Development of ML and IoT Enabled Disease Diagnosis Model for a Smart Healthcare System

Navita Mehra[1†] and Pooja Mittal[2††]

navitamehra55@gmail.com mpoojamdu@gmail.com

[1]Ph. D Scholar, DCSA [2]Assistant Professor, DCSA

M. D. University Rohtak, Haryana, India M. D. University, Rohtak, Haryana, India

Summary

The current progression in the Internet of Things (IoT) and Machine Learning (ML) based technologies converted the traditional healthcare system into a smart healthcare system. The incorporation of IoT and ML has changed the way of treating patients and offers lots of opportunities in the healthcare domain. In this view, this research article presents a new IoT and ML-based disease diagnosis model for the diagnosis of different diseases. In the proposed model, vital signs are collected via IoT-based smart medical devices, and the analysis is done by using different data mining techniques for detecting the possibility of risk in people's health status. Recommendations are made based on the results generated by different data mining techniques, for high-risk patients, an emergency alert will be generated to healthcare service providers and family members. Implementation of this model is done on Anaconda Jupyter notebook by using different Python libraries in it. The result states that among all data mining techniques, SVM achieved the highest accuracy of 0.897 on the same dataset for classification of Parkinson's disease.

Keywords:

Internet of Things, machine learning, smart healthcare, disease diagnosis, support vector machine.

1. Introduction

The recent advancement in communication and information technology has introduced a novel technology named the Internet of Things (IoT). IoT allows objects and people in the physical world to interact and communicate with each other to transfer valuable information for correct and real-time decision-making [1]. IoT becomes very fascinating among people due to its enormous number of applications in various fields like smart parking, smart city, smart farming, smart healthcare, smart industry, etc. Due to a great improvement in IoT-based wearable sensors and medical devices, healthcare becomes the most likely area of research nowadays. The traditional healthcare system is unable to provide sufficient services to everyone in need because of the incredibly growing population and the increasing number of diseases worldwide [2]. Despite that, the rising costs of healthcare and inappropriate medical services are not affordable by everyone. Therefore, to overcome all such problems, there is an urgent need for the transformation of the traditional healthcare system to a smart, intelligent, and affordable healthcare system. Recently, IoT is considered one of the most important paradigms to support patient monitoring. In the current scenario of a busy schedule of the individual and growing number of high-risk diseases like diabetes, heart attack, cancer, Covid-19, etc. Doctors also suggest using various IoT-based health monitoring devices like a smartwatch, pulse oximeter, blood sugar, weighing scale, smart pillboxes, wearable band, etc. [3]. Recent wireless sensors, technologies, and IoT play a major role in developing a smart healthcare monitoring system for elderly people living alone. These techniques are also helpful in developing applications that can easily move health-related data, which is further used for analysis. Such analytical information will be highly applicable in the diagnosis of a disease like Diabetes, Hypertension, Cancer, Heart Attack Dementia, Parkinson's, etc.[4]. Medical IoT devices offer remote monitoring of patients from anywhere and anytime. Medical devices attached to the patient continuously transfer vital health data to the medical team and doctors so that immediate action must be taken at any risk [5]. The combination of machine learning with IoT offers a great opportunity in the healthcare sector. An extensible amount of data generated through IoT-based medical devices must be analyzed for accurate diagnosis and prediction of disease. Different machine learning-based model has been designed for the diagnosis of various diseases like Diabetes, Hypertension, Cancer, Heart attack, etc. [6][7].

Major contribution of this work is as follows:

(i) This work proposes a framework/structure followed for accurate diagnosis of disease in IoT environments
(ii) This work also highlights the review of various data mining techniques utilized for classification of diseases in IoT environment.
(iii) This study adopts various performances metrics to check the performance of different classification methods on collected data patterns.
(iv) The work also focuses on the comparison of various classification methods for classification and diagnosis of disease in IoT environment.

The rest of the paper is organized as Section 2 details the related work done by different researchers in this area.

Manuscript received July 5, 2022
Manuscript revised July 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.7.1

Section 3 introduces the proposed model for the diagnosis of diseases with a detailed description of each technique and method associated with it. Section 4 details about experimental work and result discussion. At last, the conclusion and future work are described in Section 5.

2. Related Study

In recent years, a large number of researchers have worked towards IoT-based healthcare models and developed different disease diagnosis models for the diagnosis of different diseases like diabetes, chronic kidney, heart disease, Breast Cancer, hypertension, etc. This section mainly presents a set of studies done by different researchers towards the disease diagnosis models which are presented below.

