RESEARCH ARTICLE OPEN ACCESS

A Comparative Study of Predictive Modelling of Diabetes Using Machine Learning Algorithms

Vishal Kumar[1], Varun Bansal[2], Naman Saini[3], Arun Saini[4]

[1](Computer Science Engineering, Shobhit University, Gangoh
Email: Vishal.kumar@shobhituniversity.ac.in)

[2] (Computer Science Engineering, Shobhit University, Gangoh
Email : Varun.bansal@shobhituniversity.ac.in)

[3](Computer Science Engineering, Shobhit University, Gangoh
Email : Naman.saini@shobhituniversity.ac.in)

[4] (Computer Science Engineering, Shobhit University, Gangoh
Email : Arun.saini@shobhituniversity.ac.in)

--************************----------------------------------

Abstract:

Diabetes is a condition that occurs when your blood sugar (glucose) levels are too high. It occurs when the body cannot make enough insulin or cannot use insulin effectively, which then enters the bloodstream. Insulin is a hormone that helps the body to let glucose enter our cells and use it as fuel. It is secreted by the pancreas. Diabetes occurs when you do not make enough insulin or do not make it at all, or when your body is not responding properly to the effects of insulin. Diabetes affects people of all ages. Most forms of diabetes are long-term (lifelong), and all forms can be controlled with medicines and/or lifestyle changes. Glucose (sugar) comes mainly from carbohydrates in your food and drinks. It is your body's energy source. Your blood carries glucose to all the cells in your body for energy. The main types of diabetes are: type 1 diabetes, type 2 diabetes, and gestational diabetes.[14] It occurs during pregnancy. If left untreated, diabetes can lead to a number of serious health problems, such as heart disease, kidney disease, vision loss, and nerve damage. This study demonstrates how machine learning can be used to predict the likelihood of developing diabetes. Machine learning has a significant impact on the diagnosis and early detection of diabetes.

***Keywords* — Diabetes, Type 1 Diabetes, Type 2 Diabetes, gestational diabetes, Insulin, Glucose, Machine Learning.**

--************************----------------------------------

I. INTRODUCTION

To understand diabetes, it is important to first know what happens in a normal, healthy body after eating. When we eat roti, rice, fruits or any carbohydrate-rich food, the body breaks them down and converts them into glucose. Glucose is the main source of energy for the cells of our body. Our brain also needs glucose to function properly. The glucose that is not used immediately is stored in the liver so that it can be used later. In this entire process, a hormone called insulin is very important, which is produced by the pancreas. Insulin is actually like a key that opens the doors of the body's cells and allows glucose to enter. If the body is unable to produce insulin properly or is unable to use it properly, the amount of glucose in the blood starts increasing, which we call diabetes. Its effect gradually affects the entire body - frequent thirst, fatigue, weight loss, frequent infections, and effects on eyesight are some common symptoms. If not identified in time, this disease can become serious and cause damage to the heart, kidneys or eyes.

The initial symptoms of diabetes may start slowly, but if left untreated, they can become severe. Some common symptoms are as follows: When the sugar level in the body increases, the person starts urinating frequently. Along with this, constant thirst and frequent hunger are also common symptoms. Due to lack of energy in the body, the person feels tired or sleepy. Sometimes weight starts decreasing rapidly without any specific reason. Things start appearing blurry. Apart from this, the person may also experience mood swings, difficulty in concentrating and frequent infections or delayed healing of wounds. All these signs can indicate that the sugar level in the body is not in balance.

Type 1 diabetes is a condition in which the body's immune system mistakenly attacks the beta cells in the pancreas. These cells make insulin, which helps the body convert glucose into energy. But when these cells are destroyed, the body stops making insulin, leading to a build-up of glucose in the bloodstream — a condition known as hyperglycemia. [1] Type 1 diabetes is most commonly found in people under the age of 30, but it can occur at any age. Symptoms come on suddenly and rapidly, and patients often find out when their condition worsens enough to require hospitalization. Because the body cannot make insulin on its own, external insulin administration becomes necessary.

In the United States, 304,000 children and adolescents under the age of 20 and 1.7 million adults have type 1 diabetes, according to the American Heart Association report. And in 2021, an estimated 9.5 million people worldwide were living with type 1 diabetes, including 800,000 people aged 15 to 19 in 2019, according to Nature and this number is projected to reach 2.1 million in low-income countries in 2025[2].

