

Analysis and Comparison for Prediction of Diabetic among Pregnant Women using Innovative Support Vector Machine Algorithm over Random Forest Algorithm with Improved Accuracy

Venkata Sai Kumar Pokala, Neelam Sanjeev Kumar

Department of Biomedical Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai. Tamilnadu. India. Pincode:602105.

ABSTRACT

Aim: To achieve accuracy, sensitivity, and precision in AI (Artificial Intelligence) calculations for the prediction of diabetes among pregnant women, a Support Vector Machine and Random Forest algorithms were utilized. **Materials and Methods:** Research looked at diabetes in pregnant women using accessible data sets such as the Pima Indian dataset from the UCI website to assess the technique's usefulness. Support Vector Machine (N=40) and Random Forest (N=40) are the two groups in this study, each having a sample size of 40. A pretest power of 80%, a threshold of 0.05, and a confidence interval of 95% were used to compute the sample size. **Results:** The accuracy, sensitivity, and precision of an algorithm are used to evaluate its performance. The Support Vector Machine has a 75% accuracy rate, whereas the Random Forest has a 74% accuracy rate. The sensitivity rate of the Support Vector Machine is 65%, whereas the sensitivity rate of the Random Forest is 68%. The Support Vector Machine has a precision rate of 80%, whereas the Random Forest has a precision rate of 76%. The accuracy rate is significantly different with $p=0.466, p>0.05$. **Conclusion:** When compared to the innovative Support Vector Machine Algorithm, the Random Forest approach predicts superior classifications in identifying the accuracy, sensitivity, and precision for accessing the rate for diabetes prediction among pregnant women.

Keywords

Diabetes prediction, Innovative Support Vector Machine Algorithm, Random Forest Algorithm, Artificial Intelligence, Accuracy.

Imprint

Venkata Sai Kumar Pokala, Neelam Sanjeev Kumar. **Analysis and Comparison for Prediction of Diabetic among Pregnant**

Women using Innovative Support Vector Machine Algorithm over Random Forest Algorithm with Improved Accuracy. *Cardiometry*; Special issue No. 25; December 2022; p. 956-962; DOI: 10.18137/cardiometry.2022.25.956962; Available from: <http://www.cardiometry.net/issues/no25-december-2022/support-vector-machine-algorithm>

INTRODUCTION

Diabetes mellitus (DM) is one of the world's deadliest noncommunicable illnesses. Diabetes prognosis and prevention have become a serious problem, according to previous medical records. The medicines are still insufficient to control the condition as the number of individuals diagnosed rises (Rout and Kaur 2020). As a result, greater forecast analysis is necessary to treat diabetes at a beginning phase, by which it can assist address less difficulties and allow the patient to be treated with fewer drugs and at a lower cost (Rodriguez-Romero et al. 2019). Hyperglycemia caused by defects in insulin production, insulin action, or both comes within the group of metabolic disorders known as diabetes (Hasan and Al Mehedi Hasan 2020). Diabetes is linked to persistent hyperglycemia as a result of long-term organ damage, malfunction, and failure, in the eyes, kidneys, nerves, heart, and blood arteries. The spread of diabetes is caused by a number of pathogenic mechanisms. These reach from immune system obliteration to the pancreatic b-cells and subsequent insulin inadequacy to anomalies leading to insulin resistance (Xie et al. 2019). Glucose levels in the human body typically vary from 70 to 99 mg per deciliter. Diabetes is diagnosed when the blood glucose level exceeds 126 mg/dl. If a person's blood glucose level is between 100 and 125 mg/d, they have prediabetes (Khanam and Foo 2021). Diabetes does not have a permanent cure. The most frequent consequences of long-term diabetes are macrovascular and microvascular diseases (Costea, Moisi, and Popescu 2021). Application of prediction of Breast cancer in study the authors compared machine learning algorithms since they are specialists in machine learning algorithms and deep learning technologies (Theerthagiri, Usha, and Vidya, n.d.). Our team has extensive knowledge and research experience that has translate into high quality publications (Chellapa et al. 2020; Lavanya, Kannan, and Arivalagan 2021; Raj R, D, and S 2020; Shilpa-Jain et al. 2021; S, R, and P 2021; Ramadoss,

