Chapter 5

Diabetes Prediction Approaches: A Survey

SUJA A. ALEX, GERALD BRIYOLAN, GABRIEL CAUMO VAZ, AND GABRIEL GOMES DE OLIVEIRA

Worldwide, the disease diabetes has been growing a lot. Certain technical tools have been established, such as Machine learning techniques, which ensures the sign of diabetes in the past decades. Then, the approaches in the Deep Learning algorithm are the inspiration of many researchers to overcome the problem of complexity due to its increasing rate. The development of computer networks or cloud computing, which are empowered by e-Health care, developed the attainability of clinical data ease. This article is an overview of machine learning (ML) and deep learning (DL) approaches based on diabetes predictions.

5.1 Introduction

Involved in the quality of human lifestyle, diabetes has become most common in the everyday lives of people. Diabetes is habitually well known as diabetes mellitus, which is an incurable disease that rises the glucose levels in the blood because of a pancreas dysfunction that paves the way for the secretion of none or little insulin, which causes, respectively, type-1 and type-2 diabetes [1, 2]. Quite a few amounts of physical and chemical tests have been undergone to determine the circumstances. Globally recent studies conducted with the help of the World Health Organization (WHO) reveal enhancement in the numeral and mortality suffering from diabetes. The WHO anticipates that, by 2030, the major cause of death will be diabetes [3, 4, 5]. The Internal Diabetes Federation (IDF) has observed that across the world a total of 537 million persons are affected by diabetes and this is predicted to enhance to 643 million by 2030 [6, 7, 8]. The origin of diabetes is still mysterious, even though scientist believes that the ecological way of living is a factor that takes a significant part. Although diabetes is incurable, by undertaking proper treatment and medication it can be managed under control. The field of bioinformatics has tried to address this disease, which leads to the creation of new systems and tools. They either establish predicting models using various kinds of Machine Learning (ML), like association or classification algorithms [9, 10, 11].

In the manual diagnosis, the main advantage is that there is no need for any machine. However, in the worst case, at the initial stages of diabetes, the symptoms are low, and even a well-experienced doctor would not come up with a conclusion. Thus, this falls in the condition of the need for advanced technologies to detect the disease in it prime stage by utilizing a self-operating program which is more potential and efficient than manual diabetes acknowledgment [12, 13, 14]. The benefits of these technologies is that they reduce the workload for medical diagnosis, and human errors can be avoided. The decision support system established in the computers may perform a vital role and helps a tonne for effectual identification and benefits administration. The field of diabetes generates big data, which is on laboratory evaluation and includes reports of a person, treatment, check out, meds, and so on [15]. The data must be extracted and processed effectively and efficiently. An automated system that is capable of the identification of diabetes and which handles abnormality for superior intelligibility and dependability differentiates from manual spotting [16].

All machine Learning and Deep Learning methodologies have their convenience and own restrictions [17]. Artificial Neural Network (ANN) is a kind of machine learning feature with high performance, and high accuracy can be observed and obtained. The introduction of deep learning has improved ANN [18], and the present cases are deals with the past outcomes to make smart decisions with machine learning [19, 20].

5.2 Background

5.2.1 Diabetes Mellitus

If a person is affected by diabetes, the person's body does not produce the required amount of insulin. This results in an increase in glucose levels. The disorder is commonly known as high sugar and may spark severe or hazardous health concerns [21, 22, 23, 24].

Prediabetes

Prediabetes takes place whenever the blood sugar level rises above the normal level, it cannot be recognized by the doctor or the healthcare taker. In this case, there is a possibility of Type-II diabetes. Proper sporting will provide a better outcome in the reduction of excess calories in the body, and 5% - 7% of the body weight will decrease during prediabetes [25, 26, 27, 28].

Type-I Diabetes

Insulin-dependent diabetes, expressed as Type-I, is determined as Juvenile-onset Diabetes Mellitus, as it habitually happens in infancy. It is an autoimmune condition that happens because of the threat to the pancreas with antitoxin, which makes it not to produce insulin [29, 30, 31, 32]. This sort of diabetes may be a genetic disease [33, 34]. Major health syndrome that happens in type-I results in the disturbance in the narrow blood vessels in the diabetic nephropathy (kidney), diabetic retinopathy (eyes), and diabetic neuropathy (nerves)

[35, 36]. The person who is affected by Insulin-dependent diabetes is put at risk of developing cardiac and brain diseases [37, 38].

Type-II Diabetes

Non-insulin-dependent diabetes is also expressed as Type-II diabetes (adult-onset diabetes). In the last 20 years, type-II diabetes has become more common among children and teenagers [39, 40, 41, 42].

In Type-II diabetes, the pancreas will release some insulin even though it is inadequate or the insulin is unused. It is slightly related to type-I, and health issues might occur, especially in blood vessels of the neuro, kidneys, and eyes [43]. The related Type-II also highly has a chance of stroke and cardiac-related problems. Persons who are obese by 20% or above their preferable weight to their height will have a high chance of it and there might occur healthiness issues. Obesity includes insulin resistance. The pancreas must have to function forcefully to produce more amount of insulin [44, 45].

