Heart Disease Prediction Using Machine Learning and Data Analytics Approach

Rina S. Patil and Mohit Gangwar

Abstract The heart is the consequent major part connecting to the brain, which has stronger precedence in the Human evidence. It elevates the blood and accumulations to all devices of the entire body. Forecast of circumstances of heart conditions in the medical field is important to work. Data analytics is beneficial for divining more knowledge, and it helps the medical center predict various conditions. A huge number of patient-related data is prepared each month. The collected data can be beneficial for the source of predicting the emergence of future weakness. Unusual data mining and machine learning procedures have been used to indicate heart disease. This research proposed heart disease prediction using various modified Recurrent Neural Network (mRNN) deep learning algorithms. Numerous feature extraction and selection methods have been used to get important features and data collection using custom-generated IoT environments. The system effectively provides heart risk scores with the highest accuracy in a runtime environment.

Keywords Internet of things · Deep learning · Feature selection · Feature extraction · Optimization · Classification · Supervised learning

1 Introduction

Health data mining provides immense promise to discover the secret trends in the scientific domain names data sets. For psychiatric disorders, these structures can be used. The providing raw patient records are therefore widely dispersed, homogenous, and weighty. It is useful to acquire such data in a structured manner. The health monitoring system, data gathered can then be incorporated. A user-oriented methodology to new and secret trends in the information is generated by business intelligence. It

R. S. Patil
Department of Computer Science and Engineering, Bhabha University, Madhya Pradesh, Bhopal, India

M. Gangwar (✉)
Department of IT, IET, Dr. Rammanohar Lohia Avadh University, Ayodhya, Uttar Pradesh, India
e-mail: mohitgangwar@rmlau.ac.in

V. Goyal et al. (eds.), *Proceedings of International Conference on Communication and Artificial Intelligence*, Lecture Notes in Networks and Systems 435,
https://doi.org/10.1007/978-981-19-0976-4_29

has been estimated from the WHO that 12 million people die internationally each year due to heart failure.

Related to cardio-related disease, half the deaths in the USA and other developing nations occur. It is also the primary cause of casualties in many developed countries. Most of all, it is considered the main reason for deaths worldwide. The word High Cholesterol includes the different diseases that cause the condition. The main cause of fatalities in various countries such as Sri Lanka was breast cancer. In the Western World, heart illness kills one adult every 34 s. Other types of heart disorders include coronary heart disease, cardiomyopathy, and cardiovascular disease. A wide variety of conditions influencing the cardiovascular framework and how air is filtered and distributed through the body are included in the word 'cardiovascular disease.' Several diseases, injuries and deaths arise from cardiovascular disease (CVD). A critical and complicated role in medication is the diagnosis of disorders.

Psychological care is an essential but complex task that remains to be improved precisely and effectively. This program's automation would be highly beneficial. Regrettably, in any sub-specialty, all physicians do not have the experience. Besides, there is a lack of staff at some sites. Therefore, by putting all of them together, and automated medical diagnosis scheme would possibly be extremely helpful. Reasonable judgment on computer-based knowledge help networks will support the achievement of clinical studies at a lower cost. A comparative analysis of the different technologies applied includes effective and precise implementations of the electronic machine. This paper aims to analyze the various statistical and machine learning strategies for evaluating machine learning introduced in past months.

2 Literature Survey

A new deep learning model for sex and age identification using a standard inertial sensor is suggested [1]. This approach is assessed with data obtained from more than 700 individuals using the highest sensor-based training repository value. Ten experiments of Monte-Carlo test sets were given to confirm the robustness and efficacy of the presented method. In protection applications, such as cell phone authorization, biometrics has been commonly embraced. Various biometric features has been tested for each individual object, based on extracted parameters btIoT modules system automatically detect the gender as well as age of user. A supervised neural technique uses a single inertial sensor attached to the lower back of gender and age recognition issues. Machine learning-based techniques have been commonly used throughout vision-based gender detection. Still, they haven't been used for electromagnetic sensor-based gender acknowledgment to our best understanding.

