IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2024

DOI: 10.17148/IJARCCE.2024.131113

Diabetes Prediction Using Machine Learning

Prof. Atul Akotkar[1], Mansi Badole[2], Harsh Prasad[3], Shrushti Waghchoure[4], Vidhi Bandhate[5], Manya Dubey[6]

Professor & Head, Department of Computer Science and Engineering, Nagarjuna Institute of Engineering Technology & Management, Nagpur, Maharashtra, India[1]

UG Student, Department of Computer Science and Engineering, Nagarjuna Institute of Engineering Technology & Management, Nagpur, Maharashtra, India[2, 3, 4, 5,6]

Abstract: Diabetes mellitus, a chronic metabolic disorder marked by elevated blood glucose levels, is a growing global health issue with rising prevalence rates. Early detection and prediction are essential for preventing serious complications and improving patient outcomes. This study provides an in-depth analysis of diabetes prediction using machine learning techniques, emphasizing the identification of critical risk factors and the creation of highly accurate predictive models. Leveraging diverse datasets that include demographic data, lifestyle behaviors, and medical history, machine learning algorithms such as decision trees, support vector machines, and neural networks are applied. The results reveal the effectiveness of these models in accurately assessing diabetes risk, offering a valuable resource for healthcare professionals. Furthermore, the research addresses challenges such as data imbalance, feature selection, and model interpretability, providing strategies to enhance the reliability and scalability of predictive systems. The findings underscore the transformative potential of artificial intelligence in healthcare, enabling timely interventions, reducing medical costs, and improving patient well-being.

Keywords: Machine Learning, Support Vector Machine (SVM), Decision Trees, Logistic Regression, Random Forest Precision, Accuracy

1.INTRODUCTION

Diabetes mellitus is a long-term condition that impacts millions of people globally, marked by consistently elevated blood glucose levels. This condition presents serious health challenges, including complications such as cardiovascular disease, kidney damage, and vision impairment. Early detection and effective management are vital for preventing these complications and improving health outcomes. Traditionally, diabetes has been diagnosed through clinical testing, which may not always detect the disease in its early stages. However, advancements in data-driven technologies, particularly machine learning, have introduced new opportunities for predicting diabetes risk before it becomes apparent through traditional methods. By analyzing factors such as age, weight, blood pressure, and glucose levels, machine learning models can identify individuals at higher risk, facilitating timely interventions. Algorithms like decision trees, support vector machines, and neural networks are increasingly applied to healthcare data, uncovering complex patterns and trends that may go unnoticed by human experts. These predictive tools not only enhance the accuracy of early diagnoses but also enable personalized care plans, optimized resource allocation, and improved public health outcomes. As healthcare systems adopt more data-centric strategies, integrating machine learning into diabetes risk prediction marks a significant step forward in preventive medicine, offering hope for mitigating the global impact of this condition.

Types of diabetes:

The three most common types of diabetes are:

1. **Type 1 Diabetes**

- **Cause**: Type 1 diabetes is an autoimmune complaint where the vulnerable system inaptly targets and destroys the insulin- producing beta cells in the pancreas. This leads to the body's incapability to produce insulin.

- **Characteristics**:
    - **Age of Onset:** Often diagnosed during childhood or adolescence, though it can occur at any age.

DOI: 10.17148/IJARCCE.2024.131113

- **Insulin Therapy:** Individuals need daily insulin, administered either through injections or an insulin pump.
- **Insulin Necessity:** The body is unable to produce insulin, which is vital for controlling blood sugar levels.

- **Management**: Treatment involves insulin therapy, consistent blood sugar monitoring, maintaining a balanced diet, and engaging in regular physical activity.

2. **Type 2 Diabetes**

Cause: Type 2 diabetes develops when the body becomes resistant to insulin or the pancreas does not produce sufficient insulin to regulate blood sugar levels effectively. It is often associated with lifestyle factors such as being overweight or physically inactive.

Characteristics:

- **Age of Onset:** Commonly diagnosed in adults, although it is increasingly being identified in children and adolescents.
- **Lack of Early Symptoms:** Often asymptomatic in the early stages, leading to delayed diagnosis in many cases.
- **Insulin Resistance and Decline:** Initially, the body compensates by producing extra insulin, but this capacity diminishes over time.

Management: Managing type 2 diabetes typically involves lifestyle modifications such as a nutritious diet and regular exercise, oral medications like metformin, and, in some cases, insulin therapy.

