Tikrit Journal of Pure Science
ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)
Journal Homepage: http://tjps.tu.edu.iq/index.php/j

An Intelligent Gestational Diabetes Mellitus Recognition System Using Machine Learning Algorithms

Rasool Jader, Sadegh Aminifar
Computer Science, Faculty of Science, Soran University, Erbil, Iraq
https://doi.org/10.25130/tjps.v28i1.1269

ARTICLE INFO.

Article history:
-Received: 26 / 7 / 2022
-Accepted: 28 / 9 / 2022
-Available online: 20 / 2 / 2023

Keywords: Artificial Intelligence, Machine Learning, Clustering, Classification, Gestational Diabetes.

Corresponding Author:
Name: Rasool Jader
E-mail:
rasool.jader@gmail.com
s.aminifar@yahoo.com
Tel:

ABSTRACT

Diabetes mellitus is also called gestational diabetes when a woman has high blood sugar while she is pregnant. It can show up at any time during pregnancy and cause problems for the mother and baby during or after the pregnancy. If the risks are found and dealt with as soon as possible, there is a chance that they can be reduced. The healthcare system is one of the many parts of our daily lives that are being rethought thanks to the creation of intelligent systems by machine learning algorithms. In this article, a hybrid prediction model is suggested as a way to find out if a woman has gestational diabetes. In the recommended model, the amount of data is reduce by using the K-means clustering method. Predictions are made using a number of classification methods, such as decision tree, random forests, SVM, KNN, logistic regression, and naive bayes. The results show that accuracy goes up when clustering and classification are used together.

1. Introduction

Diabetes mellitus during pregnancy, also known as gestational diabetes, is one of the most frequent complications that may arise, affecting about one in every six live births globally [4]. Any degree of glucose inability, with the onset of symptoms occurring during pregnancy, is the definition of this condition. Conway at 2012 [9], discussed that GD may strike at any point during pregnancy, producing complications for the mother and baby throughout pregnancy and after the baby is born. If the risks are recognized and mitigated at an early stage, the potential consequences will be reduced. Algorithms for machine learning are being used in the creation of intelligent systems, which are bringing about changes in every facet of our life. By automating tasks that would normally be carried out by humans, artificial intelligence makes the lives of patients, clinicians, and hospital managers simpler. The current goals for AI include reducing the number of incorrect diagnoses of diabetes and developing more effective medical treatment methods [8]. These days, healthcare systems create vast volumes of data; thus, more complex systems are reliant on this data in order to construct more precise models. The purpose of this work is to attempt to construct a model that can examine both current and historical instances of diabetes in order to forecast and diagnose future cases using machine learning techniques, which may be of use to both patients and hospital administrators. And we attempted to determine, via the use of machine learning algorithms, which model has the highest degree of precision with the least amount of mistakes. In the Iraqi Kurdistan Region, we attempted to work with a data set that was obtained in both governmental and private labs. It is important to note that this was our goal. The information that we have gathered consists of the subjects' ages, as well as their weights, heights, gestational numbers, family medical history, and the results of their diabetes tests. This information reveals at what ages and in what circumstances pregnant women are more likely to develop gestational diabetes. The remains of this

paper are structured as follows: Section 2 explains several studies of related work. Section 3 describes the methodology of the proposed model, and Section 4 some discussion about the results. Finally, section 5 expresses the conclusion.

2. Literature review

In this part, we talk briefly about some studies that have been done on intelligent systems and machine learning techniques for modeling and predicting different types of diabetes mellitus.

[1] To figure out how to predict diabetes, the results of different machine learning strategies were looked at. The classification learner tool in MATLAB was used for this study. This tool had a decision tree, K Nearest Neighbor, Support Vector Machines, and Logistic Regression. When it comes to measuring performance, the results are judged by how well they can be categorized.

[3] The authors looked at 768 samples of data from the Pima Indian Diabetes Dataset (PIDD) and used predictive algorithms like K-Nearest Neighbor, Nave Bayes, Random Forest, and J48 to create an ensemble learning by combining different machine learning techniques into one. This improved the accuracy of the suggested system's performance. This was done so that different machine learning techniques could be combined to make an ensemble learning.

