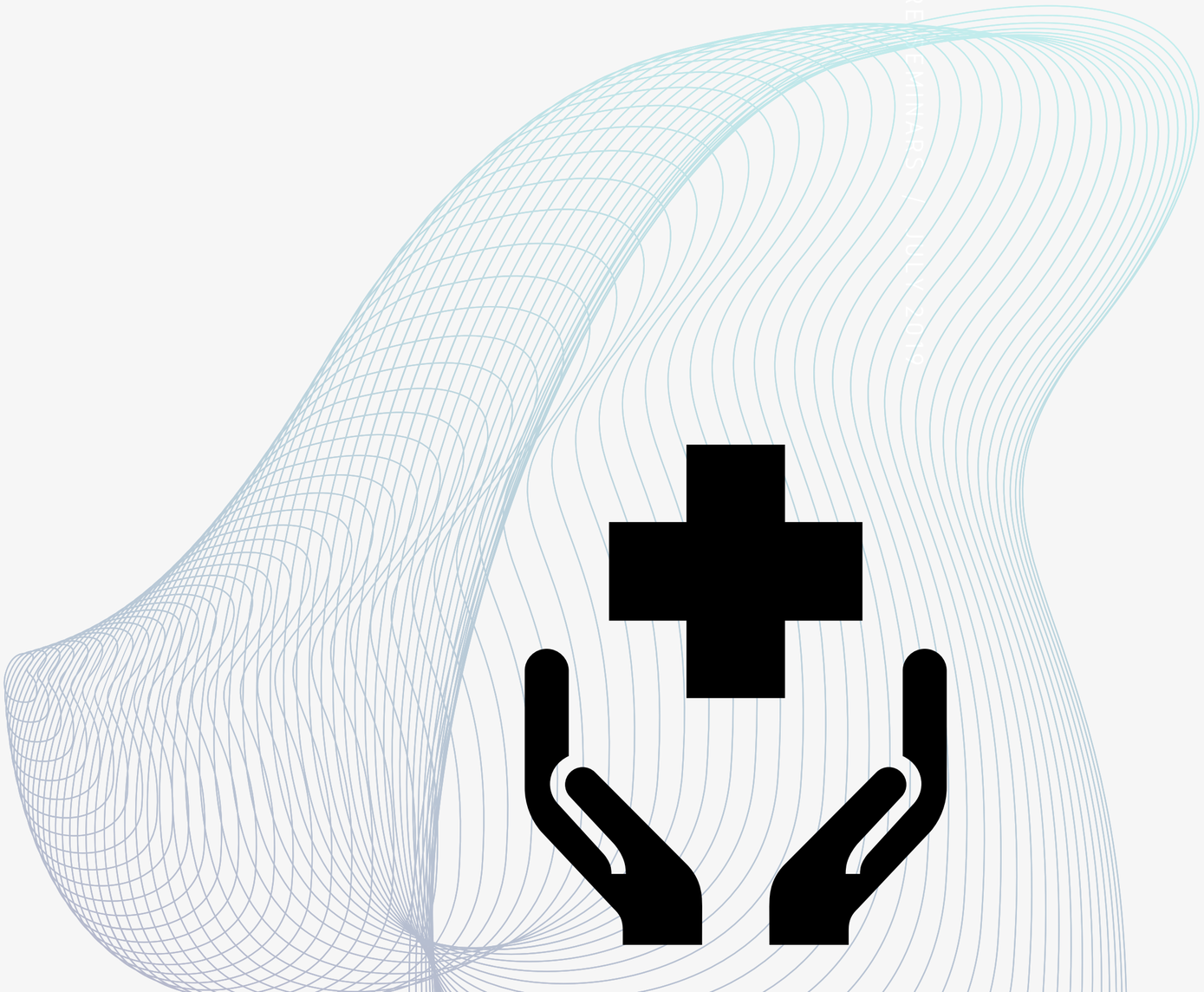
PROBLEM STATEMENT

Machine Learning Models for Alzheimer Disease Prediction, Risk Analysis and Recommendations

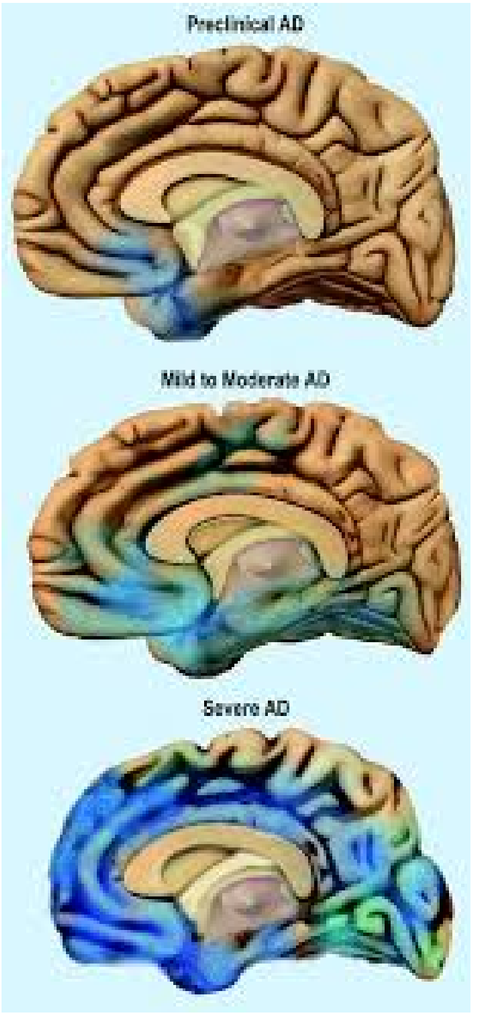
Table of Contents



INTRODUCTION

INTRODUCTION

- Alzheimer's disease is a **progressive neurodegenerative disorder** that affects memory, thinking, and behavior. It is the most **common form of dementia**, impacting millions of people worldwide.
- **Primary Symptoms:** Memory loss, confusion, difficulty completing familiar tasks, language problems, and mood swings.



OBJECTIVES

OBJECTIVES

To develop and train machine learning models for early prediction of Alzheimer’s disease using clinical, behavioral, and lifestyle-related input features.

To provide personalized recommendations based on parameters to help users take preventive measures and manage potential Alzheimer’s risk more effectively.

To design a user-friendly web application that allows individuals to input relevant data and receive real-time risk assessment along with prediction results.

