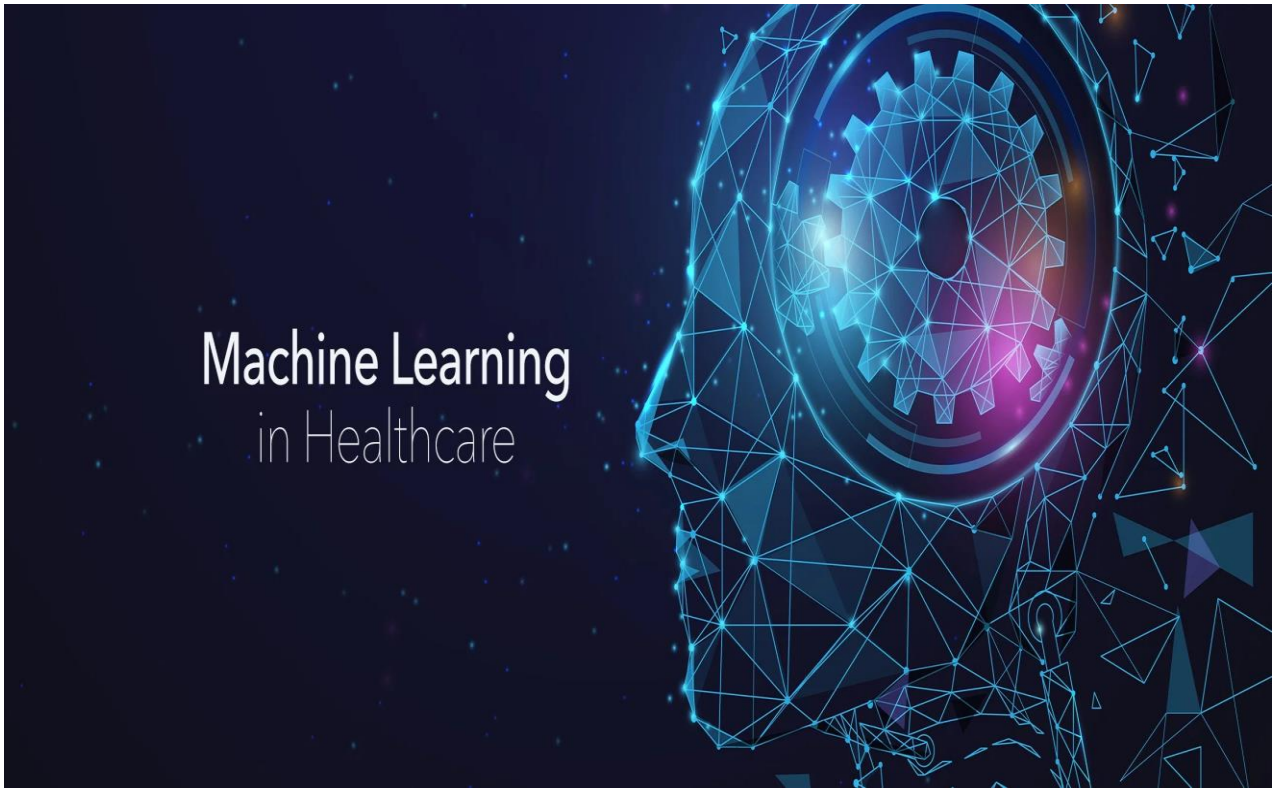
الكلية متعددة التخصصات - ورزازات
FACULTÉ POLYDISCIPLINAIRE DE OUARZAZATE

Intelligence Artificielle et Ingénierie Logiciel

Réalisé par : Amina El Helymy



Medical Risk Assessment Application

Année universitaire 2025-2026

Table des matières

Introduction

Healthcare risk prediction is one of the most important applications of Artificial Intelligence in modern medicine.

Early detection of diseases such as Heart Disease, Diabetes, and Hypertension can significantly reduce complications and improve patient outcomes.

This project, Medical Risk Assessment, is an interactive web application developed using Python and Streamlit, designed to:

- ❖ Predict disease risk levels
- ❖ Provide explainable AI results
- ❖ Assist clinical decision-making (educational purpose)

⚠ Disclaimer:

This application is for educational and academic use only and must not be used as a real medical diagnostic tool.

Project Objectives

The main objectives of this project are:

- ➢ Build a unified platform for **multiple disease risk predictions**
- ➢ Apply **Machine Learning (Logistic Regression)** on medical datasets
- ➢ Provide **risk levels** (*Low* / *Medium* / *High*)
- ➢ Explain predictions using **feature contribution analysis**
- ➢ Create an **interactive and user-friendly interface**

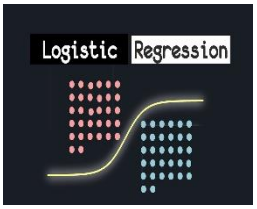
Technologies & Tools Used

Python: Used as the main programming language to build the entire machine learning pipeline, from data processing to model training and application deployment.

Pandas: Used for data loading, manipulation, preprocessing, and structuring datasets into clean tabular formats suitable for machine learning models.

Scikit-learn: Used to build, train, and evaluate machine learning models, providing tools for preprocessing, model pipelines, and performance optimization.

Plotly-Express: Used to generate interactive data visualizations and feature contribution charts to enhance model interpretability and user understanding.



Logistic-Regression : Applied as the core predictive model due to its interpretability and suitability for medical risk classification problems.

Streamlit: Used to develop an interactive web application that allows real-time user input and displays medical risk predictions with clear visual feedback.

VSCode(Visual-Studio-Code): Used as the primary development environment for writing, testing, debugging, and organizing the project code efficiently.

User Interface Overview

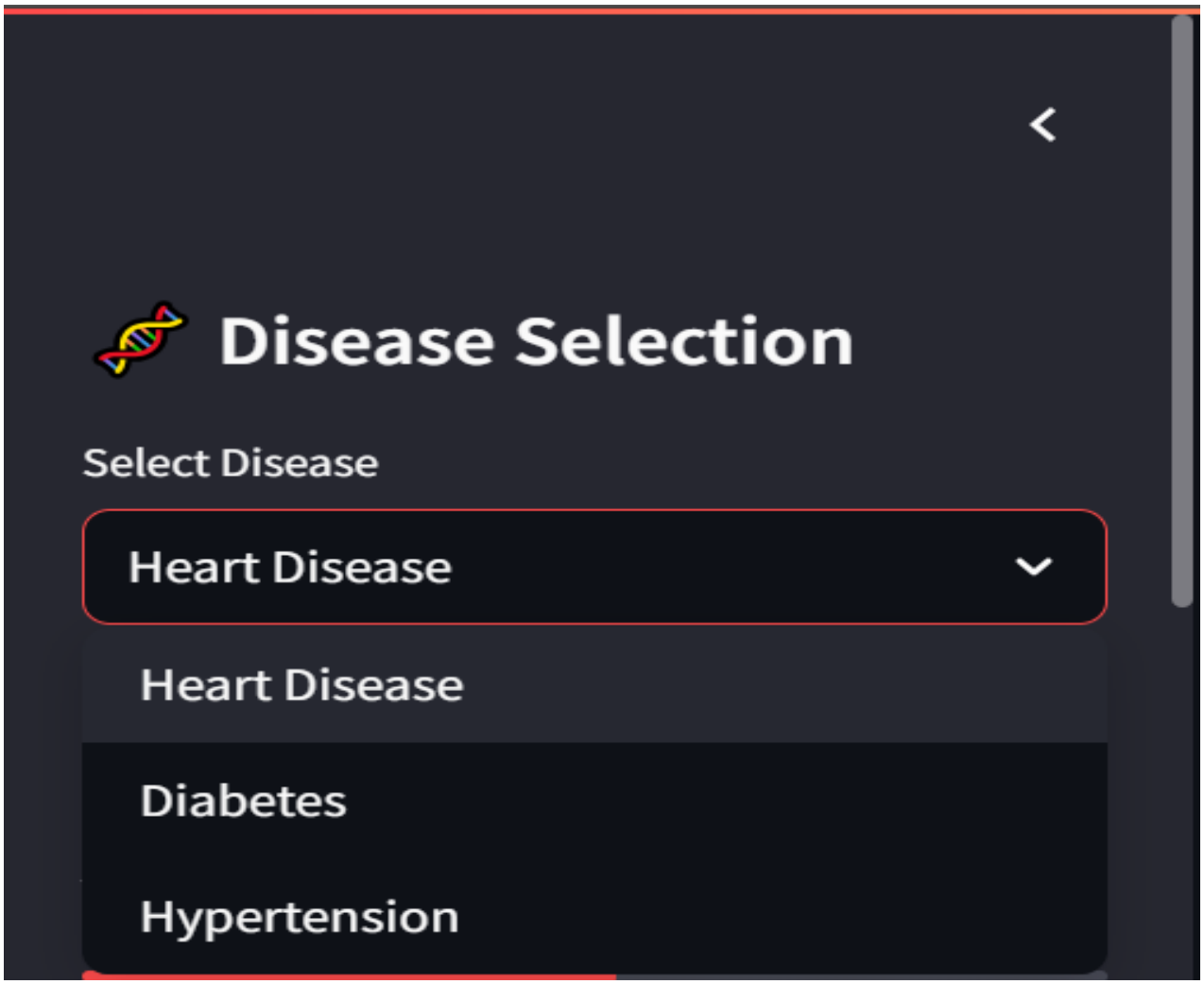
Disease Selection (Sidebar) :

The application provides an intuitive sidebar that allows the user to select the disease to assess.

Three medical conditions are available:

- Heart Disease
- Diabetes
- Hypertension

Once a disease is selected, the interface dynamically updates to display the corresponding input form, prediction button, and results section.



Disease selection sidebar showing the three available diseases

Heart Disease Risk Assessment Module

Dataset

The Heart Disease module is based on the UCI Heart Disease dataset, which contains clinical attributes commonly used for cardiovascular diagnosis.

The dataset includes the following features:

- *Age*
- *Sex*
- *Chest Pain Type*
- *Resting Blood Pressure*
- *Cholesterol Level*
- *Maximum Heart Rate*
- *ST Depression (oldpeak)*
- *Number of Major Vessels*
- *Thalassemia*

Model Used

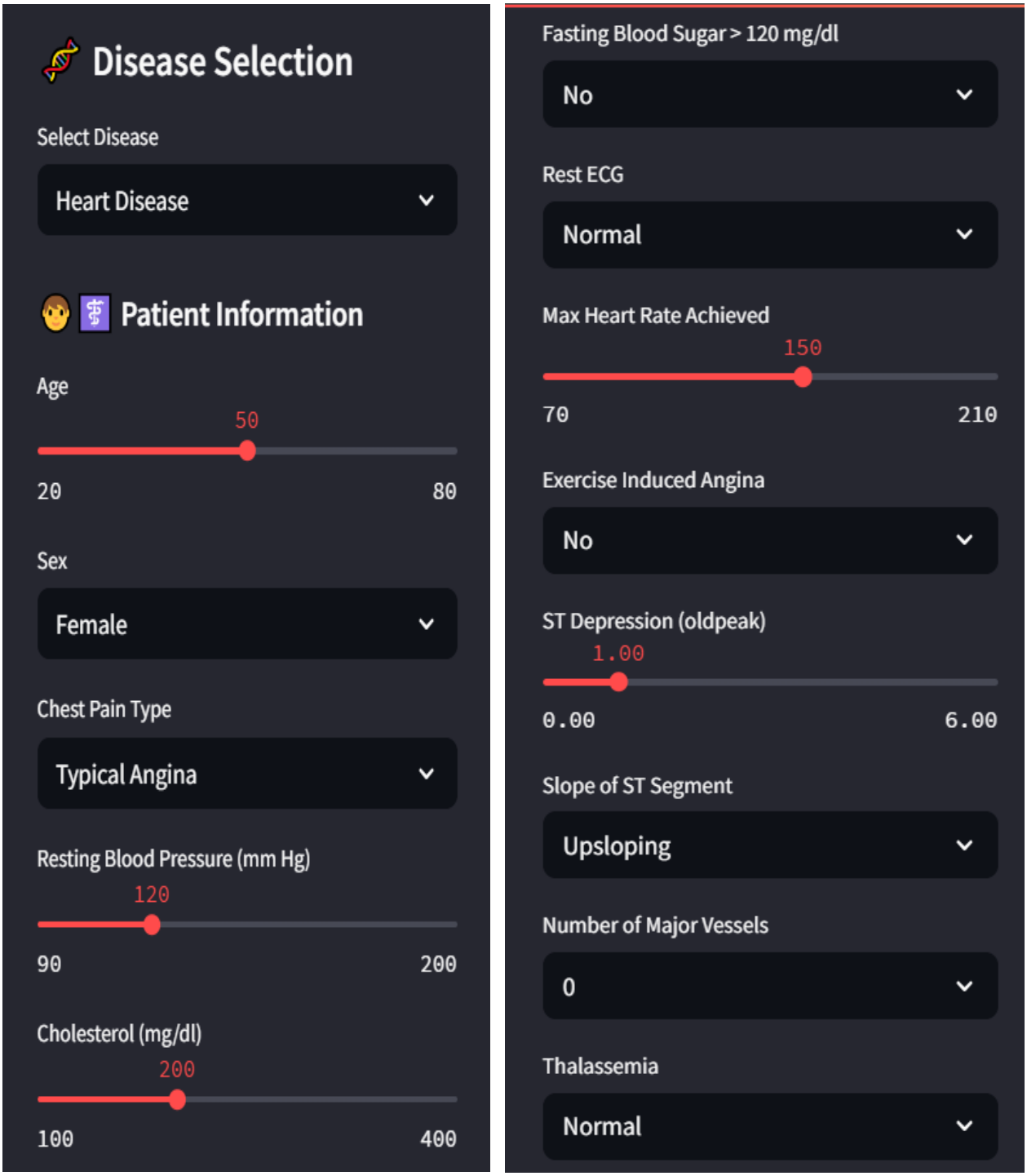
A Logistic Regression model is used to predict the probability of heart disease occurrence. The model is trained using a machine learning pipeline that includes:

- *Feature scaling using StandardScaler*
- *Logistic Regression classifier*
- *Class balancing to handle data imbalance*

User Inputs

The user inputs patient clinical data using interactive sliders and dropdown menus in the sidebar.

