OriginalArticle

Analysis and Comparison for Prediction of Diabetic among Pregnant Women using Innovative Decision Tree algorithm over Support Vector Machine Algorithmwith Improved Accuracy

VenkataSai Kumar Pokala[1], NeelamSanjeev Kumar[2]
[1]Research Scholar,Department of Biomedical Engineering,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences.Saveetha University, Chennai. Tamilnadu. India. Pincode: 602105
[2]Project guide, Corresponding authorDepartment of Biomedical Engineering,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences.Saveetha University, Chennai. Tamilnadu. India. Pincode:602105

Abstract

Aim: A decision tree algorithm and Support vector machine were employed in machine learning algorithms for the prediction of diabetes among pregnant women to achieve accuracy, sensitivity, and precision. **Materials and Methods:** To test the technique's utility, researchers used open data sets such as the Pima Indian dataset from the UCI website to look at diabetes in pregnant women. This study has two groups, each with a sample size of 40: Decision tree (N=40) and Support vector machine learning (N=40). The sample size was calculated using a pre-test power of 80%, a threshold of 0.05, and a confidence interval of 95%. **Results**: Algorithm performance is measured by its accuracy, sensitivity, and precision. The accuracy rate of the Decision tree is 65%, whereas the accuracy rate of the Support vector machine is 67%. The decision tree has a sensitivity rate of 54%, whereas the support vector machine sensitivity rate of 67%. The decision has a precision rate of 75%, whereas the support vector machine has a precision rate of 63%. The accuracy rate differs by a considerable amount p=0.366 with p>0.05. **Conclusion**: The Support vector machine method predicts superior classifications in identifying the accuracy, sensitivity, and precision for accessing the rate for diabetes prediction among pregnant women when compared to the Innovative Decision Tree algorithm.

Keywords: Diabetes prediction, Innovative Decision tree algorithm, Support vector machine algorithm, Artificial Intelligence, Accuracy.

DOI: 10.47750/pnr.2022.13.S03.017

INTRODUCTION

Diabetes disease is a disorder that occurs when there is a significant increase in blood glucose, often known as blood sugar, according to diabetes medical history. Unhealthy eating habits raise blood glucose levels and influence diabetes. Insulin is a pancreatic hormone that permits glucose from meals to enter cells and be used for energy. It is also referred to as "borderline diabetes" or "a touch of sugar"(Kanmani and Murugan 2020). Diabetic Mellitus (DM) has become a highly serious illness in developing nations like India, given the current situation. This is a non-communicable disease (NCD) that affects a large number of people. According to 2017 data, around 425 million people have diabetes. Diabetes claims the lives of around 2-5 million people each year. It is estimated that by 2040, the population would have risen to 629 million (Mujumdar and Vaidehi 2019). High blood sugar affects several areas of the human body, including blood vessels and nerves, as well as causing symptoms such as increased thirst, appetite, and weight loss. Diabetes patients often require ongoing therapy or face a slew of life-threatening consequences (Brisimi et al. 2019). Diabetes is diagnosed when the 2-hour post-load plasma glucose level is at least 200 mg/dL, and different research on diabetes detection emphasises the need of recognising diabetes early. Application of prediction of Breast cancer in study as a result, medical research is becoming increasingly interested in the prediction and prevention of diabetes mellitus (Bindiya 2020).

In recent years, a number of machine learning methods for diabetes prediction have been created. There were 540 results on Google Scholar, and 23 research publications were available on ScienceDirect. Important decisions and predictions may be made using analytics on healthcare data. In this paper, we use machine

learning methods in a Hadoop-Map Reduce environment to forecast the forms of diabetes that are common, their complications, and the therapy that may be given as a result (Kalyankar, Poojara, and Dharwadkar 2017). To identify valuable patterns within datasets, machine learning methods are utilized. A categorization issue is always present in a medical diagnosis (Verma and Mishra 2017). In the medical and healthcare fields, classification is one of the most commonly utilized data mining and machine learning techniques. In the medical and healthcare fields, the earning approach is used (R and Manimaran 2017). There is a large repository of diverse algorithms and approaches used in data mining and machine learning, especially for supervised machine learning techniques (Kour et al. 2021). As a result, choosing the best algorithm or approaches to implement DD detection and early diagnostic systems has been a problem for researchers (Hasan and Al MehediHasan 2020). Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021)