According to the American Heart Association report, 30.3 million Americans had diabetes in 2015, which is 9.4% of the US population.[3] By 2015, the global prevalence of diabetes (DM) had increased from 30.3 million in 2005 to 435 million. Heart disease (HF) affects at least 26 million people worldwide [4] and its prevalence is increasing; it is very important that we keep getting periodic medical checkups and do not take the body's signals lightly. If diabetes is not treated, it can lead to heart disease, kidney failure and other serious problems. Diabetes and heart disease often occur together.

II. LITERATURE REVIEW

At present, diabetes and heart disease are two major and rapidly growing health problems, which have affected millions of people in the world. In recent years, solutions based on machine learning have emerged. These techniques make the identification and prediction of diseases accurate, fast and quick.

1. A Classification Algorithm-Based Hybrid Diabetes Prediction Model

Patel et al. (2017)

Abstract: In the proposed method, a hierarchical system is used to combine two or more classifiers even working as an ensemble and deals with them at higher level. In our case, we did it in two steps, this implies we first trained a Decision Tree and Logistic Regression model followed by passing the output of these models into Neural Network. This is combined with another objective (to increase the overall accuracy of all previous classifiers) which is learned through DNN. [5] To validate our hypothesis, we have used PIMA Indian diabetes database as benchmark problem. We set new records with classification accuracies exceeding 83% (see Table I) on the proposed model while remaining competitive to other state of the art methods in literature.

Algorithm	**Accuracy (%)**	**Remarks**
Support Vector Machine (SVM)	83.1%	Highest accuracy among all
Random Forest (RF)	80.5%	Used as an additional comparison
Naive Bayes (NB)	78.5%	Performed better than Decision Tree
Decision Tree (DT)	70.7%	Lowest accuracy among compared models

Table I. (Highest accuracy in SVM)

In these experiments we have split the PID into 30% test and 70% training portion. We could achieve 65% accuracy using decision tree; the regression model achieved better accuracy, over 80%. However, by combining the output of previous models and trained an Artificial Neural Network (ANN), the overall accuracy jumped to 83%. (see Table II).

Method	Overall Accuracy	Label	Precision	Recall
Decision Tree (ID3)	65.84%	Positive (Diabetes)	65.84	33.33
		Negative (Normal)	84.7	95.53
Logistic Regression	80.71%	Positive (Diabetes)	51.97	81.48
		Negative (Normal)	94.38	80.51
Our proposed Ensemble model: (Artificial Neural Network + Logistic Regression + Decision Tree)	83.08%	Positive (Diabetes)	25.00	82.35
		Negative (Normal)	98.57	83.13

Table II (Overall accuracy in ANN+ LR+ DT)

The experiment results suggest that the proposed model can effectively combine the Decision Tree and Regression model and improve the overall accuracy. In comparison to other methods in the literature, our proposed model also demonstrated improved performance.

In this paper authors designed to perform a review of comparison of KNN, Naive Bayes, Decision Tree and SVM techniques on PIMA Indian Diabetes Dataset. [9] SVM gave the highest accuracy of around 83.1%, Random Forest achieved around 80.5% accuracy when used. Decision Tree achieved around 70.7% accuracy and Naive Bayes achieved around 78.5% accuracy.[13]

2. Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning. Saxena et al. (2019)

Abstract: Cardiovascular diseases (CVD), especially heart attacks, are one of the major reasons for death throughout the globe, emphasizing that early detection and intervention strategies should be improved. In this paper, we compare the Random Forest (RF) and Logistic Regression (LR) techniques to identify the more powerful one for predicting risk of heart attack from a varying dataset containing heterogeneous patient samples.[12] It is addressed to unveil a set of low signals patterns and risk factor correlations traditional methods might miss; however, the same comparative approach is taken to evaluate accuracy and performance of both models. One of the most important contributions of this study is that it attempts to make some top algorithms interpretable, as there is a dearth of such interpretability in previous works. Moreover, the problem of dataset imbalance (common in medical data) is tested and solutions to enhance model reliability in practical settings are suggested.[6] These results fit into the current debate on how to optimize machine learning in healthcare, and support a customized choice which trades-off predictive power versus interpretability. This study therefore sought to compare the strengths and weaknesses of RF and LR in predicting heart attacks; findings that similar assessments might have important clinical implications for clinicians, researchers and thereby improving decision-making in cardiovascular care and interventions. [15]

