A NOVEL APPROACH TO PREDICT HEART DISEASES IN PATIENTS USING HYBRID MACHINE LEARNING ALGORITHMS BASED ON STACKING ENSEMBLE TECHNIQUE IN AN IOT SYSTEM

Afroz Pasha[1], Dr. Nagaraja S R[2]

Article History: Received: 15.03.2023 **Revised:** 31.04.2023 **Accepted:** 14.06.2023

Abstract

Cardiovascular disease is the reason for global death nowadays is a crucial disease, since it takes away the lives of children, youngsters, and aged people. Early prediction is mandatory to avert the loss and many researchers had effectuated heart disease prediction models. The prediction accuracy must be high to provide effective treatment. In context with this, we propose a novel stacking approach for the IoT-based prediction of heart diseases with the utilization of K-Nearest Neighbors (KNN), Naïve Bayes (NB), Extreme Gradient Boosting (XGBoost), and Feed –Forward Neural network (FNN). The parameters such as Blood pressure, SPO2, and ECG are collected from the respective sensors and stored in cloud storage for further processing. The collected data are pre-processed using Min-Max Normalization to delete the unnecessary and blank data. Simulations are conducted in Python and validate the performances of the proposed approach with state-of-art works such as OCI-DBN, SMOTE, CNN, AND EDCNN with the metrics such as log-loss, specificity, F1-score, ROC, accuracy, Matthews' correlation coefficient and precision.

Keywords: Heart Disease Prediction, KNN, Stacking Machine Learning Approaches, and FFNN.

[1]Assistant Professor School of Computer Science and Engineering Presidency University, Bangalore
[2]Associate Professor School of Computer Science and Engineering Presidency University, Bangalore.

DOI: 10.31838/ecb/2023.12.6.47

1. INTRODUCTION

The most vital organ that a person has is the heart [1], which uses a combination of veins and arteries to circulate oxygen-rich circulation to specific other different organs. Heart illness is any condition that harms our hearts. Over 17 million individuals globally pass away from cardiovascular disease annually, based on a study released by the World Health Organization that was presented in 2019 [2]. Health complications come in many different forms, including fibrillation, congenital cardiac defects, and heart valve disease. Heart attack patients have much range of signs, including shortness of breath, lightheadedness, and excessive sweating. The main contributors to heart disease [3] include nicotine, hypertension, diabetes, being overweight, and others. Using rigorous methods to diagnose the illness is costly and uncomfortable. Thus, a method that can identify heart disease non-invasively and inexpensively is required.

The main reason for the overwhelming majority of mortality has occurred globally during the last few decades. It is regarded as a particularly prevalent health issue, and the heart's inability to flow adequate blood to fulfill the requirements of the body is the primary contributing factor [4]. Binge food, emotional stress, inherited conditions, or a sedentary lifestyle is possible causes. It infects a client gradually and is extremely challenging to treat when it has not advanced to a hazardous stage. The patient could suffer if the necessary corrective actions aren't really implemented. Hence, in order to encourage individuals to take good care of themselves as soon as a coronary heart manifestation is noticed, a cardiovascular disease [5] forecast is required.

Using a medical records system to precisely regulate the quality of medical services serves as the most popular technique to obtain all the documents relevant to treatment. It is hard to pinpoint blood vessel function because a number of variables, including pulse rate, saturated fat, creatinine, etc., have an impact on it. The several risk factors for heart disease [6] and discovered those that can be changed, including alcohol consumption, nicotine, diabetes, excessive triglycerides, and insufficient exercise. Rheumatic, arterial, and genetic heart diseases are only a few of the numerous varieties of what are referred to as cardiovascular disorders. Hence, the body's functioning throughout training, relaxation, as well as a function has already been examined [7].

Heart failure, uneasiness, difficulty breathing, excessive perspiration, heart palpitations, nausea, and exhaustion are all indications of coronary artery disease (CAD). Although significant has been accomplished in comprehending the intricate pathogenesis of cardiac insufficiency, extremely effective assessment is now required due to the sheer volume and unpredictable way that information and statistics must be analyzed and tracked. A reasonably precise method that can be used as an assessment tool to find underlying knowledge of heart disorders in healthcare records and forecast cardiac arrests earlier than they occur is thus continually needed. It is thought that such action might lead to more effective cardiac arrest treatment. The heart can be hurt following a coronary.

