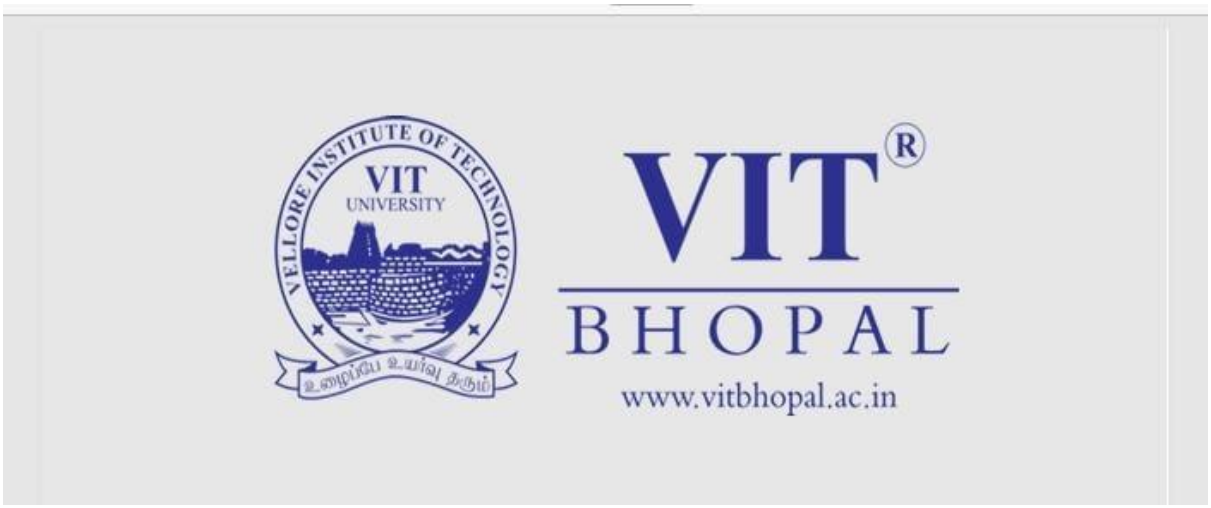

PROJECT REPORT

ALZHEIMER'S DISEASE DETECTION

Alzheimer's Disease Detection Using Patient Data

Tanisha Deshpande

CONTENTS

1)INTRODUCTION

Alzheimer's Disease (AD) is a progressive neurological disorder that affects memory, cognition, and behavior. Early detection is important for timely interventions. With the availability of longitudinal patient data, machine learning can be developed to predict dementia using clinical and demographic features without relying on expensive MRI scans.

This model develops a machine learning pipeline to predict whether a patient is demented or non-demented using tabular clinical data.

2)PROBLEM STATEMENT

here is a dataset containing clinical features of patients along with a label indicating if the patient is demented or not. The goal of this model is to

- **Train a machine learning model that identifies patterns in the data.**
- **Predict dementia for new patients.**
- **Identify the important features influencing dementia risk.**

This enables early detection and aids healthcare providers in decision-making.

3)PREREQUISITES: FUNCTIONAL

1. Data Loading Module: Load patient data from CSV or Excel.

2. Data Preprocessing Module: standardize numerical features.

3. Model Training Module: Train a RandomForest classifier using processed data.

4. Evaluation Module: Analyze the model using accuracy, confusion matrix, and classification report.

5. Prediction Module: Predict dementia for new patient data.

PREREQUISITES:NON-FUNCTIONAL

1. Performance: Model should train efficiently.

2. Usability:easy to run in Colab or local environment

3. Reliability: Handles missing data and unexpected formats (CSV/Excel).

4. Maintainability: Modular code structure for easy changes

5. Scalability: Pipeline can handle large datasets

6. Error Handling: Alerts for incorrect file formats or missing columns.

4)SYSTEM ARCHITECTURE

1. Input Layer: Patient data (longitudinal.csv)

2.Processing Layer: Preprocessing, encoding, scaling

3.ML Layer: RandomForest classifier

4.Output Layer: Accuracy,confusion matrix, feature importance, predicted labels.