[10] In this study, the PIMA Indian dataset was used to show the different ways that machine learning can be used to predict whether or not a pregnant woman has gestational diabetes. The first thing they do is clean up their data so that it is as accurate as possible. The accuracy of each method was compared using the scores for the Receiver Operating Characteristic (ROC) and the Area Under the Curve (AUC). A picture of the results of the confusion matrix of algorithms was used to judge both how well the models worked and where they fell short. The results of this study show that adjusting the parameters of machine learning algorithms can make them more accurate.

[19] The other goal of this study was to find a better way to predict diabetes risk. Models are put into groups with Decision Tree (DT), Artificial Neural Network (ANN), Naïve Bayes (NB), and Support Vectors Machine (SVM). About 75% of the time, all four models are about right. The scientists' plan would use ML algorithms to predict data and show a comparison to find the most accurate treatment for diabetes. Without data, SVMs are great. Even without data, SVMs work well. SVM can work with both semi-structured and unstructured data, like images and text. SVM needs certain parameters to be set because the algorithm needs it. The DT is changing the best way to organize data. Naive Bayes skips over missing-value probabilities. This is good for large datasets. Preparedness-minded. Adding more data to a training set makes bias worse. ANN is accurate and easy to use. It takes time to process complex data.

[20] has talked about how high blood sugar levels can lead to diabetes. Also, different intelligent systems were shown using classifiers to predict and predict diabetes using machine learning algorithms like decision tree, SVM, Naive Bayes, and ANN algorithms. Diabetes could be predicted based on these groups.

3. Methodology

Combinations of machine learning techniques are used in the model that has been suggested. According to the flow chart shown in **Figure 1**, the model employs the supervised approach for the purpose of data reduction. For the purpose of prediction, the model also incorporates unsupervised techniques such as the classification algorithms: decision tree, random forest, SVM, KNN, logistic regression, and naive bayes.

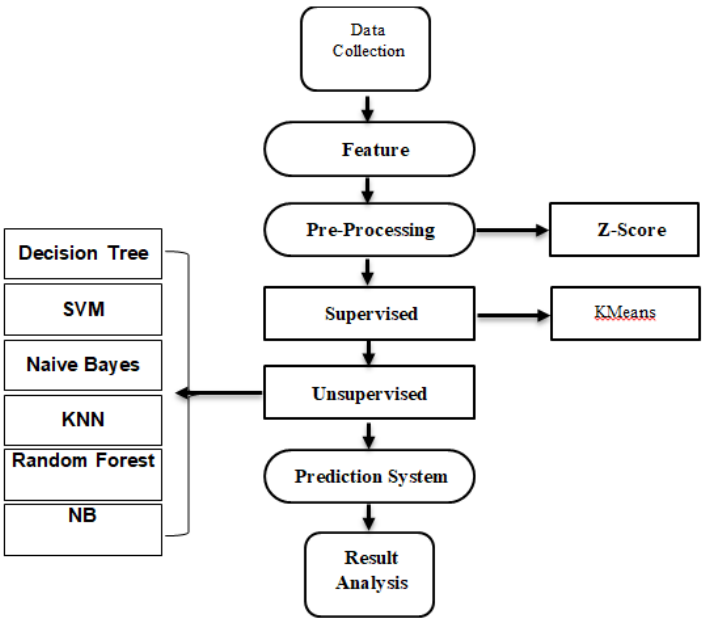

Fig. 1: The Flow Chart of Proposed Model

3.1. Data Collection:

Large volumes of data have been collected by healthcare systems and labs all over the globe, and improved apps depend on that data to function better and provide more accurate findings [12]. Our models were trained using data obtained from a variety of governmental and commercial labs located inside the Iraqi Kurdistan Region. The data collection had a total of 730 different instances and 7 different characteristics. **Table 1** presents the properties together with the data types that correspond to them.