The pulse of one's heart and such capacity can push oxygen throughout the body may be impacted by this. Furthermore, users could be in danger of having repeat cardiac events or developing ailments like peripheral arterial disease (PAD) [8], kidney damage, or strokes. Reduces the necessity of the cardiovascular system to circulate additional blood to the tissues by improving the muscular ability to extract oxygen from the circulatory system. to overcome these issues in the prediction our work proposed a novel stacking machine learning approach that utilizes K-Nearest Neighbors (KNN), Naïve Bayes (NB), Extreme Gradient Boosting (XGBoost), and Feed –Forward Neural network (FNN) techniques. main contributions are,

- The data are collected from the IoT based sensors that are attached to the patients and the corpuses of the data are stored in the cloud storage for future use.
- The irrelevant images and outliers are removed with the help of Min-Max normalization-based image pre-processing.
- The pre-processed images are forwarded to the stacked machine learning model in which the first learner learns the data and outputs the results and generated a new dataset from this.
- The generated new dataset is forwarded to the second-level learners and predicts the data. Bagging, Boosting and Ada Boosting approaches are used for the stacking of machine learning methods.

The rest of the work is organized as follows, In Section 2, the relevant works are analyzed and listed their merits and demerits, and the background of the work is machine learning approaches are listed in Section 3. The proposed approach is explained in Section 4. Simulation details and results are enclosed in Section 5. Finally, the summarization of the work is done in Section 6.

2. LITERATURE SURVEY

Rani et al. [9] have described a hybridized feature selection algorithm depending on the client's patient features, which should contribute to the timely identification of cardiac disease. Especially in this particular system, the professional can sometimes receive assistance making a choice. The virtue of this strategy was that it will just be founded on scientific data and won't need a doctor specializing in heart disease. There are several, inaccurate heart disease decision assistance systems, the component selection strategy, and the concern of incomplete data jointly. The proposed system would diagnose heart disease more accurately than current systems. However, with this approach, the extent of cardiac disease cannot be determined.

Ali et al. [10] have presented an OCI-DBN (optimally configured and improved deep belief network) to fix certain issues and enhance network efficiency. Determining the ideal depth of each hidden node resolves the routing efficiency and deployment issue. After employing a straightforward optimization technique to arrive at the best return, users included a remote security operation. The data are returned in order to enhance the evolutionary system's potential to dedicate itself. It increases cardiac disease forecast with a recognition rate of 94.61% and can assist healthcare professionals in making wise judgments. Hence, the healthcare system is unsustainable now at the current time of complication.

Ishaq et al. [11] have implemented Synthetic Minority Oversampling Technique (SMOTE) to identify key characteristics and efficient data mining approaches that can improve the precision of the mortality estimate for cardiology disorders. Health data becomes tough and difficult physically manipulate in secure information extraction due to its huge different databases. Computational developments have also introduced precisely calibrated solutions for clinical applications while briefly examining medical data. The chance to improve the medical field and help doctors forecast how long a patient with heart failure will remain. Thus, the model is enhanced to perform better feature selection problems.

Mehmood et al. [12] have evaluated the convolutional neural networks (CNN) which forecast the likelihood that a service user will have a cardiac illness. Early detection of heart conditions could help avoid heart attack-related losses. A good grading system may enable the doctor to identify metabolic syndrome until it manifests itself. This present study uses image classification with a cutting-edge database from the data set to forecast the possibility of developing heart problems. The suggested model performs an overall accuracy is 97%. However, it is insufficient to predict other serious illnesses.

Pan et al. [13] describe the Enhanced Deep learning-assisted Convolutional Neural Network (EDCNN) to aid and enhance cardiovascular disease treatment prognostics. Using specialized learning methods, the expanded computational intelligence model parameters and identification have mostly been built including a deep multi-layer interpretation integrated to generate a reliable but instead changed for the better training dataset floating point numbers nonlinear transfer generalizations. It is reliable and precise in diagnoses of heart disease. Hence, for various applications, the precision should be increased.

Aggrawal et al. [14] highlighted a sequential feature selection algorithm often a significant aspect for recognizing death outcomes in patients with coronary artery disease following therapy. Typically, there are two sorts of methodologies they are channel and method-based methods. The exposure technique may usually create more optimal groupings than the broadcast technique, as well as the channel-based method with a high processing cost. The highlights can shorten the classifier model's runtime and increase classification precision. Thus, heart illness is identified to produce fundamental classifiers.

Khan et al. [15] suggested a MDCNN (Modified Deep Convolutional Neural Network) to rather reliably assess cardiac disease. The inputs were split into testing and training sets after selected features had been normalized. The generated information was then exposed to choosing and attribute grading using an analytical strategy. The approach employed an identical characteristic set for the evaluation of the information that it had used for retraining. The suggested classifier delivers greater accuracy than the current methods. Moreover, there is a lack of enhanced efficacy in the predictive classifier.

Ahmed et al. [16] have demonstrated a machine-learning algorithm that has a high level of precision in forecasting cardiac disease. Building an interactive system that can analyze and retrieve features about heart disorders from individuals broadcasting, Twitter messages to evaluate the probability that the client is in danger of acquiring heart disease. It can achieve great reliability in asynchronous pattern construction. Hence, the time complexity is more.