Padmanaban, and Subramanian 2022; Wu et al. 2020; Kalidoss, Umapathy, and Rani Thirunavukkarasu 2021; Kaja et al. 2020; Antink et al. 2020; Paul et al. 2020; Malaikolundhan et al. 2020)

Several machine learning approaches for diabetes prediction have been developed in recent years. On Google Scholar, there were 1040 hits, and on ScienceDirect, there were 18 research papers. Predictive analysis is a method for extracting information from current and verifiable informational indexes and determining future events that consolidates different information mining strategies, AI calculations, and insights (Pasha 2020). Prescient Analysis is a philosophy that consolidates a scope of AI calculations, information mining strategies, and factual ways to deal with reveal data and estimate future events utilizing current and verifiable information (Costea, Moisi, and Popescu 2021). Regression techniques may be used to do predictive analytics. Predictive analytics attempts to improve clinical outcomes by increasing illness diagnosis accuracy, patient care, resource optimization, and resource optimization (R et al. 2020). Machine learning is one of the most significant artificial intelligence aspects since it enables the construction of computer systems that can learn from previous experiences without the need for programming in every situation (R et al. 2020; Kumar, Suresh Kumar, and Pranavi 2017). This study focuses on developing a diabetes prediction model utilizing machine learning algorithms and data mining approaches.

Inefficient early diabetes diagnosis and human mistakes in existing diabetes detection techniques led me to undertake this research on increasing the accuracy of machine learning and diabetes prediction in pregnant women in the early stages. The current study's primary issue is that diabetes prediction algorithms are frequently inaccurate. The major goal is to use cutting-edge machine learning algorithms like (SVM) Support Vector Machine and Random Forest algorithms to test and evaluate diabetes deduction strategies.

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 study, using clinical.com, the sample size was estimated using prior study findings (An et al. 2020) with an alpha error-threshold of 0.05, enrollment ratio

of 0:1, 95 percent confidence interval with statistically significant $p < 0.05$, and power of 80 percent (Rout and Kaur 2020). Group 1 consisted of a Support Vector Machine algorithm ($N=20$) and Random Forest ($N=20$). A total of 40 samples were included in this study.

The data samples used in this study are collected from the Kaggle website. The data set undergoes data reduction techniques to obtain the absolute data required. The data should be given as input to MatLab 2021a to perform classification learning techniques. Input data should be imported to classification learning tools to perform training. The imported data is trained separately for each algorithm i.e, once for the Support Vector Machine varying with validations from 5 to 24 and similarly for the Random forest with validations from 5 to 24. After validation of data for an algorithm, the confusion matrix should be obtained for each validation, which involves the TP, TN, FP, FN expressed as true positive, true negative, false positive and false negative respectively. Precision, Sensitivity, and Accuracy are calculated with the help of these values. sed as.

The data under review is the Pima Indian Diabetes Dataset, which contains information from almost 760 medical records, 268 of which were positive and the rest were negative, all of which came from a community in Phoenix, Arizona, USA. The patient is diabetic if the result is positive, and the patient is not diabetic if the result is negative. There are eight properties in each instance, all of which are numeric data types. These values include personal medical information as well as information gleaned from medical examination results. Number of times pregnant (preg), plasma glucose focus at 2h in an oral glucose resilience test (in addition to), Diastolic pulse (pres), Triceps skin-fold thickness (skin), 2-h serum insulin (Insu), Body mass record (BMI), Diabetes family work (Pedi), Age (Age), and Class variable are the finished subtleties of the qualities utilized in the dataset (Class).

Because the model's predictions are heavily reliant on data quality, preprocessing is a critical activity that should not be overlooked. Many filters in MATLAB offer preprocessing, and the most appropriate approaches are chosen for the optimization of the original dataset. The medical implications of each characteristic are first assessed in respect to diabetes mellitus (DM). The attribute "number of pregnancies" was found to have a model influence on the DM, thus it was converted to

a nominal value by allocating 0 for nonpregnant and 1 for pregnant. As a consequence, the data complexity was reduced to a minimum.