Algorithm	Accuracy (%)	Remarks
Decision Tree (DT)	93.19%	Highest accuracy in the study
Support Vector Machine (SVM)	92.30%	Very close to Decision Tree in performance
Random Forest (RF)	85–86%	Good performance, supports ensemble methods
Logistic Regression	Not specified	Used, but exact accuracy not mentioned
Naive Bayes (NB)	Not specified	Used, but exact accuracy not mentioned

Table III. (Highest accuracy in DT)

The implementation of Pearson correlation analysis reduced the initial 13 features to 7, focusing on those with the highest correlation coefficients, as shown in Table 3. For instance, features like cp (chest pain type) and exng (exercise induced angina) demonstrated the highest correlations, with coefficients of 0.432 and 0.436, respectively. The results of this study indicate that both the Random Forest (RF) and Logistic Regression (LR) models achieved comparable test accuracy rates of 86%. Despite the higher training accuracy of RF 100%, its test performance did not exceed that of LR, suggesting the possibility of overfitting. This section delves into the potential reasons for this similarity, examines additional performance metrics, and evaluates the models' real-world applicability.

Table IV: Pearson coefficient analysis for selected features.

Feature	Pearson Coefficient
cp	0.432080
thalachh	0.419955
exng	0.435601
oldpeak	0.429146
slp	0.343940
caa	0.408992
thall	0.343106

Table IV. (Highest accuracy in DT)

Several factors may explain the comparable performance of RF and LR models:

1. **Feature Selection Overlap:** Pearson's correlation analysis identified features (cp, exng, etc.) that were highly predictive for both models. This shared feature space may have led to similar predictive outcomes.
2. **Dataset Characteristics:** With only 303 samples, the limited size of the dataset may have restricted RFs ability to take advantage of its complexity advantages over LR. Furthermore, the linear relationships in the dataset between certain features and the target variable may have favored the simpler modeling approach of LR.
3. **Class Balance:** Although the dataset exhibited some imbalance, both models likely handled it effectively due to their inherent robustness and the moderate size of the dataset.
4. **Noise and Missing Data:** The intentional decision not to impute missing values may have influenced both models similarly, as they had to learn patterns from incomplete data.

In this research, techniques like Logistic Regression, Decision Tree, Random Forest, Naive Bayes and SVM have been used on Cleveland Clinic or Framingham heart disease data.[11] Decision Tree has shown an accuracy of up to 93.19%, while SVM has shown an accuracy of 92.30% and Random Forest has shown an accuracy of about 85–86%.

3. Deep Cardio Sound—An Ensembled Deep Learning Model for Heart Sound Multi-Labelling

Uddin et al. (2020)

Abstract— Heart sound diagnosis and classification play an essential role in detecting cardiovascular disorders, especially when the remote diagnosis becomes standard clinical practice.[10] Most of the current work is designed for single category based heard sound classification tasks. To further extend the landscape of the automatic heart sound diagnosis landscape, this work proposes a deep multilabel learning model that can automatically annotate heart sound recordings with labels from different label groups, including murmur's timing, pitch, grading, quality, and shape.[7] Our experiment results show that the proposed method has achieved outstanding performance on the holdout data for the multi-labelling task with sensitivity=0.990, specificity=0.999, F1=0.990 at the segments level, and an overall accuracy=0.969 at the patient's recording level.

Metric	Value	Remarks
Accuracy	96.9%	Overall classification performance
F1 Score	0.99	Balance between precision and recall
Sensitivity	0.99	True positive rate (recall)
Specificity	0.999	True negative rate
Model Used	CNN + RNN	Deep learning architecture (ensemble)
Task Type	Multi-label classification	For heart sound diagnosis

Table V. (Highest accuracy in CNN+RNN)

Since there are no standard test benches to which we can directly compare this work, we carried the performance comparison between our models and the existing results focused on the heart sound classification task (single label). Table 9 shows that even though our work deals with many more labels, its performance outperforms those designed for the simple classifications with the best sensitivity=0.990, specificity=0.999 and the top 2 for the accuracy. That said, there are still a few limitations of this work.

Secondly, this work has only focused on labelling the systolic murmurs due to the inadequate data samples for the diastolic periods. We do not think this is a significant challenge. As with our network architecture, each label group is separated from the rest. Given more samples, extending this work to support diastolic murmur labelling is straightforward without affecting the existing model's performance for the systolic murmurs.