3. BACKGROUND

This section reveals the ML algorithms such as Naïve Bayes, K-nearest neighbor (KNN), Extreme

Gradient Boosting and Feed-Forward Neural Network.

3.1 K-nearest neighbor (KNN)

KNN [17] is the classifier that relies on the concept that uses nearest patterns to the pattern that uses the label information. Based on the majority of the K-nearest patterns the classes are assigned. It can be defined as,

$$\left\|m' - m_j\right\|^p = \left(\sum_{i=1}^{q}\left|(m_i)' - (m_i)_j\right|^p\right)^{1/p} \quad (1)$$

The Euclidean distance p=2, the Hamming distance is B^q and the label set of binary classification lie in the interval -1 to 1.

$$fKNN(m') = \begin{cases} 1 & if \ \sum_{i\in N}(m') \quad z_i \geq 0 \\ -1 & if \ \sum_{i\in N}(m') \quad z < 0 \end{cases} \quad (2)$$

Here m represents the observations and y denotes the labels in the q-dimensional patterns.

3.2 Naïve Bayes

This model [18] is effectively helps to detect heart disease with non-intricate nature of the parameters and is widely used as an algorithm that outperforms most classification approaches. It ads to evaluate the posterior probability $P(m|y)$ from the $P(m)$, $P(y)$, and $P(y|m)$. Based on conditional independence, the predictor (y) over the class is considered independent of the Bayes classifier. It can be expressed as,

$$P(m|y) = \frac{P(y|m)P(m)}{P(y)} \quad (3)$$

$$P(m|y) = P(y_1|m) \times P(y_2|m) \times ..... \times P(y_n|m) \times P(m) \quad (4)$$

The posterior probability of the targeted class over the predictor is $P(m|y)$, $P(m)$ is the class probability prior to the class, $P(y|m)$ the predictor likelihood probability, and the prior probability of the predictor is $P(y)$ [23].

3.3 Extreme Gradient Boosting

The format of XGBoost [19] follows the training format $y_i(i = 1,2,3,...,n)$ and the sequence of $z_i(i = 1,2,3,...,n)$. Mostly the XGBoost performs the allocation of observations' weight. The steps used in this are depicted in Algorithm 1.

Algorithm 1: Steps of XGBoost Algorithm
Initialize the inputs $y_i(i = 1,2,3,...,n)$ Initialize the parameters β (learning rate) and *l*() (Weak classifier) Output can be evaluated as $z_i(i = 1,2,3,...,n)$ Uniform distribution up-gradation $U_{t+1}(i) = \frac{U_t \exp(-\beta_t z_i l_t(y_i))}{a_t}$ Here, $a_t = \sum_{i=1}^{n} U_t(i)\exp(-\beta_t z_i l_t(y_i))$

3.4 Feed-Forward Neural Network (FFNN)

The FFNN [20] has a multilayer framework known as the input layer, hidden layers, and output layers. The output layer depends on the output that is required. If one output is required then one neuron is connected in the output layer and more. The structure is outlined in figure 1.

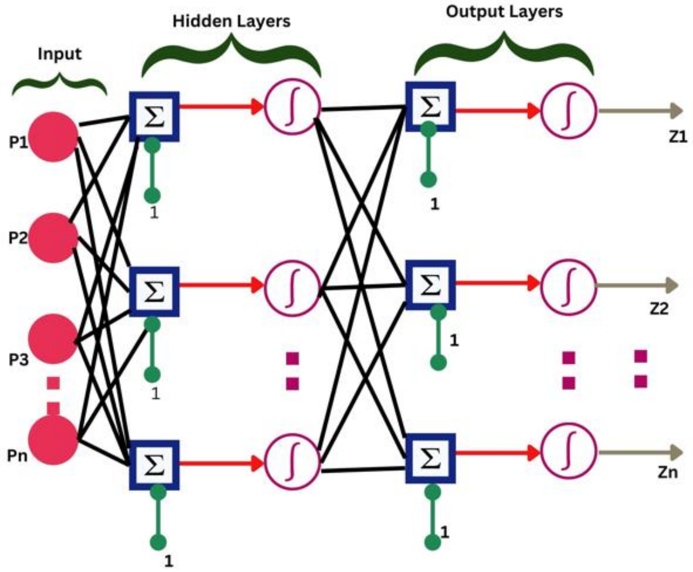

Fig 1: FFNN Architecture

4. PROPOSED METHODOLOGY

The proposed IoT based prediction of heart disease includes a collection of data, pre-processing of the data to remove irrelevant data, and noises, and conversion into the required format. The third step is utilizing our proposed hybrid machine learning stacked approach to predict the data. The data collection of our approach is based on the IoT sensors which are attached to the patients. From those sensors the data are collected remotely. The corpuses of the data are stored in the cloud for easy maintenance and accessibility. The overlay of proposed approach is framed in figure 2.