The erroneous and some missing values in the dataset, which are the major source of many non-correct outcomes in most trials, are detected and deleted. For example, the value for diastolic blood pressure and body mass index cannot be 0, and if it is in dataset 95, it implies that the true value is missing.

STATISTICAL ANALYSIS

The statistical software program utilized in this investigation was IBM SPSS 26.0.1. The independent sample T-Test was used to determine the mean, standard deviation, and standard error mean statistical significance between the groups, and 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 83.78 percent, mean value 0.8378, and standard deviation value 0.04104. While picture size is an independent variable, image accuracy is a dependent variable, image size is an independent variable (An et al. 2020).

RESULTS

Table 2 illustrates how to use Support Vector Machine algorithms (SVM) and Random Forest to predict diabetes among pregnant women. When comparing the performance of the SVM with the Random Forest algorithm exhibits better values in Accuracy, Sensitivity, and Precision. The accuracy rate of the SVM and Random Forest algorithms are shown in Table 2. When compared to Support Vector Machine, the accuracy, sensitivity, and precision rate of a Random Forest are greater as shown in Table 1a and Table 1b. Random Forest results have a 79.02% accuracy, 78.25% sensitivity, and precision of 83.78%, while Support Vector Machine (SVM) results have a 77.67% accuracy, 76.67% sensitivity, and 83.54% precision rate. Table 2 shows that Principal Component analysis has a lower error rate than in Support Vector machine.

Table 3 indicates that using the independent sample T-test, there appears to be a statistically insignificant difference $P=0.445$ ($p>0.05$) for accuracy, $P=0.041$ ($p<0.05$), for sensitivity, $P=0.616$ ($p>0.05$) for precision, $P=0.048$. In both techniques $P=0.445$ ($p>0.05$) and for accuracy, for sensitivity $P=0.616$ ($p>0.05$) with insignificant value. But both methods are getting Statistically significant value $P=0.041$ ($p<0.05$), for sen-

Table 1a

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 1b

Diabetes prediction samples using Random Forest Algorithm

Samples	Accuracy	Sensitivity	Precision
1	0.81	0.80	0.85
2	0.82	0.80	0.8
3	0.77	0.7	0.8
4	0.77	0.76	0.83
5	0.78	0.77	0.85
6	0.81	0.81	0.83
7	0.77	0.76	0.83
8	0.78	0.78	0.83
9	0.75	0.7	0.8
10	0.78	0.78	0.83
11	0.8	0.8	0.83
12	0.82	0.80	0.8
13	0.8	0.77	0.8
14	0.82	0.80	0.8
15	0.75	0.77	0.77
16	0.71	0.71	0.71
17	0.77	0.76	0.83
18	0.82	0.79	0.89
19	0.78	0.78	0.83
20	0.8	0.8	0.8
21	0.82	0.81	0.84
22	0.83	0.79	0.82

Table 2

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

GROUP STATISTICS					
PARAM-ETERS	GROUP	N	MEAN	STD. DEVIATION	STD. ERROR MEAN
ACCU-RACY	RANDOM FOR-EST	20	0.7767	0.02248	0.00503
	SUPPORT VEC-TOR MACHINE	20	0.7902	0.02720	0.00608
SENSI-TIVITY	RANDOM FOR-EST	20	0.7667	0.01721	0.00385
	SUPPORT VEC-TOR MACHINE	20	0.7825	0.02559	0.00572
PRECI-SION	RANDOM FOR-EST	20	0.8354	0.03235	0.00723
	SUPPORT VEC-TOR MACHINE	20	0.8378	0.04104	0.00918

Table 3

Independent sample T-test in predicting the accuracy, sensitivity, and precision of diabetes using the Support Vector Machine and Logistic Regression. There appears to be a statistically significant difference ($p < 0.05$) for sensitivity and insignificant difference in both the methods for Accuracy and precision with $p > 0.05$

