**A red square with a white border

Description automatically generated with medium confidence**

|  |  |
| --- | --- |
| Student Name and Number: | Yeo Xiu Juan (34678307)  Roldan Aquino (34641153)  Kanagaraj Prakash (35113797) |
|  |  |
| Unit Code: | ICT 619-TJA 25 |
| Unit name: | Artificial Intelligence |
| Date: | 15th Feb 2025 |
| Assignment name: | Dia Sense AI: Transforming Diabetes Care |
| Tutor | Dr.Poh Kok Loo |

**ELECTRONIC ASSIGNMENT COVERSHEET**

Table of Contents

[Abstract 1](#_Toc194153759)

[Introduction 2](#_Toc194153760)

[Background 3](#_Toc194153761)

[Significance of the problem 5](#_Toc194153762)

[Uniqueness & Innovation 5](#_Toc194153763)

[Data Collection 6](#_Toc194153764)

[Data Pre-Processing 6](#_Toc194153765)

[AI Technique: 7](#_Toc194153766)

[1.Multi-Layer Perceptron (MLP) 7](#_Toc194153767)

[2.MLP + Logistic Regression (MLP+LR) 9](#_Toc194153768)

[3. MLP + Random Forest (MLP+RF) 10](#_Toc194153769)

[Comparing Decay Resistance Across Models 12](#_Toc194153770)

[Evaluation Method 13](#_Toc194153771)

[Model Optimization 15](#_Toc194153772)

[MLP + LR Results 17](#_Toc194153773)

[Learning curve 19](#_Toc194153774)

[Comparative results with theoretical optimum 20](#_Toc194153775)

[Dia Sense Integration and Maintenance Plan 21](#_Toc194153776)

[Quality Assurance & System Validation 21](#_Toc194153777)

[Key Integration tech Slack Stakeholders: 22](#_Toc194153778)

[Functionality Elements 22](#_Toc194153779)

[Non-Functionality Elements 22](#_Toc194153780)

[Testing & Deployment 23](#_Toc194153781)

[Post-Release: Proactive Maintenance 23](#_Toc194153782)

[Discussion 24](#_Toc194153783)

[Key Findings 24](#_Toc194153784)

[Implications 25](#_Toc194153785)

[Comparison with Previous Studies 26](#_Toc194153786)

[Expected Outcome: 26](#_Toc194153787)

[Unexpected Outcome: 26](#_Toc194153788)

[Key Lesson Learnt 26](#_Toc194153789)

[What went well, and what went poorly? 27](#_Toc194153790)

[Limitation: Statistical Methods vs. Expert Knowledge in Diabetes Prediction 28](#_Toc194153791)

[References 31](#_Toc194153792)

[Appendix 33](#_Toc194153793)

[*Appendix 1: PIMA Indian Diabetes Dataset* 33](#_Toc194153794)

[*Appendix 2: Data Pre-Processing and Post-Processing for Voting Classifier (MLP+LR)* 35](#_Toc194153795)

[*Appendix 3 : List of the three models code from Hyperparameter Tuning* 39](#_Toc194153796)

[*Appendix 4 : Final model MLP+LR* 43](#_Toc194153797)

[*Appendix 5 : Dia sense UI Applications* 45](#_Toc194153798)

Dia sense AI: Transforming Diabetes Care

# Abstract

DiaSense is an AI-powered system designed for the early diagnosis and proactive management of Type 2 Diabetes (T2D), with a focus on women. This study investigates an AI-driven approach to diabetes prediction using the Pima Indian dataset, employing a hybrid model that combines Multi-Layer Perceptron (MLP) and Logistic Regression (LR). By integrating advanced machine learning techniques with statistical evaluation methods, including Youden’s J statistic, the model aims to enhance early detection and risk assessment of diabetes.

The study identifies key factors influencing diabetes progression, such as genetic predisposition, insulin resistance, and glucose regulation. The MLP+LR model achieved an AUC-ROC of 0.84, along with competitive G-Means and AUC-PRC scores of 75%. The model's interpretability was further enhanced through SHAP global explanations, identifying top predictors such as pregnancies, glucose levels, and blood pressure.

Despite its success, the model faced challenges, including class imbalance, which affected its generalization performance. A comparative analysis with state-of-the-art models revealed a performance gap of approximately 15%, suggesting the need for further optimization to improve accuracy and robustness.

# Introduction

Type 2 Diabetes (T2D) is a chronic metabolic disorder affecting millions worldwide, with women facing unique health risks, including increased cardiovascular complications, gestational diabetes, and pregnancy-related issues. Traditional diagnostic methods, such as fasting plasma glucose (FPG) and hemoglobin A1c (HbA1c) tests, are often costly, time-consuming, and dependent on clinical visits. These limitations contribute to delayed interventions and suboptimal disease management(Health., 2024).

T2D is characterized by insulin resistance, where the body’s cells become less responsive to insulin, coupled with a gradual decline in insulin production. Unlike Type 1 Diabetes, which results from the autoimmune destruction of insulin-producing beta cells, T2D arises from a combination of genetic, environmental, and lifestyle factors. Managing the disease requires continuous monitoring and lifestyle modifications, including dietary changes, physical activity, and medication adherence. However, the complexity of T2D, its heterogeneous progression, and the presence of comorbidities make diabetes management particularly challenging.

Diabetes is a disease that directly impacts insulin regulation, and its effects can be detrimental when not properly controlled. While there is no cure for diabetes as of now, it can be managed through preventive measures, including medication, lifestyle adjustments, and regular monitoring. If left untreated or poorly managed, diabetes can lead to severe complications, such as cardiovascular disease, kidney failure, blindness, and lower limb amputations. The global prevalence of diabetes has been rising rapidly, with nearly 200 million people affected, and more than half of these cases occurring in women. Women, especially between the ages of 25 and 44, are at a significantly higher risk of developing T2D, making early diagnosis and intervention even more critical.

Several studies include (Muhammad Mazhar Bukhari, 2021) are currently underway to explore effective ways to predict and diagnose diabetes at an early stage, with the aim of improving outcomes and preventing long-term complications. Among the promising techniques, Artificial Neural Networks (ANN) have emerged as an effective tool for diabetes prediction. By analysing key attributes and using advanced diagnostic models, ANN can aid in early identification, allowing individuals to seek timely medical advice and preventive care.

This paper focuses on leveraging **MLP with ensemble learning** to enhance the prediction and diagnosis of diabetes, offering a powerful tool for early intervention and improved disease management.

In light of the growing prevalence of diabetes and its potential to lead to serious complications, it is essential to continue advancing predictive models and diagnostic tools. By doing so, we can help mitigate the impact of this disease on both individuals and healthcare systems.

# Background

AI-driven diabetes prediction has become a significant area of research due to its potential to improve early detection and management of diabetes. The complexity of diabetes, influenced by factors such as lifestyle, genetics, and medical history, makes it a challenging condition to predict and manage. Prior Machine learning (ML) models have been explored to aid in this process, including 1) Traditional models like logistic regression, support vector machines (SVMs), and decision trees, as well as more advanced methods like 2) Deep learning models and 3) Fuzzy Logic.

**1.Traditional Machine Learning Models:**

1. ***Logistic Regression:***

Logistic regression is a linear model commonly used for binary classification tasks, such as predicting diabetes risk based on features like age, BMI, family history, and blood sugar levels. While logistic regression is easy to interpret and implement, it struggles to capture non-linear relationships in the data, which are often crucial in diabetes prediction. As a result, it may underperform when interactions between features play a significant role(Zeng, 2024).

1. ***Support Vector Machines (SVMs):***

SVMs are designed to find the optimal hyperplane that separates the data into distinct classes. SVMs are particularly effective in high-dimensional spaces and can handle non-linear decision boundaries through the use of kernel functions. However, they can be sensitive to the choice of hyperparameters and may not scale well to very large datasets, posing challenges in practical applications(Asha, 2024).

1. ***Decision Trees:***

Decision trees are intuitive and easy to interpret. They split the data into subsets based on feature values, making them valuable for understanding the decision-making process. However, decision trees are prone to overfitting, especially when the data is noisy, and they may not generalize well to unseen data. Techniques like pruning and ensemble methods (e.g., Random Forest) are often used to mitigate these issues(Zeng, 2024).

**2.Deep Learning Approaches:**

In recent years, deep learning models have gained popularity for prediction due to their ability to capture highly complex patterns in data. Models like Neural Network and convolutional neural networks (CNNs) have shown success, but they come with certain limitations.

1. ***Convolutional Neural Networks (CNNs):***

While CNNs are typically associated with image processing, they have also been used in diabetes prediction, especially when working with structured health data such as time-series sensor readings or medical imaging. CNNs can automatically learn relevant features and patterns from raw data, allowing for more sophisticated decision-making. However, the main drawbacks of CNNs are their requirement for large datasets and the substantial computational power needed for training(Al Sadi & Balachandran, 2023).

1. ***Deep Neural Network (NN)***

Through deep learning models, neural networks (NN) can process large datasets to detect early warning signs of diabetes that might not be immediately visible through traditional diagnostic methods. By analyzing real-time data from wearables and health devices, NNs can provide insights into how factors like diet, exercise, and medication affect glucose levels, offering actionable feedback for better disease management. However, despite promising results, there is still limited long-term evidence supporting the sustained effectiveness of neural networks in diabetes management compared to more traditional, clinically validated methods(Muhammad Mazhar Bukhari, 2021).

**3.Expert System: Fuzzy Logic with Mamdani Inference for Diabetes Management**

In diabetes management, fuzzy logic systems are particularly effective in dealing with data that is imprecise, vague, or uncertain — a common challenge in medical diagnostics and healthcare management. Fuzzy logic enables the system to model human-like reasoning and decision-making processes, making it well-suited to handle the inherent uncertainty in clinical data, such as fluctuating blood glucose levels, symptoms, or patient behaviors.

