International Journal of Management and Data Analytics (IJMADA)

Int. J. Management and Data Analytics, Vol. 4(1), 43-54
ISSN: 2816-9395
Journal Homepage: http://ijmada.com
https://doi.org/10.5281/zenodo.13926972

Performance Analysis of Diabetes Detection Using Machine Learning Classifiers

Hung Vu Trung Huynh, Liu Hui, Ngoc Han Nguyen, Ruixuan Qiao
University Canada West, Vancouver, BC, Canada

Received: September, 2024, Published: Oct, 2024

ARTICLE INFO
Keywords:
Classifiers; Diabetes Prediction and Diagnosis; Healthcare; Machine Learning

ABSTRACT
Diabetes is a chronic medical condition that has been causing severe public health challenges in not only Canada, but the entire world, for as long as time immemorial, impacting millions of people and putting pressure on healthcare resources. That said, conventional diagnostic procedures sometimes depend on few data points and are prone to mistakes, resulting in premature action. Additionally, the sluggish adoption of modern machine learning (ML) technologies in the healthcare industries might be due to their misunderstanding of the systems' decision-making procedures. This study purports to fill that gap by looking at various machine learning (ML) algorithms and applying them on the PIMA Indians Diabetes Dataset provided by the National Health Institute of Diabetes and Digestive and Kidney Diseases with the aim of improving the validity of diabetes prediction and diagnosis. Three types of machine learning classifiers are used: Tree-based, Function-based, and Rule-based. Results have shown that Stochastic Gradient Descent (function), Logistic Regression (function), JRip (rules) and Random Forests (trees) are among the top performing classifiers. They are judged based on different metrics, such as accuracy, precision, recall, specificity, F-1 score, MCC, and ROC area. Despite performing well in almost all the metrics, SGD's low recall score shows that it is not the most optimal algorithm. Given that recall score is prioritized in the context of clinical diagnostics, Random Forest emerges as a strong candidate due to its balanced performance across key metrics.

1. INTRODUCTION

Diabetes is a chronic medical illness that affects millions of individuals globally, posing considerable dangers that could be fatal (Mousa et al., 2023). Furthermore, the condition is incurable, and patients can only manage its symptoms through expensive personalized treatment and therapy. Identifying a person's susceptibility to diabetes is a huge undertaking. Early diagnosis and intervention will hinder the progression of the disease while also helping patients save on their medical bills. The traditional data analysis method is not efficient in dealing with the complexity and volume of data, which can lead to low efficiency in disease diagnosis and patient management. In relation to this matter, machine learning is a powerful technology that could help with decision-making, strengthen illness prediction, and thereby provide better medical treatment to patients. (Davenport & Kalakota, 2019).

Machine learning is a branch of artificial intelligence; it enables computers to learn from data and make predictions using algorithms. It can identify hidden patterns which are not obvious to human eyes. The ability to parse large, complex data efficiently is one of the core factors helping AI systems surpass humans in assessing the likelihood of disease onset (Hounguè & Bigirimana, 2022). According to Iparraguirre-Villanueva et al. (2023), ML models have proven to be accurate forecasters of diabetes' growth. These models can help with predicting the onset of diabetes or liver disorders by analyzing the patient's medical history, demographics, lifestyles and their genetic composition. With this information provided by machine learning, medical care can take preventive measures, make medical treatment plans, as well as better utilize the limited resources in healthcare.

To enhance treatment options and make healthcare more accessible to the more unfortunate, an automated computerized system is needed to diagnose diabetes. This research will, therefore, tap into various approaches for identifying diabetes mellitus using three classification methods, such as trees, rule, and function-based. This paper will further discuss the role of machine learning in overcoming the challenges in healthcare, focusing on its application in diabetes treatment. This study will highlight how the effectiveness of these technologies improves diagnosis accuracy and patient outcomes by reviewing the existing literature and analyzing different machine learning algorithms. In addition, we will discuss the impact of said algorithms on healthcare workers, patients, and decision-makers, highlighting why it is crucial to apply these innovative technologies responsibly and ethically in everyday practice.

Contact: Hung Huynh, hunghuynhza@gmail.com

Many trees-based algorithms, namely random trees, J48 (C4.5 algorithm) and random forests can explain and handle the complicated non-linear relationships with data. It can generate decision rules that are easy for medical practitioners to understand, which makes them especially suitable for clinical settings. For the same reasons, rule-based algorithms include methods such as RIPPER, OneR, and PART aim to create human-readable rules from data. These algorithms are widely used in medical settings because they provide clear and actionable interpretations to guide clinical decisions, particularly ideal in environments which demand transparency and explainability. Lastly, we have function-based algorithms such as Logistic Regression, Stochastic Gradient Descent (SGD), and Multilayer Perceptron (MLP), which are well known for their high accuracy and ability to identify critical patterns in large datasets. Their predictive power makes them highly effective in detecting subtle signs of disease that may be neglected by simpler models.

By comparing these algorithms, we aim to determine which algorithm is the most effective one in predicting diabetes diseases from a given dataset. The comparison will be based on key performance metrics such as accuracy, sensitivity/recall, specificity, precision, ROC Area, and MCC. In doing so, we seek to identify the ML algorithms with the best all-round performance in prediction and practical application so that they can be used effectively in correctly identifying positively tested diabetic patients.

The rest of the paper is structured as follows: In section three, the literature review will be covered, providing an overview of the existing ML application in healthcare, particularly in diabetes diseases. After that, a comprehensive analysis and discussion of the classifiers used in this study will be conducted, as well as the comparison of different classifier performances across various key performance metrics. Finally, section five will summarize our findings and recommendations for ML practical implementation and discuss the limitations of this study while also mentioning how future studies are needed to address these limitations.

2. LITERATURE REVIEW

A. AI in Healthcare Context

AI entails the techniques which enable computers to mimic human intelligence, allowing them to acquire knowledge from data and experiences while carrying out complex tasks that would normally call for human intellect (L. P. Nguyen et al., 2023). ML is a subset of AI techniques that focuses on using statistical methodologies to help computer systems improve with experiences. Leveraging information from people's regular physical assessments can assist in producing a preliminary evaluation, serving as an instrument for healthcare professionals in early diagnoses and predictions.

The continual evolution of today's world has brought with it an exponential growth of data whose increased availability makes them readily accessible for building AI and machine learning (ML) algorithms. Within the context of healthcare, the growing availability of data, coupled with the ubiquity of smart devices and data-driven resources, has made it easier for healthcare providers to develop an accurate representation of a patient's condition over time, while also offering novel perspectives on unprecedented cases (Kolasa et al., 2023).

