*Grenze International Journal of Engineering and Technology, Jan Issue*

# Empathy of Diabetics through Supervised Machine Learning Models

Dileep Kumar Kadali[1], P.LV.D Ravi Kumar[2], V.Leela Prasad[3], Srinivasa Rao Dangeti[4] and Kondepudi Gayathri[5]
[1-5]Shri Vishnu Engineering College for Women, Bhimavaram, India-534202
Email: dileepkumarkadali@gmail.com, {plvdravikumarit, leelaprasadv, dsrinivasaraoit , 21b01a1280}@svecw.edu.in
ORCID: 0000-0003-4058-3215

***Abstract*— Present-era diabetes is the most dangerous problem for the majority of people. It will be affected irrespective of age fact. In this case, a person's lifestyle can estimate the next stage of affected by comparing it with diabetic patients' reports for future prediction on his diabetic affected stage is one of the approaches. This automation can be done using machine learning approaches with good prediction methodologies. We proposed a study to compare all classification models in machine learning for diabetes prediction using lifestyle data. Based on lifestyle-related and demographic characteristics variables predicting diabetes, all classification models in supervised machine learning. Models are evaluated performance metrics for all models to produce a result as accuracy, sensitivity, specificity, precision, F1 score, and ROC curve with time complexity. These machine-learning models assist medical institutions in identifying diabetes patients.**

***Index Terms*— Supervised Machine Learning, Classification, accuracy, sensitivity, specificity, precision, F1 score, and ROC.**

## I. INTRODUCTION

Machine Learning (ML) is one of the sub-parts of computerised reasoning AI (artificial intelligence) that arrangements with the manners by which machines gain as a matter of fact [1], [2], [9]. In any case, some researchers are of the assessment that the terms man-made intelligence and ML are indistinguishable due to the chance of gaining, which is the principal component of the element called keen framework [4], [7]. A nitty gritty meaning of the term machine learning was given as "a framework is said to have gained for a fact E as for a few classes of undertakings T and execution measure P, in the event that its presentation at task in T, as estimated by P, improves with experience E" [17], [23]. For a long time, the machine has tackled many refined and complex true issues in application regions, for example, promoting, business and retail applications, normal language handling, medical care, independent vehicle frameworks, wise robots, environmental change, picture handling, voice, and gaming, among others [10], [32]. Machine learning procedures have been utilized for the expectation and analysis of numerous illnesses like the Coronavirus pandemic, jungle fever, typhoid, coronary conduit infections, and diabetes mellitus, among others [2], [10], [19]. Machine Learning calculations are normally founded on the experimentation approach which is very inverse to traditional calculations that adhere to the programming directions in view of if-else choice explanations [6], [13]. Machine learning undertakings are ordered into four general classifications, in particular administered learning, unaided learning, dynamic learning and support learning [23], [34] [22]. Managed gaining deduces a capability from the named preparation information,