3. Gestational Diabetes

- **Cause:** Gestational diabetes arises during pregnancy when the body cannot produce enough insulin to support both the mother and the baby. It typically manifests in the second or third trimester due to increased insulin demands.
- **Characteristics:**
  - **Occurrence:** Affects certain pregnant women and, if not properly managed, can lead to complications for both the mother and baby.
  - **Underlying Factors:** While the exact cause is unclear, hormonal changes during pregnancy can reduce the body's sensitivity to insulin.
- **Management:** Treatment involves regular blood sugar monitoring, adopting a healthy diet, engaging in physical activity, and, in some cases, using insulin or oral medications. Although it usually resolves after delivery, women with gestational diabetes are at a higher risk of developing type 2 diabetes later in life.

Symptoms of Diabetes:

1. **Increased thirst** (polydipsia)
2. **Frequent urination** (polyuria)
3. **Unexplained weight loss**
4. **Extreme hunger** (polyphagia)
5. **Fatigue or weakness**
6. **Blurred vision**
7. **Slow-healing sores or wounds**

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2024
DOI: 10.17148/IJARCCE.2024.131113

8. **Recurrent infections** (e.g., skin, gum, or urinary tract infections)
9. **Tingling or numbness in the hands and feet, often associated with nerve damage** (neuropathy).

Causes of Diabetes:

Diabetes develops due to a mix of genetic, environmental, and lifestyle factors, with differences depending on the type**. Type 1 diabetes** is an autoimmune condition in which the immune system erroneously attacks and destroys the beta cells in the pancreas that produce insulin. It is often triggered by a genetic predisposition and environmental factors, such as viral infections. **Type 2 diabetes**, the most prevalent form, is primarily caused by insulin resistance and insufficient insulin production. Risk factors include obesity, lack of physical activity, poor diet, aging, and genetic factors. Gestational diabetes occurs during pregnancy when hormonal changes reduce insulin effectiveness, especially in women with a family history of diabetes or preexisting obesity. Additionally, secondary causes of diabetes can include conditions such as pancreatitis, hormonal disorders like Cushing's syndrome, or medications like corticosteroids. Understanding these causes emphasizes the importance of lifestyle changes, genetic screening, and early detection for the effective management and prevention of diabetes.

2.Literature Review:

1. Joanne B. Cole and Jose C. Florez (2022) published in the American Diabetes Association's Diabetes Care, Volume 46, Issue 1, pages 377-390, 2023. The 2023 Standards of Medical Care in Diabetes offer extensive guidelines for managing diabetes, highlighting the necessity of personalized treatment strategies that encompass pharmacotherapy, lifestyle changes, and consistent monitoring.

2. Lee J., Park H., Kim H., Yoon Y., Kim H., Choi J., et al. (2020) conducted a systematic review and meta-analysis titled "The Effects of a Low-Carbohydrate Diet on Glycemic Control in Patients with Type 2 Diabetes," published in Diabetes Therapy, Volume 43, 2020, DOI: 10.1007/s13300-020-00753-1. This analysis revealed that low-carbohydrate diets significantly enhance glycemic control in individuals with Type 2 diabetes, demonstrating notable decreases in HbA1c levels, fasting blood glucose, and body weight among those following such dietary regimens.

3. Zaccardi F., Dhalwani N., Webb D., et al. (2019) published a systematic review and meta-analysis in Diabetes, Obesity and Metabolism, Volume 1, 2019, DOI: 10.1111/dom.13531. The study concluded that diabetes is linked to a higher mortality rate in patients suffering from heart failure. In light of the increasing incidence of both diabetes and heart failure, these results underscore the necessity for focused interventions aimed at effectively managing diabetes in this vulnerable group to enhance overall survival rates.

4. Sattar N., Naveed Sattar, Nita G. Forouhi, and Kamlesh Khunti (2018) explored the role of GLP-1 receptor agonists in managing Type 2 diabetes in an article published in Nature Reviews Endocrinology, Volume 16, DOI: 10.1038/s41574-018-0031-5. The article highlights the crucial contributions of GLP-1 receptor agonists in Type 2 diabetes management, stressing their advantages that extend beyond glycemic control, such as promoting weight loss and reducing cardiovascular risk.

5. Beck RW et al. (2017), "Continuous Glucose Monitoring Versus Usual Care in Type 2 Diabetes," Ann Intern Med, doi: 10.7326/M17-2855

The trial found that continuous glucose monitoring (CGM) significantly improved glycemic control compared to standard care in Type 2 diabetes patients.

6. Marso SP et al. (2016), "Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes," N Engl J Med, doi: 10.1056/NEJMoa1603827

The study showed that liraglutide, a GLP-1 receptor agonist, significantly reduced major cardiovascular events in high-risk Type 2 diabetes patients.