As stated, modern technology and its automated processes facilitate the handling of huge volumes of data, making machine learning an ideal tool to help doctors and other medical personnel make informed decisions. Doctors can look into a patient's illness using medical measurements that include body temperature and arterial pressure and prescribe remedies after having gone through a series of iterative analyses (Bhat et al., 2023). Moreover, Davenport and Kalakota (2019) asserted that 'precision medicine' is the most applied machine learning model in healthcare as it predicts which form of treatment is most likely to be effective on a patient. This is evaluated based on a vast collection of patients' attributes, their previous therapies, their family's medical history, and the setting in which they are treated. In addition, although a healthcare practitioner and a Machine Learning algorithm can arrive at the same conclusion based on the same information given, the latter's response rate is considerably much faster, yielding results more efficiently, which enables intervention to take place sooner (Javaid et al., 2022). Javaid et al. (2022) further pointed out that ML techniques are more favored because they reduce the possibility of human error.

B. ML/AI in Diabetes Diagnoses

For a chronic metabolic condition like diabetes, Artificial Intelligence (AI) technologies in the likes of machine and deep learning serve a crucial role in facilitating decision-making processes for healthcare professionals (Chang et al., 2022; Baadel, et al. 2020). Not only that, but they also help clinicians monitor and manage their patients accordingly, aided by customizable treatments in accordance to each and every patient's health record. Real life approaches see ML/AI use genomic data to assess and predict diabetes risks and use electronic health records (EHR) for diabetes diagnoses (Salazar-Reyna et al., 2020).

In account for the prevalence of diabetes, several machine learning algorithms, for instance Random Forest (RF), Decision Trees (DT), Neural Networks, Logistic Regression, and ensembles, have been proposed to detect the condition risks early on, using features such as age, blood glucose levels, body mass index (BMI), blood

pressure, and other pertinent variables (Bhat et al., 2023). Different methodologies are adopted by researchers to diagnose and predict whether a patient has diabetes or not. For instance, researchers Shetty et al. (2017) used KNN and Naive Bayes methods for said prediction whereby they input patient records via a software program to see whether the person in question has the condition or not. To add on this, Ahmed (2016) also leveraged patient records and treatment plans to classify the disease through Naive Bayes, J48, and logistic regression algorithms. Other algorithms such as Random Forest, function-based multilayer perceptron (MLP) have also proven to be successfully applied in this regard after data preprocessing, after which a correlation-based feature selection process can be initiated to eliminate redundant features (Alam et al., 2019).

1. Tree-based Models

The goal of developing a machine learning model that can classify an individual's condition based on the occurrence of diabetes, which can be done using decision trees. The algorithm follows a hierarchical structure made up of branches, indicating qualities that impact the end result, and nodes where various possibilities are considered. Two categories are derived from this method: classification and regression. According to Dudkina et al. (2021), classification trees are more favored in medical diagnostics because they organize symptoms in accordance with the known target class. This form of learning with labeled training - the known target class - is referred to as supervised learning. To successfully categorize data, the tree separates it into categories and constructs sequences of "if... then..." rules (Dudkina et al., 2021). The traversal through various branches helps the computer to match the symptoms to the target class, thereby predicting diabetes.

Evidently, decision-tree analysis was employed by Chang et al. (2023) to determine the relationships between risk variables that influence the levels of HbA1c, also known as glycated hemoglobin in diabetic patients. The researchers further found depression to be a core component among type 2 diabetes patients. The decision tree algorithm used in the study indicated three pathways with risk features linked to poor blood glucose control in patients with diabetes mellitus.

However, to improve upon this method with less overfitting risks, Random Forest (RF) can be considered. Random Forest entails a training process in which multiple decision trees are aggregated to produce a single output. This machine learning algorithm is particularly ideal for medical diagnostics because of its ability to model complex and multi-feature data. Additionally, there are other reasons why this algorithm gains traction from the healthcare industry. Haripriya et al. (2021) stated that RFs are competent with both category and numerical data in predictive tasks. Its integrated cross-validation capacity allows features/predictors to be ranked accordingly, based on the impact score they have on the dependent variable. In the context of disease prediction, this function could be advantageous for clinicians as they can discover which predictor is affecting the outcome variable the most and promptly act from there.

On a related note, Pal et al. (2022) investigated several supervised learning algorithms for reliable diabetic predictions. The study found that the Random Forest model has the highest accuracy rate compared to the rest. Allen et al. (2022)'s similar approach in using Random Forest to diagnose type 2 diabetes mellitus also resulted in a high accuracy score of 82%, outperforming traditional statistical methodologies.

2. Rule-based System

Natural Language Processing (NLP) applications in medical systems have been hailed as a critical component, helping health services to fulfill the demands of personalized healthcare. Berge et al. (2023) also mentioned that in the 1970s, rule-based NLP systems designed for retrieving structured clinical data from narrative texts were implemented and have since been effectively employed. That said, although this methodology is proven to have a good accuracy record, these systems may not perform well if terms appearing in the narrative cannot be found in their lexicon. It must be stated that the medical language contains multitudes of jargons, technical terms, and grammatical structures, not to mention abbreviations and incorrect spellings (Berge et al., 2023). Hence, according to Jonnalagadda et al (2011) for these systems to be effective at information extraction, they must rely on clinical experts to verify the legitimacy of deterministic rules, as well as to develop and maintain the quality of medical lexical resources.

C. State of the Art Approaches in Diabetes Prediction

1. Deep Learning in Diabetic Retinopathy Detection

In instances of detecting diabetic retinopathy (DR), deep learning techniques have proven to be accurate and effective in their diagnoses, far more superior and error-proof compared to traditional manual methods. DR is a chronic medical condition that requires early identification to avoid grave consequences (Das et al., 2022). Furthermore, according to Das et al. (2022), many studies have shown that deep learning models surpass machine learning algorithms when dealing with large and complex datasets for better DR generalizations. Convolutional neural networks (CNN) and ensemble learning methods are some of the more popular approaches in this regard that use retinal pictures to make early predictions. Recent developments in CNNs have rendered them a popular method in image classification using a hierarchy of features (Xu et al., 2017). Through Xu et al. (2017)'s study, the efficacy of this method was investigated based

on real retina data with results showing 94.5% accuracy, the highest among prior handmade feature-based classifiers.

In short, CNNs have a considerable advantage over human feature selection approaches since they can extract hierarchical features from raw images. They may identify intricate patterns in data that human specialists or simpler models may overlook. CNN designs, including ResNet and VGGNet, are increasingly often employed in medical image analysis (Mall et al., 2023). Early identification of diabetic retinopathy can avert serious complications including blindness. In this regard, CNNs' ability to detect minuscule abnormalities in retinal scans far quicker than standard approaches makes them a valuable tool in diabetes control.