In this paper authors designed to perform a review of Deep Learning based model (CNN + RNN) to better understand the multi-label classification of heart sound data. This model has achieved an overall accuracy of around 96.9% (see table 5), F1-score = 0.99, sensitivity = 0.99 and specificity = 0.999, which is very good.

4. Diabetes Prediction Model Based on GA-XGBoost, and Stacking Ensemble Algorithm Khan et al. (2021)

Abstract: Diabetes, as an incurable lifelong chronic disease, has profound and far-reaching effects on patients. Given this, early intervention is particularly crucial, as it can not only significantly improve the prognosis of patients but also provide valuable reference information for clinical treatment. This study selected the BRFSS (Behavioral Risk Factor Surveillance System) dataset, which is publicly available on the Kaggle platform, as the research object, aiming to provide a scientific basis for the early diagnosis and treatment of diabetes through advanced machine learning techniques. Firstly, the dataset was balanced using various sampling methods; secondly, a Stacking model based on GA-XGBoost (XGBoost model optimized by genetic algorithm) was constructed for the risk prediction of diabetes; finally, the interpretability of the model was deeply analyzed using Shapley values. The results show: (1) Random oversampling, ADASYN, SMOTE, and SMOTEENN were used for data balance processing, among which SMOTEENN showed better efficiency and effect in dealing with data imbalance. (2) The GA-XGBoost model optimized the hyperparameters of the XGBoost model through a genetic algorithm to improve the model's predictive accuracy (see table 6). Combined with the better-performing LightGBM model and random forest model, a two-layer Stacking model was constructed. This model not only outperforms single machine learning models in predictive effect but also provides a new idea and method in the field of model integration. (3) Shapley value analysis identified features that have a significant impact on the prediction of diabetes, such as age and body mass index. This analysis not only enhances the transparency of the model but also provides more precise treatment decision support for doctors and patients. In summary, this study has not only improved the accuracy of predicting the risk of diabetes by adopting advanced machine learning techniques and model integration strategies but also provided a powerful tool for the early diagnosis and personalized treatment of diabetes.[8]

Component	**Details**	**Remarks**
Dataset Used	BRFSS (Behavioral Risk Factor Surveillance System)	Large-scale public health dataset
Model Type	GA-XGBoost + Stacking Ensemble	Hybrid of feature selection and ensemble classification
Optimization Technique	Genetic Algorithm (GA)	Used for feature selection
Data Imbalance Handling	SMOTE + ENN + other oversampling techniques	Improves balance in class distribution
Performance	Higher than traditional ML models	Exact accuracy not given, but noted as superior
Key Features Identified	Age, BMI	High interpretability due to XGBoost

Remarks	Focus on both accuracy and model explainability	Suitable for real-world health data applications

Table VI. (GA-XGBoost + Stacking Ensemble)

In this paper authors designed to perform a review of uses the Behavioural Risk Factor Surveillance System dataset to optimize the Extreme Gradient Boosting model with a Genetic Algorithm and then build a stacking ensemble model. The model also uses Synthetic minority over-sampling technique (SMOTE) [14] and edited nearest neighbour (ENN) and other oversampling techniques to correct for data imbalance. It provides higher accuracy and interpretability than other machine learning models, and particularly well explains feature importance such as age and BMI.

III. RESULT

This study explored diverse machine learning and deep learning techniques across multiple medical prediction tasks, emphasizing model optimization, feature selection, and class imbalance handling. In diabetes prediction, combining traditional models—Decision Tree and Logistic Regression—with an Artificial Neural Network led to an ensemble model that improved overall accuracy to 83.08%, with significantly higher recall for diabetic cases (82.35%). A separate investigation using the XGBoost model revealed the limitations of training on imbalanced data, where high accuracy (86.52%) was paired with poor recall (17.34%). This was effectively addressed by applying sampling techniques, particularly the SMOTEENN method, which elevated the model's performance to 93.25% accuracy and 94.73% recall, demonstrating the value of data preprocessing for minority class enhancement.

In cardiovascular research, Pearson correlation analysis was used to identify the most predictive features for heart disease diagnosis, leading to comparable results between Random Forest and Logistic Regression models (both with 86% accuracy), highlighting the benefits of model simplicity when supported by strong feature selection. Further, a deep learning architecture integrating CNN and RNN was introduced for multi-label heart sound classification. Despite challenges such as dependence on pre-segmentation and limited data for certain murmur types, the model achieved exceptional performance metrics—sensitivity of 0.990, specificity of 0.999, and 96.9% accuracy. These findings underline the importance of tailored model design, preprocessing, and ensemble learning to achieve robust, generalizable outcomes in medical AI applications.