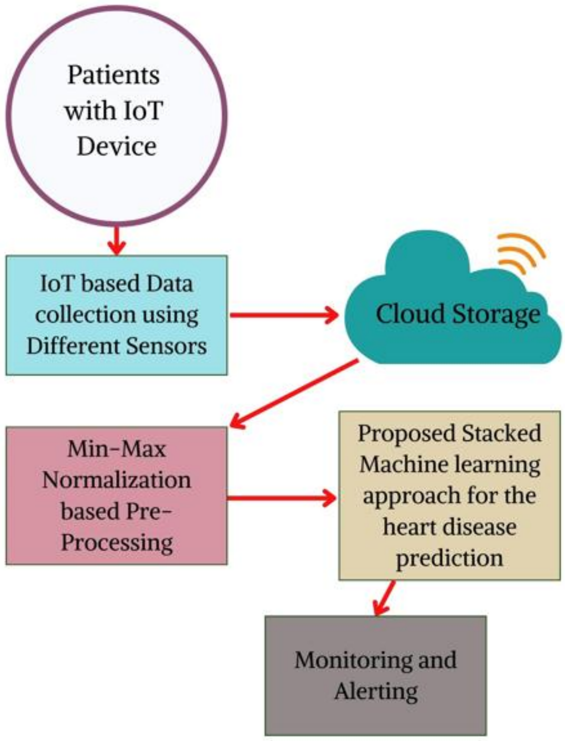

Fig 2: Overlay of proposed approach

4.1 Data Collection

The sensors are used to collect the heart rate and other details of about 2500 patients. For this, AD8232 sensor which collects the ECG signals, glucose level, blood pressure, blood fat level, peripheral pulse oximetry level, and pulse rates from the patients. AD8232 is an integrated signal conditioning block for ECG and biopotential applications and utilized to extract the signals, amplifies it and filters the needed signal from it. The collected data are forwarded to the cloud storage.

4.2 Cloud storage

Cloud storage is used to store both structured and unstructured data of any capacity and can be retrieved at anytime from anywhere. Since the collected data are unstructured the work utilizes cloud storage.

4.3 Data Preprocessing

It is an important process in data science to remove the noises from the images, and convert the images into the required format. Here we use the min-max normalization approach to effectuate the pre-processing of images. It also ignores the outliers identified, and missing values, and imputes the median values.

4.4 Proposed IoT based stacked machine learning model for the heart disease prediction

The proposed approach stacked ML approach is used for heart disease prediction. Our proposed approach utilizes several machines learning approach such as KNN, Naïve Bayes, XGBoost, and FFNN. Using these several models enhances the prediction accuracy, and mitigates the ensemble complexity. The combination of the machine learning methods is performed with the Bagging, Boosting [21], and Ada boost [22] which act as the meta-classifier for the merging of methods. Furthermore, the stacking hybrid machine learning approach is depicted in figure 3.

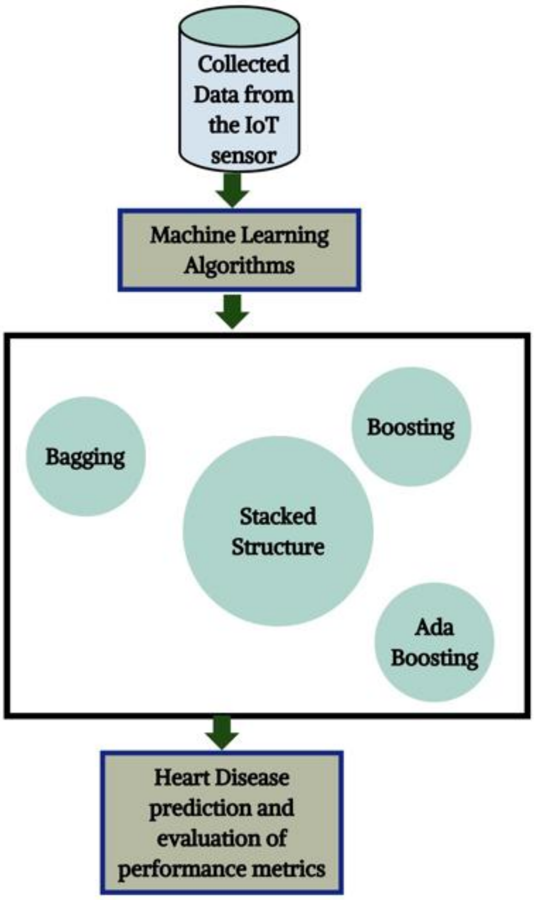

Fig 3: Proposed IoT based Hybrid Machine learning stacked approach for the heart disease prediction

Stacking of different machine learning approaches is combining the approaches stated above which includes two phases,

- Basic learners are at 1st level and stacking method learners are at level 2.
- At 1st level, the proposed models are used to learn the datasets and the outcome from these models are utilized to make a new dataset and the steps are explained in Algorithm 2.

