

# SMART PREDICTORS: ADVANCING HEART STROKE DETECTION THROUGH MACHINE LEARNING TECHNIQUES

## Abstract

Currently, the number of deaths that are attributed to heart strokes is steadily rising on a daily basis. Sadly, identifying such disorders in human beings is a difficult and time-consuming endeavour. Data sets are a useful tool for managing jobs that are of such a complicated nature. Because it is necessary to continue monitoring the heart rate, the only way to accurately anticipate the incidence of heart strokes is through the use of automation. This particular article makes use of the MITBIH arrhythmia dataset since it is one of the datasets that is helpful to us. A number of different data mining strategies can be utilised in order to achieve automation for the work that was specified.

In this article, the methodologies that were utilised include decision trees, naïve bayes, k-nearest neighbors algorithm, and Random Forest algorithm. Therefore, the purpose of this research is to establish a comparison between the algorithms that have been discussed above and to determine which one is more accurate in terms of completing the assignment. At the end of the day, after all of the evaluations, we are able to declare that the algorithm Random Forest has achieved a level of accuracy that is 99%, which is the greatest level of accuracy among all of the algorithms.

**Keywords:** Disease Prediction, Arrhythmia; Stoke Predictor; Heart Stoke; Machine Learning; Random Forest; Decision Tree;

## Authors

### **S. Bavan Kumar**

Assistant Professor  
Department of Computer Science & Engineering  
St. Martin's Engineering College  
Hyderabad, India.  
bavankumar42@gmail.com

### **Dr. P. Deepan**

Associate Professor & Head  
Department of Computer Science & Engineering  
St. Peter's Engineering College  
Hyderabad, India.  
deepanp87@gmail.com

### **Dr. R. Santhoshkumar**

Associate Professor & Head  
Department of Computer Science & Engineering  
St. Martin's Engineering College  
Hyderabad.  
santhoshkumar.aucse@gmail.com

### **Dr. B. Rajalingam**

Associate Professor & Head  
Department of AI&DS  
St. Martin's Engineering College  
Hyderabad.  
rajalingam35@gmail.com

## I. INTRODUCTION

Cardiovascular Disease (CVD) is the most well-known and very dangerous sickness that occurs all over the world. Each year, a bigger percentage of the population passes away due to Cardiovascular Disease (CVD) than they do from any other illness [1]. Over the course of the year, 17.9 million people lost their lives due to cardiovascular disease (CVD), which accounts for approximately 31% of all deaths that occur around the globe. More than eighty-five percent of these fatalities are the result of heart failure and heart stroke. More than three-quarters of fatalities from cardiovascular disease occur in countries with low yields. Eighty-two percent of the 17 million premature deaths that occurred in people under the age of seventy in 2015 due to non-infectious diseases occurred in countries with a low birth rate, and thirty-seven percent of these deaths were caused by cardiovascular disease (CVD) [2]. It is possible to eliminate the majority of cardiovascular diseases (CVD) by addressing the risk factors that are readily apparent. These risk factors include the use of tobacco, unhealthy eating habits and obesity, physical inactivity, and the harmful consumption of alcohol in settings that are prevalent in the general population [3].

Those individuals who are diagnosed with cardiovascular disease (CVD) or who are at high cardiovascular risks (due to the presence of at least one risk factor, such as hypertension, diabetes, hyperlipidemia, or an illness that has effectively settled) are required to receive an early introduction and direction through the use of brief prescriptions, as specified [4]. This is necessary in order to prevent cardiovascular complications. The growth of greasy stores inside the conduits (atherosclerotics) and the formation of blood clusters are the root causes of cardiovascular disease (CVD), which ultimately leads to the development of the disease [5].

There is also the possibility that it is connected to damage to the blood vessels in organs such as the brain, the heart, the kidneys, and the eyes. Despite the fact that cardiovascular disease (CVD) is one of the primary causes of mortality and disability in the United Kingdom, it is possible to avoid it to a significant degree by leading a healthy lifestyle routine [6]. The majority of the time, coronary events and strokes are brought on by intense events, and the majority of the time, they are brought on by a blockage that prevents blood from flowing to the heart or the brain. The growth of greasy stores, which are the most inner dividers of veins, is the most well recognised reason associated with this phenomenon. In most cases, the presence of a combination of risk factors is the cause of cardiovascular failures and strokes. These risk factors include the use of tobacco, an unhealthy eating routine, and being overweight [7].

