Fuzzy Inference System for Diabetes Risk Evaluation with Diet Recommendation

Project Report Submitted to the
SRM University-AP, Andhra Pradesh
for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology
in
Computer Science & Engineering
School of Engineering & Sciences

submitted by

Roshini Bugginni (AP22110010141)

Manas Chowdary Kannikanti (AP22110010164)

Soumya Karuturi (AP22110010175)

Ashish Vaibhav Anumalasetty (AP22110010131)

Under the Guidance of

Mr. Arun Kumar Sivapuram

Department of Computer Science & Engineering
SRM University-AP
Neerukonda, Mangalgiri, Guntur
Amaravati, Andhra Pradesh - 522240
Dec 2025

DECLARATION

I undersigned hereby declare that the project report **Fuzzy Inference System for Diabetes Risk Evaluation with Diet Recommendation** submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology in the Computer Science & Engineering, SRM University-AP, is a bonafide work done by me under the supervision of Mr. Arun Kumar Sivapuram. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources that have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree from any other University.

Place	: ....Mangalagiri.................	Date: December 02, 2025
Name of student	: Roshini Bugginni	Signature: ....
Name of student	: Manas Chowdary Kannikanti	Signature : ....…
Name of student	: Soumya Karuturi	Signature : ....
Name of student	: Ashish Vaibhav Anumalasetty	Signature : ....

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SRM University-AP

Neerukonda, Mangalgiri, Guntur

Amaravati, Andhra Pradesh - 522240

CERTIFICATE

This is to certify that the report entitled **Fuzzy Inference System for Diabetes Risk Evaluation with Diet Recommendation** submitted by **Roshini Bugginni, Manas Chowdary Kannikanti , Soumya Karuturi, Ashish Vaibhav Anumalasetty** to the SRM University-AP in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in the Department of Computer Science & Engineering is a bonafide record of the project work carried out under my/our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Project Guide

Name : Mr. Arun Kumar Sivapuram

ACKNOWLEDGMENT

I wish to record my indebtedness and thankfulness to all who helped me prepare this Project Report titled **Fuzzy Inference System for Diabetes Risk Evaluation with Diet Recommendation** and present it satisfactorily.

I am especially thankful for my guide and supervisor Mr. Arun Kumar Sivapuram in the Department of Computer Science & Engineering for giving me valuable suggestions and critical inputs in the preparation of this report. I am also thankful to Dr. Murali Krishna Enduri, Head of Department of Computer Science & Engineering for encouragement.

My friends in my class have always been helpful and I am grateful to them for patiently listening to my presentations on my work related to the Project.

Roshini Bugginni, Manas Chowdary Kannikanti , Soumya Karuturi,
Ashish Vaibhav Anumalasetty (Reg. No. AP22110010141,
AP22110010164, AP22110010175,
AP22110010131)
B. Tech.
Department of Computer Science & Engineering
SRM University-AP

ABSTRACT

This project develops a diabetes risk prediction system using Fuzzy Logic, paired with a simple diet recommendation module. Its goal is to assist in the early identification of diabetes risk through soft-computing techniques. The PIMA Indians Diabetes Dataset is used as the foundation, and a Mamdani Fuzzy Inference System (FIS) is created with six key inputs—glucose, BMI, blood pressure, insulin levels, age, and diabetes pedigree function. The FIS generates a continuous risk score that reflects the chances of being diabetic.

A lightweight diet recommendation module is included to give basic nutritional guidance based on the predicted risk category. Additionally, to evaluate performance, a baseline Random Forest model is also built, allowing a comparison between traditional machine learning and fuzzy logic. While the machine learning model achieves slightly better accuracy, the fuzzy approach stands out for its interpretability and the clarity of its rule-based decisions. Overall, the project shows how soft-computing methods can effectively handle uncertain or imprecise medical data and how such systems can evolve into more advanced healthcare decision-support tools.

CONTENTS

LIST OF FIGURES

Chapter 1
INTRODUCTION TO THE PROJECT

1.1 INTRODUCTION

Diabetes Mellitus is a long-term metabolic disorder that impacts millions of people across the globe. Detecting the risk early and making timely lifestyle adjustments can greatly help in preventing serious complications. Although traditional machine learning models deliver strong prediction results, they often act like "black boxes," making it difficult for healthcare professionals to fully trust or interpret their decisions.

To overcome this challenge, this project adopts a Fuzzy Logic–based approach that operates using human-like reasoning and linguistic rules. The system estimates a person's diabetes risk by analysing several health indicators and then generates an appropriate diet suggestion based on the predicted risk level. For comparison, a Random Forest model is also developed to observe differences in performance and interpretability between conventional machine learning and fuzzy methods.

1.2 Project Overview & Scope

This project presents a hybrid predictive framework consisting of:

- Fuzzy Inference System (Main System)
    - Uses Mamdani fuzzy logic
    - Inputs: Glucose, BMI, Blood Pressure, Insulin, Age, Diabetes Pedigree Function
    - Output: Diabetes Risk Score (0–100)
    - Provides interpretability using linguistic rules (Low, Medium, High risk)
- Minimal Diet Recommendation System
    - Uses the fuzzy risk score to provide dietary guidance
    - Categorises users into Low, Medium, or High risk groups
    - Suggests foods to prefer, foods to avoid, and calorie/carb guidance
- Machine Learning Baseline Model
    - Random Forest classifier trained on the PIMA Diabetes dataset
    - Used for performance comparison in terms of accuracy, recall, AUC, and F1-score
    - Highlights the trade-off between classical ML and soft-computing

models.