*Grenze ID: 01.GIJET.10.1.18*

unaided gaining induces a capability from unlabelled preparation information, and dynamic learning gathers a capability by picking the most useful example for marking to prepare the model. In contrast, support learning connects with a unique climate [20], [31]. The flowchart of the preparation interaction of machine learning errands incorporates directed learning, solo learning and dynamic realizing, while preparing information is named, the preparation cycle is called administered while in any case it is called an unaided preparation process [3]. Conversely, when both named and unlabelled information are utilized for the preparation handling, this preparation system is called semi-directed [21], [28], [16]. In any case, the errand of administered machine learning is the one most ordinarily utilized in genuine applications, particularly for the conclusion and expectation of sicknesses [18], [35]. In this review, the directed machine models were created involving managed machine learning calculations for the expectation of diabetes mellitus [5], [33]. Diabetes mellitus (DM), usually alluded to as diabetes, is a gathering of metabolic problems of carb digestion in which glucose is underutilized, delivering hyperglycaemia (expanded glucose focus in the blood) [12], [22], [24]. The significant side effects of DM incorporate incessant peeing, expanded thirst, and yearning. Whenever left untreated, diabetes can cause numerous confusions, like diabetic ketoacidosis and non-ketoacidosis hyperosmolar unconsciousness [25], [36]. The drawn-out inconveniences brought about by DM incorporate cardiovascular illnesses, strokes, persistent kidney disappointments, foot ulcers, and harm to the eyes and nerves [21]9. Diabetes happens because of either the pancreas not creating sufficient insulin or the cells of the body not answering as expected to the insulin delivered [5]. DM is perhaps one of the deadliest sicknesses on the planet, particularly in created countries [14]. It is, notwithstanding, turning out to be more widespread in emerging countries like Nigeria, while presenting a larger number of dangers to people in the last option than those in the previous [15], [37]. More than 415 million individuals were experiencing diabetes mellitus overall starting around 2015, 8.3% being essential for the grown-up populace, with equivalent rates in all kinds of people [21]. DM type 2 comprises roughly 90% of the cases. DM is assessed to have brought about 1.5-5.0 million passings every year in the time of 2012-2015 and it duplicates an individual's gamble of biting the dust [29], [30]. The quantity of individuals with diabetes is supposed to ascend to 592 million by 2035. DM is one of the developing general well-being worries in Nigeria [8], [24]. A long time back, South Africa and Ethiopia had a bigger number of instances of diabetes than Nigeria, yet presently Nigeria has the most noteworthy frequency of diabetes in sub-Saharan Africa [16]. In this review, the analytic dataset of DM type 2 was gathered from the Murtala Mohammed Medical clinic, Kano-Nigeria and used to foster the prescient managed machine model in light of calculated relapse, support vector machine, K-closest neighbour, arbitrary woods, guileless Bayes, and slope helping calculations [26] [33].

## II. Literature Review

Many works have been done utilizing regulated machine learning to construct prescient models in the medical care area to supplement and enhance crafted by well-being labourers throughout diagnosing numerous sicknesses. In crafted by, Coronavirus forecast models that utilization directed machine learning (ML) were created. The model was created in view of direct relapse (LR), support vector machine (SVM), least outright shrinkage and determination (Rope), and remarkable smoothing (ES) calculations [2], [9]. The review exhibited the ability of the administered machine learning calculations to foresee the quantity of impending Coronavirus patients that were impacted. A directed machine learning approach, which consolidated the conventional calculation and weighted K-closest neighbour (WKNN) calculations to foresee and group DM type 2 as per the presence or nonappearance of coronary conduit illness difficulties, was created in crafted [16], [32]. The managed machine learning prescient model for intense ischemic stroke post intra-blood vessel treatment was created in crafted by

. The model showed a promising exactness of the expectation and the concentrate additionally proposed a powerful learning model that might possibly streamline the choice cycle for clinical treatment and endovascular action in the administration of intense strokes [8]. In the work, the prescient-managed machine learning model for the expectation of post-acceptance hypotension was created [23], [37]. The consequence of the review showed that the achievement kept in expectation post-enlistment hypotension exhibits the capacity of administered machine learning models for prescient examination in the area of anesthesiology [21], [29]. The prescient model for hospitalization because of coronary illness was created involving administered machine learning calculations in the work. The dataset utilized for the improvement of the model was gathered from a metropolitan clinic in Boston and five models were created utilizing SVM, AdaBoost, LR, Gullible Bayes, and probability proportion test calculations [20]. A directed machine learning model for the fast location of intensity rate discontinuity and cardiovascular arrhythmias was created in the review [21]. An irregular woodland calculation and a dataset of 300 examples of arrhythmic, non-arrhythmic coronary conduit sickness, and people with next to no restoratively huge cardiovascular circumstances were utilized to foster a prescient model. The model was assessed with 104 autonomous cases and ended up being exceptionally proficient [7], [22]. In the work, directed machine learning was utilized to create a sub-atomic mark that can characterize metastatic hepatocellular carcinoma patients and distinguish qualities that were pertinent to the metastatic and endurance of the patients [30]. In the work, a survey of regulated AI for populace hereditary qualities was completed and the survey investigation discovered that there is a promising heading nearby [17]. The investigation further discovered that directed AI is a significant and underutilized procedure that has extensive potential for transformative genomics. A regulated machine-learning model for distinguishing proof of mosquitoes from the backscattered optical sign was created in the review [35], [16]. The review showed that the optical sensor combined with directed machine learning can be a reasonable elective means for observing the mosquito populace [8]. The prescient managed machine learning approach for the assessment of the gamble repeat in the beginning phases of oral tongue squamous cell carcinoma has been created in the work [19]. The review's consequence showed regulated ML's capacity to anticipate locoregional repeats [26], [27] [34]. Regulated machine learning calculations which incorporate help vector machines, straight discriminant examination, and K-closest neighbour calculations were utilized to distinguish dementia in the work [13]. The consequence of the review showed that the calculations are fit for anticipating dementia [14], [37].

