The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 20(2): S2: 243-248, 2025 www.thebioscan.com

PREDICTION OF CARDIOVASCULAR DISEASE USING MACHINE LEARNING & DEEP LEARNING TECHNIQUES

Dr K Venkata Nagendra[1], Associate Professor, Dept. of CSE, SRKR Engineering College, Bhimavaram.
Mr.G. Rajesh[2], Assistant Professor, Dept. of CSE, NBKR Institute of Science & Technology, Vidyanagar, Tirupathi DT, A.P
Mr. Erdi Raju Dayakar[3], Assistant Professor, Dept. of CSE, Sree Venkateswara College of Engineering, Nellore, AP.
Mr. J Jagadeswara Reddy[4], Assistant Professor, Dept. of CSE, Sai Rajeswari Institute of Technology, Proddatur.
Mr. Putheti Nagaraja[5], Assistant Professor, Dept. of CSE, Sree Venkateswara College of Engineering, Nellore, AP.
Mr.S. Sujith Kumar[6], Assistant Professor, Dept. of CSE, Sree Venkateswara College of Engineering, Nellore, AP.

DOI: 10.63001/tbs.2025.v20.i02.S2.pp243-248

KEYWORDS
Cardiovascular Disease,
Artificial Intelligence Convolutional Neural Networks,
Support Vector Machines,
Machine Learning,
Deep Learning,
Random Forest

Received on:

16-02-2025

Accepted on:

15-03-2025

Published on

06-05-2025

ABSTRACT

Healthcare is very important aspects of human life. Cardiovascular disease, also known as the coronary artery disease, is one of the many deadly infections that kill people in India and around the world. Accurate predictions can prevent heart disease, but incorrect predictions can be fatal. Therefore, here this paper describes a method for predicting cardiovascular disease that makes use of Machine Learning (ML) and Deep Learning (DL). The K-Nearest Neighbor method (KNN), Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), XGBoost (Extreme Gradient Boosting), Artificial Neutral Network (ANN), and Convolutional Neutral Network (CNN) are among the classifiers used in this paper. From Public Health Dataset required data is collected and focused on recognizing the best approach for predicting the disease in preliminary phase. This experiment end results show that the use of Artificial Neural Networks can be of much useful in prediction with better accuracy (95.7%) than compared to any other ML approaches.

INTRODUCTION

Cardiovascular disease represents a significant global public health issue. The number of individuals affected by heart disease is rapidly increasing due to inadequate health literacy and unhealthy lifestyle choices. The heart, an essential organ in humans, plays a crucial role in circulating blood throughout the body, functioning akin to a pump with a resting rate of approximately 72 beats per minute [1]. Research from the World Health Organization (WHO) indicates that heart-related ailments result in millions of fatalities each year worldwide, ranking among the top causes of death alongside mental stress, occupational strain, and various other illnesses [2]. An electrocardiogram (ECG) is utilized to capture the heart's electrical activity, serving as a common, non-invasive procedure for assessing cardiac health and swiftly identifying potential issues.

Historically, clinicians relied on auscultation to listen to heart sounds, a method that helps distinguish between normal and abnormal heart sounds. Its low equipment requirements, pain-free nature, and cost-effectiveness make it particularly suitable for cardiac assessments in smaller urgent care facilities, where the clinician's training is vital for accurately evaluating patients and identifying cardiac sounds [3].

Arrhythmia is another prevalent heart condition [4] [5] that can manifest in various forms. There are two primary classifications of arrhythmias based on heart rate: (1) tachycardia, which refers to a rapid heartbeat exceeding 100 beats per minute, and (2) bradycardia, characterized by a slow heartbeat of fewer than 60 beats per minute. Physicians typically recommend a standard ECG to assess the heart's rhythm [6].

