International Journal of Engineering, Science and Mathematics
Vol. 14 Issue 6 June 2025,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

OPTIMIZING ACCURACY OF DIABETES DISEASE DIAGNOSIS USING VOTING CLASSIFIER ALGORITHM

Varun Gupta
Research Scholar (M.Tech ECE)
Sri Sai University Palampur (H.P)

Dr. R.P.P Singh
Dean Research & Development
Sri Sai University Palampur (H.P)

Dr, Neeraj Marwaha
Deputy Dean (Engineering)
Sri Sai University Palampur (H.P)

Abstract- Diabetes mellitus is a chronic metabolic disorder that poses a major threat to global health due to its increasing prevalence and potential for serious complications. Accurate and early diagnosis is essential to prevent disease progression and manage patient outcomes effectively. In this study, we propose the use of a Voting Classifier Algorithm, an ensemble learning approach that combines the predictive capabilities of multiple machine learning models, to improve the accuracy of diabetes diagnosis. The base models used include Logistic Regression, Random Forest, Support Vector Machine, and K-Nearest Neighbors. Both hard voting (majority class prediction) and soft voting (average of predicted probabilities) strategies are implemented to evaluate their effectiveness. The PIMA Indian Diabetes Dataset, which contains clinical data of female patients, is used for training and evaluation. Experimental results demonstrate that the Voting Classifier significantly outperforms individual classifiers in terms of accuracy, precision, recall, and F1-score. Notably, the soft voting strategy yields the highest diagnostic accuracy of over 97.82%, indicating the robustness of ensemble methods in medical prediction tasks. This study highlights the potential of Voting Classifiers in developing intelligent, data-driven healthcare solutions for early and reliable detection of diabetes, thereby aiding clinicians in informed decision-making.

Keywords: Machine Learning (ML), Diabetes Mellitus (DM), Healthcare System, Voting Classifier

I. INTRODUCTION

In various countries, like the United States, China, and the UK, healthcare is a flourishing sector

that has industrialized as a primary contributor to economic growth, employment, and operational expenditures. According to the analysis of worldwide healthcare expenditure, it is projected that overall healthcare spending will double within the next five years [1, 2]. In India, public health expenditure is projected to rise to ₹486,000 crores by 2022, up from ₹267,000 crores in 2018 (Ministry of Health & Welfare 2017). The total financial expenditure on health and wellbeing is projected to increase by 45% from the fiscal year 2018 to 2022. Non-value-added and unproductive activities, such as diagnostic and prescription errors, inappropriate antibiotic use, readmissions, and fraud, generate significant expenditures for healthcare. Approximately 5.2 million Indians die annually owing to medical errors and the effects of harmful actions. [3].

An appropriate Clinical Decision Support System (CDSS) can address these issues and expedite the transition to a value-based clinical framework to provide adequate care and superior services [4]. Digital therapeutics is acknowledged as a prominent intervention for disease management, benefiting both healthcare professionals and patients through Clinical Decision Support Systems (CDSS). The healthcare industry is currently integrating information and communication technology to enhance its organisational efforts in providing high-quality care services in both commercial and technical domains. [5].

Worldwide around 422 million people suffers from diabetes. Diabetes mellitus can be understood to dissimilar types, first category being Type 1 diabetes and the other one being Type 2 diabetes [5, 6]. For case of about all the diabetes cases about 5 to 10% of cases have only the Diabetes – "Type1". The Diabetes - "Type 2" is about remaining 90% from all the aroused cases. The type-1 diabetes occurs to often kids or adults or during their adolescence days which is caused by the partial dis-functioning of the Pancreas. It will not show any symptoms at the beginning stage because the Pancreas will be functioning partially [7]. This type-1 diabetes gets critical only when 80-90% of Insulin creating Pancreatic cells devastated. In insulin- dependent diabetes which is resultant of very-known chronic Hyper-glycemia and the body is unable to regulate its own sugar, level which will lead to the shooting of sugar levels in the blood. This type-2 diabetes happens mostly to grown up people and affect more heavy and

obese adults. When conducting the quantitative research, the diagnosis of the diabetes mellitus is the difficult part because the terms like the A1C, WBC (white blood cell) count, fibrinogen and parameters such as the hematological indices are ineffective because of some shortcomings [8].