2. *Hybridization of Deep Learning and Traditional Methods*

Hybrid models can be used to further enhance the predictability and interpretability in diagnosing diabetes through the combination of deep learning frameworks (for example, CNNs) with classic machine learning methods such as 'Random Forest' or 'Logistic Regression'. For instance, Simaiya et al. (2022) conducted a novel, multistage ensemble technique that blends neural networks and decision trees for diabetes prediction, demonstrating greater accuracy and recall than solo models.

Such a hybridized approach can leverage deep learning's capacity to detect complicated patterns in unstructured data (scans, images, genomic data, and so forth) while utilizing more interpretable models for structured data, offering a balanced mix of predictive accuracy and clinical interpretability. In clinical settings, these hybrid models can offer medical practitioners with a comprehensive yet easy-to-understand decision framework and simultaneously harness deep learning's predictive power to the fullest.

D. AI-powered Patient Care

The introduction of chatbots in healthcare services has paved the way in medical communication, fostering relationships between clinics and patients. The adoption of AI, deep learning, and artificial neural networks in chatbots technology helps the bots improve their ability to deliver their services in a sensible and empathetic manner (Siddique & Chow, 2021). For instance, chatbot systems (for example, the Mandy and Nurse Chatbot) use AI-powered natural language processing (NLP) to support patient intake, offer assistance 24/7, and provide personalized care through intelligent dialogue systems that replicate human conversation.

AI and ML have become prevalent in radiology and radiotherapy, most notably through chatbots and virtual support networks (Siddique & Chow, 2021). NLP is implemented in these systems to cluster patient data for further analysis on patient behavior and clinical parameters. Radiotherapy uses cloud computing and machine learning algorithms, including Monte Carlo simulations to enhance treatment planning and optimize dosage computation, especially in demanding instances like breast cancer.

E. Implications and Future Development

Within the context of diabetes prediction, early detection of those at risk is critical for successful intervention strategies. However, widespread diabetes testing would be expensive, laborious, and stressful for medical personnel (Chowdhury et al., 2024). Perhaps, one of the more pressing concerns related to the use of AI/ML healthcare is transparency. For instance, the application of deep learning in image processing is complex and difficult to interpret fully. If a patient is notified that a scan has led to a severe diagnosis, they will want to know how this is possible. Even healthcare professionals who are accustomed to the workings of deep learning may struggle to explain how they work (Davenport & Kalakota, 2019).

It must be stated that AI/ML is not completely error-free and will likely make mistakes in diagnosis and predictions but holding them accountable would be a challenging task. Therefore, it is important that medical institutions develop a coherent structure to monitor risks and set up proper governance to prevent adverse outcomes from fully realizing.

The most urgent matter in this regard is not so much about determining the algorithms would be effective as validating their adoption in routine clinical settings. For ML to be widely accepted and thrive in medical operations, the systems must be approved by governing bodies, EHR-system incorporated, have a standardized process, taught to healthcare workers, and updated constantly (Davenport & Kalakota, 2019).

3. METHODOLOGY

The dataset in this study is taken from the Pima Indian Diabetes Dataset, which is administered by the National Institute of Diabetes and Digestive and Kidney Diseases. The dataset consists of several diagnostics indicators, featuring 768 instances across 9 features/attributes. The predictors are as follows:

1. **Preg -** The number of times the subjects had been pregnant.
2. **Plas -** Plasma glucose concentration level measured 2 hours after consuming a glucose solution.
3. **Pres -** Diastolic blood pressure in mmHG
4. **Skin -** Tricep skinfold thickness (mm)
5. **Insu -** 2-hour serum insulin
6. **Mass -** Body mass index
7. **Pedi -** Diabetes pedigree function

8. **Age -** Age of the subject measured in years

The aim of the study is to use all of the above-mentioned attributes to predict the 'Class' of the patient. 'Class' is the dependent variable divided into 'tested positive' and 'tested negative'. To carry out the analysis, the data was uploaded onto Waikato Environment for Knowledge Analysis software, otherwise known as 'WEKA'. All the attributes are summarized in Figure 1 below.

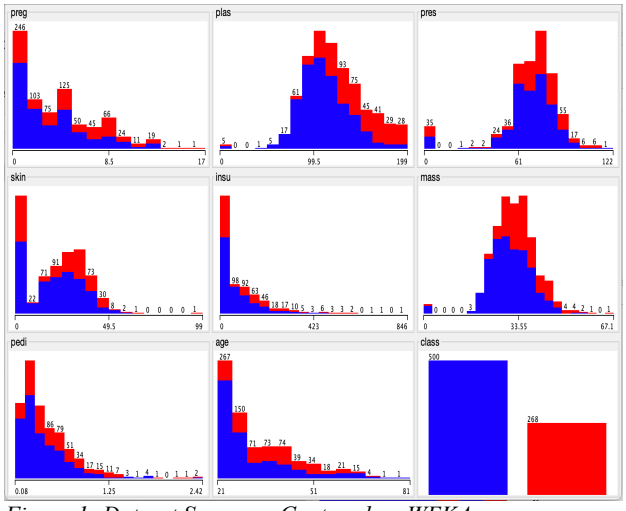

Figure 1: Dataset Summary Captured on WEKA

The findings were evaluated using seven metrics: accuracy, precision, sensitivity, specificity, F-score, ROC Area, and MCC (Matthew Correlation Coefficient). These variables are created from the confusion matrix, demonstrating the proportions of the actual and expected result classes on the testing set. The formulas for the metrics are as follows:

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

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

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

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$\text{F-1} = 2 \text{ x } \frac{Precision \; x \; Recall}{Precision + Recall} \quad (5)$$

$$\text{MCC} = \frac{TP \; x \; TN - FP \; x \; FN}{\sqrt{(Tp + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6)$$

3.1 Data Pre-processing

Prior to training the models, data processing techniques are used to significantly increase the overall quality of the training. Data preprocessing is an important phase in the information discovery process since quality data is the cornerstone for improving decision making. This step involved several strategies such as data normalization, transformation, and reduction (see Feature Engineering). From the dataset in Figure 1, one can see an imbalance distribution between positive and negative classes, with the former significantly outnumbered by the latter (500 versus 268). Such an imbalance might cause some biases toward the majority class, resulting in better classifier prediction for the majority than the minority class (Jadhav et al., 2022). The data preparation process could be helpful in improving model performance with imbalanced data. Data normalization helps standardize feature magnitudes, ensuring that features are on a comparable scale to minimize the impact of the majority class on the training process.