Parameter	Equal Vari-ances	Levene's Test for Equality of Variances		T-test for Equality of Means					
		F	Sig	t	df	Significance (one-Sided p)	Mean Dif-ference	95% Confi-dence interval (Lower)	95% Confi-dence interval (Upper)
Accuracy	Assumed	0.595	0.44	-1.71	38	0.048	-.01349	-.02946	.00248
	Not assumed			-1.71	36.70	0.048	-.01349	-.02946	.00250
Sensitivity	Assumed	4.461	0.04	-2.28	38	0.014	-.01572	-.02968	-.00176
	Not assumed			-2.28	33.26	0.015	-.01572	-.02975	-.00169
Precision	Assumed	0.255	0.61	-0.20	38	0.420	-.00238	-.02604	.02127
	Not assumed			-0.20	36.03	0.420	-.00238	-.02608	.02132

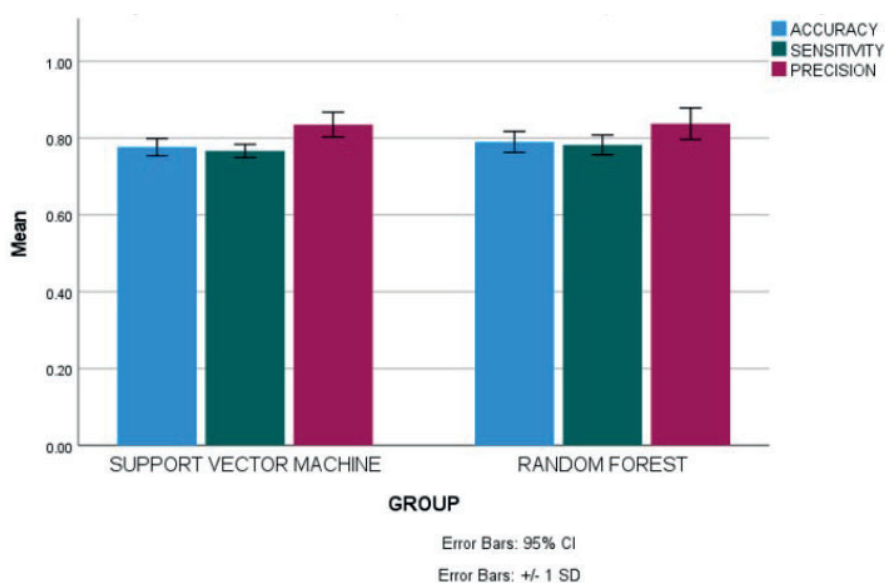


Fig. 1. Graph representing the comparison between mean accuracy, sensitivity, and precision of Diabetes prediction with the Support Vector Machine algorithm and the Random Forest algorithm. Both the techniques appear to create the same variable results with accuracy ranging from 76% to 84%. X-axis: Support Vector Machine vs Random Forest. Y-axis: mean accuracy, sensitivity, and precision detection ± 1 SD.

sitivity. These findings indicated that the Random Forest method outperforms the Support Vector Machine in predicting Diabetes illness. Figure 1 shows a bar chart depicting the comparison of Support Vector Machine and Random Forest 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 Support Vector Machine and Random Forest.

DISCUSSION

When compared to SVM accuracy (77.67%), sensitivity (76.67%), and precision (83.54%), Random Forest performed better with accuracy (79.02%), sensitivity (78.25%), and precision (83.78%) in this research for predicting diabetes among pregnant women. The substantial difference appears to have somewhat

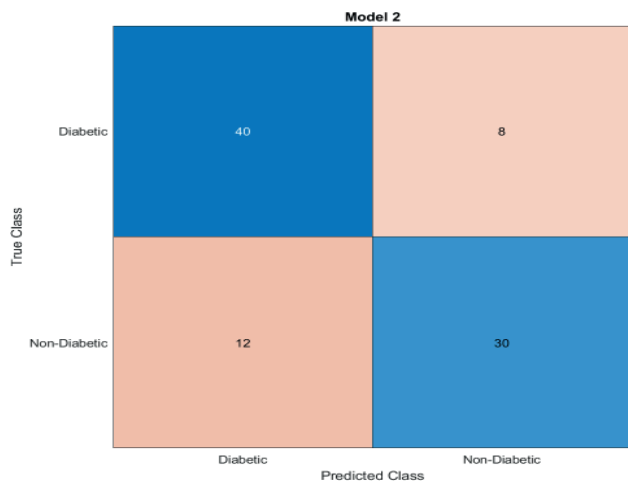


Fig. 2a. Confusion matrix for Support Vector Machine Algorithm K=5. True Positive is found to be 40% and false positive is found to be 8%, true negative is found to be 30% and false negative is found to be 12%.

