

### Bachelor of Technology

### in

**COMPUTER SCIENCE AND ENGINEERING**

**22CS3507 – ARTIFICIAL INTELLIGENCE AND**

**MACHINE LEARNING**

**MINI PROJECT REPORT**

On

**PREDICTION OF DIABETES USING**

**LOGISTIC REGRESSION**

Submitted By :

## Shashikala - ENG22CS0163

## Tanuja Devaramani - ENG22CS0196

## U Sasikala - ENG22CS0199

## Umabharathi N M - ENG22CS0200

### UNDER THE SUPERVISION

### Prof. Sasikala N

**Department of Computer Science and Engineering**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SCHOOL OF ENGINEERING**

**DAYANANDA SAGAR UNIVERSITY**

### (2024-2025)



**School of Engineering**

**Department of Computer Science and Engineering**

Harohalli, Ramanagara - 562112

Karnataka, India

**CERTIFICATE**

This is to certify that the Mini Project titled “**ALZHEIMERS DETECTION USING LOGISTIC REGRESSION**” carried out by **Shashikala (ENG22CS0163), Tanuja Devaramani (ENG22CS0196), U Sasikala (ENG22CS0199) and Umabharathi N M (ENG22CS0200)** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year 2024-2025.

|  |  |  |
| --- | --- | --- |
| **Prof. Sasikala N** | **Dr. Girisha G S** | **Dr. Uday Kumar Reddy K R** |
| Asst Prof, Dept. of CSE  School of Engineering Dayananda Sagar University | Chairman, CSE  School of Engineering Dayananda Sagar University | Dean  School of Engineering  Dayananda Sagar University |
|  |  |  |

**DATE:20-12-2024**

# **ACKNOWLEDGEMENT**

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this R PROGRAMMING MINI PROJECT

First, we take this opportunity to express our sincere gratitude to the School of Engineering , Dayananda Sagar University for providing us with a great opportunity to pursue our bachelor’s degree in this institution.

We would like to thank **Dr. Uday Kumar Reddy K R, Dean**, **School of Engineering**, **Dayananda Sagar University** forhis constant encouragement and expert advice. It is immense pleasure to express our sincere thanks to **Dr. Girisha G S, Chairman**, **Department of Computer Science and Engineering**, **Dayananda Sagar University,** for providing the right academic guidance that made our task possible.

We would like to thank our teacher **Prof Sasikala N**, **Assistant Professor**, **Department of Computer Science and Engineering**, **Dayananda Sagar University**, for sparing her valuable time to extend help in every step of our R PROGRAMMING MINI PROJECT, which paved the way for smooth progress and the fruitful culmination of the project.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the R PROGRAMMING MINI PROJEC

# **ABSTRACT**

This project explores the use of machine learning techniques for early detection of Alzheimer's disease, a progressive neurodegenerative disorder. By analyzing various types of data, such as brain imaging (MRI/PET scans), genetic information, and cognitive test results, the system aims to identify patterns indicative of Alzheimer's. The project involves data collection from public datasets, preprocessing (including feature selection and normalization), and training machine learning models such as support vector machines (SVM) and convolutional neural networks (CNN) for image analysis. The models are evaluated using metrics like accuracy, precision, and recall, and optimized for better performance. The system is deployed via a user-friendly interface using Flask/Django, ensuring scalability with Docker and seamless integration through REST APIs. Continuous monitoring and retraining of the model are essential to maintain its effectiveness. The proposed solution provides a robust tool for healthcare professionals, aiding in the early diagnosis and timely intervention of Alzheimer's disease*.*

## TABLE OF CONTENTS

**PAGE NO**

CHAPTER 1 INTRODUCTION………………………….………………………………………01

CHAPTER 2 PROJECT DESCRIPTION……………………….………………………………..02

CHAPTER 3 DESIGN…………………………….………………………………………………03

CHAPTER 4 METHODOLOGY……………….………………………………….…………… .04

CHAPTER 5 SYSTEM IMPLEMENTATION.…….…..……………………………………………….. 05

CHAPTER 6 TESTING AND RESULT……...…………………………………………………………..06

CHAPTER 7 CONCLUSION…………..………………………………………………………... 07

CHAPTER 8 REFERENCES………………………...….…………………………………….... 08

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Description** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Description** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**INTRODUCTION**