Mamdani Inference, one of the most widely used methods for fuzzy inference, plays a crucial role in translating fuzzy input variables (such as blood sugar levels) into actionable output decisions, such as determining whether to increase, decrease, or maintain insulin dosage. This method uses a set of rules that mimic human reasoning, making it intuitive for healthcare providers to understand and interpret the results.

When combined with Harris Hawks Optimization (HHO), fuzzy logic systems can significantly enhance their performance. HHO optimizes the fuzzy system’s parameters, improving its ability to predict glucose levels accurately and adjust treatment plans based on dynamic patient conditions. This dynamic optimization process ensures that the fuzzy system can adapt in real-time, providing more personalized and effective care for individuals with diabetes(Asghari Varzaneh, 2023)

However, despite the advantages of fuzzy logic in diabetes management, designing a robust fuzzy inference system (FIS) can be time-consuming and requires expertise. The accuracy of the system depends heavily on the quality of the rule base and the careful tuning of the membership functions. This calibration process can be challenging, particularly when scaling the system for large patient populations or complex real-world scenarios.

The integration of AI and advanced optimization methods holds great promise for revolutionizing diabetes care, improving early detection, and supporting real-time management. Each technique offers unique strengths but also comes with its limitations. Traditional machine learning models like logistic regression, SVMs, and decision trees provide interpretability but may fail to capture complex, non-linear relationships. Deep learning models like CNNs excel at handling large datasets and complex patterns but require significant computational power and large volumes of labeled data.

Fuzzy logic systems, particularly with Mamdani Inference and Harris Hawks Optimization, offer significant potential for personalized and adaptive diabetes management, especially in dealing with uncertain and fluctuating clinical data. However, the design and tuning of such systems require careful consideration and expertise, particularly when scaling for large, diverse patient populations.

# Significance of the problem

Existing studies often fail to address the specific health risks faced by women, particularly those with maternal health concerns. In the context of the Pima Indian population(Pearson, 2015), there is a notably high prevalence of gestational diabetes mellitus (GDM)—a condition that develops during pregnancy in women without prior diabetes diagnoses. Pregnant women with GDM are at increased risk of elevated blood sugar levels, which can have long-term health implications for both mother and child.

Research indicates that the prevalence of GDM in Pima women may reflect similar trends in other ethnic groups across the United States. This suggests that diabetes in pregnancy, initially identified among the Pima Indians, could be an emerging public health concern on a broader scale.

Additionally, obesity is prevalent in the Pima Indian population, with younger individuals exhibiting higher average body mass index (BMI) values. Over the past 25 years, modest increases in BMI have been observed across different age and sex groups, further compounding the risk of diabetes(Bashir I. A, 2019; Nelson et al., 2021).

# Uniqueness & Innovation

DiaSense is a groundbreaking AI-powered system designed to transform diabetes risk assessment, particularly for high-risk populations such as the Pima Indian community. Unlike traditional diagnostic tools that rely on static, periodic clinical assessments, DiaSense leverages machine learning to provide real-time, predictive insights tailored to each individual.

Unlike conventional methods that primarily focus on biochemical markers, DiaSense analyzes a holistic combination of physiological, lifestyle, and genetic factors to assess diabetes risk with greater accuracy. Its AI-driven predictive analytics go beyond simple screenings, offering adaptive risk profiling that evolves with an individual’s changing health data.

What truly sets DiaSense apart is its dynamic feedback mechanism, which continuously refines risk predictions based on new information. This makes it more responsive and accurate than conventional screening methods, which often fail to account for evolving risk factors over time. By integrating cutting-edge AI technology with personalized health monitoring, DiaSense is paving the way for early intervention and proactive diabetes prevention, ultimately improving health outcomes for at-risk populations**.**

Key **innovations** of DiaSense include:

* AI-Powered Predictive Modeling – Uses advanced machine learning algorithms to analyze health data and predict diabetes risk with higher accuracy.
* Real-Time Monitoring & Integration – Seamlessly integrates with wearable devices and mobile applications, enabling continuous glucose monitoring and proactive intervention.
* Accessibility & Remote Assessment – Designed with a mobile-friendly interface, empowering individuals—especially those in remote or underserved areas—to self-assess their risk without requiring frequent clinical visits.
* Explainable AI (XAI) for Transparency – Unlike traditional black-box AI models, DiaSense provides clear, interpretable insights, ensuring that both patients and healthcare providers understand the reasoning behind its risk assessments.
* Personalized, Data-Driven Insights – Combines machine learning with expert-driven knowledge to deliver customized recommendations based on an individual’s lifestyle, medical history, and real-time data.

By merging cutting-edge AI with real-world clinical expertise, DiaSense is more than just a predictive tool—it represents a paradigm shift in proactive, accessible, and personalized diabetes management. With its ability to provide real-time, adaptive risk assessments, DiaSense sets a new standard for AI-driven healthcare solutions, offering unprecedented support for individuals at risk of diabetes.

# Data Collection

The dataset used in this study was obtained from Kaggle. Initial data exploration included Spearman’s rank correlation analysis and multivariate correlation analysis. Further details can be found in: *Appendix 1: PIMA Indian Diabetes Da**taset*.

# Data Pre-Processing

The pre-processed data must demonstrate the anticipated output, ensuring that the model is prepared for effective training and evaluation. The following steps were applied during the pre-processing phase:

1. **Min-Max Scaling**: All numerical features were scaled using the Min-Max scaling technique to normalize the data within a specific range, ensuring that each feature contributes equally to the model's performance.
2. **Feature Engineering through PCA**: Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset and extract the most significant features, enhancing the model’s ability to capture important patterns while reducing computational complexity.
3. **Train-Validation-Test Split**: The dataset was split into training, validation, and test sets to ensure proper model evaluation. Stratification was applied to the "Diabetes" variable to account for class imbalance, ensuring that each subset of the data (train, validation, and test) maintains a similar distribution of the dependent variable.

Further details can be found in : *Appendix 2: Data Pre-Processing and Post-Processing for Voting Classifier (MLP+LR)*

In the following sections, we will explore the selected AI techniques, including **MLP, MLP+LR** and **MLP+RF**, to understand their potential contributions to diabetes prediction and management.

# AI Technique:

This section explores three deep learning techniques aimed at improving predictive accuracy in complex tasks. Specifically, the effectiveness of Multi-Layer Perceptron’s (MLPs) combined with ensemble learning methods, such as soft voting, will be investigated to enhance model performance. By combining MLP with Logistic Regression (MLP+LR) and MLP with Random Forest (MLP+RF), the goal is to leverage the strengths of both individual models and ensemble strategies to achieve more robust and accurate predictions.

1. Multi-Layer Perceptron (MLP)

**Multilayer Perceptron (MLP)** is a widely used algorithm in machine learning, particularly within the domain of artificial neural networks, known for its ability to automatically identify complex patterns in data. MLPs are particularly effective for handling non-linear relationships between input features and output predictions.

**Architecture and Mechanism**

The MLP operates on a **feedforward** and **backpropagation** mechanism:

* **Feedforward phase**: The network is structured in layers. Each neuron in a given layer is densely connected to the neurons in the previous layer, enabling the input data to be passed through successive layers. Each layer performs transformations based on learned weights and biases, gradually refining the data's features.
* **Backpropagation phase**: After the network generates an output, the error is calculated by comparing the predicted output with the actual values, typically using a loss function such as mean squared error for regression or cross-entropy for classification. The error is then propagated backward through the network, adjusting the weights and biases of each neuron in an attempt to minimize the error. The optimization process is carried out using gradient descent (or its variants), where the gradients of the loss function with respect to the parameters are used to update the weights iteratively, helping the model converge toward a more accurate prediction.

A diagram of a network

Description automatically generated

Figure 1: MLP

During the **training phase**, the weights and biases are adjusted according to the gradients of the loss function with respect to these parameters. Once the network has been sufficiently trained, it can be utilized to make predictions on unseen test data.

Previous studies by (Zhang et al., 2024) conducted an experiment using various diabetes datasets, where the Backpropagation Neural Network (BPNN) achieved an impressive accuracy of 89.81% through 5-fold cross-validation.

**Key Strengths of MLP:**

1. **Nonlinear Mapping**: MLPs are capable of approximating nonlinear functions, which enables them to model complex, intricate relationships between input features and target variables. This characteristic makes MLPs particularly well-suited for tasks where linear models would be insufficient.
2. **Feature Learning**: Through the use of multiple hidden layers, MLPs are able to automatically learn hierarchical and abstract representations of features from raw input data. This ability allows the model to extract and construct meaningful patterns from unprocessed data, facilitating improved performance in a variety of predictive tasks.
3. **Scalability**: With recent advancements in computational hardware, such as GPUs, and the development of more efficient optimization algorithms, MLPs are increasingly able to scale to accommodate large and high-dimensional datasets. This scalability allows MLPs to effectively handle more complex model architectures and larger volumes of data.

**Key Limitations Of MLP:**

1. **Overfitting:** MLPs are prone to overfitting, without proper regularization (e.g., dropout or early stopping), the model may memorize the training data, leading to poor generalization to unseen data.
2. **Gradient Vanishing/Exploding**: Deep MLPs may suffer from vanishing or exploding gradients during backpropagtion.
3. **Hyperparameter Sensitivity**: The performance of MLPs is highly dependent on the selection of hyperparameters, such as the learning rate, batch size, and the network architecture (e.g., the number of hidden layers and neurons per layer).
4. **Computational Intensity**: Training MLPs can be resource-heavy, requiring significant computational power, especially with large datasets.
5. **Black-box Nature**: MLPs lack interpretability and transparency, making it difficult to explain model decisions, which could be a drawback in high-stakes environments like healthcare.