Table 1: The Attributes and Data Types of Training Dataset

	Attributes	Data Type
1	Age	Numeric
2	Weight	Numeric
3	Height	Numeric
4	BMI	Numeric
5	Pregnancy Number	Numeric
6	Heredity	Boolean
7	Blood Sugar Test	Boolean

3.2. Feature Extraction:

Feature extraction is useful when you need to limit the number of characteristics that are necessary for processing without losing information that is vital or relevant. Feature extraction helps you do this [2]. The elimination of redundant information within the dataset is another potential benefit of feature extraction [19]. The datasets that we have for weight and height represent and are used to calculate the body mass index (BMI). As can be seen in Equation 1, there was a significant and positive association between BMI and weight, whereas there was a significant and negative correlation between BMI and height. The level of body mass index has been shown to have a correlation with the prevalence of diabetes mellitus, hypertension, and dyslipidemia. According to the findings of both studies, a rise in the body mass index (BMI) is often related with a considerable increase in the prevalence of diabetes mellitus.

Equation 1: $$BMI = \frac{Weight}{Height^2}$$

$$\boldsymbol{BMI = \frac{65\ kg}{(154\ cm)^2} * 10000 = 27.4}$$

3.3. Data Pre-Processing:

One of the ways for processing data is known as data normalization [15]. This method is used to convert unintelligible data into a data collection that is understandable. Normalize data is a scaling or mapping method for transforming unnormal data to normal data [16]. The Z-score method, which is the process of normalizing every value in the dataset, was included in our model by the use of the Z-score methodology. The formula for normalizing z-scores can be found in the Equation 2, while the formula for calculating standard deviation, which was required in order to produce a z-score, can be found in Equation 3.

Equation 2: $$z = (x_i - \mu)/\sigma$$

Equation 3: $$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N - 1}}$$

3.4. Clustering Algorithm:

Clustering is an algorithm of machine learning that separates a set of data points or a collection of data points into several groups, with data points in the same group being more similar to one another than data points in other groups. The clustering algorithm is described below. To put it another way, the goal is to create clusters out of groupings that have characteristics in common with other groups [2]. Clustering is a useful methodology in the field of data science. It is a technique for determining whether or not a data collection has a cluster structure by analyzing the degree of dissimilarity that exists between the clusters and the level of similarity that exists within each cluster [18]. In order to reduce the amount of data we needed to analyze for our study, we made use of the k-means approach, which is the most well-known and widely used clustering algorithm. In the research that has been done, k-means has been extended in a number of different ways [29]. Initializations will always have an effect on the k-means technique and its expansions with a necessary minimum number of clusters. On the other hand, in pattern recognition and machine learning, clustering is accomplished by unsupervised learning [13]. In an unsupervised method, selecting the appropriate number of clusters in which the data may be categorized is of the utmost importance [25]. We used The Elbow Method, which is one of the most well-known ways for identifying the value of k that should be used for KMeans clusters, in order to establish the optimal k for those clusters. In the Elbow technique, what we are doing is really counting the number of clusters that are represented by (K), as shown in **Figure 2**, and calculating the WCSS (Within-Cluster Sum of Square) points for each value of K.

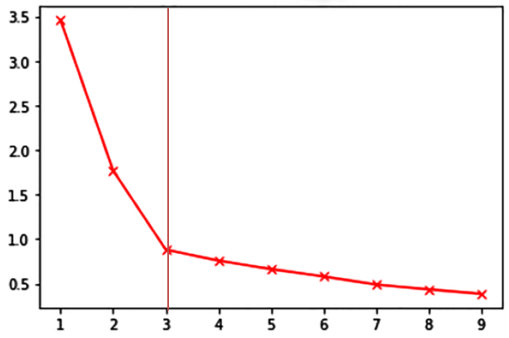

Fig. 2: Determine The Knee Value of The Elbow Method

3.5. Classification Algorithms:

The results of the clustering stage, in which data is reduced and grouped into a particular group, act as an input for the classifier techniques. In these techniques, a sample of the data from the stage before this one is allocated to a cluster of new input data that

is similar to the cluster that the sample was allocated in the stage before this one. A decision tree, a random forest, a support vector machine (SVM), a key-value network (KNN), logistic regression, and naive bayes are some of the several classification techniques that are included in the system that has been developed.