## III. Proposed Model

One more mixture method of learning called semi-regulated learning includes a mix of both managed and solo learning over a dataset that comprises of little part of marked preparing models with residual larger part of the models being un-named. It unites the exceptionally productive and exact speculation capacity of directed calculations with sensible guesses given by solo ones. This approach is relevant in the event of datasets containing a huge number of preparing models with many elements where speculation over such information is typically undeniably challenging. In such a setting, managed learning is applied over the little part of the named preparing information and solo learning is utilized to bunch the un-marked preparing information into groups of comparable models. At last, the information acquired during managed learning is used to sum up over-grouped information. The new patterns of always expanding scale and intricacy of information have led to one more profoundly refined learning worldview called as profound learning. A famous sub-space of ML models along with creating better improvements of the forecasts. The idea driving profound learning calculations is to copy the abilities to learn of the human cerebrum which includes datasets and has picked up a ton of speed.

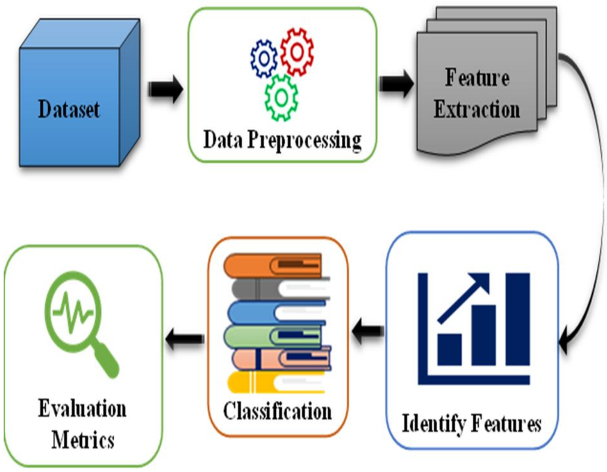


Figure 1. The Architecture of ML-based Diabetics Prediction

### *A. Data Sets*

Clinical choice emotionally supportive network alludes to electronic medical care frameworks that give help/help to clinicians, clinical staff or patients themselves to improve well-being and medical care benefits, and executed through present-day advancements like artificial intelligence, ML and information mining by taking advantage of clinical information. When applied for illness anticipation and the board, clinical choice help utilizing ML includes savvy information examination and deduction to help the doctors in clinical navigation by working with individualized risk profile study, customized medicines, prognostic displaying of patient's wellbeing state and so on. The overall design of the ML framework for offering such clinical choice help with normal constituent advances is portrayed as making sense of underneath

*B. Data Preprocessing*

Information obtaining Assortment of medical services information from accessible sources like emergency clinic data sets, online public stores given by research associations, legislatures or researchers, medical care studies and reviews information, information given by protection organizations, drug organizations and thusly. The idea of the information relies upon the issue and the goals of the review. EHR, clinical pictures from various modalities, remedies, and information gathered by clinical sensors are exceptionally well known and profoundly take advantage of wellsprings of medical care information in research concentrates on these days. Information pre-handling This alludes to handling crude info information and applying reasonable changes to refine it further for better or ideal outcomes. Clinical datasets are especially portrayed by heaps of missing qualities, anomalies as well and class unevenness. Disposal of such commotion is achieved by different pre-handling activities.