## II. LITERATURE SURVEY

In order to determine the extent to which various supervised machine learning (ML) algorithms are useful for predicting clinical occurrences, the purpose of this study is to evaluate and contrast the accuracy and internal validity of these algorithms [8]. Comparisons were also made between the results, which were produced via the use of two different statistical software platforms. The Framingham Heart Study, which began in 1948 in Framingham, Massachusetts, as a prospective study of risk factors for cardiovascular disease, provided the data that was utilised in this investigation. The data was obtained from the open database of the Framingham Heart Study. A comparative methodological research comparing

the various machine learning algorithms, including decision tree, random forest, support vector machines, neural networks, and logistic regression, was carried out. This was accomplished through the use of data mining procedures, which resulted in the development of three data models [9]. The area under a curve (AUC) was the global selection criterion that was used to choose the appropriate collection of hyperparameters and the type of data modification.

The fluctuation of physiological signals can provide valuable insights into the activities of the cardiovascular system as well as clinical cardiovascular disorders. Two of the most significant time series variables are the heart rate variability (HRV) and the pulse transit time variability (PTTV) [10]. The combination of HRV and PTTV, on the other hand, has the potential to improve the classification accuracy for heart failure, which is currently unclear. This experiment aimed to examine the enhancement of HRV-based heart failure identification with the assistance of PTTV analysis. This was accomplished by performing a simultaneous analysis of HRV and PTTV on both normal individuals and heart failure patients. The results of this analysis were analysed both simultaneously and simultaneously. Forty healthy volunteers and forty patients suffering from heart failure were recruited for the study [11].

The radial artery pressure waveforms and the standard limb lead-II electrocardiogram were both captured at the same time. HRV and PTTV analysis were carried out on the obtained RR and PTT time series by using the conventional time domain indices (MEAN, SDNN, and RMSSD), frequency domain indices (LF, HF, and LF/HF), and non-linear domain indices (SD1, SD2, sample entropy, and fuzzy measure entropy). The findings of the study indicated that all HRV indices, with the exception of MEAN ( $P = 0.1$ ) and LF/HF ( $P = 0.9$ ), exhibited noteworthy variations (all  $P < 0.01$ ) between the two groups [12]. However, it is worth noting that only MEAN in PTTV exhibited a substantial reduction in heart failure patients ( $P < 0.01$ ). Furthermore, when the HRV and PTTV indices were combined with the predicted probabilities that were generated from the distance distribution matrix-based convolutional neural network models, the support vector machine classifier was able to achieve the highest classification performances. It was able to produce a sensitivity of 0.93, a specificity of 0.88, and an accuracy of 0.90. The potential of PTTV analysis for the diagnosis of clinical heart failure was established via the undertaking of this study [13].

One of the most frequent types of cardiomyopathy, known as dilated cardiomyopathy (DCM), is often linked with unfavourable results. With regard to the short-term follow-up, it has been shown that individuals with DCM who have a low ejection fraction have a bad prognosis [14]. Following an examination of the link between a variety of characteristics and the results, machine learning (ML) may be of assistance to physicians in the process of risk assessment and patient management. Through the use of machine learning, the current study intended to assist physicians in risk assessment and patient management by predicting the cardiovascular events that would occur in patients with severe DCM over a period of one year [15].

There were 98 patients with severe DCM (LVEF < 35%) from two different centres who provided the dataset that was utilised to create the machine learning model [16]. A whole set of 32 characteristics derived from clinical data were fed into the machine learning algorithm. The information gain (IG) technique was then used to determine the major aspects

that were highly relevant to the cardiovascular events. A naive Bayes classifier was constructed, and its prediction ability was tested by calculating the area under the curve (AUC) of the receiver operating characteristics by the use of 10-fold cross-validation [17].