These inputs simulate real-world medical parameters used in cardiovascular assessment.



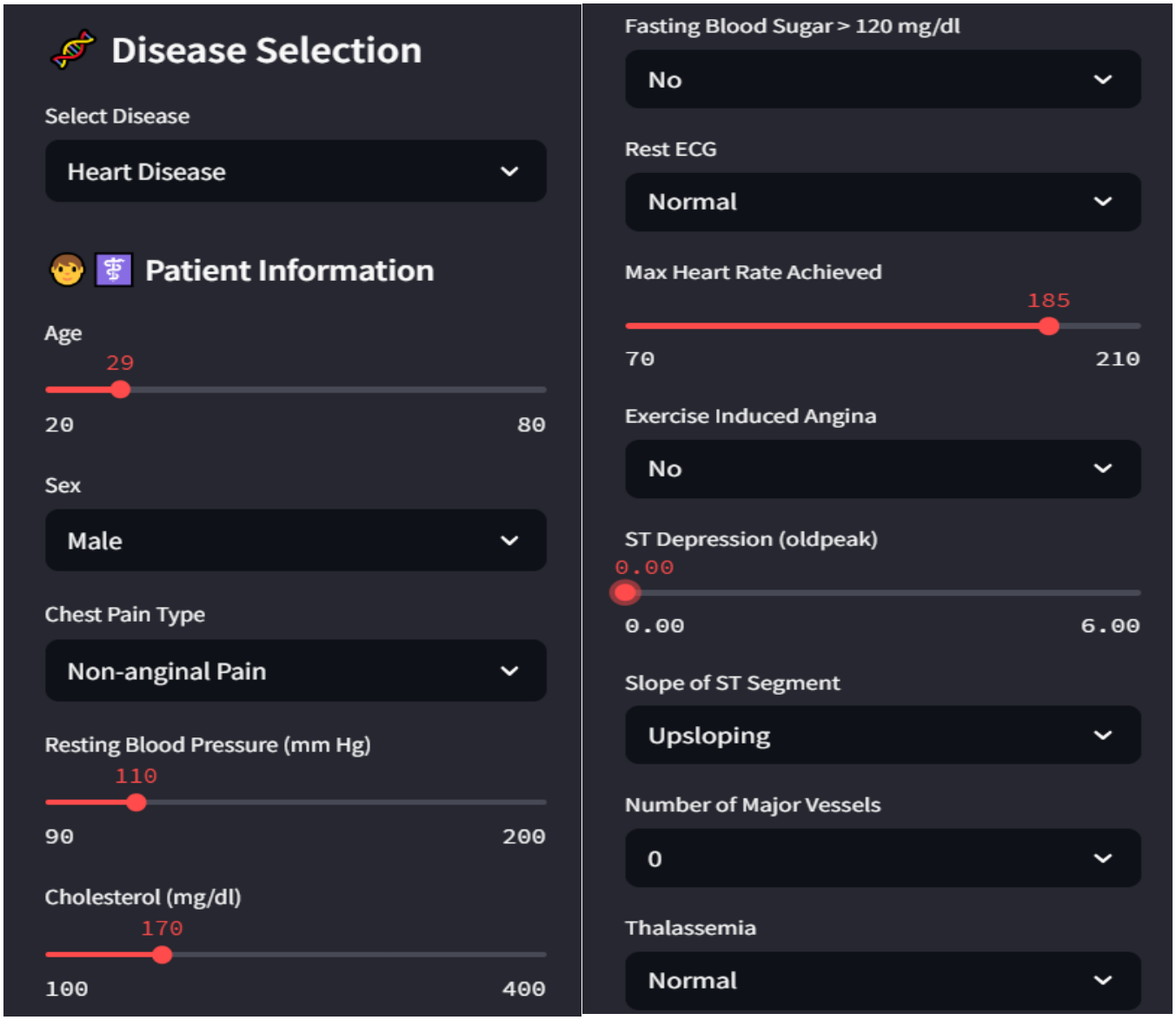
Heart disease patient input interface.

The predicted probability is interpreted into three clinical risk levels:

Risk Level	Probability
LOW	< 50%
MEDIUM	50% – 80%
HIGH	≥ 80%

LOW Risk Test :

This test represents a patient with normal clinical values, resulting in a low cardiovascular risk prediction.



This test case represents a patient with generally healthy cardiovascular indicators. The AI model predicts **no presence of heart disease**, with a very low estimated risk probability of **4.93%**, which is classified as **LOW risk**.

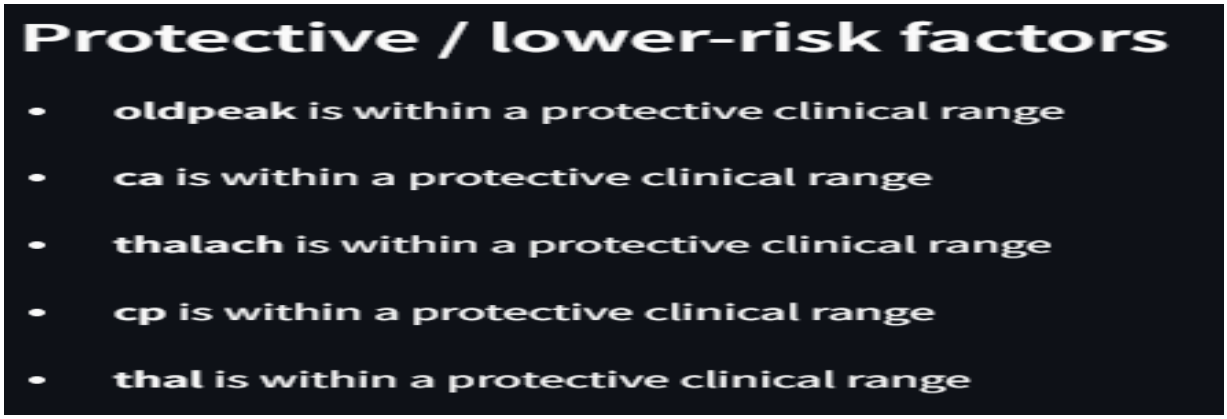
Factors Increasing Disease Risk

Although a few features such as slope, sex, and resting ECG slightly contribute to increasing cardiovascular risk, their impact remains minimal and does not significantly affect the overall prediction.

On the other hand, several important protective factors strongly reduce the risk, including:

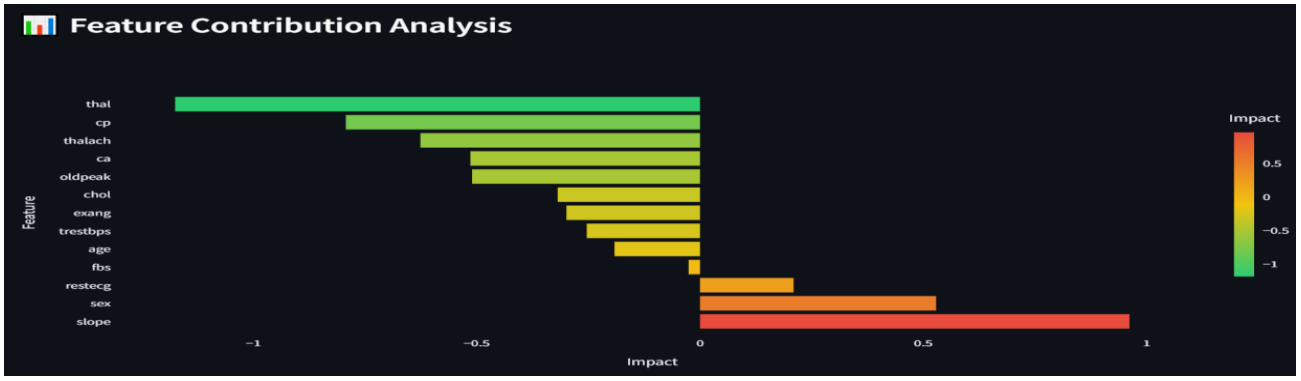
- ST depression (*oldpeak*) within a normal range
- Number of major vessels (*ca*) indicating no arterial blockage
- Maximum heart rate achieved (*thalach*) reflecting good cardiac performance
- Chest pain type (*cp*) associated with lower risk
- Thalassemia status (*thal*) within normal clinical values
- The feature contribution analysis confirms that protective factors outweigh risk factors, leading to a confident LOW cardiovascular risk prediction.

Protective / Lower-Risk Factors



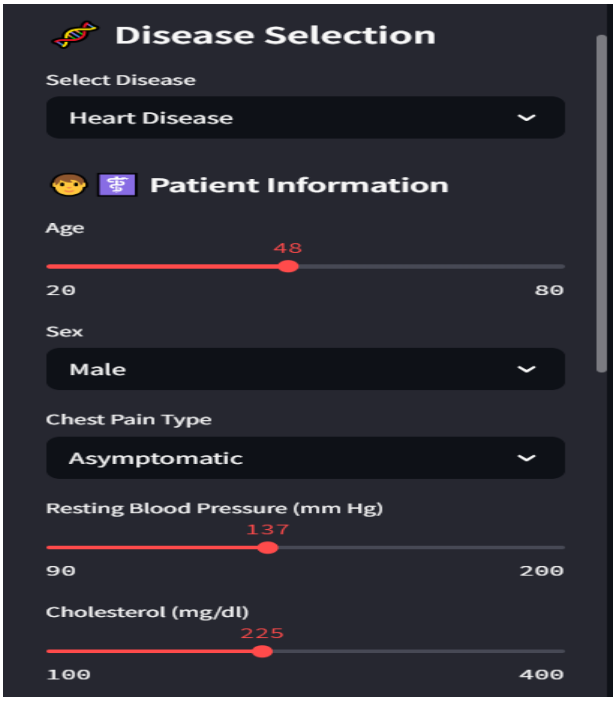
Model Explainability

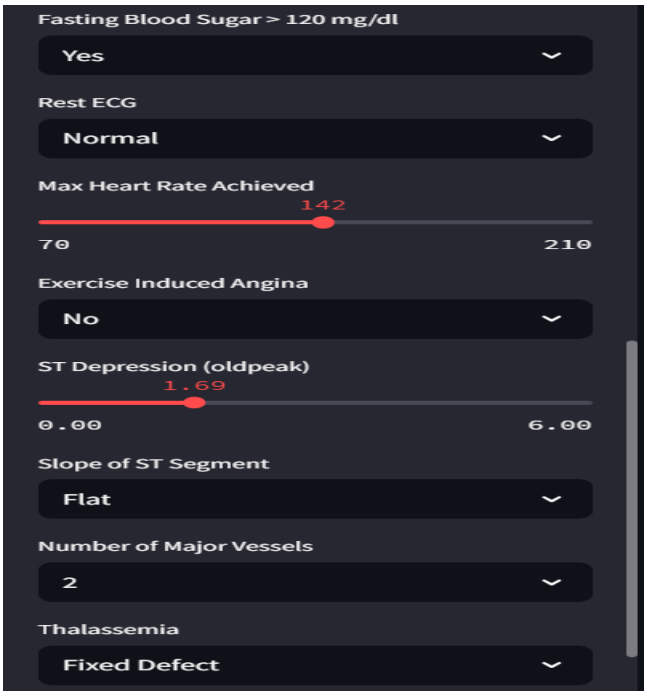
The **Feature Contribution Analysis** provides transparency by highlighting how each clinical feature influences the prediction, supporting interpretability and trust in the model's decision.