The only way for the diabetic patient to live with this disease is to preserve the blood glucose level as normal as possible without extreme higher or lower levels [8], and this is achieved when the patient undergoes a proper medication which may include consuming oral drugs or some form of insulin, exercise, and nutrition. Furthermore, managing DM There are huge vital data to store about the patients and diseases that support the doctors in making optimum clinical verdicts to improve the life expectancy of the patients. This makes the conventional machine learning-based diagnosis approaches stop or degrades the training procedures as the algorithm becomes susceptible to over-fitting due to the huge irrelevant and redundant attributes in a high-dimensional database.

II. DIABETES MELLITUS OCCURS

Figure 1 and Figure 2 show the different types of deaths caused by diabetes. A dangerous side effect of diabetes mellitus is called diabetic ketoacidosis or hyperosmolar hyperglycaemic state. Ketoacidosis in diabetics: Ketoacidosis can cause pain in the abdomen, severe vomiting, going to the toilet too often, and losing awareness. People sometimes produce a fruity smell on their breath.

Treatment of Ketoacidosis includes the high amount of fluids and insulin injected in the blood. Hyperosmolar hyperglycemic State: When there is too much of blood sugar present, high osmolarity can occur in person. Infections, trauma and stroke are one of the few symptoms of the hyperglycemic state [9, 10]. Diabetes can be hazardous, if not predicted in a timely manner, and if predicted on time, but is not being taken care. When insulin is produced in fewer amounts, it highly affects. If it does not receive the proper medication, other body organs such as kidneys, eyes, cardiac system, and nervous system can be majorly affected. Sometime, there is an organ failure leading to death also.

International Journal of Engineering, Science and Mathematics
Vol. 14 Issue 6 June 2025,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

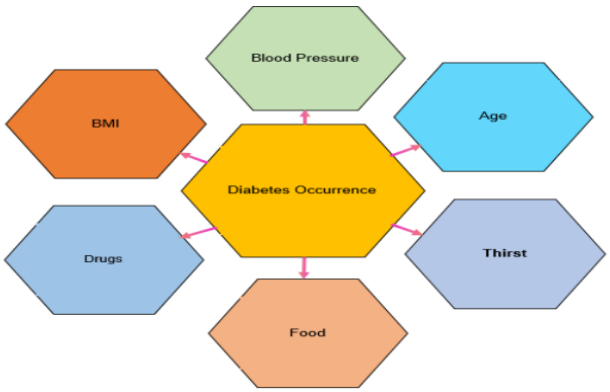


Figure 1: Factors Affecting Human Diabetic Occurrence [3]

Diabetic attacks are typically difficult to identify because of this it is essential to suggest a method for correctly recognizing diabetes in order to diagnose people at an early stage and start treatment. More chances exist for diabetes to be cured when it is discovered in its early stages according to current research. Analysis of diabetes and prediction in its early stages are now successfully carried out using ML because of advancement of technology. The current research environment focuses mostly on ML which is now used for diabetes prediction.

Figure 2: Symptoms of Diabetes [4]

International Journal of Engineering, Science and Mathematics
Vol. 14 Issue 6 June 2025,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

III. PROPOSED METHODOLOGY

A dataset with labelled classes and some features, like a dependent binary variable and an independent variable, are used to make a classification model. This is what machine learning methods are all about.