Hosseinzadeh et al. [8] proposed a chronic kidney disease prediction model in an IoT environment by utilizing the concept of smart sensor technology. The proposed model specifies the severity of disease by utilizing the Glomerular filtration rate (GFR) method. The performance of the model was evaluated on multimedia data set using different classification algorithms like support vector machine (SVC), multilayer perceptron (MLP), Decision Tree (DT), and Naïve Bayes (NB). The results revealed that the decision tree has the highest performance and low execution time among all classification methods. Along with that, Jabeen et al. [9] also presented an efficient hybrid recommendation system for cardiovascular patients based on IoT. The presented model predicts cardiovascular disease into eight classes and suggests a dietary and physical plan to the patient depending upon age and gender. The performance of the model is improved by the utilization of sequential forward feature (SFF) with SVC, NB, Random Forest, and MLP classification algorithms. The presented system gives an accuracy of 98%. Another cloud and IoT integrated healthcare framework was proposed by Verma et al. [10] for the prediction of potential diseases with their level of severity. The proposed system deals with the general health issues faced by students living alone, suffering from obesity, waterborne or infectious diseases, heart-related diseases, hypertension, respiratory index and stress index, etc. SVC, Decision Tree, K- nearest neighbor (K-NN), and NB were utilized for classification purposes. Results revealed that the decision tree performed well for the patients suffering from infectious disease and K-NN outperformed for the heart disease dataset. 4-Fold cross-validation method was utilized for result optimization purposes. A fuzzy-based classification method was developed by Satpathy et al. [11] for detecting the pathological condition of heart disease patients. Regression analysis techniques were utilized for parametric reduction. The developed method was implemented on Field Programmable Gate Array (FPGA) and named Fuzzy-FPGA. The proposed model experimented on the UCI data set and revealed the highest accuracy compared to previous models with low execution time. Abdelaziz et al. [12] proposed a machine learning model by combining IoT and cloud computing technology for chronic kidney disease. Cloud computing features support the prediction of disease at any time and from anywhere. The proposed chronic kidney disease (CKD) model hybrid two intelligent techniques named linear regression and neural network. Linear regression is used as a feature selection technique to select critical features that influence the CKD. The proposed hybrid model obtained an accuracy of 98%, which is superior to most of the previous models. Next, a breast cancer-based diagnostic model was proposed by Memon et al. [13] in an IoT environment. The performance of the proposed system was improved by using the Recursive Feature Selection method (RFS). SVC with different kernel values was used for classification purposes and experimented on Wisconsin diagnostic breast cancer dataset. Results found that SVC linear kernel achieved higher accuracy compared to other kernels and the proposed system is reliable in all aspects of IoT healthcare. A student health monitoring system was proposed by Souri et al. [14] to detect biological and behavioral changes in student health based on vital signs. All data regarding student health was collected through IoT devices and analyzed by using different data mining techniques like MLP, decision tree, SVC, and random forest. The proposed model attained the highest accuracy of 99.1% by utilizing SVC as the classification algorithm. Next, Kaur and Chana[15] proposed a Cloud-based intelligent healthcare system called CBIHCS by using the concept of cloud computing that supports online monitoring of patient health data for the diagnosis of diabetes disease. Patient health data is collected through various health sensors. The proposed system used Principal Component Analysis (PCA) as a feature selection technique and K-NN and naïve Bayes as the classification techniques. Evaluated results found that K-NN achieved a higher accuracy of 92.59% than NB for the prediction of chronic disease in real-life scenarios. A three-tier IoT-based healthcare model was proposed by Kumar and Gandhi [16] for heart disease (HD) patients. The presented system mainly focused on the storage and processing of a large number of sensor data. The most significant clinical parameter affecting heart disease was identified by using receiver operating characteristic curve (ROC) analysis. Along with that, a cloud and IoT-based disease diagnosis and prediction system were also proposed by Kumar et al. [17] to offer online healthcare services to people. A fuzzy rule-based neural classifier was also proposed and applied to the UCI dataset as well as real health records of diabetic patients. The experiment results found that the proposed model attained higher accuracy in comparison with the existing model. An online medical decision support system named OMDSS for the prediction of CKD has been proposed by Arulanthu and Perumal [18]. The developed model worked with Logistic Regression (LR) as a classification model and Adaptive Moment Estimation (Adam) and adaptive learning optimization algorithms for tuning the classification algorithm. The experimentation resulted that the presented model provides an accuracy of 97.75% when applied to CKD dataset. A new Breast Cancer Diagnosis model was also presented by Zhang et al. [19] by utilizing the concept of feature extraction. Feature selection is done using K-means and SVC for classification purposes. The established model resulted in a classification accuracy of 97.38% with reduced CPU utilization. A breast cancer disease diagnosis model was developed by Kamel et al. [20]. The model performance was improved by utilizing Gray Wolf as a feature selection technique. Finally, classification was done by SVC, the result stated that the established model attained an accuracy of 100%. Soumiya and Sumitra [21] introduced a heart disease diagnosis model by utilizing Ant Colony Optimization (ACO) as a feature selection technique and Hybrid KNN as a classification technique. The proposed model achieved an accuracy of 99.2%. Mohammed et al. [22] Proposed multiagent feature selection algorithm (MAFT) by combining the Genetic Algorithm (GA), Adam Optimizer (AO), and minibatch gradient descent (MBGD), and classification was done utilizing Convolution neural network (CNN). Khan et al. [23] proposed an IoT-based heart disease diagnosis framework by utilizing a Modified Deep Convolution Neural Network (MDCNN) in combination with a Mapping-based cuttlefish optimization algorithm (MCFA). The result obtained yields that the proposed model achieved an average accuracy of 98.2%, which is higher than other existing models. Mansour et al. [24] proposed a smart healthcare system using IoT and ML for the diagnosis of diabetes and heart disease. The proposed model is named as CSO-LSTM model, in which the crow search optimization technique is used for feature selection and the LSTM model. The proposed model reveals that it achieves better accuracy than other existing models, that is, 96.16 % for heart disease and 97.26% for diabetes. Fitriyani et al. [25] presented a novel model for early prediction of hypertension and type 2 diabetes by focusing on individual's risk parameters. The presented DPM focused on outlier detection and data balancing concept. For the outlier detection and isolation forest (iForest) based method is used and for data balancing purpose the synthetic minority oversampling technique Tomek link (SMOTE Tomek) is utilized in combination with an ensemble approach to predict the diseases.

Table 1: Summary of previous work done by different researchers

Sr. No.	*Author Name*	*Disease*	*Data Set*	*Feature Selection Techniques*	*Data Mining Techniques*	*Accuracy Achieve*
1	Hosseinzadch et al., 2020	CKD	Sensor data	GFR	SVC, J48, MLP, NB	J48
2	Abdelaziz et al., 2019	CKD	UCI	LR (Linear Regression)	Fuzzy Classifier	98%
3	Arulanthel & Perumal, 2020	CKD	UCI	-	LR (Logistic Regression) tuned with Adaptive moment estimation, adaptive learning rate optimization algorithm.	97.75%
4	Zhang, 2013	CKD	UCI	SVC	K-means	97.38%
5	Elhonseny, 2019	CKD	UCI	ACO	DFS	95.00%
6	Kumar and Gandhi, 2017	CVD	Real time	Logistic Regression	ROC	Model identify most critical parameters
7	Jabeen et al., 2019	CVD	UCI	SFS	MLP, SVC, RF, NB	98%
8	Satpathy, 2019	CVD	UCI	Regression analysis	Fuzzy classifier	the model requires less computational time and high accuracy

9	Sowmiya and Sumitra	CVD	UCI	ACO	Hybrid K-NN	99.2%
10	Khan et al., 2020	CVD	UCI, sensor data	MCFA	MDCNN	98.2%
11	Mansour et al., 2021	CVD, Diabetes	UCI	CSO	LSTM	96.6%, 97.26%
12	Memon et al., 2019	Breast Cancer	UCI	REF	SVC with diff. Kernel	99.1% (Linear SVC kernel)
13	Kamel, 2019	Breast Cancer	UCI	GWA	SVC	100%
14	Kaur et al., 2014	Diabetes	Real-time	PCA	K-NN, NB	K-NN (92.59%)
15	Kumar et al., 2018	Diabetes	UCI	-	Fuzzy neural classifier	Achieved high accuracy as compared to others
16	Fitriyani et al., 2019	Diabetes, Hypertension	UCI	Filter techniques (IG)	IForest + SMOTE+ Ensemble classifier	High accuracy achieved
17	Ijaz et al., 2018	Diabetes, Hypertension	UCI, Real-time	IG	DBSCAN+SMOTE+ RF	83.64%
18	Mohammed et al., 2021	Parkinson	UCI	MAFT	CNN	93.7%
19	Balaji et al., 2021	Parkinson	Psyionet	Correlation	KNN, SVC, NB, EC,	SVC 98.4%

3. Proposed Work

This section is partitioned into two segments, the first segment will describe the system architecture followed by a smart healthcare system, and the second segment will describe the proposed model for disease diagnosis using ML.