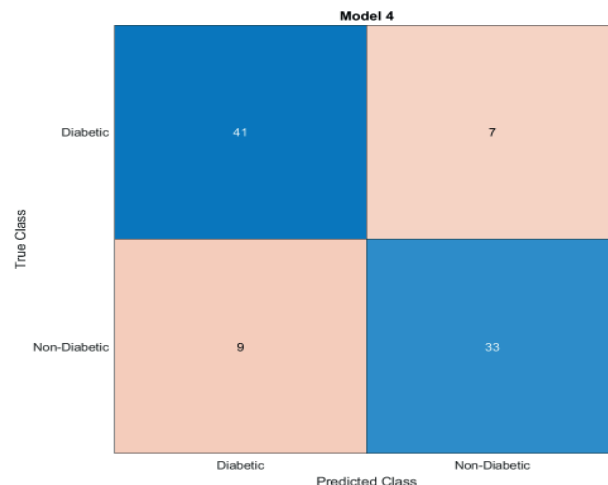


Fig. 2b. Confusion matrix for Random Forest algorithm K=5. True Positive is found to be 41% and false positive is found to be 7%, true negative is found to be 33% and false negative is found to be 9%.

grown, despite the fact that it is not statistically significant (Table 3). Early diagnosis of diabetes relies heavily on machine learning techniques.

Many researchers have suggested a model based on ensemble approaches utilizing machine learning algorithms, with the main goal of evaluating the model in terms of accuracy, precision, and sensitivity. The researchers used SVM and Random forest algorithms to predict diabetes with the SVM model having 79% (Shafi and Ansari, n.d.). (Mujumdar and Vaidehi 2019) did other research, and developed a computer-assisted identification approach based on Random Forest and Support vector machine classifiers (Pobi 2006). (Sohail et al. 2019) According to the findings, the Random forest model had the best accuracy (79.2%), sensitivity (79.49%), and specificity (81.4%) of three algorithms (Ayyadevara and Kishore Ayyadevara 2018). (Lyngdoh, Choudhury, and Moulik 2021) machine-learning techniques SVM and Gradient Boosted Decision Tree were used to build this ensemble model, which has an accuracy of 83.3 percent.

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 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. The proposed approach, in

conjunction with the recommended Machine learning classification algorithms, may be useful in the prediction or diagnosis of novel disorders in the near future. The study work, as well as a few other Machine learning techniques, might be updated and enhanced for diabetes prediction analysis. In future research, metaheuristic algorithms will be used to entirely learn the missing data. The algorithms have been improved to learn how to anticipate missing data in the future. Learning algorithms like Grey Wolf Optimizer (GWO) and other nature-inspired computer algorithms can help research with unique swarm-based meta-heuristic features. Furthermore, by gathering data from a variety of locations throughout the world and developing a more precise and common discriminating framework, the study may be expanded to predict diabetes. The work might be changed and enhanced to make the diabetes analysis more automated.

CONCLUSION

The Matlab-based Random Forest (79.02) outperformed the Support Vector Machine in this diabetes prediction study (77.67). Furthermore, unlike previous approaches, the algorithm's performance improved as the amount of data increased. This model is quite efficient and shows a lot of potential in terms of predicting and assessing diabetes, so it may be utilized in hospitals and testing facilities.