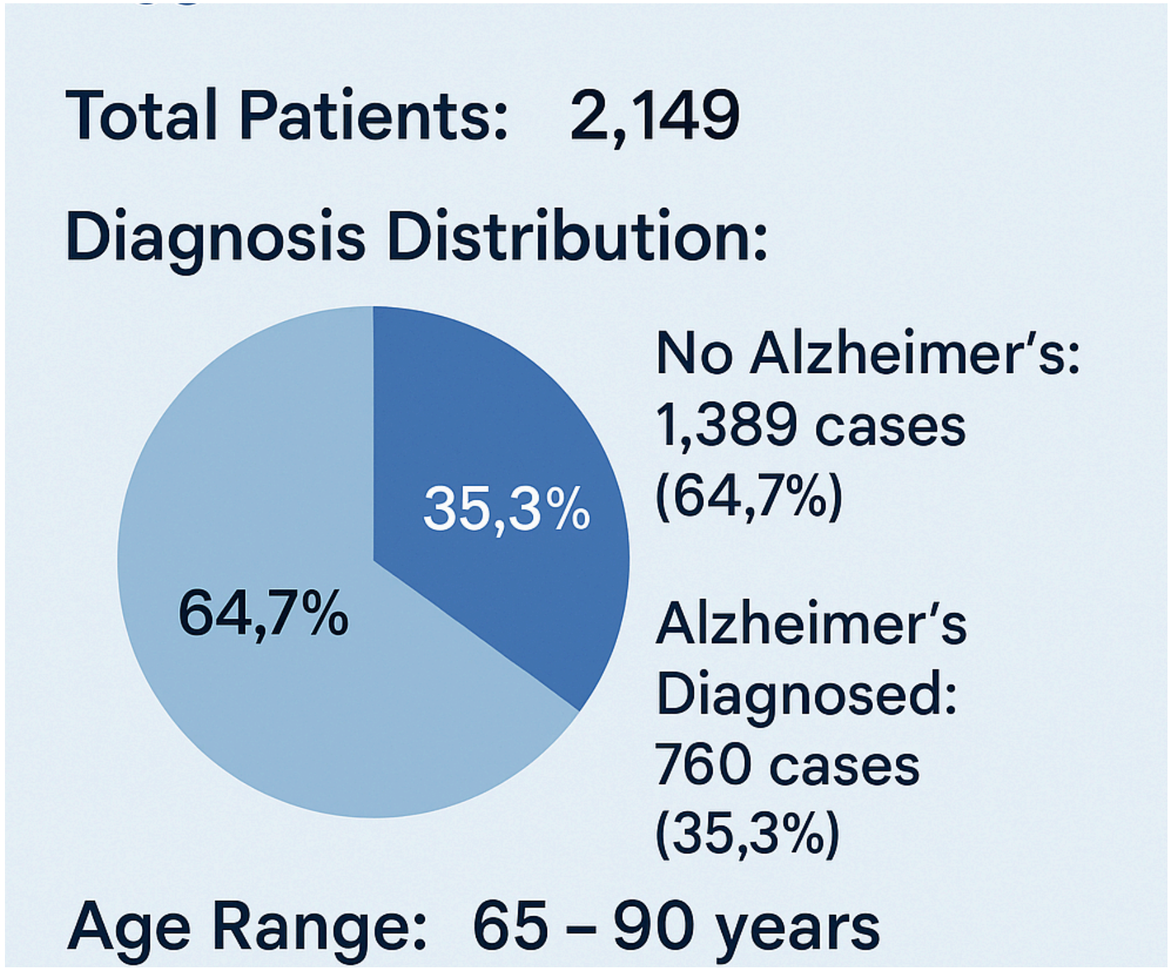
LITERATURE SURVEY

Author(s)	Technique	Data Type	Advantage	Limitation
Machine learning techniques for early detection of Alzheimer's disease, Brown & White (2019)	Decision Tree, KNN	Clinical + Cognitive	Good baseline models	Poor generalization across datasets
Artificial Intelligence for Healthcare: Applications in Alzheimer's Disease, Thomas & Clark (2020)	Impact of Preprocessing on ML Models	Structured Alzheimer's data	Effective preprocessing (e.g., cleaning, transformation) significantly improved model accuracy and reduced noise.	Preprocessing steps are data-specific and time-consuming, requiring domain knowledge and manual effort.
Comparison of machine learning algorithms for Alzheimer's disease prognosis, Patel et al. (2021)	Random Forest, GBM	Structured Data	Ensemble models improved accuracy	Need for large training data
Evaluation metrics for Alzheimer's disease classification models, Patel (2021)	Evaluation Standards in AD ML Models	Various structured datasets	Promoted use of standard metrics for reliable and comparable model evaluation.	Inconsistent metrics make cross-study comparison difficult.

Author(s)	Technique	Data Type	Advantage	Limitation
Deep learning for Alzheimer's disease classification using PET and MRI scans, Gupta & Rai (2021)	Deep CNN	PET + MRI	State-of-the-art detection	Risk of overfitting on small datasets
Improving Alzheimer’s diagnosis using AI-powered recommendation systems, Rodriguez (2021)	AI Recommender System	Clinical History	Personalized care suggestions	Limited adoption in hospital setups
Dietary patterns and Alzheimer's risk: A lifestyle-based correlation study, Hossain et al. (2023)	Diet–Risk Correlation	Lifestyle Survey	Highlighted diet’s role in AD risk	Not used in AI modeling yet
The impact of sleep patterns on cognitive decline in Alzheimer’s patients, Kim et al. (2022)	Sleep–Cognition Study	Sleep + Cognitive Data	Linked sleep to cognitive decline	Not integrated into diagnostic models

Dataset

DATASET



Rabie El Kharoua. (2024). Alzheimer's Disease Dataset [Data set].
Kaggle. https://doi.org/10.34740/KAGGLE/DSV/866827

DATASET DESCRIPTION

👫 **Demographics**

- Age, gender, ethnicity, and education level of patients.

🏋️‍♂️ **Lifestyle & Behavioral Factors**

- Body Mass Index (BMI), smoking habits, alcohol consumption, physical activity levels, diet quality, and sleep patterns.

🩺 **Medical History & Risk Factors**

- Family history of Alzheimer’s, cardiovascular conditions, diabetes, depression, head trauma, and hypertension.

📊 **Clinical & Biometric Data**

- Blood pressure (systolic & diastolic), cholesterol levels (LDL, HDL, triglycerides), and other vital health metrics.

🧠 **Cognitive & Functional Assessments**

- Mini-Mental State Examination (MMSE) scores, memory evaluation, behavioral patterns, and daily living activity assessments.

⚠️ **Neurological & Behavioral Symptoms**

- Confusion, disorientation, personality changes, forgetfulness, and difficulty in performing routine tasks.

🔬 **Diagnosis & Classification**

- Alzheimer’s status: Binary classification (0 = No Alzheimer’s, 1 = Diagnosed with Alzheimer’s).

Feature	Description	Example Values
Age	Person’s age (numerical)	65, 75, 80
Gender	Biological sex	Male, Female
Ethnicity	Ethnic background	Asian, Hispanic, Caucasian
EducationLevel	Highest education achieved	High School, Bachelor’s, PhD
BMI	Body Mass Index (weight category)	22.5 (Normal), 30.1 (Obese)
Smoking	Smoking status	Yes, No
AlcoholConsumption	Frequency of alcohol intake	Never, Occasionally, Daily
DietQuality	Nutritional quality of diet	Poor, Good
SleepQuality	Quality of sleep	Poor, Good
FamilyHistoryAlzheimers	Family history of Alzheimer’s	Yes, No
Diabetes	Presence of diabetes	Yes, No
Depression	Presence of depression symptoms	Yes, No
HeadInjury	Past head injuries	Yes, No

Hypertension	Presence of high blood pressure	Yes, No
SystolicBP	Systolic blood pressure reading	120 mmHg, 150 mmHg
DiastolicBP	Diastolic blood pressure reading	80 mmHg, 95 mmHg
CholesterolLDL	Low-density lipoprotein (bad cholesterol)	100 mg/dL, 190 mg/dL
CholesterolHDL	High-density lipoprotein (good cholesterol)	60 mg/dL, 30 mg/dL
CholesterolTriglycerides	Triglyceride levels in blood	120 mg/dL, 250 mg/dL
MMSE	Mini-Mental State Examination (cognitive function)	28/30 (Normal), 15/30 (Impaired)
ADL_FunctionalCombined	Ability to perform Activities of Daily Living (ADL)	High (Independent), Low (Needs help)
Behavioral_Issues_Combined	Summary of behavioral health issues	Mild anxiety, Severe agitation

IMPLEMENTATION

DATA PREPROCESSING

- **Removed Irrelevant or Redundant Features**

  Dropped columns like 'PatientID', 'CardiovascularDisease', "MemoryComplaints', 'Forgetfulness', 'PhysicalActivity' DoctorInCharge', etc. using df.drop() to reduce noise and improve model focus.