Alternative forms of Diabetes

This type is found in 1% - 5% of individuals, this is mainly due to pancreatic disorder, drugs, and illnesses. In this situation, the medical advisers must have close conduct with the patient, and the doctor must to keep a sharp eye on the sugar levels of the patient [46, 47, 48, 49].

The research on diabetes has achieved extensive substantial data above the antecedent decennary on:

- Etiology (the cause of a disease) or Pathology;
- Detection of Diagnosis;
- Unfolded, explained, and defined limits.

5.2.2 Machine Learning Techniques for Diabetic Prediction

Machine learning algorithms are well famed and best in the sector of medicine in the prediction of disease. Many researchers have utilized ML approaches to resolve diabetes effect as an advance to achieve better and most accurate results [50].

The multiple classifiers of Kandhasamy and Balamurali use Random Forest, KNN (K-Nearest Neighbour), Support Vector Machine, and J48. With the results obtained from the classifiers, the evaluation metrics are accuracy, sensitivity, and specificity. Two cases of classifications are done: with and without pre-processing and non-pre-processing, including the utilization of 5-fold cross-validation. However, pre-processing steps which are applied to the dataset are not explained by the researchers. It is just denoted that the error of noise is detected and removed from the data. The decision tree of the J48 classifier produced a maximum precision of 73.82% without pre-processing. KNN classifier in which value k = 1 and the maximum of 100% precision is obtained from the Random Forest for following the pre-processing of the data [51, 52].

An application for diabetes prediction is proposed by Yuvaraj and Sripreethaa utilizing Naïve Bayes, Random Forest, and decision tree, which are three different methods of ML algorithms. They adopted the dataset of Pima Indian Diabetes (PID) and pre-processed it. The discussion of the Information Gain method is being utilized for the aspect which is considered in the extraction of the applicable element. The consumption of the attributes is eight among a total of thirteen. Then, 70% of specifications are divided into training sets along with testing sets of 30%. The algorithm of Random Forest provides the highest precision of 94% [53, 54].

The SVM and Naïve Bayes are the advanced models which are submitted by Tafa et al to detect diabetes. The sets were evaluated using data from three locations in Kosovo. 80 of the 402 patients have type-II diabetes. Attributes that are utilized in the survey have not been investigated previously without including a balanced diet, exercises, and an account of patients' sugar levels. In the authentication process, the data has been split 50% for instruction and examine sets. A combination of methods that have been proposed eventually has enhanced the precision by 97.6% with accuracy of 95.52% and 94.52%, which were obtained by SVM and Naïve Bayes [55, 56].

The authors Deepti and Dilip utilized three different methods for predicting: SVM, Decision Tree, and Naïve Bayes. The major intention is to determine the classifiers for maximum precision. The dataset of Pima Indian is concluded for the research, and the partition is completed utilizing 10-fold cross-validation for the datasets. The execution was estimated with a measure of validity, recall, and F-measure. An accuracy of 76.3% was obtained using Naïve Bayes [57, 58].

Mercaldo et al. used a total of six classifiers for the prediction: J48, Hoeffding Tree, Multilayer Perceptron, Random Forest, JRip, and Bayes Net. The data set of Pima Indian was utilized in the research. They worked with Greedy Stepwise to indicate prejudice of the assign with an assist in the improvement of the classifiers execution, which includes the attributes of BMI, diabetes pedigree function, hemoglobin glucose concentration, and lifetime. Cross-validation of 10-fold is applicable in a set of data. Analysis and their contrast are made premised on precision, F-Measure, and recall. The outcome of correctness is 75.7%, recall equals 76.2%, and F-measure equals 75.9% [59, 60].

The authors Negi and Jaiswal use SVM as the main objective to predetermine diabetes. The datasets of Pima Indians and Diabetes 130-US are considered combinational data. Total samples of 102,538, 49 attributes, 64,419 positive samples, and 38,115 negative samples. The misplaced values are replaced by post-processing of statistics, and the statistics which are not in the scope with values of zero, the non-digital are converted into numeric significance. At last, the data is standardized between 0 and 1. The script of F is preferred in the LIBSVM package in which four attributes are chosen. The methods of Wrapper and Ranker are elected nine out of twenty attributes. The validation process of the 10-fold cross technique is used. With an accuracy of 72% the combined dataset is more reliable in diabetes prediction [61, 62, 63].

Olaniyi and Adnan have utilized the classifiers of Multi-layered Neural Networks of feed forwarding. Error correction methods are utilized for instruction purposes. A key objective is to improvise precision. The database of Pima Indian Diabetes has been utilized and the normalization has been done before processing the classifiers to attain numerical stability. Each of the attributes is divided by its amplitude to construct the data between the values of

0 and 1. Among the total dataset, 500 samples are training sets, and 268 as testing sets. The maximum accuracy is about 82% [64, 65].

Soltani and Jafarian apply the method of pattern recognition of PNN for pre-determination. The datasets of Pima Indian Diabetes were utilized. The technique of pre-processing is applied. In the total division, 90% is considered for the instruction set, and the leftover 10% is considered as the examined set. The accuracy is obtained at about 89.56% and 81.49% for both sets [66, 67].