Moreover, given that the dataset's parameters are of different scales, they must be normalized in order to achieve a better outcome when executing the models. The data was normalized with the aim of rescaling the attributes to the range of 0 and 1. Normalization is particularly apt when the data distribution is unknown or not Gaussian (a bell curve), which is applicable in this study (see Figure 1). The process was done on WEKA using and applying the 'Normalize' filter.

Following normalization, the dataset was processed through 10-fold cross validation (the default setting) and nine algorithms based on tree, function, and rule-based classifiers. Tree classifiers include 'Random Forest', 'J48', and 'Random Tree'; Function classifiers consist of 'Logistic Regression' (Log Reg), 'Multilayer Perceptron' (MLP), and 'Stochastic Gradient Descent' (SGD); whereas rules-based classifiers encompass 'Java Repeated Incremental Pruning' (JRip), 'One Rule' (OneR), and 'Partial Decision Tree' (PART). Each algorithm was tested 10 times (10 runs of 10-fold cross validation) using various random number seeds. This yielded 10 slightly varied outcomes for each assessed algorithm, a tiny population that may be examined by statistical methods later.

4. ANALYSIS AND DISCUSSION

This section will provide a comprehensive look into how different machine learning algorithms perform on the dataset. The classifiers are categorized into three types - Trees, Function, and Rules. These classifiers' performances were recorded in tabular form, using metrics such as accuracy, sensitivity/recall, specificity, precision, f-score, ROC, and MCC for assessing class label.

Table 1. Performance of Tree-based Algorithms

Tree-Based Method			
Metric	Random Forest	J48	Random Tree
Accuracy	0.758	0.738	0.681
Specificity	0.836	0.814	0.746
Precision	0.754	0.735	0.684
Recall	0.612	0.632	0.560
F1-Score	0.755	0.736	0.682
MCC	0.458	0.417	0.303
ROC Area	0.820	0.751	0.653

Note: J48 - C4.5 Algorithm; MCC - Matthews Correlation Coefficient; ROC - Receiver Operating Characteristic

Table 2. Performance of Function-based Algorithms

Function-based Method			
Metric	Log	MLP	SGD
Accuracy	0.772	0.754	0.78
Specificity	0.718	0.832	0.896
Precision	0.767	0.75	0.776
Recall	0.571	0.608	0.56
F1-Score	0.765	0.751	0.771
MCC	0.48	0.449	0.497
ROC Area	0.832	0.793	0.73

Note: Log - Logistic Regression; MLP - Multilayer Perceptron; SGD - Stochastic Gradient Descent; MCC - Matthews Correlation Coefficient; ROC - Receiver Operating Characteristic.

Table 1 looks at the performance of Random Forest, J48, and Random Tree classifiers. Random Forest is an ensemble method that aggregates the outputs of multiple decision trees to enhance the model's accuracy while also reducing overfitting. J48, also known as the implementation of Quinlan's 4.5 algorithm, is noted for its ability to simplify decision trees and make them more interpretable (Chang et al., 2022). Interpretability is a crucial factor in explaining how a prediction or diagnosis is reached and so, for this reason, J48 was considered for this study. Random Trees is another straightforward algorithm used in this study where each tree is constructed by selecting random features. The simplicity of the model makes it less computationally demanding than Random Forest, which might be suitable for cases where quick decision-decision making is required.

Overall, Random Forest emerges as the strongest performer among all tree-based algorithms, ranking ahead of J48 and Random Trees across almost all metrics, except for Recall where it is bested by J48. Furthermore, its ability to work with large feature spaces, as well as its resistance to overfitting, makes it the preferable tree-based model for this task.

However, comparing these findings to a comparable study done by Bhat et al. (2023), it was discovered that XGBoost and LightGBM, ensemble learning techniques, outperform standard Random Forest due their superior boosting techniques. By using XGBoost, their diabetes prediction score was as high as 85% (Bhat et al., 2023). These models are intended to handle big, imbalanced datasets more effectively and provide quicker training times than RF.

After testing the tree-based algorithms, function classifiers were looked at - Logistic Regression, Multilayer Perceptron, Stochastic Gradient Descent (Table 2). Logistic Regression (LR) is a highly interpretable model employing a linear combination of input data to compute the final binary outcome, for instance, having or not having diabetes. Conversely, Multilayer Perceptron is tasked with modeling complex, non-linear relationships between input features and output, making it more ideal for detecting intricate particulars that logistic regression might have missed. Finally, Stochastic Gradient Descent (SGD) was considered for the investigation due to its computational efficiency, as it is capable of handling large datasets and high-dimensional data (Fjellström & Nyström, 2022). The algorithm's high scalability makes it applicable to a wide range of models, from linear to neural networks.

Based on the results observed in Table 2, SGD is the top performing algorithm for diabetes prediction. Its performance is relatively high across almost all metrics, scoring particularly well in Precision and Specificity, two elements that are crucial in chronic health diagnosis (Grabler et al., 2017). That said, Logistic Regression nevertheless remains a good model due to its simplicity and interpretability, achieving the highest ROC Area score. Although MLP's scores are the lowest among all function classifiers, the algorithm still shows promise thanks to its relatively high recall score in relation to the other two classifiers. Its low performance in other aspects might suggest that this dataset is not large enough to fully make use of its capabilities.

As observed, one can see that the function-based method might have performed better than the tree-based. These

three function classifiers, especially SGD, can deal with high dimensional data. However, when comparing them to deep learning models, they are not quite effective. This can be seen in a study done by Xu et al. (2017) who reported that they achieved an accuracy score of over 94% when implementing CNNs to diagnose diabetic retinopathy from retinal pictures. Nevertheless, SGD's scalability and speed are some of its perks. In contrast, deep learning models, while more powerful, would typically require more computer resources and are less interpretable.

Table 3. Performance of Rule-based Algorithms

Rules-based Method			
Metric	**JRip**	**OneR**	**PART**
Accuracy	0.76	0.715	0.753
Specificity	0.856	0.866	0.844
Precision	0.755	0.703	0.747
Recall	0.582	0.433	0.582
F1-Score	0.755	0.699	0.748
MCC	0.457	0.334	0.441
ROC Area	0.739	0.649	0.794

Note: JRip - Java Repeated Incremental Pruning to Produce Error Reduction; OneR - One Rule; PART - Partial Decision Trees; MCC - Matthews Correlation Coefficient; ROC - Receiver Operating Characteristic.