2.MLP + Logistic Regression (MLP+LR)

As discussed in previous sections, a notable limitation of Multi-Layer Perceptron’s (MLP) is its "black box" nature, which results in limited interpretability of the model's decision-making process. To address this challenge, we propose leveraging an ensemble learning method, specifically the Voting Classifier, to combine the strengths of MLP with the simpler yet effective Logistic Regression (LR) model.

Logistic Regression (LR) is widely used in binary classification tasks and is known for its simplicity and interpretability. The LR model operates by classifying a dependent variable into two categories (e.g., 0 or 1) using a linear decision boundary. It employs a sigmoid activation function, which computes the weighted sum of input features and a bias term, producing a probability output between 0 and 1. LR’s parameters (weights and bias) are optimized using gradient descent.

A Voting Classifier combines multiple models by aggregating their predictions, typically via a majority voting scheme for classification tasks. This ensemble method allows us to leverage both the non-linear modeling ability of MLP and the linear classification power of LR. The result is a more robust, flexible model that can handle complex data while maintaining some level of transparency through the Logistic Regression component.

Several researchers, including (Tunç, 2012) have demonstrated the success of hybrid models that combine MLP with LR. For example, in the context of lung cancer prediction, the MLP+LR ensemble achieved an Area Under the Curve (AUC) of 92.31%, with sensitivity of 80% and specificity of 96.45%.

The combination of Multi-Layer Perceptrons (MLP) and Logistic Regression (LR) in an ensemble model presents several notable strengths, which enhance the predictive capabilities of the model. However, like any hybrid approach, it also comes with certain limitations that must be carefully managed.

**Key Strengths:**

1. **Improved accuracy on complex data**

The combination of MLP and LR allows for enhanced accuracy on datasets with intricate, non-linear relationships. MLP excels at learning complex, non-linear patterns, while LR provides a simple and efficient linear decision boundary. This hybrid model can capture a wider range of patterns, improving predictive performance compared to using either model independently.

1. **Diverse and versatile for various domain**

(Pietukhov et al., 2023) demonstrated versatility of MLP+LR across different fields, including healthcare, finance, and supply chain. By combining MLP’s ability to learn intricate patterns with LR's interpretability, this approach can be effectively applied in a variety of domains requiring both predictive power and some level of model explainability.

**Key Limitations:**

1. **Increase model complexity**

The combination of MLP with LR introduces additional complexity to the model. While MLP is already computationally intensive, adding a logistic regression layer further complicates the system. This may lead to longer training times and more challenging deployment, especially for large datasets.

1. **Overfitting risk**

Similar to standalone MLP models, the hybrid MLP + LR model is vulnerable to overfitting, particularly when the model’s complexity exceeds what the data can support. For instance, an overly complex MLP with many hidden layers might memorize the training data rather than learning generalizable patterns. This overfitting can lead to poor performance on unseen data. Regularization techniques, such as dropout, early stopping, and L1/L2 penalties, are essential to mitigate overfitting. Moreover, LR’s regularization strength is often influenced by parameters like the C parameter, which requires careful tuning to maintain the right balance between bias and variance in the model.

1. **Tuning complexity**

The tuning process for a combined MLP+LR model can be significantly more challenging than tuning a single model. Finding the optimal hyperparameters for both the MLP and LR components requires substantial experimentation. This can be computationally expensive, particularly when dealing with large datasets or deep network architectures. The process may involve tuning multiple hyperparameters, such as learning rates, number of layers, and the regularization strength of both models. This can be time-consuming and might require significant computational resources, especially in real-world applications.

3. MLP + Random Forest (MLP+RF)

Multi-Layer Perceptron’s (MLPs), a class of deep learning models, are highly effective for handling high-dimensional data and capturing non-linear relationships. However, these models often suffer from a lack of interpretability and transparency, which can be significant barriers to their adoption in clinical and real-world settings where understanding the decision-making process is critical. The opacity of MLPs, though powerful in predictive tasks, complicates their use in domains requiring clear explanations of model behavior.

In contrast, Random Forest (RF), a tree-based method, excels at handling both categorical and continuous data, and is well-suited for capturing non-linear relationships between variables. One of the key advantages of RF models is their interpretability. The decision trees within the random forest ensemble allow for the identification of the most influential features, thus providing a more transparent view of the decision-making process. Moreover, RF is relatively less prone to overfitting, making it a robust model when applied to diverse datasets.

Several studies have demonstrated the effectiveness of ensemble approaches in machine learning. For example, (Choudhury, 2023) employed a combination of MLP, RF, and Deep Belief Networks (DBN) to classify breast cancer, achieving an impressive 96.5% accuracy. Similarly,(Rajawat et al., 2018) explored the use of an ensemble comprising MLP, Extreme Gradient Boosting (XGBoost), and RF for Pima Indian diabetes classification, achieving an accuracy of 87.3%. These results underscore the potential of ensemble methods in improving model performance across various domains.

To further capitalize on the strengths of both MLP and RF, an ensemble approach can be employed, such as the Voting Classifier, which aggregates predictions from both models. This hybrid approach can provide several benefits:

* **Handling Diverse Data Types:** By combining MLP and RF, the ensemble model leverages the ability of RF to handle various data types (categorical and continuous) and the power of MLP to model complex, non-linear patterns in high-dimensional data.
* **Improved Accuracy:** The ensemble model benefits from the complementary strengths of MLP and RF, often resulting in higher overall accuracy than either model used in isolation. The fusion of both models allows the combined system to leverage the predictive power of deep learning while maintaining interpretability through the Random Forest component.
* **Reduced Overfitting:** While deep learning models, such as MLP, are prone to overfitting, particularly with small datasets, RF's ensemble approach offers inherent regularization. By combining these models, the ensemble classifier can achieve better generalization and robustness to overfitting.
* **Robustness to Outliers:** RF’s inherent ability to handle noisy or outlier-prone data complements MLP’s susceptibility to overfitting on such data, thereby increasing the overall stability of the ensemble model.
* **Reduced Computational Complexity:** When compared to standalone deep learning models, which can be computationally expensive and require substantial resources for training, the ensemble approach combining MLP and RF offers a more balanced trade-off between predictive accuracy and computational efficiency. While still powerful, the ensemble model can be optimized to reduce training time and resource usage.

While combining Multi-Layer Perceptron (MLP) and Random Forest (RF) in an ensemble framework offers several advantages, there are also some inherent drawbacks associated with this hybrid approach:

1. **Difficulty in Tuning Hyperparameters**:

Both MLP and RF have several hyperparameters that need to be carefully tuned to achieve optimal performance. For example, MLP requires tuning the number of layers, neurons per layer, and the learning rate, while RF needs tuning for the number of trees, maximum depth, and minimum samples split. When combining the two models, the complexity of hyperparameter tuning increases significantly, making it more challenging to find the best set of parameters for the ensemble model.

1. **Model Interpretability**:

One of the advantages of RF is its interpretability, as it is based on decision trees, which allow for the identification of important features. However, MLP, being a deep learning model, lacks transparency in decision-making due to its "black-box" nature. When combined, the interpretability of the RF component can be overshadowed by the opaque nature of the MLP. This can limit the overall transparency of the ensemble model, which is a critical factor in domains like healthcare, where model interpretability is important for trust and adoption.

Comparing Decay Resistance Across Models

Even though leveraging the power of a hybrid model has the potential to significantly improve performance, it is crucial to consider the decay resistance, as highlighted in Table 1. Understanding decay resistance is essential, as it helps to identify how well each model maintains its predictive power over time and with varying data inputs.

This understanding will guide future experimental choices and ensure that the model remains robust and reliable under different conditions. **MLP+LR** provides a **balanced trade-off**, while MLP alone is the most vulnerable to performance decay, particularly in cases of overfitting or class imbalance. MLP+RF offers the highest decay resistance across all key aspects (accuracy, precision/recall, and generalization)

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **MLP** | **MLP+LR** | **MLP+RF** |
| **Accuracy Decay** | High (prone to overfitting) | Moderate (more stable due to LR) | Low (stable with Random Forest) |
| **Precision/Recall Decay** | Moderate (sensitive to class imbalance) | Low (better with LR classifier) | Low (Random Forest handles imbalance) |
| **AUC-ROC/AUC-PRC** | Moderate to High | High | High |
| **Training Time** | Fast | Moderate | Moderate to Slow |
| **Inference Time** | Fast | Moderate | Moderate |
| **Resource Consumption** | Low to Moderate | Moderate | Moderate |
| **Model Size** | Small to Medium | Medium | Large |

Table 1: Comparing Decay Resistance across Models

# Evaluation Method

The evaluation method we have chosen, which involves assessing the performance of our proposed AI models using multiple benchmarks and performance measurement indices, is suitable because it offers a comprehensive view of how well the models will perform in real-world applications, particularly in healthcare settings.

By incorporating both efficiency metrics and conventional performance metrics, we ensure that our models are not only accurate but also practical for deployment, especially in resource-constrained environments like mobile health devices and remote healthcare settings. This holistic evaluation approach allows us to consider both model performance (such as accuracy, sensitivity, specificity, AUC-ROC) and resource efficiency (such as CPU usage, memory usage, energy consumption), which are critical factors in determining whether a model can be effectively used in healthcare applications where resources like CPU power, memory, and energy are often limited.

**Traditional metrics:**

* **AUC-ROC**: Provides an overall assessment of how well the model distinguishes between the two classes (diabetic and non-diabetic)(Hajian-Tilaki, 2013).
* **AUC-PRC**: Important in the case of imbalanced datasets, as it focuses on the precision and recall of the positive class (diabetes), helping to avoid missing at-risk patients(Saito & Rehmsmeier, 2015).
* **Sensitivity**: Minimizes false negatives, which is crucial for early detection of diabetes and preventing complications.
* **Specificity**: Minimizes false positives, preventing unnecessary treatment or concern for patients without diabetes.