Rakshit et al. have utilized the neural network of two-class for the prediction using datasets of Pima Indian. The data has been pre-processed and is obtained by the attributes which have been normalized in the sample which has been done by utilizing the average and the root-mean-square deviation, to gain numeric stability of each of the attributes. The extract of the applicable attributes makes use of the method of relation. The 314 are illustrative of the instruction set and 78 illustrative of the examine the coordinates. The highest accuracy which has been obtained in this method is 83.3% [68, 69, 70].

The researchers Mamuda and Sathasivam have used the three supervised instructions: Scaled Conjugate Gradient (SCG), Levenberg Marquardt (LM), and Bayesian Regulation (BR). The research contains 768 illustrative samples and eight attributes for estimating performance. The authentication of a 10-fold cross is being done to separate the data in instructions and examine coordinates. The best performance is obtained by the Levenberg Marquardt (LM) validation set which is obtained by Mean Squared Error (MSE) equal to 0.00025091 [71, 72, 73].

5.2.3 Deep Learning Techniques for Diabetic Prediction

Considering a large number of datasets, Deep Learning may be used to process them. This paves the way in the prediction of diabetes. Several studies proposed Deep Learning.

The Deep Neural Network (DNN) technique is used by the researcher Ashiquzzaman et al. The DNN architecture consists of General Regression Neural Network (GRNN), Multilayer Perceptron (MLP), and Radial Basis Function (RBF). The pre-processing is not done since the Deep Neural Network acts as a strainer of the statistics and can obtain biases. For the testing set and the training set the statistics are divided into 192 samples. Precision is obtained up to a percent of 88.41% [74, 75, 76, 77, 78].

A study was done by Swapna using two Deep Learning techniques to improve the exactness of the forecast. The Electrocardiogram is individual data that is utilized to appraise the execution of Convolutional Neural Network and Convolutional Neural Network-Long Short-term memory, which includes 142,000 samples and a total of eight attributes. The method of cross-validation of five-fold is being utilized and divided into testing along with priming sets. Due to the untutored Deep Neural Networks, the researchers neither include pre-processing data nor a selection method. The accuracy rates of the model are about 90.9% and 95.1%. Alex et al. put forward a method of DCNN (Deep Convolutional Neural Network) in the pre-determination of diabetes. Execution of classifications is set side by side with Machine and Deep Learning (ML and DL) models. Alex et al. prefer deep LSTM for diabetes prediction and got good classification results [79, 80].

5.2.4 Comparative Analysis of Diabetes Prediction

The comparative analysis is taken into account in this section and expressed in Table 5.1. Chatrati et al. used SVM, KNN, DT, and LR for the pre-determination of diabetes and hypertension on the statistics of Pima Indian Diabetes, and SVM secures high accuracy of 75% [41]. Maniruzzaman et al. used Linear Regression-Random Forest (LR-RF) as a combined feature selection, Decision Tree (DT), Naïve Bayes (NB), AdaBoost, and Random Forest (RF) on National Health of the Nutrition Examination Survey data which is to divine diabetes patients and bagged an accuracy of 94.25% [43]. Kumari et al. applied NB, RF, and LR on the PIDD dataset and got a 79.08% precise outcome on the statistics of PIMA [44]. Rajendra et al. used Logistic Regression on PIDD and Vanderbilt datasets with accuracy of 78% and 93% respectively [46, 81].

Table 5.1: Comparative analysis of existing classifiers.

Author / Year	Objective	Allocation used	Datasets	Evaluation framework
Chatrati et al., 2020 [41]	Pre-determine the existence of diabetes and high blood pressure	K-Nearest Neighbour, Vector Support Machine, Linear Regression, and Decision Tree	Pima Indian Diabetes Dataset (PIDD)	Accuracy and ROC curve
Maniruzzaman et al., 2020 [43]	Generate a structure that makes use of machine learning (ML) to divine diabetes	Linear Regression – Random Forest merging of the attributes, Naïve Bayes, Decision Tree, Random Forest, and AdaBoost	Internal Healthy and nourishment investigation	Accuracy and Area Under Cover (AUC)
Kumari et al., 2021 [44]	Improvise precision of detection of the diabetes mellitus utilizing a union of ML	Random Forest, Naïve Bayes, and Linear Regression	Pima Indian Diabetes Dataset (PIDD)	Accuracy, Precision, Recall, F1-score, and Area Under Cover (AUC)

Continued on next page

Table 5.1 – continued from previous page

Author / Year	Objective	Allocation used	Datasets	Evaluation framework
Rajendre et al., 2021 [46]	Generate a detection and look over various methodologies to improvise the presentation and precision	Linear Regression	Pima Indian Diabetes Dataset (PIDD) and Vanderbilt	Precision, Recall, and F1–score
Yadav et al., 2021 [47]	Utilize a classifier for predetermination	JRIP, Chi-Square for feature selection, OneR, Decision Tree, Bagging, and Boosting	UCI depository, and 9 features	Accuracy, Recall, extractions, and F1-Score
Goyal et al., 2021 [48]	The development in the Type-II pre-determine the enhanced execution for the premature diabetes	Using the crossvalidation of the 10-fold perspective and the ensemble	Pima Indian Diabetes Dataset (PIDD)	Accuracy
A. Prakash, 2021 [51]	Utilize the example of various machine learning techniques in the predetermination	J48, NB, RF, RT, and SimpleCART	Pima Indian Diabetes Dataset (PIDD)	Accuracy, computational time, Precision, F-Measure, ROC, and PRC