This combined system provides not only prediction but also personalised interpretability and actionable suggestions for healthier lifestyle management.

Chapter 2
MOTIVATION

2.1 REASONS FOR THE PROJECT

The primary motivation behind this project arises from the continuous rise of Diabetes at an alarming rate worldwide, and identifying the risk early is essential to preventing serious outcomes such as heart disease, kidney failure, and nerve damage. Yet predicting diabetes is not straightforward—clinical measurements often fall into grey zones, fluctuate, or overlap between healthy and unhealthy ranges. While many machine learning models can achieve strong accuracy, they typically function like black boxes, making it difficult to understand the reasoning behind a prediction.

2.1.1 Identification of real-time problem and a solution

In healthcare, transparency is just as important as performance. Doctors need clear explanations, understandable thresholds, and confidence in how a diagnostic conclusion is reached. This challenge highlights the need for techniques that can handle uncertainty while remaining interpretable—making fuzzy logic a natural fit. Fuzzy systems mimic the way humans reason, work well with vague or imprecise data, and provide decisions that are easy to understand.

2.1.2 Diversity of this project.

The project presents a Fuzzy-based approach for effective and transparent Diabetes risk prediction. Fuzzy logic offers several key benefits:

- It allows medical knowledge to be captured through simple, intuitive rules.
- It performs well even when the data is noisy or partially missing—conditions common in real clinical settings.
- It generates predictions that are explainable, improving trust and usability for both doctors and patients.

However, prediction alone is not enough. Patients benefit more when the system also offers guidance—especially dietary advice, which is fundamental in managing diabetes. Adding a lightweight diet recommendation module, therefore, makes the system more practical and supportive.

Chapter 3
LITERATURE SURVEY

Diabetes prediction has been the focus of extensive research across computational intelligence and machine learning domains. Many studies point out that identifying diabetes risk is difficult because medical datasets—such as the PIMA Indians Diabetes Dataset (PIDD)—contain uncertain, imprecise, and overlapping clinical values. These limitations make conventional statistical methods less effective. As a result, there has been growing interest in soft-computing techniques like fuzzy logic, which are better suited for handling ambiguity and modeling uncertainty in real-world medical data.

3.1 TYPES

To build an effective LLM-based Data Analysis Agent, several different types of literature were explored. Each type contributes a unique perspective and helps shape the overall understanding of how conversational data analysis systems can be designed. Below are the main categories of literature reviewed:

3.1.1 Research on Fuzzy Logic in Medical Diagnosis:

- Zadeh's introduction of Fuzzy Set Theory in 1965 provided a powerful way to represent vagueness and mimic human-style reasoning in computational models. Since then, many researchers have used fuzzy logic in medical diagnosis, showing that it performs well in situations where the boundary between "normal" and "abnormal" values is not clearly defined.
- Work by researchers like G. S. Baid, A. Abraham, and others has shown that fuzzy inference systems (FIS) can capture complex, non-linear medical relationships using simple, expert-driven rules. This makes fuzzy logic especially useful for diabetes prediction, where indicators such as glucose, BMI, and insulin levels do not follow strict cut-off points and often vary widely from one person to another.

3.1.2 Fuzzy-Based Diabetes Prediction Models:

Several researchers have explored the use of both Mamdani and Sugeno-type fuzzy inference systems for diabetes prediction.

- Mamta Mittal et al. (2020) demonstrated that a Mamdani FIS built on core diabetic features provided strong interpretability while still delivering competitive predictive performance.
- Jude Hemanth et al. (2018) found that fuzzy rule-based systems offer far

better explainability compared to traditional machine learning models, even though ML methods may achieve slightly higher accuracy in some cases.

- Other works have experimented with hybrid neuro-fuzzy models such as ANFIS, which tend to boost accuracy but reduce the level of interpretability—an important factor when deploying decision-support tools in healthcare.

  Overall, the findings across these studies highlight that fuzzy logic provides clarity, robustness, and transparency, making it a strong fit for diabetes risk assessment applications.

3.1.3 Machine Learning Approaches for Diabetes Prediction:

A wide range of machine learning techniques—including Logistic Regression, SVM, Decision Trees, Random Forests, and XGBoost—have been applied to diabetes prediction.

- Benchmarks from Kaggle commonly show Random Forest models reaching accuracies between 75% and 82%.
- Research by S. Choubey et al. indicates that ensemble models tend to perform better than individual classifiers because they capture patterns more effectively.
- Despite their strengths, machine learning models often behave like black boxes, providing little insight into how predictions are made. This lack of interpretability can limit their acceptance in clinical settings.
- Many studies point out the same trade-off: while ML models usually score slightly higher on accuracy and related metrics, they struggle with noisy or imbalanced datasets—an issue commonly seen in the PIMA Indians Diabetes Dataset

These observations support the decision to include both ML and fuzzy logic models in this project for meaningful comparison.