Algorithm 2: Steps involved in stacking
Initialize the inputs 1st–level learning algorithms (L1) 2nd level learning algorithms (L2)

For t=1,….maxT
Training of 1st level individuals using $j_t = L_t(D)$
Applying 1st level
Dataset
End
Generation of a new dataset
For i=1,2,….,m
For t=1,2,…,maxT
Classification of training data
End
Update the dataset
Train the 2nd level learner
Applying 2nd level
Return the output.

Based on the predicted output the alert message is forwarded to the physician.

5. RESULT AND DISCUSSION

This section discusses the performance of the proposed heart disease prediction model and it is implemented in the JAVA platform. The experimental investigations belong to dataset collection, and performance metrics with its analysis are discussed in the below sub-sections.

5.1 Dataset description:

This study utilizes a data collected explicitly from the IoT sensors that are attached to the heart patients. The sensors are used to collect the heart rate and other details of about 2500 patients in which the ECG data signals. The cardiac problem presence or absence in the patients indicates the label coronary angiography, which includes 283 isnatnces with 71 features. The heart disease presence is represented by combining the values of original datasets and the dataset includes the characteristics of the patient's age, sex, BP level, type of chest pain, heart rate, blood sugar and etc. Examining practitioners and history information provides the participating patients in the investigation.

5.2 Performance Metrics:

The performance measures like precision, recall, accuracy, F-score, sensitivity, specificity, Matthews' correlation coefficient, log-loss and ROC to evaluate the HD detection performance. The accurately classified class measures its accuracy. A ratio of the total number of positive cases to the true positive (false positive Fp and truly positive Tp) cases defines precision. A good specificity and sensitivity or recall should be maintained for any medical predictive model. The metric recall for gauging how well an ability to form true positive instances. The number of actual negative (false negative Fn and truly negative Tn) case that the model correctly identified as negatives is known as specificity.

$$Accuracy = \frac{Tn + Tp}{Tp + Fp + Tn + Tp} \tag{5}$$

$$Sensitivity = \frac{Tp}{Tp + Fn} \tag{6}$$

$$Specificity = \frac{Tn}{Tn + Fp} \tag{7}$$

$$Precision = \frac{Tp}{Tp + Fp} \tag{8}$$

$$Recall = \frac{Tp}{Tp + Fp} \tag{9}$$

The harmonic mean of recall and precision score is the F-score metric.

$$F - score = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{10}$$

The measurable investigation of prediction is Matthews' correlation coefficient (MCC). The value of -1 (mediocre expectations) and 1 (exact expected value) is for MCC values.

$$MCC = \frac{(Tn * Tp - Fn * Fp)}{\sqrt{(Fp + Tp)(Fn + Tp)(Fp + Tn)(Fn + Tn)}} \tag{11}$$

The feature model performance is estimated with logarithm loss (log loss) in which the range of 0 and 1 for predictive input. Where zero is the ideal model for logarithmic loss. From the true value, the expected probability is separated by an increase in the logarithm loss. The expected to be terrible is 0.012 probability if is marked as a true perception.

5.3 Performance Investigation:

Figure 4 plots the comparison based on accuracy. The state-of-art techniques such as OCI-DBN, SMOTE, CNN, EDCNN and proposed to validate the effectiveness of accuracy. The overall performance of accuracy for OCI-DBN is 76%, SMOTE is 80%, CNN is 82%, EDCNN is 86% and the proposed method is 97%. However, the proposed framework outperformed a higher accuracy rate than other existing techniques of OCI-DBN, SMOTE, CNN, and EDCNN for heart disease prediction.