Table 3 above shows the performance of three rules-based algorithms - JRip, OneR, and PART. These algorithms are particularly beneficial in situations when interpretability and simplicity are essential. They generate simple decision rules for healthcare predictions in various contexts. The first among them is JRip, an efficient rule-learning algorithm that builds sets of rules on association principles and reduced error pruning (Simaiya et al., 2022). It strikes the right balance between the model's complexity and predicted accuracy, a critical aspect in medical diagnostics where accuracy and interpretability are equally vital. The second rule classifier is OneR, which creates a single rule according to the characteristic with the highest classification accuracy. It evaluates each attribute individually and picks the one that produces the greatest results. This method is highly interpretable, and its simplicity makes it easy to apply. Lastly, PART is also referred to as 'Partial Decision Tree', an algorithm that incorporates components of decision trees and rule-based classifiers. As its name might suggest, PART develops partial decision trees, which are subsequently converted into a set of rules.

Based on the results indicated in Table 3, JRip performs the best in most metrics, achieving the highest scores among all the rules-based classifiers in accuracy, precision, F-1 measure, and recall (same score as PART), making it a fine choice for diabetes diagnosis where balanced performance is deemed necessary. OneR has the highest specificity score but falls short in other metrics. Meanwhile, PART scores the highest on the ROC Area, which highlights its ability to distinguish between positive and negative cases. From this observation, it can be deduced that JRip and Part high performances are attributable to their complex algorithms compared to OneR. JRip's iterative pruning enables it to make better generalizations, evidenced by a balance of accuracy and precision, whereas PART's hybridization of partial decision trees and rules-based learning provides enough flexibility to deal with complex data.

Table 4. Comparative Performance of All Algorithms

	Tree-Based Method			**Rule-based Method**			**Function-based Method**		
Metric	**Random Forest**	**J48**	**Random Tree**	**JRip**	**OneR**	**PART**	**Log**	**MLP**	**SGD**
Accuracy	0.758	0.738	0.681	0.76	0.715	0.753	0.772	0.754	0.78
Specificity	0.836	0.814	0.746	0.856	0.866	0.844	0.718	0.832	0.896
Precision	0.754	0.735	0.684	0.755	0.703	0.747	0.767	0.75	0.776
Recall	0.612	0.632	0.56	0.582	0.433	0.582	0.571	0.608	0.560
F1-Score	0.755	0.736	0.682	0.755	0.699	0.748	0.765	0.751	0.771
MCC	0.458	0.417	0.303	0.457	0.334	0.441	0.48	0.449	0.497
ROC Area	0.82	0.751	0.653	0.739	0.649	0.794	0.832	0.793	0.73

In terms of the three methods' overall performance, function-based classifiers, such as Logistic Regression and SGD slightly edge ahead of the trees and rules-based methods, offering high accuracy and F-1 scores that signify a great balance between precision and recall. 'Rule-based Method' stands out with respect to interpretability. JRip and OneR create rules that are easy to understand but also have enough complexity to handle intricate patterns. This interpretable factor is of great value in clinical diagnostics where understanding the basis for a prediction is critical. Lastly, with reference to robustness, Random Forest, which is the best performing classifier in the tree-based method, can also be considered given its adeptness at handling complex datasets with very minimal chance of overfitting.

As for comparing each individual classifier across all methods, it is evident that SGD, Stochastic Gradient Descent, emerges as the most effective algorithm based on the metrics' comprehensive overview. It achieves the highest scores in nearly all metrics, except for Recall and ROC Area. Yet, the algorithm's high specificity and precision minimize the likelihood of false positives, but

its sensitivity/recall remains doubtful compared to other algorithms.

After SGD, Logistic Regression is another excellent algorithm in the function-based method, especially given its large ROC Area, which implies its effectiveness in identifying diabetes and non-diabetic patients (true positive and false positive cases). It also has a higher recall score than SGD, which is deemed the most important metric in medical diagnosis,

In instances where the simplicity of rules is valued, JRip proves to be the best alternative. Within the context of healthcare, particularly when one must account for the degree of which the model will be applied in practice, JRip is most likely the best choice because of its interpretability, which is extremely useful for medical decision-making. However, if predictive performance is prioritized, Random Forest should also be considered with reference to its higher Recall and robustness, a factor that is exemplified by the larger ROC Area.

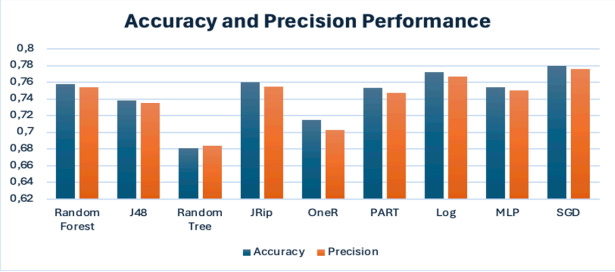

Figure 2: Accuracy & Precision Performance

In terms of accuracy, SGD, Logistic Regression, JRip are the top 3 performers, scoring 0.78, 0.772, 0.76, respectively. This means that these three algorithms can predict the correct outcome. However, accuracy alone can be misleading, especially within the context of medical diagnosis where there are many other crucial factors that need to be considered, such as Recall, Specificity, F-1, and Precision. For precision, these 3 algorithms remain as top performers, indicating that when a diabetic case is predicted as positive, the prediction is likely correct.

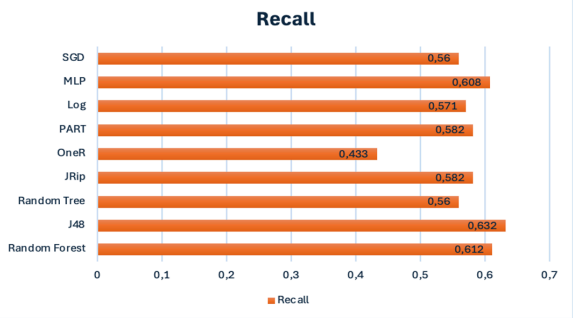

Figure 3: Recall Performance

J48 and Random Forest are most effective at identifying positive diabetic cases, considering their high recall scores. Surprisingly, despite performing well in accuracy and precision, SGD's recall is lower than expected as this suggests that this algorithm might misclassify many diabetic cases (false negatives). This metric is particularly important in disease prediction because it minimizes the probability of false negatives—when a diabetic patient is mistakenly identified as non-diabetic, possibly leading to the condition left untreated (Government of Canada, 2023).

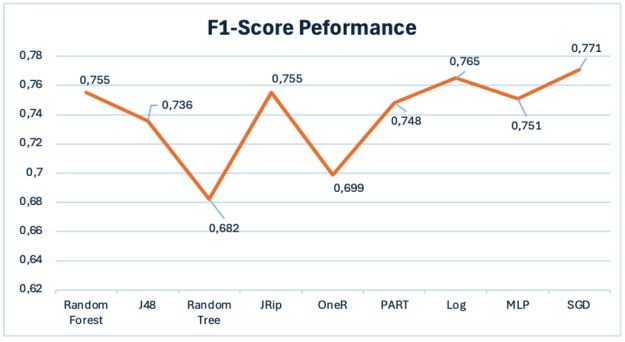

Figure 4: F1-score Performance

SGD once again comes out on top with the highest F1-score, seconded by Logistic Regression, and then JRip. This implies that these algorithms maintain a balance between precision and recall. On top of this, as referred to the ROC Area performance in Figure 6, Logistics Regression yet demonstrates a strong performance with the largest ROC Area (0.832), showing its capability at identifying true positive cases and false positive cases.

