INTERNATIONAL JOUNAL OF: AFRICAN SUSTAINABLE DEV. RESEARCH

TIMBOU-AFRICA ACADEMIC PUBLICATIONS
INTERNATIONAL JOURNAL, FEBRUARY, 2023 EDITIONS
VOL. 12 NO. 2 ISSN: 2067-4112

EXISTENTIAL RISK PREDICTION MODELS FOR DIABETES MELLITUS

MOKO, ANASUODEI; & AMANNAH, CONSTANCE IZUCHUKWU

Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Nigeria.

ABSTRACT

This paper reviewed, and analysed risk prediction models for variants of diabetes mellitus. The descriptive approach method was explored. It clearly described the various deep learning and machine learning risk prediction models. For diabetes mellitus classification and forecasting problems, Deep Neural Network Model algorithms have the highest score in terms of accuracy. The Deep Learning models outweighed machine learning

Introduction

Diabetes a multi-pronged locution for how your body transforms energy into food. It is regarded as a serious condition connected with an abdominal illness of the body whereby blood glucose levels have been incongruent because of a deficiency in the pancreas leading to low insulin or no insulin production at all, triggering insulin resistance to type 1 diabetes or to cells, causing type 2 diabetes. The major source of diabetes is still unclear, nevertheless scientists believe that diabetes is majorly ascribed to both genetic and environmental factors. It can be handled through treatment and medication, although it is incurable. Individuals with diabetes mellitus are vulnerable to some secondary health problems, for example heart disease and nerve damage. (Larabi-Marie-Sainte *et al.* 2019)

There are three major diabetes forms: type 1, type 2 and gestational diabetes. Type 2 diabetes among these 3 is the extremely common and accounts for 90% to 95% of all cases. Type 2 diabetes mellitus can be predicted, prevented, and reported based of lifestyle (for instance, low physical activity, obesity) and additional possibilities (for example, age, sex, race, family history) usually develop later in life. Many models for predicting Type 2 diabetes have been developed, (Xie *et al.* 2019). However, The Type 1 known as the aetiological type includes most cases, mainly due to pancreatic β - cells killing and ketoacidosis, its cases are attributed to an auto-immune

models in terms of performance. There's also the issue of other algorithms' precision. It is recommended that when conducting a classification and risk prediction survey, researchers should consider using the algorithms that explicitly describe performance while paying close attention to their advantages and disadvantages, as well as their potential outcomes. It is possible to combine deep learning techniques and these algorithms to create ensemble models, which will improve prediction performance.

Keywords: Diabetes, Type 1, Type 2, Gestational, Deep Neural Network, Machine Learning

process, or that are ketoacidosis-prone and have no known anethogenesis or pathogenesis.