- **Combined Related Features**

  Merged 'ADL' and 'FunctionalAssessment' into a single composite feature 'ADL_FunctionalCombined' by averaging and rounding values.

- **Summarized Behavioral Issues**

  Created 'Behavioral_Issues_Combined' by summing up behavioral indicators (e.g., 'BehavioralProblems', 'Disorientation', etc.) into one meaningful score.

- **Rounding Float Values**

  Rounded all float-type columns to 2 decimal places using .round(2) for cleaner data and consistent precision.

ALGORITHMS USED

- Logistic Regression
- Random Forest
- Gradient Boosting
- XGBoost
- LightGBM
- CatBoost
- AdaBoost
- KNN
- Naive Bayes

LOGISTIC REGRESSION

Metric	Training	Testing
Precision	0.7096	0.7843
Recall	0.5977	0.5298
Accuracy	0.7708	0.7837
F1 Score	0.6488	0.6324

CONFUSION MATRIX:
[[257 22]
[71 80]]

RANDOM FOREST

Metric	Training	Testing
Precision	0.9091	0.9375
Recall	0.6568	0.4967
Accuracy	0.8551	0.8116
F1 Score	0.7626	0.6494

CONFUSION MATRIX:
[[274 5]
[76 75]]

GBM

Metric	Training	Testing
Precision	0.8571	0.9383
Recall	0.5287	0.5033
Accuracy	0.8015	0.8140
F1 Score	0.6540	0.6552

CONFUSION MATRIX:
[[274 5]
[75 76]]

XGBOOST

Metric	Training Set	Testing Set
Precision	0.8571	0.9500
Recall	0.5287	0.5033
Accuracy	0.8015	0.8163
F1 Score	0.6540	0.6580

CONFUSION MATRIX:
[[275 4]
[75 76]]

LIGHT GBM

Metric	Training	Testing
Precision	0.9918699187	0.7777777778
Recall	1.0	0.6026490066
Accuracy	0.9970909091	0.8
F1 Score	0.9959183673	0.6791044776

CONFUSION MATRIX:
[[253 26]
[60 91]]

CATBOOST

Metric	Training	Testing
Precision	0.9762419006	0.8736842105
Recall	0.9262295082	0.5496688742
Accuracy	0.9658181818	0.8139534884
F1 Score	0.9505783386	0.6747967480

CONFUSION MATRIX:
[[267 12]
[68 83]]

ADABOOST

Metric	Training	Testing
Precision	0.8547	0.9500
Recall	0.5184	0.5033
Accuracy	0.7978	0.8163
F1 Score	0.6454	0.6580

CONFUSION MATRIX:
[[275 4]
[75 76]]

KNN

Metric	Training	Testing
Precision	0.738	0.429
Recall	0.248	0.139
Accuracy	0.702	0.633
F1 Score	0.371	0.210

CONFUSION MATRIX:
[[251 28]
[130 21]]

NAIVE BAYES

Metric	Training Value	Testing Value
Precision	0.7273	0.8
Recall	0.5902	0.5563
Accuracy	0.7760	0.7953
F1 Score	0.6516	0.6563

CONFUSION MATRIX:
[[258 21]
[67 84]]

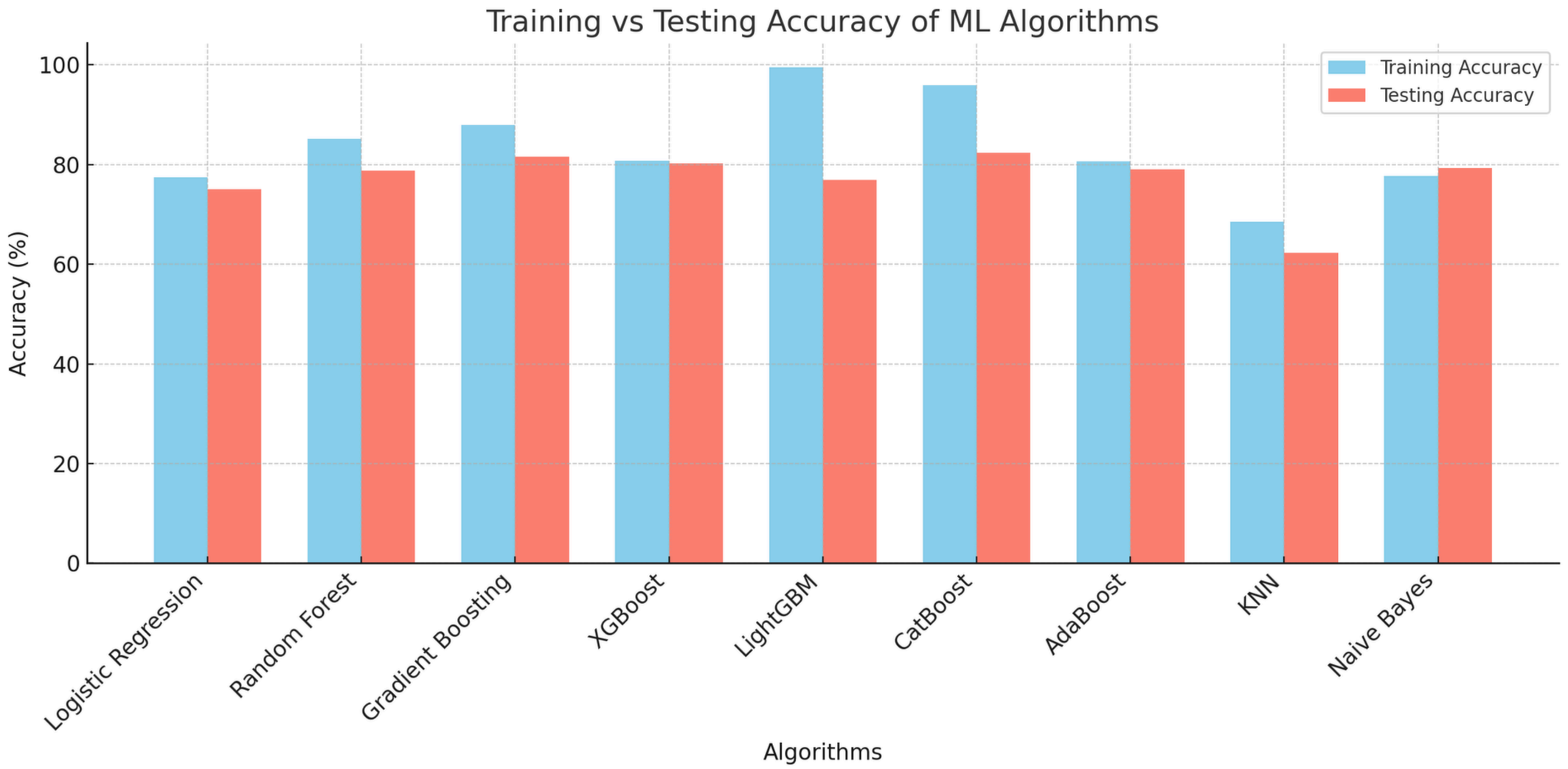
ALGORITHM PERFORMANCE COMPARISON

- BEST PERFORMING ALGORITHM (ACCURACY, PRECISION, RECALL, F1 SCORE): XGBOOST GENERALLY PERFORMED WELL ON BOTH TRAINING AND TESTING DATA WITH DECENT PRECISION AND RECALL, FOLLOWED BY CATBOOST AND LIGHTGBM. THESE MODELS SHOWED HIGH TRAINING AND TESTING ACCURACIES, MAKING THEM GOOD CANDIDATES FOR YOUR CLASSIFICATION PROBLEM.