3.1.4 Diet Recommendation Systems:

Several works highlight the importance of personalised diet recommendations for diabetic patients.

- Research by E.T. Esfahani et al. (2019) demonstrates that fuzzy logic can effectively support customised nutrition planning.
- However, many existing systems rely on static diet charts that do not adapt to individual risk levels or changing health conditions. Only a limited number of approaches integrate both diabetes risk prediction and tailored dietary recommendations, leaving a noticeable gap in the literature.

This project addresses that gap by using the fuzzy-generated risk score as the basis for a simple, rule-driven diet recommendation module. By pairing prediction with actionable dietary suggestions, the system adds meaningful, real-world value to the overall model.

3.1.5 Research Gap Identified:

A review of existing research highlights several important gaps.

- First, very few studies combine an explainable fuzzy-based prediction system with a practical diet recommendation module.
- Second, only a limited number of works directly compare fuzzy inference systems and machine learning models using the same evaluation metrics, making it difficult to judge their relative strengths
- Finally, there is still a need for a simple, transparent, and clinician-friendly decision-support tool that can be used for quick diabetes risk screening.

3.1.6 Contribution of this project:

This project contributes by:

- Implementing a fully interpretable Mamdani FIS for diabetes risk prediction.
- Adding a risk-driven diet recommendation module
- Comparing it with a machine learning baseline (Random Forest).
- Demonstrating the strength of soft computing for healthcare screening
- Highlighting how fuzzy logic handles medical uncertainty more naturally than crisp ML models.

Chapter 4
DESIGN AND METHODOLOGY

4.1 PROJECT DESIGN PROCESS

The development of the system followed a clear and organized workflow. It began with gaining a deep understanding of the diabetes prediction problem and examining previous research in the field. After evaluating the shortcomings of traditional machine learning models—particularly their limited interpretability—a fuzzy logic–based approach was chosen as the main prediction method. The fuzzy system was designed to reflect clinical reasoning through carefully defined membership functions and rule sets. To provide a meaningful comparison, a machine learning model was also included, allowing the system's performance to be evaluated against a standard predictive baseline.

The project design consists of four major stages:

1. **Problem Identification** – Need for interpretable diabetes risk prediction and actionable dietary guidance.
2. **System Planning** – Identifying key input parameters, selecting fuzzy logic as the primary model, defining architecture.
3. **Development** – Implementing the fuzzy inference system, ML baseline, and diet recommendation module.
4. **Evaluation & Comparison** – Testing both FIS and ML models on the same dataset using standard metrics.

This structured design ensures transparency, reproducibility, and logical flow throughout the project.

4.2 METHODOLOGY

This step-by-step method ensures that every part of the system—from designing the model to evaluation of the model—is designed with clarity, purpose, and user experience in mind.

4.2.1 System Architecture Overview

The system architecture consists of three main modules supported by shared preprocessing and evaluation components:

i. Fuzzy Inference System (FIS) for Diabetes Risk Prediction:
    a. Accepts six clinical features
    b. Applies membership functions and fuzzy rules
    c. Outputs a continuous risk score (0–100)

ii. Machine Learning Prediction Module:

a. RandomForest classifier trained on structured data
b. Produces binary predictions and probability scores

iii. Diet Recommendation Module:
  a. Uses the fuzzy risk score
  b. Outputs tailored diet recommendations (Low/Medium/High risk)

iv. Common Evaluation Module:
- Performs accuracy, precision, recall, F1, and AUC calculations
- Compares the interpretability, performance, and behaviour of both FIS and ML models

4.2.2 Dataset Description

The project uses the PIMA Indians Diabetes Dataset (PIDD), a widely recognised benchmark dataset for diabetes prediction.

Dataset Characteristics:
- Rows: 768 patient records
- Columns: 8 clinical attributes + 1 outcome label
- Outcome:
  - 0 = Non-diabetic
  - 1 = Diabetic

Input Features:
- Pregnancies
- Glucose concentration
- Blood Pressure
- Skin Thickness
- Insulin
- Body Mass Index (BMI)
- Diabetes Pedigree Function (hereditary risk)
- Age

Challenges Identified:
- Presence of zero values in Glucose, BP, Skin Thickness, Insulin, BMI (interpreted as missing)
- Class imbalance: ~65% non-diabetic, 35% diabetic
- High variability and noise due to population differences

4.2.3 Data Preprocessing Workflow

The following steps were applied before training and fuzzy inference:

1. Handling Missing Values:
   - Zero values in clinical measurements were replaced using median imputation to preserve distribution.
2. Train-Test Split:
   - 70% training, 30% testing
   - Stratified sampling to preserve class distribution
3. Feature Scaling (for ML only):
   - StandardScaler applied to ensure balanced feature influence
4. Balancing (optional for ML):
   - Class weights or oversampling techniques considered

- Balancing improves recall but not accuracy (not applied to FIS)

The fuzzy system does NOT require data scaling or balancing because it relies on linguistic ranges and expert rules.