MEDIUM Risk Test

This test represents a patient with moderate risk factors, leading to a medium risk prediction.





This test case represents a patient with moderate cardiovascular risk.

The AI model predicts the presence of heart disease, with a risk probability of 61.36%, which falls into the MEDIUM risk category.

Factors Increasing Disease Risk

Several clinical features significantly contribute to increasing cardiovascular risk, including:

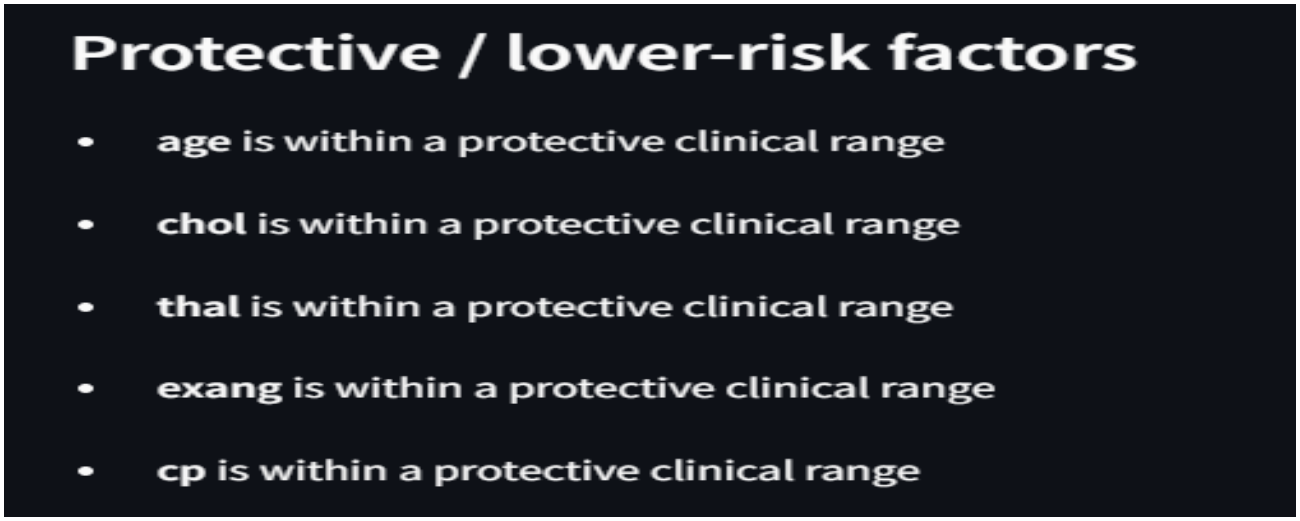
- Number of major vessels (*ca*), suggesting possible arterial narrowing
- Sex, which is a known non-modifiable cardiovascular risk factor
- ST depression (*oldpeak*), indicating abnormal cardiac stress response
- Slope of the ST segment, associated with myocardial ischemia
- Resting ECG (*restecg*), showing potential electrical abnormalities

Protective / Lower-Risk Factors

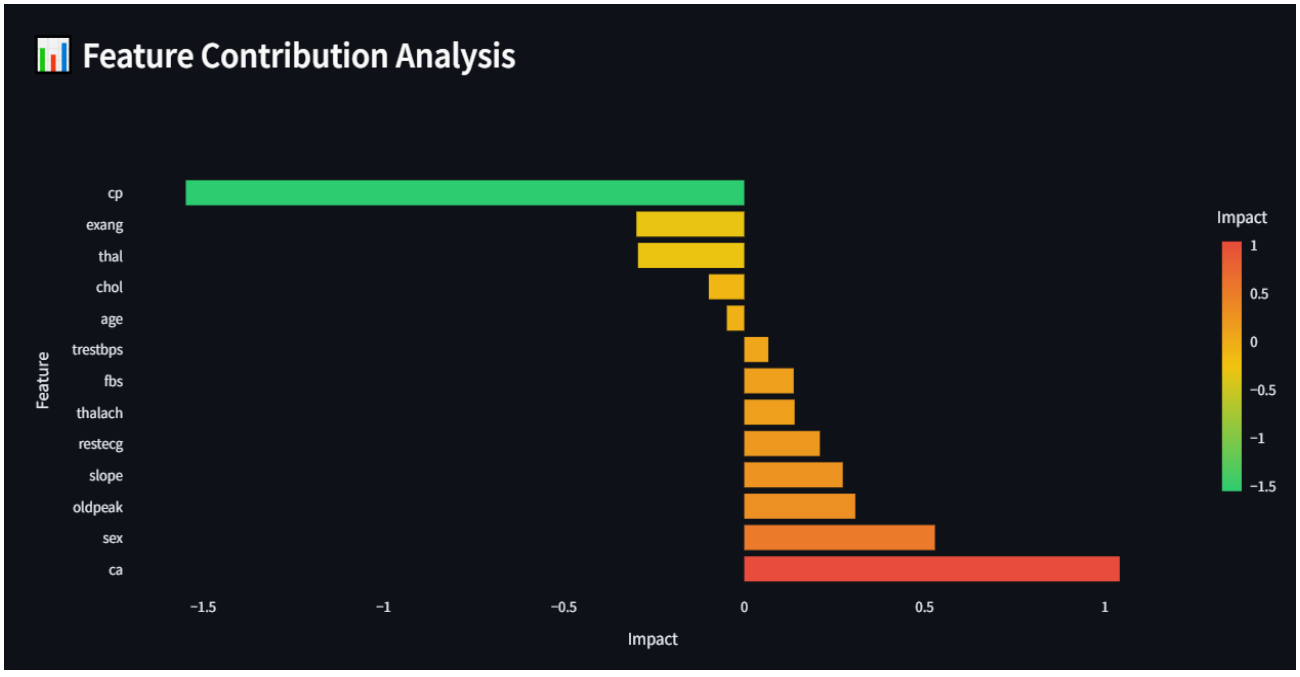
Despite these risk-increasing factors, multiple protective features help reduce the overall severity of the condition:

- Age within a lower-risk range
- Cholesterol level (*chol*) within acceptable clinical limits
- Thalassemia status (thal) indicating no major blood disorder
- Exercise-induced angina (*exang*) absent
- Chest pain type (*cp*) associated with lower cardiovascular risk

The balance between risk factors and protective factors results in a moderate risk classification, highlighting the need for clinical monitoring rather than immediate critical intervention.

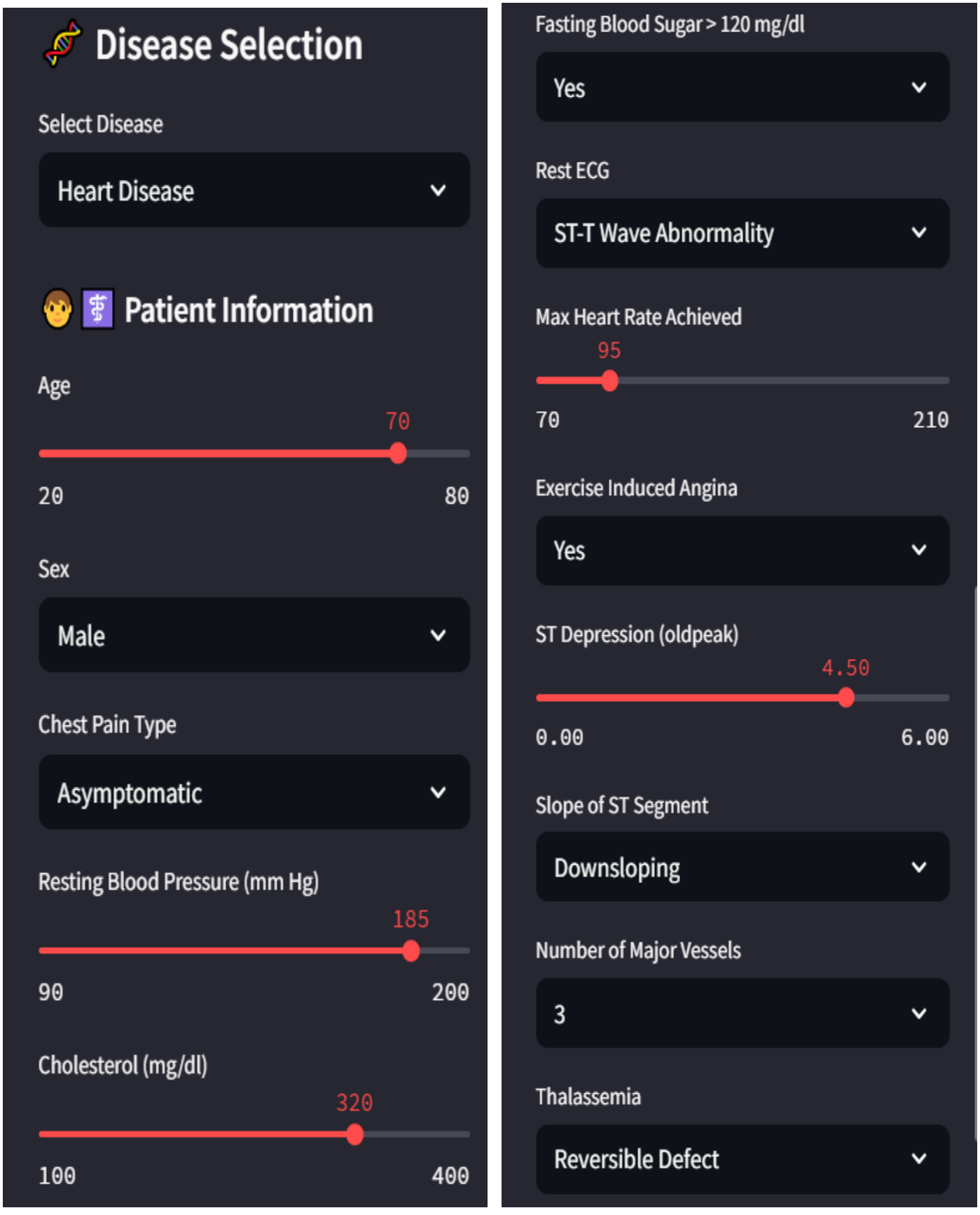


The feature contribution analysis visually explains how each parameter influences the prediction, ensuring transparency and interpretability of the AI model.



HIGH Risk Test

This test simulates a high-risk patient with multiple abnormal clinical indicators.



This test case represents a high-risk cardiovascular profile.

The AI model predicts the presence of heart disease with a very high probability of 99.46%, which clearly places the patient in the HIGH risk category.

Factors Increasing Disease Risk

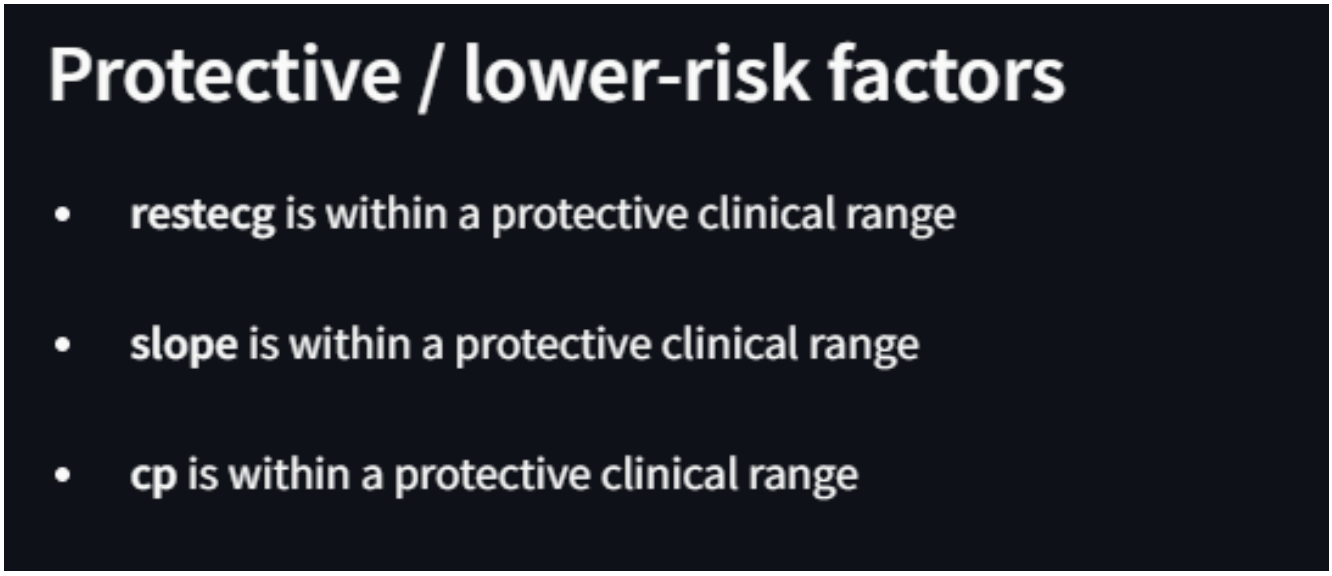
Several critical clinical factors strongly increase the cardiovascular risk:

- Number of major vessels (*ca*), indicating significant coronary artery involvement
- ST depression (*oldpeak*), reflecting severe cardiac stress abnormalities
- Maximum heart rate achieved (*thalach*), suggesting abnormal cardiac response to exertion
- Resting blood pressure (*trestbps*), which is considerably elevated
- Thalassemia (*thal*), associated with impaired blood oxygen transport

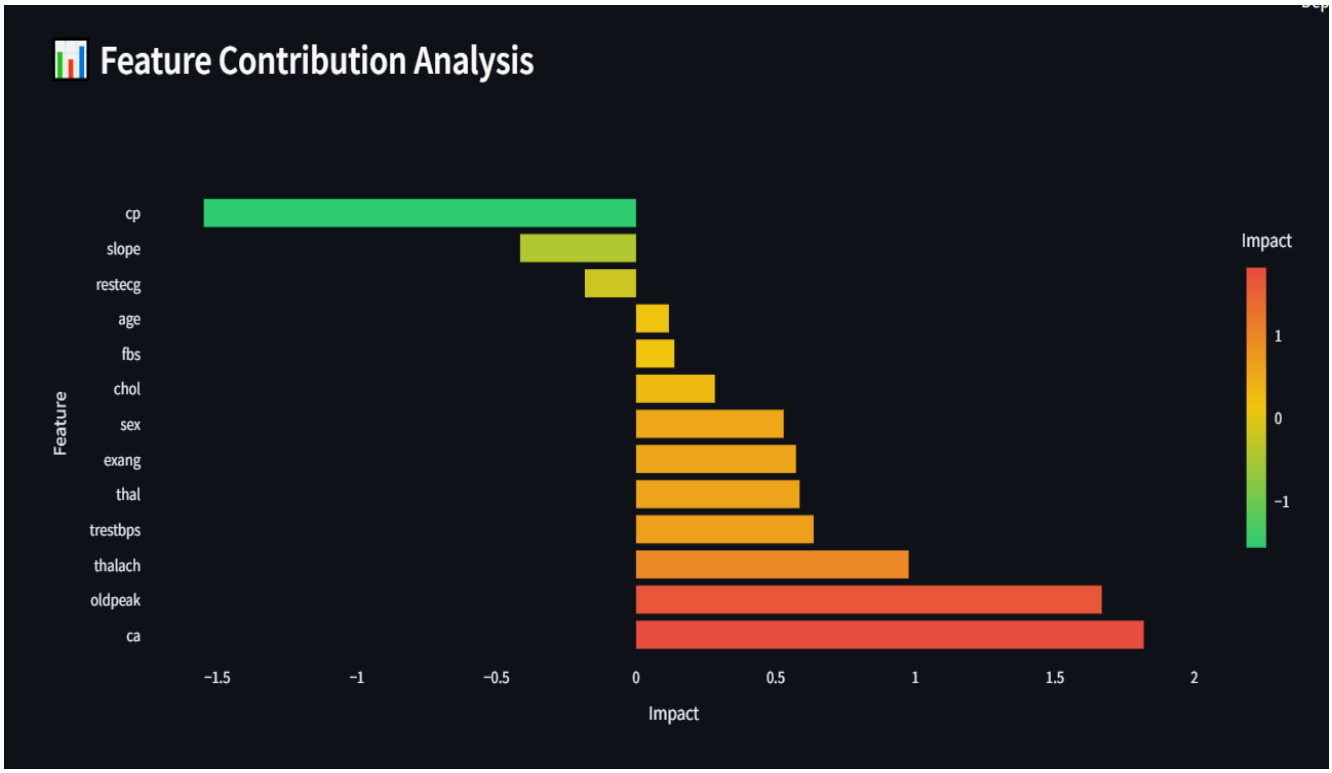
Protective / Lower-Risk Factors

Although a few parameters remain within protective clinical ranges—such as resting ECG (restecg), ST segment slope, and chest pain type (cp)—their positive impact is insufficient to counterbalance the severity of the high-risk features.

The model therefore classifies this patient as HIGH cardiovascular risk, indicating an urgent need for clinical evaluation and medical intervention.



The feature contribution analysis clearly highlights the dominant influence of high-risk parameters, providing transparent and explainable AI-driven decision support.



Diabetes Risk Assessment Module

Dataset

The Diabetes module uses the Pima Indians Diabetes Dataset, which includes metabolic and demographic features such as:

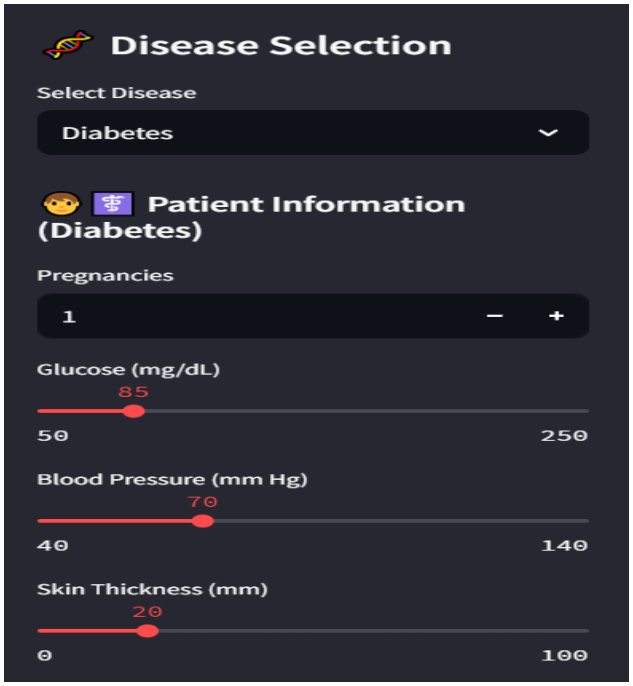
- *Glucose*
- *BMI*
- *Insulin*
- *Age*
- *Pregnancies*
- *Blood Pressure*

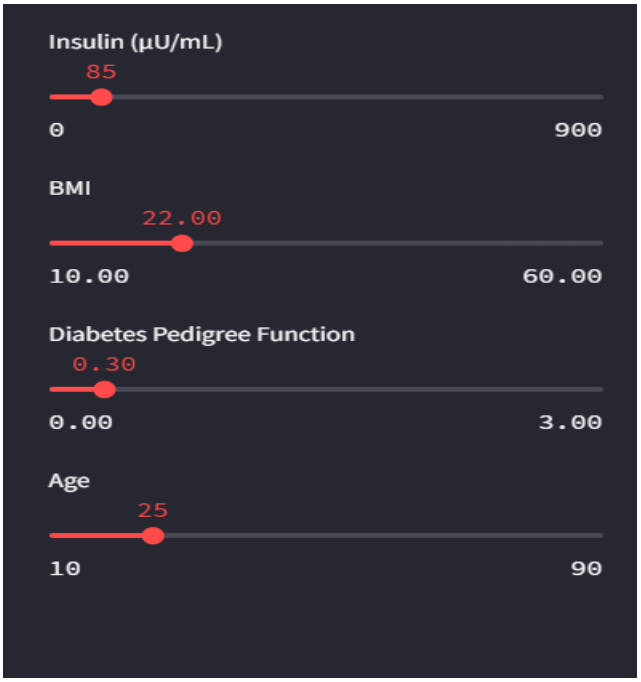
Risk Thresholds

Risk Level	Probability
LOW	< 45%
MEDIUM	45% – 75%
HIGH	≥ 75%

Diabetes – Test Scenarios

LOW Risk





The AI-assisted system predicts no diabetes, with a very low risk probability of 4.36%, classifying the patient as LOW RISK.

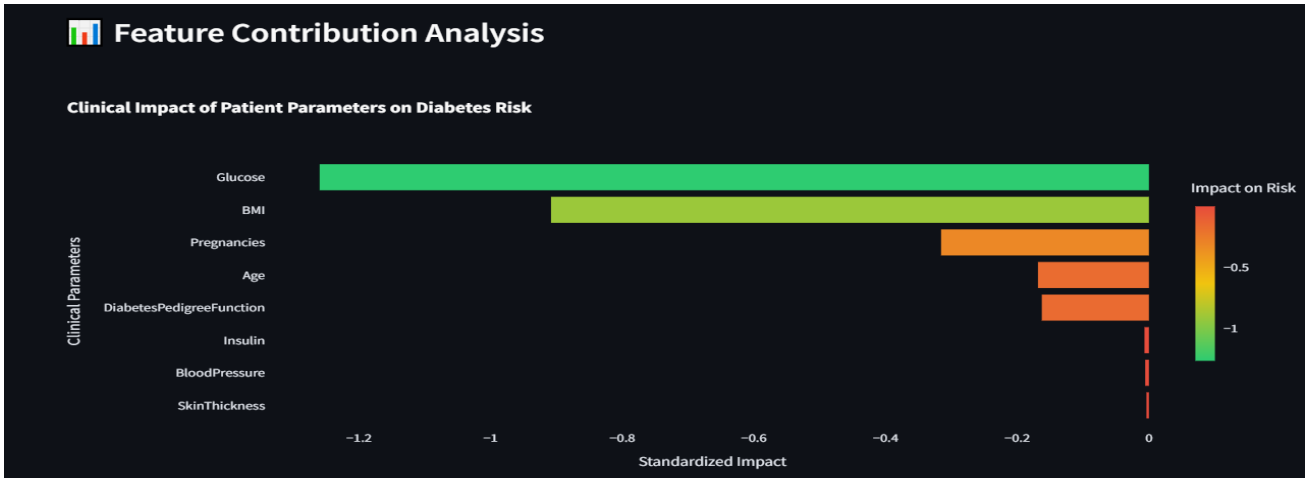
Factors Increasing Diabetes Risk & Protective / Lower-Risk Factors

All major clinical indicators fall within healthy and protective ranges:

- ❖ Glucose level is normal, indicating good blood sugar regulation
- ❖ BMI reflects a healthy body weight
- ❖ Age does not contribute to metabolic risk
- ❖ Pregnancies count is within a safe range
- ❖ Diabetes Pedigree Function suggests low genetic predisposition
- ❖ No significant risk-increasing factors are detected.

The feature contribution analysis confirms that protective factors dominate the prediction, resulting in a stable and reassuring clinical outcome.

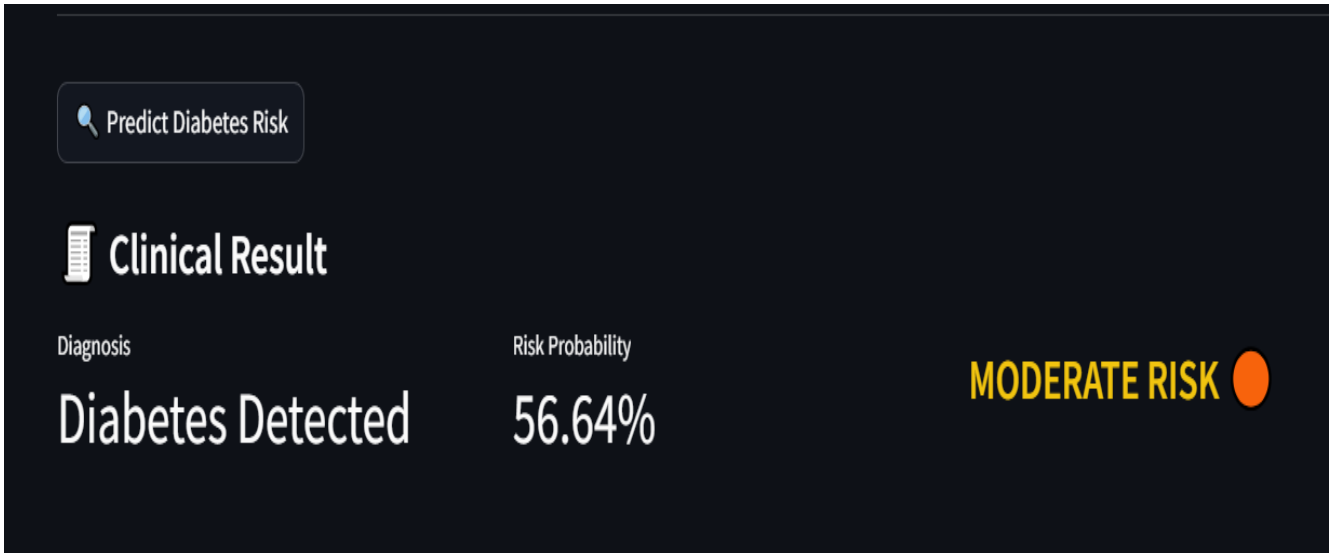
Overall, the model concludes that the patient has a very low likelihood of developing diabetes, and no immediate medical intervention is required.