The major problem that motivated me to do this research on improving the accuracy of machine learning and diabetes prediction in pregnant women in the early stages is inefficient early diabetes diagnosis and human errors in existing diabetes detection approaches. The main flaw in the present research is that diabetes prediction algorithms are often incorrect. Because the authors are experts in machine learning algorithms and deep learning technologies, they compared machine learning algorithms("An Efficient Approach for Weblog Analysis Using Machine Learning Techniques" 2020; Goyal and Rathore 2020; Chan and Chin 2016) The main objective is to test and evaluate diabetic deduction methods using cutting-edge machine learning techniques like the Decision Tree and Logistic Regression algorithms.

MATERIALS AND METHODS

The study was carried out at the University simulation lab, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. In this paper, the sample size was determined using clinical.com, using an alpha error-threshold of 0.05, enrollment ratio of 0:1, 95 percent confidence interval, and power of 80 percent, based on earlier study findings (Hasan and Al MehediHasan 2020). A Decision Tree algorithm in Group 1 (N=20) and Support Vector Machine (N=20) were in Group 2. This research includes a total of 40 samples.

The Kaggle website provided the data samples utilized in this investigation. To acquire the absolute data necessary, the data set is subjected to data reduction procedures. To execute the classification learning technique, the data should be fed into Matlab 2021a. To train input data should be loaded into categorization learning systems. The improved data is trained twice, once for the Decision Tree with validation ranging from 5 to 24, and once for the Support Vector Machine with validation ranging from 5 to 24. After data validation of an algorithm, acquire the confusion matrix for each validation, which includes the TP (true positive), TN (true negative), FP (false positive), FN (false negative). These variables are used to calculate accuracy, sensitivity, and precision.

Preprocessing is an important task that should not be ignored because the model's predictions are largely based on data quality. Many MATLAB filters enable preprocessing, and the most suited techniques are chosen for the original dataset's optimization. Each characteristic's medical consequences are first examined in relation to diabetes mellitus (DM). Because the attribute "number of pregnancies" was discovered to have a significant impact on the DM, it was changed to a nominal value of 0 for nonpregnant and 1 for pregnant. As a result, data complexity was reduced to an absolute minimum. Erroneous and missing values in the dataset are discovered and removed, which are a key cause of many non-correct results in most trials. The number for diastolic blood pressure and body mass index, for example, cannot be zero, and if it is in dataset 63, it means the real value is missing.

STATISTICAL ANALYSIS

IBN SPSS 26.0.1 is the statistical software package that was used in this study. The mean, standard deviation, and standard error mean statistical significance between the groups were determined using the independent sample T-Test, and then a comparison of the two groups using SPSS software yielded accurate values for the two different algorithms that will be used with the highest level of accuracy 83.54 percent, mean value 0.8354, and standard deviation value 0.03235. The image size is an independent variable, while the image size is an independent variable, while the image accuracy is a dependent variable ((Hasan and Al MehediHasan 2020)).

RESULTS

Table 2 shows how to predict diabetes in pregnant women using Decision Tree and Support Vector Machine methods. When comparing the accuracy, sensitivity, and precision of the Decision Tree with the Support Vector Machine, the Support Vector Machine approach outperforms the Decision Tree. Table 2 shows the accuracy of the Support Vector Machine and Decision Tree methods. A Support Vector Machine has higher accuracy, sensitivity, and precision rate than a Decision Tree as shown in Table 1a and Table 1b. Decision Tree findings have 72.35 percent accuracy, 73.27 percent sensitivity, and 75.97 percent precision rate, while Support Vector Machine results have a 77.67 percent accuracy, 76.67 percent sensitivity, and precision of 83.54 percent. Support Vector Machine has a lower error rate than Decision Tree, as seen in Table 2.