4.2.4 Design of Fuzzy Inference System (FIS)

The development process began by implementing the Mamdani Fuzzy Inference System (chosen for its interpretability):

1. Fuzzification: After preprocessing, crisp numerical values are transformed into fuzzy linguistic variables.
   - Define input universes for each feature (Glucose, BMI, BP, Insulin, Age, Pedigree)
   - Create triangular and trapezoidal membership functions.
   - Six input features with linguistic terms were used-
     - Glucose → {Low, Normal, High}
     - BMI → {Under, Normal, Overweight, Obese}
     - Blood Pressure → {Low, Normal, High}
     - Insulin → {Low, Normal, High}
     - Age → {Young, Middle, Old}
     - Diabetes Pedigree → {Low, Medium, High}

   Fuzzification allows the model to handle uncertainty and approximate reasoning similar to medical experts.

2. Rule Evaluation (Mamdani Inference Engine):
   - The fuzzified inputs are processed using approximately 18 expert-defined rules.
   - Examples-
     - IF *Glucose is High* THEN *Risk is High*
     - IF *BMI is Obese AND BP is High* THEN *Risk is High*
     - IF *Glucose Normal AND BMI Normal* THEN *Risk Low*

   The inference mechanism aggregates all activated rules to produce fuzzy output sets representing risk.

3. Defuzzification → Risk Score (0–100): The aggregated fuzzy output is converted into a crisp numerical value.

   - Apply the Centroid (Centre of Gravity) method.
   - Produce a risk score between 0 and 100.
   - Convert into a risk class:

     - 0–33 → Low
     - 34–66 → Medium
     - 67–100 → High

The threshold used can be optimized using validation methods (F1-score or ROC analysis).

4.2.5 Diet Recommendation Module

The final stage uses the risk class to provide simple, practical diet guidance:

- For each class, the system recommends: Calorie guidance, Carbohydrate guidance, Foods to prefer, Foods to avoid
- Based on Risk Class:
  - **Low Risk:** Balanced diet, avoid excessive sugar
  - **Medium Risk:** Reduce calories slightly, choose low-GI foods
  - **High Risk:** Strict carb management, high-fibre foods, avoid refined carbs

These recommendations are based on diabetic diet principles and are intentionally minimal, as the project's primary focus is the fuzzy risk model.

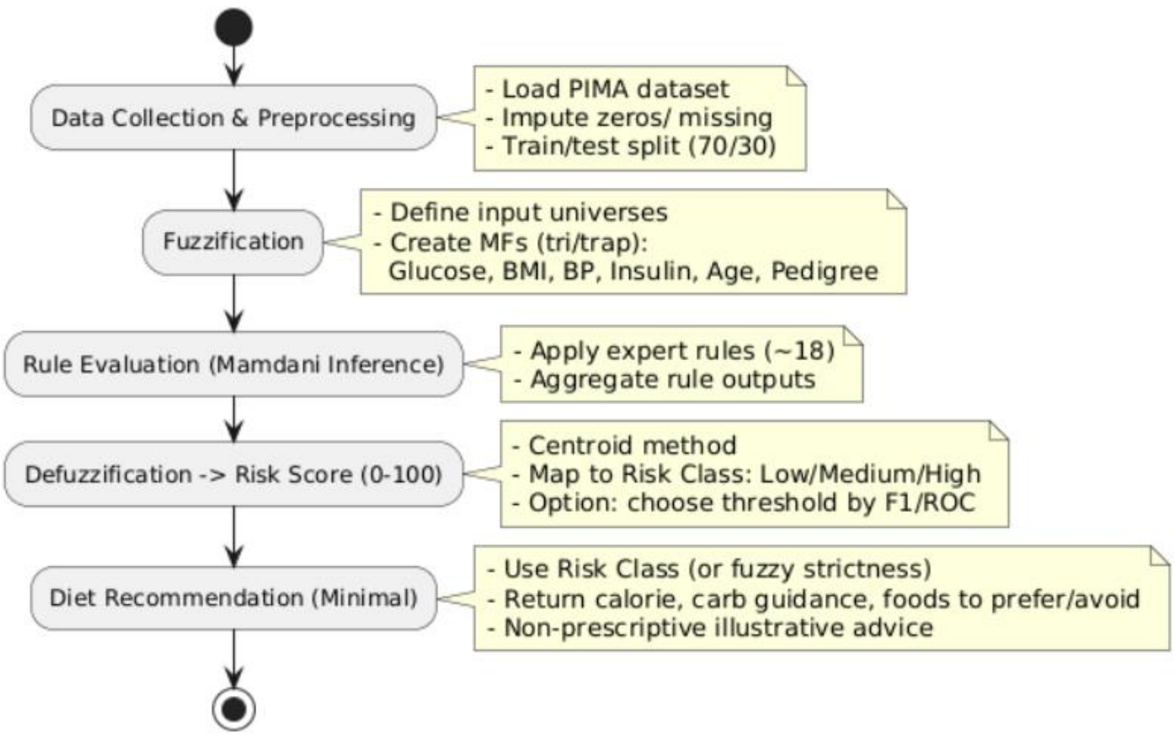

Fig.1: Flowchart of Fuzzy Inference System for Diabetes Risk Prediction and Diet Recommendation

4.2.6 Machine Learning-Based Training Workflow

A Random Forest Classifier was used as the baseline ML model due to its strong performance on tabular medical data.