DECLARATION

Conflicts of Interest

No conflict of interest in this manuscript

Author Contributions

Author VSKP was involved in data collection, data analysis & manuscript writing. The author guide 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. An, Chansik, Hyunsun Lim, Dong-Wook Kim, Jung Hyun Chang, Yoon Jung Choi, and Seong Woo Kim. 2020. "Machine Learning Prediction for Mortality of Patients Diagnosed with COVID-19: A Nationwide Korean Cohort Study." *Scientific Reports* 10 (1): 18716.
2. Antink, Christoph Hoog, Joana Carlos Mesquita Ferreira, Michael Paul, Simon Lyra, Konrad Heimann, Srinivasa Karthik, Jayaraj Joseph, et al. 2020. "Fast Body Part Segmentation and Tracking of Neonatal Video Data Using Deep Learning." *Medical & Biological Engineering & Computing* 58 (12): 3049–61.
3. Ayyadevara, V. Kishore, and V. Kishore Ayyadevara. 2018. "Random Forest." *Pro Machine Learning Algorithms*. https://doi.org/10.1007/978-1-4842-3564-5_5.
4. Chellapa, L. R., S. Rajeshkumar, M. I. Arumugham, and S. R. Samuel. 2020. "Biogenic Nanoselenium Synthesis and Evaluation of Its Antimicrobial, Antioxidant Activity and Toxicity." *Bioinspired Biomimetic and Nanobiomaterials*, July, 1–6.
5. Costea, Naomi Estera, Elisa Valentina Moisi, and Daniela Elena Popescu. 2021. "Comparison of Machine Learning Algorithms for Prediction of Diabetes." 2021 16th International Conference on Engineering of Modern Electric Systems (EMES). <https://doi.org/10.1109/emes52337.2021.9484116>.
6. 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>.
7. Kaja, Rekha, Anandh Vaiyapuri, Mohamed Sherif Sirajudeen, Hariraja Muthusamy, Radhakrishnan Unnikrishnan, Mohamed Waly, Samuel Sundar Doss Devaraj, Mohamed Kotb Seyam, and Gopal Nambi S. 2020. "Biofeedback Flutter Device for Managing the Symptoms of Patients with COPD." *Technology and Health Care: Official Journal of the European Society for Engineering and Medicine* 28 (5): 477–85.
8. Kalidoss, Ramji, Snehalatha Umapathy, and Usha Rani Thirunavukkarasu. 2021. "A Breathalyzer for the Assessment of Chronic Kidney Disease Patients' Breathprint: Breath Flow Dynamic Simulation on the Measurement Chamber and Experimental Investigation." *Biomedical Signal Processing and Control* 70 (September): 103060.
9. Khanam, Jobeda Jamal, and Simon Y. Foo. 2021. "A Comparison of Machine Learning Algorithms for Diabetes Prediction." *ICT Express*. <https://doi.org/10.1016/j.ict.2021.02.004>.
10. Kumar, P. Suresh, P. Suresh Kumar, and S. Prana-vi. 2017. "Performance Analysis of Machine Learning Algorithms on Diabetes Dataset Using Big Data Analytics." 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS). <https://doi.org/10.1109/ictus.2017.8286062>.
11. Lavanya, M., P. Muthu Kannan, and M. Arivalagan. 2021. "Lung Cancer Diagnosis and Staging Using Firefly Algorithm Fuzzy C-Means Segmentation and Support Vector Machine Classification of Lung Nodules." *International Journal of Biomedical Engineering and Technology* 37 (2): 185.
12. Lyngdoh, Arwatki Chen, Nurul Amin Choudhury, and Soumen Moulik. 2021. "Diabetes Disease Prediction Using Machine Learning Algorithms." 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). <https://doi.org/10.1109/iecbes48179.2021.9398759>.
13. Malaikolundhan, Harikrishna, Gowsik Mookkan, Gunasekaran Krishnamoorthi, Nirosha Matheswaran, Murad Alsawalha, Vishnu Priya Veeraraghavan, Sura-