The objective of this interdisciplinary industrial research project is to accomplish the development of a hybrid clinical decision support mechanism that is inspired by ontology and machine learning driven approaches [18]. This will be accomplished by merging evidence that has been extrapolated from historic patient data in order to assist cardiovascular preventative treatment. The cardiovascular clinical decision support framework that has been suggested is made up of two more critical components that are novel: (1) The Ontology-Driven Clinical Risk Assessment and Recommendation System, often known as ODCRARS (2) A medical prognostic system that is powered by machine learning (MLDPS) [19]. In order to accomplish the goals of the prognostic modelling, the most cutting-edge machine learning and feature selection techniques are utilised. In order to conduct out cardiac risk assessment for a variety of cardiovascular conditions, the ODCRARS is a knowledge-based system that is based on the knowledge of clinical experts and is encoded in the form of a clinical rules engine [20].

The MLDPS is a data-driven and knowledge-free system that was built by applying cutting-edge machine learning and feature selection algorithms to actual patient datasets [21]. It is a non-knowledge-based system. For the goals of development and clinical validation, clinical case studies in the RACPC, heart disease, and breast cancer domains are taken into consideration. According to the objectives of this research, a clinical case study in the RACPC/chest pain domain will be explored in great depth from the point of view of development and validation [22]. Validation of the clinical decision support framework that has been proposed is accomplished through the utilisation of clinical case studies in the cardiovascular domain. As part of this study, an effective cardiovascular decision support mechanism is demonstrated for the purpose of addressing inconsistencies in the clinical risk assessment of patients experiencing chest pain [23]. Additionally, this mechanism assists physicians in efficiently distinguishing patients suffering from acute angina and cardiac chest pain from those who are experiencing chest pain due to other reasons. A very excellent predictive power was achieved by the newly developed clinical models after they were examined in clinical practise. This resulted in an overall increase in performance when compared to the multivariate statistical classifiers that served as the benchmark. In preparation for more clinical studies, a number of different prototypes for assessing the risk of chest discomfort have been created and made available online [24, 25].

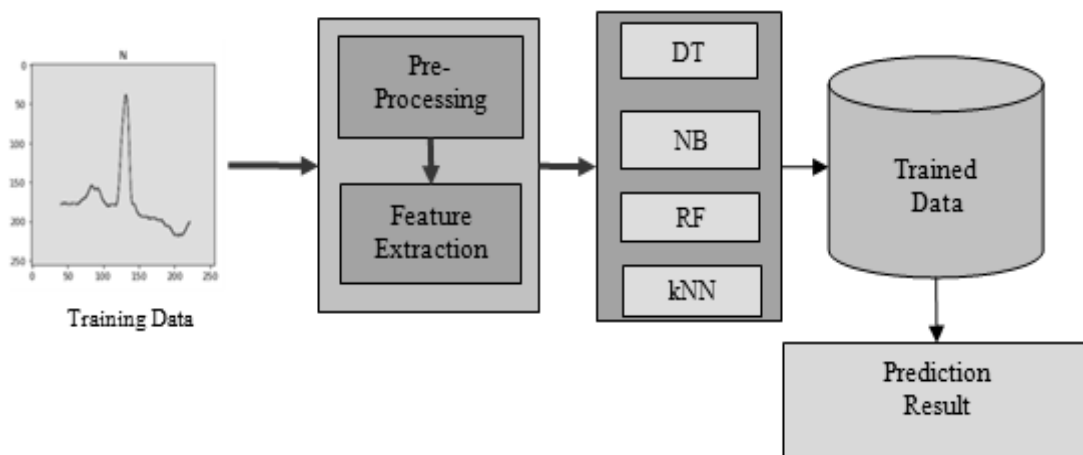
Wearable ECG monitoring sensors that are capable of real-time monitoring are among the most promising possibilities for aiding in the detection of cardiovascular disease. Within the scope of this research, we want to provide an innovative real-time machine learning system for the categorization of arrhythmias. The parallel Delta modulation and QRS/PT wave detection algorithms serve as the foundation for the architecture of the system [26,27]. Using three different types of feature vectors that are derived directly from the Delta modulated bit-streams, we propose a patient-dependent rotating linear-kernel support vector machine classifier. This classifier combines the global and local classifiers [28,29]. For the classification of SVEB, the F1 score, sensitivity, specificity, and positive predictivity value of the recommended SkP-32 technique are 0.83, 79.3%, 99.6%, and 88.2%, respectively. On the other hand, for the classification of VEB, the values are 0.92%, 92.8%, 99.4%, and 91.6%,

respectively. Based on the findings, it can be concluded that the performance of our suggested method is equivalent to that of research that has been published [30-32]. The low-complexity method that has been described has the potential to be utilised as a solution for machine learning that is performed on sensors.