3.5.1. Decision Tree:

A decision tree is a kind of algorithmic method that may be used to segment data based on particular characteristics [10]. One of the algorithms for supervised learning looked somewhat like this. The objective is to get familiar with the fundamental decision tree instructions in order to construct a mental model that can accurately forecast the value of a target variable. It is well suited for lifelong education [27]. The rules are often expressed in the form of if-then-else phrases, and the decision tree follows these [19]. Decision tree, which does not need a great deal of computer power, are used in the classification process. The decision tree is a useful tool for processing continuous data [5][6].

3.5.2. Random forest:

Random Forest is a method for the reduction of dimensionality that makes use of a large number of decision trees in order to construct a classification [9]. This is an example of an ensemble approach that may be used for classification, as well as other problems [8]. It is possible to use it to order the variables in terms of how significant they are [7].

3.5.3. Support Vector Machine:

The purpose of support vector machines is to locate a hyperplane line that divides the positive and negative samples by the greatest possible margin [6][26]. Finding a hyperplane in the Multidimensional Features space that effectively categorizes data points is the goal of the support vector machine (SVM) [31]. An SVM model in which the instances are modeled as points in space and the space is mapped in such a way that the examples of the different categories are separated by as much space as is practically possible. After then, new instances are mapped into the same area, and a category is assigned to them depending on where they are located in the gap [7].

3.5.4. K-Nearest Neighbors:

K-Nearest Neighbors, or KNN, is a method of classification in which the function is only approximated locally, and all computation is postponed until the function is evaluated. As the classification of this method is based on distance, if the characteristics represent various physical units or arrive at wildly different sizes, then normalizing the training data may significantly enhance the algorithm's performance. . KNN is only simulated locally, and computations are postponed until the process of classification is finished. One of the most fundamental types of machine learning algorithms accessible is called the KNN algorithm [11].

3.5.5. Logistic Regression:

The generalized linear model is also known as logistic regression, which is another term for it. The linear component of nonlinear functions and the link function are both subdivided into the nonlinear function. The link function is responsible for delivering the linear portion of the output that is generated by the classification model. In logistic regression, a logistic function is applied to the linear output in order to deal with it. The logistic function never produces results outside of the range of 0 to 1 [1].

3.5.6. Naive Bayes:

NB is a classification strategy that can deal with missing values while it is in the process of classifying data. In order to categorize them, a supervised learning approach is used. The premise that Naive Bayes works on the basis of conditional probability is the central tenet of this statistical methodology. Conditional independence breaks down for linked attributes, which causes Naive Bayes performance to degrade as a direct consequence [20]. When it comes to calculating conditional probabilities, the Naive Bayes method is quite effective against noise points.

3.6. Prediction System:

The proposed model uses a clustering approach to partition the data without missing any. KMeans methods are used instead of other clustering techniques because the DBSCAN algorithm may identify samples as noise and discard these objects, while KMeans generally cluster all the objects. And also, it is worth mentioning that the numeric data set prevents us from using KModes clustering. And to improve the KMeans clustering algorithm, the proposed model uses the elbow technique to find the optimal k for clustering the data. And After data reduction by KMeans, The Classification Algorithms used for prediction the New Samples. The code for the suggested model was written using the computer language Python, and a confusion matrix was used so that several classification techniques could be evaluated and compared to one another. In the course of our investigation, we made use of a number of different comparison criteria, including accuracy which shown in Equation 4, precision which shown in Equation 5, and recall which shown in Equation 6.

Equation 4: $$Accuracy = \frac{total\ currect\ predictions}{total\ predictions}$$

Equation 5: $$Precision = \frac{TP}{TP + FP}$$

Equation 6: $$Recall = \frac{TP}{TP + FN}$$

4. Results and Discussion:

Using a combination of supervised and unsupervised algorithms in this research presents a method for the reliable prediction of models that may be used in recognition of gestational diabetes mellitus. In the beginning, we solely used classification techniques; after that, we used a combination of unsupervised, which is KMeans clustering algorithm and classification methods in the model that was

recommended. According to **Table 2**, the proposed model provides more accurate results than the currently used classification methods for pregnant women who have diabetes. The rustles of Table 2 obtained from our research.