After reviewing all this research, we can say that Support Vector Machine (SVM): It is mostly used for diabetes detection. This technique works very well when the data is limited and clean. Its approximate accuracy is 80% to 85%, which makes it a good and reliable option.

Random Forest: This is a technique that can work by combining very small decision trees. It is used a lot in the detection of heart diseases. It can mostly give information about accuracy of 85% to 90% and can work well even on complex data.

Naive Bayes: This is a simple and fast model that is mostly used in early detection or screening. It is very good and works even on small data sets, but its accuracy is only around 70% to 75% which is a bit low.

Deep Learning (CNN): When the data is complex - such as voice, image or many types of health record data is complex, then CNN based deep learning models are used. Their accuracy mostly exceeds 90% and they are used for large data sets (see Table 7).

Logistic Regression: This is one of the most basic techniques, mostly used for initial risk assessment. Its accuracy can be measured around 75% to 80% and it is useful only when explainability is important and needs to be well explained.

Technique	Use Case	Accuracy (Approx.)
Support Vector Machine	Diabetes Prediction	80–85%
Random Forest	Cardiovascular Disease Detection	85–90%
Naive Bayes	Early Screening	70–75%

Deep Learning (CNN)	Multi-modal Health Data	90%+
Logistic Regression	Baseline Risk Assessment	75–80%

Table VII. (Highest accuracy)

Machine Learning algorithms such as Random Forest, SVM, XGBoost, and Neural Networks are accurate in disease detection and for Feature Selection (such as PCA, Genetic Algorithms) is good at accurate disease detection. However Hybrid models, which use more than one technique, are more robust and scalable.

IV. CONCLUSIONS

The In this study, we explored a variety of machine learning and deep learning approaches to enhance medical predictions for conditions like diabetes and heart disease. Using the PIMA Indian Diabetes dataset, we developed an ensemble classification method combining Logistic Regression, Decision Tree (ID3), and an Artificial Neural Network (ANN), which significantly improved prediction accuracy compared to individual models. Additionally, we investigated advanced techniques such as XGBoost optimized through genetic algorithms and balanced with hybrid sampling methods like SMOTEENN. These strategies effectively addressed challenges like class imbalance and improved generalization, leading to high predictive performance with AUC values reaching 98.35% and accuracy above 93%. To further increase the transparency of complex ensemble models, we incorporated SHAP-based model interpretation, providing valuable insights into feature importance and supporting clinical applicability.
For cardiovascular disease prediction, we evaluated the performance of both Random Forest and Logistic Regression models using a reduced feature set selected through Pearson correlation analysis. Despite the small dataset, both models achieved comparable test accuracy (86%), suggesting that simpler, interpretable models can be just as effective as more complex ones under certain conditions. Moreover, we extended our investigation to the domain of heart sound classification, where we proposed a novel deep learning architecture for multi-label classification. This model achieved remarkable performance, including a sensitivity of 0.990 and specificity of 0.999, despite facing limitations such as dependence on signal segmentation and limited data for diastolic murmurs. Overall, the study highlights the importance of combining robust algorithmic approaches with effective data preprocessing, model interpretability, and adaptability to real-world clinical data to build reliable predictive systems for medical diagnostics and risk assessment.

It seems as if many studies have focused on only one disease. Limited work has been done on multi-disease classification. Working on models with real-time medical data is often lacking. This is seen here as well. More research is needed on local/regional data sets so that the model can become more generalizable and provide better results.

ACKNOWLEDGMENT

Vishal Kumar like to thank the School of Engineering and Technology, Shobhit University, Gangoh, India for supporting and encouraging him to work on this research.