Table 3 shows that using the independent sample T-test, there appears to be a statistically insignificant difference (P=0.286 for accuracy with p>0.05, P=0.055 for sensitivity with p<0.05, P=0.299 for precision with p>0.05, in both techniques (P=0.0286 for accuracy, P=0.055 for sensitivity, for precision P= 0.299, P=<0.001). The Support Vector Machine technique outperformed the Decision Tree in predicting diabetes disease, according to these data. Figure 1 shows a bar chart depicting the comparison of Decision Tree and Support Vector Machine mean accuracy, sensitivity, and precision values.

Figures 2a and 2b represent the true positive, true negative, false positive, and false negative values are utilized to derive the accuracy, sensitivity, and precision values from the confusion matrix of the Decision Tree and Support Vector Machine.

DISCUSSION

Support Vector Machine performed better than Decision Tree accuracy (72.35 percent), sensitivity (73.27 percent), and precision (75.97 percent) in this research paper for predicting diabetes among pregnant women, with accuracy (77.67 percent), sensitivity (76.67 percent), and precision (83.54 percent). Despite the fact that it is not statistically significant, the significant difference appears to have expanded. Machine learning techniques are widely used in the early detection of diabetes.

The researchers used Decision tree and SVM algorithms to predict diabetes with the SVM model having 79% (S.Soliman et al. 2014). The authors have proposed a model based on ensemble methods using machine learning algorithms, with the objective of assessing the model's accuracy, precision, and sensitivity (Saradha and Sujatha 2018). According to the findings, the SVM model had the best accuracy (79.2?%), sensitivity (79.9%), and specificity 80.4% of three algorithms (Nagaraj and Deepalakshmi 2021). SVM and Gradient Boosted Decision Tree are two of the machine learning techniques used to construct this ensemble model, which has an 82.33 percent accuracy (Nagaraj and Deepalakshmi 2021). (Abu-Farha, Abubaker, and Tuomilehto 2021) Related work on SVM, Gradient Boosted Decision Tree are among the machine learning approaches used to create this ensemble model which produces an accuracy of 82.21%. Clinical prediction models are evaluated with the use of machine learning and laboratory data, and accuracy, precision, and recall are 80.0 percent, 81.3 percent, and 84.5% respectively (Naz and Ahuja 2020)

This study is hampered by a lack of data. Higher accuracy, sensitivity, and precision may be achieved by increasing the sample size. Cleaning and preparing the data for diabetes prediction takes additional time. Soon, an effective classification method will be created that combines the efficiency of the best-performing algorithms to improve diabetes prediction accuracy in pregnant women. Better performance may be achieved by combining a large data set of real-time applications with various machine learning and deep learning approaches. Overall, the outcomes of the study are extremely promising for the future. In the near future, the proposed technique, in combination with the suggested Machine learning classification algorithms, may be beneficial in the prediction or diagnosis of new illnesses. For diabetes prediction analysis, the study work, as well as a few other Machine learning approaches, might be updated and improved. Metaheuristic algorithms will be utilized to completely learn the missing data in future studies. The algorithms have been enhanced so that they can learn how to forecast missing data in the future. With novel swarm-based meta-heuristic characteristics, learning algorithms like Grey Wolf Optimizer (GWO) and other nature-inspired computer algorithms can aid research. Furthermore, the study may be broadened to predict diabetes by collecting data from several sites throughout the world and creating a more accurate and common discriminating framework. To make the diabetes analysis more automated, the work might be modified and improved.

CONCLUSION

In this diabetes prediction research, Matlab-based Support Vector Machine (77.67 percent) generated superior results than Decision Tree (72.35). Furthermore, the algorithm's performance improved as the amount of data increased, unlike prior techniques. This model is quite efficient and has a lot of promise for predicting and assessing diabetes, thus it may be used in hospitals and testing centers.