3.1 System Architecture

The proposed model is composed of three different layers named as data collection layer, data processing layer, and alert generation layer. The complete model is shown in Fig.1

1. Data Collection Layer

Within the data collection layer, the collected data may be clinical or real-time data. Clinical data may be collected through the UCI repository or other health-related data-providing sites. Real-time data is only collected by using IoT devices or wearable sensors in real time. Data collected from smart sensors is stored on some device. Along with this, the gateway is used to communicate with the data repository by utilizing different wireless communication technologies including 5G, CDMA/GPRS, mobile n/W.

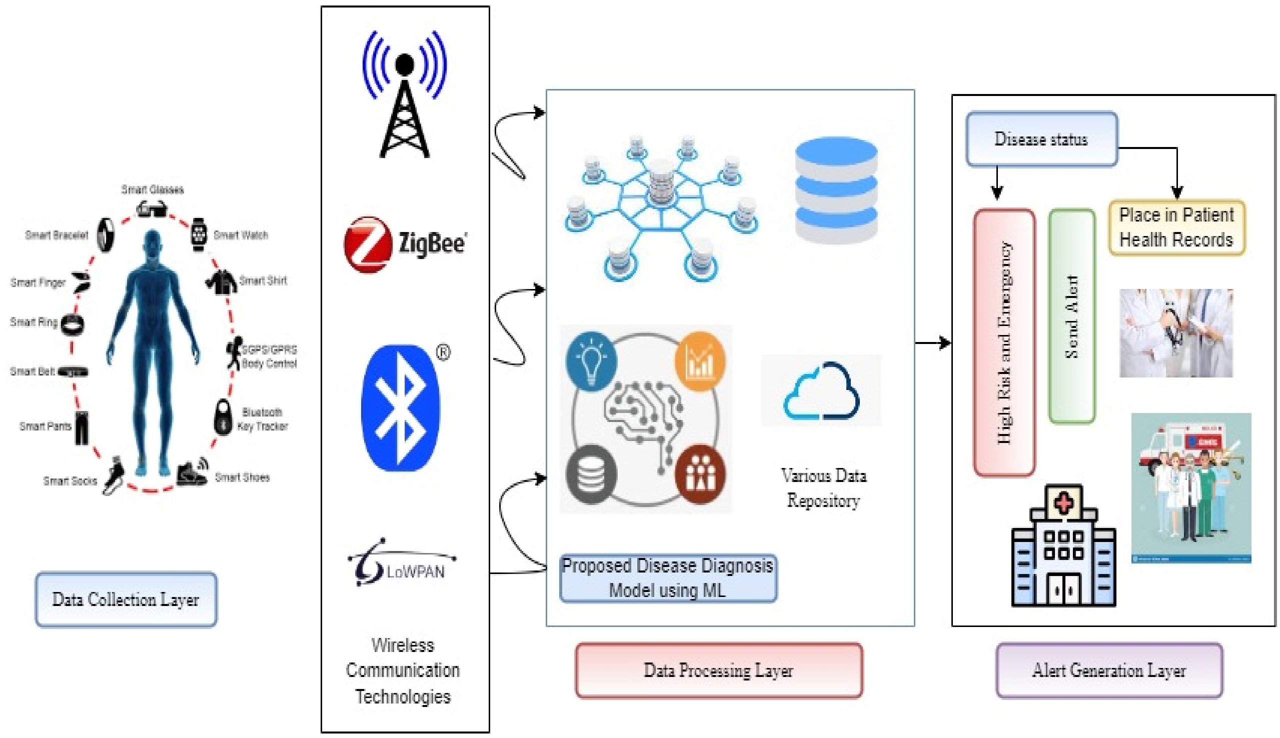

Fig.1. Proposed system architecture for disease diagnosis model

Algorithm 1: Data Collection Layer
Step 1: Clinical data gathering
Data gathered from the UCI repository
Step 2: Real-Time data gathering
Data collected from wearable sensors
ECG/EEG, Blood pressure monitoring device, Glucose monitoring, etc.
Collect physiological data which may be in structured and unstructured form
Step 3: Data Transmission
Data is transmitted to the data repository using wireless 5G, CDMA/GPRS technologies
Apply security mechanism

2. Data Processing Layer

In this layer, the data collected regarding patient health, which is stored in the service repository, is saved in cloud data centers. The data collected either from clinical sources or sensors or IoT devices have equal importance in the evaluation of our system. The data processing layer provides a preprocessing capability to handle structured as well as unstructured data.

For preprocessing, different preprocessing techniques such as data normalization, handling missing values, format data, and feature extraction and selection can be used. After that, the classification process starts, in which different classification algorithms can be used for the effective diagnosis of a disease, which gives information for making actions in the next layer.

Algorithm 2: Data Processing Layer

Step 1: Place the data retrieved from the data collection layer to System Storage or cloud storage
Step 2: Transfer data from data centers to a cloud service repository for effective decision making
Step 3: Apply data preprocessing, remove missing values, and normalize data in a range of 0 to 1
Step 4: Apply the feature selection method to select only relevant features
Step 5: Apply the proposed hybrid model for knowing the status of patient health

3. Alert Generation Layer

The third layer or the last layer is devoted to achieving patient health status and performing vital actions for improving patient health and storing data in the patient health record. After finding the final result by applying the classification algorithm alert generation layer, decide what to do next? If a patient's well-being status is "less critical" or a little serve condition, in that situation, the system will alert the patient or doctor and if the health status is the critical or "highly critical" then the system will automatically send an alert to the nearby hospital to offer emergency services like ambulance and urgent doctors to the patient's location otherwise the information stored in a patient health record.

