International Journal of Multidisciplinary Research and Growth Evaluation.

Predicting Diabetes in Libya Using AI: A case Study

Esam Ali [1, 2], **Ibrahim Hamammu** [1]
[1] Department of Physics, Faculty of Science, University of Benghazi, Benghazi 9480, Libya
[2] Robotic Process Automation and Machine Learning, Binaryway LLC, Houston, US

* Corresponding Author: **Esam Ali**

Article Info

ISSN (online): 2582-7138
Volume: 05
Issue: 05
September-October 2024
Received: 01-07-2024
Accepted: 05-08-2024
Page No: 14-19

Abstract
This study investigates the application of artificial intelligence (AI) techniques for predicting diabetes in Libya, a region with limited healthcare data. Utilizing a dataset from the Diabetes and Endocrinology Clinic in Al-Marj, Libya, we examined records of 2,009 patients between April 2019 and May 2023. Our methodology involved applying various AI algorithms, including logistic regression, decision trees, random forests, support vector machines (SVMs), and neural networks, to predict diabetes outcomes. These algorithms were assessed using metrics like precision, recall, F1 score, and the area under the receiver operating characteristic (ROC) curve. The findings indicate a promising potential of AI in forecasting diabetes, particularly when analyzing factors such as fasting blood sugar levels, HbA1c levels, hypertension, heart disease, age, and gender. The majority of the algorithms demonstrated high accuracy, suggesting their utility in enhancing healthcare outcomes in Libya. This research not only provides insights into the effectiveness of AI in diabetes prediction but also underscores the importance of such technologies in regions with scarce health data. It opens pathways for further exploration in the use of AI for healthcare improvement in similar contexts.

DOI: https://doi.org/10.54660/.IJMRGE.2024.5.5.14-19

Keywords: Artificial Intelligence, Diabetes Prediction

1. Introduction

The modern lifestyle has significantly evolved, impacting global population growth and contributing to increased life expectancy. While these advancements have led to a more comfortable lifestyle, they have simultaneously increased the need for comprehensive healthcare. However, this progress and development are accompanied by their challenges affecting human health, particularly the increased prevalence of chronic diseases such as heart disease, diabetes, and hypertension [1]. This underscores the need for an overall enhancement of healthcare.

As Libya faces a growing prevalence of diabetes, a challenge compounded by limited healthcare infrastructure, there is a pressing need for innovative solutions. In this context, Artificial Intelligence (AI) emerges as a pivotal technology. AI's capability to integrate statistical analysis with sophisticated programming makes it particularly suited to address the complexities of diabetes management and prediction in Libya. By harnessing AI, healthcare professionals can overcome data limitations and enhance diagnostic accuracy, offering a transformative approach to tackling this escalating health issue.

Recent technological advancements in healthcare have aimed to elevate health standards. Among these developments, Artificial Intelligence (AI) is a notable technological advancement that has seamlessly integrated statistical analysis with sophisticated programming, streamlining diagnostics and medical report generation. AI applications have proven effective in predicting diseases by studying various factors, reducing medical errors, and identifying various disease types [2].

The integration of AI in healthcare, especially in predicting diseases like diabetes, showcases the power of combining medical expertise with advanced statistical and programming techniques [3]. Advanced AI algorithms, by analyzing a range of health indicators such as blood pressure, heart diseases, gender, age, fasting and cumulative sugar levels, obesity, and genetic predisposition, have demonstrated high accuracy in early diabetes detection [4]. This capability is crucial for predicting diabetes onset before it leads to complications affecting vital organs. The diverse array of AI models, from logistic regression to neural networks, has been instrumental in enhancing disease prediction, thereby improving healthcare management and quality of life for patients at risk of diabetes [5-7]. These technological advancements in AI provide a roadmap for effective disease management, especially in regions with limited healthcare resources, underscoring the potential for AI to revolutionize healthcare in Libya and similar contexts.