### III. PROPOSED METHODOLOGY

This section can be examined advancing Heart Stroke Detection through Machine Learning Techniques. The working of proposed system is as follows:

- Dataset Collection
- Pre-processing operation
- Feature Extraction
- Train the Heart Stroke Detection model
- Classification model
- Predicting the Heart Stroke



**Figure 1:** Architecture of Heart Stroke Prediction using ML Model

#### 1. Collecting the Dataset

The heart stroke prediction dataset that was provided by Kaggle was utilised in the course of this research project. There are eleven characteristics that make up the dataset. These parameters include the following: identification, age, gender, hypertension, work type, residence type, heart disease, average glucose level, body mass index (BMI), marriage status, smoking status, and stroke.

The dataset comprises 11 characteristics, each of which indicates whether the data is category or numerical.

- **Id:** This element represents an individual's distinct identify. Quantifiable information that can be determined by mathematical calculations.

- **Age:** This attribute denotes the chronological age of an individual. This trait serves as an indicator of the person's gender. Quantifiable information that can be determined using mathematical operations.
- **Hypertension:** This characteristic determines the presence or absence of elevated blood pressure in an individual. Details on the categories.
- **Employment Type:** This attribute denotes an individual's occupational status.
- **Residence Type:** This attribute represents the individual's current living circumstance.
- **Heart Illness:** The presence of this feature suggests that the person has heart disease. Quantifiable information that can be computed.
- **Average Glucose Level:** This characteristic indicates the mean value of an individual's blood sugar concentration. Quantifiable information that can be determined using mathematical operations.
- **BMI:** BMI is an acronym for "Body Mass Index," which is a numerical measurement used to assess a person's body weight in relation to their height. This characteristic pertains to an individual's BMI (body mass index).
- **Marital Status:** data obtained from the category. This attribute defines a person's marital status. Smoking status: Statistical data categorising individuals based on their smoking habits. This attribute denotes whether an individual engages in smoking or not.
- **Stroke:** Quantitative information. This characteristic indicates whether an individual has experienced a stroke or not. The decision class encompasses all qualities except for the response class. The collection contains around 62,000 records. The input dataset is divided into two categories: the train dataset and the test dataset. The train dataset, which represents 80% of the total dataset, is used for training the model. A training dataset refers to a collection of data that is utilised to train a machine learning model. The test datasets are utilised to showcase the performance of the trained model.

**Table 1: Dataset Attributes and Corresponding Type**

S. No.	Attribute Name	Type
1	Gender	Numerical type
2	Age	Category type
3	Hypertension is there or not	Category type
4	Heart disease stroke	Numerical type
5	Marital status	Category type
6	Work type	Category type
7	Residence type	Numerical type
8	Average_glucoselevel	Numerical type
9	Bmi (body mass index)	Numerical type
10	Smoking status	Category type
11	Stroke	Numerical type

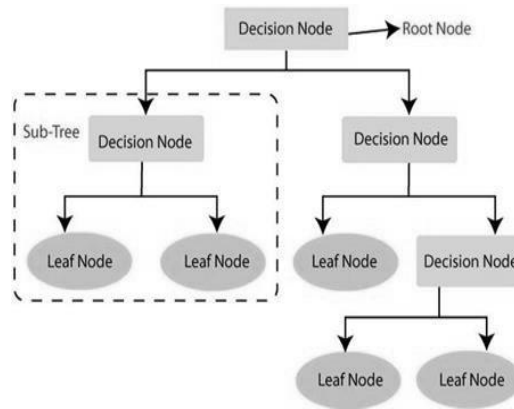
## 2. Perform Pre-Processing

The second phase in the operational process is the pre-processing of the data. It processes the dataset, if necessary, in order to get it ready for training. In this stage, the CSV file datas are pre-processed like removing duplicates, filling the dummy items using forward filling or backward filling. For the purpose of preprocessing, there are a few methods that may be utilised, including determining the mean of a column, finding the median of a column, and finding the mode of a column. Confirming the absence of values, When examining the dataset for values that are unique, determining the number of people who have been impacted by stroke, Assessing the gender distribution, By categorising the genders according to the stroke, Determine the number of people who smoke, categorise the smoking status according to the number of strokes.

