

**AN EXPERT SYSTEM FOR INSULIN DOSAGE PREDICTION USING  
ARTIFICIAL NEURAL NETWORK AND RANDOM FOREST****Shaik Riyaz, Pereketla Naga Nithin****Guide: Koyya Avinash**Department of Computer Science and Engineering, K L University (KLU)  
Vijayawada Andhra Pradesh**ABSTRACT**

Diabetes is a common health problem that requires proper monitoring and correct insulin dosage. In this project, an expert system is developed to help in predicting diabetes and suggesting insulin dosage using machine learning techniques.

First, an Artificial Neural Network (ANN) is used to predict whether a person has diabetes based on medical data. If diabetes is detected, a Random Forest Regression model is used to predict the required insulin dosage. The system is trained using datasets such as the PIMA Indian Diabetes dataset and UCI datasets.

This approach helps in making better decisions by providing accurate predictions. Doctors and patients in managing diabetes more effectively and reducing health risks.

**Keywords:**

Health Care, Diabetes, ANN, Random Forest, Insulin Dosage, Machine Learning, Expert System

**Dataset Details**

| Parameter          | Description  |
|--------------------|--|
| <b>Data Source</b> | Publicly available datasets obtained from online machine learning repositories, including Kaggle and UCI       |
| <b>Categories</b>  | Medical attributes such as Glucose Level, Blood Pressure, BMI, Age, Insulin Level, Pregnancies                 |
| <b>Data Type</b>   | Structured Dataset (CSV/Tabular Data)  |
| <b>Content</b>     | Patient medical records including health parameters used for diabetes prediction and insulin dosage estimation |
| <b>Purpose</b>     | Diabetes prediction and insulin dosage estimation using Machine Learning techniques                            |

**Novelty Statement:**

The novelty of this project lies in the integration of classification and regression techniques for complete diabetes management. The system first predicts whether a patient is diabetic using Artificial Neural Networks and then estimates the required insulin dosage using Random Forest Regression. This combined approach provides a more accurate and personalized solution, helping in better decision-making and reducing health risks associated with improper insulin dosage.

**INTRODUCTION**

Diabetes mellitus is a chronic disease that affects millions of people worldwide and requires proper monitoring. Managing diabetes is challenging because it involves accurate diagnosis. Traditional methods of insulin calculation are often manual and may lead to errors. With the advancement of Artificial Intelligence, machine learning techniques can be used to improve healthcare decision-making. In this project, an expert system is developed using Artificial Neural Networks (ANN) to predict whether a person has diabetes. If diabetes is

# IJETRM

## International Journal of Engineering Technology Research & Management (IJETRM)

### Journal Article

<https://ijetrm.com/issue/>

detected, a Random Forest Regression model is used to estimate the required insulin dosage. This system helps in providing accurate and data-driven support for better diabetes management.

### OBJECTIVES

The main objective of the study is to develop an expert system for the prediction of diabetes and estimation of insulin dosage using machine learning techniques. It aims to accurately identify diabetic patients using Artificial Neural Networks and, based on the prediction, determine the appropriate insulin dosage using Random Forest Regression. The study also focuses on utilizing standard datasets such as PIMA and UCI to train and evaluate the models, ensuring reliability and accuracy. Furthermore, the system is designed to support healthcare decision-making and improve diabetes management by reducing risks associated with incorrect insulin administration.

### PROPOSED WORK

This work presents a machine learning-based expert system for predicting diabetes and estimating insulin dosage. The proposed approach utilizes patient health data to generate accurate predictions and supports better decision-making in diabetes management. By applying intelligent algorithms, the system aims to provide reliable outputs that can assist in monitoring and treatment planning.

Patient medical attributes such as glucose level, BMI, age, blood pressure, and insulin values are used as inputs to the model. An Artificial Neural Network (ANN) is employed to perform binary classification for diabetes prediction.

Once diabetes is detected, a Random Forest Regression model is applied to estimate the required insulin dosage. This two-stage approach ensures both accurate diagnosis and appropriate dosage recommendation. The models are trained using datasets such as the PIMA Indian Diabetes dataset and UCI Machine Learning Repository.

In addition, the system provides a user-friendly interface that allows users to upload data, execute models, and visualize results such as prediction outputs and performance graphs.

Overall, the proposed system assists healthcare professionals and patients by providing data-driven insights, improving treatment accuracy, and reducing risks associated with incorrect insulin dosage.