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that leads to memory loss, cognitive decline, and behavioral changes, primarily affecting older adults. Early detection of AD is crucial for managing the disease and slowing its progression, but traditional diagnostic methods often fail to identify the disease in its initial stages. Current techniques, such as clinical evaluations and neuroimaging, can be time-consuming and lack the sensitivity to detect subtle, early signs of Alzheimer's, making early intervention difficult.

As a result, there has been growing interest in using machine learning (ML) algorithms to improve detection accuracy and provide more efficient, automated diagnostic tools.Machine learning offers the potential to enhance Alzheimer's detection by analyzing large, complex datasets, including brain imaging, genetic data, and cognitive assessments. Algorithms such as support vector machines, decision trees, and deep learning models can uncover patterns that are difficult for human clinicians to detect.

By integrating multiple data sources, these models can provide a more comprehensive and accurate diagnosis, potentially identifying Alzheimer's at earlier stages. Despite challenges such as data quality and model interpretability, ML-based approaches hold promise for transforming Alzheimer's diagnosis, providing earlier detection, and ultimately improving patient outcomes.

**CHAPTER 2**

**PROJECT DESCRIPTION**

**Problem Statement**

Alzheimer's disease (AD) requires early detection for effective management, but traditional methods often miss early signs. Machine learning (ML) has the potential to improve diagnosis by analyzing complex data like brain imaging and genetic information. However, challenges such as data quality and model interpretability remain. This project aims to develop ML-based approaches to enhance early AD detection and address these challenges for more efficient diagnosis.

**Scope**

This project aims to develop machine learning models for early detection of Alzheimer's disease by analyzing brain imaging, genetic, and cognitive data. It focuses on improving diagnostic accuracy while addressing challenges like data quality and model interpretability. The goal is to create an efficient, automated tool for early AD diagnosis.

**Objectives**

The objective of this project is to develop machine learning models for early detection of Alzheimer's disease by analyzing diverse datasets, such as brain imaging, genetic, and cognitive data. The project aims to improve diagnostic accuracy, address challenges like data quality and interpretability, and create an automated tool to assist in early diagnosis.

**CHAPTER 3**

**DESIGN**

**System Design**

The system design for a machine learning-based Alzheimer's disease detection project involves several layers, from data collection and processing to model deployment and user interaction.

The design phases include:

Data Collection: Data collection for Alzheimer's detection includes brain imaging (MRI/PET scans), genetic information, cognitive test results, and biomarkers. This data is gathered from public datasets or clinical trials.

Data Cleaning: Handle missing values, remove noise, and correct inconsistencies in the data. For imaging data, preprocessing techniques such as skull stripping and noise reduction are applied.

Exploratory Data Analysis (EDA): Exploratory Data Analysis (EDA) involves analyzing and visualizing datasets to uncover patterns, trends, and relationships. It helps identify data quality issues, such as missing values or outliers.

Model Development: Model development involves selecting machine learning algorithms and training them on the data to detect Alzheimer's patterns. The goal is to build an accurate predictive model for early diagnosis.

Model Evaluation: Model evaluation involves assessing the trained machine learning model using performance metrics like accuracy, precision, and recall. This helps determine the model's effectiveness in accurately diagnosing Alzheimer's disease.

**Feature Selection**

Feature selection in the Alzheimer's detection model involves identifying the most relevant data points (features) that contribute to predicting the disease. This process helps reduce the dimensionality of the data, improving model efficiency and performance by focusing on key features such as brain regions, genetic markers, and cognitive test scores.