According to [2], a point fulfilment network utilizes a specific dimensional image taken from every suitable angle of view to perform 3d model of objects. The suggested technique has resolved many primary problems in vision-based disease diagnosis relative to previous methods, such as view deformation and scale uncertainty. For testing this method, various experiments were carried out. Still, the feasibility of

SVM and NB, the algorithm was demonstrated in the robust prediction. Tested objects have asked to remember the complete description of food products eaten over the last 24 h and the corresponding portion sizes to measure the calorie consumption. The size of both the food component relies mainly on the entity's personal opinions, which is often too unreliable. Under the circumstances of view occlusion, this approach demonstrates the effectiveness and accuracy of food volume estimation. The finished pixel value of the obstructed food products can be accessed by using the point fulfilment network.

In [3], a comprehensive approach to healthcare monitoring activity recognition using body receptors and a complex Convolutional Neural Network called CNN. They analyze signals from various body sensors in healthcare services, like ECG, wearable sensors, and control system sensors. A CNN based operation is equipped based on the features after removing essential characteristics from the wearable sensors based on PCA. Eventually, the deeply educated CNN used for the identification of data testing events. A publicly accessible standard dataset is applied to the defined solution and then contrasted with other traditional methods. The experimental analysis of this system show it is superior to other traditional classification systems while system works like sensor-based intelligent health systems for behavioral assisting. As a growing field of study, esteemed solutions to implementing a comprehensive smart healthcare delivery system for individuals to extend their independent lives must now be explored. The detectors were positioned on the neck, right wrist, or left ankle of the subject, separately, and connected using elastic bands. Several sensors help us quantify the movement experienced by different body parts, including momentum, turn frequency, and direction of the gravitational flux, thereby better measuring the body's nature.

A new cloud and IoT-based healthcare application was created by Ganesan and Sivakumar [4] to monitor and detect critical disorders. The classifier was trained using data from the benchmark dataset during the training phase. During the testing phase, real patient data was utilized to detect illness and the presence of disease. According to Majumder et al. [5], a numerous sensors system employing a smart IoT that provides an early warning of heart disease risk is suggested. The system gathers data from users through IoT devices and communicates it to a smartphone over Bluetooth using the Body Area Sensor (BAS) system. All the processing and data analysis took place in the application to view real-time user plots of future cardiac arrest. An IoT system with a low power consumption communication model developed regularly collects body temperatures and heart rate using a smartphone. Here ML and signal processing techniques were used to analyze sensor data and predict high accuracy cardiac arrests. A wearable device was implemented based on a smartphone for heart rate detection. It used a combination of ECG and body temperature. A heart rate analysis is done on the Android platform where users can view body temperature and plots of real-time ECG signals.

An adversarial training strategy to multitask learning is developed by Yu et al. [6] to calculate multi-type cardiovascular indices in MRI and CT. These task requirements are shared and taught via multitask forms of collaboration. Finally, they used CT to transfer characteristics learnt from MRI. A set of tests were carried out. The authors

used ten-fold cross-validation to improve system efficiency across 2900 myocardial MRI images. The network was then tested on a separate data set of 2360 cardiac CT images. All of the studies on the suggested reverse mapping show that it is quite good at calculating numerous cardiac indices.

In Ali et al. [7] implemented an Optimally Configured and Improved Deep Belief Network (OCI-DBN) to resolve the problems of existing IoT models and boost system performance. System used Ruzzo-approach Tompa's to delete features that do not contribute enough to boost processing speed. They stacked evolutional algorithms stacking two genetic algorithms to have an optimally designed DBN to obtain maximal configuration settings.

In Kumar et al. [8], cloud-based and IoT-based m-healthcare applications have been produced and updated to observe and diagnose the genuine level of severity. To gather the body parametric data using remote devices, like BP, pulse rate, heart rate, pulse pressure, hemogram etc. the approximate estimate can be collected as vital information, connected to the human body, gained from IoT devices. They were using the UCI Repository dataset and clinical sensors to foresee the average person with a major influence on non-communicable diseases medicinal knowledge is produced. The resulting information is securely stored by implementing a new federal storage method in five different steps, such as data storage, data recovery, data gathering, database division, and file merging.

