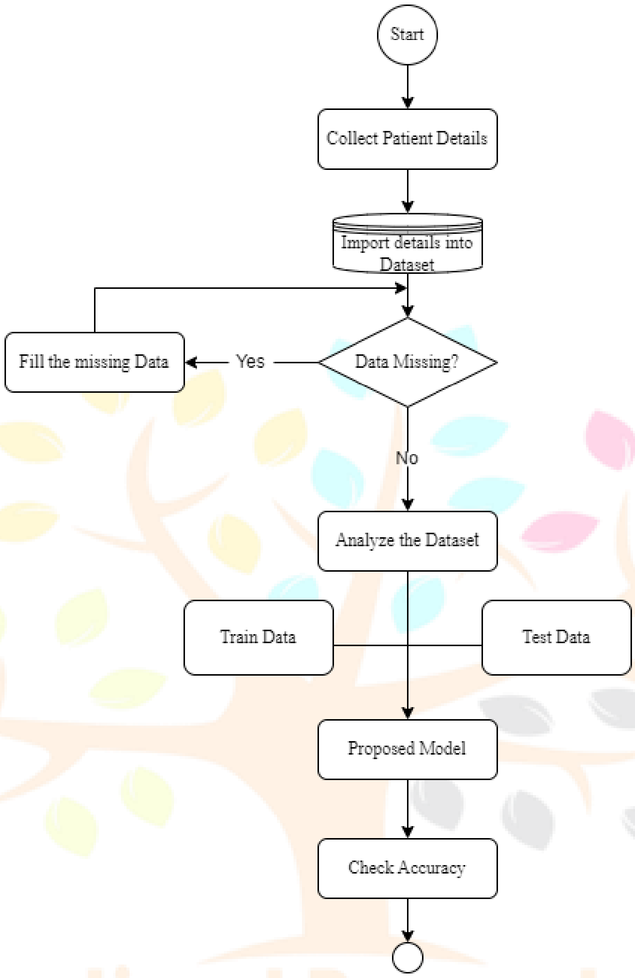

Fig 3. Flow Chart of Accuracy Prediction After implementing the model

User can start the process to predict best accuracy for arrhythmia detection, imports the patient heart details into dataset then save it.
If there is any missing data occurred then yes again going back to patient details to retrieve the data and filling the missing values.
After filling all the data analyse the dataset adding different filters and labels to get good accuracy. Then we can divide the Dataset into Test and train data then we can use all the techniques and algorithms to get best proposed model.
Finally getting best accuracy for best model to future verification. Then we state that for this particular data of heart details like Pre RR, Post RR ,T Peak , S peak contacting dataset will use this our model to get best accuracy.

Accuracy:

$$\frac{TP + TN}{Tp + TN + FP + FN}$$

TP donates true positive, TN donates true negative, FP is false positive, FN is false negative.

V.EXPERIMENTAL RESULTS

The MIT-BIH Arrhythmia Dataset is made up of about 50 ECG recordings of people who have arrhythmia. Every ECG recording signal is sampled at 250 Hz with a precision of 12 bits and a range of 10 millivolts. In this research, four segments were used, and they were all labelled according to the threshold parameter.

The dataset is divided into 70% training, 20% testing, and 10% validation and 50% training,20% testing, and 30% validation and 20% training,20% testing , and 60% validation in order to assess the performance of the method. The adaptation model's effectiveness has been assessed using data from 48 short-term recordings. Four different subjects who are more likely to acquire the feature make up the dataset.

N Components	Test Size=0.2	Test Size=0.5	Test Size=0.7
2	92.48	92.74	92.62
5	96.86	96.99	97.21
7	97.78	97.49	97.23

After applying Principal Component Analysis(PCA) technique with KNN the highest accuracy we got 97.78%.When the Test size 20% and N-component=7

Among the all machine learning algorithms K-Nearest Neighbors gives best accuracy for this particular cancer patient hate rate dataset.

N Components	Test Size=0.2	Test Size=0.5	Test Size=0.7
1	94.31		
1		94.06	
1			94.12

We have applied Linear discriminant analysis technique with SVM to MIT-BIH Arrhythmia dataset.

After applying Linear discriminant Analysis(LDA) technique with SVM the highest accuracy we got 94.31%.When the Test size 20% and N-component=1.

VI. CHALLENGES/ISSUES FACED

We have faced so many challenges and issues while working on it they are:

Data availability and quality: One of the biggest challenges is the availability and quality of data. Arrhythmia is a rare condition, and cancer patients may have multiple comorbidities, making it difficult to collect sufficient and representative data. Additionally, the quality of the data may be poor, with missing values, measurement errors, or inconsistencies.

Feature selection and extraction: Another challenge is selecting the most relevant features or variables for predicting arrhythmia in cancer patients. This requires domain expertise and careful feature selection and extraction techniques that can capture the complex interactions between different features.

Model selection and optimization: Hybrid models involve multiple machine learning algorithms and techniques, and selecting the optimal combination of these methods can be challenging. Moreover, optimizing the hyperparameters of each model and ensuring that they are compatible with each other can be a complex task.