**Efficiency Metrics:**

* **CPU Usage Before Training (%)**:This metric identifies whether the **training process** places excessive load on the system's central processing unit (CPU). High CPU usage before training may indicate an inefficient environment setup, which could impact the overall performance of the system.
* **CPU Usage After Training (%)**: This measures the CPU load **after the model has been trained**. High CPU usage post-training might suggest that the model has not been optimized, and it could be consuming more resources than necessary, especially if it is designed to run on **resource-constrained devices** like **healthcare equipment**.
* **Memory Usage Before Training (%):** This metric tracks how much system memory (RAM) is being utilized before model training begins. It helps assess the environment setup, including the libraries and tools loaded into memory.
* **Memory Usage After Training (%):** This indicates the memory consumption after training, which reflects how much memory is required to store the trained model’s parameters (e.g., weights, activations). This metric is particularly useful for ensuring that the model can fit into the memory of devices or systems where memory is limited.
* **Training Time (Seconds):**This metric tracks how long it takes to train the model, measured in seconds. Training time can be influenced by the size of the dataset and model complexity**.**
* **Inference Time (Seconds):**This measures the time it takes for the model to make predictions once trained. In real-world applications like diabetes prediction or medical diagnostics, inference time must be minimized to allow for real-time decision-making.
* **Model Size (MB):**This refers to the disk space required by the trained model, including all parameters (e.g., weights and biases). The model size is an important consideration, particularly for deployment in low-storage environments.
* **Scalability Index (Training Time / Dataset Size):** In healthcare applications, datasets can grow rapidly as more patient data is collected. A model with poor scalability may require disproportionate computational resources and time as the dataset expands, making it less practical for long-term use or continuous updates.
* **Energy Consumption Index (Joules per F1 Score):** Energy efficiency is crucial in resource-constrained environments such as remote healthcare settings or wearable health devices(Santos et al., 2024).

Model Optimization

During the hyperparameter tuning phase, various strategies were employed to optimize model performance. **Grid Search and Random Search** were used to explore different parameter combinations systematically and stochastically, ensuring a comprehensive search for the best configuration. **SHAP analysis** was applied to enhance model interpretability *(*Figure 2.4: SHAP: Interpretability & Explainability (MLP+LR)*) by* identifying the most influential features.

The dataset was split into 70% training, 15% validation, and 15% testing to ensure a well-balanced evaluation and prevent overfitting. Overall, the **MLP+LR** model was selected for deployment as it provides a better balance between accuracy, generalization, and interpretability. While MLP+RF may yield slightly higher predictive performance, its trade-offs in specificity, error rate, and computational demand make it less practical for real-world implementation, particularly in healthcare applications where decision-making must be both reliable and explainable.

**Model Generalization and Performance Trade-offs**

One of the key considerations in model selection is generalization performance, ensuring that the model performs well not only on training data but also on unseen test data. While MLP+RF exhibited higher AUC-ROC (81.10%) and AUC-PRC (66.98%), it struggled with lower specificity (70%) and higher error rates (27%), indicating a higher likelihood of false positives. This suggests that MLP+RF, despite capturing complex patterns, may have overfitted to the training data, leading to reduced reliability on new samples.

In contrast, MLP+LR demonstrated a more balanced trade-off between complexity and interpretability, achieving:

* Lower variance and better generalization, as indicated by comparable performance across training and test sets.
* Improved interpretability through the Logistic Regression component, making it more suitable for healthcare applications where explainability is essential.
* More stable sensitivity and specificity, ensuring a reliable balance between detecting true positives and minimizing false positives.

**Performance vs. Complexity Trade-offs**

* Model Complexity: MLP+RF is inherently more complex due to the ensemble structure, leading to longer training times and increased computational cost. In contrast, MLP+LR maintains a lightweight architecture, making it more efficient for deployment.
* Interpretability vs. Accuracy: While deep models (MLP+RF) excel in capturing intricate data patterns, their "black-box" nature makes them difficult to interpret. MLP+LR offers a compromise by combining neural networks with a transparent linear model, making it more interpretable without a significant loss in accuracy.
* Overfitting Risk: The MLP+RF model is more prone to overfitting due to its higher complexity, as evidenced by its lower specificity and higher error rate. MLP+LR mitigates this issue through regularization and logistic regression constraints, allowing better generalization to new data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Split | 70 (Train Set):15 (Validation Set) | | |  |
| Model | **MLP** | **MLP + LR** | **MLP + RF** |  |
| Performance Metrics |  |  |  |  |
| *Accuracy* | 75% | 75% | 73% |  |
| *G-Means* | 75% | 75% | 74% |  |
| *Sensitivity* | 75% | 75% | 77% |  |
| *Specificity* | 75% | 75% | 70% |  |
| *AUC-ROC* | 80% | 80% | 81% |  |
| *AUC-PRC* | 65% | 65% | 67% |  |
| *Error Rate* | 25% | 25% | 27% |  |

*Table 2 : Model Optimization and Performance Evaluation*

**Selection of Hyperparameters**

The following key hyperparameters were chosen based on their impact on model performance:

* **MLP Hyperparameters:**
  + **Hidden Layer Size = (5,)** → A single hidden layer with 5 neurons was selected to balance model complexity and computational efficiency, preventing overfitting while capturing key patterns in the data.
  + **Learning Rate (0.01)** → A moderately high learning rate was chosen to speed up convergence while avoiding instability in weight updates.
  + **Batch Size (64)** → This batch size was selected to ensure stable training updates while maintaining computational efficiency.
  + **Alpha (0.001, L2 Regularization)** → Regularization was applied to reduce overfitting and improve generalization.
  + **Early Stopping (Validation Fraction = 0.2)** → This setting was used to **monitor validation loss** and stop training when no further improvement was observed, preventing unnecessary computation and overfitting.
* **Logistic Regression Hyperparameters:**
  + **C (0.001, L2 Regularization Strength)** → A low **C value** was chosen to **penalize large coefficients**, reducing the risk of overfitting while maintaining interpretability.
  + **Solver (‘liblinear’)** → This solver was selected because it efficiently handles small datasets and supports **L1 and L2 regularization**, which is beneficial for feature selection and reducing model complexity.
  + **Class Weight (‘balanced’)** → This was applied to compensate for class imbalance, ensuring that the minority class (diabetes-positive cases) was appropriately weighted during training.

# MLP + LR Results

Below are the optimal parameters for the Voting Classifier using soft voting with the MLP+LR model after applying SMOTE, demonstrating significant improvements over the baseline performance.

On the **test set**, the model improved, reaching **77% accuracy**, with **higher specificity (80%)** but slightly lower sensitivity (71%). The **AUC-ROC (84%) and AUC-PRC (75%)** suggest strong discrimination, particularly for positive cases. In terms of **efficiency**, the model is lightweight (**0.02 MB**), trains rapidly (**0.03 seconds**), and has minimal inference time (**0.000 seconds**). Its resource usage remains stable, with a moderate CPU and memory footprint.

|  |  |
| --- | --- |
| **Voting Classifier** | **Best Hyperparameters by Randomized + Grid Search 10-cv** |
| MLP+LR | MLPClassifier(alpha=0.001, batch\_size=64, early\_stopping=True,  hidden\_layer\_sizes=(5,), learning\_rate\_init=0.01, max\_iter=500,  random\_state=200, validation\_fraction=0.2,activation=’relu’)  LogisticRegression(C=0.001, class\_weight='balanced', random\_state=42,  solver='liblinear') |

*Table 3 : MLP+LR Best Parameters*

|  |  |
| --- | --- |
|  | Test Set  (15%) |
| Performance Metrics |  |
| *Accuracy* | 77% |
| *G-Means* | 75% |
| *Sensitivity* | 71% |
| *Specificity* | 80% |
| *AUC-ROC* | 84% |
| *AUC-PRC* | 75% |
| *Error-Rate* | 23% |
| *10-fold CV* | 77% |
| Efficiency Metrics |  |
| *CPU Usage Before Training (%)* | 54% |
| *CPU Usage After Training (%)* | 33% |
| *Memory Usage Before Training (%)* | 49% |
| *Memory Usage After Training (%)* | 49% |
| *Training Time (Seconds)* | 0.03 |
| *Inference Time (Seconds)* | 0.000 |
| *Model Size (MB)* | 0.02 |
| *Scalability Index (Training Time / Dataset Size)* | 0.0000 |
| *Energy Consumption Index (Joules per F1 Score)* | 14.48 |

*Table 4 : MLP+LR Model Performance*

Learning curve

Figure 2 shows that after epochs 62, the model reaches validation loss of 0.50, test loss 0.47 and test accuracy of 77%.

A graph of a graph

AI-generated content may be incorrect.

*Figure 2: MLP Train/ Test Validation loss curve*

The Voting Classifier (MLP+LR) achieves a testing accuracy of approximately **0.77** and training score of around 0.80 with a training set size of 500.

**A graph of a graph

AI-generated content may be incorrect.**

*Figure 3: Voting Classifier : MLP + LR Learning Curve*

Comparative results with theoretical optimum

Our proposed MLP+LR voting classifier achieved an accuracy of 77% when tested on the Pima Indian Diabetes (PID) dataset. In comparison, the average accuracy of deep learning models on similar datasets is around 87%, resulting in a performance gap of 10%.

Regarding ROC-AUC, our model showed a 5% gap compared to other approaches. It is worth noting that the PID dataset has multiple versions, some of which contain missing values, requiring imputation. However, the version we used, sourced from Kaggle, had no missing values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Author / Articles** | **Deep Learning Approach** | **Accuracy** | **AUC-ROC** |
| **Our Proposed:** | MLP+LR | 77% | 84% |
| **PID** | **Av. acc** | **87%** |  |
| (Tumgoyev, 2022) | MLP |  | 90% |
| (Zhang et al., 2024) | BPNN | 89% |  |
| (Butt et al., 2021) | MLP | 86.08% |  |
| (Rajawat et al., 2018) | MLP+RF+XGB | 87.33% |  |

*Table 5 : Dia sense Benchmarks Comparison*

# Dia Sense Integration and Maintenance Plan

### **Quality Assurance & System Validation**

* **Preventing Bottlenecks:** The integration process should anticipate potential bottlenecks in data processing or system communication, particularly when managing real-time medical information. Cloud resources and data streamlining techniques should be employed in an efficient manner.