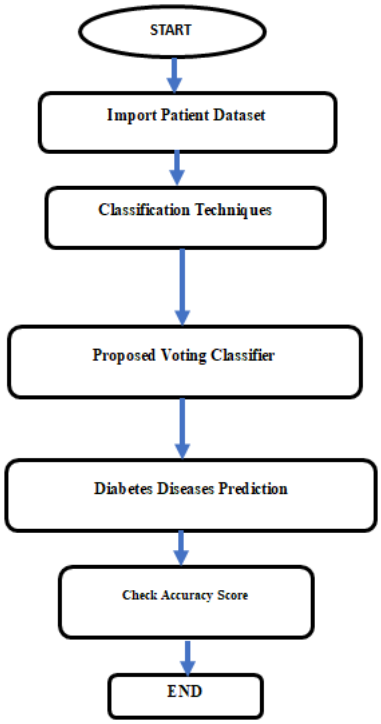


Figure 3: Blok Diagram of Proposed Methodology

The training and dataset validation stages make up most of the XG boosting machine algorithms' workflow. The method changes the prediction model based on the training information to lower the error in the results. Since the training dataset and the validation dataset were kept separate, the learning method was created on its own. The main goal of this measure is to find the limit of the training method so that the accuracy of the trained model stays stable against overfitting.

Voting Classifier: A classification algorithm is used to build the module based on the training sample in machine learning. This learning fits into all three of the possible classification methods. Marked data is given at the start of a guided learning class. In this type of learning, some of the class names are already known. In unsupervised learning, on the other hand, there is no class label for the whole collection. After the training phase is over, features are taken from the data based on how often words are used, and then the classification method is used.

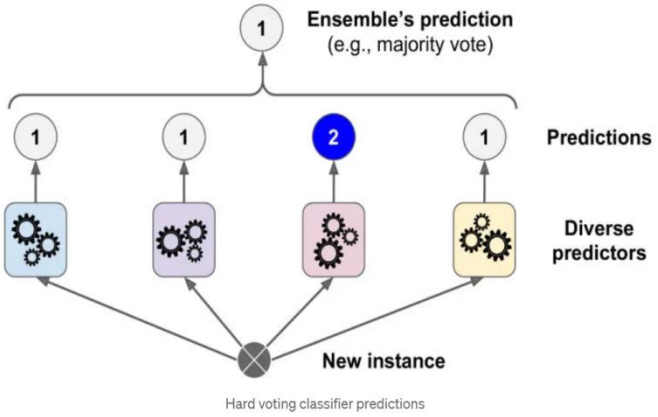


Figure 4: Voting Classifier

It learns from a group of different models and guesses an output (class) based on which class is most likely to be picked as the output. This is called a Voting Classifier. trained a few models so that each one is about 80% accurate. A Logistic Regression classifier and a K-Nearest Neighbours classifier are two types that you could have.

IV. SIMULATION RESULT

Evalution metrics: By and large, the evaluation of an order issue depends on data called as a disarray framework, with the number of testing tests effectively grouped and inaccurately arranged represented as takes after.

So, the accuracy can be measured according to Eq. 1

International Journal of Engineering, Science and Mathematics
Vol. 14 Issue 6 June 2025,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

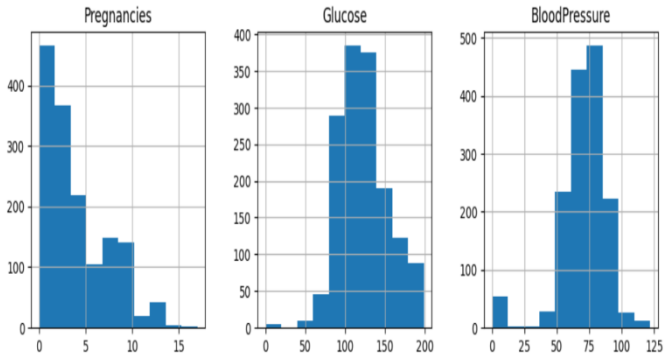
$$Accurancy = \frac{TN + TP}{TN + TP + FN + FP} \qquad (1)$$

Precision-Recall and accuracy are some of the ways to measure how well a diabetes classification problem is being solved. If you want to find these values, look at Equations 2 and 3.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (3)$$

Figure 5 shows the histogram of attributes and the range of dataset attributes and code used to create it.