Algorithm 3: Alert Generation Layer

The layer follows the following steps
Step 1: If (Patient wellbeing status== less critical)
Notify the doctors or physicians about the patient health status
Else if (Patient health status== highly critical)
Send an alert to a nearby hospital about the patient health status
Else
Save information to patient health record

3.2 Proposed Disease Diagnosis Model using ML

This section mainly describes all the methodological steps involved in the proposed diagnosis model, including preprocessing, feature selection, and classification methods and techniques, and can be considered as a subpart of the cloud layer and shown by Fig. 2.

1. Data Acquisition

This mainly refers to the procedure of collecting data from different data sources. Data may be collected from different sensors, IoT-based medical devices, and publically available data set repositories like UCI and Psyionet.

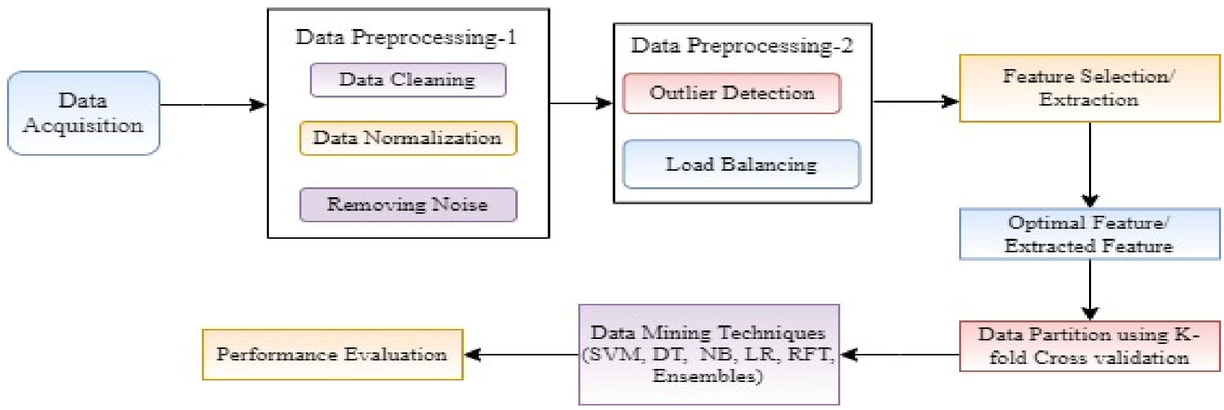

Fig.2. Working process of proposed disease prediction model using different DMT

2. Data Pre-processing -1

This is one of the most significant steps that must be followed during the classification task. In this model, the data preprocessing step is divided into phases. The first phase mainly deals with missing values, noise removal, and normalization. Because data is often collected from multiple sources and that data is not reliable to use directly for classification purposes. Collected data may contain missing values and unformatted data. Different methods are used for the preprocessing purpose such as finding missing values, min-max scalar, standard scalar, etc. The attribute which has a missing value is removed from the dataset. Standard scalar makes sure that every feature has a mean of 0 and variance of 1. Mix-max scalar moves the data so that all features have a range between 0 and 1. Therefore, before performing any classification task, data preprocessing must be done. The main objective behind data preprocessing is to improve classification performance and reduce computation time.

3. Data Preprocessing -2

Data preprocessing in phase-2 deals with inconsistent and imbalanced data. Data collected from different sources may face challenges like high dimensionality, data imbalance, and inconsistencies. Different techniques are associated to deal with such types of challenges and used by different researchers like Z-test, SMOTE, iForest, DBSCAN, etc. [25] [27]. Removing outliers from the training data will improve classification accuracy.

4. Feature Extraction/Selection

After preprocessing phase, feature extraction and selection techniques are followed to improve the performance of overall classification model.

5. Data Mining Techniques:

- K-Nearest Neighbor (K-NN): The working principle of K-NN is much similar to the nearest neighbor technique, then again, it looks for the nearest k instances to the unclassified instances. This approach classifies the unknown instance by using a distance parameter and then associates it with the known instance to make the final classification. The classification is done based on the class which has the highest number of neighbors depending upon the value of K. The proximity is calculated by calculating the Euclidean distance (ED) such that:

$$Euclíean\ Distance = \sqrt{\sum_{i=1}^{n}(ai - bi)^2} \qquad (1)$$

- Decision Tree (DT): DT is a classification method that builds a tree from the given input data. The tree could be utilized to extract a set of rules which help in finding the class and label of the attribute. DT utilizes sequential data to make dissimilar groups and to maximize the distance between each group. DT algorithm using the knowledge base concept, the tree is generated as an output from a different set of adjective values. The most common advantage associated with a decision tree is its high flexibility and comprehensibility. In this technique, the tree is built for each applied input based on the greatest entropy value.

$$\text{Entropy} = \sum_{j=1}^{m} \text{pij log2pij} \qquad (2)$$

- Radom Forest Tree (RFT): RFT is the most common data mining technique that often provides better results without adjusting its parameters. Due to its simplicity and high usability, it works well for both classification and regression purposes. This method also comes under the ensemble approach as it combines multiple decision trees. This method was mainly introduced to resolve the pruning problem that occurred in the decision tree approach. Besides searching for the most important feature while dividing a node, this algorithm looks for the best feature among a random set of attributes.
- Naïve Bayes (NB): NB classification algorithm depends on Bayes Theorem for its working procedure. This classification technique is highly suitable for problems that contain input data with multiple dimensions. By using the Bayes classifier, the probability of feature F belonging to disease class Ci is determined as follows;

$$\text{P(Ci|F)} = \frac{\text{P(Ci)P(F|Ci)}}{\text{P(F)}} \qquad (3)$$

Here, Ci|F is the posterior probability, P(Ci) is the prior probability, P(F|Ci) is the likelihood, and P(F) is evidence.

6. Performance Evaluation Measures: Performance evaluation measures mainly deal with all performance measuring methods through which the performance of the model can be evaluated. The most commonly used evaluation measures are sensitivity, specificity, accuracy, F1_Score, etc.

4. Experimental Result and Discussion

This section will detail the experiments done for the prediction of diseases using different classification methods.

4.1 Data Set

The global predicting system consists of several phases as shown in Fig.2. The first step consists of collecting data from different sources. For implementation purposes, the dataset named" Parkinson's dataset" was collected from the UCI repository [42]. This dataset contains 195 records with 24 attributes. This dataset was composed of voice measurements from 31 people using phones [43]. The chosen data set is already in cleaned form, and balanced and no outliers are detected. The heatmap representation of this data set is shown in Fig. 3.

