PROJECT REPORT

"HealthGuard: Diabetes Prediction System"

Table of Contents

1. Introduction

The Diabetes Prediction system is an AI-based solution designed to identify the possibility of diabetes in users based on health metrics. The application integrates machine learning (Random Forest Classifier) and an expert system (Experta) for reasoning, combined with a user-friendly GUI for interaction. This system aids individuals in early risk assessment and provides actionable insights to moderate health risks.

2. Problem Statement

Diabetes is a growing global health concern that often goes undetected until significant symptoms develop. Accurate prediction tools are needed to assist individuals in identifying their risk levels early. Current solutions lack interpretability and user engagement, leaving individuals uncertain about their health decisions. This system aims to address these issues by providing accurate predictions coupled with an explanation of the results.

3. Objectives

- **Prediction Accuracy**: Leverage machine learning to deliver reliable predictions of diabetes risk.
- **Explainable Reasoning**: Integrate an expert system to provide interpretable reasoning for predictions.
- **User-Centric Design**: Create an interactive and intuitive interface for health data input and results visualization.
- **Risk Awareness**: Educate users on health factors contributing to diabetes.

4. Models and Algorithms

1. **Machine Learning Model**: ○ **Algorithm**: Random Forest Classifier.
    - **Purpose**: Predict diabetes possibility based on user input. ○ **Implementation**:
        - Feature standardization with `StandardScaler`.
        - Model trained on **Pima Indians Diabetes** dataset with 80-20 train-test split.
2. **Expert System**:
    - **Framework**: Experta (Rule-based system).
    - **Purpose**: Provide explanations for predictions by analyzing individual health metrics. ○ **Rules**: Evaluate glucose levels, BMI, age, and family history.

5. System Architecture

5.1 Components:

1. **Data Layer**:
   - ***Input:*** User-provided health metrics.
   - ***Output:*** Preprocessed data for the ML model and expert system.
2. **Processing Layer**:
   - ***Machine Learning:*** Generates a binary prediction (Positive/Negative for diabetes).
   - ***Expert System:*** Analyzes input data to provide reasoning.
3. **Presentation Layer**:
   - ***GUI (Tkinter):*** Facilitates user interaction and displays predictions with reasoning.

5.2 Workflow:

1. User inputs health metrics through the GUI.
2. Data is preprocessed and standardized.
3. Machine Learning model predicts diabetes risk.
4. Expert System analyzes inputs for risk factors.
5. Results are displayed in the GUI, including the prediction and reasoning.

6. Knowledge Representation Techniques

1. **Rule-Based Representation**:
   o Implemented using the Experta framework. o Encodes medical knowledge into actionable rules (e.g., "High glucose level indicates risk").
2. **Procedural Knowledge**: o Captures model-specific processes for feature standardization and prediction.
3. **Decision Trees** (Underlying the Random Forest Model): o Represents probabilistic relationships between features and outcomes.

7. Implementation Details

7.1 Data Handling:

- **Dataset**: Pima Indians Diabetes dataset.
- **Features**: Includes glucose, BMI, age, pregnancies, and other health indicators.
- **Preprocessing**: Standardized using `StandardScaler`.

7.2 ML Model:

- **Algorithm**: Random Forest Classifier with 100 estimators.
- **Training**: 80% of the dataset.

- **Testing**: Evaluated on 20% of the dataset.

7.3 Expert System:

- **Rules**: Defined using the `Rule` and `Fact` classes in Experta.
- **Reasoning Process**:
  - Declares facts about the user's health metrics.
  - Triggers rules based on conditions.

7.4 GUI:

- **Framework**: Tkinter.
- **Features**:
  - Dynamically adjusts input fields based on user gender.
  - Provides prediction and reasoning outputs.

8. Testing Scenarios and Results

8.1 Scenarios:

1. **Valid Inputs**: Ensure accurate predictions and reasoning for varied health profiles.
2. **Missing/Empty Inputs**: Check system response to incomplete data.
3. **Boundary Conditions**: Test with extreme values (e.g., high glucose or BMI).
4. **Gender-Specific Inputs**: Validate dynamic field adjustments for male and female users.

8.2 Results:

- **Accuracy**: Achieved over 72% accuracy on test data.
- **Reasoning**: Rules correctly identified risk factors based on user inputs.
- **GUI**: Responsive and user-friendly, with clear results and explanations.