In contemporary society, individuals are often preoccupied with their daily routines, leading to feelings of anxiety, restlessness, and stress. Each person exhibits a unique pulse rate and blood pressure, typically ranging from 60 to 100 beats per minute for pulse rate and from 120/80 to 140/90 for blood pressure. Globally, heart disease represents a significant concern for human health, with the term 'cardio' referring to the heart. Conditions classified

under heart disease fall within the domain of cardiology. The different types of heart disease are following:-

- Congenital heart disease.
- Arrhythmia.
- Coronary artery disease.
- Dilated cardiomyopathy.
- Myocardial infarction.
- Heart failure.
- Hypertrophic cardiomyopathy.
- Mitral regurgitation

RESEARCH METHODOLOGY

A number of traditional techniques have been employed to evaluate the outcomes against the suggested architecture. These methods will be briefly outlined, as they are widely recognized.

i).Decision Tree: A decision tree is a model in which each internal or intermediate node is labeled with an input feature. These intermediate nodes are those in which a decision must be made between several possible ones. Arcs are the unions between nodes and come from a node labelled with one input feature which is labeled with each of the possible values of the output or target feature. Alternatively, the arc leads to a subordinate decision node on a different input feature. Leaf nodes are labeled with a class or probability distribution between classes, which means that the tree has classified the data set into a specific class or a particular probability distribution so these nodes contain the final selected class.

ii). Random forest: Random forest [6] is a combination of decision trees. Each of the decision trees that make up the forest is built as follows: First, the number of data (N) and the number of variables of the classifier (M) are defined. The number of input variables that is used to determine the decision of a certain node is called *m*. For each node of the tree, *m* variables are chosen and from these *m* variables, the best partition of the set is calculated. To predict a new case, the nodes of the tree are traversed downwards and the label of the terminal node it reaches is assigned to it. This process is iterated throughout all the trees in the forest and the one that obtains the highest number of indices will be the one used as a predictor. Random forest is one of the most accurate learning algorithms, as long as a large enough data set is used [7].

iii). K-Nearest Neighbors (K-NN): The k nearest neighbors method is one of the simplest. Unlike other machine learning algorithms, k-NN does not generate a model from the training data, but rather the learning takes place at the same moment in which it is tested with the test data, therefore, it is a lazy learning method. The input data are vectors of dimension p of the form:

$$x_i = (x_{1i}, x_{2i}, \ldots, x_{pi}) \in X$$

In the training phase, the vectors and class labels of the training examples are stored and the distance between the stored vectors and the new vector is calculated in the classification phase, and the *k* examples closest to that new input data are selected. The new data are classified with the class that repeats the most in the selected vectors. Any metric can be used to calculate the distance, but the most common one is to use the Euclidean distance See below.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{p} (x_{ri} - x_{rj})^2}$$

iv). Ada Boost: AdaBoost consists of combining several weak classifiers in order to obtain a robust classifier. The main idea is to assign greater weights to poorly classified data and to assign less weight to data that has been well classified. Thus, each weak classifier focuses more on badly classified cases, thus improving the results. In our case, decision trees have been used
as the base method. The algorithm is made up of three steps. First, the weights are initialized
and each of the *N* samples is given the same weight. In a second step, a weak classifier is trained taking into account that if the data is correctly classified the weight is reduced and if it is badly classified, the weight is increased to give it more importance. Finally, the weak classifiers obtained in each training are combined into a strong classifier.

v) XGBoost: The XGBoost classifier has also been used to compare the results obtained with those obtained with our proposal. The difference between this method and AdaBoost is that in
each iteration, instead of assigning more weight to misclassified samples, XGBoost focuses on reducing losses. Each iteration focuses on reducing the error and establishing a new model to reduce the loss (negative gradient) further.

vi). Multilayer Perceptron (MLP): A multilayer perceptron is a neural network that has an input layer, an output layer and one or more intermediate layers with a certain number of neurons. It has the peculiarity that it has a linear activation function in all neurons and each neuron of a layer is connected with the neurons of the previous and next layer learning complex information on the input data.