MEDIUM Risk Test

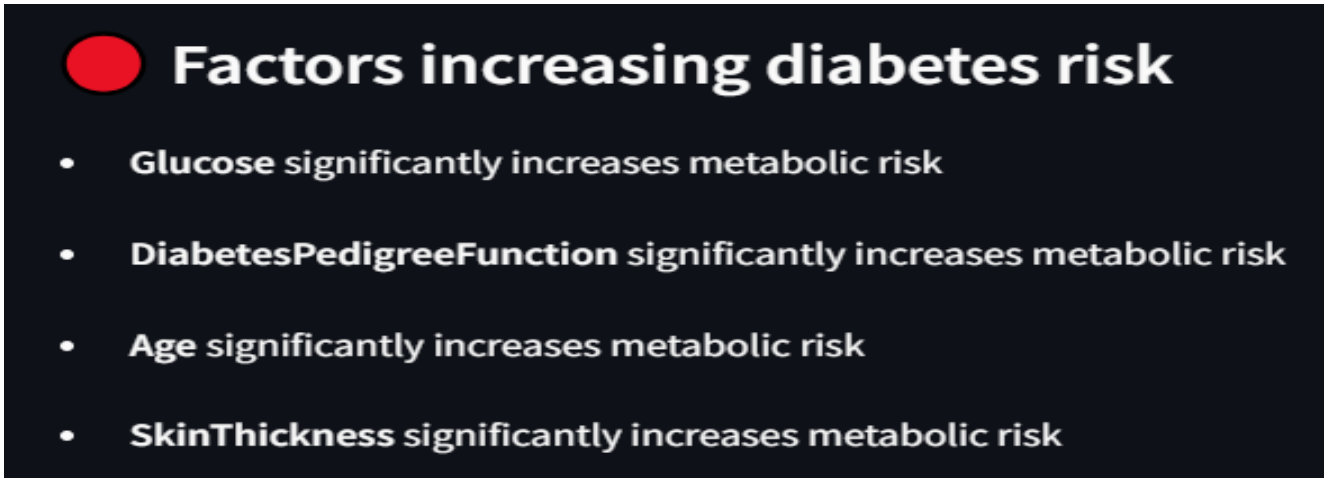
This test case corresponds to a moderate diabetes risk profile.

This assessment indicates a **moderate probability of diabetes**. The patient presents **several metabolic risk factors** that significantly increase the likelihood of developing diabetes, although some protective clinical indicators are still present.



Factors Increasing Diabetes Risk

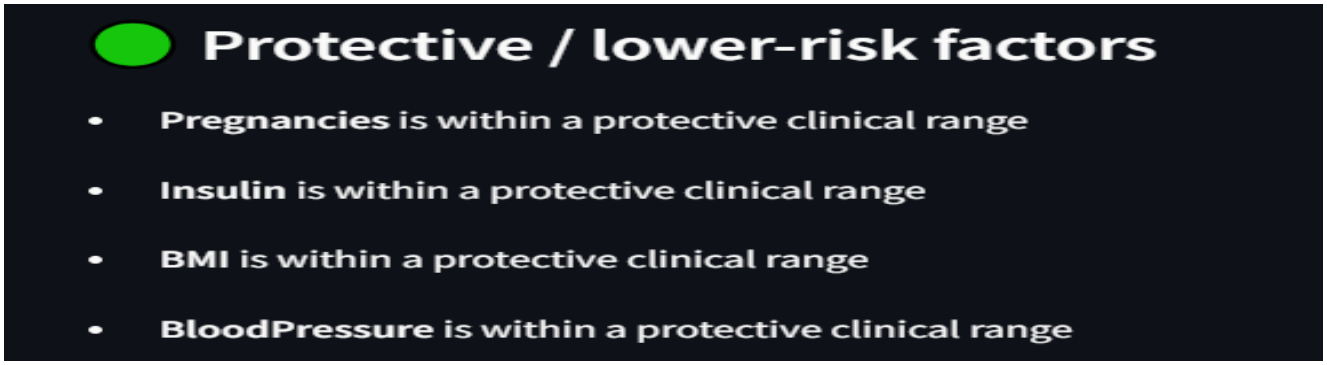
- **Glucose** levels are significantly elevated, indicating impaired glucose regulation and a high metabolic risk.
- **Diabetes Pedigree Function** suggests a strong genetic predisposition to diabetes.
- **Age** contributes to increased risk due to reduced metabolic efficiency over time.
- **Skin Thickness** indicates possible insulin resistance and abnormal fat distribution.



These factors collectively raise concern and justify closer medical monitoring.

Protective / Lower-Risk Factors

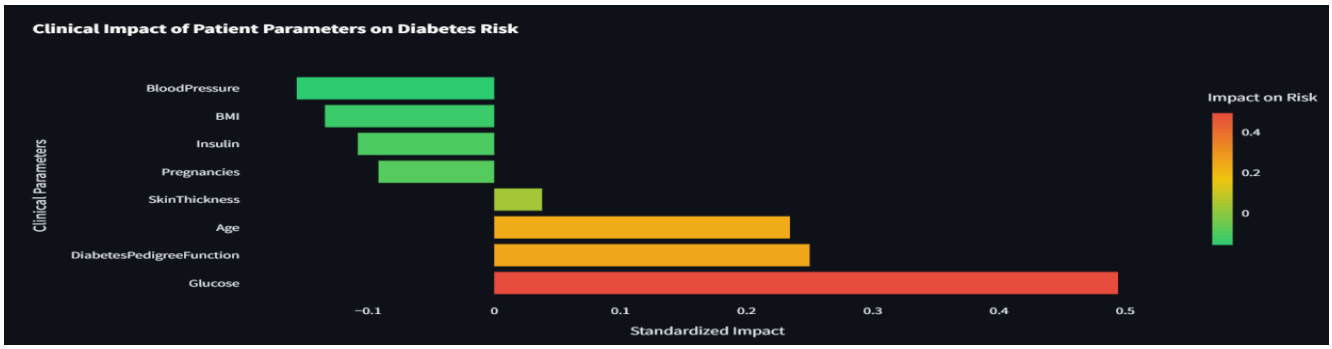
- **Pregnancies** count remains within a clinically safe range.
- **Insulin** levels are within normal limits, suggesting preserved insulin production.
- **BMI** is within an acceptable range, reducing obesity-related risk.
- **Blood Pressure** is controlled and does not contribute to additional metabolic stress.



These protective factors help **moderate** the overall risk level.

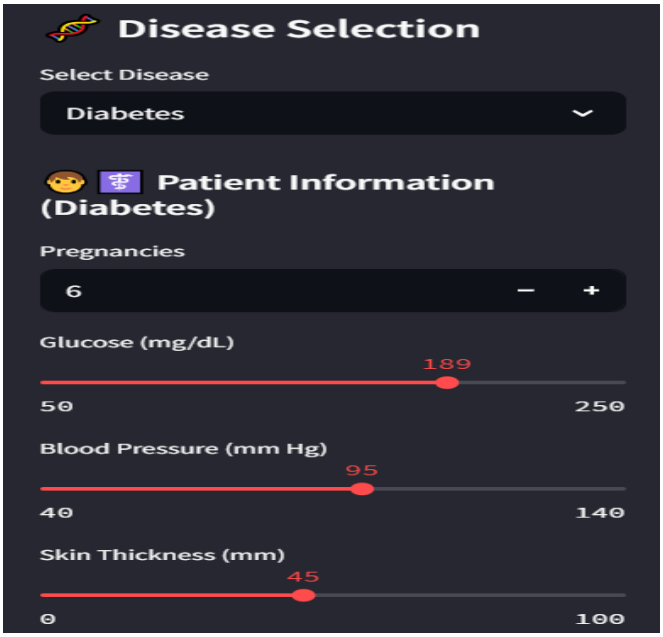
Feature Contribution Analysis

The model highlights **glucose concentration, genetic background, age, and skin thickness** as the most influential features driving the prediction toward diabetes, while BMI, insulin, and blood pressure help reduce the final risk score.



HIGH Risk Test

This test case represents a high-risk diabetes profile.



This assessment indicates an extremely high probability of diabetes.

The patient exhibits multiple severe metabolic and physiological risk factors, strongly suggesting an active diabetic condition that requires urgent medical evaluation and intervention.

Factors Increasing Diabetes Risk

Glucose levels are critically elevated, representing a major indicator of diabetes and poor glycemic control.

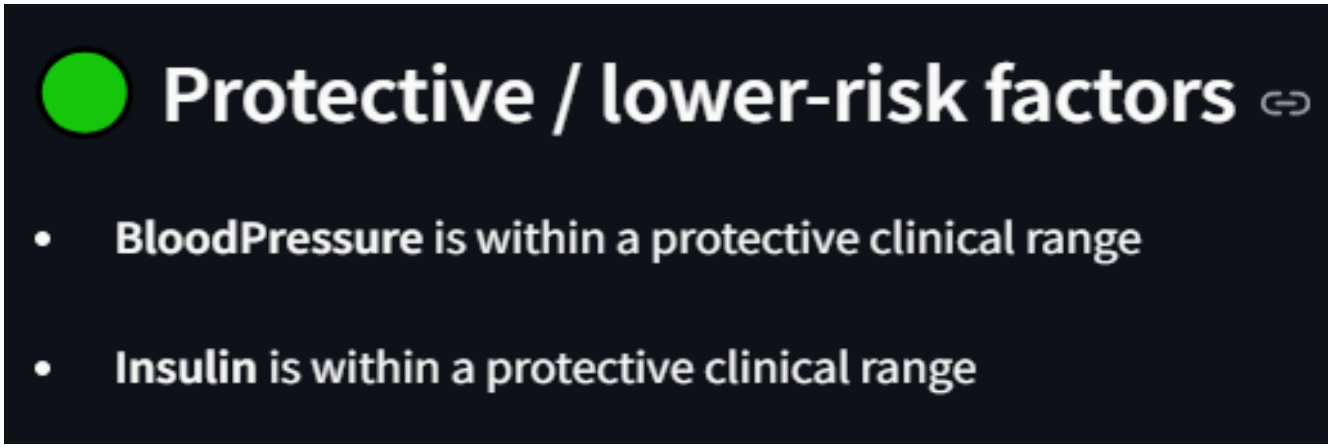
- ❖ Diabetes Pedigree Function shows a strong genetic susceptibility to diabetes.
- ❖ BMI is significantly high, reflecting obesity-related insulin resistance.
- ❖ Age increases metabolic vulnerability and reduces glucose regulation efficiency.
- ❖ Pregnancies count contributes to elevated hormonal and metabolic stress.

These combined factors overwhelmingly push the risk toward a confirmed diabetic profile.

Protective / Lower-Risk Factors

Blood Pressure remains within a clinically acceptable range.