A diagram of a process flow

AI-generated content may be incorrect.

Figure 4: Full scale Integration Plan

### Key Integration tech Slack Stakeholders:

* **Project Manager:** Oversees the deployment, ensuring timelines and resources are managed.
* **Software Engineer:** Handles API setup, deployment configuration and any coding-related tasks.
* **Data Scientist:** Fine-tunes the predictive model, monitors performance, and adjusts hyperparameters as needed.
* **Quality Assurance (QA)**: Conducts testing for accuracy, functionality, and security.
* **Test Review Team:** Reviews the results from testing to ensure the system functions as intended in a real-world healthcare setting.
* **Audit Team:** Ensuring the system complies with essential healthcare regulations such as **HIPAA** (Health Insurance Portability and Accountability Act) in the U.S., **GDPR** (General Data Protection Regulation) in Europe, and other local regulations related to patient data privacy and security.
* **Medical Professional/Researcher:** Assists with model validation, providing feedback on diagnostic accuracy, and ensuring that the system supports clinicians in their decision-making processes.

### Functionality Elements

#### Scalability & Performance

* **Volume:** Utilize cloud computing platforms to ensure the predictive system can scale and handle large amounts of healthcare data.
* **Data warehouse:** Hadoop, smooth data integration.
* **Containerization:** Utilize **Docker & Kubernetes** for deployment consistency.
* **Velocity:** Optimize API integrations and real-time data processing using NVIDIA TensorRT.
* **Veracity and Variety:** Use advanced analytics and machine learning models to ensure data quality and effectively handle different types of health-related data.

### Non-Functionality Elements

#### Risk Assessment:

* Identify potential risks and challenges, such as system failures, data breaches, or inaccuracies in diagnoses.
* Implement contingency plans to address catastrophic events like network outages, including offline modes or redundant systems, with data backup.

#### Mitigating User Feedback:

* Continuously collect and integrate user feedback for system improvements, including UI updates and clinician-driven feature adjustments.
* Use A/B testing to compare different interface design and assess which is more interactive.

#### Regulatory Compliance:

* Ensure the system adheres to healthcare regulations such as GDPR (General Data Protection Regulation) or HIPAA (Health Insurance Portability and Accountability Act), ensuring the protection of patient data and privacy.

#### Security Protections:

* Implement robust cybersecurity measures to safeguard sensitive patient information. This includes encryption, access control, and secure data storage.

### Testing & Deployment

#### Test Review & Validation

The Test Review Team should consist of:

* **Technical Team**: Experts responsible for assessing system architecture, coding quality, and ensuring the expert system’s algorithms meet clinical standards.
* **Competence Assessment**: Gather feedback from technicians and healthcare professionals to ensure the system meets medical standards before releasing it for full use.
* **Test System Report Release**: After testing is complete, a detailed report should be prepared to confirm that the system is ready for general release.
* **A/B testing** compares two different variants of the prototype to determine which one performs better based on predefined criteria.

#### Deployment & Monitoring

* **Continuous Integration (CI/CD):** **Automate deployments using GitHub Actions.**
* **Kubernetes Orchestration:** Auto-scale API services based on load.
* **Serverless Deployment:** Utilize **AWS Lambda/Google Cloud Functions** for real-time inference.
* **Deployment Platform :** Implement and deploy the system through Streamlit.

### Post-Release: Proactive Maintenance

Once the system is deployed, continuous monitoring and proactive maintenance are crucial to ensure its long-term effectiveness. Several strategies should be implemented:

#### Performance Monitoring

* Model Drift Detection: Implement Kolmogorov-Smirnov tests.
* **Real-Time Metrics Dashboard:**
  + Build an interactive **Power BI-based real-time dashboard** to track:
    - **Model Performance**: Accuracy, Sensitivity, Specificity, AUC-ROC.
    - **System Health**: API response time, error rates, request volume.
    - **Usage Insights**: Number of predictions, peak usage times.
    - **Automated Alerts**: Use **Prometheus Alertmanager** to notify about performance drops or system failures.

#### Automated Model Retraining

* Data Pipeline: Integrate Airflow/Kubeflow for scheduled retraining.
* Model Selection: If MLP+LR accuracy falls below 77%, switch to MLP+RF fallback model.

#### User Feedback & System Updates

* Feedback Loop: Collect clinician feedback for continuous improvement.
* Training & Documentation: Provide user manuals and conduct periodic training.
* Weekly & Monthly Reports: Generate reports on system accuracy and user adoption.

# Discussion

The SHAP (SHapley Additive exPlanations) analysis applied to our MLP+LR model provides key insights into how different features contribute to diabetes-related conditions. Based on SHAP values, individuals were categorized into **Early Diabetes, Diabetes, Hypoglycemia, and Early Hypoglycemia**, and visualized in Tableau for better interpretability.

## Key Findings

**1. Early Diabetes**

* Individuals in this category have **elevated glucose levels (104-113 mg/dL)**, indicating **impaired glucose regulation** but not full diabetes.
* **High DPF values (0.543 - 0.741)** suggest a **strong genetic predisposition** to diabetes.
* **Insulin levels vary:**
  + Some have **high insulin (156 µU/mL at age 41)**, indicating **insulin resistance**.
  + Others have **no insulin (0 µU/mL at ages 62 and 21)**, which may indicate **beta-cell dysfunction or depletion** over time.

**2. Diabetes**

* A 70-year-old individual has a **glucose level of 145 mg/dL**, crossing the diabetes threshold.
* **Insulin is 0**, indicating that the pancreas may have stopped producing insulin, **suggesting full-blown Type 2 Diabetes with beta-cell failure**.
* The **low DPF (0.235)** indicates a weaker genetic component, implying that **age and other factors (e.g., lifestyle, metabolic changes) may have contributed to the condition**.

**3. Hypoglycemia**

* A 37-year-old individual shows **glucose = 0 mg/dL**, which may indicate **severe hypoglycemia or a measurement error**.
* **Insulin is also 0**, suggesting possible **pancreatic failure, insulin overuse, or extreme metabolic dysregulation**.
* DPF is **moderate (0.346)**, meaning genetics might play a partial role, but other factors like diet, medication, or stress may be involved.

**4. Early Hypoglycemia**

* A 26-year-old individual has **glucose = 78 mg/dL**, which is **lower than normal but not critically low**.
* **Insulin is 88 µU/mL**, indicating **excess insulin production**, which can lead to **further drops in blood sugar levels** if not managed properly.
* **Low BP (50 mmHg)** suggests possible **circulatory issues**, often linked to **hypoglycemia-related dizziness or fainting**.

A screenshot of a computer screen

AI-generated content may be incorrect.

## Implications

* **Progression Risk:** Early diabetes cases could progress to **full diabetes** over time, especially in individuals with **high genetic risk (DPF ≥ 0.7) and beta-cell dysfunction (insulin = 0 at older ages).**
* **Metabolic Dysfunction:** The **drop from high insulin (156 µU/mL at age 41) to 0 (at age 62 and 70)** suggests a shift from **insulin resistance to insulin deficiency**, a common progression in **Type 2 Diabetes.**
* **Hypoglycemia Risk:** Cases with **low glucose levels (78 mg/dL and 0 mg/dL)** indicate potential risks of **severe hypoglycemia**, which could lead to **serious health complications** if not managed properly.

## Comparison with Previous Studies

The performance of the MLP+LR hybrid model in this study aligns with certain expectations while also presenting some unexpected findings.

### Expected Outcome:

The model achieved an accuracy of 77% and an AUC-ROC of 84%, which was in line with expectations for a hybrid approach that combines Multi-Layer Perceptron (MLP) and Logistic Regression (LR). MLP effectively captures non-linearity, while LR enhances interpretability in decision-making. Given the nature of the dataset and model architecture, this level of performance was reasonable, though further optimization strategies could potentially improve the results.

### Unexpected Outcome:

Despite achieving a competitive AUC-ROC score, the model’s accuracy falls short compared to several previous studies that applied deep learning techniques to the Pima Indian Diabetes (PID) dataset. Studies such as Zhang et al. (2024) reported higher accuracy scores of 90% , using BPNN (Backpropagation Neural Network). Similarly, models that integrated ensemble learning techniques—such as MLP+RF+XGB (Rajawat et al., 2018) achieved 87.33% accuracy, suggesting that ensemble methods may offer better predictive performance.

Interestingly, while our model underperformed in terms of accuracy, its AUC-ROC of 84% suggests a strong ability to distinguish between positive and negative cases, making it a reliable classifier despite its lower accuracy. This discrepancy highlights the potential benefits of exploring feature selection, additional training data, or ensemble techniques to enhance predictive performance.

## Key Lesson Learnt

Through the experiment, several key learnings emerged:

1. **Hybrid Models Can Offer a Good Trade-off:** Combining MLP with Logistic Regression provided a balance between predictive power and interpretability. MLP is effective in capturing non-linear patterns, while LR helps interpret the results, making the model more transparent, especially for healthcare professionals​.
2. **Class Imbalance is a Major Issue:** The class imbalance (35% diabetic vs. 65% non-diabetic cases) in the dataset posed challenges for the model, although SMOTE helped balance the dataset and improve performance​.
3. **Hyperparameter Tuning Matters:** The combination of randomized search and grid search for hyperparameter optimization demonstrated how fine-tuning model parameters can significantly impact the results, improving overall performance​.
4. **Bridging the Gap Between Machine Learning and Real-World Application**

Through the experiment, we have learned that building an interactive UI to help users understand the implications of Type 2 Diabetes (T2D) is a key takeaway. The Pima Indian Dataset provided a solid base for understanding the impact of diabetes, especially among Pima Indian women in North America. Using MLP techniques combined with voting classifiers helped gain deeper insights into model performance. Additionally, the process of integrating complex machine learning techniques into a user-friendly interface helped bridge the gap between model findings and their real-world application.