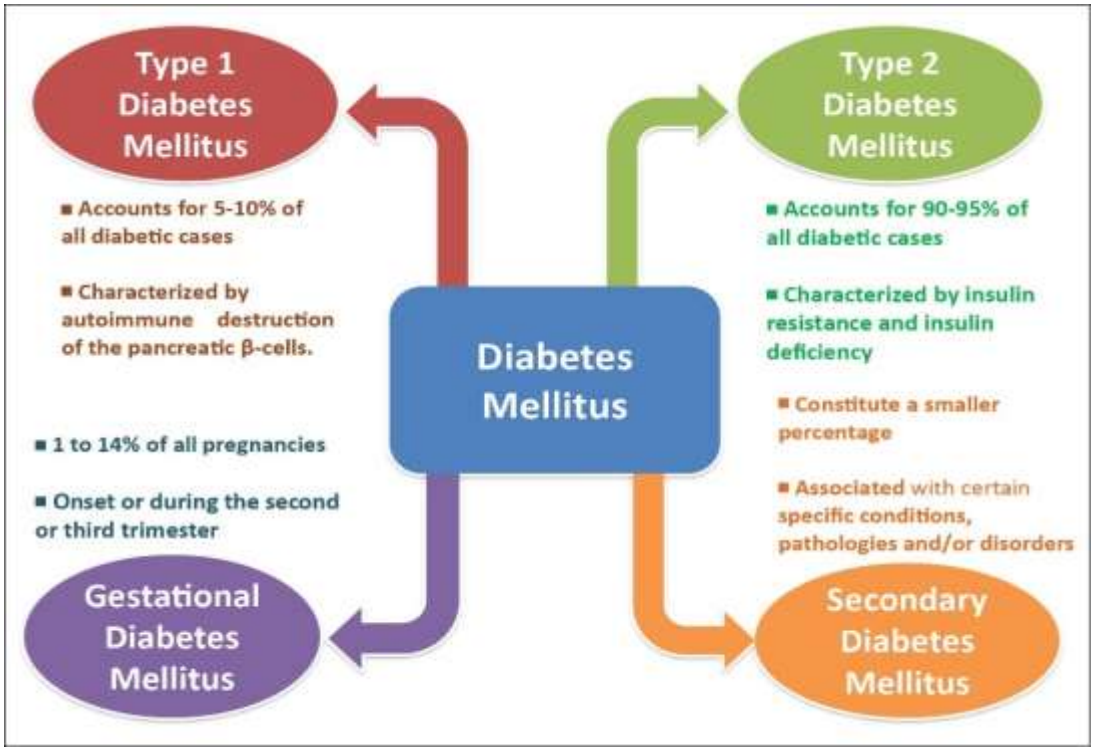

Figure 1. Classification of Diabetes Mellitus (Source: Banday et al 2020)

Diabetes in broad sense, and in specific diabetes in genetic mutations or related genetic abnormalities, proper and timely molecular diagnoses can help analyze disease risk and help predict the disease, and timely identify individuals who, family members, are at

greater risk for a disorder. In such cases, predictive molecular/genetic testing and preventive management can play a key role.

Problem Statement

Diabetes mellitus has progressive effects including the advanced growth of specific retinopathic complications which can cause probable blindness, renal failure-related nephropathy and/or foot ulcer risk neuropathy, amputation, Charcot joints and autonomous dysfunction, including sexual dysfunction. The risk of cardiovascular, peripheral vascular and brain vascular disease is increasing for people with diabetes, and as such this study looks at the different existing risk prediction models and recommends the suitable model for the prediction of diabetes mellitus.

Aim and Objectives

The study was desined to review a number of existential risk predictions models for diabetes mellitus.

The paper intended to:

i. review with comparative evaluation risk prediction models for the different variants of diabetes mellitus.
ii. recommend the most efficient model for the different variants of diabetes mellitus risk prediction.

Scope of the study

The study was limited to risk prediction models for predicting diabetes mellitus. It explored descriptive comparative technique. It was delineated to Type 1, Type 2, and Gestational Diabetes Mellitus.

Significance of the study

Diabetes Mellitus risk prediction models can be used for a range of things in medical practice, along with:

i. predicting the advancement of the distinct variants of diabetes mellitus.
ii. predicting therapeutic efficacy or patient prognosis of the disease.
iii. can commonly be used to predict the risk of adverse outcomes after intervention when performing surgeries.
iv. models of risk prediction are now a significant part of clinical decision-making, that will offer a fast and easy way to evaluate the risk of the patient from diabetes or event that can then direct the treatment.

Methodology

The method adopted in this study is exploratory descriptive approach. The study explores the existential predictive models and subsequently describe them analytically.

Review of Related works

Variety of predictive models have been developed for variants of diabetes mellitus. This section reviews the existing risk prediction models based on the different variants of diabetes mellitus.

Type 1 and Type 2

Xie *et al.,* (2019) study investigated on 138,146 people, including 20,467 people with type 2 diabetes, took part in a cross-sectional data experiment. Type 2 diabetes mellitus was predicted in the study using Support Vector Machine, Decision Tree, Neural Network, Logistic Regression, and Guassian Naive Bayes classifier models. To further investigate the probable key risk factors of diabetes, univariable and multivariable weighted logistic regression were also applied. Area under the curve ranged from 0.7182 to 0.7949, with neural network generating the highest accuracy score of 82.4%, as shown by the study's findings.