- MODERATE PERFORMERS: RANDOM FOREST, LOGISTIC REGRESSION, AND ADABOOST PERFORMED REASONABLY WELL BUT SHOWED LOWER TESTING PRECISION AND RECALL WHEN COMPARED TO XGBOOST AND CATBOOST, ALTHOUGH THEIR F1 SCORES WERE STILL RESPECTABLE.

- LOWER PERFORMERS: NAIVE BAYES AND KNN HAD LOWER PERFORMANCE METRICS, PARTICULARLY FOR PRECISION AND RECALL, MEANING THEY STRUGGLED WITH CLASSIFYING THE POSITIVE CLASS EFFECTIVELY.

ALGORITHMS USED

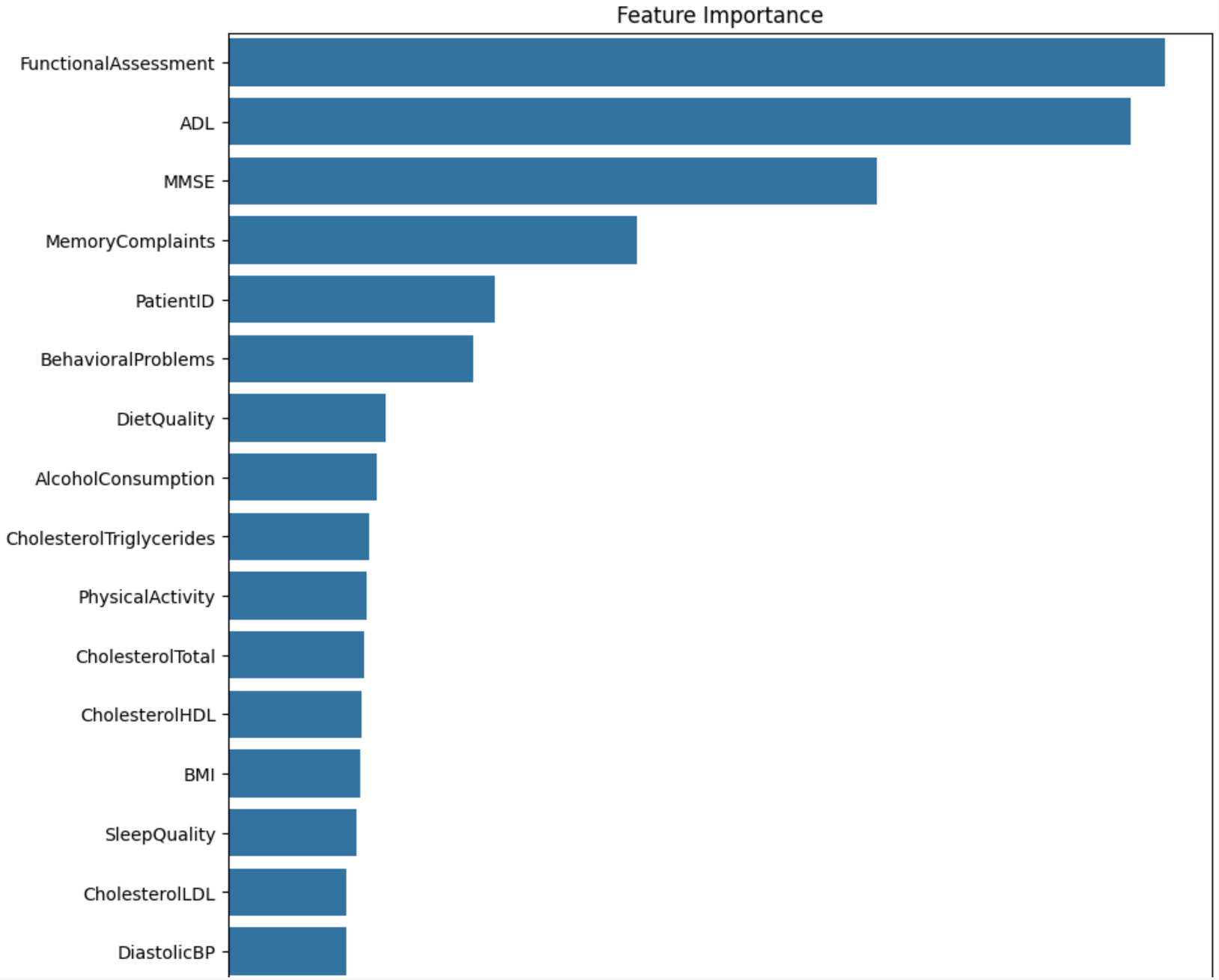


IMPLEMENTATION PART 2

ALGORITHMS USED

- Random Forest for Feature Selection
- XGBoost

FEATURE IMPORTANCE



FEATURE SELECTION

- Feature selection identifies and retains only the most influential features based on Random Forest feature importances.
- Features with an importance score below 0.01 are excluded to avoid irrelevant or low-impact variables, improving model efficiency.
- Excluding the 'PatientID' column prevents overfitting as it’s not predictive.
- Reducing the dataset to key features makes the model less complex, faster to train, and easier to interpret.
- The approach improves model accuracy and enhances generalizability to unseen data by focusing on the most relevant patterns.
- Selected 17 features based on their importance in predicting the target variable.
- Selected features :
- Age, BMI, AlcoholConsumption, PhysicalActivity, DietQuality, SleepQuality, SystolicBP, DiastolicBP, CholesterolTotal, CholesterolLDL, CholesterolHDL, CholesterolTriglycerides, MMSE, FunctionalAssessment, MemoryComplaints, BehavioralProblems.

FEATURE SELECTION

- Built an initial XGBoost model and evaluated it using accuracy, precision, recall, and F1-score.

- In the first round, we trained the XGBoost model with 100 estimators, achieving a training accuracy of 96.04% and a testing accuracy of 97.21%.

- In the second round, we continued training with 250 estimators, leading to a slight improvement in performance, with a training accuracy of 97.09% and a testing accuracy of 97.44%.

- Finally, in the third round, we further increased the number of estimators to 300, resulting in a training accuracy of 98.72% and maintaining the testing accuracy at 97.44%.

RESULTS

(100 ESTIMATORS)

Metric	Training Score	Testing Score
Accuracy	0.960	0.972
Precision	0.96	0.97
Recall	0.96	0.97
F1-Score	0.96	0.97

RESULTS

(250 ESTIMATORS)

Metric	Training Score	Testing Score
Accuracy	0.971	0.974
Precision	0.97	0.98
Recall	0.97	0.97
F1-Score	0.97	0.97

RESULTS

(300 ESTIMATORS)

Metric	Training Score	Testing Score
Accuracy	0.987	0.974
Precision	0.99	0.98
Recall	0.99	0.97
F1-Score	0.99	0.97

Recommendations

RECOMMENDATIONS

Provides personalized health recommendations based on parameters like BMI, blood pressure, cholesterol levels, and mental health assessments. For each parameter, we offer suggestions, such as adjusting diet, increasing exercise, or managing stress. By considering multiple health factors.