DECLARATION

Conflicts of Interest

No conflict of interest in this manuscript

Author Contributions

Author VSKP was involved in data collection, data analysis & manuscript writing. Author NSK was involved in conceptualization, data validation, and critical review of manuscripts.

Acknowledgment

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

Funding:

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Venus Electronics Tamilnadu.
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Abu-Farha, Mohamed, Jehad Ahmed Abubaker, and JaakkoTuomilehto. 2021. *Diabetes in the Middle East*. Frontiers Media SA.
2. Adhinarayanan, Rajesh, AravindhRamakrishnan, GopalKaliyaperumal, Melvinvíctor De Poures, Rajesh Kumar Babu, and DamodharanDillikannan. 2020. "Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
3. "An Efficient Approach for Weblog Analysis Using Machine Learning Techniques." 2020. *Predictive Analytics Using Statistics and Big Data: Concepts and Modeling*. https://doi.org/10.2174/9789811490491120010009.
4. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. "Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre–reinforced Epoxy Composite." *Biomass Conversion and Biorefinery*, August. https://doi.org/10.1007/s13399-020-00938-0.
5. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. "Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique." *Biomass Conversion and Biorefinery*, May. https://doi.org/10.1007/s13399-021-01551-5.
6. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik R. Bhagat. 2021. "Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for PhotocatalyticTransesterification of Castor Oil." *Fuel* 304 (November): 121490.
7. Bindiya, A. R. 2020. "Diabetes Mellitus Prediction Using Machine Learning Algorithms." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2020.30632.
8. Brisimi, Theodora S., TingtingXu, Taiyao Wang, Wuyang Dai, and IoannisChPaschalidis. 2019. "Predicting Diabetes-Related Hospitalizations Based on Electronic Health Records." *Statistical Methods in Medical Research* 28 (12): 3667–82.
9. Chan, T. K., and C. S. Chin. 2016. "Data Analysis to Predictive Modeling of Marine Engine Performance Using Machine Learning." *2016 IEEE Region 10 Conference (TENCON)*. https://doi.org/10.1109/tencon.2016.7848391.
10. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. "Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement." *Biomass Conversion and Biorefinery*, February. https://doi.org/10.1007/s13399-021-01312-4.
11. Goyal, Neha, and Sanghmitra Singh Rathore. 2020. "Predictive Visual Analysis of Speech Data Using Machine Learning Algorithms." *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*. https://doi.org/10.1109/icetce48199.2020.9091770.
12. Hasan, Kazi Amit, and Md Al Mehedi Hasan. 2020. "Prediction of Clinical Risk Factors of Diabetes Using Multiple Machine Learning Techniques Resolving Class Imbalance." *2020 23rd International Conference on Computer and Information Technology (ICCIT)*. https://doi.org/10.1109/iccit51783.2020.9392694.
13. Jayanth, BellappuVenkat, Melvin Victor Depoures, GopalKaliyaperumal, DamodharanDillikannan, DilipsinghJawahar, KumaranPalani, and Ganesha Prasad MeravanigeeShivappa. 2021. "A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
14. Kalyankar, Gauri D., Shivananda R. Poojara, and Nagaraj V. Dharwadkar. 2017. "Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop." *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. https://doi.org/10.1109/i-smac.2017.8058253.
15. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and ShantanuPatil. 2020. "Melatonin

Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration." *Process Biochemistry* 99 (December): 36–47.
16. Kanmani, K., and A. Murugan. 2020. "Prognosticate and Diagnosis of Diabetes Using Data Preprocessing and Null Value Removal on the Modified Data Set with Possible Outcome of a Decision Tree Construction through R- Programming." *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2020.10.631.
17. Kour, Hardeep, Munish Sabharwal, Shakhzod Suvanov, and Darpan Anand. 2021. "An Assessment of Type-2 Diabetes Risk Prediction Using Machine Learning Techniques." *Proceedings of International Conference on Big Data, Machine Learning and Their Applications*. https://doi.org/10.1007/978-981-15-8377-3_10.
18. Mujumdar, Aishwarya, and V. Vaidehi. 2019. "Diabetes Prediction Using Machine Learning Algorithms." *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2020.01.047.
19. Nagaraj, P., and P. Deepalakshmi. 2021. "Diabetes Prediction Using Enhanced SVM and Deep Neural Network Learning Techniques." *International Journal of Healthcare Information Systems and Informatics*. https://doi.org/10.4018/ijhisi.20211001.oa25.
20. Naz, Huma, and Sachin Ahuja. 2020. "Deep Learning Approach for Diabetes Prediction Using PIMA Indian Dataset." *Journal of Diabetes & Metabolic Disorders*. https://doi.org/10.1007/s40200-020-00520-5.
21. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. "Collective Influence of 1-Decanol Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends." *Fuel* 277 (October): 118166.
22. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. "Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications." *Fuel* 278 (October): 118315.
23. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. "Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour." *Biomass Conversion and Biorefinery*, March. https://doi.org/10.1007/s13399-021-01427-8.
24. R, Manimaran, and R. Manimaran. 2017. "Prediction of Diabetes Disease Using Classification Data Mining Techniques." *International Journal of Engineering and Technology*. https://doi.org/10.21817/ijet/2017/v9i5/170905319.
25. Saradha, S., and P. Sujatha. 2018. "Prediction of Gestational Diabetes Diagnosis Using SVM and J48 Classifier Model." *International Journal of Engineering & Technology*. https://doi.org/10.14419/ijet.v7i2.21.12395.
26. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). https://doi.org/10.1002/ep.13696.
27. Shanmugam, Rajasekaran, DamodharanDillikannan, GopalKaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 43 (23): 3064–81.
28. S.Soliman, Omar, Omar S. Soliman, Faculty of Computers and Information, Cairo University, and Eman AboElha. 2014. "Classification of Diabetes Mellitus Using Modified Particle Swarm Optimization and Least Squares Support Vector Machine." *International Journal of Computer Trends and Technology*. https://doi.org/10.14445/22312803/ijctt-v8p108.
29. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. "Pretreatment and Process Optimization of Bioethanol Production from Spent Biomass of GanodermaLucidum Using Saccharomyces Cerevisiae." *Fuel* 306 (December): 121680.
30. Verma, Deepika, and Nidhi Mishra. 2017. "Analysis and Prediction of Breast Cancer and Diabetes Disease Datasets Using Data Mining Classification Techniques." *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. https://doi.org/10.1109/iss1.2017.8389229.

TABLES AND FIGURES

Table 1a. Diabetes prediction samples using Decision Tree algorithm

Samples	Accuracy	Sensitivity	Precision
1	0.76	0.76	0.81
2	0.67	0.67	0.7
3	0.7	0.71	0.72
4	0.71	0.72	0.72
5	0.72	0.73	0.7
6	0.76	0.75	0.83
7	0.73	0.7	0.7
8	0.74	0.72	0.83
9	0.75	0.78	0.7
10	0.73	0.73	0.79
11	0.73	0.7	0.77

12	0.7	0.71	0.72
13	0.75	0.7	0.81
14	0.7	0.70	0.7
15	0.68	0.69	0.73
16	0.71	0.7	0.7
17	0.7	0.72	0.70
18	0.72	0.74	0.72
19	0.73	0.7	0.77
20	0.72	0.75	0.70
21	0.74	0.77	0.74
22	0.76	0.76	0.72

Table 1b. Diabetes prediction samples using Support Vector Machine

Samples	**Accuracy**	**Sensitivity**	**Precision**
1	0.71	0.73	0.79
2	0.76	0.76	0.81
3	0.81	0.77	0.91
4	0.78	0.78	0.83
5	0.8	0.78	0.85
6	0.77	0.76	0.83
7	0.74	0.75	0.77
8	0.75	0.75	0.8
9	0.76	0.76	0.8
10	0.76	0.76	0.8
11	0.78	0.77	0.85
12	0.77	0.76	0.83
13	0.78	0.77	0.85
14	0.77	0.76	0/83
15	0.8	0.77	0.8
16	0.8	0.77	0.8
17	0.77	0.76	0.83

18	0.76	0.76	0.8
19	0.77	0.77	0.83
20	0.78	0.77	0.85
21	0.78	0.78	0.81
22	0.79	0.76	0.83

Table 2: Comparison of mean accuracy, sensitivity, and precision using Principal Component Analysis and Support Vector Machine algorithms.