Table 2: The Comparison of Classification Accuracy and Combination Accuracy which obtained from our research (The Proposed Model)

	Algorithms	Accuracy of Classification	Accuracy of Clustering vs. Classification
1	Decision Tree	**0.95**	**0.96**
2	Random Forest	**0.94**	**0.96**
3	SVM	**0.77**	**0.83**
4	KNN	**0.93**	**0.93**
5	Logistic Regression	**0.74**	**0.76**
6	Naïve Bayes	**0.73**	**0.72**

It is possible that we will evaluate the recommended model in light of results gained from earlier categorization models. The results shown in **Table 3** demonstrate that the suggested model is, on average, more accurate.

Table 3: The Machine Learning Accuracy of Existing Works

	Algorithms	Accuracy	Reference
1	Decision Tree	**78.1768 %**	[5] (AlJarullah, 2011)
2	Random Forest	**91 %**	[14] (Mujumdar and Vaidehi, 2019)
3	SVM	**77.3 %**	[19] (Sonar and Jaya Malini, 2019)
4	KNN	**77 %**	[17] (Sarwar et al., 2018)
5	Logistic Regression	**77.9 %**	[1] (Al-Zebari and Sengur, 2019)
6	Naïve Bayes	**79.84 %**	[11] (Jeevan Nagendra Kumar et al., 2019)

5. Conclusion

The Dataset that we used to build our proposed model came from both public and private laboratories located in the Kurdistan Region of Iraq. The data collection had a total of 730 different instances and 7 different characteristics. For the purpose of recognizing gestational diabetes, a hybrid prediction model within the framework of the suggested model has been created. The K-means clustering algorithm is employed in the recommended model for the purpose of reducing the amount of data, and several classification methods, including decision tree, random forests, SVM, KNN, logistic regression, and naive bayes, are used for the purpose of classification and prediction. The accuracy that was acquired via each technique of supervised and unsupervised algorithms improved performance when compared to previous works. According to the results, combining the K Means clustering algorithm for data reduction and classification algorithms for prediction has a great effect on accuracy. The Random Forest with 96% and the Decision tree with 96% have the most incredible accuracy.

6. Future Work

Healthcare systems have generated large amounts of data, so advanced systems depend on that data to make more accurate models. This paper will try to develop a model to analyze the current and old diabetic cases to predict and diagnose the new issues using machine learning algorithms to help patients and hospital administrators. And we tried to find the performance of applied machine learning algorithms and data mining techniques to generate a prediction model to get the best accuracy with a minor error. In future work, we will try to develop the proposed model for an application for predicting diabetes instances in public and private hospitals and laboratories. The application will be designed to get a new sample and add it to the dataset. This method updates the database daily, so each time the model is trained, it will have more data to work with.

References

[1] Al-Zebari, A. and Sengur, A. (2019) 'Performance Comparison of Machine Learning Techniques on Diabetes Disease Detection', 1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings, pp. 2–5. doi: 10.1109/UBMYK48245.2019.8965542.

[2] Alapati, Y. and Sindhu, K. (2016) 'Combining Clustering with Classification: A Technique to Improve Classification Accuracy', International Journal of Computer Science Engineering, 5(06), pp. 336–338. Available at: https://en.wikipedia.org/wiki/Feature_selection.

[3] Alehegn, M., Joshi, R. and Alehegn, M. (2017) 'Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach.', International Research Journal of Engineering and Technology, 4(10), pp. 426–436. Available at: www.irjet.net.