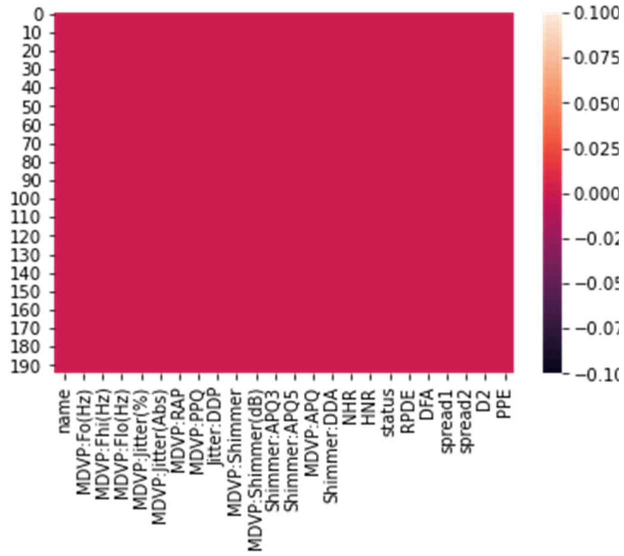

Fig. 3. Data heatmap representation

4.2 Implementation Tool and Techniques

The implementation of the model is done by using Anaconda jupyter notebook, and Python as a programming language. Different python libraries like pandas, NumPy, sns, matplotlib, sklearn metrics are used for different purposes. Pandas and NumPy libraries are used for numerical computation, matplotlib and Sns used for visualization purposes, sklearn used for using different machine learning techniques, and metrics used for performance evaluation measures. For prediction and analysis purposes, different supervised data mining techniques like k-nearest neighbor (K-NN), support vector machine (SVM), decision tree (DT), and random forest tree (RFT) are used. The dataset is used to analyze and predict whether the patient suffers from Parkinson's disease or not. Accuracy, precision, recall, F1_score, and confusion matrix are used as evaluation measures to measure the performance of different data mining techniques.