## 3. Classification

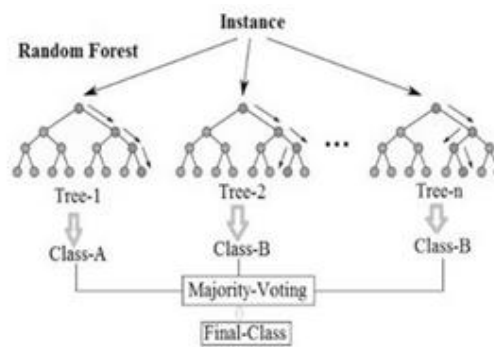
We have used four different types of ML classification algorithms such as Decision tree algorithm, Naïve Bayes algorithm, Random Forest algorithm and K- Nearest neighbor algorithm to find the performance metrics for this work.

- **Decision Tree:** The Decision Tree algorithm falls under the category of Supervised Learning and may be applied to address classification and regression issues. There is a significant emphasis placed on choosing the sequence in which the nodes (attributes) are arranged. The Gini index, impurity, entropy, and information gain are some of the particular metrics that we are able to analyse.



**Figure 2:** Working Principles of Decision Tree

- Naïve Bayes:** The Naïve Bayes Classifier is a statistical classifier that adopts the Bayes theorem as its reference. Bayesian classifiers are also considered to be statistical classifiers. There is proof, a hypothesis, and a likelihood included inside it.  $P(B|A)$  is the likelihood, and A stands for the hypothesis, B stands for the evidence, and The probability  $P(A|B)$  is a posterior probability, whereas the probability  $P(H)$  is the prior probability.
- Random Forest:** A random forest method is a type of classifier that takes a large number of decision trees from various subsets of a dataset and averages them out to determine the predictability of the dataset as a whole. It is a regulated machine learning algorithm that is typically utilised for the purpose of resolving classification and regression issues.



**Figure 3:** Random Forest Classifiers

- K- Nearest Neighbors:** Following the search for space training patterns, the K-classifier nearest neighbour in the pattern chooses the pattern that is the most similar to the pattern that is unknown. In the Euclidean distance, the closeness between samples is defined as the distance between them. A non-parametric approach known as the K-Nearest Neighbour method uses all of the data that is already available and organises a new data point depending on how similar it is to a previous data point. With the help of the K-NN algorithm, it is possible to categorise newly acquired data in a straightforward manner into the right category.



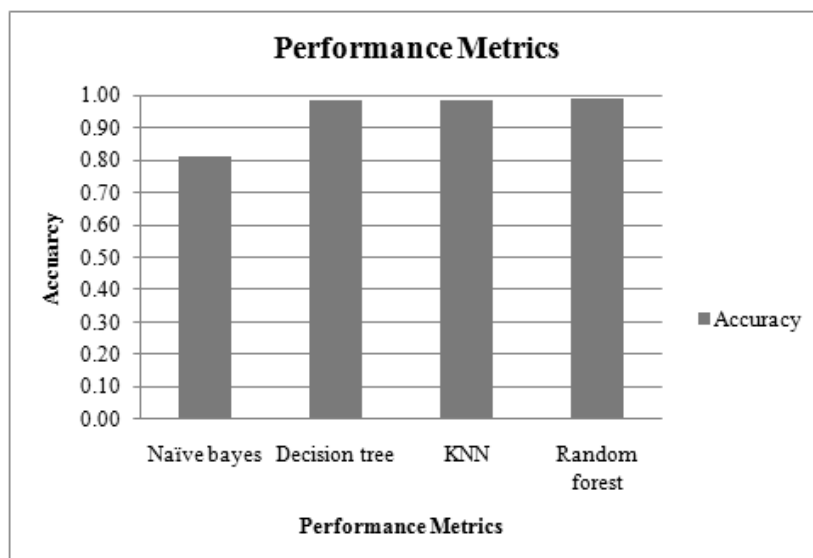
#### IV. EXPERIMENTAL DISCUSSION AND ANALYSIS

Experiments are carried out and data are gathered over the course of the research, and this section provides an explanation of how these processes are carried out. The Python programming language was utilised in order to carry out this experiment, and the methodologies of machine learning were utilised in order to carry out this research. We have utilised csv data and ecg iamges data to get this project off the ground. We have imported the dataset into CSV format by using the techniques in Python, and we have used the preprocess methodology. Additionally, we have applied the entire machine learning classifier models, and we have forecasted the performance metrics for that dataset. We divide the train, test split function in such a way that the train split has eighty percent of the data from the dataset, while the test split contains twenty percent of the data.