Continued on next page

Table 5.1 – continued from previous page

Author / Year	Objective	Allocation used	Datasets	Evaluation framework
Singh Ashima et al., 2021 [53]	To compare several classifiers and the features selection technique	SVM, NN, DT, RF, and XGBoost	Pima Indian Diabetes Dataset (PIDD)	Precision, vulnerability, Specificity, Precision < Area Under Cover (AUC), Gini Index, AUCH, minimum weighted coefficient, minimum error rate
Saxena et al., 2022 [55]	More accuracy to predict diabetes	MLP, DT, KNN, and RF	Pima Indian Diabetes Dataset (PIDD)	Sensitivity, Specificity, Accuracy, and Area Under Cover (AUC)
K. Hasan, 2021 [57]	Put forward a sturdy framework for prediction	K-Nearest Neighbour, Vector Support Machine, Decision Tree, Naïve Bayes, MLP, AdaBoost, and XGBoost	Pima Indian Diabetes Dataset (PIDD)	Sensitivity, Specificity, and Area Under Cover (AUC)
Tigga et al., 2022 [59]	Different kinds of ML methodologies were encompassed to pre-determine the risk of Type-II diabetes	Naïve Bayes and Random Forest	Pima Indian Diabetes Dataset (PIDD)	Accuracy, Precision, Recall, and F1-score

Continued on next page

Table 5.1 – continued from previous page

Author / Year	Objective	Allocation used	Datasets	Evaluation framework
Jashwath Reddy et al., 2022 [61]	The model with the highest degree of accuracy for the prediction of diabetes	K-Nearest Neighbour, Vector Support Machine, Linear Regression, GB, Naïve Bayes, and Random forest	Pima Indian Diabetes Dataset (PIDD)	Accuracy, ROC, Precision, Recall, and FM
Jackins et al., 2022 [64]	To come across a recognition of diabetes, coronary heart disease, and cancer among the available statistics	Random Forest and Naïve Bayes	Pima Indian Diabetes Dataset (PIDD)	Accuracy
Raghavendran et al., 2022 [68]	Analysis of statistics to determine the precision of Type-II diabetes	K-Nearest Neighbour, Vector Support Machine, Linear Regression, Random Forest, AdaBoost, and Naïve Bayes	Pima Indian Diabetes Dataset (PIDD)	Accuracy, Precision, Recall, F1-score, and CM
Laila et al., 2022 [71]	Increase the machine learning ensemble standard algorithms accuracy	AdaBoost, Bagging, and RF	Pima Indian Diabetes Dataset (PIDD)	Precision, Recall, Accuracy, F1-score

As a neural network of multilayer perceptron and convolutional neural networks, Mohebbi et al use logistic regression. Depending on the datasets obtained in continuous glucose monitoring (CGM) signals of the diabetes patients are monitored, including 9 patients, total days of 10,800 continuous glucose monitoring (CGM) data, which provides a total simulation of 97,200 continuous glucose monitoring (CGM). The training, validation, and testing dataset are split-up. Six patients are included in the training, validation, and others in testing sets. The highest accuracy is about 77.5% [82].

A proposed framework of unsupervised Deep Neural Networks done by Miotto et al. is known as Deep Patient. Electronic health records are used as the framework which includes a patient of 704,857. The prediction of different diseases can be done by using this method. During the process of validation, the statistics are divided into 5000 as testing illustrations and the remaining for training, 76,217 patients are included. The accuracy is determined by the Area Under Cover (AUC) and is measured at about 0.91. To enhance the prediction performance data must be pre-processed, and the extract of relevant attributes use PCA before performing the Deep Learning. [83]

Three different DL techniques are used by Pham et al. to collect the dataset manually from regional Australian hospitals. The 12,000 samples in the dataset contain 55.5% of males. They used some pre-processing techniques (not mentioned by them) to reduce the dataset to 7191 samples. 2/3 dataset is divided as the set of training, 1/6 as the set of authentications, and 1/6 as the test set. The methodologies which are used are Long-Short-Term Memory, Markov, and Plain RNN. To differentiate the execution of methodology, precision value is used. The LTSM method is used to attain the best precision of 59.6% [77].

The two different kinds of diabetes can be determined by the methodology of Recurrent Neural Network (RNN), and Ramesh et al. preferred it [84]. The 768 samples are utilized by the datasets of Pima Indian with eight attributes. According to research, "Glucose, BMI, Age, Diabetes Pedigree Function, Pregnancies, Skin Thickness, Insulin, and Blood Pressure" is used to indicate the attributes and it is ordered according to the highest importance. For better validation, the set of data is divided into 80%, and 20% as training and training sets, providing the precision of Type-I at 78% and Type-II at 81% [85].