The global surge in population growth has led to an exponential increase in medical data, presenting both challenges and opportunities [8]. Concurrently, breakthroughs in biotechnology have revolutionized various fields, with a prominent impact on big data analytics. AI algorithms have emerged as a solution, showcasing rapid tracking capabilities, and facilitating the analysis of vast datasets in diabetes research [9]. This ease of analysis contributes to the advancement of healthcare systems through early detection, raising the effectiveness of preventive interventions before problems arise [10]. Furthermore, these algorithms extend their utility to both predicting diseases and determining the success of various treatment.

This study focuses on gathering and analyzing real-world data from Al-Marj City in Libya. Given the scarcity of statistical health data in Libya, this dataset offers valuable insights into the local health landscape, contributing significantly to healthcare development initiatives. The following sections will delve into the specific algorithms employed in this study, compare their effectiveness, and present a comprehensive summary of our findings.

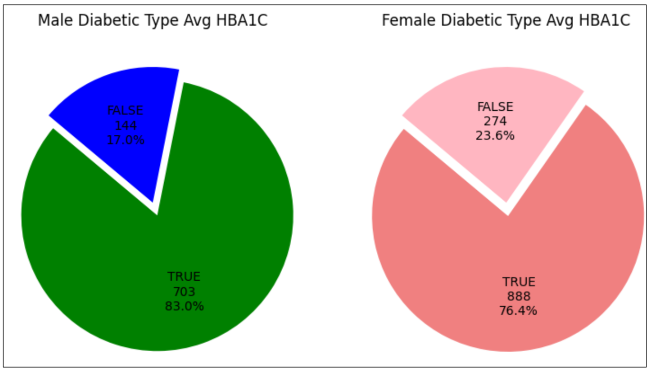

Fig 1: Percentage of Diabetes Incidence among Male and Female Categories

Data Collection

Data were collected continuously over four years from April 26, 2019, to May 9, 2023, at the Diabetes and Endocrinology Clinic in Al-Marj, Libya. The dataset includes both new diagnosed cases and individuals with a history of the disease. In this manuscript, we will focus on computing the average cumulative diabetes test results for each patient, with a more detailed discussion about patient history reserved for a subsequent publication. The study involved 2,009 patients diagnosed with diabetes, analyzing their health status and disease progression. Before exploring patterns or predicting outcomes, it is crucial to understand the data thoroughly, providing readers with insight into the research rationale and the purpose of this specific dataset analysis. Figure 1 illustrates the distribution of diabetes among male and female patients based on their average HBA1C levels (where an average HBA1C level of 6.5% or higher indicates diabetes). It shows that 83.0% of male patients were diabetic, with an average HBA1C level above 6.5%, while 17.0% were not. Among female patients, 76.0% had diabetes, indicated by an average HBA1C level above 6.5%, and 23.6% did not have the disease.

Figure 2 presents a 3D scatter plot visualizing the relationship between average HbA1c levels, fasting blood sugar (FBS) readings, and age, differentiated by gender. Age is plotted along the x-axis, average HbA1c along the y-axis, and average FBS along the z-axis. This graphical representation reveals the variation in HbA1c and FBS levels with age in both genders, which could be indicative of underlying trends valuable for predicting and managing diabetes. Clusters of data points may signify shared characteristics in diabetes progression or its management within the cohort studied.

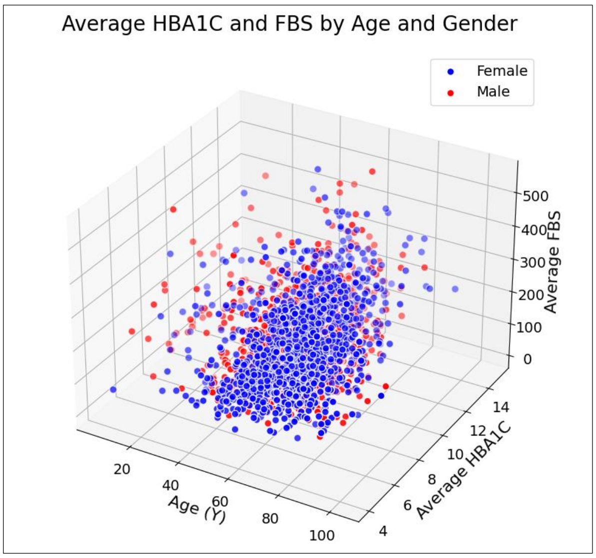

Fig 2: Gender-Based Distribution of Average HbA1c and FBS Levels by Age in a Diabetic Population