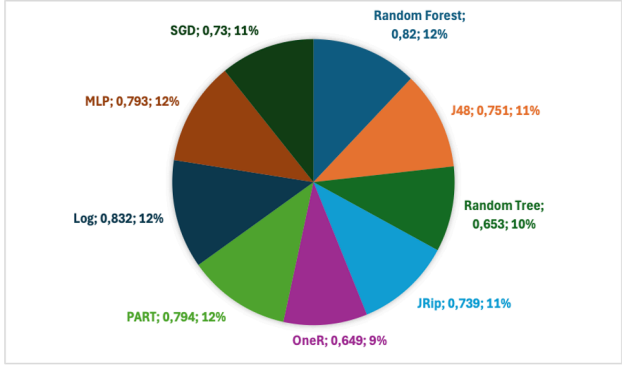

Figure 5: ROC Area Performance

4.1 Feature Engineering

Feature engineering is the process of choosing, transforming, and producing useful input features from raw data for supervised machine learning (Patel, 2024). For evaluating whether the machine learning model could be improved by feature selection, 'InfoGainAttributeEval'

was chosen as an evaluator employed to rank the usefulness of the features.
The following features were ranked in order:
Plas, mass, age, insu, skin, preg, pedi, and pres.

The attribute score above shows that **plasma glucose concentration** level is the most critical feature in diagnosing diabetes. Based on research, type 2 diabetes can develop as early as four to seven years of age before a clinical diagnosis (Gurung et al., 2024). The plasma glucose level can be used as the basis for monitoring the disease and predicting complications among Type 2 diabetic patients. On a related note, the American Diabetes Association (n.d.) has established the limits for prediabetes (100-125 mg/dL fasting) and diabetes (>126 mg/dL). For this reason, it is not unexpected that this attribute is placed first, as it serves as the foundation for diabetes diagnosis.

The second most important feature is **'Body Mass Index'**. This is particularly the case when obesity is often deemed as a key risk factor for Type 2 diabetes (Daley & Yashi, 2023). BMI is an indicator of body fat based on height and weight. A high BMI score is closely associated with an increased risk of being diagnosed with diabetes. The condition elevates insulin resistance, immobilizing the body's cells' ability to respond to insulin, resulting in increased blood sugar levels. Research done by Ganz et al. (2014) shows that people who score more than 30 on their BMI are much more likely to develop Type 2 diabetes.

Ranked behind BMI is **'Age'**. Research has shown that the likelihood of developing diabetes grows with age (Mordarska & Godziejewska-Zawada, 2017). Clinical studies have shown that people over the age of 45 have a high risk of developing diabetes (Flores et al., 2020). This is attributable to the insulin secretion deficit and insulin resistance that develop with age and changes in body composition.

Next is the **'Insu'** feature, which stands for 2-hour Serum Insulin, that measures an individual's insulin levels following a 2-hour glucose tolerance test, showing how the body might react to glucose. Insulin levels found in insulin-resistant individuals are usually high because the pancreas has to overwork to produce more insulin to overcome the resistance (Wilcox, 2005). However, insulin production may decrease over time as the pancreas loses effectiveness. Tracking this alteration can be vital for detecting prediabetes and Type 2 diabetes.

Ranked fifth is the '**Skin'** feature that uses skinfold thickness as a baseline for measuring body fat percentage. Its lower ranking may be because skinfold thickness is not as direct or effective a predictor of diabetes as BMI, which provides a more comprehensive picture of a person's weight in relation to their height. Skinfold thickness may give supplementary information, although it is not usually a primary factor in diabetes risk assessment.

The last three attributes are **'Preg'** (pregnancy)**, 'Pedi' (**diabetes pedigree function)**,** and '**Pres' (**diastolic blood pressure) show that they are not significant in predicting diabetes. Although their attribute scores are low, they help add context to the model in fine-tuning the prediction, especially certain subgroups such as older individuals and women with a history of gestational diabetes.

Overall, the 'Info Gain Evaluator' facilitates better understanding of the features' influence on the outcome variable. Judging from the evaluator, feature **'Plas'** (plasma glucose concentration level) is ranked first with the highest score, whereas feature **'Pres'** (diastolic blood pressure) comes last. The Random Forest classifier was used as the basis for the study to see whether the removal of the lowest score feature would improve the model's performance.

4.2 Feature Reduction Analysis

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         582               75.7813 %
Incorrectly Classified Instances       186               24.2188 %
Kappa statistic                          0.4566
Mean absolute error                      0.3106
Root mean squared error                  0.4031
Relative absolute error                 68.3405 %
Root relative squared error             84.5604 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.836    0.388    0.801      0.836    0.818      0.458    0.820     0.886     tested_negative
               0.612    0.164    0.667      0.612    0.638      0.458    0.820     0.679     tested_positive
Weighted Avg.  0.758    0.310    0.754      0.758    0.755      0.458    0.820     0.814

=== Confusion Matrix ===

   a   b   <-- classified as
 418  82 |   a = tested_negative
 104 164 |   b = tested_positive
```

Figure 6: RF's Performance Before Feature Removal

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         586               76.3021 %
Incorrectly Classified Instances       182               23.6979 %
Kappa statistic                          0.4664
Mean absolute error                      0.3153
Root mean squared error                  0.4081
Relative absolute error                 69.3804 %
Root relative squared error             85.6197 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.844    0.388    0.802      0.844    0.823      0.468    0.814     0.891     tested_neg
               0.612    0.156    0.678      0.612    0.643      0.468    0.814     0.669     tested_pos
Weighted Avg.  0.763    0.307    0.759      0.763    0.760      0.468    0.814     0.813

=== Confusion Matrix ===

   a   b   <-- classified as
 422  78 |   a = tested_negative
 104 164 |   b = tested_positive
```

Figure 7: RF's Performance After Feature Removal

As evident from the above figures, the performance of Random Forest slightly improves across almost all fronts. The recall score also increases, signifying the model's stability in fighting against misclassification. Although the ROC Area decreases marginally from 0.820 to 0.814, it nevertheless suggests that feature reduction does help streamline the model without compromising its predictive ability. After this was done, other algorithms were immediately tested to see if they would have the same outcome as Random Forest.