*C. Feature Extraction*

Highlight determination Elements are viewed as the structure blocks of any ML calculation, overall and are fundamentally the natural properties or attributes of the dataset itself. Highlight determination is a course of extricating a pertinent subset of elements that best address the different classes in the dataset as well as expand the exactness of the learning calculation. There are three classes of component determination methods a) Channel techniques Concentrate significant highlights in light of the innate, measurable properties of the information to get an element subset that is profoundly pertinent to the objective. Connection, shared data, and fluctuation are regularly utilized to highlight significance estimates in channel strategies. b) Covering strategies Include a pursuit through the element space and an order calculation to choose a component subset that outcomes in the most extreme grouping exactness. Dissimilar to channel techniques, coverings are viewed as profoundly reliant upon the characterization calculation. c) Inserted techniques Include a blend of channels along with covering strategies where the quest for the ideal element subset is a necessary piece of classifier preparation, as opposed to a free step.

*D. Identify the Feature*

Expectation All the more usually called as an arrangement in ML phrasing, it contains two sub-steps in particular preparing and testing. Here, the first dataset is partitioned into preparing and testing datasets. During preparation, the preparation dataset is given as the contribution to the learning calculation, which figures out how to sum up or foresee the results through design acknowledgement. Testing includes deciding the forecast execution of the gaining calculation over beforehand concealed models from the testing dataset. The clinical choice help given by ML prescient calculations incorporates a forecast of illness risk, type, progress, treatment, complexities, etc.

*E. Classification Models*

Approval Alludes to the most common way of approving the classifier execution for exactness and dependability of its expectations. Frequently, it is completed on an inward test dataset, got from the first dataset known as inside approval. At the point when an outer dataset, which isn't a piece of the information dataset is utilized to approve the outcomes, it is called outside approval. Approval is a fundamental stage, particularly in symptomatic clinical choice help applications assuming it is to be embraced in clinical settings.

*F. Confusion Matrix*

A confusion matrix is well-defined as a table that is frequently used to define the concept of a classification model proceeding an established of test data for which the true values are known. It is enormously convenient for measuring the precision, recall, AUC-ROC curves and accuracy.

TABLE I. TYPE SIZES FOR CAMERA-READY PAPERS

| | | ***Actual Values*** | |
|---|---|---|---|
| | | *Positive (1)* | *Negative (0)* |
| ***Predicted Values*** | *True (1)* | TP<br>True Positive | TN<br>True Negative |
| | *False (0)* | FP<br>False Positive | FN<br>False Negative |

*True Positive (TF):* The count value of predicted positive from the cross table and its value is true
*True Negative (TN):* The count value of predicted negative from the cross table and its value is true.
*False Positive (FP) (Type 1 Error):* The count value of predicted positive from the cross table and its value is false.

*False Negative (FN) (Type 2 Error):* The count value of predicted negative from the cross table and its value is false.

*G. Accuracy*

Accuracy purely procedures how frequently the model classifier appropriately expects the prediction, and accuracy is the proportion of the count of correct expects of the values of the prediction and the total count of overall prediction values.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

*H. Precision*

The precision value from the target is well-defined by way of the count of true positive values divided by the overall count of predicted positive values.

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

*I. Recall (Sensitivity)*

The recall value from the target is well-defined by way of the count of true positives divided by the total count of actual positive values.

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

*J. F1 Score*

The F1 score will give a united knowledge of recall and precision metrics. It is thoroughgoing that the minute precision is identically equal to recall.

$$\text{F1} = 2 \text{ X } \frac{\text{Precision X Recall}}{\text{Precision + Recall}} \quad (4)$$

## IV. Experiment Result

The dataset contains information about female Pima Indians aged 21 and older. It consists of various attributes such as age, number of pregnancies, BMI (Body Mass Index), blood pressure, skin thickness, insulin level, diabetes pedigree function, and whether or not the individual developed diabetes (binary classification). The dataset was originally collected from the National Institute of Diabetes and Digestive and Kidney Diseases and is available through various machine learning repositories, including the UCI Machine Learning Repository. The primary objective of this dataset is to develop machine learning models that can predict whether an individual has diabetes based on the provided attributes. The dataset is often used to practice and demonstrate data preprocessing, feature engineering, model selection, hyper parameter tuning, and evaluation techniques. It's a common dataset for teaching and learning machine learning concepts, especially binary classification. The dataset may have missing values and outliers, which need to be handled during data preprocessing. As with real-world medical data, the dataset might have some complexities and noise, which could impact model performance.

Steps to Work with the Dataset:

*Step-01:* Load the dataset and understand its structure.