### Model Analysis

| Model                 | Technique           | Advantages  | Limitations   |
|-----------------------|---------------------|---|---|
| <b>Proposed Model</b> | ANN + Random Forest | Provides reliable prediction results, supports non-linear data relationships, enables patient-specific insulin dosage estimation, and offers a user-friendly implementation | Model performance is influenced by the quality of input data and requires appropriate parameter optimization during training. |

### METHODOLOGY

This study is conducted using a quantitative and experimental research design. Developed using machine learning techniques to predict diabetes and estimate insulin dosage. The datasets used for this study include the PIMA Indian Diabetes dataset and UCI Machine Learning Repository datasets.

Initially, the collected data is cleaned and normalized to remove inconsistencies and missing values. Model is then trained to classify whether a patient is diabetic or not based on input features. Once diabetes is detected, a Random Forest Regression model is applied to predict the appropriate insulin dosage.

The performance of both models is evaluated using standard metrics such as accuracy for classification. The developed system aims to provide reliable and data-driven predictions.

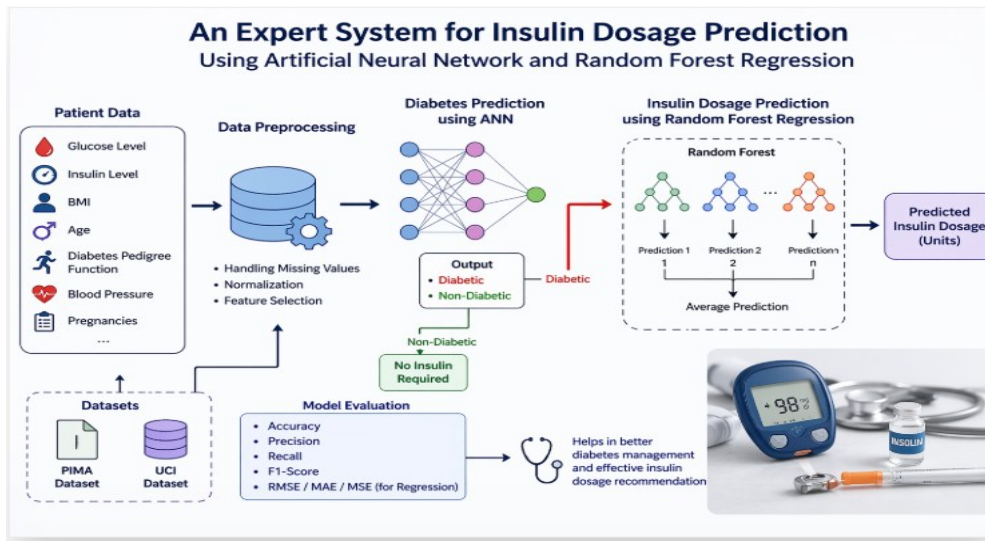


Figure 1.1 Architecture for proposed expert system for insulin dosage prediction

This figure illustrates the overall architecture of the proposed system, including data input, preprocessing, diabetes prediction using ANN, and insulin dosage estimation using Random Forest. It shows the step-by-step flow from patient data to final prediction output, highlighting how both models work together for accurate results.

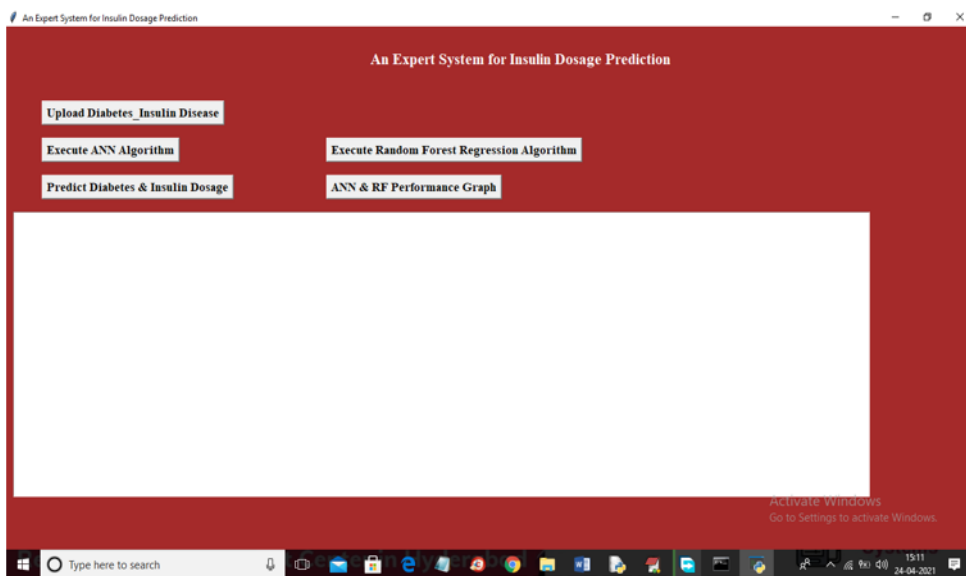


Figure 1.2 Interface

This interface allows the user to upload the dataset and execute ANN and Random Forest algorithms for prediction. It provides options to predict diabetes status and estimate insulin dosage along with performance visualization.