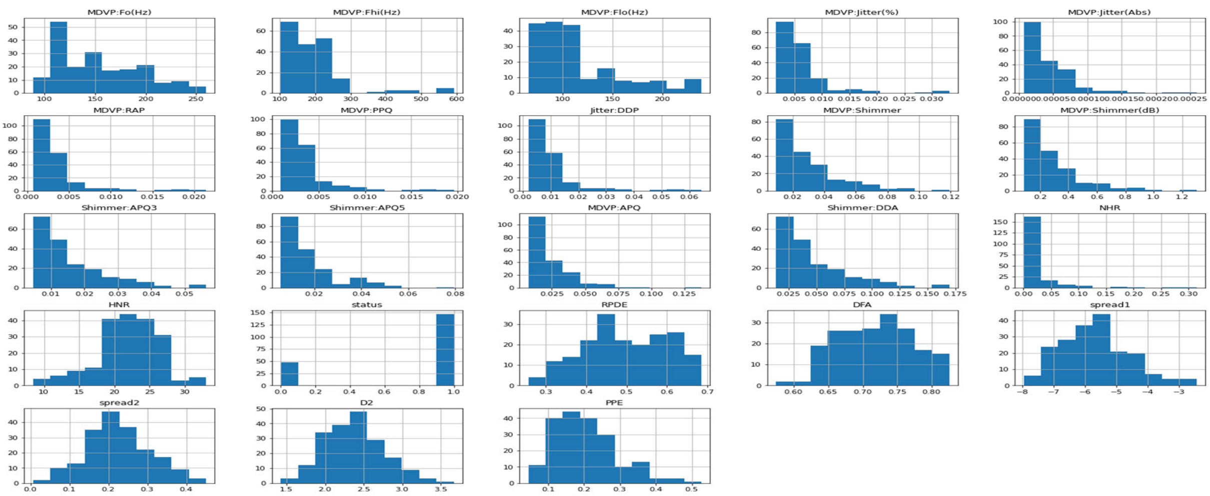

Fig.4. Histogram representation of dataset

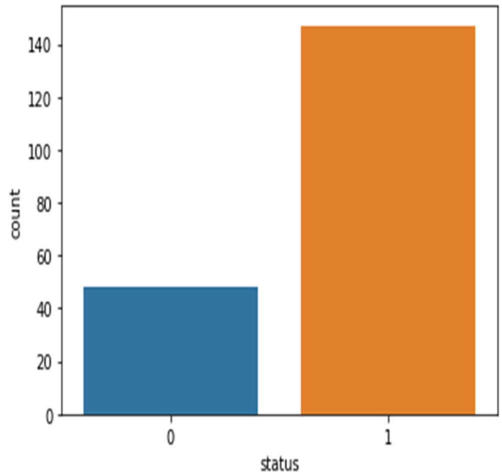

Fig. 5. Count plot representation of dataset by category/class

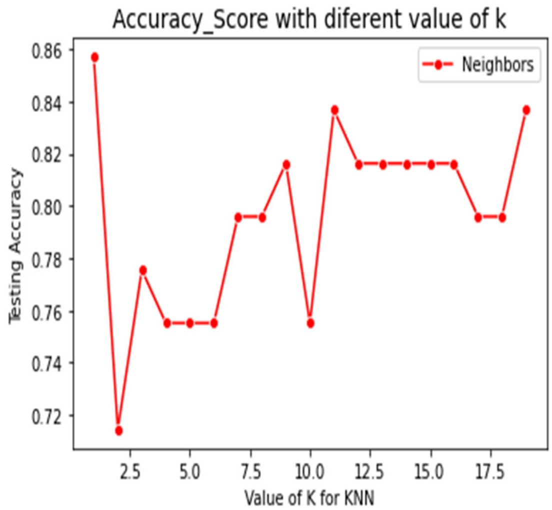

Fig.6. Accuracy plot of K-NN with different values of k

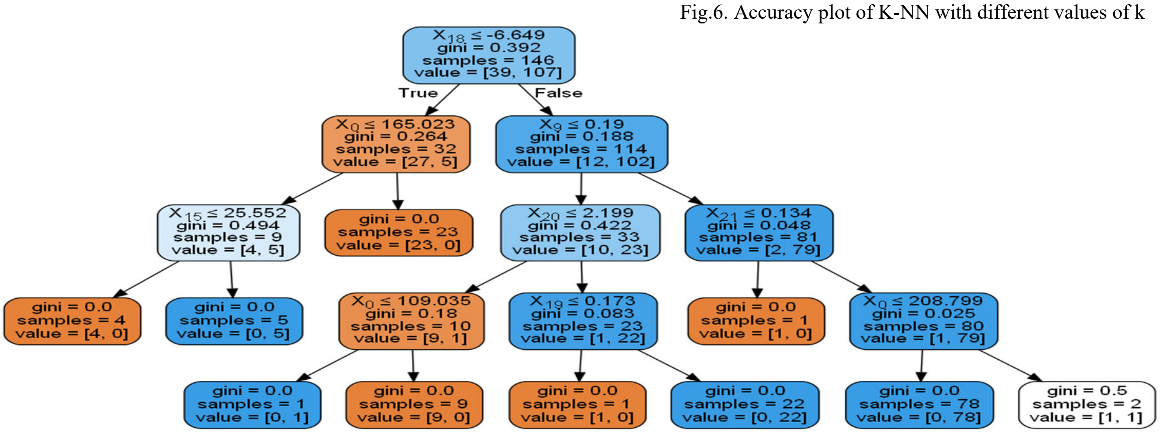

Fig. 7. Accuracy plot of a decision tree using Gini as splitting criteria

4.3 Experimental Result

Visualization of the dataset is shown in Fig. 4 and Fig. 5. A graphical view of the accuracy of K-NN is plotted in Fig. 6 with various number of k. A decision tree is represented in Fig.7. After the analysis of all data mining techniques, the result is represented in Table 2. The overall performance of all data mining techniques is represented in Fig. 8 through the confusion matrix and graphically in Fig. 9. The experimentation done on the Parkinson's disease dataset revealed that among all data mining techniques, SVM achieved the highest accuracy of 0.897 and after that K-NN achieved an accuracy of 0.877 and the decision tree achieved the lowest accuracy.

Table 2: Performance result of different DMT

DMT	Accuracy	Precision	Recall	F1_Score
RFT	0.836	0.85	0.84	0.84
SVM	0.897	0.91	0.9	0.88
KNN	0.877	0.91	0.82	0.84
DT	0.755	0.87	0.76	0.78

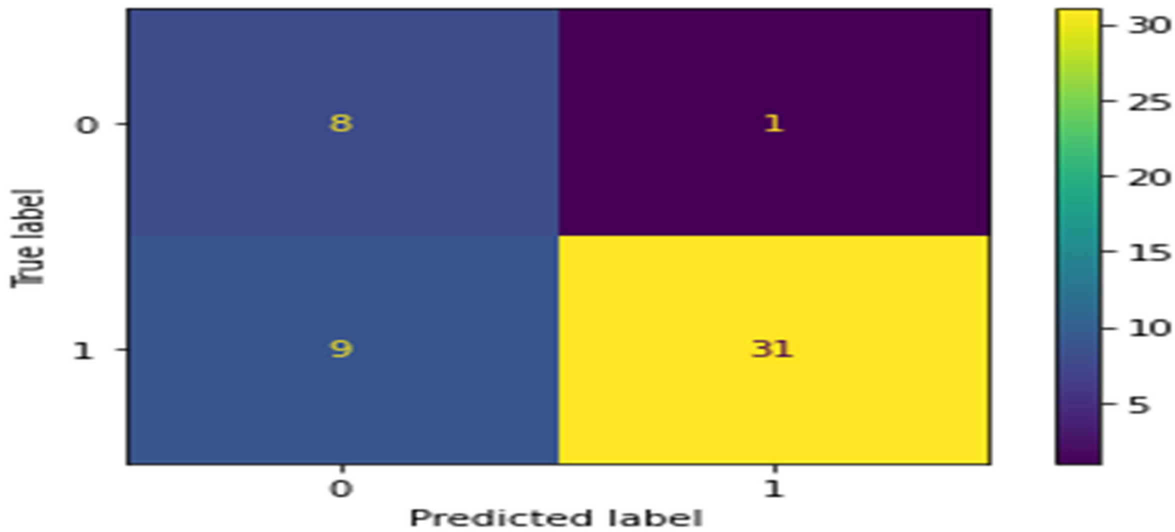

Fig. 8. Confusion matrix

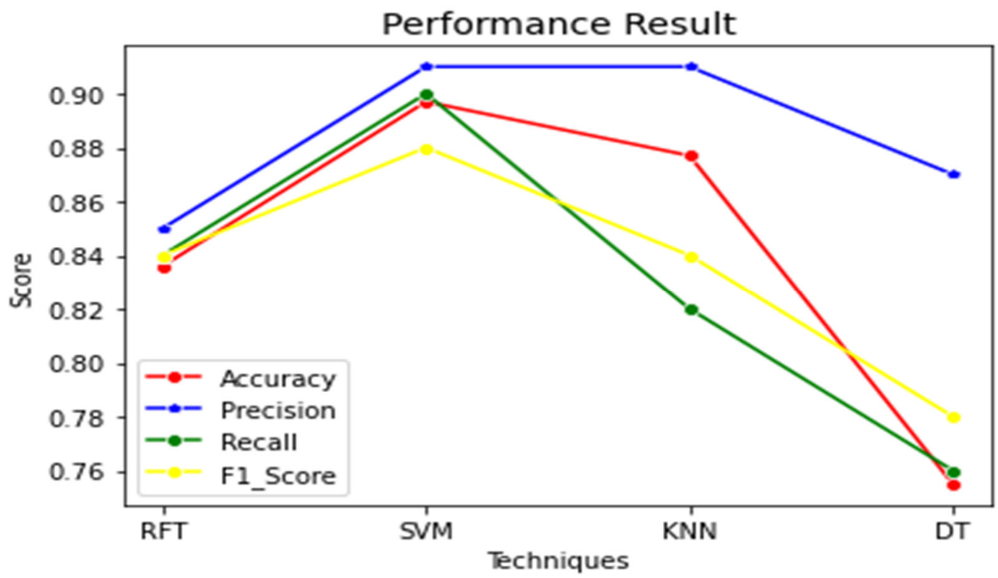

Fig. 9. Overall performance analysis results of different DMT

Fig. 9 shows the graphical representation of the score achieved by different data mining techniques. The graphical plot shows that among all techniques, SVM achieved the highest performance in terms of accuracy, precision, recall, and F1_score.