The figure 3 shows a series of histograms analyzing the distribution of age, average HbA1c, and average FBS levels within a dataset, likely of diabetic patients. The top row of histograms, with overlaid density plots, offers a probability distribution of the data, indicating the likelihood of data points falling within certain ranges. The bottom row provides a frequency distribution, showing how many data points fall into each bin or interval.

The first column represents the distribution of patients' ages, with the age on the x-axis. It shows a bell-shaped distribution, suggesting a majority of the patients fall into a middle-age bracket. The second column displays the distribution of average HbA1c levels, which are clinical measurements used to assess glucose control over time. The distribution appears slightly right-skewed, indicating a tail of patients with higher HbA1c levels. The third column shows the distribution of average FBS levels, a measure of blood sugar after a period of fasting. This distribution is also right-skewed, suggesting a subset of patients with elevated FBS levels. These histograms are valuable for understanding the demographic and clinical characteristics of the study population.

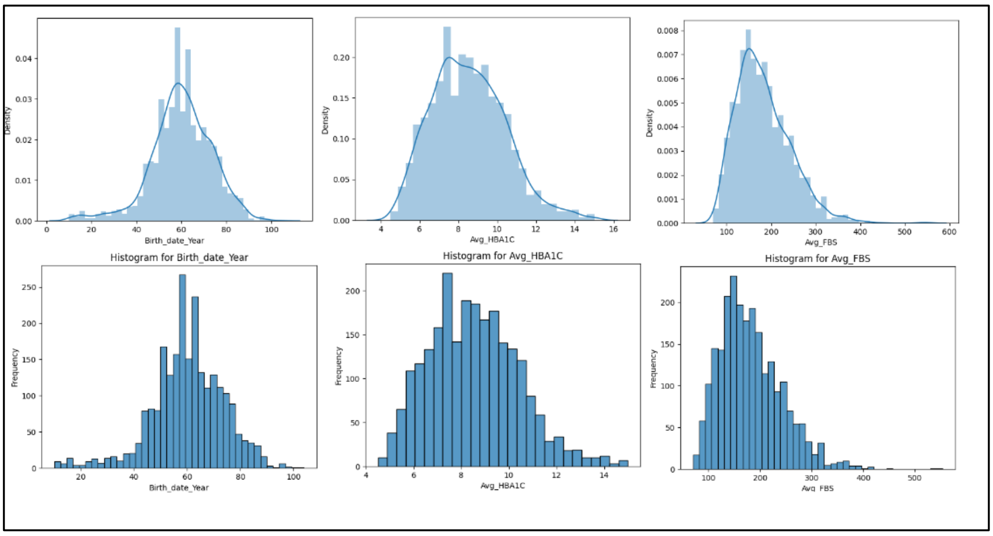

Fig 3: Distribution Analysis of Age, Average HbA1c, and Average FBS in a Diabetic Cohort.