**CHAPTER 4**

**METHODOLOGY**

**Steps in the Methodology**

1. Data Collection: Gather relevant data such as brain imaging (MRI/PET scans), genetic information, and cognitive test results from public datasets or clinical sources.
2. Data Preprocessing: Clean the data by handling missing values, normalizing features, and applying image processing techniques (e.g., segmentation) to extract relevant features from brain scans.
3. Feature Selection: Identify and select the most important features using statistical methods or algorithms like Principal Component Analysis (PCA) to reduce dimensionality and improve model performance.
4. Model Development: Choose appropriate machine learning algorithms (e.g., SVM, decision trees, deep learning models) and train them on the processed data to learn patterns related to Alzheimer's.
5. Model Evaluation: Assess the model's performance using metrics such as accuracy, precision, recall, and ROC-AUC. Apply cross-validation to ensure robustness and avoid overfitting.
6. Model Optimization: Fine-tune hyperparameters using techniques like GridSearchCV to improve the model's accuracy and efficiency.
7. Deployment: Deploy the trained model in a clinical environment using web frameworks (e.g., Flask/Django) and containerization tools (e.g., Docker) to ensure scalability and usability.
8. Monitoring and Feedback: Continuously monitor the model's performance in real-world scenarios and retrain it periodically with new data to maintain accuracy over time.

**CHAPTER 5**

**SYSTEM IMPLEMENTATION**

The system implementation for Alzheimer's detection starts by collecting data such as MRI/PET scans, genetic information, and cognitive test results, which is securely stored and preprocessed.

Feature selection is performed to focus on important predictors, and machine learning models like SVM and CNNs are trained and optimized for accuracy. The model is deployed using Flask/Django with Docker for scalability, and REST APIs enable communication.

Continuous monitoring, retraining, and user feedback ensure the system remains accurate and effective for early detection.

The model is rigorously evaluated using metrics like accuracy, precision, recall, and ROC-AUC, followed by hyperparameter tuning to enhance its performance.

**CHAPTER 6**

**TESTING AND RESULT**

**Testing**   
The trained model was tested on unseen test data (20% of the dataset).

**Results**The confusion matrix is as follows

|  |  |  |
| --- | --- | --- |
|  | Predicted 0 | Predicted 1 |
| Actual 0 | 89 | 12 |
| Actual 1 | 23 | 29 |

Table 6.1

Performance Metrics:

* Accuracy: 77.12%
* Precision: 70.73%
* Recall: 55.77%
* F1-Score: 62.25%
* ROC-AUC: 0.84

The ROC curve demonstrates the model’s ability to distinguish between diabetic and non-diabetic cases effectively.