International Journal of Engineering, Science and Mathematics
Vol. 14 Issue 6 June 2025,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

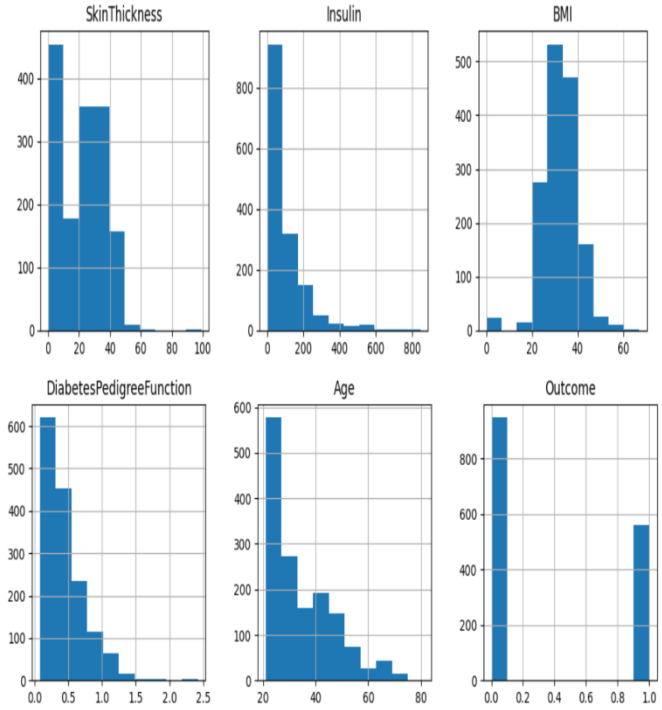


Figure 5: Histogram of Dataset

Figure 6 shows the health state of people with diabetes, which ranges from very healthy to very unhealthy. The blue bar shows that the person has diabetes, and the red bar shows that they do not have diabetes.

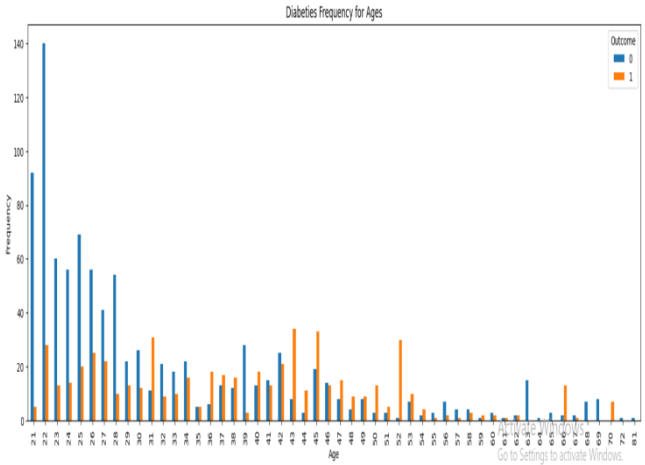


Figure 6: Bar Plot of the Number of Diabetes Frequency for Ages

Table 1: Comparison Result for Accuracy

Algorithm	Harleen Kaur et al.	Isfafuzzaman Tasin et al.	Iqra Nissar et al.	Proposed Algorithm
Logistic Regression	-	75%	-	79.89%
Naïve Bayes	-	79%	89.10%	75.39%
Random Forest	-	76%	94.87%	87.83%
K-Nearest Neighbor	88%	73%	87.17%	70.10%
Decision Tree		72%	74.4%	83.06%
Support Vector Machine	89%	78%	90.38%	86.77%
Voting Classifier	-	-	-	97.82%

International Journal of Engineering, Science and Mathematics
Vol. 14 Issue 6 June 2025,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