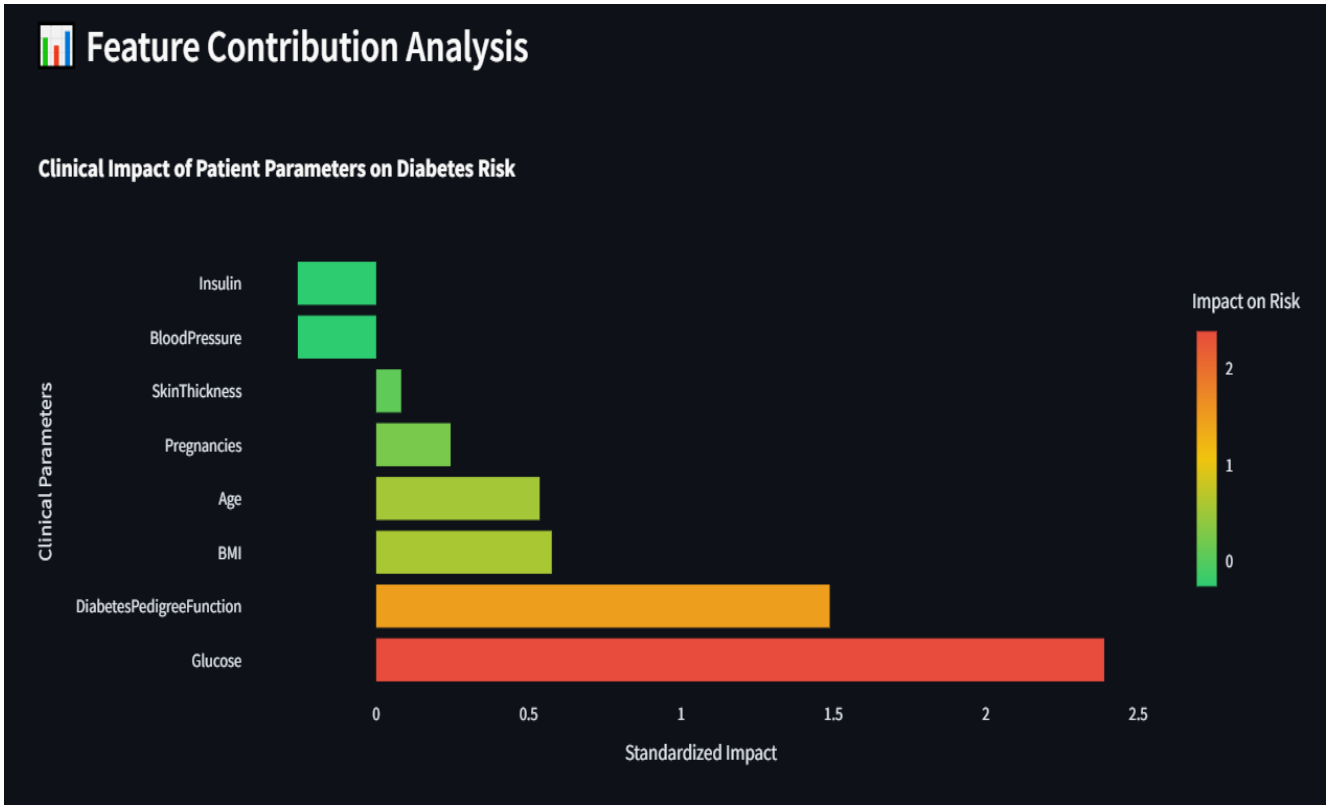
Insulin levels are still within a protective range, suggesting partial preservation of pancreatic function.



Despite these protective elements, their impact is insufficient to counterbalance the severity of the risk factors.

Feature Contribution Analysis

The prediction is primarily driven by glucose concentration, genetic predisposition, BMI, age, and pregnancy history, which dominate the model's decision and result in a near-certain diabetes classification.



Hypertension Risk Assessment Module

AI-assisted blood pressure risk prediction

This tool evaluates the risk of **hypertension (high blood pressure)** using a machine learning model trained on clinical and lifestyle-related features.

Dataset

The Hypertension module uses a clinically-inspired hypertension dataset that includes **demographic, lifestyle, and physiological features** associated with elevated blood pressure risk.

The dataset contains the following variables:

- **Age** – Patient age (*years*)
- **Salt_Intake** – Daily salt consumption (*grams/day*)
- **Stress_Score** – Psychological stress level (0–10 *scale*)
- **BP_History** – Previous history of high blood pressure (*Yes/No*)
- **Sleep_Duration** – Average sleep duration (hours/night)
- **BMI** – Body Mass Index
- **Medication** – Use of antihypertensive medication (*Yes/No*)
- **Family_History** – Family history of hypertension (*Yes/No*)
- **Exercise_Level** – Physical activity level (*Low / Moderate / High*)
- **Smoking_Status** – Smoking behavior (*Yes/No*)

🎯 **Target variable :**

- **Has_Hypertension** (*0 = No Hypertension, 1 = Hypertension*)

Risk Thresholds

Hypertension risk is estimated using a **Logistic Regression model** trained on standardized clinical-data.

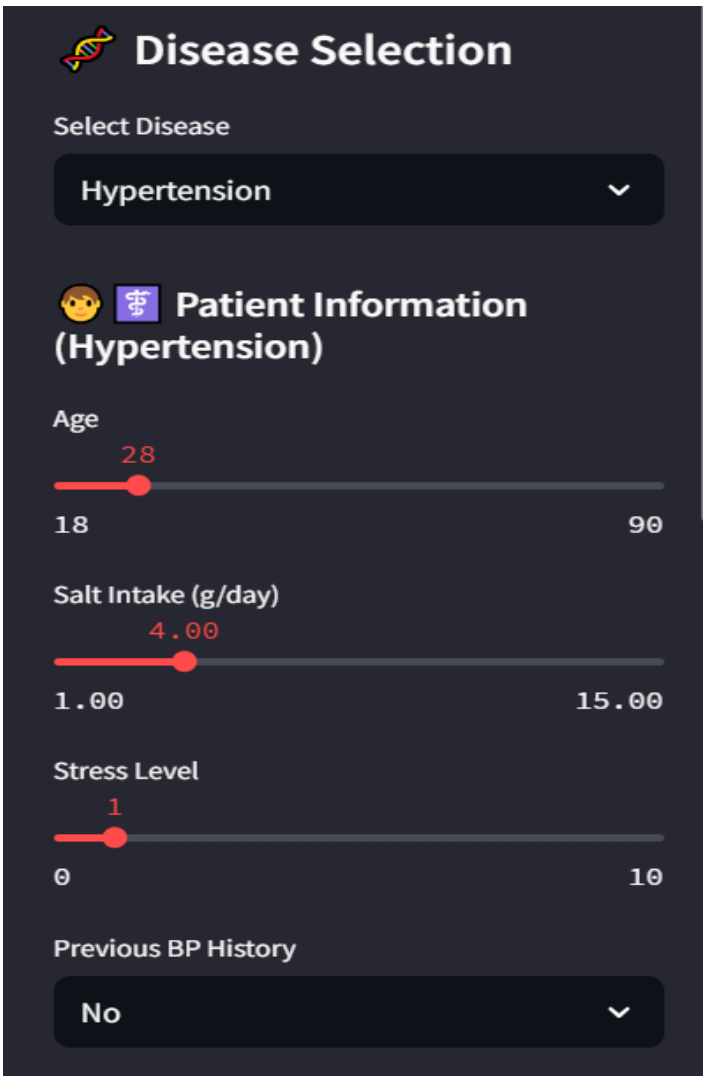
The predicted probability is mapped to three clinically meaningful risk categories:

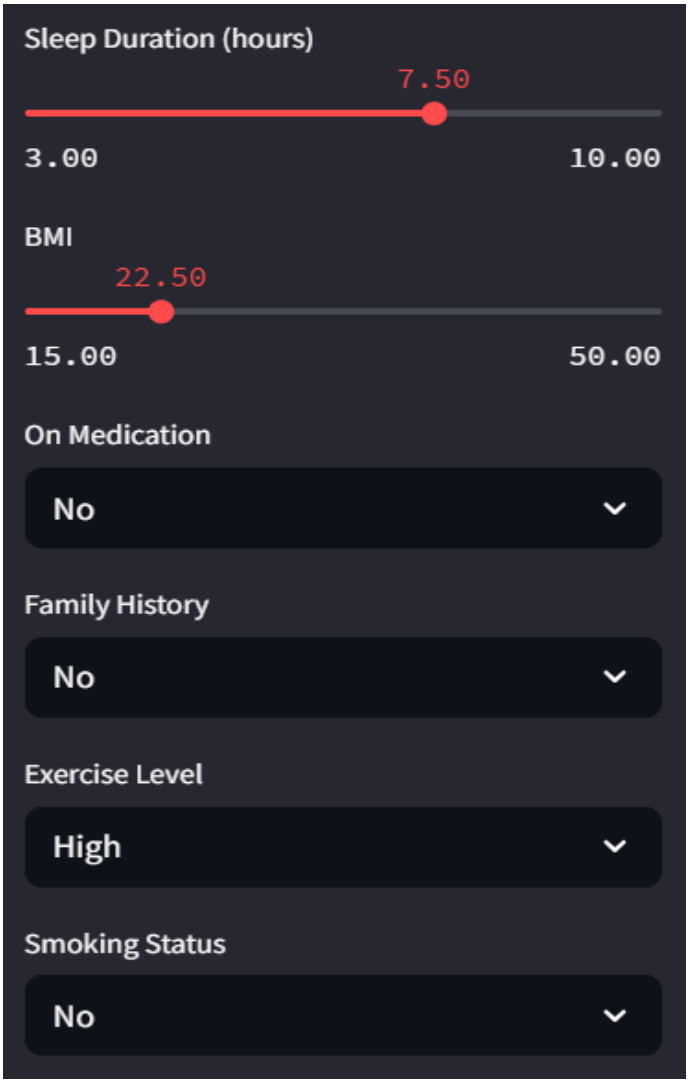
Risk Level	Probability
LOW	< 40%
MEDIUM	40% – 70%
HIGH	≥ 70%

Hypertension – Test Scenarios

LOW Risk

This test case represents a low-risk hypertension profile.





The AI-assisted system predicts no hypertension, with a very low risk probability of 4.37%, classifying the patient as LOW RISK.

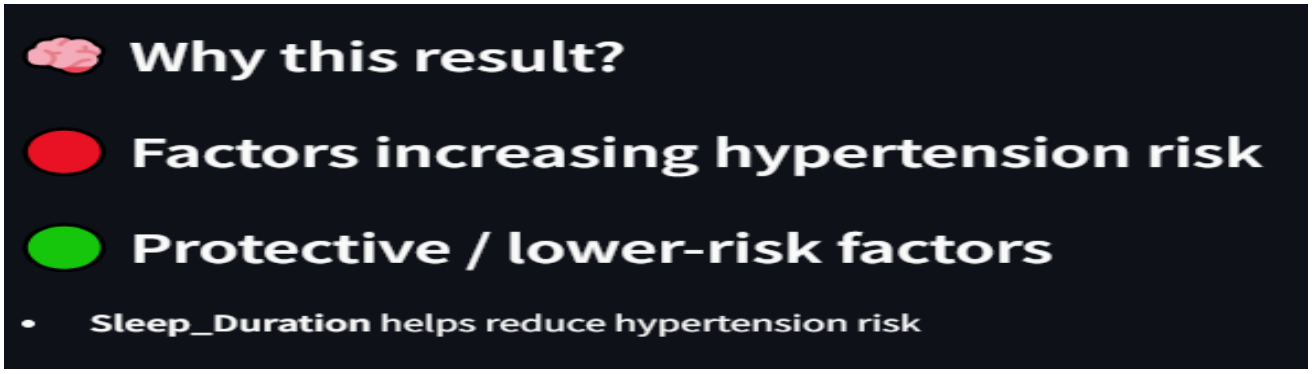
Protective / Lower-Risk Factors

Most clinical and lifestyle indicators fall within healthy and protective ranges:

- ❖ Sleep duration is adequate and contributes positively to blood pressure regulation
- ❖ No prior blood pressure history, reducing baseline cardiovascular risk
- ❖ BMI remains within a healthy range
- ❖ Stress level is controlled and does not significantly impact blood pressure
- ❖ No smoking behavior, eliminating a major hypertension risk factor
- ❖ Physical activity level supports cardiovascular health

Factors Increasing Hypertension Risk

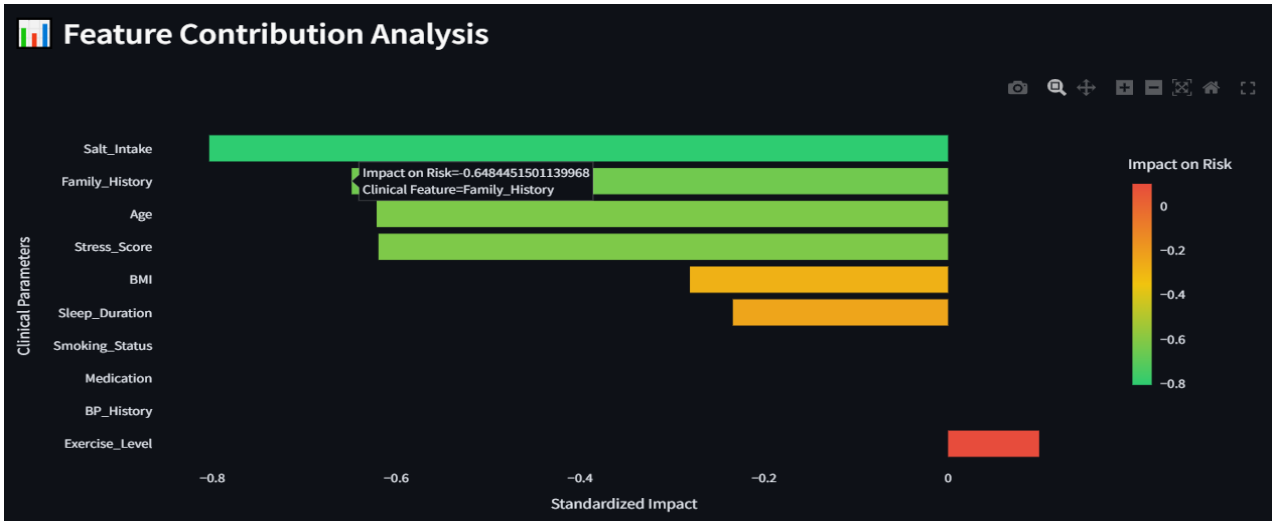
Only minimal risk-increasing factors are detected, none of which have a significant clinical impact on the final prediction.