[4] Ali, N. et al. (2021) 'Effect of gestational diabetes mellitus history on future pregnancy behaviors: The Mutaba'ah study', International Journal of Environmental Research and Public Health, 18(1), pp. 1–12. doi: 10.3390/ijerph18010058.
[5] Hamad, A., Aminifar, S. and Daneshwar, M. (2020) 'An interval type-2 FCM for color image segmentation', International Journal of Advanced Computer Research, 10(46), pp. 12-17. doi: 10.19101/ijacr.2019.940114.
[6] Jader, R.F., Abd, M.H.M. and Jumaa, I.H., 2022. Signal Modulation Recognition System Based on Different Signal Noise Rate Using Artificial Intelligent Approach. Journal of Studies in Science and Engineering, 2(4), pp.37-49.
[7] AlJarullah, A. A. (2011)' Decision tree discovery for the diagnosis of type II diabetes', 2011 International Conference on Innovations in Information Technology, IIT 2011, pp. 303–307. doi: 10.1109/INNOVATIONS.2011.5893838.
[8] Barakat, N., Bradley, A. P. and Barakat, M. N. H. (2010) 'Intelligible support vector machines for diagnosis of diabetes mellitus', IEEE Transactions on Information Technology in Biomedicine, 14(4), pp. 1114–1120. doi: 10.1109/TITB.2009.2039485.
[9] Aminifar, S. and Marzuki, A. (2013) 'Uncertainty in interval type-2 fuzzy systems', Mathematical Problems in Engineering, 2013. doi: 10.1155/2013/452780.
[10] Benbelkacem, S. and Atmani, B. (2019) 'Random forests for diabetes diagnosis', 2019 International Conference on Computer and Information Sciences, ICCIS 2019, pp. 1–4. doi: 10.1109/ICCISci.2019.8716405.
[11] Chen, W. et al. (2018) 'A hybrid prediction model for type 2 diabetes using K-means and decision tree', Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, 2017-Novem(61272399), pp. 386–390. doi: 10.1109/ICSESS.2017.8342938.
[12] Aminifar, S., 2014. Design and implementation of fuzzy controllers for handling uncertainty in an industrial application (Doctoral dissertation, Universiti Sains Malaysia).
[13] Jader, R.F., Aminifar, S. and Abd, M.H.M., 2022. Diabetes Detection System by Mixing Supervised and Unsupervised Algorithms. Journal of Studies in Science and Engineering, 2(3), pp.52-65.
[14] Aminifar, S. (2020) 'Uncertainty Avoider Interval Type II Defuzzification Method', Mathematical Problems in Engineering. Edited by J. V. Salcedo, 2020, p. 5812163. doi: 10.1155/2020/5812163.
[15] Conway, D. L. (2012) 'Gestational Diabetes Mellitus', Queenan's Management of High-Risk Pregnancy: An Evidence-Based Approach: Sixth Edition, 26, pp. 168–173. doi: 10.1002/9781119963783.ch20.
[16] Gnanadass, I. (2020) 'Prediction of Gestational Diabetes by Machine Learning Algorithms', IEEE Potentials, 39(6), pp. 32–37. doi: 10.1109/MPOT.2020.3015190.
[17] Jeevan Nagendra Kumar, Y. et al. (2019) 'Prediction of diabetes using machine learning', International Journal of Innovative Technology and Exploring Engineering, 8(7), pp. 2547–2551. doi: 10.35940/ijrte.e6290.018520.
[18] Jiang, F. et al. (2017) 'Artificial intelligence in healthcare: Past, present and future', Stroke and Vascular Neurology, 2(4), pp. 230–243. doi: 10.1136/svn-2017-000101.
[19] Likas, A., Vlassis, N. and Verbeek, J. (2011) 'The global k-means clustering algorithm Intelligent Autonomous Systems', ISA technical report series.
[20] Marzuki, A., Tee, S. Y. and Aminifar, S. (2014) 'Study of fuzzy systems with Sugeno and Mamdanitype fuzzy inference systems for determination of heartbeat cases on Electrocardiogram (ECG) signals', International Journal of Biomedical Engineering and Technology, 14(3), pp. 243-276. doi: 10.1504/IJBET.2014.059673.
[21] Rasool Jader, Sadegh Aminifar, "Predictive Model for Diagnosis of Gestational Diabetes in the Kurdistan Region by a Combination of Clustering and Classification Algorithms: An Ensemble Approach", Applied Computational Intelligence and Soft Computing, vol. 2022, Article ID 9749579, 11 pages, 2022. https://doi.org/10.1155/2022/9749579
[22] Patro, S. G. K. and sahu, K. K. (2015) 'Normalization: A Pre-processing Stage', Iarjset, pp. 20–22. doi: 10.17148/iarjset.2015.2305.
[23] Saravana Kumar, N. M. et al. (2015) 'Predictive methodology for diabetic data analysis in big data', Procedia Computer Science, 50, pp. 203–208. doi: 10.1016/j.procs.2015.04.069.
[24] Sarwar, M. A. et al. (2018) 'Prediction of diabetes using machine learning algorithms in healthcare', ICAC 2018 - 2018 24th IEEE International Conference on Automation and Computing: Improving Productivity through Automation and Computing, (September), pp. 1–6. doi: 10.23919/IConAC.2018.8748992.
[25] Jader, R. and Aminifar, S., 2022. Fast and Accurate Artificial Neural Network Model for Diabetes Recogni-tion. NeuroQuantology, 20(10), pp.2187-2196.
[26] Aminifar, S. and Bin Marzuki, A. (2013) 'Horizontal and vertical rule bases method in fuzzy controllers', Mathematical Problems in Engineering, 2013. doi: 10.1155/2013/532046.
[27] Sinaga, K. P. and Yang, M. S. (2020) 'Unsupervised K-means clustering algorithm', IEEE Access, 8, pp. 80716–80727. doi: 10.1109/ACCESS.2020.2988796.
[28] Sonar, P. and Jaya Malini, K. (2019) 'Diabetes prediction using different machine learning approaches', Proceedings of the 3rd International