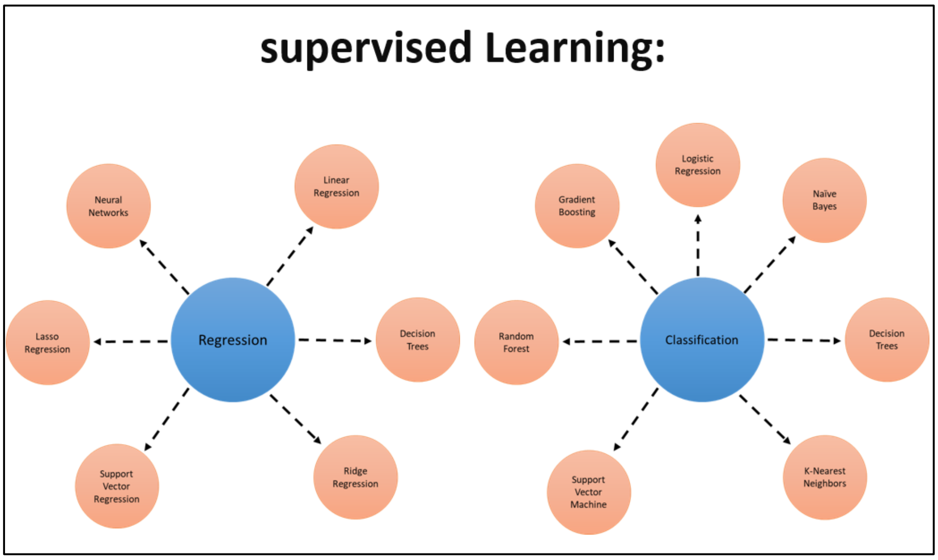

Fig 4: Illustrates the Incidence Rate of Diabetes among Male and Female Categories

Artificial Intelligence in Diabetes Detection

Artificial Intelligence (AI) has emerged as a pivotal tool in the realm of diabetes prediction and early detection, analysing intricate input data to uncover patterns indicative of the disease. This study investigates the utilization of AI methodologies, focusing on patterns linked to common diabetes comorbidities such as heart and kidney diseases, age, and high blood pressure. The AI model employed in this research facilitates the prediction of diabetes, contributing to advancements in forecasting the disease's trajectory.

In recent years, the integration of AI in diabetes research has doubled, yielding promising outcomes in the accurate disease prediction and progression projection. This manuscript explores three primary AI learning methodologies: supervised, unsupervised, and reinforcement learning, It particularly focuses on supervised learning applied to binary and labeled data, training the model to distinguish between individuals with diabetes (True) and those in a healthy state (False), based on relevant medical indicators.

Within the supervised learning paradigm, several algorithms are examined: logistic regression, decision trees, random forests, support vector machines, and neural networks, as illustrated in Figure 4. The evaluation of these algorithms in predicting diabetes involves assessing performance using metrics such as accuracy, precision, recall, F1-score, and cross-validation. The confusion matrix in table 1, a crucial tool in this evaluation, helps discern classification algorithm performance and elucidating error types, with columns representing actual values and rows depicting expected values.

Table 1: Confusion Matrix for Classification Model Performance

		Actual Value	
		True	**False**
Predicted Values	True	True Positive (TP) Correct Decision	False Positive (FP) Type I Error
	False	False Negative (FN) Type II Error	True Negative (TN) Correct Decision

Accuracy

Accuracy is a fundamental measure of the classification model's effectiveness. It indicates the proportion of correct predictions made out of all predictions. This metric provides a straightforward overview of the model's overall ability to predict outcomes accurately. The formula for accuracy is:
Accuracy = (Number of Correct Predictions) / (Total Number of Predictions) or more specifically:
Accuracy =(TP + TN)/(TP+TN+FP+FN)
Where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

Precision

Precision quantifies the accuracy of positive predictions made by the model. It is the ratio of true positive predictions to the total number of positive predictions (including both true positives and false positives). The formula for precision is:
Precision = (TP)/(TP+FP)
Precision focuses specifically on the model's accuracy in predicting positive cases.

Recall

Also known as sensitivity or the true positive rate, recall measures the model's ability to identify all relevant instances within the dataset. It calculates the proportion of actual positives correctly identified. The formula for recall is:

Recall =(TP)/(TP+FN)

Recall is crucial in situations where missing a positive instance (true positive) is significantly worse than incorrectly labeling negative instances as positive.