The feature contribution analysis confirms that protective lifestyle factors dominate the model's decision, resulting in a stable and reassuring clinical outcome.

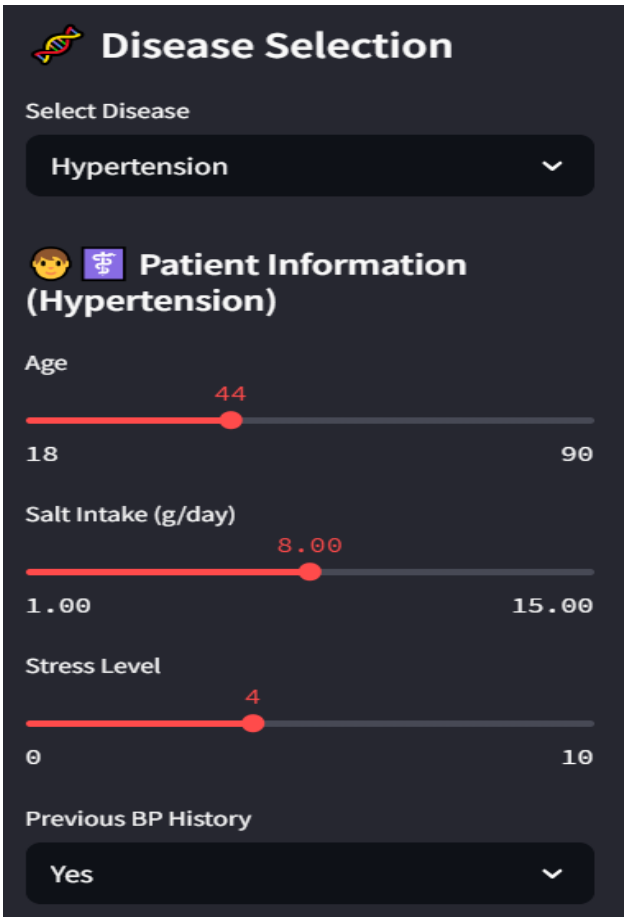
Feature Contribution Analysis

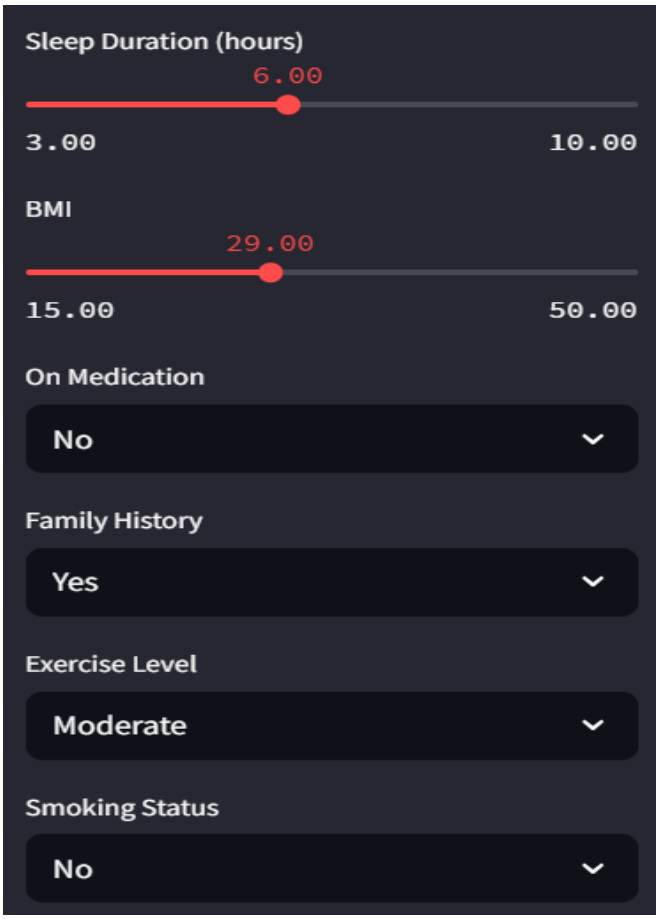
The model prediction is dominated by **protective clinical features**, leading to a confident low-risk classification.



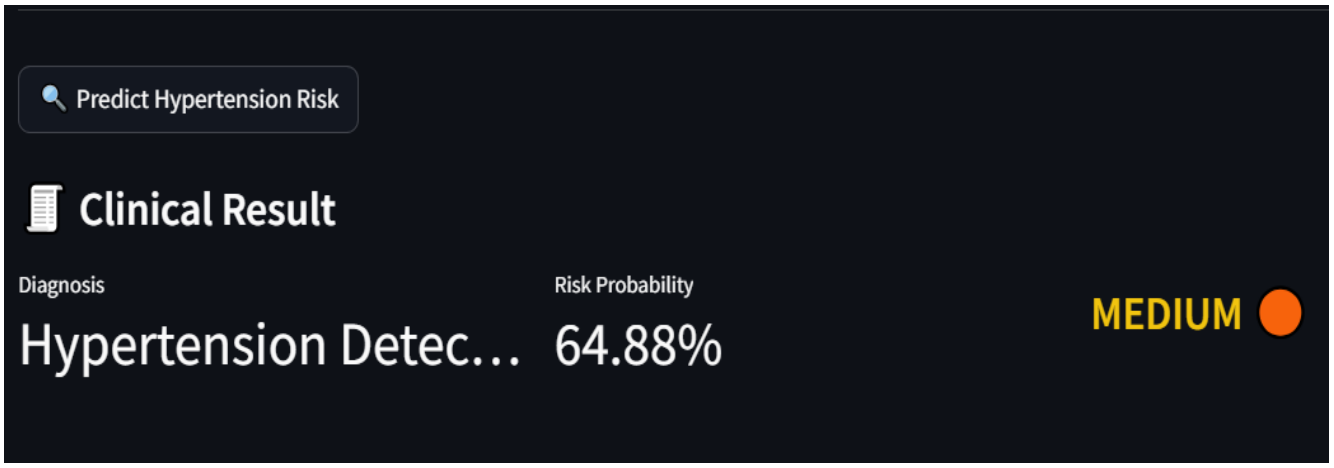
Overall, the model concludes that the patient has a very low likelihood of developing hypertension, and no immediate medical intervention is required, aside from maintaining healthy lifestyle habits.

MEDIUM Risk Test





This result indicates a **moderate risk of hypertension**, suggesting that elevated blood pressure is present but not yet at a severe stage. The condition appears to be **influenced mainly by lifestyle and genetic factors**, which may progressively increase cardiovascular strain if not addressed.



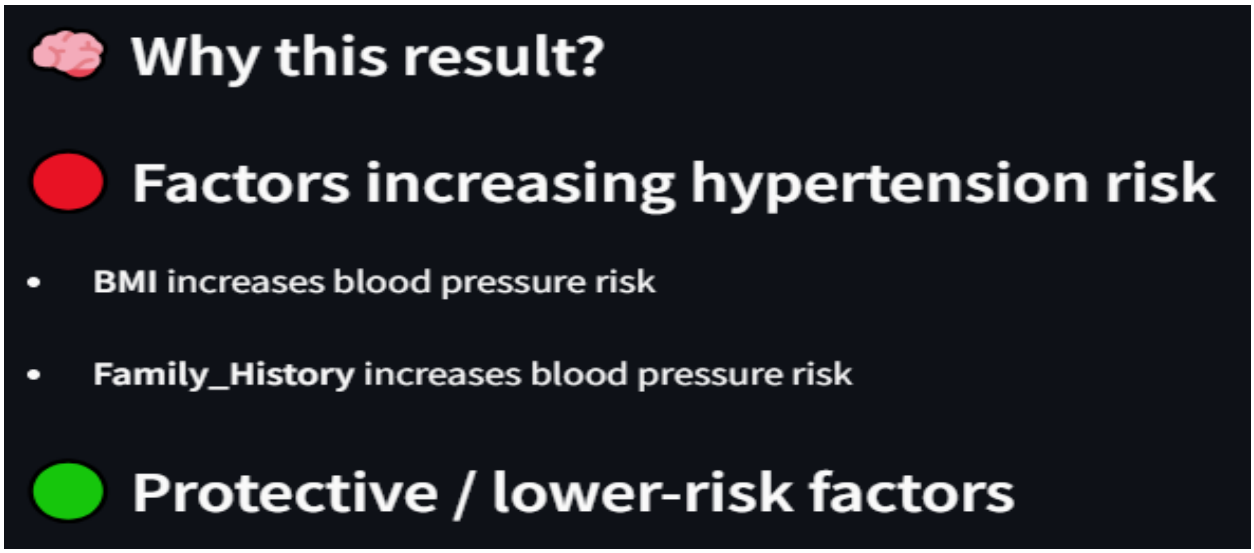
Factors Increasing Hypertension Risk

BMI increases blood pressure risk by placing additional workload on the cardiovascular system.

Family_History increases blood pressure risk, indicating a genetic predisposition to hypertension.

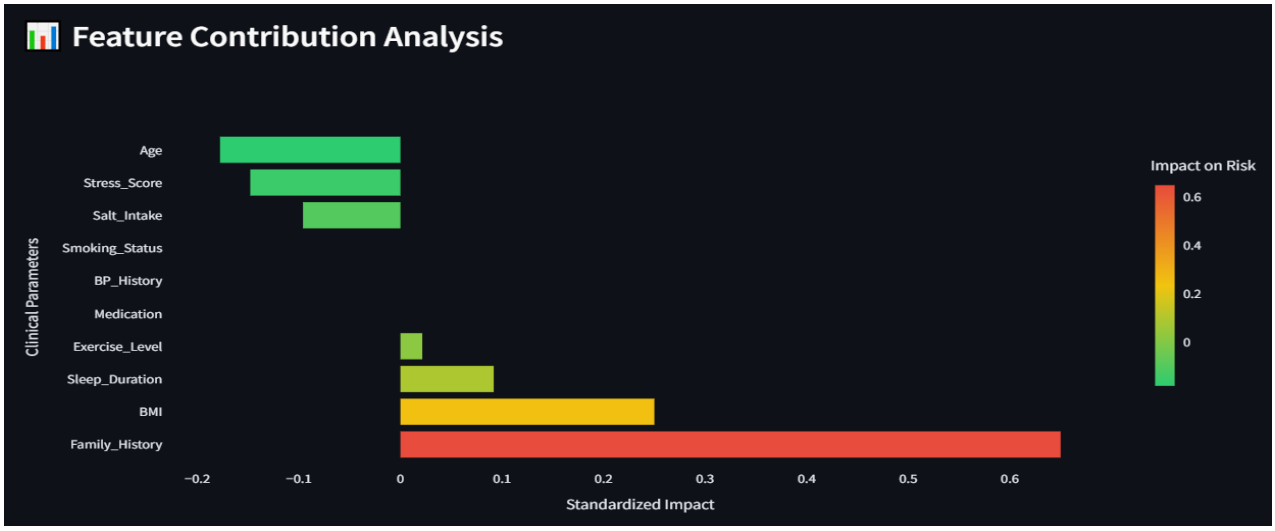
Protective / Lower-Risk Factors

Other clinical indicators remain within acceptable ranges, helping to partially regulate blood pressure levels.



Absence of severe vascular or metabolic abnormalities reduces immediate complication risk.

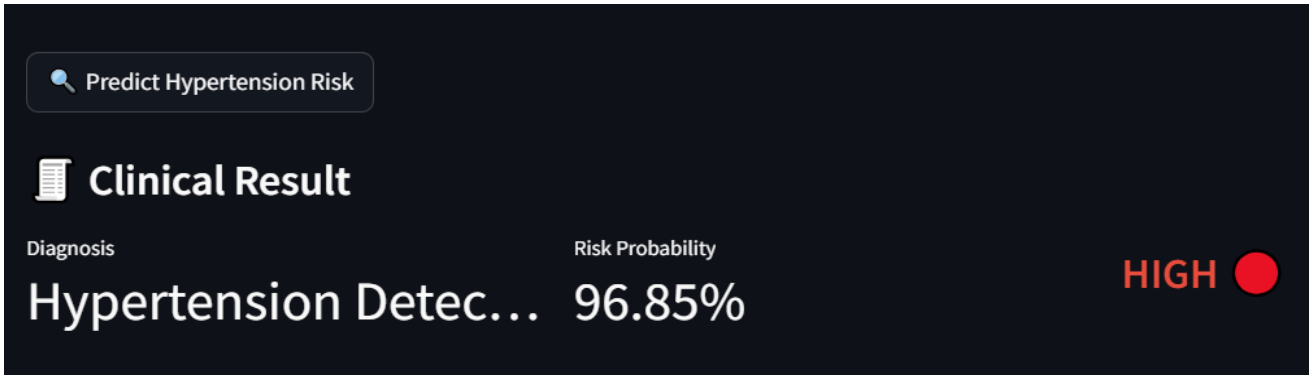
The model prediction is primarily driven by body mass index and genetic background, while remaining features act as stabilizing factors.