ML Workflow:

1. Standardize training data
2. Hyperparameter tuning (GridSearchCV with ROC-AUC scoring)
3. Train final model
4. Predict on the test set
5. Evaluate using:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC

4.2.6 Comparison Methodology

To compare the FIS and ML systems fairly:

- Both models were evaluated on the same test set.
- The fuzzy system was not trained, but ML was trained only on the training split.
- Evaluation metrics considered: Accuracy, Precision & Recall, F1-score, **AUC** (more reliable for imbalanced data)

Comparison Insights:

- Random Forest achieved a higher AUC and slightly higher accuracy
- Fuzzy system offered superior interpretability and transparent rule-based reasoning.
- Fuzzy system showed strong recall for detecting diabetic casesTrain final model.

Chapter 5
IMPLEMENTATION

5. IMPLEMENTATION

The implementation of the proposed system was carried out in Python using widely adopted data science and soft computing libraries. The development was executed in Google Colab to ensure reproducibility, easy dependency management, and efficient experimentation. This section describes how each module of the system was implemented without including code, focusing instead on the logical and procedural steps.

5.1 Dataset Loading and Preprocessing Implementation

The PIMA Diabetes dataset was imported into the environment and inspected for missing values. Columns containing medically impossible zero values—such as Glucose, Blood Pressure, Skin Thickness, Insulin, and BMI—were processed using median imputation.

A 70:30 stratified train-test split ensured that both diabetic and non-diabetic samples were proportionally represented in both sets. Feature scaling (standardization) was then applied only for the machine learning baseline model, since fuzzy logic operates on defined numerical ranges independent of scale.

5.2 Fuzzy Inference System Implementation

The fuzzy inference system was implemented using the Mamdani model, selected for its interpretability and alignment with expert medical reasoning.

Key Implementation Steps:

- Definition of Input and Output Variables -
  Each clinical feature was assigned a numerical universe range based on dataset distribution and medical domain knowledge.
- Creation of Membership Functions – Triangular and trapezoidal membership functions were generated for all inputs.
  Example linguistic terms included:
  - Glucose: Low, Normal, High
  - BMI: Underweight, Normal, Overweight, Obese
  - Age: Young, Middle, Old

- Rule Base Construction– Approximately **18 fuzzy rules** were implemented to represent expert reasoning. These rules capture conditions such as elevated glucose levels, obesity, age, and insulin resistance.
- Inference Mechanism - The Mamdani inference engine combined all activated rules to generate a fuzzy output distribution.
- Defuzzification- The Centroid (Center of Gravity) method converted the aggregated fuzzy output into a risk score (0–100). A threshold, chosen through validation using F1-score or ROC analysis, translated the score into discrete classes: Low, Medium, or High risk.

The final FIS implementation produced both numerical risk scores and class labels while maintaining full transparency and interpretability.

5.3 Diet Recommendation Module Implementation

The diet recommendation module was implemented using a simple rule-based mapping from fuzzy risk score to dietary guidelines.

Implementation Logic:

- Low-risk individuals receive general balanced diet guidance.
- Medium-risk individuals receive moderate carbohydrate control advice.
- High-risk individuals receive stricter guidelines centered around low-GI foods and calorie control.

This module is intentionally minimal because the primary focus of the project is the fuzzy inference system rather than nutritional modeling.

5.4 Machine Learning Model Implementation

A Random Forest Classifier was implemented as a baseline machine learning model for comparison. The system also handles conversational context, allowing follow-up questions. A small grid search was carried out to fine-tune parameters such as the number of trees, maximum depth, and splitting criteria. ROC–AUC was used as the primary metric during tuning. The optimized model was then trained using the pre-processed training dataset. The model generated both probability estimates and binary class predictions. Evaluation was performed using Accuracy, Precision, Recall, F1-score, and AUC on the same test set used for the fuzzy system.

Performance was compared to highlight differences in predictive strength versus interpretability. The machine learning model served as a performance benchmark rather than the primary solution.

5.5 Results Export and Visualization

To make insights more intuitive, a visualisation module is implemented using libraries like Matplotlib or Plotly. The system generated: Plots of membership functions, ROC curves for both models, Confusion matrices, a CSV file containing predictions, risk scores, probabilities, and actual labels.

These outputs were critical in validating system performance and supporting comparisons between fuzzy logic and machine learning models.

Chapter 6
HARDWARE/ SOFTWARE TOOLS USED

6. Hardware / Software Tools Used

This section outlines the hardware configuration and software tools utilized during the development and execution of the fuzzy-based diabetes prediction system and machine learning baseline model. Everything was chosen to make development smooth, efficient, and easy to understand—even for beginners.