The other studies which are conducted by Lekha and Suchetha used the one-dimensional (1-D) modified CNN to pre-determine diabetes, depending on breath signals. Data of breath signals are collected, which includes 11 healthy persons, nine patients affected by type-II diabetes, and five patients affected by type-I diabetes. Pre-processing is not done. The cross-validation of Leave-One Out is considered in the validation of data. Receiver Operating Characteristics (ROC) determine the actual performance and were evaluated, which reached 0.96 [86].

5.3 Conclusion

This study reviewed various ML and DL methodologies. Development of different classifiers and built-in new models to improvise accuracy. For more than six years, the classifiers of Machine Learning (ML) and Deep Learning (DL) reviewed their frequency of use. This work also focused on the datasets used. The existing works are compared based on various performance metrics.

Bibliography

[1] Yun-lei Sun and Da-lin Zhang. Machine learning techniques for screening and diagnosis of diabetes: a survey. *Tehnički vjesnik*, 26(3):872–880, 2019.

[2] Umm e Laila, Khalid Mahboob, Abdul Wahid Khan, Faheem Khan, and Whangbo Taekeun. An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors*, 22(14):5247, 2022.

[3] Parampreet Kaur, Neha Sharma, Ashima Singh, and Bob Gill. Ci-dpf: A cloud iot based framework for diabetes prediction. In *2018 IEEE 9th Annual Information Technology,*

Electronics and Mobile Communication Conference (IEMCON), pages 654–660. IEEE, 2018.

[4] Talha Mahboob Alam, Muhammad Atif Iqbal, Yasir Ali, Abdul Wahab, Safdar Ijaz, Talha Imtiaz Baig, Ayaz Hussain, Muhammad Awais Malik, Muhammad Mehdi Raza, Salman Ibrar, et al. A model for early prediction of diabetes. *Informatics in Medicine Unlocked*, 16:100204, 2019.

[5] Umm e Laila, Khalid Mahboob, Abdul Wahid Khan, Faheem Khan, and Whangbo Taekeun. An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors*, 22(14):5247, 2022.

[6] ID Federation. Idf diabetes atlas, tenth. *International Diabetes*, 2021.

[7] J Pradeep Kandhasamy and SJPCS Balamurali. Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, 47:45–51, 2015.

[8] Aiswarya Iyer, S Jeyalatha, and Ronak Sumbaly. Diagnosis of diabetes using classification mining techniques. *arXiv preprint arXiv:1502.03774*, 2015.

[9] Neha Sharma and Ashima Singh. Diabetes detection and prediction using machine learning/iot: A survey. In *Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2*, pages 471–479. Springer, 2019.

[10] Ankita Gupta and Rita Chhikara. Diabetic retinopathy: Present and past. *Procedia computer science*, 132:1432–1440, 2018.

[11] Goutham Swapna, Soman Kp, and Ravi Vinayakumar. Automated detection of diabetes using cnn and cnn-lstm network and heart rate signals. *Procedia computer science*, 132:1253–1262, 2018.

[12] Imran Qureshi, Jun Ma, and Qaisar Abbas. Recent development on detection methods for the diagnosis of diabetic retinopathy. *Symmetry*, 11(6):749, 2019.

[13] Julia Carracedo, Matilde Alique, Rafael Ramírez-Carracedo, Guillermo Bodega, and Rafael Ramírez. Endothelial extracellular vesicles produced by senescent cells: pathophysiological role in the cardiovascular disease associated with all types of diabetes mellitus. *Current Vascular Pharmacology*, 17(5):447–454, 2019.

[14] Nidhi Bansal. Prediabetes diagnosis and treatment: A review. *World journal of diabetes*, 6(2):296, 2015.

[15] Anastasia Katsarou, Soffia Gudbjörnsdottir, Araz Rawshani, Dana Dabelea, Ezio Bonifacio, Barbara J Anderson, Laura M Jacobsen, Desmond A Schatz, and Åke Lernmark. Type 1 diabetes mellitus. *Nature reviews Disease primers*, 3(1):1–17, 2017.

[16] Ralph A DeFronzo, Ele Ferrannini, Leif Groop, Robert R Henry, William H Herman, Jens Juul Holst, Frank B Hu, C Ronald Kahn, Itamar Raz, Gerald I Shulman, et al. Type 2 diabetes mellitus. *Nature reviews Disease primers*, 1(1):1–22, 2015.

[17] Jason Flannick, Stefan Johansson, and Pål R Njølstad. Common and rare forms of diabetes mellitus: towards a continuum of diabetes subtypes. *Nature Reviews Endocrinology*, 12(7):394–406, 2016.

[18] Rahul C Deo. Machine learning in medicine. *Circulation*, 132(20):1920–1930, 2015.

[19] N Yuvaraj and KR SriPreethaa. Diabetes prediction in healthcare systems using machine learning algorithms on hadoop cluster. *Cluster Computing*, 22(Suppl 1):1–9, 2019.

[20] Gabriel Gomes de Oliveira, Lucas Alves Rodrigues de Sá, Yuzo Iano, and Gabriel Caumo Vaz. Security in smart home using blockchain. pages 306–313, 2023.