## What went well, and what went poorly?

**What went well:**

* **Hybrid Approach:** The combination of MLP and LR leveraged the strengths of both models. While MLP handled complex, non-linear relationships, LR provided transparency and interpretability, which is crucial in healthcare applications where decisions must be explainable​.
* **SMOTE Application:** SMOTE effectively addressed class imbalance, leading to better model generalization and fewer misclassifications of the minority class (diabetic cases)​.
* **Fine-Tuning:** The use of hyperparameter tuning methods like randomized search and grid search significantly improved the model's ability to generalize by finding the optimal set of parameters​.
* **SHAP for Enhanced Model Interpretability**

Integrating SHAP for model interpretability helped enhance the transparency of predictions, making it easier to understand which features had the greatest impact on model outcomes.

**What went poorly:**

* **Computational Complexity:** Combining MLP and LR introduced additional computational complexity, increasing the model’s training time. Tuning both models simultaneously required significant computational resources​.
* **Performance Gap:** The model lagged other state-of-the-art models (e.g., BPNN, MLP+RF+XGB), which achieved higher accuracy, highlighting that the hybrid approach used here needs further refinement to compete with more advanced methods​.

## Limitation: Statistical Methods vs. Expert Knowledge in Diabetes Prediction

**Statistical Methods**

Currently, the model utilizes Youden's J statistic to evaluate performance, optimizing the balance between sensitivity and specificity. This approach relies entirely on statistical techniques to predict outcomes, using only the features present in the dataset. In the absence of expert domain knowledge, such as medical insights into diabetes risk factors, models like the Voting Classifier (MLP+LR) estimate the probability of diabetes based purely on observed data.

These statistical models work well in many cases, particularly when the relationships between features and the outcome are relatively straightforward. However, they may struggle when handling complex dependencies between features or when expert knowledge of causal relationships is necessary for more accurate predictions.

**Bayesian Network**

If expert domain knowledge were available, we could enhance the model by incorporating it into a Bayesian Network. Bayesian networks are powerful probabilistic models that allow reasoning based on conditional dependencies between variables such as age, BMI, glucose levels, family history, and other risk factors. By using a Bayesian network, we could capture the uncertainty and interactions between these features, providing more nuanced predictions. However, developing and using Bayesian networks requires a deep understanding of the domain to define prior probabilities and conditional relationships accurately.

**Future Improvement**

1.**Advanced Hybrid and Ensemble Methods**

* **MLP with XGBoost or Random Forest (RF):** Exploring more sophisticated hybrid approaches could improve both linear and non-linear pattern recognition, enhancing predictive accuracy.
* **Fuzzy Logic Integration:** Incorporating fuzzy logic would allow the model to handle uncertainty and imprecise data, particularly for variables like glucose levels that fluctuate continuously.
* **Deep Belief Networks (DBN):** Combining **MLP with DBN** could enhance feature extraction, leading to better data representation and classification.
* **Genetic Algorithms (GA):** Leveraging GA for **hyperparameter optimization and feature selection** could help identify optimal configurations, reducing overfitting and improving generalization.

2. **Privacy-Preserving Learning Techniques**

* **Federated Learning:** This approach enables model training across decentralized data sources without transferring sensitive patient information, ensuring compliance with **privacy regulations (e.g., GDPR, HIPAA)** while maintaining model performance.
* **Differential Privacy:** Applying differential privacy techniques would add noise to sensitive data, enhancing security while preserving model utility.

3. **Data Augmentation & Class Imbalance Handling**

* Exploring advanced resampling techniques like **ADASYN or borderline SMOTE** could further improve dataset balance and enhance minority class recognition.

**4. Enhanced Feature Engineering**

* **Domain-Specific Features:** Integrating additional features, such as genetic markers or more detailed medical history, could refine the model's ability to detect subtle patterns in diabetes risk.

**Conclusion**

The MLP+LR hybrid model serves as a solid foundation for predicting diabetes in healthcare applications. While it demonstrated reasonable performance in terms of accuracy and sensitivity, it also faced challenges such as, class imbalance, and computational complexity. The performance gap compared to more advanced models suggests that while the hybrid approach is promising, further refinement is needed.

Future iterations can benefit from:

* **Advanced Techniques**: Incorporating Fuzzy Logic, Deep Belief Networks (DBN), Genetic Algorithms (GA), and ensemble methods to enhance predictive power.
* **Model Optimization**: Reducing complexity without sacrificing accuracy to improve efficiency for real-world applications.
* **Data Privacy & Ethics**: Leveraging federated learning to ensure patient data privacy while maintaining model performance.
* **Real-Time Deployment:** Optimizing the model for low-latency, high-efficiency predictions in clinical decision-making.
* **Consider Fairness and Bias**:Evaluate models for fairness and bias by analyzing performance across different subgroups and sensitive attributes.

As the field of AI in healthcare continues to evolve, DiaSense represents a significant step forward in improving patient outcomes. Furthermore, its development is grounded in humanized AI, ensuring that its implementation prioritizes ethical principles, fairness, transparency, and respect for individual privacy.

This work lays the foundation for AI-driven diagnostic tools in healthcare, with the potential for even greater advancements and more personalized, accessible care in the future.

# References

Al Sadi, K., & Balachandran, W. (2023). Revolutionizing Early Disease Detection: A High-Accuracy 4D CNN Model for Type 2 Diabetes Screening in Oman. *Bioengineering (Basel)*, *10*(12). <https://doi.org/10.3390/bioengineering10121420>

Asghari Varzaneh, Z., & Hosseini, S. (2023). An Intelligent Fuzzy System for Diabetes Disease Detection using Hawks Optimization. *Artificial Intelligence and Data Mining*, *11(2)*, 187-194. <https://doi.org/10.22044/JADM.2023.12383.2383>

Asha, V., Lamani, M;R., Padmaja, K., & Gondhale, T. A. (2024). Enhancing Diabetes Prediction Accuracy with AdvanSVM: A Machine Learning Approach Using the PIMA Dataset. *Seybold Report Journal*, *19(05)*, 51-72. <https://seyboldpublications.com/wp-content/uploads/2024/05/Asha-V.pdf>

Bashir I. A, M. A. A., Banu R, Elyas T. B. (2019). Significance of Health Related Predictors of Diabetes in Pima Indians Women. . *7(2)*. <https://doi.org/http://dx.doi.org/10.12944/CRNFSJ.7.2.05>

Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. *J Healthc Eng*, *2021*, 9930985. <https://doi.org/10.1155/2021/9930985>

Choudhury, Z. (2023). *Integrating Random Forest, MLP and DBN in a Hybrid Ensemble Model for Accurate Breast Cancer Detection*. <https://doi.org/10.31219/osf.io/sdjqf>

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*, *4*(2), 627-635. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3755824/>

Health., H. (2024). *Type 2 diabetes mellitus*. Retrieved 30 January 2025 from <https://www.health.harvard.edu/diseases-and-conditions/type-2-diabetes-mellitus-a-to-z>

Kaggle. (2016). *PIMA Indians Diabetes Database*. Retrieved January 27,2025 from <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>

Muhammad Mazhar Bukhari, B. F. A., Saddam Hussain, Abdu Gumaei, Adel Assiri, Syed Sajid Ullah. (2021). An Improved Artificial Neural Network Model for Effective Diabetes Prediction. <https://doi.org/https://onlinelibrary.wiley.com/doi/10.1155/2021/5525271>

Nelson, R. G., Knowler, W. C., Kretzler, M., Lemley, K. V., Looker, H. C., Mauer, M., Mitch, W. E., Najafian, B., & Bennett, P. H. (2021). Pima Indian Contributions to Our Understanding of Diabetic Kidney Disease. *Diabetes*, *70*(8), 1603-1616. <https://doi.org/10.2337/dbi20-0043>

Pearson, E. R. (2015). Dissecting the Etiology of Type 2 Diabetes in the Pima Indian Population. *Diabetes*, *64*(12), 3993-3995. <https://doi.org/10.2337/dbi15-0016>

Pietukhov, R., Ahtamad, M., Faraji-Niri, M., & El-Said, T. (2023). A hybrid forecasting model with logistic regression and neural networks for improving key performance indicators in supply chains. *Supply Chain Analytics*, *4*, 100041. <https://doi.org/https://doi.org/10.1016/j.sca.2023.100041>

Rajawat, P. S., Gupta, D. K., Rathore, S. S., & Singh, A. (2018). Predictive Analysis of Medical Data using a Hybrid Machine Learning Technique. *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 228-233. <https://doi.org/10.1109/ICSCCC.2018.8703302>

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, *10*(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>

Santos, S. O. S., Skiarski, A., García-Núñez, D., Lazzarini, V., Moral, R. D. A., Galvan, E., Ottoni, A. L. C., & Nepomuceno, E. (2024). Green Machine Learning: Analysing the Energy Efficiency of Machine Learning Models. *2024 35th Irish Signals and Systems Conference (ISSC)*, 1-6. <https://doi.org/10.1109/ISSC61953.2024.10603302>

Streamlit. (2025). *Diabetes Risk Prediction*. <https://ict619diasense-bmoqpf3atdcrw7ush9podz.streamlit.app/>

Tumgoyev, Y. (2022). *Diabetes prediction using Multilayer Perceptron* NAZARBAYEV UNIVERSITY]. <https://nur.nu.edu.kz/server/api/core/bitstreams/88aee630-02ae-40d9-a299-f58efcfc097e/content>

Tunç, T. (2012). A new hybrid method logistic regression and feedforward neural network for lung cancer data. *Mathematical problems in engineering*, *2012*(1), 241690. <https://onlinelibrary.wiley.com/doi/full/10.1155/2012/241690>

Zeng, F. N. J. (2024). Evaluating Traditional Machine Learning Models for Predicting Diabetes Onset Using the Pima Indians Dataset. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4878052>

Zhang, Z., Ahmed, K. A., Hasan, M. R., Gedeon, T., & Hossain, M. Z. (2024). A Deep Learning Approach to Diabetes Diagnosis. <https://arxiv.org/html/2403.07483v1>

# Appendix

### *Appendix 1: PIMA Indian Diabetes Dataset*

The Pima Indian Diabetes dataset (Kaggle, 2016) comprises 769 observations and 9 variables, which include both predictor (independent) variables and a target (dependent) variable labelled "Outcome." The target variable indicates whether an individual has diabetes, with a binary encoding: 0 represents no diabetes, and 1 represents the presence of diabetes.