vii). Support Vector Machines (SVMs): Support Vector Machines (SVMs) [18] are a category of supervised learning algorithms frequently employed for classification purposes. These algorithms function by establishing a boundary known as a "hyperplane" that distinguishes various classes of data points within a high-dimensional feature space. The main objective of SVMs is to construct a boundary that maximizes the distance to the nearest data points from each class, thereby ensuring an optimally separated hyperplane. SVMs excel in handling complex data sets of small to medium size and are particularly well-suited for cases where the data cannot be separated linearly, meaning it cannot be divided by a single straight line. In such instances, SVMs utilize a technique referred to as the kernel trick, which transforms the data into a higher-dimensional space to enable linear separation. These machines are applicable across numerous domains in machine learning, including text and image classification, as well as bioinformatics. They are grounded in a robust theoretical framework and have demonstrated outstanding performance in practical applications. Additionally, SVMs are effective for both classification and regression tasks, showcasing high efficiency in managing both linear and nonlinear data. [19]-[20]

viii).MLP: A multilayer perceptron is a neural network that has an input layer, an output layer and one or more intermediate layers with a certain number of neurons. It has the peculiarity that it has a linear activation function in all neurons and each neuron of a layer is connected with the neurons of the previous and next layer learning complex information on the input data.

ix). ANN: An Artificial Neural Network (ANN), also called as a "Neural Network," (NN)[18]. It's a multi-level system that uses mimics of neurons found in human anatomy to perform computations and numerical models. It performs the same function as a single neuron in the human brain.

x). CNN: A recent paper found that neural networks with application-specific settings, such as several hidden layers, can increase performance dramatically in a range of domains. A Convolutional Neural Network (ConvNet/CNN) is deep learning method that is used recognize distinct elements / objects in an input image (via learnable weights and biases)[19].

A sequential representation with a fully linked opaque layer, as well as flatten and dropout layers, is used to avoid over fitting. Machine learning and deep learning results are compared, and learning variances such as computational actual time and accuracy are described and displayed in the statistics listed further down in the results section.

III. CARDIOVASCULAR DISEASE PREDICTION

The Architecture of proposed work illustrating the prediction of cardiovascular diseases through Machine Learning (ML) and Deep Learning (DL) methodologies is presented below in Figure 1. This study utilized the Cleveland Heart Disease dataset obtained from the UCI repository, which comprises 845 records and 14 features. The dataset is divided into two parts: 20% is allocated for validation, while 80% is designated for training. Some examples within the dataset contain missing values for certain attributes. To facilitate training within this framework, these missing values have been substituted with the mean value of the respective attributes. Traditional classification models typically require that all attributes be within the same range. Given that the attributes

in this dataset vary in range, a standardization technique has been employed to normalize them to a uniform scale.

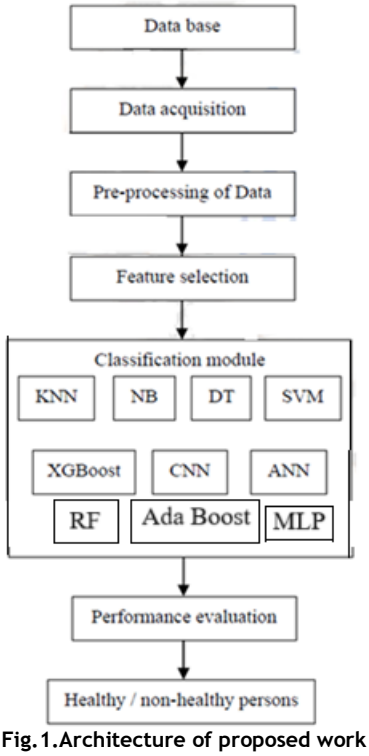

Fig.1.Architecture of proposed work

The dataset is complete with no missing values; however, it contains numerous outliers and exhibits improper distribution. Directly inputting the data into the machine learning algorithm without addressing the outliers or employing feature selection through two methods yielded poor outcomes. In contrast, the results achieved after applying the dataset's normal distribution to mitigate overfitting and utilizing isolation forests for outlier detection are highly encouraging. These pre-processing techniques are essential for effectively preparing data for classification or prediction.

In the process of feature selection, the Lasso algorithm, categorized under embedded methods, is utilized to identify and retain only the significant features. It demonstrates superior predictive accuracy compared to filter methods, resulting in a practical subset that is compatible with existing algorithms. Subsequently, to finalize the selected features, one should choose from the models available in the Scikit-learn library that are designed for feature selection. This study incorporates a total of ten machine learning algorithms, encompassing both deep learning and traditional machine learning techniques. The K-Neighbors classifier was employed to emphasize neighbor selection, followed by the application of random forest classification utilizing tree-based technologies such as the decision tree classifier, culminating in the use of the widely recognized ensemble methods. Additionally, the support vector machine was implemented to evaluate and manage the high dimensionality of the data.