Conference on Computing Methodologies and Communication, ICCMC 2019, (Iccmc), pp. 367–371. doi: 10.1109/ICCMC.2019.8819841.
[29] Vijayan, V. V. and Anjali, C. (2016) 'Prediction and diagnosis of diabetes mellitus - A machine learning approach', 2015 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2015, (December), pp. 122–127. doi: 10.1109/RAICS.2015.7488400.
[30] Brereton, R. G. (2015) 'The mahalanobis distance and its relationship to principal component scores', Journal of Chemometrics, 29(3), pp. 143–145. doi: 10.1002/cem.2692.
[31] Tuysuzoglu, G., Birant, D. and Pala, A. (2019) 'Majority voting based multi-task clustering of air quality monitoring network in Turkey', Applied Sciences (Switzerland), 9(8), pp. 1–21. doi: 10.3390/app9081610.
[32] Jader, R.F., Aminifar, S. and Abd, M.H.M., 2022. Journal of Studies in Science and Engineering.
[33] Bansal, M., Goyal, A. and Choudhary, A. (2022) 'A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning', Decision Analytics Journal, 3(May), p. 100071. doi: 10.1016/j.dajour.2022.100071.
[34] Chung, J. and Teo, J. (2022) 'Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges', Applied Computational Intelligence and Soft Computing. Edited by A. Minutolo, 2022, p. 9970363. doi: 10.1155/2022/9970363.

نظام ذكي للتعرف على داء السكرى الحمل باستخدام خوارزميات التعلم الآلي

رسول فاخر جادر ، صادق عبدالله

قسم علوم الحاسوب، كلية العلوم ، جامعة سوران، اربيل، العراق

الملخص

سكري الحمل هو نوع من ارتفاع السكر في الدم يحدث أثناء فترة الحمل. يمكن أن يحدث في أي مرحلة من مراحل الحمل ويمكن ان يسبب مشاكل لكل من الأم والطفل أثناء الولادة او بعدها. يمكن تقليل مخاطر هذا المرض إذا تم اكتشافها وإدارتها مبكرًا ، خاصة في المناطق التي لا تتوفر فيها سوى الاختبارات الدورية للنساء الحوامل. الأنظمة الذكية المصممة بواسطة خوارزميات التعلم الآلي تعيد صياغة جميع مجالات حياتنا ،خاصة في نظام الرعاية الصحية. يقترح هذا البحث نموذج تنبؤ مشترك لتشخيص سكري الحمل. تم الحصول على مجموعة البيانات من مختبرات العراق – إقليم كردستان، والتي جمعت المعلومات من النساء الحوامل المصابات وغير المصابات بداء السكري. النموذج المقترح يستخدم تقنية KMeans لتجميع وكذلك لتقليل البيانات. ويستخدم طريقة الكوع للعثور على قيمة k المثلى في خوارزمية KMeans. وطرق التصنيف مثل شجرة القرار ، الغابة العشوائية، SVM، KNN والانحدار اللوجستي، والخلايا الساذجة كلها مستخدمة للتنبؤ. أظهرت نتائج النموذج المقترح أن استخدام مزيج من طريقة KMeans وطريقة الكوع وتقنية التصنيف يحسن بشكل كبير دقة التنبؤ.