Both electrocardiogram (ECG) photographs and csv files have been used in the execution of this project. For the categorization of the csv file, we have utilised a number of different techniques associated with machine learning. These algorithms include the Decision tree algorithm, the Naïve Bayes strategy, the k-Nearest Neighbour algorithm, and the Random Forest algorithm. According to the measures used to measure accuracy, the scores are as follows.

**Table 2: Performance Analysis of Different Classification**

S. No.	Model	Accuracy
1	Naïve bayes	81.2%
2	Decision tree	98.5%
3	KNN	98.7%
4	Random forest	99%



**Figure 4: Perfoamnce Metrics of All the Four Classifiers**

## V. CONCLUSIONS

There has been a significant step forward in the field of medicine with the implementation of machine learning in the identification of cardiac strokes. A promising route for more precise diagnoses and personalised therapies is offered by machine learning, which has the capacity to analyse detailed patterns and anticipate hazards at an early stage. While technological advancements continue to be made, there is a significant possibility that stroke prevention may undergo a revolution, which will ultimately lead to improvements in patient outcomes and healthcare practices.

According to the results of all of the evaluations, we are able to conclude that the algorithm Random Forest has achieved a level of accuracy of 99%, making it the algorithm with the greatest accuracy rating among all of the algorithms. In future, we are planning to implement the heart stroke disease prediction model using deep learning techniques as well as planning to implement the work in Google Colaboratory for reduce the training time.

## REFERENCES

- [1] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation," Apr. 2018, doi: 10.1109/ICHI.2018.00092.
- [2] D. Zhang, Y. Chen, Y. Chen, S. Ye, W. Cai, and M. Chen, "An ECG Heartbeat Classification Method Based on Deep Convolutional Neural Network," *Journal of Healthcare Engineering*, vol. 2021, 2021, doi: 10.1155/2021/7167891.
- [3] T. M. Geethanjali, D. M. D, M. S. K, S. M. K, and A. Professor, "STROKE PREDICTION USING MACHINE LEARNING," JETIR, 2021. [Online]. Available: [www.jetir.org/d710](http://www.jetir.org/d710)
- [4] B. Deepak Kumar, S. Yellaram, S. kothamasu, S. Puchakayala, and A. Professor, "Heart Stroke Prediction using Machine Learning," 2021. [Online]. Available: [www.ijcrt.org](http://www.ijcrt.org)
- [5] 5. W. Ullah, I. Siddique, R. M. Zulqarnain, M. M. Alam, I. Ahmad, and U. A. Raza, "Classification of Arrhythmia in Heartbeat Detection Using Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2021, 2021, doi: 10.1155/2021/2195922.
- [6] 6. G. Sailasya and G. L. Aruna Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms." [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [7] L. Ladha and T. Deepa, Feature Selection Methods and Algorithms, *International Journal of CSE*, p.1787-1797, (2011).
- [8] J. Emakhu, S. Shrestha and S. Arslanturk, Prediction System for Heart Disease Based on Ensemble Classifiers, *Proceedings of the 5th International Conference on Industrial Engineering and Operations Management Detroit, Michigan, USA, August p. 10 - 14, (2020)*.
- [9] H. Kathleen, H. Julia, and J. George, Diagnosing Coronary Heart Disease Using Ensemble Machine Learning Diagnosing Coronary Heart Disease Using Ensemble Machine Learning, *International Journal of Advanced Computer Science and Applications*, p. 1-12, (2016).
- [10] K. Mohamad Almustafa, Prediction of heart disease and classifiers sensitivity analysis, *BMC Bioinformatics*, p. 1-5, (2020).
- [11] X. Gao, A. Amin Ali, H. Shaban Hassan and E.M. Anwar, Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method, p. 1-12, (2021).
- [12] P. Deepan, R. Santhosh Kumar, B. Rajalingam, P. Santhosh Kumar Patra and S. Ponnuthurai, "An Intelligent Robust One Dimensional HAR-CNN Model for Human Activity Recognition using Wearable Sensor Data," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 1132-1138, doi: 10.1109/ICAC3N56670.2022.10073991.
- [13] I.D. Mienye, Y. Sun, Z. Wang, An improved ensemble learning approach for the prediction of heart disease risk, *Informatics in Medicine Unlocked*, p. 312-322, (2020).
- [14] T.K. Ho, The random subspace method for constructing decision forests, *IEEE Transaction on Pattern Analysis Mach Intelligence*, p. 832-844, (1998).