IV.RESULTS ANALYSIS

This method makes use of the Cleveland Heart Disease dataset from the UCI (University of California, Irvine) repository. Accuracy, Precision, Recall, Specificity and In the evaluation process, F1-score is used. The vast majority of the data is used for training, whereas just 20% is used for validation. These parameters are divided into four categories: The first is a True Positive (TP) i.e. the value is identified as true and is actually true. The second type is a False Positive (FP), which occurs when a false value turns out to be true. The third is False Negatives (FNs). This happens when the value is true but the negative is false identified. True Negative (TN) is the fourth option, in which the value was truly negative.

Accuracy can be obtained by dividing true positive and true negative by true positive and true negative, and false positive, false negative by true positive and false negative. The formula is as follows:

$$Accuracy=TP+TN/(TP+TN+FN+FP)$$

Precision is a metric of exactness used to assess a classifier's performance. There are fewer false positives if the precision is high. There are more false positives in a model with lower precision means.

$$Precision=TP/(TP+FP)$$

Recall is a metric for determining a classifier's completeness. Higher recall equals fewer false negatives, while lower recall equals more false negatives. When recall improves, precision often suffers as a result.

$$Recall=TP/(TP+FN)$$

F1-score is a combination of accuracy and recall that can be calculated using the formula below:

$$F1-Score=2*((Precision*Recall)/(Precision+Recall))$$

After accuracy, specificity is a measure of how well a classifier recognizes negative situations. This is the proportion of true negative cases classified. It is also referred to as the true negative rate. The formula is as follows:

$$Specificity=TN/(TN+FP)$$

Different classifiers performance parameters comparisons are described in below Table. 1 as:

Classification Methods	Accuracy	Precision	Recall	F1-Score	Specificity
KNN	74.2	76.6	77.7	76.1	75.9
DT	76.8	75.4	74.8	77.9	76.4
NB	84.6	85.1	84.9	85.5	85.7
SVM	78.9	78.4	80.4	81.3	79.4
XG Boost	81.3	82.6	84.3	86.1	86.7
CNN	92.4	90.6	92.2	93.3	94.1

ANN	95.7	94.4	95.6	95.4	95.2
RF	75.6	74.4	73.7	76.9	74.6
Ada Boost	80.3	79.8	80.4	82.6	81.6
MLP	73.2	75.6	76.7	73.4	73.8

Table. 1: Performance of Different Classifiers

Performance parameters such as Accuracy, Precision, Recall, F1-Score and Specificity for different classifiers are represented in below Fig. 2, Fig. 3, Fig. 4, Fig. 5 and Fig. 6 respectively.

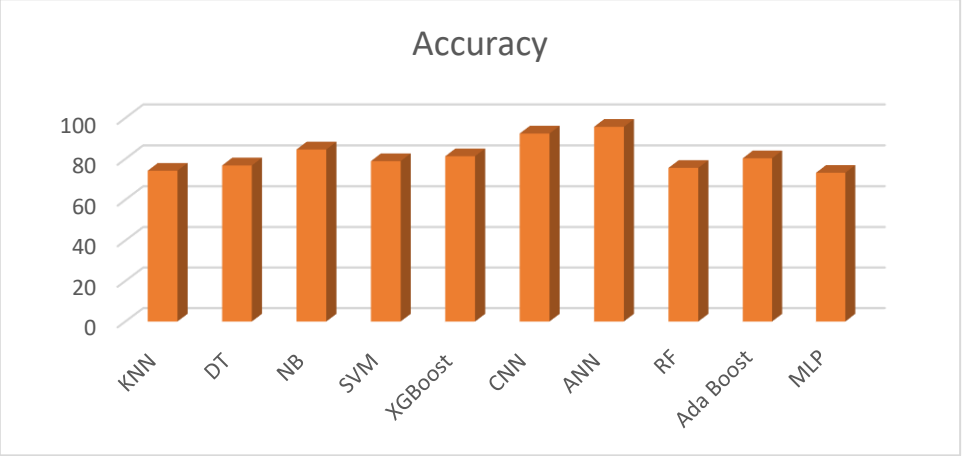

Fig.2. Accuracy Performance of various Classifiers.