*Step-02:* Preprocess the data, handling missing values and outliers.

*Step-03:* Split the dataset into features (attributes) and targets (diabetes classification) and Perform feature selection and engineering if needed.

*Step-04:* Choose a machine learning algorithm then Split the data into training and testing sets.

*Step-05:* Train the chosen model on the training data.

*Step-06:* Evaluate the model's performance on the testing data using appropriate metrics (accuracy, precision, recall, F1-score, etc.).

The diabetes dataset is a great resource for learning and practising machine learning concepts, particularly for binary classification tasks. It's important to approach the analysis ethically and responsibly, especially when dealing with medical data.

The following table shows metrics evaluation like accuracy, balanced accuracy, ROC AUC and F1 Score for all classification models in supervised machine learning models. These values are very helpful in predicting future values.

TABLE II. METRICS FOR CLASSIFICATION MODELS

| Model | Accuracy | Balanced Accuracy | ROC AUC | F1 Score |
|---|---|---|---|---|
| NuSVC | 0.78 | 0.75 | 0.75 | 0.78 |
| SVC | 0.78 | 0.75 | 0.75 | 0.77 |
| ExtraTreesClassifier | 0.78 | 0.75 | 0.75 | 0.77 |
| NearestCentroid | 0.73 | 0.74 | 0.74 | 0.73 |
| KNeighborsClassifier | 0.75 | 0.74 | 0.74 | 0.75 |
| LinearDiscriminantAnalysis | 0.76 | 0.73 | 0.73 | 0.76 |
| DecisionTreeClassifier | 0.74 | 0.73 | 0.73 | 0.75 |
| BaggingClassifier | 0.76 | 0.73 | 0.73 | 0.75 |
| CalibratedClassifierCV | 0.76 | 0.73 | 0.73 | 0.76 |
| LinearSVC | 0.76 | 0.73 | 0.73 | 0.75 |
| LogisticRegression | 0.76 | 0.73 | 0.73 | 0.75 |
| RidgeClassifierCV | 0.76 | 0.72 | 0.72 | 0.75 |
| RidgeClassifier | 0.76 | 0.72 | 0.72 | 0.75 |
| GaussianNB | 0.74 | 0.72 | 0.72 | 0.74 |
| AdaBoostClassifier | 0.74 | 0.72 | 0.72 | 0.74 |
| XGBClassifier | 0.73 | 0.71 | 0.71 | 0.73 |
| QuadraticDiscriminantAnalysis | 0.74 | 0.71 | 0.71 | 0.74 |
| RandomForestClassifier | 0.74 | 0.71 | 0.71 | 0.73 |
| BernoulliNB | 0.7 | 0.7 | 0.7 | 0.7 |
| LGBMClassifier | 0.72 | 0.7 | 0.7 | 0.72 |
| SGDClassifier | 0.7 | 0.68 | 0.68 | 0.71 |
| PassiveAggressiveClassifier | 0.7 | 0.67 | 0.67 | 0.7 |
| Perceptron | 0.69 | 0.66 | 0.66 | 0.69 |
| ExtraTreeClassifier | 0.67 | 0.64 | 0.64 | 0.67 |
| LabelSpreading | 0.63 | 0.62 | 0.62 | 0.64 |
| LabelPropagation | 0.63 | 0.62 | 0.62 | 0.64 |
| DummyClassifier | 0.64 | 0.5 | 0.5 | 0.5 |