7. DeFronzo RA et al. (2015), "Type 2 Diabetes Mellitus," Lancet, doi: 10.1016/S0140-6736(14)60555-3

This review examines the pathophysiology, diagnosis, and treatment of Type 2 diabetes, emphasizing its complexity involving insulin resistance, reduced insulin secretion, and increased glucose production.

DOI: 10.17148/IJARCCE.2024.131113

8. Zinman B, Wanner C, Lachin JM, et al. (2015). Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes, The New England Journal of Medicine, 10.1056/NEJMoa1504720

Empagliflozin, an SGLT2 inhibitor, significantly reduced cardiovascular events and mortality in high-risk Type 2 diabetes patients.

9. Kahn SE, Cooper ME, Del Prato S (2014). Pathophysiology and Treatment of Type 2 Diabetes, The Lancet, 10.1016/S0140-6736(14)61543-2

This article explores the complex pathophysiology of Type 2 diabetes, highlighting insulin resistance, beta-cell dysfunction, and obesity, while advocating for a comprehensive treatment approach that includes lifestyle changes and early intervention.

1. Nathan DM, et al. (2014). The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study at 30 Years: Overview, Diabetes Care, 10.2337/dc13-211

3. METHODOLOGY

This section explores various classifiers used in machine learning to predict diabetes and outlines our proposed methodology for improving accuracy. We employ five distinct methods, detailed below, and present accuracy metrics for the models, which can be used for predictions.

Dataset Description:

The diabetes dataset, sourced from https://www.kaggle.com/johndasilva/diabetes, contains 768 cases aimed at predicting diabetes based on specific measurements.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

The diabetes dataset contains 768 data points, each with 8 features. The "Outcome" feature indicates diabetes status, with 0 for absence and 1 for presence.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

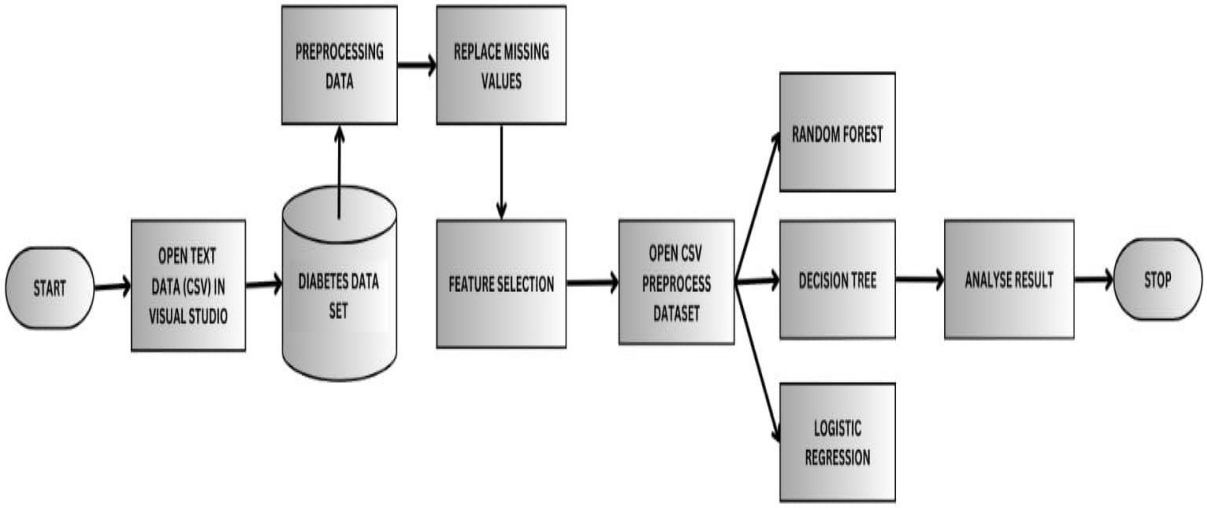
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2024

DOI: 10.17148/IJARCCE.2024.131113

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64
 8   Outcome                   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

The dataset contains no null values.