In Park et al. [9] a smart wheelchair method was installed that monitors but visualizes the user's location through a mobile app that attempts to address the inconsistent position of the user. They used strain gauges and shift sensing, using IoT and wireless techniques to communicate with low consumption. It's an extension of Arduino that detects various user stances. This integration complements the user by offering real-time interactive and visualized data for mobile apps to sit properly and understand their present incarnation. With pressure displayed in red, yellow, pink and green triangles, this is a great example of Information systems.

Kumar et al. [10] researchers introduced IoT's three-tier architecture with early detection using a supervised learning algorithm to detect cardiac disease. This system also suggested multi-tier architectures that stored and processed massive quantities of data from wearable IoT environments. First-tier demonstrates on data collection from multiple analog sensors; tier two describes Apache Hadoop that store massive data in cloud storage generated from IoT module, while tier three uses Apache-Mahout to build a reinforcement learning detection and prediction model. In conclusion, ROC demonstrates to get entire system analysis of heart disease prediction.

In Nashif et al. [11] researchers designed a disease prediction system based on the cloud. A real-time monitoring system sensing health parameters like blood pressure, temperature, heartbeat, and humidity was developed using an Arduino microcontroller. The proposed system can detect heart disease using ML techniques as recorded data transmitted to a central server is updated every 10 s.

Suresh et al. [12] implemented an optimized prediction model using a genetic algorithm. System describes various prediction models and important feature selection algorithms. The accuracy and performance of system was better than other traditional prediction models. This method was tested against several prediction models using

heart disease datasets and was verified using real-time datasets. To create a balanced training and testing data set, the n–cross-validation process is used.

Researchers employed the rapid miner tool and several ML techniques to enhance the prior accuracy score and forecast heart illness in Alotaibi [13]. The heart disease dataset from UC Irvine was put to the test. The suggested approach enhanced the accuracy score previously obtained.

3 Proposed System

According to many factors, the medical remote management forecast period demand has expanded significantly. When it is very common today in developing countries that older adults typically live separately in their own homes, the elderly population is growing. Besides, the Internet of Things (IoT) makes these remote patient monitoring systems theoretically feasible (IoT as the idea of a capable and modifiable environment of surveillance in which controls or actuators to human and non-living particles) and makes it financially viable because of the even lower expense of sensors. The advancement of smart mobile technology, there are some inbuilt applications are available that provides automatic disease prediction after taking few input parameters, such as temperatures, blood glucose, breathing, mass spectrometry sensor etc. For example, personal health observers such as smart beds instantly notify who is utilizing them. Moreover, they are sure to enlighten about different patients' physical levels, making real smart home drug dispensers automatically warn when medication is not taken, for example. Several virtual care monitoring devices use various technologies inside facilities and their residences to monitor and track clinicians or diagnostic applications. Regrettably, most of these systems are not versatile when new sensors are introduced during runtime, as far as we know. Neither has it allowed regular users with the smart technologies added to generate ad-hoc notifications automatically.

Distributed data processing is one of the processes involved that can resolve a number of the protection, distribution, integration, and management challenges of aggressive data innovation. This propels current research toward thinking about pervasive clinical frameworks focused on the internet. The Predictive Analytic System hardware implementation predicts that disease relative to body temperature, pulse rate, and individual tension. The sensors are mounted on the human chest to monitor the physician's metabolic rate, heart rate, cholesterol levels. It sends data from the information it predicts and diagnoses the disease to the Arduino microcontroller. This information is hosted in the cloud via the end receiver. It can be monitored and anywhere whenever through the Internet. The condition of the physician will be registered on the database and retained. A smartphone framework for machine learning is designed for viewing data from sensors and sending emergency alerts. Healthcare practitioners can track, anticipate, diagnose, and inform their clients at any moment by using this method.

The conceptual system can control several predictor variables on wearable technology, integrated with such a handheld device. Sensor networks work together again to obtain the necessary data, except encountered some problems. It operates by reducing less human involvement to preserve its precision. Using cardiovascular sensors, it is a legitimate diagnostic device for remote areas heart-prone clinicians that monitor heart rate, cholesterol levels, metabolic rate, and many other metrics. N number of measurements obtained can be preserved and viewed subsequently by practitioners to correct acute and chronic treatment.