F1-Score

The F1-score is a balanced measure that considers both precision and recall. It is particularly useful when you need to balance the importance of precision and recall. The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both values. The formula for the F1-score is:
F1-Score= (Precision X Recall)/ (Precision + Recall)
This metric is especially valuable when dealing with imbalanced classes where one class might be underrepresented in the dataset.
In Figure 5, the workflow for a typical machine learning task is outlined, starting with data exploration and preprocessing. This phase is crucial for ensuring data quality and includes cleaning, normalizing, and possibly transforming the data. (false negatives)

After preprocessing, the data is labeled and undergoes feature engineering, which is a critical step for enhancing model training. The data is then divided, usually in a 70:30 ratio, with the larger portion for training the model and the smaller for testing its predictive power. The diagram showcases an array of machine learning algorithms that are employed, ranging from linear approaches like Linear and Ridge Regression to more complex ones like Decision Trees, Random Forest, Logistic Regression, and Gradient Boosting. Support Vector Machines and K-Nearest Neighbors are also listed, indicating a breadth of methods suited to different kinds of data patterns. Post-training, the model undergoes a thorough evaluation using the test set to measure various performance metrics such as accuracy, precision, and recall. The concluding step in the process is the analysis of results, where the effectiveness of the model is interpreted, providing insights that inform further actions or decisions in practical scenarios.

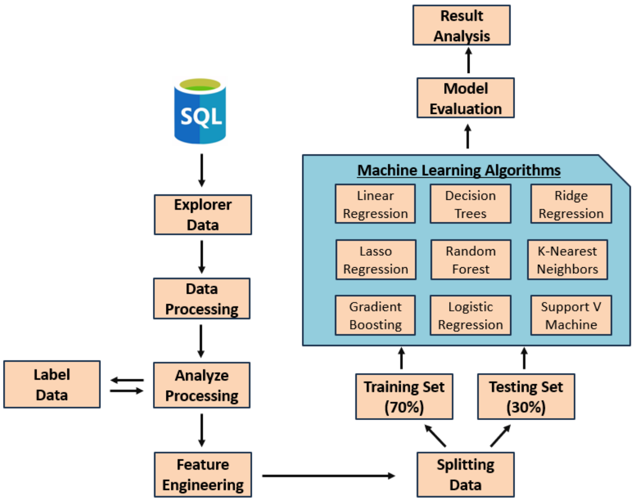

Fig 5: Machine learning workflow for diabetes prediction using a standard training-testing split.

Results and discussion:

The performance of various machine learning models was evaluated using the confusion matrix, which provides insight into the types of errors each model makes. Table 1 outlines the confusion matrix results for each model, showing the number of True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). These values are critical in understanding the balance between different types of errors, such as Type I errors (false positives) and Type II errors

Table 1: The confusion matrix results for each model.

Model	True Positive	False Positive	False Negative	True Negative
Support Vector Machine	471	85	15	32
KNN	475	98	11	19
Decision Tree	465	84	21	33
Random Forest	459	70	27	47
Multilayer Perceptron	467	75	19	42
XGBoost	462	72	24	45
Logistic Regression (LR)	474	78	12	39
LightGBM	462	66	24	51

Table 2: Performance Metrics of Various Machine Learning Models for Diabetes Prediction

Model	Accuracy	Precision	Recall	specificity	F1 Score	F2-Score	Cohen's Kappa	AUC
Support Vector Machine	83%	85%	97%	27%	90%	94%	31%	73%
KNN	82%	83%	98%	16%	90%	94%	19%	76%
Decision Tree	83%	85%	96%	28%	90%	93%	30%	80%
Random Forest	84%	87%	94%	40%	90%	93%	40%	84%
Multilayer Perceptron	84%	86%	96%	36%	91%	94%	39%	79%
XGBoost	84%	87%	95%	38%	91%	93%	40%	84%
Logistic Regression (LR)	85%	86%	98%	33%	91%	95%	39%	77%
LightGBM	85%	88%	95%	44%	91%	93%	45%	85%

The classification performance of the models was systematically evaluated using a suite of key metrics, including Accuracy, Precision, Recall, Specificity, F1 Score, F2-Score, Cohen's Kappa, and the Area Under the ROC Curve (AUC). These metrics collectively offer a detailed and comprehensive assessment of each model's ability to correctly classify instances, providing insights into their strengths and weaknesses across different dimensions of performance**.**