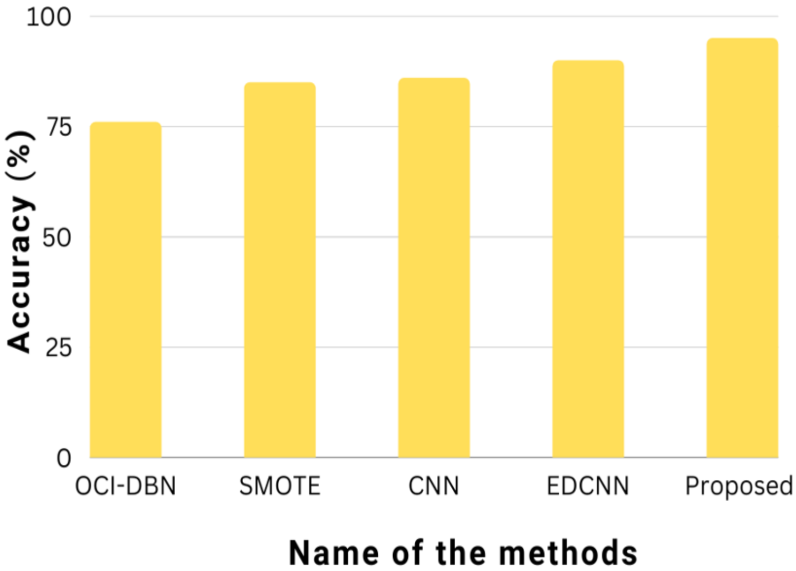

Fig 4: Comparison based on the accuracy

The comparison based on specificity and sensitivity is depicted in Figure 5. The performances of specificity and sensitivity are varied with respect to varying the performance of percentages. The overall specificity performance of OCI-DBN is 75%, SMOTE is 82%, CNN is 84%, EDCNN is 92%, and the proposed method is 97%. The overall sensitivity performance of OCI-DBN is 78%, SMOTE is 82%, CNN is 84%, EDCNN is 82%, and the proposed method is 96%. However, the proposed framework outperformed other existing techniques for heart disease prediction, such as OCI-DBN, SMOTE, CNN, and EDCNN, in terms of specificity and sensitivity.

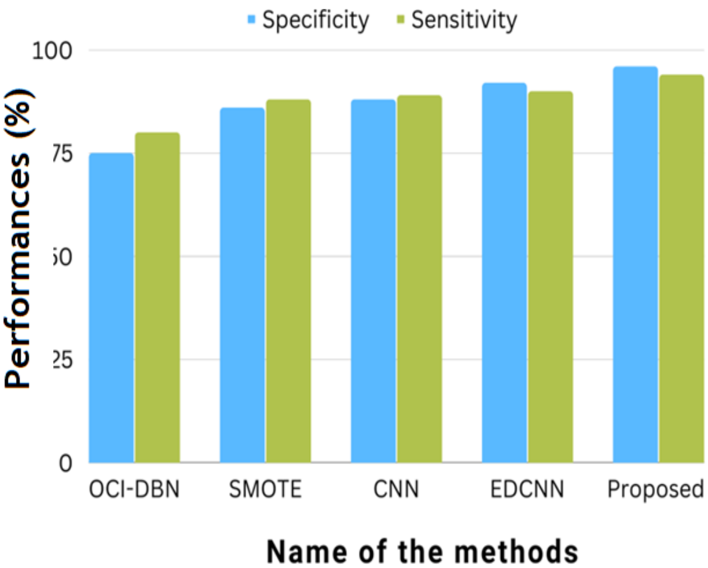

Fig 5: Comparison based on specificity and sensitivity

Figure 6 depicts the comparative study in terms of parameters such as F-score, precision, and recall. The performances of F-score, recall and precision are varied with respect to varying the performance of percentages. The overall precision performance of OCI-DBN is 81%, SMOTE is 80%, CNN is 90%, EDCNN is 93%, and the proposed method is 97%. The overall recall performance of OCI-DBN is 80%, SMOTE is 88%, CNN is 89%, EDCNN is 92%, and the proposed method is 96%. The overall specificity performance of OCI-DBN is 75%, SMOTE is 82%, CNN is 84%, EDCNN is 92%, and the proposed method is 96%. The overall F-score performance of OCI-DBN is 79%, SMOTE is 86%, CNN is 90%, EDCNN is 94%, and the proposed method is 97.2%. However, the proposed framework outperformed superior performances to other existing techniques such as OCI-DBN, SMOTE, CNN, and EDCNN for heart disease prediction in terms of precision, recall and F-score.