- paneni Krishna Mohan, and Aiting Di. 2020. "Anticarcinogenic Effect of Gold Nanoparticles Synthesized from Albizia Lebbeck on HCT-116 Colon Cancer Cell Lines." *Artificial Cells, Nanomedicine, and Biotechnology* 48 (1): 1206–13.
14. 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>.
 15. Paul, M., S. Karthik, J. Joseph, M. Sivaprakasam, J. Kumutha, S. Leonhardt, and C. Hoog Antink. 2020. "Non-Contact Sensing of Neonatal Pulse Rate Using Camera-Based Imaging: A Clinical Feasibility Study." *Physiological Measurement* 41 (2): 024001.
 16. Pobi, Shibendra. 2006. A Study of Machine Learning Performance in the Prediction of Juvenile Diabetes from Clinical Test Results.
 17. Raj R, Kathiswar, Ezhilarasan D, and Rajeshkumar S. 2020. "β-Sitosterol-Assisted Silver Nanoparticles Activates Nrf2 and Triggers Mitochondrial Apoptosis via Oxidative Stress in Human Hepatocellular Cancer Cell Line." *Journal of Biomedical Materials Research. Part A* 108 (9): 1899–1908.
 18. Ramadoss, Ramya, Rajashree Padmanaban, and Balakumar Subramanian. 2022. "Role of Bioglass in Enamel Remineralization: Existing Strategies and Future Prospects-A Narrative Review." *Journal of Biomedical Materials Research. Part B, Applied Biomaterials* 110 (1): 45–66.
 19. Rodriguez-Romero, Violeta, Richard F. Bergstrom, Brian S. Decker, Gezim Lahu, Majid Vakilynejad, and Robert R. Bies. 2019. "Prediction of Nephropathy in Type 2 Diabetes: An Analysis of the ACCORD Trial Applying Machine Learning Techniques." *Clinical and Translational Science* 12 (5): 519–28.
 20. Rout, Madhusmita, and Amandeep Kaur. 2020. "Prediction of Diabetes Risk Based on Machine Learning Techniques." 2020 International Conference on Intelligent Engineering and Management (ICIEM). <https://doi.org/10.1109/iciem48762.2020.9160276>.
 21. R, Sripreetha K., K. R. SriPreetha, N. Yuvaraj, and G. Jenifa. 2020. "Ensemble Machine Learning Approach for Diabetes Prediction." *Innovations in Information and Communication Technology Series*. https://doi.org/10.46532/978-81-950008-1-4_105.
 22. Shafi, Salliah, and Gufran Ahmad Ansari. n.d. "Early Prediction of Diabetes Disease & Classification of Algorithms Using Machine Learning Approach." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3852590>.
 23. Shilpa-Jain, D. P., Jogikalmat Krithikadatta, Dinesh Kowsky, and Velmurugan Natanasabapathy. 2021. "Effect of Cervical Lesion Centered Access Cavity Restored with Short Glass Fibre Reinforced Resin Composites on Fracture Resistance in Human Mandibular Premolars- an in Vitro Study." *Journal of the Mechanical Behavior of Biomedical Materials* 122 (October): 104654.
 24. Sohail, Muhammad Noman, Ren Jiadong, Musa Muhammad Uba, Muhammad Irshad, Wasim Iqbal, Jehangir Arshad, and Antony Verghese John. 2019. "A Hybrid Forecast Cost Benefit Classification of Diabetes Mellitus Prevalence Based on Epidemiological Study on Real-Life Patient's Data." *Scientific Reports* 9 (1): 10103.
 25. S, Sudha, Kalpana R, and Soundararajan P. 2021. "Quantification of Sweat Urea in Diabetes Using Electro-Optical Technique." *Physiological Measurement* 42 (9). <https://doi.org/10.1088/1361-6579/ac1d3a>.
 26. Theerthagiri, Prasannavenkatesan, Ruby A. Usha, and J. Vidya. n.d. "Diagnosis and Classification of the Diabetes Using Machine Learning Algorithms." <https://doi.org/10.21203/rs.3.rs-514771/v3>.
 27. Wu, Shuang, Shanmugam Rajeshkumar, Malini Madasamy, and Vanaja Mahendran. 2020. "Green Synthesis of Copper Nanoparticles Using Cissus Vitifolia and Its Antioxidant and Antibacterial Activity against Urinary Tract Infection Pathogens." *Artificial Cells, Nanomedicine, and Biotechnology* 48 (1): 1153–58.
 28. Xie, Zidian, Olga Nikolayeva, Jiebo Luo, and Dongmei Li. 2019. "Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques." *Preventing Chronic Disease*. <https://doi.org/10.5888/pcd16.190109>.