Figure 1 demonstrates propose system architecture for synthetic and real-time IoT data processing environment. It describes data preprocessing at the first phase, that does null values elimination, class balancing, data acquisition etc. Then various features extraction and selection has been done for next process. In training phase similar features has been considered for module training while for testing from test data. The outcome of system, it provides patient health report according to current parameters. Two major physician and patient interfaces would allow data to be transmitted to each other. In actual environments, this system recognizes fundamental heart problems for the sake of the health of the individual. The system able to provide heart risk score for both genders as well as it having an ability to provide systematic prediction of heart disease based on cardiac index.

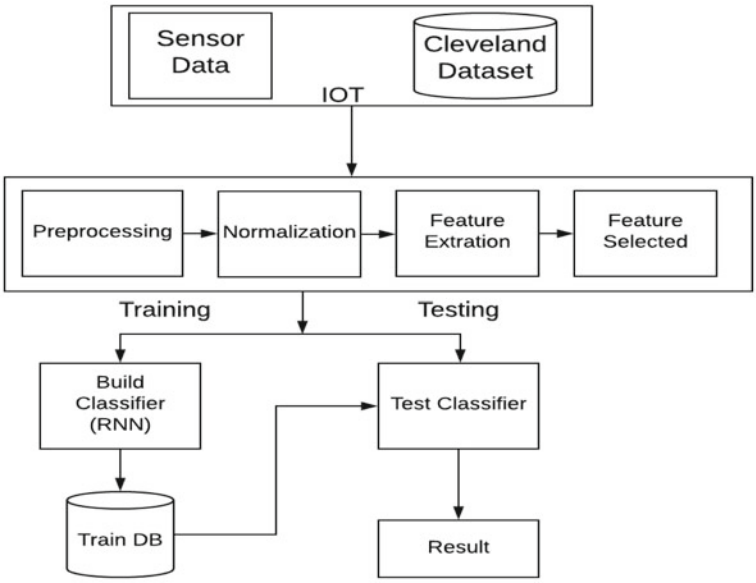

Fig. 1 Proposed system architecture

4 Algorithm Design

Algorithm 1: Proposed modified Recurrent Neural Network Algorithm (mRNN).

Input: Train_Feature set [], // Set of training dataset
Test_Feature set [] //Set of test dataset
Threshold denominator Th
Collection List cL

Output: Generate class label for all test instances based on classification results.
Step 1: Read all attributes from testing dataset using the below function

$$\text{Test\_Feature} = \sum_{j=1}^{n} (T[j])$$

Step 2: Read all attributes from the training dataset using the below function

$$\text{Train\_Feature} = \sum_{k=1}^{m} (T[k])$$

Step 3: Read total attributes from train instances using below function.
Step 4: Calculate the similarity index and generate weight for both feature set

Weight = classifyInstance(Train_Feature, Test_Feature)

Step 5: Verify with Th.
optimized_Instance_result = Weight >Th ?1: 0;
Add each optimized_instance to cL, when instances = null.
Step 6: Return cL.

5 Results and Discussions

Please It makes sense to interrelate quantitative assessment and clinical scores and could potentially address many challenges in decision-making. The generated training repository was applied with six machine learning models to establish patterns of common, questionable and dangerous activities. To test and rate the machine learning strategies, the threefold, fivefold and tenfold cross-validation model was employed using the behavior classification, training-database. Figure 2 below displays the threefold classification technique used on all parameters and explains all implementations' consistency.