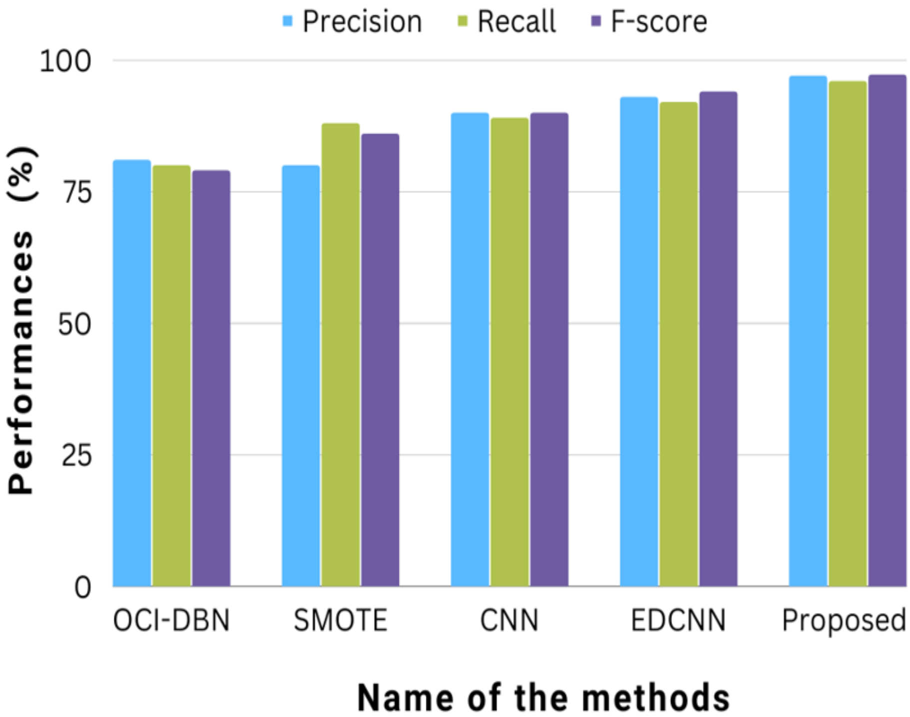

Fig 6: Comparison based on precision, recall and F-score

The state-of-art MCC, ROC and log loss are tabulated in Table 1. The state-of-art techniques such as OCI-DBN, SMOTE, CNN, EDCNN and are proposed to validate the effectiveness of log loss and MCC. The proposed method demonstrated 0.55% MCC, 4.26% log loss and 0.94% ROC, which provides superior performances in the case of ROC, MCC and log loss when compared to OCI-DBN, SMOTE, CNN and EDCNN methods.

Table 1: The state-of-art MCC, ROC and log loss

Methods	MCC (%)	Log loss (%)	ROC (%)
OCI-DBN	0.13	15.73	0.87
SMOTE	0.09	13.24	0.74
CNN	0.33	18.23	0.68
EDCNN	0.11	17.32	0.82
Proposed	0.55	4.26	0.94

6. CONCLUSION

This study presented a novel approach to predicting heart diseases in patients using hybrid stacking machine learning algorithms in IoT systems. Blood pressure, SPO2, and ECG parameters are among those that are gathered from the appropriate sensors and stored in cloud storage for later processing. To remove the useless and blank data, the collected data are pre-processed using Min-Max Normalization. Simulations are used to compare the proposed approaches performance to state-of-art approaches. The evaluation of the proposed framework is investigated using various kinds of evaluation criteria like Matthews' correlation coefficient, log-loss, sensitivity, ROC, specificity, F1 score, precision and accuracy, which are compared with the existing methods of OCI-DBN, SMOTE, CNN and EDCNN. When compared to the previous methods, the proposed framework demonstrated an accuracy of 97%, specificity of 97%, the sensitivity of 96%, precision of 97%, recall of 96%, F-score of 97.2%, 0.55% MCC, 4.26% log loss and 0.94% ROC.