- [15] K. Kalaivani, M. Priya, P. Deepan, L.R. Sudha, and J. Ganesh, "Heart Disease Prediction System Based on Multiple Feature Selection Algorithm with Ensemble Classifier", ECS Transactions, Vol. 107 (1), pp. 8049-8059, 2022. <https://doi.org/10.1149/10701.8049ecst>.
- [16] R.E. Schapire and Y. Singer, Improved boosting algorithms using confidence-rated predictions, Mach Learn, p. 297-336, (2011).
- [17] F. Leon, S.A. Floria, C. Badica, Evaluating the effect of voting methods on ensemble-based classification, IEEE International Conference on Innovations in intelligent Systems and Applications (INISTA), p. 1-6, (2017).
- [18] R.E. Banfield, L.O. Hall, K.W. Bowyer and W.P. Kegelmeyer, A comparison of decision tree ensemble creation techniques, IEEE Trans Pattern Anal Mach Intell, p. 173-180, (2017).
- [19] D. Ruta, B. Gabrys, C. Lemke, A generic multilevel architecture for time series prediction, IEEE Transaction on Knowledge Data Engineering, p. 350-359, (2011).
- [20] B. Zhang, Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images, IEEE Access, p. 110358-110371, (2019).
- [21] L. Han, S. Luo, J. Yu, L. Pan, S. Chen, Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes, IEEE Journal of Biomedical and Health Informatics, p. 728-734, (2015).
- [22] P.Deepan and L.R. Sudha, Scene Classification of Remotely Sensed Images using Ensembled Machine Learning Models, Proceedings in Lecturer Notes on Electrical Engineering, Springer Nature, p.535-550, (2021).
- [23] P.Deepan and L.R. Sudha, Deep Learning and its Applications related to IoT and Computer Vision, Artificial Intelligence and IoT: Smart Convergence for Eco-friendly Topography, Springer Nature, p. 223-244, (2021).
- [24] Chatterjee, D., & Chandran, S. (2019). Prediction and classification of heart disease using AML and power BI. Proceedings of the 11th International Joint Conference on Computational Intelligence. <https://doi.org/10.5220/0008381505080515>
- [25] H. T. K. (2023). Prediction of heart disease using machine learning with data mining. Physical Science & Biophysics Journal, 7(1), 1-6.
- [26] Pati, A., Parhi, M., & Pattanayak, B. K. (2022). IHDPM: An integrated heart disease prediction model for heart disease prediction. International Journal of Medical Engineering and Informatics, 14(6), PP. 1-12.
- [27] Shuvo, S. (2023). An early detection of heart disease using machine Learning (recurrent neural network). <https://doi.org/10.14293/s2199-1006.1.sor-.ppgv4id.v1>
- [28] VARKALA, K. (2022). Heart disease prediction system using Convolutional neural networks. <https://doi.org/10.21203/rs.3.rs-2009078/v2>
- [29] Agarwal, M. (2020). Serving very large numbers of low latency AutoML models. 2020 IEEE Infrastructure Conference. <https://doi.org/10.1109/ieeconf47748.2020.9377614>
- [30] Arzani, B., Hsieh, K., & Chen, H. (2021). Interpretable feedback for AutoML and a proposal for domain-customized AutoML for networking. Proceedings of the Twentieth ACM Workshop on Hot Topics in Networks. <https://doi.org/10.1145/3484266.3487373>
- [31] Celik, B., Singh, P., & Vanschoren, J. (2022). Online AutoML: An adaptive AutoML framework for online learning. Machine Learning, 112(6), 1897-1921. <https://doi.org/10.1007/s10994-022-06262-0>
- [32] Dai, W. (n.d.). Lower the barrier of machine learning : Meta learning for transfer learning and automl. <https://doi.org/10.14711/thesis-991012818569503412>