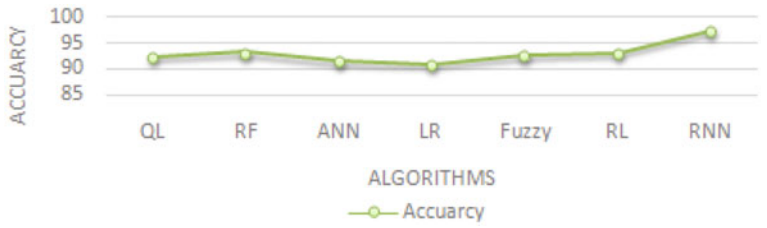

Fig. 2 Accuracy evaluation of various ML and RNN classification

Figure 2 shows the overall accuracy of all approaches, including the suggested RNN. Its accuracy percentage is 97.23%. The minimal accuracy of Linear Regression (LR) is 90.90%, which is greater than other approaches.

The True Positive (TP) of all algorithms, including the proposed RNN, is shown in Fig. 3. It has a TP ratio of roughly 97.40%. The algorithm Linear Regression (LR) has a minimum accuracy of 93.00% when compared to other methods.

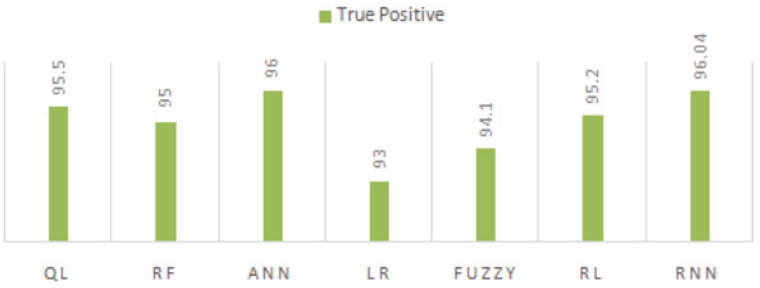

Fig. 3 True Positive evaluation of various ML and RNN classification

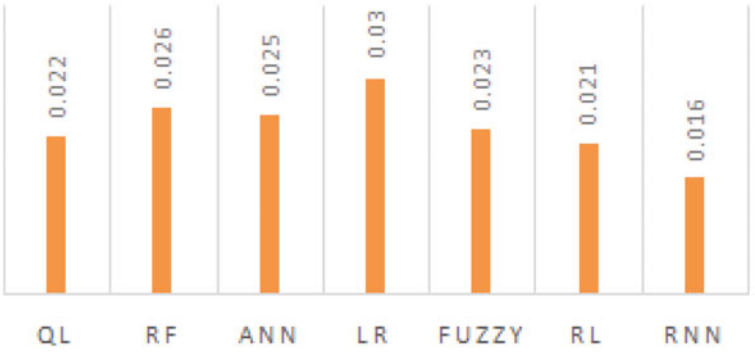

Fig. 4 False Positive evaluation of various ML and RNN classification

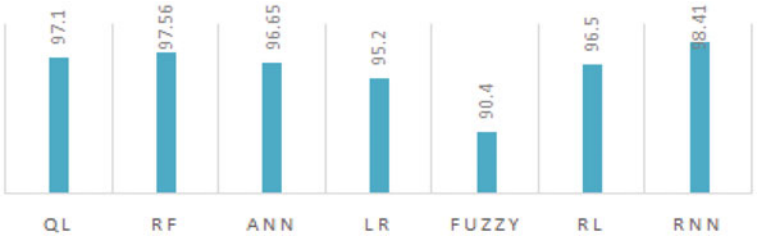

Fig. 5 True Negative evaluation of various ML and RNN classification

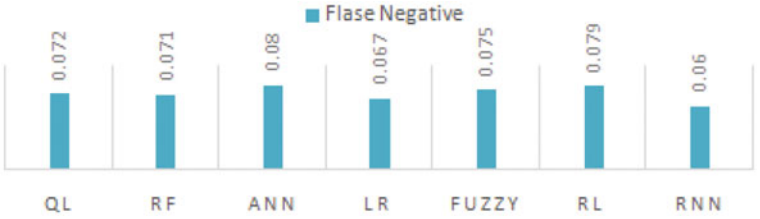

Fig. 6 False Negative evaluation of various ML and RNN classification

The total accuracy of all techniques, including the proposed RNN, is shown in Fig. 4. It has a 97.23% accuracy rate. Linear Regression (LR) has a minimum accuracy of 90.90%, which is higher than other methods.