[21] Madam Chakradar, Alok Aggarwal, Xiaochun Cheng, Anuj Rani, Manoj Kumar, and Achyut Shankar. A non-invasive approach to identify insulin resistance with triglycerides and hdl-c ratio using machine learning. *Neural Processing Letters*, pages 1–21, 2021.

[22] Deepti Sisodia and Dilip Singh Sisodia. Prediction of diabetes using classification algorithms. *Procedia computer science*, 132:1578–1585, 2018.

[23] Francesco Mercaldo, Vittoria Nardone, and Antonella Santone. Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. *Procedia computer science*, 112:2519–2528, 2017.

[24] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Pablo David Minango Negrete, and Daniel Rodrigues Ferraz Izario. Horizontal curves with transition. the use of this methodology for the calculation of a road project in the city of campinas/sp-brazil. pages 51–65, 2022.

[25] Anjli Negi and Varun Jaiswal. A first attempt to develop a diabetes prediction method based on different global datasets. In *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pages 237–241. IEEE, 2016.

[26] Souad Larabi-Marie-Sainte, Linah Aburahmah, Rana Almohaini, and Tanzila Saba. Current techniques for diabetes prediction: review and case study. *Applied Sciences*, 9(21):4604, 2019.

[27] Zahed Soltani and Ahmad Jafarian. A new artificial neural networks approach for diagnosing diabetes disease type ii. *International Journal of Advanced Computer Science and Applications*, 7(6), 2016.

[28] Juliana P da S. Ulian, Luiz Carlos Pereira da Silva, Gabriel Gomes de Oliveira, João Guilherme Ito Cypriano, Yuzo Iano, and Gabriel Caumo Vaz. Telemanagement and its benefits to energy, environment, and society: A case study in street lighting. pages 178–187, 2022.

[29] Somnath Rakshit, Suvojit Manna, Sanket Biswas, Riyanka Kundu, Priti Gupta, Sayantan Maitra, and Subhas Barman. Prediction of diabetes type-ii using a two-class neural network. In *Computational Intelligence, Communications, and Business Analytics: First International Conference, CICBA 2017, Kolkata, India, March 24–25, 2017, Revised Selected Papers, Part II*, pages 65–71. Springer, 2017.

[30] Mamman Mamuda and Saratha Sathasivam. Predicting the survival of diabetes using neural network. In *AIP Conference Proceedings*, volume 1870, page 040046. AIP Publishing LLC, 2017.

[31] Akm Ashiquzzaman, Abdul Kawsar Tushar, Md Rashedul Islam, Dongkoo Shon, Kichang Im, Jeong-Ho Park, Dong-Sun Lim, and Jongmyon Kim. Reduction of overfitting in diabetes prediction using deep learning neural network. In *IT Convergence and Security 2017: Volume 1*, pages 35–43. Springer, 2018.

[32] Lilian Regis Laraia, Yuzo Iano, Ricardo Takahira, Luiz Vicente Figueira de Mello Filho, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Technology for electric bus in the brazilian scenario: Focus on the adoption of national components. pages 276–285, 2022.

[33] Ali Mohebbi, Tinna B Aradottir, Alexander R Johansen, Henrik Bengtsson, Marco Fraccaro, and Morten Mørup. A deep learning approach to adherence detection for type 2 diabetics. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2896–2899. IEEE, 2017.

[34] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Pablo David Minango Negrete, and Juan Carlos Minango Negrete. Prop walls: A contextualization of the theme in a case study in the city of campinas (brazil). pages 128–139, 2022.

[35] Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6(1):1–10, 2016.

[36] Polyane Alves Santos, Yuzo Iano, Kelem Christine Pereira Jordão, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, Ingrid Araújo Sampaio, and Euclides Lourenço Chuma. Analysis of the relationship between maturity indicators using the multivariate linear regression: A case study in the brazilian cities. pages 203–210, 2022.

[37] Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of biomedical informatics*, 69:218–229, 2017.

[38] Ingrid Araújo Sampaio, Yuzo Iano, Aurelio Ribeiro Leite de Oliveira, Lino Marcos da Silva, Rinaldo Vieira da Silva Júnior, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Polyane Alves Santos, and Kelem Christine Pereira Jordão. The use of the elman preconditioner in the early iterations of interior point methods. pages 355–363, 2022.

[39] Sushant Ramesh, H Balaji, NCSN Iyengar, and Ronnie D Caytiles. Optimal predictive analytics of pima diabetics using deep learning. *International Journal of Database Theory and Application*, 10(9):47–62, 2017.

[40] Srinivasan Lekha and M Suchetha. Real-time non-invasive detection and classification of diabetes using modified convolution neural network. *IEEE journal of biomedical and health informatics*, 22(5):1630–1636, 2017.

[41] Saiteja Prasad Chatrati, Gahangir Hossain, Ayush Goyal, Anupama Bhan, Sayantan Bhattacharya, Devottam Gaurav, and Sanju Mishra Tiwari. Smart home health monitoring system for predicting type 2 diabetes and hypertension. *Journal of King Saud University-Computer and Information Sciences*, 34(3):862–870, 2022.