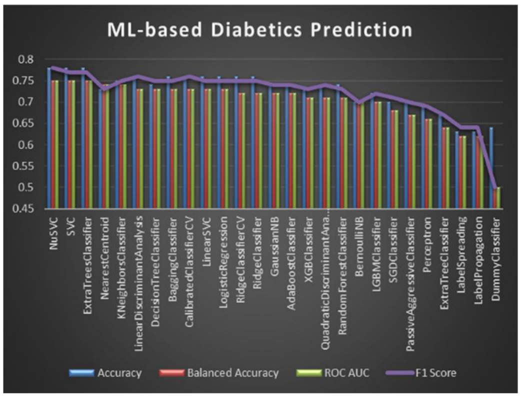


Figure 2. Evaluated the model's performance

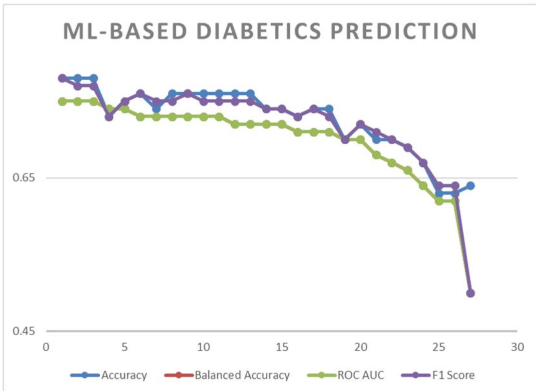


Figure 3. Comparison of evaluated the model's performance

## V. CONCLUSION

The paper focused on the automation of machine learning approaches with good prediction methodologies are compared all classification models for diabetes prediction based on lifestyle data. Finally, the models assist medical institutions in identifying diabetes patients using evaluated performance metrics for all models produced by a result as accuracy, sensitivity, specificity, precision, F1 score, and ROC curve with time complexity.


## ACKNOWLEDGEMENT

We are the authors of this paper DILEEP KUMAR KADALI, P.L.V.D RAVI KUMAR, SRINIVASA RAO DANGETI, V. LEELA PRASAD and KONDEPUDI GAYATHRI. We would like to deep gratitude to Padma Bhushan Dr. B.V. Raju, Founder Chairman of Sri Vishnu Educational Society. I would like to express our gratitude to Sri. K.V. Vishnu Raju, Chairman; Sri. R. Ravichandran, Vice-Chairman; Sri. K. Aditya Vissam, Secretary for providing an environment completed to writing this paper. We are very thankful to Dr. G. Srinivasa Rao, Principal;

Dr. P. Srinivasa Raju, Vice-Principal; My Dear Colleagues, and Students of Shri Vishnu Engineering College for Women (A)::Bhimarvaram, for their support in completing this paper.

REFERENCES

[1] U. P. Jyothi, M. Dabbiru, S. Bonthu, A. Dayal, and N. R. Kandula, "Comparative analysis of classification methods to predict diabetes mellitus on Noisy data," in Lecture notes in electrical engineering, 2023, pp. 301–313. doi: 10.1007/978-981-19-5868-7_23.

[2] Dileep Kumar Kadali, R. Mohan, N. Padhy, S. C. Satapathy, N. Salimath, and R. D. Sah, "Machine learning approach for corona virus disease extrapolation: A case study," International Journal of Knowledge-based and Intelligent Engineering Systems, vol. 26, no. 3, pp. 219–227, Dec. 2022, doi: 10.3233/kes-220015.

[3] Kadali, D. K., R. N. V. J. M. Naik M. Chandra, "Empirical Analysis on Uncertain Crime Data using Hybrid Approaches," Dec. 18, 2022. https://cims-journal.com/index.php/CN/article/view/527

[4] P. K. Vadrevu, M. R. M. Veeramanickam, S. K. Adusumalli, and S. K. Bunga, "Sign language recognition for needy people using machine learning model," in Smart Innovation, Systems and Technologies, 2022, pp. 227–233. doi: 10.1007/978-981-19-4162-7_22.

[5] D. K. Kadali, M. C. Naik, and N. V. J. M. Remani, "Estimation of data parameters using cluster optimization," in Lecture notes on data engineering and communications technologies, 2022, pp. 331–342. doi: 10.1007/978-981-19-2600-6_23.

[6] N. V. J. M. Remani, V. S. Naresh, S. Reddi, and D. K. Kadali, "Crime data optimization using neutrosophic logic based game theory," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, Mar. 2022, doi: 10.1002/cpe.6973.

[7] Ch. S. K. Raju, K. Pranitha, S. S. M, P. Samyuktha, and J. Madhumathi, "Prediction of COVID 19- Chest Image Classification and Detection using RELM Classifier in Machine Learning," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Mar. 2022, doi: 10.1109/icaccs54159.2022.9785131.

[8] D. K. Kadali and R. Mohan, "Shortest route analysis for High-Level Slotting using Peer-to-Peer," in Apple Academic Press eBooks, 2022, pp. 113–122. doi: 10.1201/9781003048367-10.