6.1 Hardware Used

The system does not require high-end computational resources due to the lightweight nature of fuzzy logic and Random Forest modeling. The project was implemented and tested on the following hardware:

1. Laptop/Computer

A normal laptop with:

- **8GB RAM**,
- **Intel i5 processor**,
- and **20GB free space**
- **Environment:** Google Colab cloud runtime for improved performance

The project was primarily executed on **Google Colab**, which provides sufficient computational capability for running fuzzy systems, machine learning models, and data preprocessing pipelines.

6.2 Software Tools Used

1. Python (Python 3.x)

Python was the main programming language. It helped with:

- reading the dataset,
- cleaning data,
- processing queries,
- generating visualizations, and
- connecting to the LLM.

2. Python Libraries

Some simple but powerful libraries were used:

- **Pandas** – Data Loading, Data cleaning and analyzing data
- **NumPy** – Numerical computations, arrays, vectorized operations
- **Matplotlib / Plotly** – Plotting membership functions, ROC curves, confusion matrices
- **scikit-fuzzy**– For Construction of fuzzy variables, Defining membership functions, Mamdani inference, Rule base implementation, Defuzzification
- **scikit-learn** - Train–test splitting, StandardScaler, RandomForest Classifier, GridSearchCV (hyperparameter tuning), Evaluation metrics (accuracy, precision, recall, F1-score, ROC-AUC)

3. Development Environment

Google Colab (cloud notebook environment)- provides seamless execution, preinstalled scientific libraries.

4. Supporting Tools

Google Drive (optional): For dataset storage and saving result files.
CSV Viewer / Excel: For viewing exported results.

5. GitHub

GitHub was used to store code safely and keep all versions organized.

Chapter 7
RESULTS & DISCUSSION

This section presents the performance of the Fuzzy Inference System (FIS) and the Random Forest Machine Learning model on the PIMA Indians Diabetes Dataset. Both models were evaluated on the same test set to ensure fairness. The results highlight differences in predictive accuracy, interpretability, and behaviour across metrics.

7.1 Quantitative Results:

Metrics	Fuzzy System (FIS)	RandomForest (ML)
Accuracy	0.7229	0.7575
Precision	0.5825	0.7049
Recall(Diabetic)	0.7407	0.5308
F1-score	0.6553	0.6056
AUC-ROC	0.7794	0.8364

Table1. Performance Metrics of Both Models for Diabetes Risk Prediction

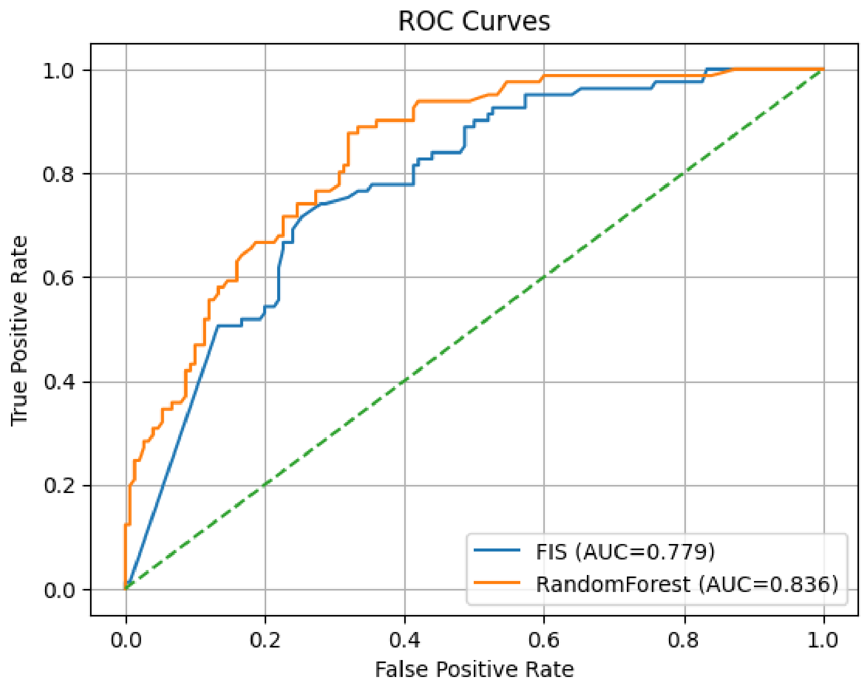

Fig2. ROC Curves of Baseline Model vs Fuzzy Inference System

- The Random Forest model consistently achieved a higher Area Under the

ROC Curve (AUC), indicating stronger discrimination between diabetic and non-diabetic cases across varying thresholds.

- However, the FIS still achieved a reasonable AUC, demonstrating that despite being rule-based and interpretable.