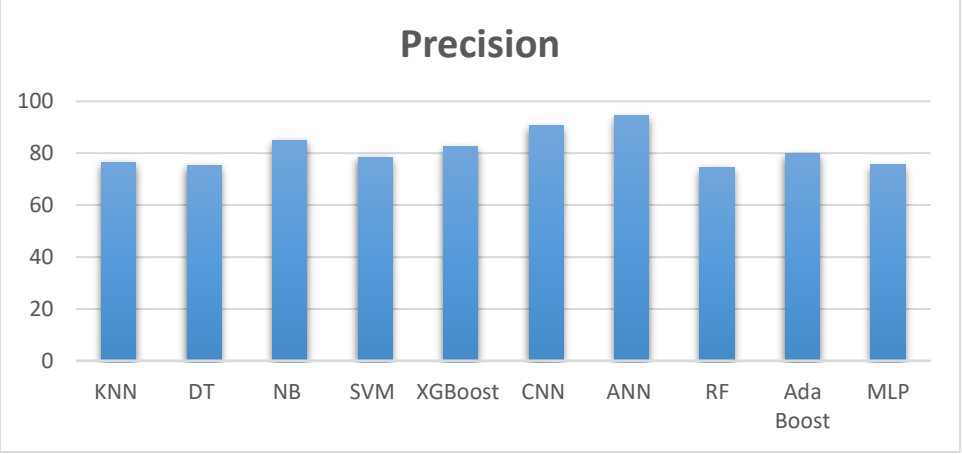

Fig.3. Precision Performance of various Classifiers.

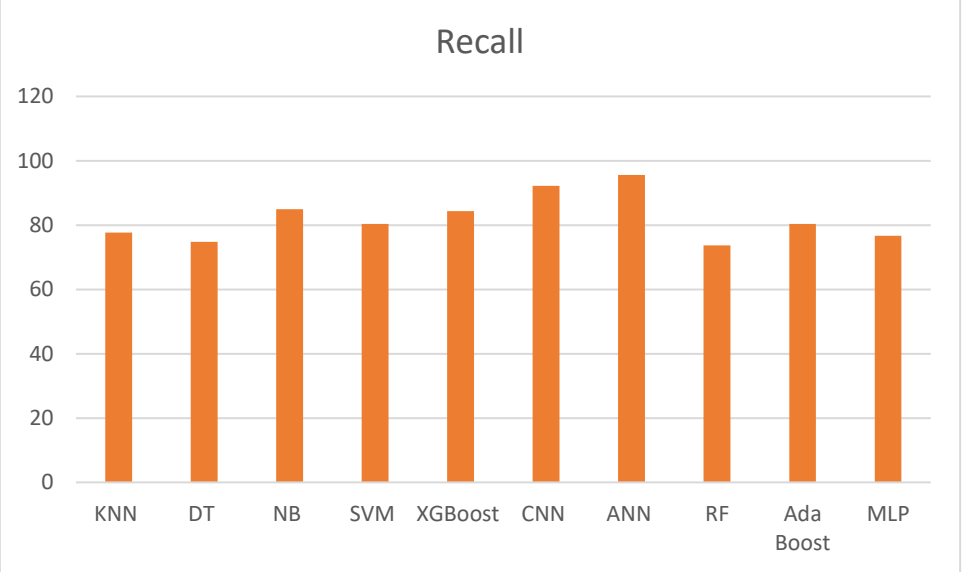

Fig.4. Recall Performance of various Classifiers.