5. Conclusion and Future Work

This paper proposed an IoT-based healthcare disease diagnosis model using different data mining techniques. The proposed system monitors the vital parameters regarding patient health using smart IoT-based technologies. The modular structure of the proposed model consists of three layers named data collection layer, cloud layer, and recommendation layer. The data collection layer gathers data regarding patient health by using smart wearable biomedical sensors and IoT-based devices. The collected data is then sent to the cloud by using different communication technologies like 3G/LTE, 5GPRS, etc. Therefore, that it can be easily accessible from any remote location, and for further analysis, different data mining techniques were utilized to make an accurate prediction about the disease and the health status of the patient. Depending upon the disease status, if the patient is at high risk, then an emergency alert may be sent to the healthcare service providers so that all emergency services will be active and immediate support must be provided to the patient and their life can be saved along with that patient's health record must be updated. Currently, the model is implemented on Parkinson's disease dataset to predict whether the patient is suffering from Parkinson's or not using different data mining techniques. The result shows that among all data mining techniques, SVM achieved the highest accuracy of 0.897. In the future, this model can be implemented on the sensor-based dataset to make a prediction in real-time or on some secondary source dataset. Further improvements must be made to the existing model.

Acknowledgment
I am highly thankful to Dr. Pooja Mittal for their comments and support in the preparation of this manuscript.

References

[1] Perissinotto, C. M.., Stijacic Cenzer, I. , Covinsk, K. E.: *Loneliness In Older Persons: A Predictor of Functional Decline and Death*. Archives of Internal Medicine, vol. 172 (14), pp. 1078–1083 (2012)

[2] Joyia, G. J., Liaqat, R. M., Farooq, A., Rehman, S. : *Internet of Medical Things (IOMT): Applications, Benefits and Future Challenges In The Healthcare Domain*. Journal of Communication, vol. 12, no. 4, pp. 240–247 (2017)

[3] Vashist, S., Schneider, E., Luong, J., *Commercial Smartphone-Based Devices and Smart Applications for Personalized Healthcare Monitoring and Management*, Diagnostics(Basel), vol. 4, no. 3, pp. 104–128 (2014)

[4] Perez, D., Memeti, S. and Pllana, S.: A *Simulation Study of a Smart Living IoT Solution for Remote Elderly Care*. 2018 3rd International Conference Fog Mobile Edge Computing (FMEC 2018), pp. 227–232 (2018)

[5] Pramanik, P. K. D., Upadhyaya, B. K., Pal, S., Pal, T.: Internet of Things, Smart Sensors, and Pervasive Systems: Enabling Connected and Pervasive Healthcare. *Advances in Ubiquitous Sensing Applications for Healthcare*, pp. 1-58 (2019)

[6] Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., Sheth, A. P.: *Machine learning for Internet of Things Data Analysis: a survey.* Digital Communiation Networks, vol. 4, no. 3, pp. 161–175 (2018)

[7] Vanani I. R., Amirhosseini, M.: *IoT-Based Diseases Prediction and Diagnosis System for Healthcare: Internet of Things for Healthcare Technologies.* Studies in Big Data, vol 73. Springer, Singapore. pp. 21–48 (2021)

[8] Hosseinzadeh, M., Koohpayehzadeh, J., Bali, A. O.: *A Diagnostic Prediction Model for Chronic Kidney Disease in the Internet of Things Platform*. Journal of Multimedia Tools and Applications, Springer, vol. 80, pp. 16933-16950 (2020)

[9] Jabeen, F., Maqsood, M., Ghazanfar, M. A., Aadil, F., Khan, S., Khan, M. F. and Mehmood., I.: *An IoT based Efficient Hybrid Recommender System for Cardiovascular Disease*. Peer-to-Peer Networking and Applications, vol. 12, no. 5, pp. 1263–1276 (2019)

[10] Verma, P., Sood, S. K.: *Cloud-Centric IoT based Disease Diagnosis Healthcare Framework*. Journal of Parallel and Distributed Computing, vol. 116, pp. 27–38 (2018)

[11] Satpathy, S., Mohan, P., Das, S., Debbarma, S.*: A New Healthcare Diagnosis System using an IoT-based Fuzzy Classifier with FPGA*. Journal of Super Computing, Springer, vol. 76, pp. 5849-5869 (2019)

[12] Abdelaziz, A., Salama, A. S., Riad, A. M., Mahmoud, A. N.: *A Machine Learning Model for Predicting of Chronic Kidney Disease based Internet of Things and Cloud Computing in Smart Cities*. A book: Security in Smart Cities: Models, Applications and Challenges, Springer International Publishing (2019)

[13] Memon, M. H., Li, J. P., Haq, A. U., Memon, M. H., Zhou, W., Lacuesta, R.: Breast *Cancer Detection in the IoT Health Environment using Modified Recursive Feature Selection.*. Journal of Wireless Communication and Mobile Computing, pp. 93-114, (2019)

[14] Souri, A., Ghafour, M. Y., Ahmed, A. M., Safara, F., Yamini, A., Hoseyninezhad, M.: *A New Machine Learning-based Healthcare Monitoring Model for Student's Condition Diagnosis in the Internet of Things Environment.* Journal of Soft Computing, vol. 24, no. 22, pp. 17111–17121 (2020)

[15] Kaur, P. D., Chana, I.: *Cloud-based Intelligent System for Delivering Health Care as a Service.* Journal of Computing Methods and Programs in Biomedicine, vol. 113, no. 1, pp. 346–359 (2014)

[16] Kumar, P. M., Gandhi, U. D.: *A Novel Three-Tier Internet of Things Architecture with Machine Learning Algorithm for Early Detection of Heart Diseases.* Computers and Electrical Engineering, vol. 65, pp. 222–235 (2018)

[17] Kumar, P. M., Lokesh, S., Varatharajan, R., Chandra Babu, G. and Parthasarathy, P.: *Cloud and IoT based Disease Prediction and Diagnosis System for Healthcare using Fuzzy Neural Classifier*. Future Generation Computer Systems, vol. 86, pp. 527–534 (2018)

[18] Arulanthu, P., and Parumal, E.: *An Intelligent IoT with a Cloud-Centric Medical Decision Support System for Chronic Kidney Disease Prediction*. International Journal of Information Management System And Technology, pp. 1–13 (2020)