The independent variables in the dataset encompass a variety of physiological and medical factors, including the number of pregnancies, age, body mass index (BMI), blood pressure (BP), insulin level, and additional related features. No missing values were found.

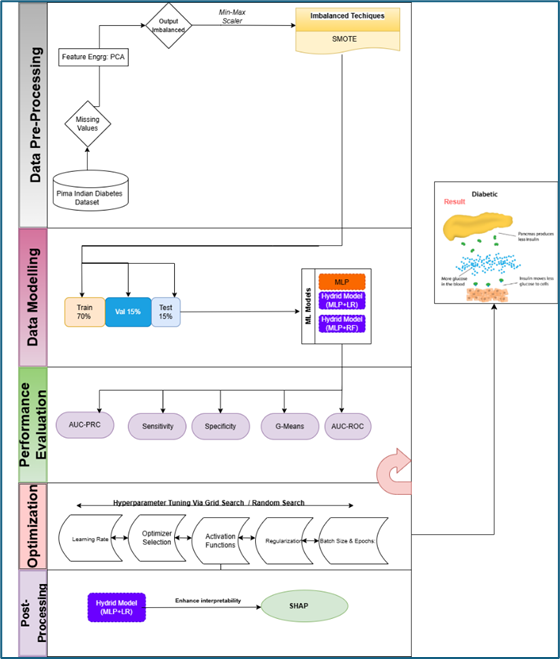
|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **Description** | **Data Type** | **Mean** |
| Pregnancies | The number of pregnancies a woman has had. | Int64 | 3.4 |
| Glucose | Blood glucose concentration after 2 hours in an oral glucose tolerance test (OGTT). | Int64 | 120.89 |
| BloodPressure | Diastolic blood pressure (mm Hg). | Int64 | 69.11 |
| SkinThickness | Thickness of the skin fold (mm). | Int64 | 20.54 |
| Insulin | 79.99 2-hour serum insulin (mu U/ml). | Int64 | 79.80 |
| BMI | Body mass index (weight in kg / height in m²). | Float64 | 31.99 |
| DiabetesPedigreeFunction | A function that scores the likelihood of diabetes based on family history. | Float64 | 0.47 |
| Age | Age of individuals | Float64 | 33.24 |
| Outcome | 0 = No Diabetes, 1 =Diabetes | Int64 | 0.35 |

The proposed machine learning pipeline is designed to enhance model performance in handling imbalanced data through a series of key steps. The process begins with data preprocessing and checking for missing values, followed by feature engineering with PCA to retain the eight principal components. SMOTE is then applied to address class imbalance, followed by splitting the data into 70% for training, 15% for validation, and 15% for testing, using 10-stratified k-fold cross-validation.

Three classifiers; MLP, Logistic Regression (LR), and Random Forest (RF)—are trained using a voting classifier to combine their outputs. The models are evaluated using multiple performance metrics, including G-Means, AUC-PRC, AUC-ROC, Sensitivity, and Specificity.

Hyperparameter optimization is performed using Grid Search and Random Search (including parameters like learning rate, optimizer, activation functions, regularization, batch size, and epochs), with cross-validation for robustness.

Once the best-performing hyperparameters are identified, the selected model is retrained on the full SMOTE-resampled training dataset to ensure it captures all available information. The optimized model is then evaluated on the held-out test set (15%) to assess its generalization ability. To enhance model interpretability, SHapley Additive Explanations (SHAP) is applied, providing insights into the contribution of each feature toward the model’s predictions. This step ensures transparency and facilitates a deeper understanding of the decision-making process.



#### Figure 1.1: Experimental design

### *Appendix 2: Data Pre-Processing and Post-Processing for Voting Classifier (MLP+LR)*

Pre-processing and post-processing are critical steps in developing an AI model to ensure data quality, enhance model performance, and improve interpretability.

**Data Pre-Processing/Data Cleaning**

Data pre-processing involves preparing raw data for training by addressing issues such as inconsistencies, and feature scaling. The following steps were applied:

1. **Min-Max Scaling**

* MLP is sensitive to the scale of input features since it relies on gradient-based optimization. Without scaling, large feature values can lead to unstable training or slow convergence.
* Logistic Regression (LR) benefits from scaling as it improves numerical stability and prevents certain features from dominating the decision boundary.

1. **Train-Test-Validation (70:15: 15)**

* **Training Set (70%):** Used for training the Voting Classifier (MLP+LR).
* **Validation Set (15%):** Used for hyperparameter tuning and model selection.
* **Test Set (15%):** Used for final evaluation of model performance.

1. **Imbalanced techniques: SMOTE**

The Pima Indian Diabetes dataset exhibits a **highly imbalanced** distribution, with 268 positive instances (indicating diabetes) and 500 negative instances, resulting in a distribution of 35% positive and 65% negative cases.

A pie chart with a pie chart and a diagram

Description automatically generated

#### Figure 2.1: Target variable (Outcome) distribution

**4.Spearman Rho Analysis**

Most features exhibit correlations below 0.8, which is within multicollinearity benchmarks**.**

As shown in Figure xx, the strongest positive correlation with the target variable (diabetes diagnosis) is observed for **'Glucose'** (0.47). This suggests that higher glucose levels are associated with a higher likelihood of diabetes diagnosis. Additionally, **BMI, Age, and being pregnant (females)** are also linked to a higher probability of a diabetes diagnosis.

To determine if there is a relationship between the categorical variable ‘outcome’ (target variable) and other independent variables, the Spearman’s Rho rank correlation test is applied. The hypotheses for this test are as follows:

**Hypotheses for Spearman's Rho correlation Test:**

*H0: There is a significant monotonic relationship between the target variable and each feature.*

*H1: There is no significant monotonic relationship between the target variable and each feature.*

***Statistical Inferences***

**1. For features with p-value < 0.05:**

For features such as **Glucose**, **BMI**, **Age**, **Pregnancies**, **Diabetes Degree Function**, **Blood Pressure**, and **Skin Thickness**, the p-values are less than the significance level (α = 0.05). Therefore, we **reject the null hypothesis** and conclude that there is a **statistically significant monotonic relationship** between these features and the target variable (diabetes diagnosis).

2.**For the feature “Insulin” (p-value > 0.05 and rho = 0.06):**

The p-value for Insulin is greater than the significance level (α = 0.05) and the correlation value is weak (rho = 0.06). Therefore, we fail to reject the null hypothesis and conclude that there is insufficient evidence to suggest a significant monotonic relationship between Insulin and the target variable (diabetes diagnosis).

***A close-up of a graph

Description automatically generated***

#### Figure 2.2 : Spearman Rho Analysis

**5.Principal Component Analysis**

Based on the cumulative explained variance we retained the first 8 principal components (PC8), which explain 100% of the variance. These components will be used for model training.

A graph with a line

Description automatically generated

A screen shot of a graph

Description automatically generated

#### Figure 2.3: PCA Components and Explained Variance

**6.Data post-processing**

To enhance the interpretability of the Multi-Layer Perceptron (MLP) within the Voting Classifier (MLP+LR), SHAP (SHapley Additive exPlanations) was applied, specifically focusing on the Logistic Regression (LR) component. This method provides insight into feature importance, helping to explain how individual variables influence the model’s predictions, thereby improving transparency and trust in the classification process.

Following model training on PCA-transformed data, SHAP values were analyzed to assess the impact of each feature on diabetes prediction. The results revealed that **Preganacies (SHAP value = 0.35)** and **Glucose (SHAP value = 0.15)** were the most influential features, significantly contributing to both positive and negative classifications.

Conversely, **Skin Thickness and Age** had the lowest SHAP values, close to zero, indicating minimal impact on the classifier’s decision-making process. This suggests a weaker correlation between these features and the target variable in the current model setup.

A graph of a graph showing a value

AI-generated content may be incorrect.

A graph with blue bars

AI-generated content may be incorrect.

#### Figure 2.4: SHAP: Interpretability & Explainability (MLP+LR)

### *Appendix 3 : List of the three models code from Hyperparameter Tuning*

**a. MLP**

To identify the best combination of MLP hyperparameters, we performed randomized search. Randomized search explores the hyperparameter space by randomly sampling combinations, which can be more efficient than a grid search, especially when dealing with a large number of parameters. The parameters tuned include:

* Number of Hidden Layers
* Activation Function
* Solver:
* Alpha (Regularization term)
* Batch Size
* Early Stopping

A screenshot of a computer program

AI-generated content may be incorrect.The hyperparameter tuning process, combined with SMOTE, successfully optimized the MLP model as shown in figure 3.1.