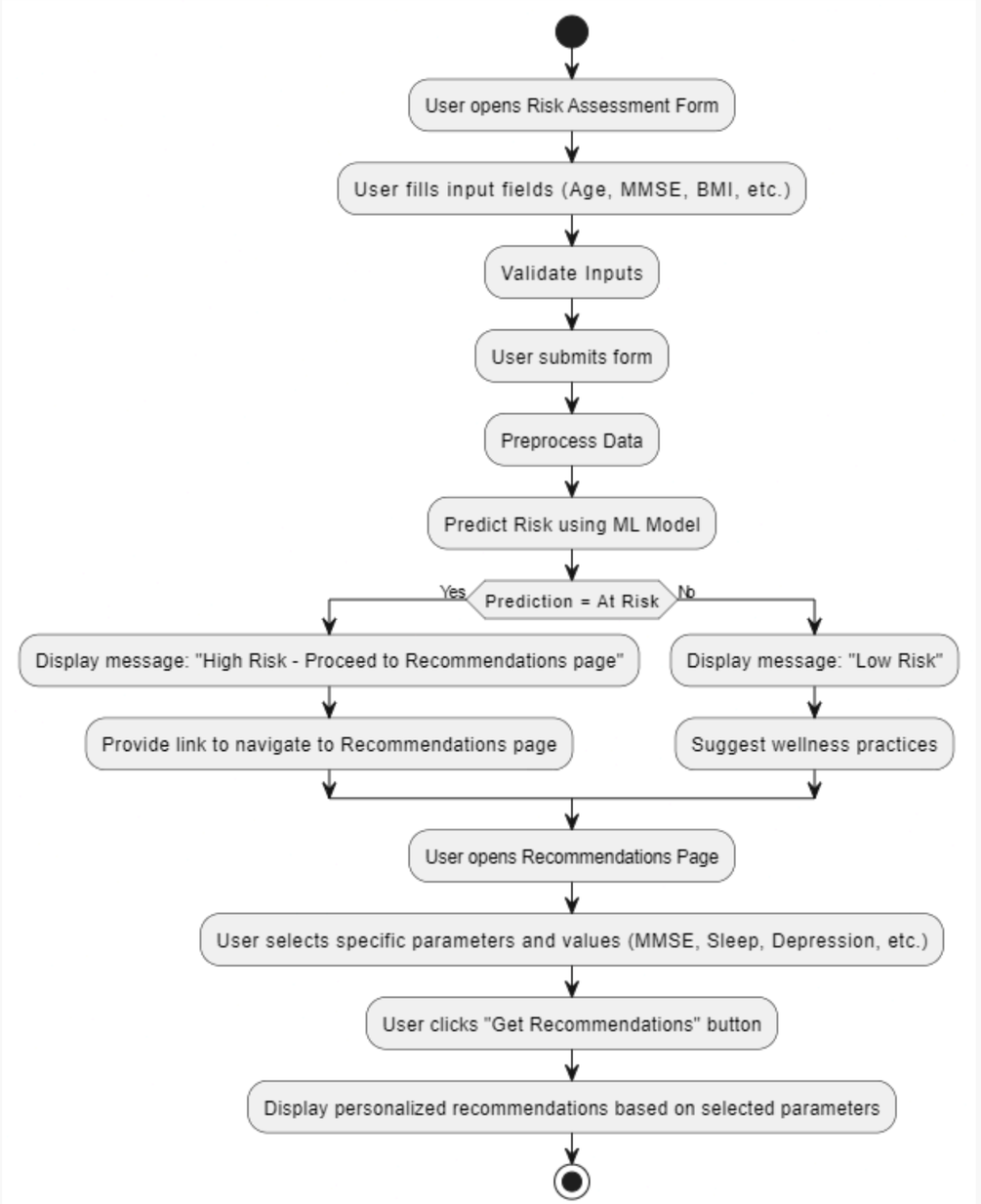
EXAMPLE

BMI

- Below 18.5: Underweight - Increase calorie intake with healthy snacks.
- 18.5 - 24.9: Normal weight - Maintain balanced diet.
- 25 - 29.9: Overweight - Increase fiber, reduce calories, exercise.
- 30 and above: Obese - Exercise, manage diet, consider medical interventions.

Web Application “AlzAware”

USER FLOW


User opens Risk Assessment Form
User fills input fields (Age, MMSE, BMI, etc.)
Validate Inputs
User submits form
Preprocess Data
Predict Risk using ML Model
Prediction = At Risk
Yes
No
Display message: "High Risk - Proceed to Recommendations page"
Display message: "Low Risk"
Provide link to navigate to Recommendations page
Suggest wellness practices
User opens Recommendations Page
User selects specific parameters and values (MMSE, Sleep, Depression, etc.)
User clicks "Get Recommendations" button
Display personalized recommendations based on selected parameters

HOME PAGE

AlzAware

Home

Risk Assessment

Get Recommendations

Brain Games

Knowledge Center

About

AlzAware

Assess, Empower, and Play for a Healthier Mind

Welcome

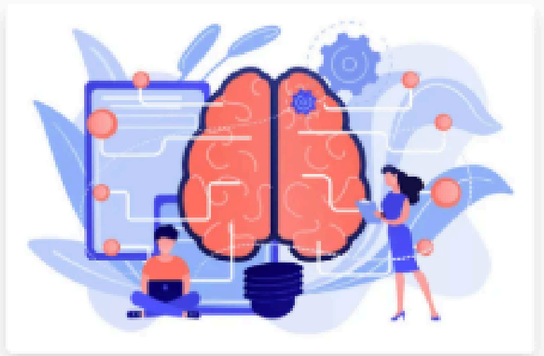
Our platform helps assess cognitive health by analyzing your input and guiding you with personalized recommendations. It serves as an early indicator tool—not a diagnosis—but a step toward better brain health awareness.

How It Works

- Click **"Risk Assessment"** to begin the assessment.
- **Get recommendations** tailored to your lifestyle and health inputs.
- Explore **"Brain Games"** to boost memory, logic, and attention through fun activities.

The Power of Early Risk Awareness

Early awareness of Alzheimer's risk empowers individuals to make meaningful changes. By identifying risk factors early, you can adopt proactive habits, improve mental fitness, and take control of your cognitive health journey.



Risk Assessment

Early Insight for a Healthier Tomorrow.

Get Recommendations

Personalized health tips and suggestions.