REFERENCES

[1] https://diabetesresearch.org/what-is-diabetes/
[2] American Heart Association. (2021). *Type 1 diabetes statistics and prevalence*. Retrieved from https://www.heart.org/
[3] Patel, J., Patel, R., & Patel, D. (2017). *A classification algorithm-based hybrid diabetes prediction model*. International Journal of Engineering and Technical Research (IJETR), 7(3), 71–75. https://www.researchgate.net/publication/343281150
[4] Saxena, K., & Sharma, P. (2019). *Estimation of prediction for getting heart disease using logistic regression model of machine learning*. Algorithms, 12(2), 88. MDPI. https://www.mdpi.com/1999-4893/16/2/88
[5] Uddin, S., Lu, H., & Khan, A. (2020). *Deep cardio sound—An ensembled deep learning model for heart sound multi-labelling*. arXiv preprint. https://arxiv.org/abs/2204.07420
[6] Khan, F., Awan, M. J., & Imran, M. (2021). *Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm*. PLoS ONE, 16(6), e0311222. https://doi.org/10.1371/journal.pone.0311222
[7] Nature. (2019). *Global estimates of type 1 diabetes prevalence*. Retrieved from https://www.nature.com/articles/s41591-019-0614-2
[8] Centers for Disease Control and Prevention (CDC). (2015). *National diabetes statistics report: Estimates of diabetes and its burden in the United States*. Atlanta, GA: U.S. Department of Health and Human Services..
[9] Jegan, Chitra. (2013). Classification Of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications. 3. 1797 - 1801.

[10] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, "Machine learning-based predictive models for detection of cardiovascular diseases," Diagnostics, vol. 14, no. 2, p. 144, 2024.

[11] R. Hasan, "Comparative analysis of machine learning algorithms for heart disease prediction," in ITM Web of Conferences, vol. 40, p. 03007, EDP Sciences, 2021.

[12] *L. Li and Z.-P. Liu, "A connected network-regularized logistic regression model for feature selection," Applied Intelligence, vol. 52, no. 10, pp. 11672–11702, 2022.*

[13] Liu WB, Liang SN, Qin XW, Dong XG, Wang C. Diabetes prediction based on iterative random forest algorithm[J]. Journal of Changchun University of Technology. 2019; 6. https://doi.org/10.15923/j.cnki.cn22-1382/t.2019.6.15

[14] Adigun O, Okikiola F, Yekini N, Babatunde RJIJoACS, Applications. Classification of Diabetes Types using Machine Learning[J]. International Journal of Advanced Computer Science and Applications.2022; 13(9).

[15] Liao HL, Zeng XX, Li HF, Yu Y, Zhao C, Chen Y. Application of machine learning in disease prediction [J]. Journal of Biomedical Engineering Research. 2021; 40(2):203–9. https://doi.org/10.19529/j.cnki.1672-6278.2021.02.17

[1]Vishal Kumar received the BCA Degree from Uttarakhand Technical University, Dehradun and MCA degree from Uttaranchal University, Dehradun. He is currently pursuing the M.Tech. degree in computer science and engineering, Shobhit University Gangoh. He is also an Assistant Professor in the Department of Computer Science Engineering, School of Engineering and Technology, Shobhit University, Gangoh, India. He is an experienced educator with over Five years of teaching and research experience in the area of computer science. He published several research articles in international journals and has presented at reputed international conferences. His research interests include Digital Image Processing, Arduino Project, and Machine Learning technology.

[2]Dr. Varun Bansal serves as a Professor in the Department of Computer Science and Engineering at Shobhit University, Gangoh, bringing over 20 years of academic and research expertise. His core research areas include Honeypot Technology, machine learning and Blended Learning. Dr. Bansal has authored several publications in prominent international journals and conferences and has mentored many postgraduate and doctoral researchers. He remains actively engaged in academic partnerships and has played a vital role in advancing curriculum innovation and promoting research-driven learning within the department.

[3]Naman Saini received the B.Tech Degree from Uttarakhand Technical University and M.Tech degree from Shobhit University , Gangoh. He is currently pursuing the Phd in computer science and engineering. He is also an Assistant Professor in the Department of CSE, School of Engineering and Technology, Shobhit University, Gangoh, India. He is an experienced educator with over Two years of teaching and research experience in the area of computer science. He has published several research articles in international journals and has presented at reputed international conferences. His research interests include Networking ,Big Data , and Machine Learning Technology.

[4]Arun Saini received the B.Tech. (I.T.) Degree from Uttar Pradesh Technical University and M.Tech. (C.S.) degree from Shobhit University Gangoh. He is an Assistant Professor in the Department of CSE, School of Engineering and Technology, Shobhit University, Gangoh, India. He is an experienced educator with over eighteen years of teaching and research experience in the area of computer science. He has published several research articles in international journals and has presented at reputed international conferences. His research interests include Internet of things (IoT), Artificial intelligence, and Machine Learning technology.