Mahboob Alam *et al*., (2019) projected that diabetes has important attributes, as is characterized by the relationship between different features. The range of tools are used for determining the range of significant attributes for clustering, prediction, and diabetes association rule mining. The main method of analysis of components was used to select significant attributes. Our results show a strong link between diabetes with BMI and glucose levels extracted via the Apriori procedure. A diabetes prevention method was developed. The ANN technique produced 75.7 percent more accurate and can be helpful in helping medical professionals make treatment decisions.

Deberneh and Kim (2021) developed a risk prediction model that will provide with outstanding prediction information of the possibility of T2D emergent. The result from the study shows that the ensemble models performed superiorly to that of the single models used.

Larabi-marie-sainte *et al.* (2019) study reviewed and recommended the usage of Machine Learning and Deep Learning techniques to solve diabetes mellitus, by reviewing and comparing the diabetes numerous prognoses that have been made in the last six years. The research paper obtained classifiers accuracy of 68% - 74% and recommended the use of the studied techniques classifiers for diabetes prediction and can be enhanced by developing combined models.

A survey by Goutham *et al*. (2020) presented a classification methodology for the use of deep learning architectures for diabetic and normal heart rate variability (HRV) signals, the study used a short-term long storage (LSTM) and a convolutionary neural network (CNN) to extract the complicated time dynamic characteristics of the input HRV. These features are them transferred to the classification support vector machine (SVM). In CNN and CNN architecture performance improvements are 0.03%, 0.06%. Lai *et al.* (2019)

evaluated and recommended two machine-learning models; Gradient Boosting Machine and Logistic Regression.

Zhou *et al.* (2020) built a risk prediction model for Type 1 and Type 2 diabetes mellitus with deep neural network hidden layers to prevent overfitting, using drop-out regularization model. The study adjusted several constraints and adopted the binary cross-entropy loss function, which gained a highly precise neural network prediction model. The outcome from the experiments demonstrates the efficiency and effectiveness of the intended DLPD model. The best exercise accuracy in the data series for diabetes type is 94.02.174%, and in Pima Indians, the exercise accuracy is 99. 4112%.

Recommending a new model for predicting type 2 diabetes mellitus based on data mining techniques for (T2DM). Wu *et al.* (2018) identified that the main problems associated with risk prediction are improving the precision of the prediction model and expanding the model's application by incorporating data from additional sources. The model is comprised of two components: an improved K-means algorithm and a logistic regression algorithm depending on a variety of preprocessing procedures. The suggested model was then tested on two additional diabetes datasets as part of the research.

Ayon and Islam (2019) investigating the distinctiveness of diabetes centered on numerous medical factors using engaged deep neural network techniques. The precision for five-fold cross validation was found to be 98.35 percent which is relatively high compared to other methods used to predict diabetes mellitus.

Abbas *et al.* (2020) created a non-linear SVM-based prediction model for identifying individuals at relatively high risk of developing type 2 diabetes. To build the model, the prediction power of features deriving from the OGTT information was first evaluated and personal data such as age, ethnicity and BMI were considered. The study showed that the properties derived from plasma glucose concentrations provide an excellent subset of properties and the toughest predictive power for the potential progress of the T2DM with a feature selection algorithm, (Kopitar *et al.* 2020). Moko and Onuodu (2021) reviewed several predictions and adjudged the SVM and Gradient Boosting of higher predictive accuracy score of 98%.

Gestational Diabetes

A new clinical and biomarker risk prediction model proposed by Sweeting *et al.* (2019) for the first quarter accurate stratification of risk for gestational diabetes mellitus (GDM) diagnosis in an Australian multiethnic population for early pregnancies. This approach to GDM risk management addresses the resource limits linked to increasing GDM prevalence by focusing on prevention and early intervention for women most at risk, with the help of multivariate regression, GDM was examined. ROC curves were used to evaluate the model's ability to forecast the diagnosis of GDM in the early stages of pregnancy (before 24 weeks of gestation) as well as the overall diagnosis of GDM. Only those women whose

risk is identified in the new prediction model may need to go to OGTT and initial diagnostic tests can be prioritized more correctly to better risk subcohorts, such as early risk at GDM, (Zhang *et al.* 2020).