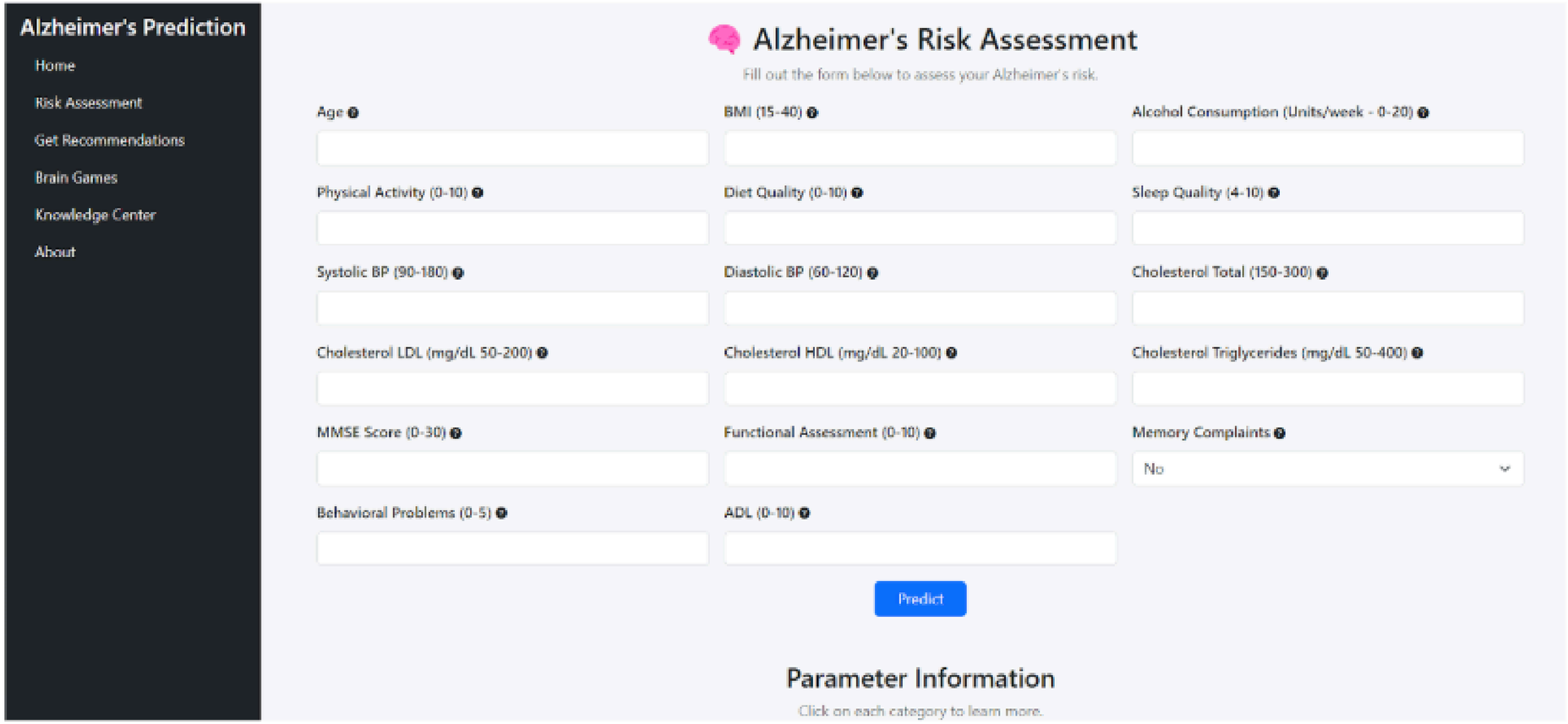
Brain Games

Engaging puzzles and memory boosters.

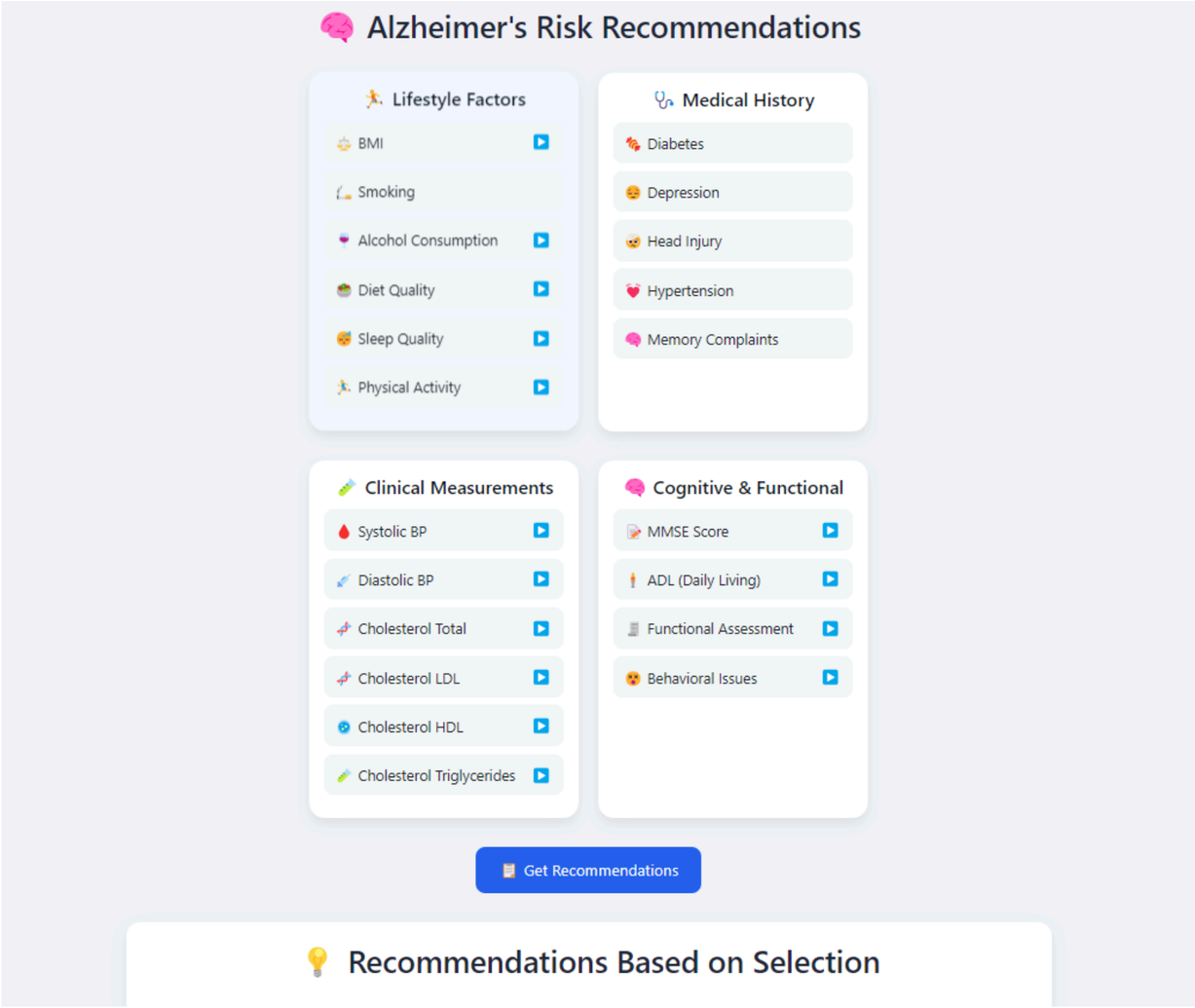
About

Learn more about our mission and our website.