GROUP STATISTICS					
PARAMETERS	**GROUP**	**N**	**MEAN**	**STD. DEVIATION**	**STD. ERROR MEAN**
ACCURACY	DECISION TREE	20	0.7235	0.02616	0.00585
	SUPPORT VECTOR MACHINE	20	0.7767	0.02248	0.00503
SENSITIVITY	DECISION TREE	20	0.7327	0.02466	0.00551
	SUPPORT VECTOR MACHINE	20	0.7667	0.01721	0.00385
PRECISION	DECISION TREE	20	0.7597	0.03831	0.00857
	SUPPORT VECTOR MACHINE	20	0.8354	0.03235	0.00723

Table 3. Independent sample T-test in predicting the accuracy, sensitivity, and precision of Diabetes using the Decision tree and SVM algorithm. There appears to be an insignificant difference in both methods p>0.05 for Accuracy and precision

Parameter	**Equal Variances**	**Levene's Test for Equality of Variances**		**T-test for Equality of Means**					
Accuracy	Assumed	F	Sig	t	df	Significance (one-Sided p)	Mean Difference	95% Confidence interval (Lower)	95% Confidence interval (Upper)

		1.172	0.28	-6.89	38	<.001	-.05316	-.06877	-.03754
	Not assumed			-6.89	37.16	<.001	-.05316	-.06878	-.03753
Sensitivity	Assumed	3.916	0.05	-5.06	38	<.001	-.03406	-.04767	-.02045
	Not assumed			-5.06	33.96	<.001	-.03406	-.04773	-.02039
Precision	Assumed	1.109	0.29	-6.75	38	<.001	-.07577	-.09846	-.05307
	Not assumed			-6.75	36.96	<.001	-.07577	-.09848	-.05305