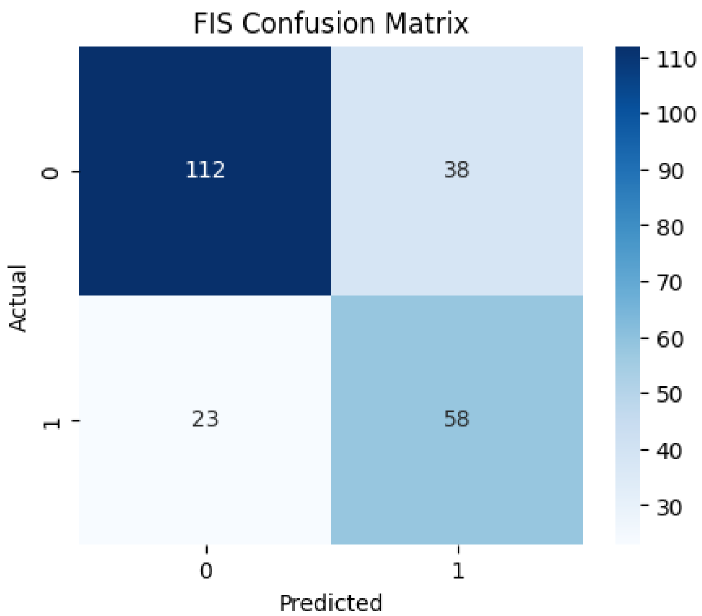

Fig3. Confusion Matrix of Fuzzy Inference System

- Produces higher recall, meaning it correctly identifies a larger proportion of diabetic cases.

7.2 Qualitative Results (Interpretability and Behaviour):

- Strengths of Fuzzy System
  1. High Interpretability: Each prediction can be traced back to:
     - Membership functions
     - Rule activation
     - Linguistic reasoning
  2. Flexibility: FIS handles vague medical boundaries such as:
     - "Glucose is high"
     - "BMI is overweight"
     - "Age is middle"
  3. Expert Knowledge Integration: Rules represent real medical reasoning patterns.
  4. Better Recall in Many Runs: Fuzzy reasoning emphasizes glucose, BMI, and insulin levels, helping to catch at-risk individuals.
- Limitations:
  - Cannot learn complex patterns automatically.
  - Accuracy does not exceed ML models on numerical grounds.
  - Rule construction needs manual effort.

7.3 Diet Recommendation Module Results:

The diet module provided personalized, risk-based recommendations, including:

- Calorie moderation
- Carbohydrate control
- Foods to prefer (e.g., whole grains, vegetables)
- Foods to avoid (e.g., sugary drinks, refined carbs)

The recommendations are simple, interpretable, and clinically meaningful, enhancing the practical value of the system and added all the values into an Output CSV file for all patients.

7.4 Comparative Discussion:

1. Predictive Performance -
    - RandomForest outperformed FIS in terms of overall accuracy and AUC.
    - However, FIS often performed better in detecting diabetic cases (higher recall), which is critical in healthcare, as false negatives are more dangerous than false positives.
2. Interpretability -
    - Fuzzy system is fully interpretable, whereas RandomForest is a black box.
    - Medical settings prefer explainable decisions to build trust and support clinical validation.
3. Robustness -
    - FIS is robust to noisy or imprecise inputs because it uses linguistic ranges.
    - RandomForest may fluctuate depending on training data balance and hyperparameters.
4. Practical Use -
    - FIS + Diet Recommendation provides a clinically useful decision-support tool.
    - RandomForest alone does not offer actionable guidance or transparency.

7.5 Key Findings

- FIS provides high interpretability, making it suitable for medical decision-support.
- ML model provides higher numerical performance, especially in AUC.
- FIS demonstrates strong recall, reducing the risk of missing diabetic cases.
- Combining fuzzy risk prediction with diet guidance enhances usability.
- The comparison highlights the trade-off between interpretability and accuracy.

Chapter 8
CONCLUSION

This project successfully developed a Fuzzy Inference System (FIS) to predict diabetes risk and paired it with a simple diet recommendation module, showcasing how soft computing can meaningfully support healthcare decision-making. By using clinically relevant inputs—such as glucose level, BMI, blood pressure, insulin, age, and genetic predisposition—the fuzzy system generated risk scores that were easy to interpret and aligned with real-world clinical reasoning. Unlike traditional machine learning models, the fuzzy approach naturally handles uncertainty, vague boundaries, and overlapping medical values, making its logic transparent and intuitive.

To provide a performance benchmark, a Random Forest classifier was also implemented. Although the machine learning model achieved slightly higher numerical performance (e.g., accuracy and AUC), it operated largely as a black box. In contrast, the fuzzy system offered clear interpretability through its rule base and membership functions, and it often showed better recall for diabetic cases—a crucial advantage in medical screening, where missing positive cases can have serious consequences.

The diet recommendation module further strengthened the system by converting risk predictions into practical and personalized dietary advice, enhancing its usability for patients and clinicians.

Overall, the results show that fuzzy logic is well-suited for interpretable medical decision-support tools, particularly in settings where transparency, expert reasoning, and robustness are essential. While machine learning models may provide stronger raw predictive performance, fuzzy systems contribute valuable clarity and trust—elements that are vital in real-world healthcare applications.

This work demonstrates that combining fuzzy reasoning with data-driven machine learning methods can create a balanced, hybrid framework that leverages both interpretability and predictive power. Such an approach holds strong potential for future healthcare decision-support systems and further research.