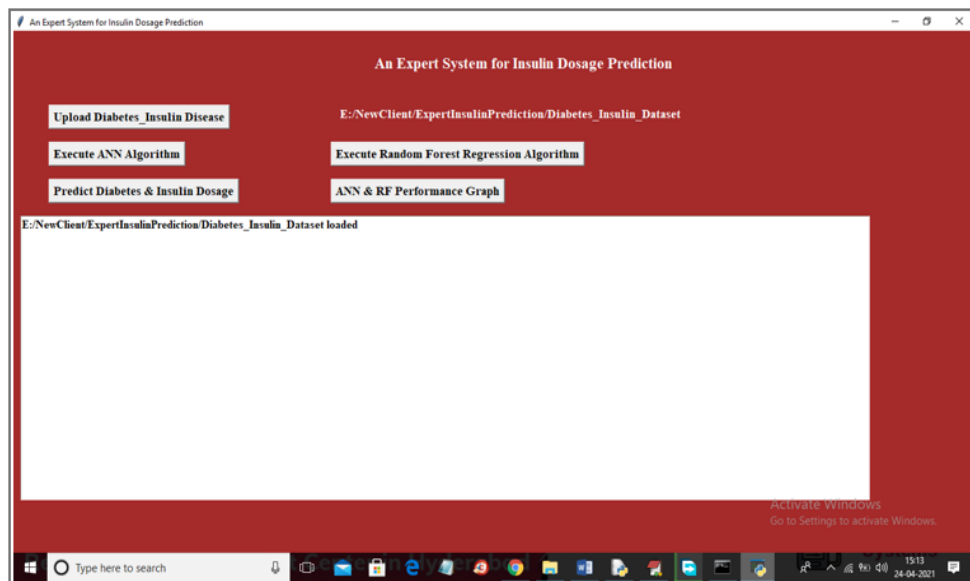


Figure 2.1 Data Uploaded Successfully

This interface shows the dataset being successfully uploaded and ready for processing. It allows the user to execute ANN and Random Forest algorithms for diabetes prediction and insulin dosage estimation.

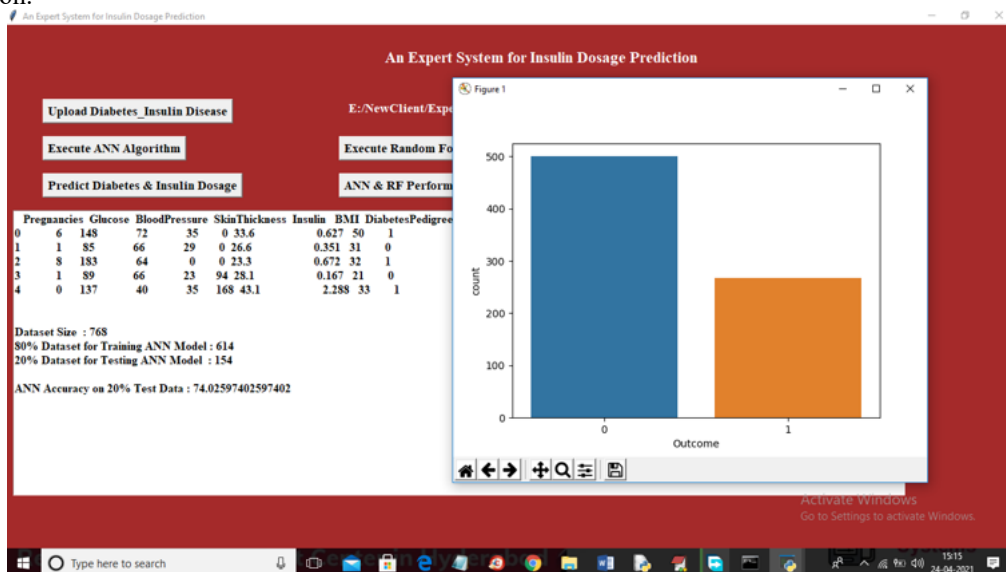
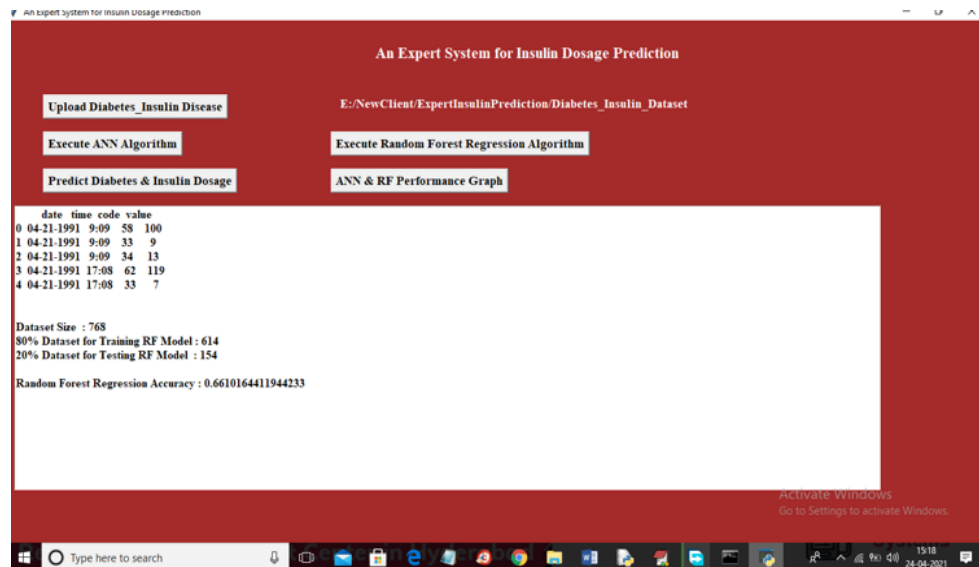