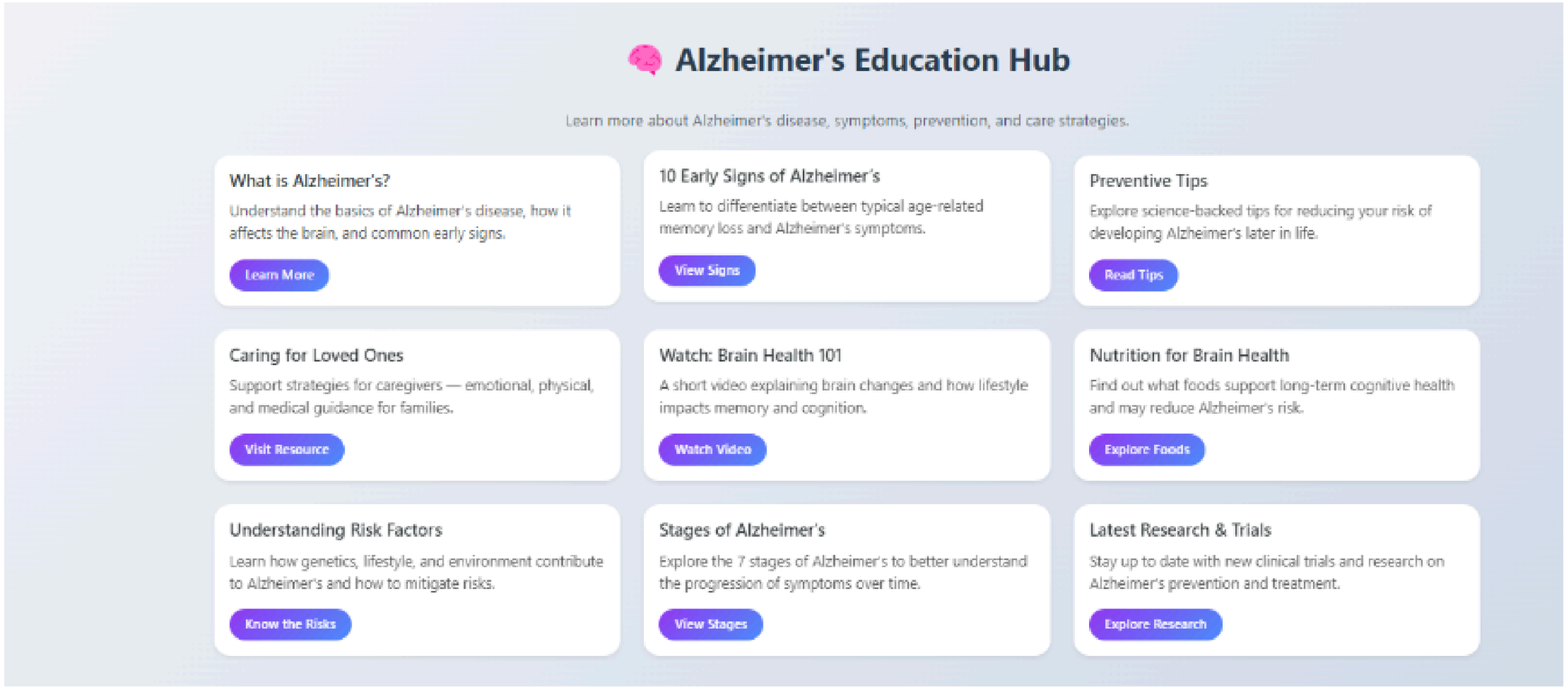
RISK ASSESSMENT PAGE



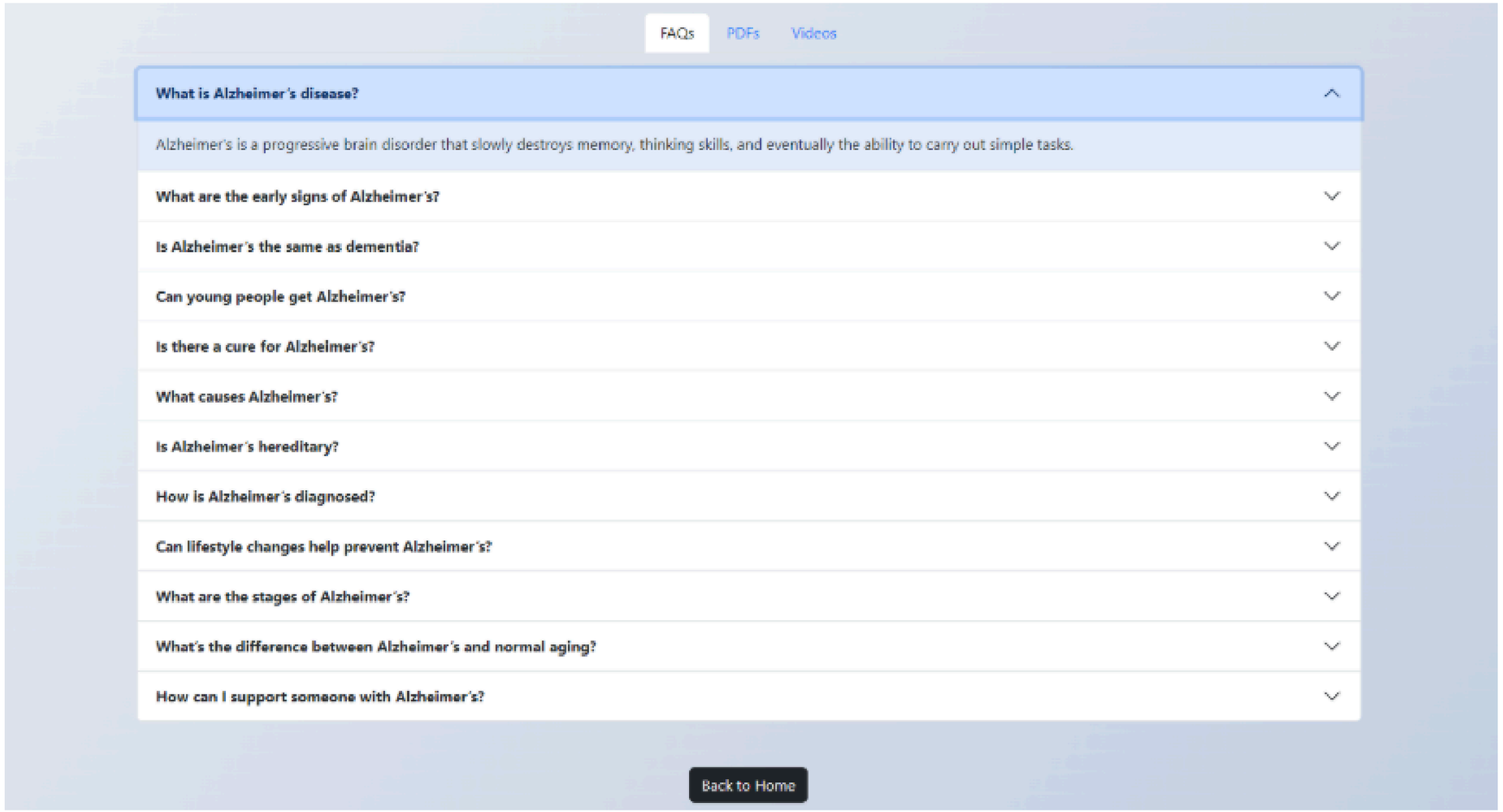
RECOMMENDATIONS PAGE



KNOWLEDGE CENTER PAGE



FAQS SECTION



ABOUT PAGE

Alzheimer's Risk Assessment and Recommendations

Know your risk and take steps towards a healthier brain.

Purpose

This tool helps assess the risk of developing Alzheimer's Disease using machine learning. We analyze various health and lifestyle factors to raise awareness and support early action.

How It Works

You provide data like age, BMI, blood pressure, and cognitive scores. Our intelligent model processes this and gives a risk assessment — helping you stay informed.

Personalized Recommendations

Based on your provided values, we generate personalized recommendations. These suggestions aim to reduce modifiable risk factors and support brain health.

Technologies Used

- Python - Machine Learning Model
- Flask - Backend API Handling
- HTML,CSS,Bootstrap - Styling

Disclaimer

Results shown are based on machine learning algorithms trained on historical data.
This application offers personalized insights derived from data-driven models based on the information you provide.
It is not intended to replace professional medical advice, diagnosis, or treatment. For any health-related concerns, please consult a qualified healthcare provider.

Dataset

https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset

Back to Home

VIDEOS SECTION

FAQs | PDFs | Videos

What is Alzheimer's Disease?

A quick intro to Alzheimer's, its symptoms, and how it progresses in the brain.

Causes & Stages

In-depth info on causes, symptoms, diagnosis, and stages of Alzheimer's.

Managing Dementia

How to manage the progression of dementia across its different stages.

Dementia & Alzheimer's Disease

An informative video explaining the differences between dementia and Alzheimer's, including biological insights.