Benhalima *et al.* (2020) prediction model was developed with easily available early pregnancy clinical variables using multivariable logistic regression analysis. Based on the WHO criteria for 2013, the prediction model had a moderately precise GDM. The proposed model for prediction can recognize women at peril of having GDM before or at early pregnancy, so that pregnancy outcomes can be improved promptly.

Diabetes Mellitus Risk Prediction Models

The speed of each machine learning and deep learning models was assessed based on the studies carried out on risk prediction for the different types of diabetes mellitus as shown in table 1 from 2017 till date. The commonest classifiers for risk prediction of diabetes shown from table 1 were deep neural network (ANN), for Deep Learning and for Machine Learning are support vector machine (SVM), Logistic Regression and Random Forest.

Table 1. Risk Prediction Model

S/N	Diabetes Type	Risk Prediction Model	Pros	Cons	Prospects
1.	Type 2	An improved K-means and Logistic Regression. *Wu et al. (2018)*	i. It avoids erasing too much original data. ii. The model can adapt to various datasets	spends additional time during the part of preprocessing.	To continuously train and optimize the proposed model with real and latest patient data to be provided by a hospital. The data set should be sufficiently large to train and predict to achieve an optimal result.
2.	Type 1 and Type 2	Deep learning predicting diabetes (DLPD) model, with best training accuracy score of 94.02174% *Zhou et al (2020)*	i. The model can project the future existence of diabetes in patients and the diabetes type experienced by a patient, type 1 or type 2. ii. The model can handle volume of data.	The presented model's precision has not been compared to those of any other models.	There is need to optimized the model to for automated diabetes analysis.
3.	Diabetes Mellitus	CNN 5-LSTM with SVM- 95.7% *Goutham et al (2018)*	non-intrusive, adaptable, and reproducible system can serve as a dependable tool to clinicians to detect diabetes.	Involves substantial amount of input data.	The deep learning approach must be improved by including omics data to predict the onset of diabetes.

4.	Type 2	Logistic Regression, Random Forest, XGBoost, Support Vector Machine, and ensemble classifiers (Classifier Integration Model (CIM), Soft Voting (SV) Algorithm) and Stacking Classifier (ST) *Deberneh and Kim (2021)*	i. it can be used to aid healthcare practices by enabling its practitioners and diabetes educators make better clinical decisions and extend the lives of their patients. ii. can help to minimize T2D risks, progression, and related complications both for clinical professionals and patients in advance.	i. Normal, prediabetes, and diabetes were all defined solely by FPG level; HbAIc and the oral glucose tolerance test (OGTT) were not included. ii. Supplementary data sources must be used to validate the models developed in this study.	the prediction models need to be made more straightforward to develop user experience for web and mobile applications.
5.	Diabetes Mellitus	Deep Neural Network with a prediction accuracy score of 98.35%. *Ayon and Islam (2019)*	Produced a more state of the art result, when compared to other models.	The precision of the output is determined by the strength of the link between units.	Shows the need to conduct anomaly detection and prediction that can be done from input data.
6.	Type 2	Glmnet (0.859%), RF, XGBoost, LightGBM *Kopitar et al (2020)*	The accessibility of more data correlates positively with improved predictive effectiveness and a more balanced weighting of variables is based on priority.	i. The population studied did not include older people whom T2DM is more predominant. ii. Since only one database with a relatively small sample size was used, a significant amount of data was left out.	The use of various ensemble construction techniques should be examined, as well as the stacking and blending of distinct prediction models.
7.	Diabetes Mellitus	Logistic Regression and Gradient Boosting Machine (GBM) techniques with AROC for the proposed GBM model is 84.7% with a sensitivity of 71.6%. *Lai et al (2019)*	i. Utilizes the adjusted-threshold and class weight methods to deal with the problem of imbalanced data. ii. The sensitivity and specificity for detecting people with diabetes mellitus in clinical settings are both substantial.	The proposed model by the study was compared to only 2 (Decision Tree and Random Forest) models	
8.	Diabetes Mellitus	RF (0.8857) J48, Neural Network, *Zou et al (2018)*	The study gives room to achieve maximum accuracy if all attributes are used.	The model cannot predict the type of	There is need to predict diabetes based on types and to