Among the evaluated models, Logistic Regression and LightGBM demonstrated the highest accuracy, both achieving a commendable 85%. This level of accuracy reflects these models' robust overall performance in correctly predicting both positive and negative cases. Precision analysis further revealed that LightGBM achieved the highest precision at 88%, indicating its strong capability to minimize false positives. Random Forest and XGBoost also performed well in this regard, with precision scores of 87%, underscoring their reliability in positive case predictions. When examining recall, which measures sensitivity, KNN and Logistic Regression led with a remarkable 98%, signifying their effectiveness in identifying nearly all positive cases. The SVM model also exhibited strong sensitivity, with a recall of 97%**.**

Specificity, which highlights the models' performance in correctly identifying negative cases, was notably highest in the LightGBM model at 44%, followed by Random Forest and XGBoost with specificity values of 40% and 38%, respectively. This metric is crucial for assessing the models' ability to avoid false positives. Furthermore, F1 Score and F2-Score were highest in the MLP, XGBoost, Logistic Regression, and LightGBM models, each achieving an F1 Score of 91%. The F2-Score, which places greater emphasis on recall, peaked at 95% in the Logistic Regression model, demonstrating its balance between precision and recall. Additionally, LightGBM outperformed other models in terms of Cohen's Kappa, with a score of 45%, indicating the highest agreement between predicted and actual labels. The AUC metric, reflecting the overall ability of the models to discriminate between classes, was also highest for LightGBM at 85%, closely followed by Random Forest and XGBoost at 84% **.**

In summary, while all models showed competitive performance across various metrics, Logistic Regression and LightGBM consistently emerged as the top performers. Both models excelled in accuracy, precision, recall, and AUC, with LightGBM particularly standing out for its superior specificity and Cohen's Kappa scores. These findings underscore the robustness of LightGBM as a model of choice for this classification task, offering a balanced approach to both sensitivity and specificity, while maintaining high overall classification accuracy**.**

Conclusion

This study explored the application of various artificial intelligence (AI) models to predict diabetes outcomes using data from the Diabetes and Endocrinology Clinic in Al-Marj, Libya. Among the models evaluated, Logistic Regression and LightGBM consistently outperformed others, achieving the highest accuracy (85%), precision (88% for LightGBM), and recall (98% for Logistic Regression). LightGBM also demonstrated superior specificity (44%) and Cohen's Kappa (45%), making it a particularly robust model for this classification task.

Future research should consider incorporating additional features, such as patient weight, BMI, and lifestyle factors, to further enhance the predictive power of these models. Expanding the dataset and exploring more advanced AI techniques could also improve generalizability and accuracy across different populations. These enhancements will be crucial for refining AI-driven diabetes prediction and expanding its utility in healthcare settings, particularly in regions with limited access to medical resources.

References

1. Boersma P, Black LI, Ward BW. Prevalence of multiple chronic conditions among US adults, 2018. Preventing Chronic Disease. 2020;17:200130. doi:10.5888/pcd17.200130.
2. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nature Medicine. 2022;28(1):31-38. doi:10.1038/s41591-021-01537-1.
3. Mohsen F, Al-Absi HRH, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. npj Digital Medicine. 2023;6 197. doi:10.1038/s41746-023-00711-7.
4. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, *et al*. Do no harm: a roadmap for responsible machine learning for health care. Nature Medicine. 2019;25(1):1337-1340. doi:10.1038/s41591-019-0548-6.
5. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(24):2402-2410. doi:10.1001/jama.2016.17216.
6. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, *et al*. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Medicine. 2018;15(11) . doi:10.1371/journal.pmed.1002686.
7. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, *et al*. Cardiologist-level arrhythmia detection and classification in ambulatory

electrocardiograms using a deep neural network. Nature Medicine. 2019;25(1):65-69. doi:10.1038/s41591-018-0268-3.
8. Sharma T, Shah M. A comprehensive review of machine learning techniques on diabetes detection. Visual Computing for Industry, Biomedicine, and Art. 2021;4 number 30. doi:10.1186/s42492-021-00075-3.
9. Körner C, Alsdorf M. Mastering Azure Machine Learning (2nd ed.). Packt Publishing; 2022. ISBN: 1803232412.
10. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare Journal. 2019;6(2):94-98. doi:10.7861/fhj.2019-0036.