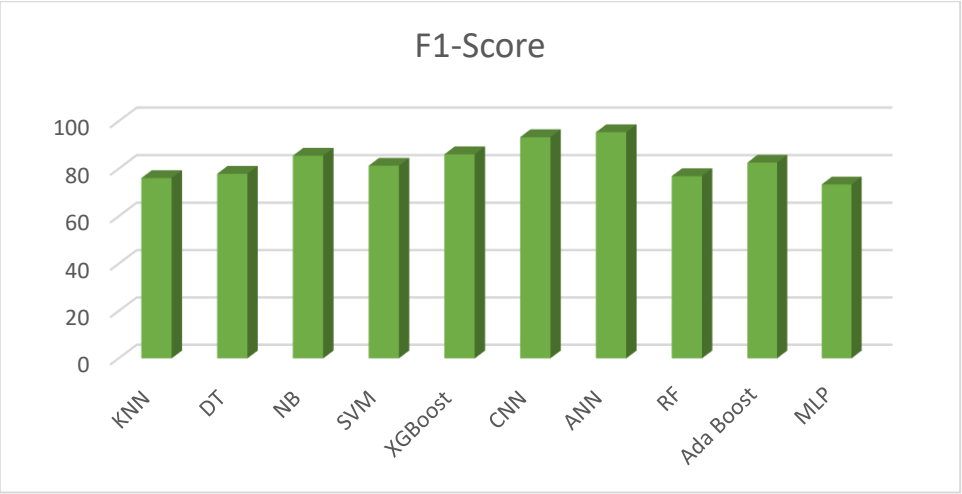

Fig.5. F1-Score Performance of various Classifiers.

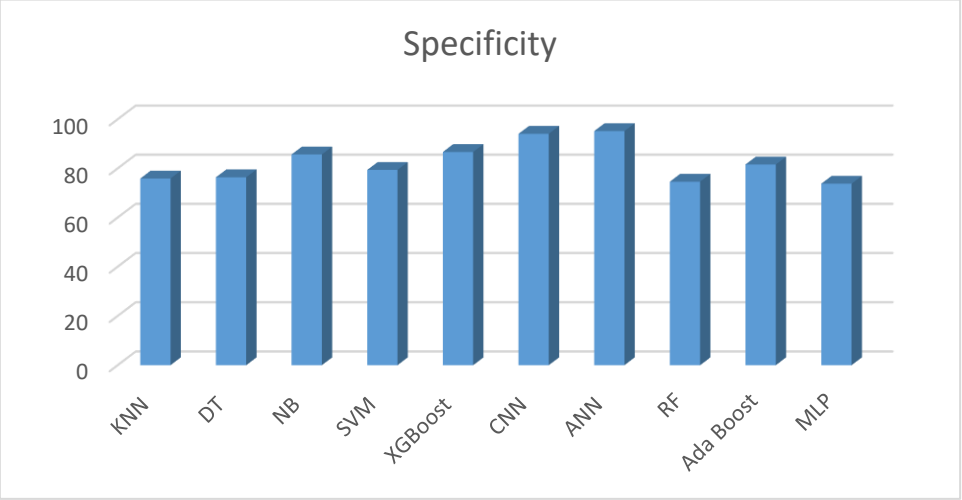

Fig.5. Specificity Performance of various Classifiers.

The results indicate that the performance of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) surpasses that of traditional machine learning models. Specifically, in terms of Accuracy, Precision, Recall, Specificity, and F1-Score, ANN demonstrates superior performance compared to CNN in the prediction of cardiovascular disease.

CONCLUSION

This paper presents machine learning (ML) and deep learning (DL) algorithms aimed at predicting cardiovascular disease. The techniques described include Synthetic Minority Oversampling Edited Nearest Neighbors (SMOTE-ENN), which were employed to balance the distribution of the training data. The classification and detection of cardiovascular disease were carried out using Ten different classifiers: KNN, SVM, NB, DT, XGBoost, RF,Ada Boost, MLP, ANN, and CNN. Performance evaluation metrics such as accuracy, precision, recall, specificity, and F1-score were utilized. The approach enhances the dataset size and incorporates deep learning along with various optimizations to yield more favorable outcomes. Notably, ANN demonstrates superior performance over CNN in predicting cardiovascular disease, achieving accuracy, precision, recall, specificity, and F1-Score metrics, with accuracy reaching as high as 95.7%.