[42] Everton Hideo Nishimura, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Application and requirements of aiot-enabled industrial control units. pages 724–733, 2022.

[43] Md Maniruzzaman, Md Jahanur Rahman, Benojir Ahammed, and Md Menhazul Abedin. Classification and prediction of diabetes disease using machine learning paradigm. *Health information science and systems*, 8:1–14, 2020.

[44] Mukesh Kumar, Karan Bajaj, Bhisham Sharma, and Sushil Narang. A comparative performance assessment of optimized multilevel ensemble learning model with existing classifier models. *Big Data*, 10(5):371–387, 2022.

[45] Gabriel Caumo Vaz, Yuzo Iano, and Gabriel Gomes de Oliveira. Iot-from industries to houses: An overview. pages 734–741, 2022.

[46] Priyanka Rajendra and Shahram Latifi. Prediction of diabetes using logistic regression and ensemble techniques. *Computer Methods and Programs in Biomedicine Update*, 1:100032, 2021.

[47] Dhyan Chandra Yadav and Saurabh Pal. An experimental study of diversity of diabetes disease features by bagging and boosting ensemble method with rule based machine learning classifier algorithms. *SN Computer Science*, 2(1):50, 2021.

[48] Priyanka Goyal and Somil Jain. Prediction of type-2 diabetes using classification and ensemble method approach. In *2022 International Mobile and Embedded Technology Conference (MECON)*, pages 658–665. IEEE, 2022.

[49] Celso Fabricio Correia de Souza, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Valéria Sueli Reis, and Josué Mastrodi Neto. Institutional development index (idi): Calculation for municipalities in the metropolitan region of campinas (brazil). pages 245–255, 2022.

[50] Domingos Teixeira da Silva Neto, Jéssica Fernandes Alves, Polyane Alves Santos, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Yuzo Iano, and Lucas dos Santos Ribeiro. Proposal mppt algorithm using the kalman filter. pages 750–759, 2022.

[51] A Prakash et al. An ensemble technique for early prediction of type 2 diabetes mellitus–a normalization approach. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(9):2136–2143, 2021.

[52] Leonardo Bruscagini de Lima, Yuzo Iano, Pedro Y Noritomi, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Data security, privacy, and regulatory issues: A conceptual approach to digital transformation to smart cities. pages 256–263, 2022.

[53] Ashima Singh, Arwinder Dhillon, Neeraj Kumar, M Shamim Hossain, Ghulam Muhammad, and Manoj Kumar. ediapredict: an ensemble-based framework for diabetes prediction. *ACM Transactions on Multimidia Computing Communications and Applications*, 17(2s):1–26, 2021.

[54] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Pablo David Minango Negrete, and Juan Carlos Minango Negrete. Structural analysis of bridges and viaducts using the iot concept. an approach on dom pedro highway (campinas-brazil). pages 108–119, 2022.

[55] Roshi Saxena, Sanjay Kumar Sharma, Manali Gupta, and GC Sampada. A novel approach for feature selection and classification of diabetes mellitus: Machine learning methods. *Computational Intelligence and Neuroscience*, 2022, 2022.

[56] Leonardo Bruscagini de Lima, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Alecssander Daniel de Almeida, Gustavo Bertozzi Motta, Gabriel Matsumoto Villaça, Matias Oliveira Schwarz, and Pedro Y Noritomi. Mathematical modeling: A conceptual approach of linear algebra as a tool for technological applications. pages 239–248, 2022.

[57] Md Kamrul Hasan, Md Ashraful Alam, Dola Das, Eklas Hossain, and Mahmudul Hasan. Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access*, 8:76516–76531, 2020.

[58] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Pablo David Minango Negrete, Juan Carlos Minango Negrete, and Euclides Lourenço Chuma. Intelligent mobility: A proposal for modeling traffic lights using fuzzy logic and iot for smart cities. pages 302–311, 2022.

[59] Neha Prerna Tigga and Shruti Garg. Prediction of type 2 diabetes using machine learning classification methods. *Procedia Computer Science*, 167:706–716, 2020.

[60] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Roger Prior Gregio, and Alessandra Cristina Santos Akkari. Analysis of the ergonomic concept of public transportation in the city of campinas (brazil). pages 453–459, 2021.

[61] D Jashwanth Reddy, B Mounika, S Sindhu, T Pranayteja Reddy, N Sagar Reddy, G Jyothsna Sri, K Swaraja, K Meenakshi, and Padmavathi Kora. Predictive machine learning model for early detection and analysis of diabetes. *Materials Today: Proceedings*, 2020.

[62] Roger Prior Gregio, Yuzo Iano, Lia Toledo Moreira Mota, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, Diego Arturo Pajuelo Castro, and Carolina Fernandes Frangeto. Energy use in urban areas using neodymium magnets. pages 988–1005, 2021.

[63] Telmo Cardoso Lustosa, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, and Valéria Sueli Reis. Safety management applied to smart cities design. pages 498–510, 2021.

[64] V Jackins, S Vimal, Madasamy Kaliappan, and Mi Young Lee. Ai-based smart prediction of clinical disease using random forest classifier and naive bayes. *The Journal of Supercomputing*, 77:5198–5219, 2021.