#### Figure 3.1. MLP Optimization code

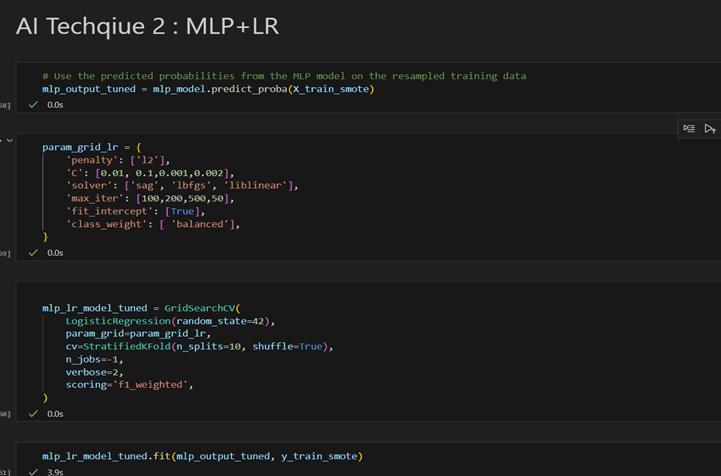
**b. MLP + LR**

The creation of a hybrid model that combines Multi-Layer Perceptron (MLP) and Logistic Regression (LR) in figure 3.2 include key steps:

1. The best MLP model was trained, and its probabilities were extracted.
2. Fitting MLP Probabilities into Logistic Regression:

* The extracted MLP probabilities were used as features for training a Logistic Regression model.
* Hyperparameters for LR (penalty, regularization strength C, and solver) were tuned using grid search.

1. SMOTE for Class Imbalance
2. Ensemble Learning with Soft Voting: The final model combines MLP and Logistic Regression using soft voting. This ensemble approach averages the predicted probabilities from both models to make the final prediction.



A screenshot of a computer

AI-generated content may be incorrect.

#### Figure 3.2. Voting Classifier: MLP + LR Optimization code

**c. MLP +RF**

The building of a hybrid model that combines Multi-Layer Perceptron (MLP) and Random Forest (RF) in figure 3.3 include key steps:

* 1. The best MLP model was trained, and its probabilities were extracted.
  2. Fitting MLP Probabilities into Random Forest:
* The extracted MLP probabilities were used as features for training a Random Forest model.
* Hyperparameters for RF (n\_estimators, max\_depth, min\_samples\_split, min\_samples\_leaf, criterion were tuned using randomised search.
  1. SMOTE for Class Imbalance

A screenshot of a computer program

AI-generated content may be incorrect.Ensemble Learning with Soft Voting: The final model combines MLP and Random Forest using soft voting. This ensemble approach averages the predicted probabilities from both models to make the final prediction.

A screenshot of a computer

AI-generated content may be incorrect.

#### Figure 3.3. Voting Classifier: MLP + RF Optimization code

Once all the models were successfully optimized, we compared their performance using the ROC-AUC and ROC-PRC curves, as shown in Figures 3.4 and 3.5. The MLP+RF model outperformed the other two models, achieving a ROC-AUC score of 81% and a ROC-PRC score of 67%. In comparison, the MLP and MLP+LR models showed similar performance, with ROC-AUC scores of 80% and ROC-PRC scores of 65%, respectively. However, due to MLP+RF's susceptibility to overfitting, the final model chosen for deployment is **MLP+LR**, as it offers a more stable performance.

A graph of a model

AI-generated content may be incorrect.

#### Figure 3.4. ROC-AUC Curve Comparison

A graph of a graph

AI-generated content may be incorrect.

#### Figure 3.5. ROC-PRC Curve Comparison

### *Appendix 4 : Final model MLP+LR*

A screenshot of a computer

Description automatically generated

#### Figure 4.1: MLP + LR Step by step code

The **confusion matrix** shows that the MLP+LR classifier correctly predicted 60 negative cases (True Negatives) and 29 positive cases (True Positives). However, it misclassified 15 negatives as positives (False Positives) and 12 positives as negatives (False Negatives). This suggests a moderate trade-off between sensitivity and specificity. The MLP+LR classifier shows relatively good overall accuracy (around 77%). However, it reveals some class imbalance. The model performs better at predicting class 0 (high specificity) than class 1 (lower recall).

The MLP+LR model demonstrates strong performance with AUC values of 0.85 (training) and 0.84 (test). The close agreement between training and test AUCs indicates good generalization and minimal overfitting.

A blue squares with white text

AI-generated content may be incorrect.

#### Figure 4.2: Confusion Matrix : MLP+LR

A graph of a curve

AI-generated content may be incorrect.

#### Figure 4.3: ROC-AUC: MLP+LR (Train vs Test)

### *Appendix 5 : Dia sense UI Applications*

**How the Dia sense AI System Works**

The Diasense AI system is a user-friendly diabetes risk assessment tool that utilizes machine learning to predict whether an individual is at high or low risk of developing diabetes. The system is built on a trained model using data from the Pima Indian Diabetes dataset, which includes key health indicators such as age, BMI, glucose level, insulin, blood pressure, diabetes pedigree function (DPF), skin thickness, and pregnancy count.

Diasense employs a hybrid AI model, integrating Multi-layer Perceptron (MLP) and Logistic Regression (LR) to enhance accuracy. This combination allows the system to analyze patterns in user-provided health data and calculate a risk probability score for diabetes.

The system is hosted on (Streamlit, 2025),allowing users to enter their data through a simple web interface.

**1. Enter Personal Health Data**: Users fill in the required health parameters.

**2. View Results**: The system instantly calculates and displays their diabetes risk probability.

**3. Take Action**: Based on the results, users can consult a healthcare professional for further evaluation and lifestyle modifications.

**Example Assessment**

**High Risk** (e.g., 85.54%) → More likely to have diabetes

* **Age: 45**
* **BMI: 27**
* **Pregnancies: 3**
* **Glucose: 189**
* **Insulin: 0**
* **Blood Pressure: 0**
* **Diabetes Pedigree Function (DPF): 0.08**
* **Skin Thickness: 0**

A Pima Indian woman aged 45, with a Diabetes Pedigree Function (DPF) of 0.08 and a glucose level of 189 mg/dL, presents a high risk of early-onset diabetes. Several clinical indicators support this assessment:

1. Elevated Glucose Levels (189 mg/dL) – This measurement significantly exceeds the normal fasting glucose range (70–99 mg/dL) and **surpasses** the prediabetic threshold (100–125 mg/dL), suggesting **hyperglycemia**, a primary **indicator** of **diabetes mellitus**.
2. Low Diabetes Pedigree Function (DPF) (0.08) – While a higher DPF is associated with a strong familial predisposition to diabetes, a low DPF does not preclude the risk of developing the condition. Instead, this finding suggests that non-genetic factors, such as environmental influences and lifestyle choices, may play a predominant role in disease progression.
3. Maternal Age (45) and Pregnancy History (3 pregnancies) – Epidemiological research indicates that women over the age of 45 who have experienced multiple pregnancies are at an **increased risk of developing type 2 diabetes.** This correlation is partly due to the physiological changes associated with pregnancy, including gestational diabetes, which can contribute to long-term metabolic dysregulation.
4. Absence of Measurable Insulin (0) and Blood Pressure (0 mmHg) – The reported absence of insulin secretion suggests potential pancreatic beta-cell dysfunction, impairing glucose regulation. Additionally, an extremely low or unrecorded blood pressure measurement may indicate an underlying circulatory or metabolic disorder, further exacerbating the likelihood of diabetes onset.

Based on these clinical parameters, the Diasense AI system estimates an 85.54% probability of diabetes for this individual, highlighting the need for immediate medical evaluation and targeted interventions to mitigate disease progression.

**A screenshot of a medical survey

AI-generated content may be incorrect.**

#### Figure 5.1: UI Application: High Risk

**Low Risk** (e.g., **29.11%)** → Less likely to have diabetes

* **Age: 20**
* **BMI: 10**
* **Pregnancies: 0**
* **Glucose: 4**
* **Insulin: 0**
* **Blood Pressure: 45**
* **Diabetes Pedigree Function (DPF): 0.47**
* **Skin Thickness: 35**

**For a Pima Indian woman aged 20 with the given health parameters, the Diasense AI system predicts a low diabetes risk (29.11%). The implications of these factors are as follows:**

1. Young Age (20 years old) – Youth is a protective factor against type 2 diabetes, as metabolic function and insulin sensitivity tend to be more efficient in younger individuals.
2. Low BMI (10) – A BMI of 10 is extremely low, potentially indicating underweight status or malnutrition. While obesity is a major risk factor for diabetes, being severely underweight may suggest other health concerns, including nutritional deficiencies.
3. No Pregnancies (0) – The absence of pregnancies eliminates any risk related to gestational diabetes, a condition that can increase susceptibility to type 2 diabetes later in life.
4. Extremely Low Glucose (4 mg/dL) – This value is abnormally low, likely an error, as normal fasting glucose ranges from 70 to 99 mg/dL. If accurate, such a low glucose level could indicate hypoglycemia, which is not typically linked to diabetes but may suggest other metabolic or endocrine disorders.
5. Insulin = 0 – The lack of measurable insulin is concerning, as it suggests a possible issue with insulin production. However, in the absence of hyperglycemia, this may not immediately indicate diabetes.
6. Blood Pressure (45 mmHg) – This is significantly below the normal range (typically 90/60 mmHg or higher). If accurate, such a low blood pressure might indicate hypotension, which could be due to dehydration, malnutrition, or another underlying condition.
7. Moderate Diabetes Pedigree Function (DPF = 0.47) – A moderate DPF suggests a moderate genetic predisposition to diabetes, meaning family history could play a role, but environmental and lifestyle factors remain critical determinants.

The Diasense AI system predicts a low diabetes risk (29.11%) for this individual. However, the extremely low BMI, glucose, and blood pressure levels may indicate potential nutritional deficiencies or underlying metabolic disorders that warrant medical attention. Despite the low risk of diabetes, maintaining a balanced diet, regular physical activity, and routine health check-ups is essential for long-term metabolic health and overall well-being.

A screenshot of a medical survey

AI-generated content may be incorrect.

#### Figure 5.2: UI Application: Low Risk