				diabetes, due to data used. There is data unbalance.	examine the proportion that may improve the accuracy of diabetes predictions.
9.	Diabetes Mellitus	Naive Bayes (NB), Artificial Neural Network (ANN), Decision Tree (DT) and Deep Learning (DL) 98.07% *Naz and Ahuja (2020)*	Provides health officials and specialists with structured diabetes knowledge.		Incorporating omics data for disease onset prediction will enhance the precision of the Deep Learning approach. To create a reliable application or website that uses the stated DL algorithm to assist healthcare professionals in the timely detection of diabetes.
10.	Gestational Diabetes	Combined multivariate prediction model with an overall score of 0.93% *Sweeting et al (2019)*	i. Preeclampsia and a huge multi-ethnic cohort with forward-looking information attained at first quarter screening. ii. The non-fasting measurement of biomarkers improves the viability of the model during pregnancy.	i. Possible bias in measuring the risk factors in maternal self-reporting of GDM. ii. There is need for external validation between the biomarkers and ethnicity. iii. The early GDM is sub cohort optimized model is derived from smaller samples, additional validation in larger sub cohorts is required.	Future validation in other cohorts should be explored. Intervention and or preventive techniques should be evaluated.
11.	Type 2	Ensemble learning method (RF-WFS and XGBoost 93.75% Accuracy) *Xu and Wang (2019)*	In light of the model's classification accuracy and performance, it's flexible enough to work with a variety of data.	Dataset used in not recent.	To verify the efficiency and performance of the proposed model, it is essential to analyze the most recent patient hospital data to be a model training and prediction dataset.
12.	Gestational Diabetes	Multivariable Logistic Regression Analysis *Benhalima et al (2020)*	The study used a large pregnant cohort with prospective data collection on obstetrical history, clinical risk factors,	Pregnancy weight was self-reported, and no information on biochemical risk factors was gathered prior to	More research is needed to diagnose and classify women at risk of GDM outlined in this model as earlier

			biochemical indicators, and comprehensive data on socioeconomic status, diet, and physical activity during early pregnancy.	the study's participants becoming pregnant.	than usual so as to minimize pregnancy outcomes.
13.	Diabetes Mellitus	Machine Learning (SVM, RF – 83.64%) and Deep Learning (CNN) Techniques. *Yahyoui et al (2019)*	It demonstrates the efficacy of DNN in assisting doctors in disease prediction.	No feature selection or calculation was done to ascertain feature importance.	Boost prediction accuracy by introducing an automatic deep feature method to the feature extraction step and by creating a better fitting model.
14.	Gestational Diabetes	Univariate analysis. The area under the curve value was 0.911 (95% CI, 0.893-0.930) *Zhang et al (2020)*	Women with GDM were found to be older, have higher BMIs, and had higher Hbalc levels than women without GDM.	The model's prediction begins in the second trimester but might not be generalizable.	Before the results may be applied to the broader public, large-scale multi-center clinical trials are essential. Pregnant rats on a high-fat diet may be utilized to research the link between liver function and adipose tissue.
15.	Type 2	Stepwise Logistic Regression, Penalized Regression and Support Vector Machine (SVM), *Bae and Park (2018)*	The model's performance was at its peak when both common and rare inputs were employed.	Its demographic and genetic based	Analysis including other constant traits, such as BMI and metabolic syndrome-related variables, necessitate the inclusion of other data.
16.	Type 2	support vector machine (SVM) 96.80% *Abbas et al (2019)*	Provides doctors with a cost-effective and efficient method for screening those who are more likely to develop T2DM in the future.	Prediction model was anchored on plasma glucose concentrations.	In comparison with other methods, the one used in this study performs much better in terms of accuracy and sensitivity combined.

Discussion

From this research, it was gathered that different risk prediction models exist for diabetes mellitus on Type 1, Type 2, and Gestational diabetes. Such that Deep Learning models specifically the Neutral Network Model in this study presented a higher performance compared with the Machine learning models, but machine learning is commonly adopted method from the more than 15 reviewed article for risk prediction for Diabetes Mellitus,

because they can predict precisely without being clearly programmed, while the deep learning models are enormously costly to train due to the multifaceted data models. As shown in Fig 2. Below, the reviewed risk prediction models for the different variants for diabetes mellitus.