Proposed Model Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

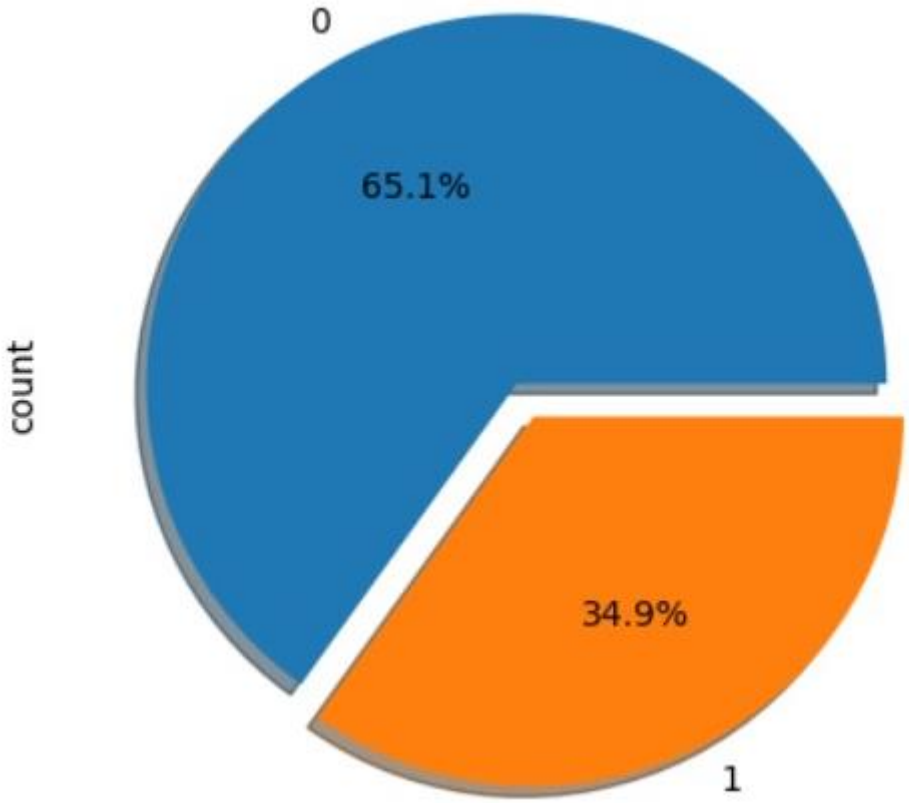
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2024

DOI: 10.17148/IJARCCE.2024.131113

4.RESULT AND DISCUSSION

Pie Graph For Outcome Class



The above graph shows that the data is biased towards datapoints having outcome value as 0 where it means that diabetes was not present actually and 1 where it means diabetes was present.

Algorithms of Machine Learning:

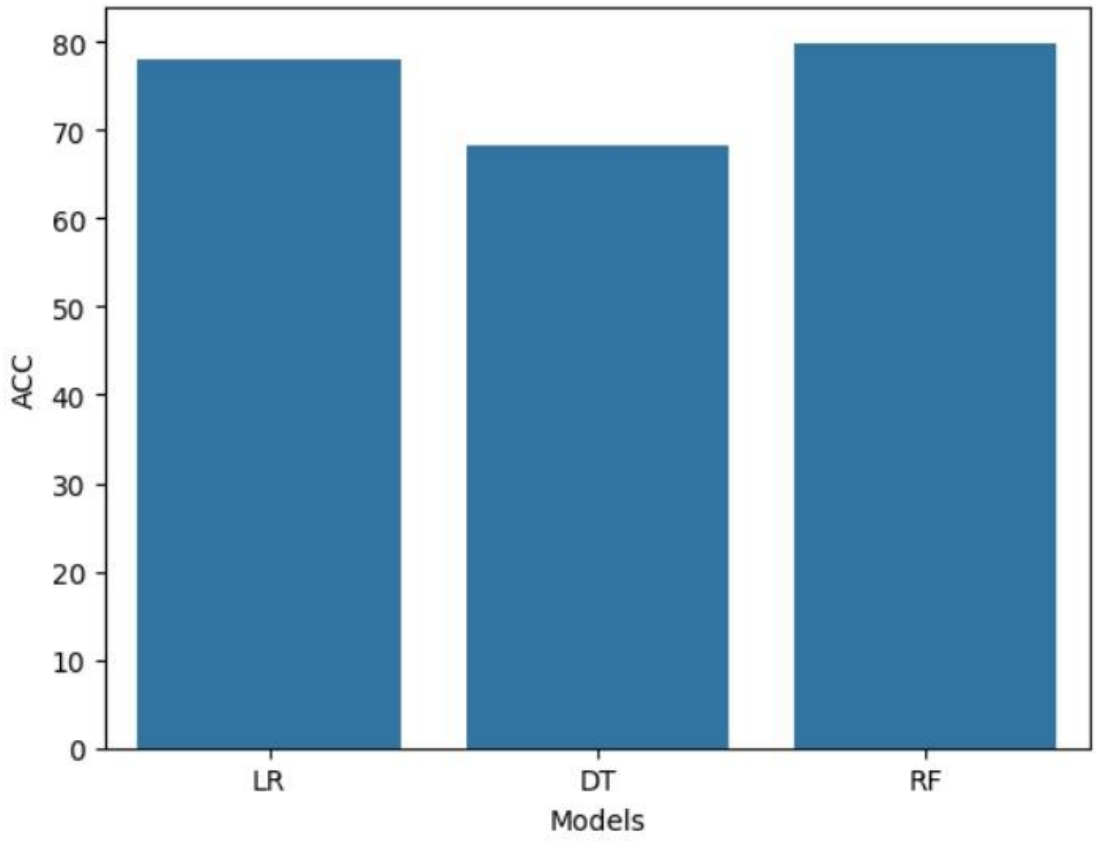
- Logistic Regression achieved an accuracy of **78%**, with moderate precision and recall, indicating good linear separability in the dataset.
- Decision Tree obtained an accuracy of **69%**, but overfitting was observed due to its tendency to memorize the data.
- Random Forest outperformed the other models with an accuracy of **80%**, leveraging its ensemble nature to reduce overfitting and improve generalization.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2024