Figure 2.2 Graphs

In above screen in text area we can dataset contains total 768 records and application using 614 records for training and 154 for testing and we got ANN accuracy on test data is 74 and in above graph we are displaying total records found with class label 1 and 0. In above graph x-axis represents class label as 0 and 1 and y-axis represents total records found in class 0 and 1. Now close above graph and then click on 'Execute Random Forest Regression Algorithm' button to build random forest on insulin dataset

**Figure 3.1 Final Predictions**

This interface displays the dataset details along with training and testing split information. It also shows the Random Forest model accuracy after processing the data.

### Experimental Results and Analysis

The proposed expert system demonstrates effective performance in predicting diabetes and estimating insulin. The Artificial Neural Network (ANN) model was able to accurately classify patients as diabetic or non-diabetic based on input medical parameters. Captures complex relationships in the dataset.

The Random Forest Regression model was applied to predict insulin dosage for diabetic patients. The dataset was divided into training and testing sets, and the model achieved satisfactory accuracy, indicating reliable prediction capability.

The results show that combining ANN and Random Forest improves overall system performance compared to traditional methods. This approach supports personalized healthcare by providing data-driven predictions. However, the performance of the system depends on the quality and size of the dataset, and further improvements can be made by using larger and more diverse data.

### ACKNOWLEDGEMENT

The completion of this project was made possible with the support of our project guide, Mr. Koyya Avinash, whose suggestions and technical insights were valuable throughout the development process.

We also recognize the role of KL University, Vijayawada, and the Department of Computer Science and Engineering for offering the necessary academic environment and resources.

The use of open-source datasets, including the PIMA Indian Diabetes dataset and UCI repository, is also acknowledged as they were essential for model training and evaluation.

### CONCLUSION

Based on the proposed expert system, the study concludes that machine learning techniques can effectively support diabetes prediction and insulin dosage estimation. Accurate classification of patients as diabetic or non-diabetic, while the Random Forest Regression model provides reliable insulin dosage predictions based on patient data.

The integration of these models improves overall system performance and offers a data-driven approach for personalized diabetes management. The system demonstrates that combining classification and regression techniques can enhance decision-making in healthcare applications.

However, the effectiveness of the system depends on the quality and size of the dataset used for training. Future improvements can include the use of larger and more diverse datasets, integration with real-time health monitoring systems, and further optimization of models to achieve higher accuracy and reliability.

# IJETRM

**International Journal of Engineering Technology Research & Management (IJETRM)**

**Journal Article**

<https://ijetrm.com/issue/>

## REFERENCES

- [1] UCI Machine Learning Repository. *PIMA Indians Diabetes Dataset*. Available: <https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes>
- [2] Kaggle. *PIMA Indians Diabetes Database*. Available: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>
- [3] Breiman, L. (2001). *Random Forests*. *Machine Learning*, 45(1), 5–32.
- [4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [5] Haykin, S. (2009). *Neural Networks and Learning Machines*. Prentice Hall.
- [6] American Diabetes Association (2020). *Standards of Medical Care in Diabetes*. *Diabetes Care Journal*.
- [7] Kotsiantis, S. B. (2007). *Supervised Machine Learning: A Review of Classification Techniques*. *Informatica*.
- [8] Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
- [9] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.