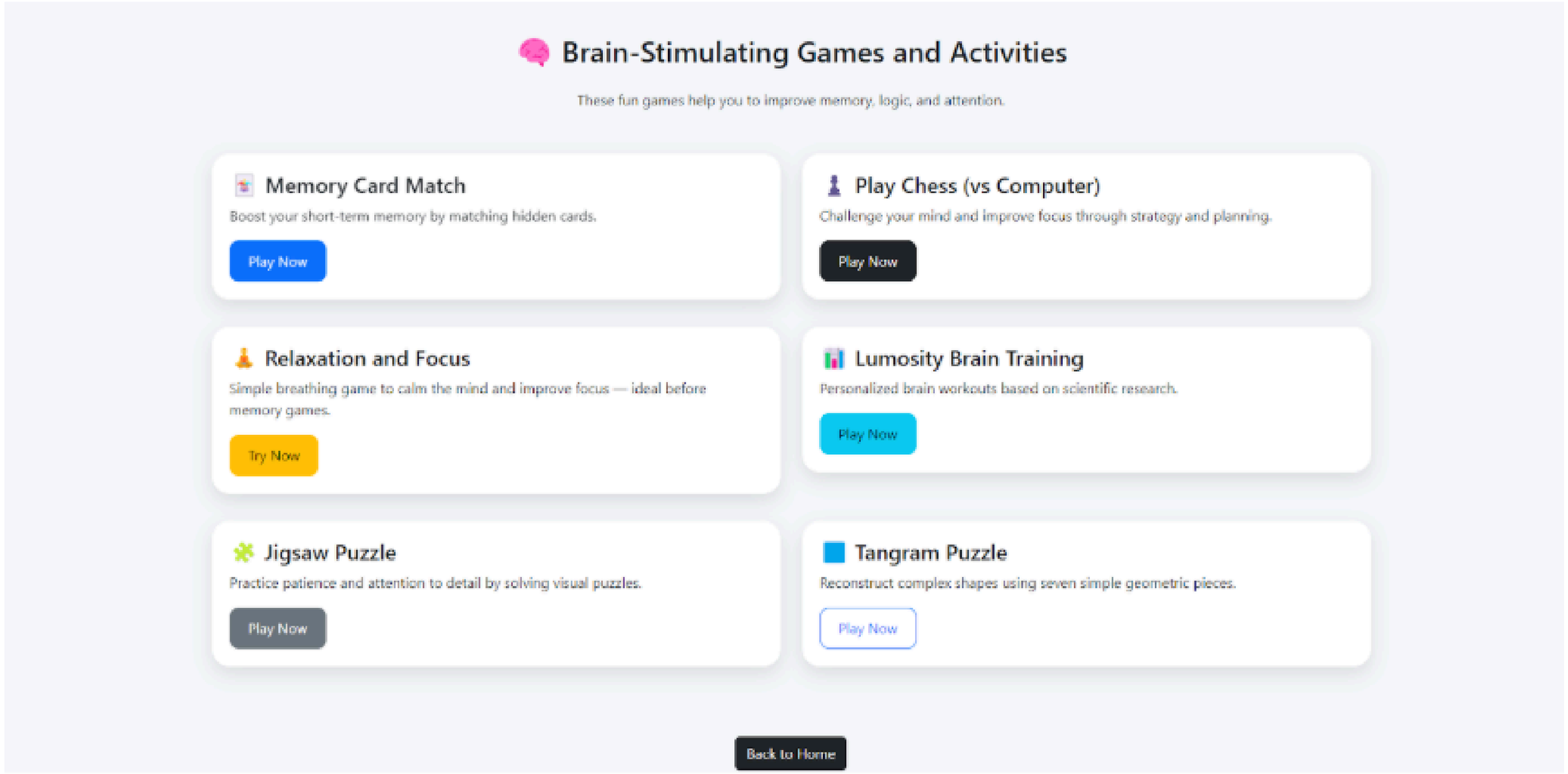
Alzheimer's Awareness - Warning Signs

A video highlighting early warning signs of Alzheimer's disease and the importance of early detection.

What is Alzheimer's Disease and Why Does it Happen?

An exploration into the causes of Alzheimer's disease and its impact on individuals and families.

BRAIN GAMES PAGE



WEBAPP WALKTHROUGH VIDEO

AlzAware

Home
Risk Assessment
Get Recommendations
Brain Games
Knowledge Center
About

AlzAware

Assess, Empower, and Play for a Healthier Mind

Welcome

Our platform helps assess cognitive health by analyzing your input and guiding you with personalized recommendations. It serves as an early indicator tool—not a diagnosis—but a step toward better brain health awareness.

How It Works

- Click **"Risk Assessment"** to begin the assessment.
- **Get recommendations** tailored to your lifestyle and health inputs.
- Explore **"Brain Games"** to boost memory, logic, and attention through fun activities.

The Power of Early Risk Awareness

Early awareness of Alzheimer's risk empowers individuals to make meaningful changes. By identifying risk factors early, you can adopt proactive habits, improve mental fitness, and take control of your cognitive health journey.

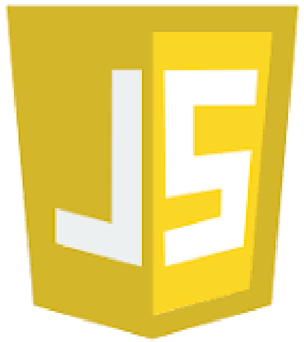
Risk Assessment
Early Insight for a Healthier Tomorrow

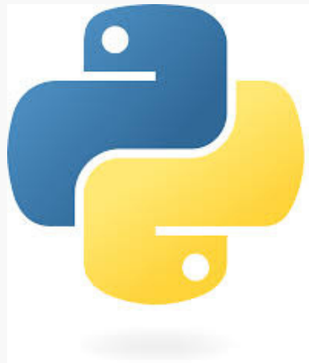
Get Recommendations
Personalized health tips and suggestions

Brain Games
Engaging puzzles and memory boosters

About
Learn more about our mission and our website

TECH STACK USED





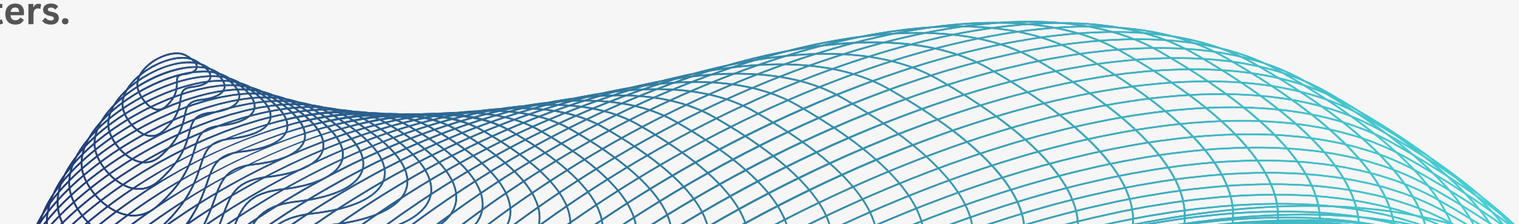
Flask

XGBoost

Render

CONCLUSION

Through iterative refinement, we improved both training and testing performance. By selecting key features and increasing the number of estimators from 100 to 300, training accuracy reached 98.72%, while testing accuracy stabilized around 97.44%. This process balanced model fit and efficiency, ensuring strong generalization for real-world predictions. To effectively support Alzheimer's patients, a personalized recommendation system based on key health parameters is essential. By analyzing various factors recommendations that improve overall well-being. XGBOOST is integrated to generate predictions, risk assessment and providing recommendations based on parameters.



THANK YOU


Column, bar, and pie charts compare values in a single category, such as the number of products sold by each salesperson. Pie charts show each category's value as a percentage of the whole.
Fundraiser Results by Salesperson
PARTICIPANT
UNITS SOLD
Column Chart
Pie Chart