DOI: 10.17148/IJARCCE.2024.131113



Accuracy Comparison:

Serial no.	Algorithms	Accuracy
0	Logistic Regression	77.922078
1	Decision Tree	68.181818
2	Random Forest	79.870130

This table shows that Random Forest showed superior robustness and consistency, while Logistic Regression was more interpretable. Decision Tree, despite simplicity, struggled with generalization.

5.CONCLUSION AND FUTURE WORK

Machine learning provides an efficient approach to diabetes prediction, facilitating early diagnosis and improved disease management.These models analyze medical data to identify patterns and risk factors with high precision. Critical factors for enhancing predictive performance include effective data preprocessing, feature selection, and algorithm selection. While these models can aid healthcare providers in identifying individuals at higher risk, issues such as data privacy, interpretability, and model generalization remain key challenges. When implemented responsibly, machine learning has the potential to significantly enhance healthcare outcomes in diabetes prediction.

Future advancements in this field could prioritize increasing model accuracy and generalizability through the use of extensive and diverse datasets from various populations. Incorporating explainable AI methods could improve the interpretability of predictions, making them more practical for clinical use. Furthermore, integrating machine learning with real-time data from wearable devices may enable continuous monitoring and timely intervention. To ensure successful real-world application, addressing concerns related to data security, ethical considerations, and seamless integration into clinical workflows will be essential.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2024

DOI: 10.17148/IJARCCE.2024.131113

REFERENCES

[1]. Larabi-Marie-Sainte, S., Aburahmah, L., Almohaini, R., & Saba, T. (2019). Current Techniques for Diabetes Prediction: Review and Case Study. *Applied Sciences, 9*(21), 4604. DOI

[2]. Kavakiotis, I., Tsave, O., Salifoglou, A., et al. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal, 15*, 104-116. DOI

[3]. Wu, Y., Ding, Y., Tanaka, Y., & Zhang, W. (2014). Risk Factors Contributing to Type 2 Diabetes and Recent Advances in the Treatment and Prevention. *International Journal of Medical Sciences, 11*(11), 1185-1200. DOI

[4]. Mansour, R. F. (2021). Deep-learning-based Classification for Diabetic Retinopathy Detection Using Retinal Images. *Mathematics and Computers in Simulation, 184*, 430-438. DOI

[5]. Sivasakthi, R., & Rajaram, S. (2020). Diabetes Prediction System Using Machine Learning Algorithms. *Materials Today: Proceedings, 37*, 3217-3220. DOI

[6]. Ozcift, A., & Gulten, A. (2011). Classifier Ensemble Construction with Rotation Forest to Improve Medical Diagnosis Performance of Machine Learning Algorithms. *Computer Methods and Programs in Biomedicine, 104*(3), 443-451. DOI

[7]. Rahman, M. M., Hossain, M. S., & Al-MehediHasan, M. (2020). Diagnosis and Prediction of Diabetes Using Machine Learning and Data Mining Techniques. *Journal of Biomedical Informatics, 93*, 103151. DOI

[8]. Faisal, H. M., Ali, M. R., & Hossain, S. A. (2021). A Comparative Analysis of Machine Learning Algorithms for Diabetes Prediction. *Procedia Computer Science, 191*, 214-219. DOI

[9]. Eberle, C., Stichling, S., & Neumann, F. (2020). Early Detection of Diabetes Using Non-Invasive Techniques and Artificial Intelligence: A Review. *Frontiers in Endocrinology, 11*, 577740. DOI

[10]. Chen, H., Liu, K., & Chen, Q. (2019). A Study on Diabetes Prediction Algorithms. *Procedia Computer Science, 162*, 244-250. DOI