[19] Zheng, B., Yoon, S. W., Lam, S. S.: *Breast Cancer Diagnosis based on Feature Extraction using a Hybrid of K-means and Support Vector Machine Algorithms*. Expert System with Applications, vol. 41, issue 4, part 1, pp. 1476-1482 (2014)

[20] Kamel, S. R., Zadeh, R. Y., Kheirabadi, M.: *Improving the Performance of Support-Vector Machine by Selecting the Best Features by Gray Wolf Algorithm to Increase the Accuracy of Diagnosis of Breast Cancer*. Journal of Big Data, pp. 1-15 (2019)

[21] Sowmiya, C., Sumitra, P.: *A hybrid approach for mortality prediction for heart patients using ACO-HKNN*. Journal of Ambient Intelligence and Humanized Computing, vol. 12 (3) (2021)

[22] Mohammed, M. B., Elhoseny, M., Abdulkareem, K. H., Mostafa, S. A., Maashia, M. S.: *Multi-Agent Feature Selection and Hybrid Classification Model for Parkinson's Disease Diagnosis*. ACM Transactions on multimedia Computinng,

Communcations and Applications, vol. 17, no. 2, pp. 1-22 (2021)

[23] Khan, M. A.: *An IoT Framework for Heart Disease Prediction based on MDCNN Classifier*. IEEE Access, vol. 8, pp. 34717-34727 (2020)

[24] Mansour, R. F., Amraou, A. El., Nouaouri, I., Diaz, V. G., Gupta, D., Kumar, S.: *Artificial Intelligence and Internet of Things Enabled Disease Diagnosis Model for Smart Healthcare Systems*. IEEE Access, vol. 9, pp. 45137-4514 (2021)

[25] Fitriyani, N. L., M., Syafrudin, Alfian, G., Rhee, J.: *Development of Disease Prediction Model based on Ensemble Learning Approach for Diabetes and Hypertension*. IEEE Access, vol. 7, pp. 144777-144789 (2019)

[26] Balaji E., Brindha D., Elumalai, V. K., Umesh K.: *Data-Driven Gait Analysis for Diagnosis and Severity Rating of Parkinson's Disease.* Medical Engineering and Physics, vol. 91, pp. 54-64 (2021)

[27] Ijaz, M. F., Alfian, G., Syafrudin, M., Rhee, J.: *Hybrid Prediction Model for Type 2 Diabetes and Hypertension using DBSCAN-Based Outlier Detection, Synthetic Minority Over sampling Technique (SMOTE), and Random Forest*. Applied Science, vol. (8), 1325 (2018)

[28] Elhoseny, M., Shankar, K., Uthayakumar, J.: *Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease.* Scientific Reports 9, 9583 (2019)

[29] Manimurugan, S., Almutairi S., Aborokbah, M. M., Narmatha, C., Ganesan, S., Chilamkurti, N., Alzaheb, R. A., Almoamari, H.: *Two-Stage Classification Model for the Prediction of Heart Disease using IoMT and Artificial Intelligence*. Sensor, 22(2), 476 (2020)

[30] Wang, Y., Chu, Y. M., Thaljaoui, A., Khan, Y. A., Chammam, W., Abbas, S. Z.: *A Multi-Feature Hybrid Data Mining Technique for Human Emotions.* BioData Mining, vol. 14(21) (2021)

[31] Zhao, X., Li..B. Daoliang, Yang, B., Chand, H., Yanga, X., Yua, C., Liu, S.: *A Two-Stage Feature Selection Method with Its Application*. Computer and Electrical Engineering, pp. 114-125, (2015)

[32] Jeon H., Oh, S.: *Hybrid-Recursive Feature Elimination for Efficient Feature Selection*. Applied Sciences, 10, 3211 (2020)

[33] Ferdousi, R., Anwar Hossain, M., Saddik, A. El.: *Early-Stage Risk Prediction of Non-communicable Disease using Machine Learning in Health CPS.* IEEE Access, vol. 9 (2021)

[34] Sarkar, I. H.: Machine Learning: *Algorithms, Real-World Applications and Research Directions*. SN Computer Science, 2:160 (2021)

[35] Jie, M. Christian, D., Johannes, F.: *Machine Learning for the diagnosis of Parkinson's disease: A Review of Literature*, Frontiers in Aging Neuroscience, vol. 13 (2021)

[36] Xie, J., Lei, J., Xie, W., Shi, Y., Liu, X.: *Two-Stage Hybrid Feature Selection Algorithms for Diagnosing Erythematous-squamous Diseases.* Health Information Science & Systems, (2013)

[37] Akhil, J., Deekshatulu, J., Chandra, P.: *Intelligent Heart Disease Prediction System using Random Forest and Evolutionary Approach*. Journal of Network and Innovative Computing, vol 4, (2016)

[38] Samagarh, J. S., Singh, D.*: Machine Learning-Based Hybrid Model for Heart Disease Prediction*. Annals of R.S.C.B., vol. 25, no. 4, pp. 2199 – 2210 (2021)

[39] Nawaz, M. S., Shoaib, B., Ashraf, M. A.: *Intelligent Cardiovascular Disease Prediction Empowered with Gradient Descent Optimization.* Heliyon 7,(2021)

[40] Alam, Md. Z., Rahman, M. S., Rahman, M. S.: *A Random Forest-Based Predictor for Medical Data Classification using Feature Ranking*. Informatics in Medicine Unlocked, vol. 15 (2019)

[41] Krishnaveni1, N. and Radha, V.: Feature Selection Algorithms for Data Mining Classification: A survey. *Indian Journal of Science and Technology*, vol 12(6), pp. 1-11 (2019)

[42] UCI Machine Learning Repository: Parkinsons Data Set

[43] Little, M. A., Sharry, P. E. Mc, Hunter, E. J., Ramig, L. O.: Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's disease. IEEE Trans Biomed Engineering, vol. 56(4) (2009)

Ms. Navita has completed her M.Tech from GJU S&T University. She is currently pursuing Ph.D in Computer Science at Department of Computer Science & Applications, M.D. University, Rohtak. Her main research area includes Internet of Things (IoT) and Data Mining.

Dr. Pooja Mittal is currently working as Assistant Professor at Department of Computer Science & Applications, M.D.University, Rohtak, India. She obtained her Ph.D from Maharshi Dayanand University. Her area of resaerch and specializtaion include Data Mining, Data Warehousing and Computer Science. She has published more than 50 resaerch papers in renowed international and SCI Journals and attended more than 30 conferences.