7. REFERENCES

1. Panwar, Madhuri, et al. "PP-Net: A deep learning framework for PPG-based blood pressure and heart rate estimation." *IEEE Sensors Journal* 20.17 (2020): 10000-10011.
2. Mienye, Ibomoiye Domor, Yanxia Sun, and Zenghui Wang. "An improved ensemble learning approach for the prediction of heart disease risk." *Informatics in Medicine Unlocked* 20 (2020): 100402.
3. Liu, Xin, et al. "Blockchain-enabled contextual online learning under local differential privacy for coronary heart disease diagnosis in mobile edge computing." *IEEE Journal of Biomedical and Health Informatics* 24.8 (2020): 2177-2188.
4. Guo, Chunyan, et al. "Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the internet of medical things platform." *Ieee Access* 8 (2020): 59247-59256..
5. Suri, Jasjit S., et al. "COVID-19 pathways for brain and heart injury in comorbidity patients: A role of medical imaging and artificial intelligence-based COVID severity classification: A review." *Computers in biology and medicine* 124 (2020): 103960.
6. Shah, D., Patel, S. and Bharti, S.K., 2020. Heart disease prediction using machine learning techniques. *SN Computer Science*, *1*, pp.1-6.
7. Avanzato, R. and Beritelli, F., 2020. Heart sound multiclass analysis based on raw data and convolutional neural network. *IEEE Sensors Letters*, *4*(12), pp.1-4.
8. Forghani, N., Maghooli, K., Dabanloo, N.J., Farahani, A.V. and Forouzanfar, M., 2021. Intelligent oscillometric system for automatic detection of peripheral arterial disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), pp.3209-3218.
9. Rani, Pooja, et al. "A decision support system for heart disease prediction based upon machine learning." *Journal of Reliable Intelligent Environments* 7.3 (2021): 263-275..
10. Ali, Syed Arslan, Basit Raza, Ahmad Kamran Malik, Ahmad Raza Shahid, Muhammad Faheem, Hani Alquhayz, and Yogan Jaya Kumar. "An optimally configured and improved deep belief network (OCI-DBN) approach for heart disease prediction based on Ruzzo–Tompa and stacked genetic algorithm." *IEEE Access* 8 (2020): 65947-65958..
11. Ishaq, Abid, et al. "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques." *IEEE access* 9 (2021): 39707-39716..
12. Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T. and Masood, M., 2021. Prediction of heart disease using deep convolutional neural networks. *Arabian Journal for Science and Engineering*, *46*(4), pp.3409-3422.
13. Pan, Yuanyuan, et al. "Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform." *Ieee Access* 8 (2020): 189503-189512..
14. Aggrawal, Ritu, and Saurabh Pal. "Sequential feature selection and machine learning algorithm-based patient's death events prediction and diagnosis in heart disease." *SN Computer Science* 1.6 (2020): 344..
15. Khan, Mohammad Ayoub. "An IoT framework for heart disease prediction based on MDCNN classifier." *IEEE Access* 8 (2020): 34717-34727..
16. Ahmed, H., Younis, E.M., Hendawi, A. and Ali, A.A., 2020. Heart disease identification from patients' social posts, machine learning solution on Spark. *Future Generation Computer Systems*, *111*, pp.714-722.
17. Murugan, A., Nair, S.A.H. and Kumar, K.S., 2019. Detection of skin cancer using SVM, random forest and kNN classifiers. *Journal of medical systems*, *43*, pp.1-9.
18. Gandhi, B.S., Megawaty, D.A. and Alita, D., 2021. Aplikasi Monitoring Dan Penentuan Peringkat Kelas Menggunakan Naïve Bayes Classifier. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *2*(1), pp.54-63.
19. Zhang, Wengang, et al. "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization." *Geoscience Frontiers* 12.1 (2021): 469-477..
20. Haldorai, Anandakumar, and Arulmurugan Ramu. "Canonical correlation analysis based hyper basis feedforward neural network classification for urban sustainability." *Neural Processing Letters* 53, no. 4 (2021): 2385-2401.
21. Ghojogh, B. and Crowley, M., 2019. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. *arXiv preprint arXiv:1905.12787*.
22. Khan, Firoz, Jinesh Ahamed, Seifedine Kadry, and Lakshmana Kumar Ramasamy. "Detecting malicious URLs using binary classification through ada boost algorithm." *International Journal of Electrical & Computer Engineering (2088-8708)* 10, no. 1 (2020).
23. Vijay, M. M., and D. Shalini Punithavathani. "Implementation of memory-efficient linear pipelined IPv6 lookup and its significance in smart cities." Computers & Electrical Engineering 67 (2018): 1-14.
24. S.Balachander , Dr.R.Chinnaiyan (2019)," Internet of Things based Agro based

Monitoring using Drones with Block Chain , ICICCS 2019 , 15-17 May 2019, Madurai. (Recommended for Publication Scopus indexed IEEE Xplore digital library)

25. S.Balachander , Dr.R.Chinnaiyan(2018) , “Internet of Things based Reliable Real –Time Disease Monitoring of Poultry Farming Imagery Analytics” in ICCBI 2018 – Springer on 19.12.2018 to 20.12.2018 (Recommended for Scopus Indexed Publication IEEE Xplore digital library)
26. Sabarmathi . G and R.Chinnaiyan (2021). Sentiment Analysis for Evaluating the Patient Medicine Satisfaction. International Journal of Computational Intelligence in Control 13(2), pp.113-118.
27. Swarnamugi , Dr.R.Chinnaiyan (2019) Context-Aware Smart Reliable Service Model for Intelligent Transport System, ICRIC 2019, SCOPUS indexed , 8th 9th March 2019, Jammu & Kashmir.
28. Tilak, Dr.R.Chinnaiyan et.al , “ IOT based Leaf Disease Detection and Fertilized Recommendation” International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue-2, December2019.
29. Yamini, Dr.R.Chinnaiyan et.al, “Smart Crop Monitor using Internet of Things, Cloud, Machine Learning and Android App” International Journal of Engineering and Advanced Technology (IJEAT), Volume- 9 Issue-3, February, 2020.
30. Hari Pranav A; M. Latha; Ashwin. M. S; R. Chinnaiyan, "BlockchainAs a Service (BaaS) Framework for Government Funded Projects e-Tendering Process Administration and Quality Assurance using Smart Contracts," *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021, pp. 1-4, doi: 10.1109/ICCCI50826.2021.9402348.
31. Hari Pranav A;M. Senthilmurugan;Pradyumna Rahul K;R. Chinnaiyan , "IoT and Machine Learning based Peer to Peer Platform for Crop Growth and Disease Monitoring System using Blockchain," *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402435.