9. Challenges and Solutions

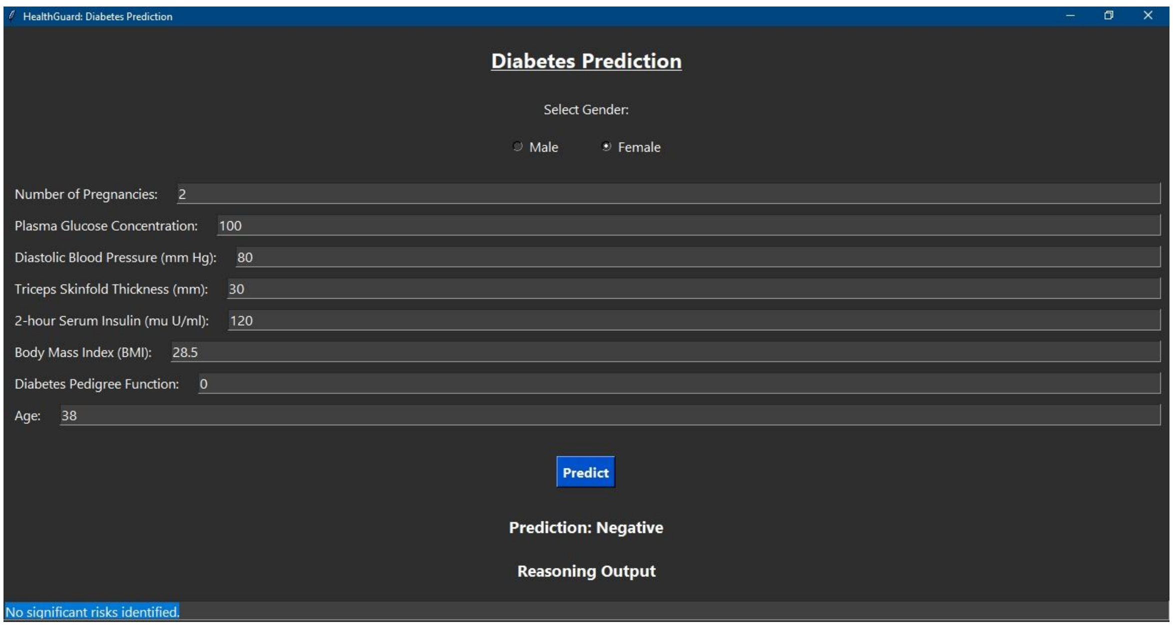
1. **Challenge**: Handling missing or invalid user inputs.
   - **Solution**: Default values and error handling in GUI (e.g., message boxes for invalid inputs).
2. **Challenge**: Aligning expert system facts with ML model inputs.
   - **Solution**: Standardized a common data pipeline for both systems.
3. **Challenge**: Complex user interactions.
   - **Solution**: Simplified GUI with dynamic field adjustments based on gender.

10. Conclusion

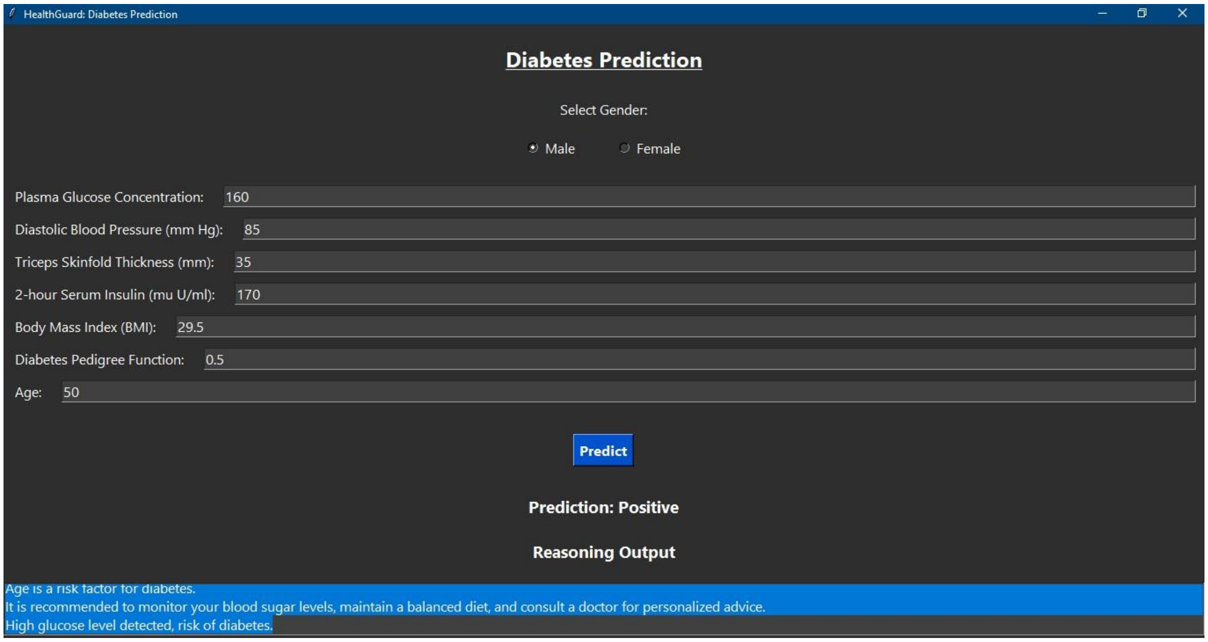
The Diabetes Prediction system successfully combines machine learning and expert knowledge to provide a complete tool for assessing diabetes risk. Its in-built GUI ensures accessibility, while explainable results enhance user understanding. This system represents a significant step toward accessible and interpretable health diagnostic, with potential for future enhancements like adding lifestyle factors and deploying on cloud platforms.

11. Output

For Female:



For Male:


HealthGuard: Diabetes Prediction
Diabetes Prediction
Select Gender:
Male
Female
Plasma Glucose Concentration: 160
Diastolic Blood Pressure (mm Hg): 85
Triceps Skinfold Thickness (mm): 35
2-hour Serum Insulin (mu U/ml): 170
Body Mass Index (BMI): 29.5
Diabetes Pedigree Function: 0.5
Age: 50
Predict
Prediction: Positive
Reasoning Output
Age is a risk factor for diabetes.
It is recommended to monitor your blood sugar levels, maintain a balanced diet, and consult a doctor for personalized advice.
High glucose level detected, risk of diabetes.