Aside from Random Forest, feature reduction has proven advantageous to J48 and Logistic Regression with slight improvement in key areas. Conversely, SGD and JRip do not have any gain from said process but rather experience a marginal drop in performance. This might be due to their reliance on the complete feature set in order to perform optimally.

From this one can see that rule-based classifiers, such as JRip and OneR, are most impacted by feature engineering because they rely on simple, interpretable rules, making them very sensitive to feature significance. Tree-based classifiers, prime example being Random Forest, see an improvement partly because of how they can take advantage of feature significance within their frameworks. Function-based classifiers, such as Logistic Regression, benefit from feature reduction since it simplifies the model and prevents overfitting, whereas SGD and MLP are more resistant to feature engineering, although feature reduction still improves performance.

5. CONCLUSION

In conclusion, this research has efficiently investigated the functionality of various machine learning algorithms in predicting and diagnosing diabetes. The primary goal was to investigate 3 machine learning methods – trees-based, rules-based, and function-based – to find the most optimal algorithm for clinical use. While SGD, a function-based algorithm, scores high in almost all the metrics, its low recall score is a major pain point since misclassifying diabetic patients (increasing the likelihood of false positives) poses grave danger in healthcare context, resulting in the disease being left untreated. Logistic Regression and Random Forest may be strong candidates in this case because of their more balanced performance. Logistic Regression, with its large ROC area, high precision, accuracy, and F-1 measure, is good for accurately identifying diabetic and non-diabetic patients across different thresholds. However, Random Forest's robustness and well-balanced metrics, and most importantly high recall score, make it an optimal classifier in a case like this where false positives must be reduced at all cost for early intervention.

Key features such as plasma glucose concentration, body mass index, and age were the top three most influential predictors in the dataset, thereby demonstrating their usefulness in assessing risks of diabetes. Feature engineering was also conducted by assessing the significance of each feature on the outcome variable. This process has proven that by modifying or removing a feature of low significance, the model's performance can be slightly improved. The 'Info Gain Attribute Eval' method was employed and proved effective in using feature reduction to stabilize classifiers' performances through relevant attributes.

The findings in this study have great implications for clinical diagnostics. Machine Learning classifiers, as shown throughout this study, can help identify diabetic patients. With early on identification, doctors can provide these patients with personalized treatment and improve their quality of life. However, despite these findings, there are a few limitations that should be noted. To start, this dataset might not accurately reflect the general population since it is leaned more toward a specific group of people (Pima Indians). This shortcoming may impair the model's ability to generalize for other groups of patients. Moreover, although the dataset is comprehensive, it might have overlooked other attributes that are more applicable to other demographics.

Therefore, future studies must address the above-mentioned constraints to improve the machine learning classifiers' credibility and make them applicable to real-world healthcare practice. For instance, other health-related datasets with comparable features to those in this can be used and assessed to broaden the scope of this study. On the same note, further research on this subject should attempt to incorporate as many data points as possible from the same group of patients over time. Doing so may culminate in the building of a sophisticated prediction and diagnosis model for diabetes intervention. According to Agliata et al. (2023), such an advanced model could be accomplished through complex pattern and context identifications, as well as neural network technologies – for example, Long-Short-Term Memory Models.

REFERENCES

[1] Agliata, A., Giordano, D., Bardozzo, F., Bottiglieri, S., Facchiano, A., & Tagliaferri, R. (2023). Machine learning as a support for the diagnosis of Type 2 diabetes. *International Journal of Molecular Sciences*, *24*(7), 6775. *https://doi.org/10.3390/ijms24076775*

[2] Ahmed, T. M. (2016). Using data mining to develop models for classifying diabetic patient control level based on historical medical records. *Journal of Theoretical and applied information Technology*, *87*(2), 316.

[3] Alam, T. M., Iqbal, M. A., Ali, Y., Wahab, A., Ijaz, S., Baig, T. I., 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. *https://doi.org/10.1016/j.imu.2019.100204*

[4] Allen, A., Iqbal, Z., Green-Saxena, A., Hurtado, M., Hoffman, J., Mao, Q., & Das, R. (2022). Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Research & Care*, *10*(1), e002560. https://doi.org/10.1136/bmjdrc-2021-002560

[5] American Diabetes Association. (n.d.). *Diabetes Diagnosis & Tests*. https://diabetes.org/about-diabetes/diagnosis

[6] Baadel, S., Thabtah, F., Lu, J. (2020). A clustering approach for Autistic trait classification, Informatics for Health and Social Care, 45 (3), 309-326.

[7] Berge, G. T., Granmo, O., Tveit, T. O., Ruthjersen, A. L., & Sharma, J. (2023). Combining unsupervised, supervised and rule-based learning: the case of detecting patient allergies in electronic

health records. *BMC Medical Informatics and Decision Making*, *23*(1). https://doi.org/10.1186/s12911-023-02271-8

[8] Bhat, S. S., Banu, M., Ansari, G. A., & Selvam, V. (2023). A risk assessment and prediction framework for diabetes mellitus using machine learning algorithms. *Healthcare Analytics*, *4*, 100273. https://doi.org/10.1016/j.health.2023.100273

[9] Chang, V., Bailey, J., Xu, Q. A., & Sun, Z. (2022). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. *Neural Computing and Applications*, *35*(22), 16157–16173. https://doi.org/10.1007/s00521-022-07049-z

[10] Chowdhury, M. M., Ayon, R. S., & Hossain, M. S. (2024). An investigation of machine learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset. *Healthcare Analytics*, *5*, 100297. https://doi.org/10.1016/j.health.2023.100297

[11] Daley S & Yashi K. (2023). Obesity and Type 2 Diabetes. https://www.ncbi.nlm.nih.gov/books/NBK592412/

[12] Das, D., Biswas, S. K., & Bandyopadhyay, S. (2022). Detection of Diabetic Retinopathy using Convolutional Neural Networks for Feature Extraction and Classification (DRFEC). *Multimedia Tools and Applications*, *82*(19), 29943–30001. https://doi.org/10.1007/s11042-022-14165-4

[13] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

[14] Dudkina, T., Meniailov, I., Bazilevych, K., Krivtsov, S., & Tkachenko, A. (2021). Classification and Prediction of Diabetes Disease using Decision Tree Method. *Symposium on Information Technologies & Applied Sciences*, 163–172. http://ceur-ws.org/Vol-2824/paper16.pdf

[15] Fjellström, C., & Nyström, K. (2022). Deep learning, stochastic gradient descent and diffusion maps. *Journal of Computational Mathematics and Data Science*, *4*, 100054. https://doi.org/10.1016/j.jcmds.2022.100054