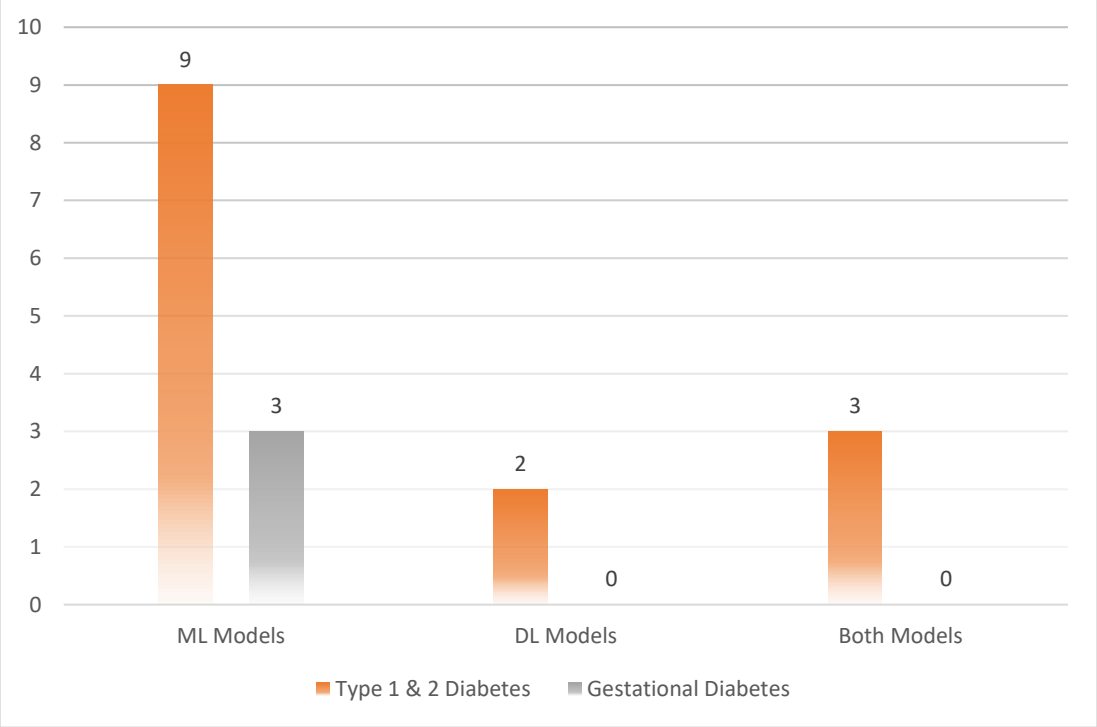

Figure 2. Reviewed Risk Prediction Models

In summary, the algorithms for the Deep Neural Network Model have achieved the highest precision and accuracy score, and as should be used in classification and forecasting problems of Diabetes Mellitus. There is also the competitive precision in other algorithms. Therefore, in the classification and risk prediction survey this study proposes considering using the algorithms explicitly explained in table 1 above, paying key attention to their advantages, disadvantages, and the prospects they proffer. These methods can also be used in conjunction with other deep learning or machine learning techniques to create ensemble models that improve prediction accuracy even further.

Conclusion and Recommendations

Diabetes mellitus (DM) is a debilitating illness that is incredibly prevalent, with the body being unable to metabolize glucose. Early diagnosis of the disease saves patients money and lowers their risk of developing serious health complications in future. The risk predictive models for Diabetes Mellitus reflect the risk patterns being developed and validated. Using either a Deep Learning or Machine Learning model, these models can be

incorporated in an automated form to assist doctors in assessing the risk of Diabetes Mellitus in their patients. The prediction model can also be used in medical studies on primary treatment for Diabetes Mellitus as a positive effects of recurrence preventive measures. The work studied different existing risk prediction model for Type 1, Type 2 and gestational diabetes by reviewing more than 10 articles and suggested the used of Deep learning models such as the Deep Neural Network model to be used for any variant of Diabetes Mellitus risk prediction, and it can also be combined with high performance machine learning model such as Support Vector Machine.