[9] A. Allen et al., "Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus," BMJ Open Diabetes Research & Care, vol. 10, no. 1, p. e002560, Jan. 2022, doi: 10.1136/bmjdrc-2021-002560.

[10] V. V. R. M. Rao, N. Silpa, G. Mahesh, and S. S. Reddy, "An enhanced machine learning classification system to investigate the status of micronutrients in rural women," in Lecture notes in networks and systems, 2022, pp. 51–60. doi: 10.1007/978-981-16-7118-0_4.

[11] N. Kari, S. K. Singh, and S. Velliangiri, "Various Machine Learning Techniques to Diagnose Alzheimer's Disease—A Systematic Review," in Lecture Notes in Electrical Engineering, 2022, pp. 557–567. doi: 10.1007/978-981-19-4364-5_40.

[12] L. Fregoso-Aparicio, J. Noguez, L. Montesinos, and J. A. García-García, "Machine learning and deep learning predictive models for type 2 diabetes: a systematic review," Diabetology & Metabolic Syndrome, vol. 13, no. 1, Dec. 2021, doi: 10.1186/s13098-021-00767-9.

[13] D. K. Kadali, R.N.V.J. Mohan, "Risk minimization Process on crime cluster data," Apr. 30, 2023. https://www.tijer.org/viewpaperforall?paper=TIJER2304443

[14] N. Silpa and V. V. R. M. Rao, "ENRICHED BIG DATA PRE-PROCESSING MODEL WITH MACHINE LEARNING APPROACH TO INVESTIGATE WEB USER USAGE BEHAVIOURY," Indian Journal of Computer Science and Engineering, vol. 12, no. 5, pp. 1248–1256, Oct. 2021, doi: 10.21817/indjcse/2021/v12i5/211205050.

[15] N. R. Eluri, G. R. Kancharla, S. Dara, and V. Dondeti, "Cancer data classification by quantum-inspired immune clone optimization-based optimal feature selection using gene expression data: deep learning approach," Data Technologies and Applications, vol. 56, no. 2, pp. 247–282, Sep. 2021, doi: 10.1108/dta-05-2020-0109.

[16] D. K. Kadali, D. Raju, and P. V. R. Raju, "Cluster query optimization technique using Blockchain," in Cognitive science and technology, 2023, pp. 631–638. doi: 10.1007/978-981-99-2742-5_65.

[17] M. Kavitha, P. V. V. S. Srinivas, P. S. L. Kalyampudi, C. S. F, and S. Srinivasulu, "Machine learning techniques for anomaly detection in smart healthcare," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Sep. 2021, doi: 10.1109/icirca51532.2021.9544795.

[18] S. Dara, S. Dhamercherla, S. S. Jadav, Ch. M. Babu, and M. J. Ahsan, "Machine Learning in Drug Discovery: A review," Artificial Intelligence Review, vol.55, no.3, pp. 1947–1999, Aug. 2021, doi: 10.1007/s10462-021-10058-4.

[19] D. K. Kadali, R. Mohan, and M. C. Naik, "Enhancing crime cluster reliability using neutrosophic logic and a Three-Stage model," Journal of Engineering Science and Technology Review, vol. 16, no. 4, pp. 35–40, Jan. 2023, doi: 10.25103/jestr.164.05.

[20] Y. Wang et al., "Genetic Risk score Increased discriminant efficiency of predictive models for Type 2 diabetes mellitus using machine learning: Cohort study," Frontiers in Public Health, vol. 9, Feb. 2021, doi: 10.3389/fpubh.2021.606711.