Figure 5 shows the True Negative of all approaches, including the suggested RNN. It has an accuracy percentage of 98.41%. At 90.40%, the Fuzzy Logic algorithm has the lowest TN of all the approaches.

The false proportion of a system with several algorithms is shown in Fig. 6; LR as well as RNN both provides the lowest negative result ratio consistently.

The above Figs. 2, 3, 4, 5 and 6 improves the importance of different experimental research focusing on various statistical tests with seven distinct algorithms: Q-Learning, RF, Fuzzy logic, Naïve Bayes, Linear Regression, RL and Random Forest with Recommended Perceptron Algorithm. For data management, the mRNN classification algorithm was used during the classifier. The neural network was shown and debated for each model. Both uncertainty metrics show the system accuracy of properly classifying, wrongly classifying, recession, and device recall.

6 Conclusion

This research presented an optimal foundation for real-time prediction models. It may be used by those who have coronary artery disease. It is capable of monitoring and predicting both, unlike many other systems. ML algorithms will be used to forecast cardiovascular disease in the system's diagnostic process. The heart disease

dataset served as the basis for the prediction results. The gadget, on the other hand, is incredibly cost-effective; we utilized an enthusiastic pulse sensor and communicated the data to mobile devices using the Arduino suite microcontroller. To monitor the variations and sound an alert if the patient's heart rate rises over the normal range. We conducted experiments using both the tracking and diagnostic methods to demonstrate the system's effectiveness. We tested QL, Linear Regression, Random Forest, Naive Bayes, ANN, and Fuzzy Logic, among other supervised machine classification approaches. The implementation process has validated with and 89% detection accuracy of the defined approach has obtained for mRNN. To extract various feature extraction and selection with embedded deep learning techniques will be an interesting future work for the system. To develop various hybrid feature selection method and evaluate with deep learning classifiers on hybrid dataset will be the interesting task in future direction.

References

1. Sun Y, Lo FP-W, Lo B (2019) A deep learning approach on gender and age recognition using a single inertial sensor. In: 2019 IEEE 16th international conference on wearable and implantable body sensor networks (BSN). IEEE
2. Lo FP-W et al (2019) A novel vision-based approach for the dietary assessment using deep learning view synthesis. In: 2019 IEEE 16th international conference on wearable and implantable body sensor networks (BSN). IEEE
3. Uddin MZ, Hassan MM (2108) Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. IEEE Sens J 19(19):8413–8419
4. Ganesan M, Sivakumar N (2019) IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. In: Proceedings of international conference on systems computation automation and networking 2019. IEEE, 978-17281-1524
5. Majumder AKM, ElSaadany YA, Young R Jr, Ucci DR (2019) An energy efficient wearable smart IoT system to predict cardiac arrest. Hindawi Adv Human-Comput Interact Volume 2019, Article ID 1507465, p 21
6. Yu C et al (2020) Multitask learning for estimating multitype cardiac indices in MRI and CT based on adversarial reverse mapping. IEEE Trans Neural Network Learn Syst. https://doi.org/10.1109/TNNLS.2020.2984955
7. Ali SA et al (2020) 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 8:65947–65958. https://doi.org/10.1109/ACCESS.2020.2985646
8. Kumar PM, Lokesh S, Varatharajan R, Gokulnath C, Parthasarathy P (2018) Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Future Gener Comput Syst
9. Park M, Song Y, Lee J, Paek J (2016) Design and implementation of a smart chair system for IoT. IEEE
10. Kumar PM, Gandhi UD (2018) A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases. J Comput Electr Eng 222–235
11. Nashif S, Raihan MR, Islam MR, Imam MH (2018) Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. World J Eng Technol, 22 Nov 2018. https://doi.org/10.4236/wjet.2018.64057

12. Suresh P, Ananda Raj MD (2018) Study and analysis of prediction model for heart disease: an optimization approach using genetic algorithm. Int J Pure Appl Math 119(16):5323–5336, ISSN: 1314-3395 (on-line version)
13. Alotaibi FS (2019) Implementation of machine learning model to predict heart failure disease. (IJACSA) Int J Adv Comput Sci Appl 10(6)