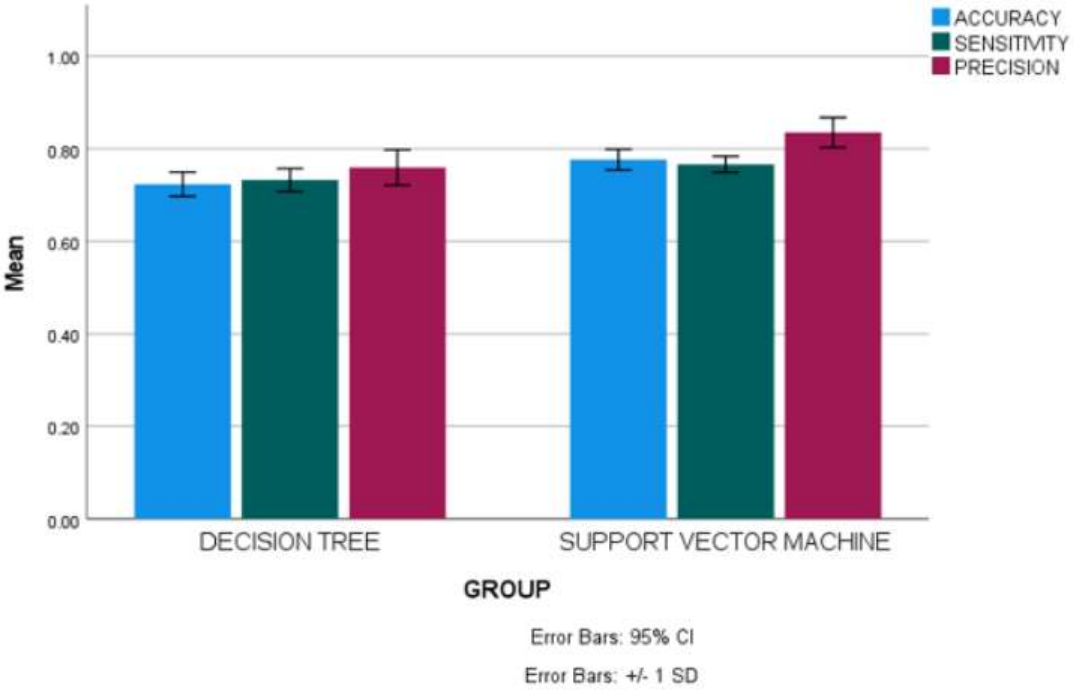

Fig. 1. Bar graph representing the comparison of mean accuracy, sensitivity, and precision of Diabetes prediction with the Decision tree algorithm and the Support vector machine algorithm. Both the techniques appear to produce the same variable results with accuracy ranging from72% to 84%. X-axis: Decision tree vs SVM. Y-axis: mean accuracy, sensitivity, and precision detection ± 1 SD.

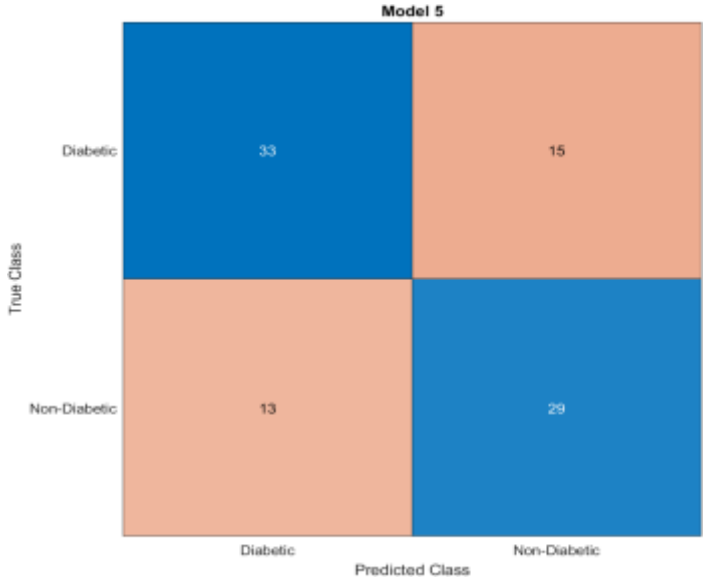

Fig. 2a. Confusion matrix for Decision tree algorithm K=5. True Positive is found to be 33% and false positive is found to be 15%, true negative is found to be 29% and false negative is found to be 13%.

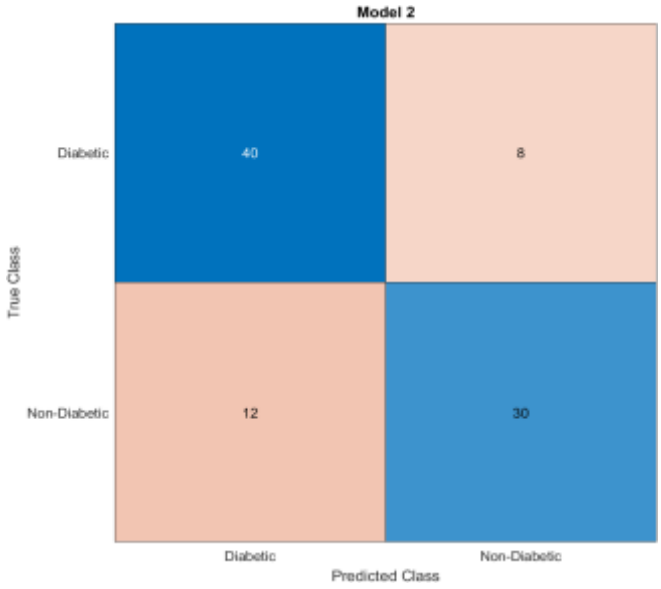

Fig. 2b. Confusion matrix for Support vector machine K=5. True Positive is found to be 40% and false positive is found to be 8%, true negative is found to be 30% and false negative is found to be 12%.