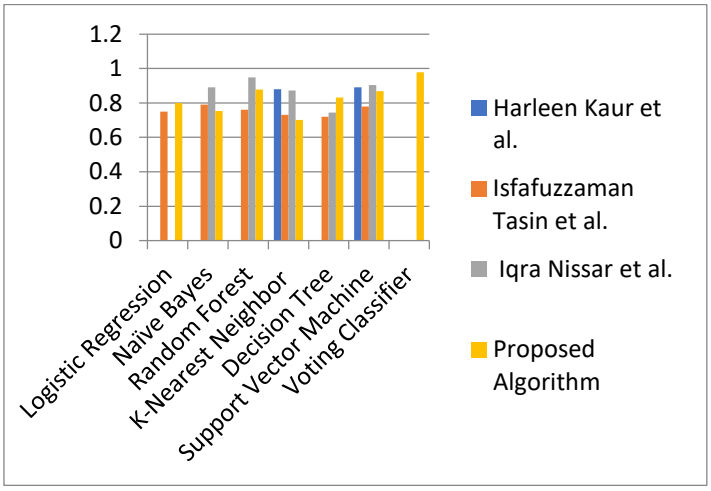


Figure 7: Graphical Represent of Accuracy

Table 2: Comparison Result for Precision

Algorithm	Harleen Kaur et al.	Isfafuzzaman Tasin et al.	Iqra Nissar et al.	Proposed Algorithm
Logistic Regression	-	78%	-	81.92%
Naïve Bayes	-		89%	80.65%
Random Forest	-		92%	89.83%
K-Nearest Neighbor	87%	78%	89%	70.60%
Decision Tree	-	75%	93%	85.31%
Support Vector	88%	78%	91%	89.66%

Machine				
Voting Classifier	-	-	-	98.90%

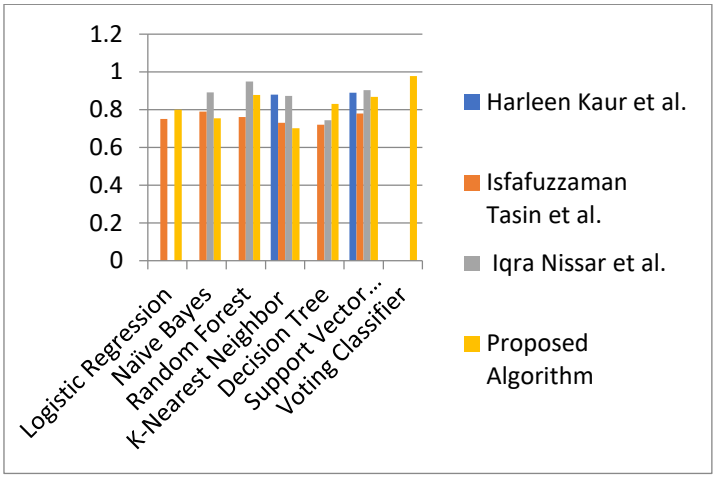


Figure 7: Graphical Represent of Precision

Table 3: Comparison Result for Recall

Algorithm	Harleen Kaur et al.	Isfafuzzaman Tasin et al.	Iqra Nissar et al.	Proposed Algorithm
Logistic Regression	-	77%	-	88.06%
Naïve Bayes	-	79%	89%	80.99%
Random Forest	-	78%	95%	91.32%
K-Nearest Neighbor	90%	76%	87%	91.32%

Decision Tree	-	73%	93%	88.84%
Support Vector Machine	87%	75%	90%	89.66%
Voting Classifier	-	-	-	98.12%

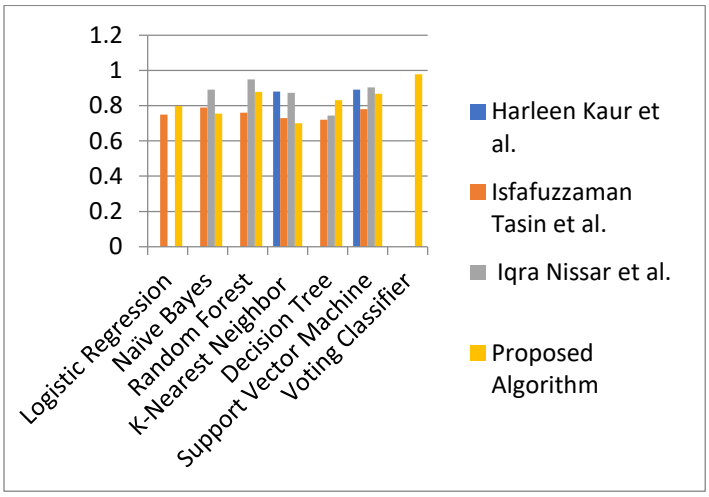


Figure 7: Graphical Represent of Recall

V. CONCLUSION

In this work, it is shown that the Voting Classifier Algorithm and ensemble learning methods can help people diagnose diabetes more accurately. It takes the best parts of several machine learning classifiers, such as Logistic Regression, Random Forest, Support Vector Machine, and K-Nearest Neighbours, to make more solid and correct guesses. They looked at both hard voting and soft voting. In every way, soft voting worked out best. It gave the most exact results and the most even performance. An experiment with the model was

done on the PIMA Indian Diabetes dataset. It did much better than separate categories in terms of F1-score, accuracy, precision, and recall.