References

Ayon, S., & Islam, M. (2019). Diabetes Prediction: A Deep Learning Approach. *International Journal of Information Engineering and Electronic Business*, *11*, 21–27.

Bae, S., & Park, T. (2018). Risk prediction of type 2 diabetes using common and rare variants. *International Journal of Data Mining and Bioinformatics*, *20*(1), 77.

Banday, M. Z., Sameer, A. S., & Nissar, S. (2020). Pathophysiology of diabetes: An overview. *Avicenna Journal of Medicine*, *10*(4), 174.

Benhalima, K., Van Crombrugge, P., Moyson, C., Verhaeghe, J., Vandeginste, S., Verlaenen, H., Vercammen, C., Maes, T., Dufraimont, E., De Block, C., Jacquemyn, Y., Mekahli, F., De Clippel, K., Van Den Bruel, A., Loccufier, A., Laenen, A., Minschart, C., Devlieger, R., & Mathieu, C. (2020). Estimating the risk of gestational diabetes mellitus based on the 2013 WHO criteria: A prediction model based on clinical and biochemical variables in early pregnancy. *Acta Diabetologica*, *57*(6), 661–671.

Deberneh, H. M., & Kim, I. (2021). Prediction of Type 2 Diabetes Based on Machine Learning Algorithm. *International Journal of Environmental Research and Public Health*, *18*(6), 3317.

Goutham, S., R., V., & K.p., S. (2018). Diabetes detection using deep learning algorithms. *ICT Express*, *4*(4), 243–246.

Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A., & Stiglic, G. (2020). Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Scientific Reports*, *10*(1), 11981.

Larabi-Marie-Sainte, Aburahmah, Almohaini, & Saba. (2019). Current Techniques for Diabetes Prediction: Review and Case Study. *Applied Sciences*, *9*(21), 4604. https://doi.org/10.3390/app9214604

Mahboob Alam, T., Iqbal, M. A., Ali, Y., Wahab, A., Ijaz, S., Imtiaz Baig, T., Hussain, A., Malik, M. A., Raza, M. M., Ibrar, S., & Abbas, Z. (2019). A model for early prediction of diabetes. *Informatics in Medicine Unlocked*, *16*, 100204.

Moko, A. & Onuodu, F. E. (2021). An Improved Model for Feature Selection on Type 2 Diabetes Risk Prediction in Nigeria, *International Journal of Basic Science and Technology, 7*(2), 123-133.

Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders*, *19*(1), 391–403.

Sweeting, A. N., Wong, J., Appelblom, H., Ross, G. P., Kouru, H., Williams, P. F., Sairanen, M., & Hyett, J. A. (2019). A Novel Early Pregnancy Risk Prediction Model for Gestational Diabetes Mellitus. *Fetal Diagnosis and Therapy*, *45*(2), 76–84.

Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, *10*, 100–107.

Xie, Z., Nikolayeva, O., Luo, J., & Li, D. (2019). Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques. *Preventing Chronic Disease*, *16*, 190109.

Xu, Z., & Wang, Z. (2019). A Risk Prediction Model for Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XGBoost Ensemble Classifier. *2019 Eleventh International Conference on Advanced Computational Intelligence (ICACI)*, 278–283.

Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019). A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques. *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 1–4.

Zhang, Y.-Z., Zhou, L., Tian, L., Li, X., Zhang, G., Qin, J.-Y., Zhang, D.-D., & Fang, H. (2020). A mid-pregnancy risk prediction model for gestational diabetes mellitus based on the maternal status in combination with ultrasound and serological findings. *Experimental and Therapeutic Medicine*, *20*(1), 293–300.

Zhou, H., Myrzashova, R., & Zheng, R. (2020). Diabetes prediction model based on an enhanced deep neural network. *EURASIP Journal on Wireless Communications and Networking*, *2020*(1), 148.

Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting Diabetes Mellitus with Machine Learning Techniques. *Frontiers in Genetics*, *9*, 515.