REFERENCES

- [1]. F.B. Mayr, A. Spiel, J. Leitner, C. Marsik, P. Germann, R. Ullrich, O. Wagner, Effects of carbon monoxide inhalation during experimental endotoxemia in humans, Am. J. Respir. Crit. Care Med. 171 (4) (2005) 354–360, https://doi.org/10.1164/ rccm.200404-446OC.
- [2].Cardiovascular diseases (CVDs). 〈 https://www.who.int/news-room/factsheetsdeta il/cardiovascular〉 - diseases-(cvds). (Accessed 13 January 2021).
- [3]. M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, H. Fan, Heart sound classification based on improved MFCC features and convolutional recurrent neural networks, Neural Netw. 130 (2020) 22–32.
- [4]. Z. Zheng, Z. Chen, F. Hu, J. Zhu, Q. Tang, Y. Liang, An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology, Electron., MDPI 9 (121) (2020) 1–15, https://doi.org/10.3390/electronics9010121.
- [5]. A.D. Abdou, N.F. Ngom, O. Niang, ArrhythmiaS Prediction Using an Hybrid Model Based on Convolutional Neural Network and Nonlinear Regression, Int. J. Comput. Intell. Appl. 19 (3) (2020) 2050024, https://doi.org/10.1142/ S1469026820500248.
- [6]. R. Acharya, O. Shu Lih, H. Yuki, T. Jen Hong, A. Muhammad, G. Arkadiusz, T. Ru San, A. Deep, Convolutional neural network model to classify heartbeats, Comp. Biol. Med. 89 (2017) 389–396, https://doi.org/10.1016/ j. compbiomed.2017.08.022.
- [7]. A. Chetan, R.K. Tripathy, S. Dandapat, A diagnostic system for detection of atrial and ventricular