[65] Euclides Lourenco Chuma, Yuzo Iano, Leonardo Lorenzo Bravo Roger, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Novelty sensor for detection of wear particles in oil using integrated microwave metamaterial resonators with neodymium magnets. *IEEE Sensors Journal*, 22(11):10508–10514, 2022.

[66] Daniel Katz Bonello, Yuzo Iano, Umberto Bonello Neto, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. A study about automated optical inspection: Inspection algorithms applied in flexible manufacturing printed circuit board cells using the mahalanobis distance method 1. pages 198–212, 2022.

[67] Y Thiagarajan, Baburao Pasupulati, Gabriel Gomes de Oliveira, Yuzo Iano, and Gabriel Caumo Vaz. A simple approach for short-term hydrothermal self scheduling for generation companies in restructured power system. pages 396–414, 2022.

[68] Ch V Raghavendran, G Naga Satish, NSL Kumar Kurumeti, and Shaik Mahaboob Basha. An analysis on classification models to predict possibility for type 2 diabetes of a patient. In *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021*, pages 181–196. Springer, 2022.

[69] Antonio Carlos Demanboro, David Bianchini, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. 6g networks: An innovative approach, but with many challenges and paradigms, in the development of platforms and services in the near future. pages 172–187, 2022.

[70] Daniel Izario, João Brancalhone, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, and Karine Izario. 5g-automation of vertical systems in the industry 4.0. pages 35–43, 2022.

[71] Umm e Laila, Khalid Mahboob, Abdul Wahid Khan, Faheem Khan, and Whangbo Taekeun. An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors*, 22(14):5247, 2022.

[72] Alex Restani Siegle, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Proposal of mathematical models for a continuous flow electric heater. pages 213–228, 2022.

[73] Y Thiagarajan, G Palanivel, ID Soubache, Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, and Himanshu Monga. Design and fabrication of human-powered vehicle-a measure for healthy living. pages 1–15, 2022.

[74] Suja A Alex, S Ponkamali, TR Andrew, NZ Jhanjhi, and Muhammad Tayyab. Machine learning-based wearable devices for smart healthcare application with risk factor monitoring. In *Empowering Sustainable Industrial 4.0 Systems With Machine Intelligence*, pages 174–185. IGI Global, 2022.

[75] Suja A Alex and J Jesu Vedha Nayahi. Deep incremental learning for big data stream analytics. In *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2018)*, pages 600–614. Springer, 2020.

[76] Suja A Alex, J Jesu Vedha Nayahi, H Shine, and Vaisshalli Gopirekha. Deep convolutional neural network for diabetes mellitus prediction. *Neural Computing and Applications*, 34(2):1319–1327, 2022.

[77] Leonardo Bruscagini de Lima, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Alecssander Daniel de Almeida, Gustavo Bertozzi Motta, Gabriel Matsumoto Villaça, Matias Oliveira Schwarz, and Pedro Y Noritomi. Mathematical modeling: A conceptual approach of linear algebra as a tool for technological applications. pages 239–248, 2022.

[78] Paolo Rodrigo de Oliveira Bacega, Yuzo Iano, Bruno Campos Simoni de Carvalho, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, and Euclides Lourenço Chuma. Study about the applicability of low latency in has transmission systems. pages 73–87, 2022.

[79] Suja A Alex, NZ Jhanjhi, Mamoona Humayun, Ashraf Osman Ibrahim, and Anas W Abulfaraj. Deep lstm model for diabetes prediction with class balancing by smote. *Electronics*, 11(17):2737, 2022.

[80] Pablo Minango, Yuzo Iano, Euclides Lourenço Chuma, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, and Juan Minango. Revision of the 5g concept rollout and its application in smart cities: A study case in south america. pages 229–238, 2022.

[81] Y Thiagarajan, Gabriel Gomes de Oliveira, Yuzo Iano, and Gabriel Caumo Vaz. Identification and analysis of bacterial species present in cow dung fed microbial fuel cell. pages 16–24, 2022.

[82] Juan Carlos Minango Negrete, Yuzo Iano, Pablo David Minango Negrete, Gabriel Caumo Vaz, and Gabriel Gomes de Oliveira. Sentiment and emotions analysis of tweets during the second round of 2021 ecuadorian presidential election. pages 257–268, 2022.

[83] Adolfo Blengini Neto, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Fabiana Silva Podeleski, Higor de Paula Kolecha, and Marcius FH de Carvalho. The bfs method in a cloud environment for analyzing distributed energy resource management systems. pages 349–362, 2022.

[84] Juan Carlos Minango Negrete, Yuzo Iano, Pablo David Minango Negrete, Gabriel Caumo Vaz, and Gabriel Gomes de Oliveira. Sentiment analysis in the ecuadorian presidential election. pages 25–34, 2022.

[85] Antonio Carlos Demanboro, David Bianchini, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Regulatory aspects of 5g and perspectives in the scope of scientific and technological policy. pages 163–171, 2022.

[86] Gabriel Gomes De Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Loureno Chuma, and Rangel Arthur. Intelligent transportation: Application of deep learning techniques in the search for a sustainable environment. pages 7–12, 2022.