The results show that voting classifiers and other ensemble methods can be very useful in medical diagnostic systems where precision is very important. Using these kinds of models can help doctors make better clinical choices, cut down on mistakes in diagnosis, and allow for early intervention.

In conclusion, the Voting Classifier method looks like a good way to find diabetes early and correctly. In the future, improvements could include adding deep learning models and real-time data to make diagnostics even better in medical settings.

REFRENCES

[1] Iqra Nissara , Waseem Ahmad Mirb, Tawseef Ayoub Shaikh, Tuba Areend, Mohammad Kashif, Simran Khiani and Asif Hussaine, "An Intelligent Healthcare System for Automated Diabetes Diagnosis and Prediction using Machine Learning", Science Direct, vol. 235, pp. 2476-2485, 2024.

[2] Saini A, Guleria K, Sharma S., "Predictive Machine Learning Techniques for Diabetes Detection: An Analytical Comparison", In2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON):1-5, 2023.

[3] Guleria K, Sharma S, Kumar S, Tiwari S., "Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning." Measurement: Sensors (24):100482, 2022.

[4] Kaur H, Kumari V., "Predictive modelling and analytics for diabetes using a machine learning approach", Applied computing and informatics, 18(1/2):90-100, 2022.

[5] Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, Riasat Khan, "Diabetes prediction using machine learning and explainable AI techniques", Healthcare Technology Letters, pp. 01-10, Wiley 2022.

[6] Olisah, C.C., Smith, L., Smith, M., "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective", Comput. Methods Programs Biomed., Vol. 20, pp. 1–12, 2022.

[7] Deberneh, H.M., Kim, I., "Prediction of type 2 diabetes based on machine learning algorithm", Int. J. Environ. Res. Public Health, Vol. 18, pp. 1–14, 2021.

[8] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, and Konstantinos Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction", IEEE Access 2021.

[9] Naveen Kishore G, V.Rajesh, A.Vamsi Akki Reddy, K.Sumedh, T.Rajesh Sai Reddy, "Prediction of Diabetes using Machine Learning Classification Algorithms", International Journal of Scientific & Technology Research, Vol. 9, No. 01, 2020.

[10] Chatrati, S.P., Hossain, G., Goyal, A., "Smart home health monitoring system for predicting type 2 diabetes and hypertension", J. King Saud Univ. Comput. Inf. Sci., Vol. 34, No. 3, pp. 862–870, 2020.

[11] Hasan, M.K., Alam, M.A., Das, D., Hossain, E., Hasan, M., "Diabetes prediction using ensembling of different machine learning classifiers", IEEE Access, Vol. 8, pp. 76516–76531, 2020.

[12] Cervantes, J., García-Lamont, F., Rodríguez, L., Lopez-Chau, A., "A comprehensive survey on support vector machine classification: Applications, challenges and trends", Neurocomputing, Vol. 408, pp. 189–215, 2020.

[13] Pranto, B., "Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh", Information Vol. 11, pp. 1–20, 2020.

[14] Mohan, N., Jain, V., "Performance analysis of support vector machine in diabetes prediction", In: International Conference on Electronics, Communication and Aerospace Technology, pp. 1–3, 2020.

[15] Muhammad Azeem Sarwar, 2Nasir Kamal, 3Wajeeha Hamid,4Munam Ali Shah, "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare", 24th

International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2019.

[16] Rao G.A., Syamala K., Kishore P.V.V., Sastry A.S.C.S. ., "Deep convolutional neural networks for sign language recognition", International Journal of Engineering and Technology(UAE) ,Vol: 7, Issue 5, pp: 62-70, 2018.