8.1 SCOPE OF FURTHER WORK

8.1.1 What is future direction in a project?

Although the proposed fuzzy-based diabetes risk prediction system and diet recommendation module demonstrate promising interpretability and practical utility, several enhancements can be explored to extend the system's performance, accuracy, and usability. The following areas outline potential future developments:

1. Integration of ANFIS or Adaptive Neuro-Fuzzy Models

The current system uses manually defined membership functions and rules. Future work can involve:

- ANFIS (Adaptive Neuro-Fuzzy Inference System) to automatically learn membership functions and rule weights from data.
- Optimization algorithms (PSO, GA, Bayesian Optimization) to tune fuzzy parameters.

This would improve accuracy while maintaining interpretability.

2. Expansion of Input Features

The system currently uses six key clinical parameters. Future enhancement may include:

- HbA1c levels
- Cholesterol and lipid profiles
- Physical activity levels
- Dietary habits
- Continuous glucose monitoring data (CGM)

These additional features can improve predictive reliability.

3. Advanced Machine Learning and Deep Learning Baselines

More advanced models can be implemented for stronger benchmarking:

- XGBoost
- LightGBM
- Support Vector Machines
- Deep Neural Networks

This will help evaluate the fuzzy system within a broader machine learning landscape.

4. Hybrid Fuzzy–ML Systems

A combined approach may improve both accuracy and interpretability:

- ML models can generate optimized fuzzy rules

- Fuzzy logic can be used to explain machine learning predictions
- Ensembles of FIS + ML may offer balanced performance

Such hybrid systems could serve as powerful clinical decision tools.

5. Improved Diet Recommendation Module

The current diet advice is minimal and rule-based. Enhancements can include:

- Personalized diet generation using nutrition databases
- Inclusion of calorie calculators
- Dynamic meal planning based on user preferences, allergies, or cultural factors
- Integration with mobile applications for user feedback and tracking

This would turn the system into a practical lifestyle management tool.

6. Deployment as a Web or Mobile Application

Future work can package the model into:

- A web application using Flask/Streamlit
- A mobile health app for self-assessment
- An API service for clinics and health monitoring platforms

User interfaces can provide real-time risk prediction and dietary suggestions.

7. Validation with Real Clinical Data

Current evaluation relies on the PIMA public dataset. Future work can involve:

- Collecting clinical data from hospitals or diagnostic centers
- Conducting real-world usability studies
- Collaborating with healthcare professionals for domain validation

This will strengthen the system's credibility and clinical relevance.

.

REFERENCES

[1] *Zadeh, L. A. (1965). Fuzzy Sets. Information and Control, 8(3), 338–353*.

[2] *Mittal, M., Saraswat, A., Kumar, R., et al. (2020). Diagnosis of Diabetes Using Fuzzy Logic Modeling. International Journal of Innovative Technology and Exploring Engineering, 9(3), 2278–3075*.

[3] *Tuly, T. A., Shafin, T. A., Hasan, M. Z., Bijoy, M. H. I., Arefin, M. S., & Uddin, M. S. (2025, February). A Mamdani Fuzzy Inference System for Comprehensive Diabetes Risk Evaluation. In 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-7). IEEE*.

[4] *Nagaraj, P., & Deepalakshmi, P. (2022). An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis. International Journal of Imaging Systems and Technology, 32(4), 1373-1396*.

[5] *Hemanth, D. J., & Anitha, J. (2018). Medical Decision Support System for Diabetes Prediction Using Fuzzy Logic. Journal of Computational and Theoretical Nanoscience, 15(1), 49–55*.

[6] *Choubey, S., Paul, S., & Jha, R. (2016). Diabetes Prediction Using Machine Learning Techniques. International Journal of Engineering Research and Technology (IJERT), 5(12), 2395–006*.

[7] *Esfahani, E. T., Ghoushchi, S. J., & Yadegaridehkordi, E. (2019). A Fuzzy Expert System for Personalized Diet Recommendation. Soft Computing, 24, 11037–11053*.

[8] *Kaggle. PIMA Indians Diabetes Dataset and Benchmark Models. Available at:* https://www.kaggle.com/uciml/pima-indians-diabetes-database

[9] *Thakkar, H., Shah, V., Yagnik, H., & Shah, M. (2021). Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. Clinical eHealth, 4, 12-23*.

[10] *Thirugnanam, M., Kumar, P., Srivatsan, S. V., & Nerlesh, C. R. (2012). Improving the prediction rate of diabetes diagnosis using fuzzy, neural network, case based (FNC) approach. Procedia engineering, 38, 1709-1718*.

[11] *Pintanel, A. T. G., Dimuro, G. P., Borges, E. N., Lucca, G., & Barcelos, C. R. G. (2023). Fuzzy Logic for Diabetes Predictions: A Literature Review. ICEIS (1), 476-483*.

[12] *Jain, V., & Raheja, S. (2015). Improving the prediction rate of diabetes using fuzzy expert system. IJ Information Technology and Computer Science, 7(10), 84-91*.

[13] *Aamir, K. M., Sarfraz, L., Ramzan, M., Bilal, M., Shafi, J., & Attique, M. (2021). A fuzzy rule-based system for classification of diabetes. Sensors, 21(23), 8095*.

[14] *Arul, K. (2025, April). Advancing Diabetes Risk Prediction Using Anfis: a Fuzzy-Logic Approach for Improved Healthcare Decision Support. In 2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA) (pp. 1-6). IEEE*