[16] Flores, Y. N., Toth, S., Crespi, C. M., Ramírez-Palacios, P., McCarthy, W. J., Briseño-Pérez, A., Granados-García, V., & Salmerón, J. (2020). Risk of developing pre-diabetes or diabetes over time in a cohort of Mexican health workers. *PLoS ONE*, *15*(3), e0229403. https://doi.org/10.1371/journal.pone.0229403

[17] Ganz, M. L., Wintfeld, N., Li, Q., Alas, V., Langer, J., & Hammer, M. (2014). The association of body mass index with the risk of type 2 diabetes: a case-control study nested in an electronic health records system in the United States. *Diabetology & metabolic syndrome*, *6*(1), 50. https://doi.org/10.1186/1758-5996-6-50

[18] Government of Canada. (2023, October 23). *Care during pregnancy: Family-centred maternity and newborn care national guidelines*. Canada.ca. https://www.canada.ca/en/public-health/services/publications/healthy-living/maternity-newborn-care-guidelines-chapter-3.html

[19] Grabler, P., Sighoko, D., Wang, L., Allgood, K., & Ansell, D. (2017). Recall and cancer detection rates for screening mammography: finding the sweet spot. *American Journal of Roentgenology*, *208*(1), 208–213. https://doi.org/10.2214/ajr.15.15987

[20] Gurung, P., Zubair, M., & Jialal, I. (2024, February 27). *Plasma glucose*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK541081/

[21] Haripriya, G., Abinaya, K., Aarthi, N., & Kumar, P. (2021). Random Forest Algorithms in Health Care Sectors: A Review of Applications.

[22] Hounguè, P., & Bigirimana, A. G. (2022). Leveraging PIMA Dataset to diabetes Prediction: case study of deep neural networks. *Journal of Computer and Communications*, *10*(11), 15–28. https://doi.org/10.4236/jcc.2022.1011002

[23] Iparraguirre-Villanueva, O., Espinola-Linares, K., Castañeda, R. O. F., & Cabanillas-Carbonell, M. (2023). Application of machine learning models for early detection and accurate classification of Type 2 diabetes. *Diagnostics*, *13*(14), 2383. https://doi.org/10.3390/diagnostics13142383

[24] Jadhav, A., Mostafa, S. M. M., Elmannai, H., & Karim, F. K. (2022). An empirical assessment of performance of data balancing techniques in classification task. *Applied Sciences*, *12*(8), 3928. https://doi.org/10.3390/app12083928

[25] Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, *3*, 58–73. https://doi.org/10.1016/j.ijin.2022.05.002

[26] Jonnalagadda, S., Cohen, T., Wu, S., & Gonzalez, G. (2012). Enhancing clinical concept extraction with distributional semantics. *Journal of Biomedical Informatics*, *45*(1), 129–140. https://doi.org/10.1016/j.jbi.2011.10.007

[27] Kolasa, K., Admassu, B., Hołownia-Voloskova, M., Kędzior, K. J., Poirrier, J. E., & Perni, S. (2024). Systematic reviews of machine learning in healthcare: a literature review. *Expert review of pharmacoeconomics & outcomes research, 24(1), 63–115. https://doi.org/10.1080/14737167.2023.2279107*

[28] Mall, P. K., Singh, P. K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., & Ganzha, M. (2023). A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthcare Analytics*, *4*, 100216. https://doi.org/10.1016/j.health.2023.100216

[29] Mordarska, K., & Godziejewska-Zawada, M. (2017). Diabetes in the elderly. Menopause Review/Przegląd Menopauzalny, 16(2), 38-43. https://doi.org/10.5114/pm.2017.68589

[30] Mousa, A., Mustafa, W., Marqas, R. B., & Mohammed, S. H. M. (2023). A comparative study of diabetes detection using the PIMA Indian Diabetes Database. *The Journal of the University of Duhok*, *26*(2), 277–288. https://doi.org/10.26682/sjuod.2023.26.2.24

[31] Nguyen, L. P., Tung, D. D., Nguyen, D. T., Le, H. N., Tran, T. Q., Van Binh, T., & Pham, D. T. N. (2023). The utilization of machine learning algorithms for assisting physicians in the diagnosis of diabetes. *Diagnostics*, *13*(12), 2087. https://doi.org/10.3390/diagnostics13122087

[32] Pal, S., Mishra, N., Bhushan, M., Kholiya, P. S., Rana, M., & Negi, A. (2022). Deep learning techniques for prediction and diagnosis of diabetes mellitus. *2022 International Mobile and Embedded Technology Conference (MECON)*. https://doi.org/10.1109/mecon53876.2022.9752176

[33] Patel, H. (2024, April 29). *Feature Engineering explained*. Built In. https://builtin.com/articles/feature-engineering#:~:text=Apr%2029%2C%202024-,Feature%20engineering%20is%20the%20process%20of%20selecting%2C%20manipulating%20and%20transforming,used%20in%20a%20predictive%20model.

[34] Salazar-Reyna, R., Gonzalez-Aleu, F., Granda-Gutierrez, E. M., Diaz-Ramirez, J., Garza-Reyes, J. A., & Kumar, A. (2020). A systematic literature review of data science, data analytics and machine learning applied to healthcare engineering systems. *Management Decision*, *60*(2), 300–319. https://doi.org/10.1108/md-01-2020-0035

[35] Shetty, D., Rit, K., Shaikh, S., & Patil, N. (2017). Diabetes disease prediction using data mining. *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. https://doi.org/10.1109/iciiecs.2017.8276012

[36] Siddique, S., & Chow, J. C. L. (2021). Machine learning in healthcare communication. *Encyclopedia*, *1*(1), 220–239. https://doi.org/10.3390/encyclopedia1010021

[37] Simaiya, S., Kaur, R., Sandhu, J. K., Alsafyani, M., Alroobaea, R., Alsekait, D. M., Margala, M., & Chakrabarti, P. (2022). A novel

multistage ensemble approach for prediction and classification of diabetes. *Frontiers in Physiology*, *13*. https://doi.org/10.3389/fphys.2022.1085240

[38] Sonia, J. J., Jayachandran, P., Quadir, A., MD, Mohan, S., Sivaraman, A. K., & Tee, K. F. (2023). Machine-Learning-Based diabetes mellitus risk prediction using Multi-Layer Neural Network No-PROP algorithm. *Diagnostics*, *13*(4), 723. https://doi.org/10.3390/diagnostics13040723

[39] Wilcox G. (2005). Insulin and insulin resistance. *The Clinical biochemist. Reviews*, *26*(2), 19–39.

[40] Xu, K., Feng, D., & Mi, H. (2017). Deep Convolutional Neural Network-Based Early Automated detection of diabetic retinopathy using FundUs Image. *Molecules*, *22*(12), 2054. https://doi.org/10.3390/molecules22122054