[21] H. Kaur and V. Kumari, “Predictive modelling and analytics for diabetes using a machine learning approach,” Applied Computing and Informatics, vol. 18, no. 1/2, pp. 90–100, Jul. 2020, doi: 10.1016/j.aci.2018.12.004.
[22] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Štiglic, “Early detection of type 2 diabetes mellitus using machine learning-based prediction models,” Scientific Reports, vol. 10, no. 1, Jul. 2020, doi: 10.1038/s41598-020-68771-z.
[23] B. M. K. P, S. P. R, R. K. Nadesh, and K. Arivuselvan, “Type 2: Diabetes mellitus prediction using Deep Neural Networks classifier,” International Journal of Cognitive Computing in Engineering, vol. 1, pp. 55–61, Jun. 2020, doi: 10.1016/j.ijcce.2020.10.002.
[24] R. García-Carretero, L. Vigil-Medina, I. Jiménez, C. Soguero-Ruíz, Ó. Barquero-Pérez, and J. Ramos-López, “Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population,” Medical & Biological Engineering & Computing, vol. 58, no. 5, pp. 991–1002, Feb. 2020, doi: 10.1007/s11517-020-02132-w.
[25] A. Aminian et al., “Predicting 10-Year risk of End-Organ complications of Type 2 Diabetes with and without metabolic surgery: a machine learning approach,” Diabetes Care, vol. 43, no. 4, pp. 852–859, Feb. 2020, doi: 10.2337/dc19-2057.
[26] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, “Discovering the type 2 diabetes in electronic health records using the Sparse Balanced Support Vector machine,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 1, pp. 235–246, Jan. 2020, doi: 10.1109/jbhi.2019.2899218.
[27] A. H. Syed and T. Khan, “Machine Learning-Based Application for Predicting Risk of Type 2 Diabetes Mellitus (T2DM) in Saudi Arabia: A Retrospective Cross-Sectional Study,” IEEE Access, vol. 8, pp. 199539–199561, Jan. 2020, doi: 10.1109/access.2020.3035026.
[28] A. P. Nguyen et al., “Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records,” Computer Methods and Programs in Biomedicine, vol. 182, p. 105055, Dec. 2019, doi: 10.1016/j.cmpb.2019.105055.
[29] Y. Liu et al., “Machine Learning For Tuning, Selection, And Ensemble Of Multiple Risk Scores For Predicting Type 2 Diabetes,” Risk Management and Healthcare Policy, vol. Volume 12, pp. 189–198, Nov. 2019, doi: 10.2147/rmhp.s225762.
[30] H. Lai, H. Huang, K. Keshavjee, A. Guergachi, and X. Gao, “Predictive models for diabetes mellitus using machine learning techniques,” BMC Endocrine Disorders, vol. 19, no. 1, Oct. 2019, doi: 10.1186/s12902-019-0436-6.
[31] A. Martínez-Millana, M. Argente, B. Valdivieso, V. T. Salcedo, and J. F. Merino-Torres, “Driving Type 2 Diabetes Risk Scores into Clinical Practice: Performance Analysis in Hospital Settings,” Journal of Clinical Medicine, vol. 8, no. 1, p. 107, Jan. 2019, doi: 10.3390/jcm8010107.
[32] S. Abhari, S. R. N. Kalhori, M. Ebrahimi, H. Hasannejadasl, and A. Garavand, “Artificial intelligence applications in Type 2 diabetes mellitus care: Focus on machine learning methods,” Healthcare Informatics Research, vol. 25, no. 4, p. 248, Jan. 2019, doi: 10.4258/hir.2019.25.4.248.
[33] N. Nirala, R. Periyasamy, B. K. Singh, and A. Kumar, “Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine,” Biocybernetics and Biomedical Engineering, vol. 39, no. 1, pp. 38–51, Jan. 2019, doi: 10.1016/j.bbe.2018.09.007.
[34] S. Perveen, M. Shahbaz, K. Keshavjee, and A. Guergachi, “Metabolic syndrome and development of diabetes mellitus: Predictive modeling based on machine learning techniques,” IEEE Access, vol. 7, pp. 1365–1375, Jan. 2019, doi: 10.1109/access.2018.2884249.
[35] D. K. Kadali, “Cluster optimization for similarity process using De-Duplication,” Sep. 01, 2016. https://ijsrd.com/Article.php?manuscript=IJSRDV4I60433
[36] A. Dayal, S. Gupta, S. Ponnada, and D. J. Hemanth, “Tornado forecast visualization for effective rescue planning,” in Springer eBooks, 2012, pp. 107–120. doi: 10.1007/978-3-030-84182-9_7.
[37] D. K. Kadali, J. Mohan, and Y. Vamsidhar, “Similarity based Query Optimization on Map Reduce using Euler Angle Oriented Approach,” https://www.ijser.org/, Jan. 2012,