This balance results in a medium-risk hypertension classification, highlighting the importance of early prevention and lifestyle adjustment.

HIGH Risk Test

This result indicates a very high risk of hypertension, suggesting a severe elevation in blood pressure with a strong likelihood of cardiovascular complications.

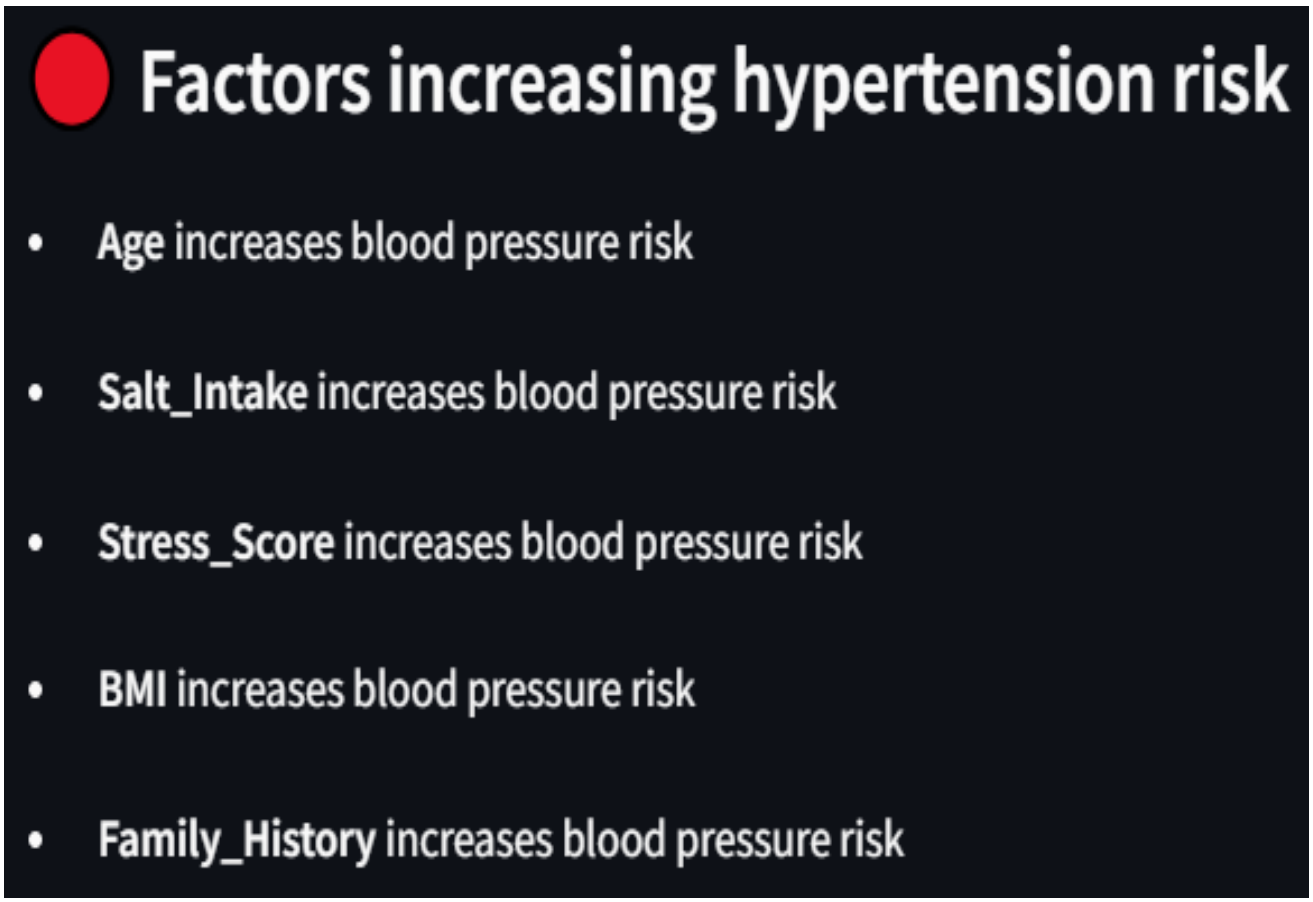


Multiple clinical, lifestyle, and genetic factors are simultaneously contributing to the risk, placing the patient in a critical health category that requires immediate medical attention.

Factors Increasing Hypertension Risk

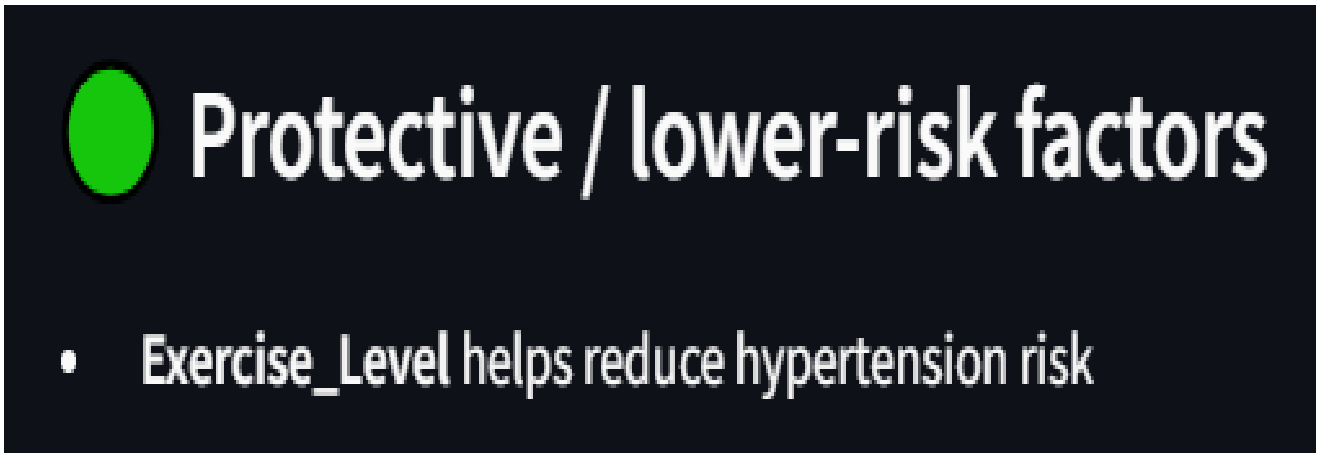
- *Age* significantly increases blood pressure risk due to reduced arterial elasticity.
- *Salt_Intake* strongly contributes to hypertension by increasing fluid retention and vascular pressure.

- *Stress_Score* elevates blood pressure through sustained hormonal and nervous system activation.
- *BMI* increases cardiovascular workload and vascular resistance.
- *Family_History* indicates a genetic predisposition to high blood pressure.



Protective / Lower-Risk Factors

Exercise_Level helps reduce hypertension risk by improving vascular health and regulating blood pressure.

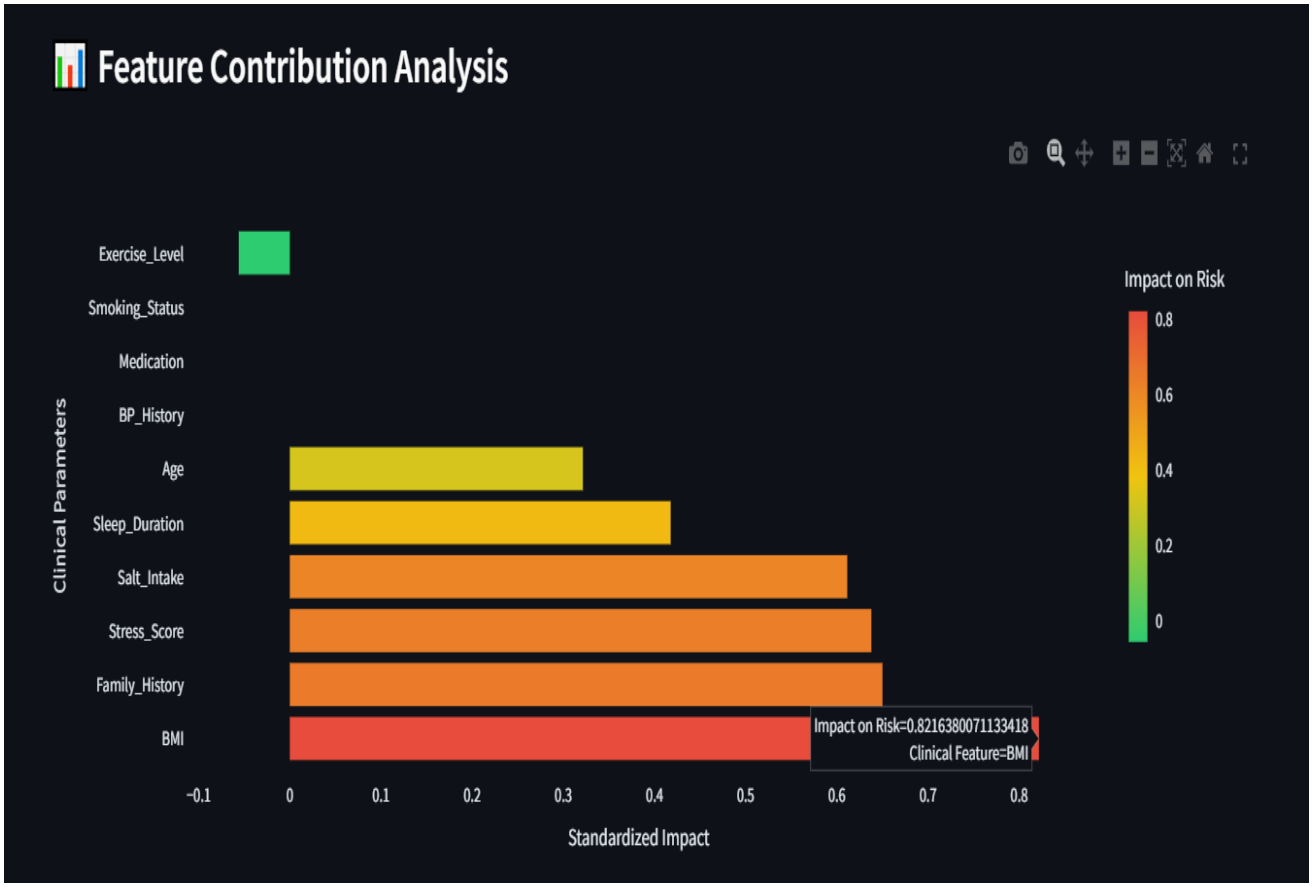


Although physical activity provides a protective effect, it is insufficient to counterbalance the accumulation of high-risk factors.

Feature Contribution Analysis

The prediction is dominated by age, salt intake, stress, BMI, and genetic history, which collectively drive the risk probability toward a high-risk classification.

Protective lifestyle factors play a secondary role and are overwhelmed by the magnitude of the contributing risk factors.



Conclusion

This project presents the design and implementation of an **AI-based medical risk assessment system** capable of predicting the risk of **Heart Disease, Diabetes, and Hypertension** using clinical and lifestyle data.

By leveraging **Machine Learning models**, particularly **Logistic Regression combined with feature scaling**, the system provides **real-time risk predictions** that are both accurate and interpretable. The integration of **Explainable Artificial Intelligence (XAI)** techniques allows users to understand not only the final prediction, but also the **individual contribution of each clinical feature** to the risk assessment.

The application's **interactive and intuitive user interface**, developed with **Streamlit**, enables healthcare students and practitioners to easily input patient data and instantly visualize results, risk levels, and feature impact analyses. This enhances trust in the model and supports informed clinical decision-making.

Experimental testing across multiple scenarios (low, medium, and high risk) for each disease demonstrates the system's **robustness, consistency, and clinical relevance**. The results confirm that combining predictive modeling with explainability significantly improves the usability and educational value of AI-driven medical tools.

Overall, this project highlights the potential of **Artificial Intelligence in healthcare**, particularly as a **decision-support system**, and lays a strong foundation for future extensions such as additional diseases, advanced models, real clinical data integration, and deployment in real-world medical environments.