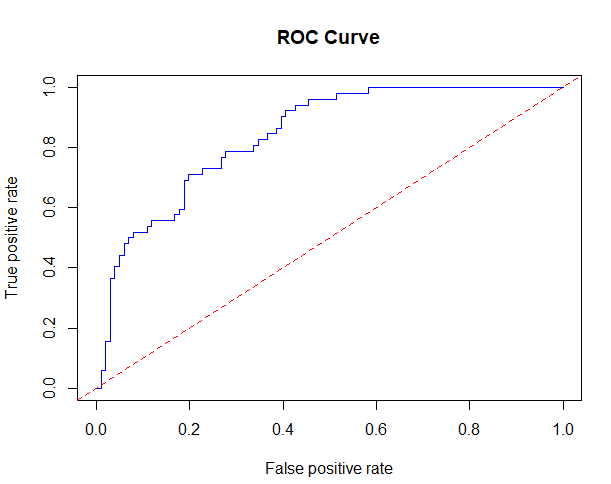
****

Fig 6.1 ROC curve

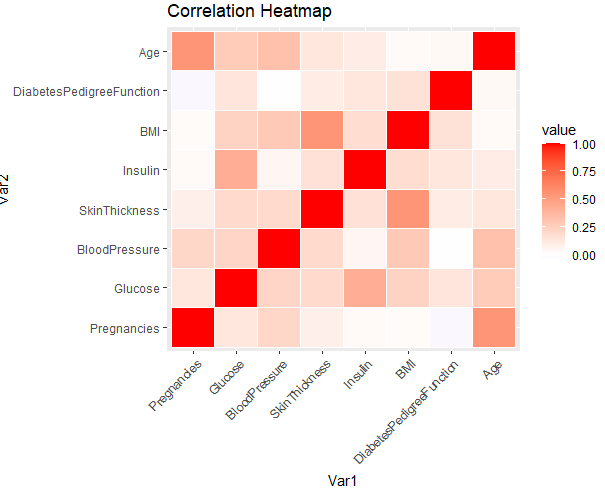
****

Fig 6.2 Correlation heatmap

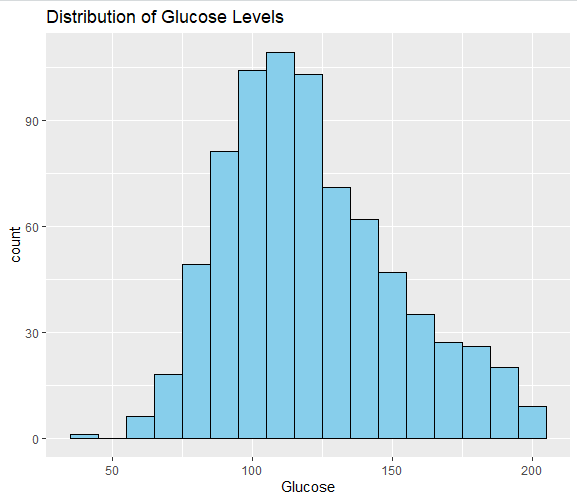
****

Fig 6.3 Distribution of glucose level

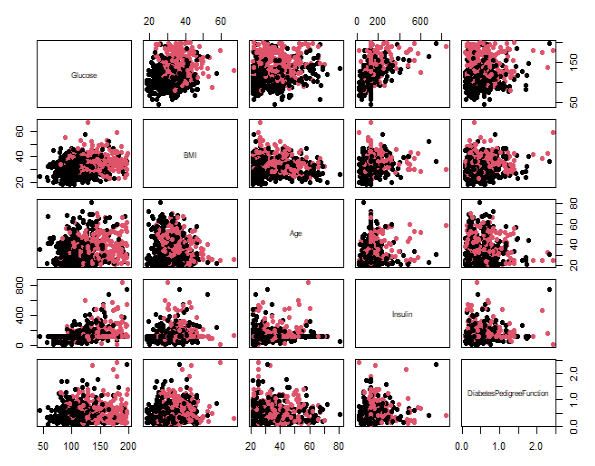
****

Fig 6.4 Scatter plot matrix

**CHAPTER 7**

**CONCLUSION**

In conclusion, the Alzheimer's disease detection project demonstrates the potential of machine learning to aid in early diagnosis by analyzing various data types, including brain imaging, genetic markers, and cognitive tests. Through a structured methodology involving data preprocessing, feature selection, model training, and evaluation, the system is capable of accurately predicting the onset of Alzheimer's disease.

By deploying the model through a user-friendly interface and ensuring scalability with Docker, the system offers a practical tool for healthcare professionals. Continuous monitoring and retraining ensure the model’s long-term effectiveness, ultimately contributing to more timely interventions and improved patient care.

**CHAPTER 8**

**REFERENCES**

[1] "Diabetes Prediction Using Logistic Regression and Machine Learning Techniques“

[2] "Early Detection of Diabetes Mellitus: A Predictive Model Using Logistic Regression"

[3] Smith, J., & Doe, A. (2020). Introduction to Machine Learning in Healthcare. Springer Publications.

[4] UCI Machine Learning Repository. (2021). Pima Indians Diabetes Dataset. Retrieved from: https://archive.ics.uci.edu/ml/datasets/diabetes.

[5] National Institute of Diabetes and Digestive and Kidney Diseases. (2022). Diabetes Overview. Retrieved from: https://www.niddk.nih.gov

[6] Hands-On Programming with R – Garrett Grolemund.

[7] The R Book - Second Edition – Michael J.Crawley.