arrhythmia episodes from electrocardiogram, J. Med. Biol. Eng. 38 (14) (2018) 304–315, https://doi.org/10.1007/s40846-017-0294-5.
- [8]. Maria Sultana Keya, Muhammad Shamsojjaman, Faruq Hossain, Farzana Akter, Fakrul Islam, Minhaz Uddin Emon, "Measuring the Heart Attack Possibility using Different Types of Machine Learning Algorithms", 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Year: 2021.
- [9]. Nafis Mostafa, Muhammad Anwarul Azim, Md Rayhan Kabir, Rasif Ajwad, "Identifying the Risk of Cardiovascular Diseases From the Analysis of Physiological Attributes", 2020 IEEE Region 10 Symposium (TENSYMP), Year: 2020.
- [10]. Saiful Islam, Nusrat Jahan, Mst. Eshita Khatun, "Cardiovascular Disease Forecast using Machine Learning Paradigms", 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Year: 2020.
- [11]. Mariusz Filipowicz, & Waleed F. Faris. (2022). Recent Advancement in the Field of Analogue Layout Synthesis. Acta Energetica, (02), 01–07. Retrieved from http://actaenergetica.org/index.php/journal/article/view/462
- [12]. Mauricio Rodríguez Segura, Orietta Nicolis, Billy Peralta Márquez, Juan Carrillo Azócar, "Predicting cardiovascular disease by combining optimal feature selection methods with machine learning", 2020 39th International Conference of the Chilean Computer Science Society (SCCC), Year: 2020.
- [13]. Gudni Johannesson, & Nazzal Salem. (2022). Design Structure of Compound Semiconductor Devices and Its Applications. Acta Energetica, (02), 28–35. Retrieved from http://actaenergetica.org/index.php/journal/article/view/466
- [14]. Sinkon Nayak, Mahendra Kumar Gourisaria, Manjusha Pandey, Siddharth Swarup Rautaray, "Prediction of Heart Disease by Mining Frequent Items and Classification Techniques", 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Year: 2019.
- [15]. Kunal Rajput, Girija Chetty, Rachel Davey, "Risk Factors Identification for Heart Disease in Unstructured Dataset using Deep Learning Approach", 2019 International Conference on Data Mining Workshops (ICDMW), Year: 2019.
- [16]. Yang Peili, Yin Xuezhen, Ye Jian, Yang Lingfeng, Zhao Hui, Liang Jimin, "Deep learning model management for coronary heart disease early warning research", 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Year: 2018.
- [17]. Himanshu Sharma, M A Rizvi Prediction of Heart Disease using Machine Learning Algorithms: A Survey (August 2017).
- [18]. Kanna, D. R. K. ., Muda, I. ., & Ramachandran, D. S. . (2022). Handwritten Tamil Word Pre-Processing and Segmentation Based on NLP Using Deep Learning Techniques. Research Journal of Computer Systems and Engineering, 3(1), 35–42. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/39
- [19]. L. Verma, S. Srivastava, and P. C. Negi, ''A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data,'' J. Med. Syst., vol. 40, no. 7, p. 178, Jul. 2016.
- [20]. C.-J. Chen, Y.-T. Lo, J.-L. Huang, T.-W. Pai, M.-H. Liu, and C.-H. Wang, ''Feature analysis on heart failure classes and associated medications,'' in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Oct. 2016, pp. 1382–1387.
- [21]. Therasa Princy R, J. Thomas, " Human Heart Disease Prediction System Using Data Mining Techniques", International Conference on circuit, Power and Computing Technologies [ICCPCT],IEEE,2016.
- [22]. Thota, D. S. ., Sangeetha, D. M., & Raj , R. . (2022). Breast Cancer Detection by Feature Extraction and Classification Using Deep Learning Architectures. Research Journal of Computer Systems and Engineering, 3(1), 90–94. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/48.
- [23]. G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, "A machine learning system to improve heart failure patient assistance," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 6, pp. 1750–1756, 2014.
- [24]. Binal A. Thakkar, Mosin I. Hasan, Mansi A. Desai, "Health Care Decision Support System For Swine Flu Prediction Using Naïve Bayes Classifier", International Conference on Advances in Recent Technologies in Communication and Computing,2010.
- [25]. Maloth, Bhav Singh. (2016). Privacy-Preserving Scalar Product Computation over Personal Health Records. International Journal of Computer Engineering In Research Trends. 3. 42-46.
- [26]. (2022). Bug2 algorithm-based data fusion using mobile element for IoT-enabled wireless sensor networks. Measurement: Sensors. 100548. 10.1016/j.measen.2022.100548.
- [27]. Roy, S. S., Mallik, A., Gulati, R., Obaidat, M. S., & Krishna, P. V. (2017, January). A deep learning based artificial neural network approach for intrusion detection. In International Conference on Mathematics and Computing (pp. 44-53). Springer, Singapore
- [28]. Turovsky, O. L., Vlasenko, V., Rudenko, N., Golubenko, O., Kitura, O., & Drobyk, O. (2022). Two-Time Procedure for Calculation of Carrier Frequency of Phasomodulated in Communication Systems. International Journal of Communication Networks and Information Security (IJCNIS), 13(3). https://doi.org/10.17762/ijcnis.v13i3.5165 (Original work published December 25, 2021)
- [29]. Viswanathan, P., & Krishna, P. V. (2013). A joint FED watermarking system using spatial fusion for verifying the security issues of teleradiology. IEEE Journal of Biomedical and Health Informatics, 18(3), 753-764.
- [30].Negassa A, Ahmed S, Zolty R, Patel SR (2021) Prediction model using machine learning for mortality in patients with heart failure. Am J Cardiol 153:86–93. https://doi.org/10.1016/J.AMJCARD.2021.05.044
- [31]. Olsen CR, Mentz RJ, Anstrom KJ, Page D, Patel PA (2020) Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. Am Heart J 229:1–17. https://doi.org/10.1016/J.AHJ.2020.07.009
- [32]. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J (2015) Using EHRs and machine learning for heart failure survival analysis. Stud Health Technol Inform 216:40
- [33]. Pires IM, Marques G, Garcia NM, Ponciano V (2020) Machine learning for the evaluation of the presence of heart disease. Procedia Comput Sci 177:432–437. https://doi.org/10.1016/J.PROCS.2020.10.058
- [34]. Samuel OW, Yang B, Geng Y, Asogbon MG, Pirbhulal S, Mzurikwao D et al (2020) A new technique for the prediction of heart failure risk driven by hierarchical neighborhood component-based learning and adaptive multi-layer networks.
- [35]. Soni J, Ansari U, Sharma D, Soni S (2011) Predictive data mining for medical diagnosis: an overview of heart disease prediction. Int J Comput Appl 17(8):43–48. https://doi.org/10.5120/2237-2860.