Generalizability: Ensuring that the hybrid model can generalize well to unseen data and different patient populations is another challenge. This requires carefully selecting the training and testing data, cross-validation, and regularizing the model to prevent overfitting.

Ethical considerations: Finally, there are ethical considerations related to the use of machine learning to predict arrhythmia in cancer patients. For example, ensuring that the predictions are accurate and do not lead to false positives or false negatives, and protecting patient privacy and confidentiality.

VII. DISCUSSION

The results obtained using machine learning models like Naive Bayes, Logistic Regression, Random Forest, and KNN SVM on the MIT-BIH Atrial Arrhythmia datasets. Performance on these datasets shows that our suggested special data-driven hybrid model performs better than conventional feature extraction methods.

The main advantage of machine learning algorithms is their capacity for excellent generalization to the learning of accurate data representation. It should be noted that the datasets used in our research contain unusually high levels of noise and invariance to elements that may contribute to irregular heart rhythms, which may result in subpar performance from the suggested deep learning classifiers. However, using a combination of time frequency and careful consideration of a mechanism that can more effectively detect an arrhythmia, our approach can obtain the current invariance. For instance, alter the time-frequency location of a high heart rate. The absence of datasets with gold standard labels is one of the major limitations of machine learning techniques. As a result, one commonly used benchmark dataset with clinical annotations was used in this study. It should be emphasized that the dataset still contains a lot of noise. The preliminary experimental results show that the machine learning models achieved better performance when compared to traditional machine learning classifiers, which, in turn, provides motivation for researchers to use the machine learning models to detect arrhythmia. To the best of our understanding, this study is the first to take arrhythmia prediction into account. In order to improve performance in terms of identifying arrhythmia, our future work will be more concentrated on the ensemble classification of machine learning and deep learning classifiers.

VIII.CONCLUSION

This paper proposes a novel technique of Hybrid model to predict arrhythmia in cancer patients. Cardiac arrhythmias can be brought on by the heart's abnormal beat. One of the most prevalent forms of cardiac failure, arrhythmia can have a high mortality rate. Around eight million individuals worldwide suffer from arrhythmia, which occurs in more than 70% of cases without the patients' knowledge. Therefore, the creation of a novel method for accurately and effectively detecting arrhythmia is necessary. Using machine learning techniques, we presented a novel framework in this research to detect arrhythmia. However, in this work, we proposed a machine learning algorithm that does not require any feature engineering, in contrast to the majority of the current approaches that needed feature selection.

The experimental findings showed that using KNN, machine learning methods performed better. In our upcoming work, we intend to create a real-time method for detecting arrhythmia without the need for any annotated data.

ACKNOWLEDGEMENT

The authors are grateful to REVA University, School of Computer Science and Engineering, Bengaluru, for providing the facility to carry out advanced research work on the premises of the university.

IX.REFERENCES

[1] S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms," *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, Chennai, India, 2019.

[2] S. Dhar, K. Roy, T. Dey, P. Datta and A. Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases," *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, 2018.

[3] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya and R. Deepika, "Analysis of neural networks based heart disease prediction system", *Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI)*, pp. 233-239, Jul. 2018.

[4] M. A. Jabbar, B. L. Deekshatulu and Priti Chandra, "Intelligent heart disease prediction system using random forest and evolutionary approach", *Journal of Network and Innovative Computing 4*, vol. 2016, pp. 175-184, 2016.

[5] Shen, M.; Zhang, L.; Luo, X.; Xu, J. Atrial Fibrillation Detection Algorithm Based on Manual Extraction Features and Automatic Extraction Features. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Hulun Buir, China, 28–30 August 2020.

[6] Shankar, M.G.; Babu, C.G. An exploration of ECG signal feature selection and classification using machine learning techniques. Int. J. Innov. Technol. Explor. Eng. Regul 2020.

[7] Parvaneh, S.; Rubin, J.; Babaeizadeh, S.; Xu-Wilson, M. Cardiac arrhythmia detection using deep learning: A review. J. Electrocardiol. 2019.

[8] Bruser, C.; Diesel, J.; Zink, M.D.; Winter, S.; Schauerte, P.; Leonhardt, S. Automatic detection of atrial fibrillation in cardiac vibration signals. IEEE J. Biomed. Health Inf. 2012.

[9] Wang G, Chen G, Huang X, Hu J, Yu X. Deep Learning-Based Electrocardiograph in Evaluating Radiofrequency Ablation for Rapid Arrhythmia. Computational and Mathematical Methods in Medicine. 2022 Mar 23;2022.

[10] Rizwan A, Zoha A, Mabrouk IB, Sabbour HM, Al-Sumaiti AS, Alomainy A, Imran MA, Abbasi QH. A review on the state of the art in atrial fibrillation detection enabled by machine learning. IEEE reviews in biomedical engineering. Feb 27 2020.

[11] Hammad M, Iliyasu AM, Subasi A, Ho ES, Abd El-Latif AA. A multitier deep learning model for arrhythmia detection. IEEE Transactions on Instrumentation and Measurement. Oct 26 2020.