5)DECISION MAKING AND RATIONALE

RandomForest: Works well with tabular data, robust to overfitting, interpretable feature importance.

StandardScaler: Makes sure all numerical features contribute equally to the model.

One-hot encoding: Handles categorical variables without introducing ordinal bias.

Modular design: Easier to maintain, debug, and extend in the future.

6)IMPLEMENTATION DETAILS

- Language: Python 3.x
- Libraries: pandas, numpy, scikit-learn, matplotlib, openpyxl, xlrd
- Modules:
  1. data_loader.py – Loads CSV/Excel
  2. preprocess.py – Cleans and encodes data
  3. model.py – Trains RandomForest
  4. evaluate.py – Computes metrics
  5. visualizations.py – Plots feature importance
  6. main.py – Pipeline orchestration

7)TESTING AND OUTPUT

Group	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
Nondemer	1	0	M	R	87	14	2	27	0	1987	0.696	0.883
Nondemer	2	457	M	R	88	14	2	30	0	2004	0.681	0.876
Demented	1	0	M	R	75	12		23	0.5	1678	0.736	1.046
Demented	2	560	M	R	76	12		28	0.5	1738	0.713	1.01
Demented	3	1895	M	R	80	12		22	0.5	1698	0.701	1.034
Nondemer	1	0	F	R	88	18	3	28	0	1215	0.71	1.444
Nondemer	2	538	F	R	90	18	3	27	0	1200	0.718	1.462
Nondemer	1	0	M	R	80	12	4	28	0	1689	0.712	1.039
Nondemer	2	1010	M	R	83	12	4	29	0.5	1701	0.711	1.032
Nondemer	3	1603	M	R	85	12	4	30	0	1699	0.705	1.033
Demented	1	0	M	R	71	16		28	0.5	1357	0.748	1.293
Demented	3	518	M	R	73	16		27	1	1365	0.727	1.286
Demented	4	1281	M	R	75	16		27	1	1372	0.71	1.279
Nondemer	1	0	F	R	93	14	2	30	0	1272	0.698	1.38
Nondemer	2	742	F	R	95	14	2	29	0	1257	0.703	1.396
Demented	1	0	M	R	68	12	2	27	0.5	1457	0.806	1.205
Demented	2	576	M	R	69	12	2	24	0.5	1480	0.791	1.186
Demented	1	0	F	R	66	12	3	30	0.5	1447	0.769	1.213
Demented	2	854	F	R	68	12	3	29	0.5	1482	0.752	1.184
Nondemer	1	0	F	R	78	16	2	29	0	1333	0.748	1.316
Nondemer	2	730	F	R	80	16	2	29	0	1323	0.738	1.326
Nondemer	3	1598	F	R	83	16	2	29	0	1323	0.718	1.327
Nondemer	1	0	F	R	81	12	4	30	0	1230	0.715	1.427
Nondemer	2	643	F	R	82	12	4	30	0	1212	0.72	1.448
Nondemer	3	1456	F	R	85	12	4	29	0	1225	0.71	1.433
Demented	1	0	M	R	76	16	3	21	0.5	1602	0.697	1.096
Demented	2	504	M	R	77	16	3	16	1	1590	0.696	1.104
Demented	1	0	M	R	88	8	4	25	0.5	1651	0.66	1.063
Demented	2	707	M	R	90	8	4	23	0.5	1668	0.646	1.052
Nondemer	1	0	M	R	80	12	3	29	0	1783	0.752	0.985
Nondemer	3	617	M	R	81	12	3	27	0.5	1814	0.759	0.968
Nondemer	4	1861	M	R	85	12	3	30	0	1820	0.755	0.964

```
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        11
           1       0.94      1.00      0.97        32
           2       0.76      0.97      0.85        32

    accuracy                           0.84        75
   macro avg       0.57      0.66      0.61        75
weighted avg       0.72      0.84      0.78        75
```

```
Data Preview:
  Subject ID         MRI ID         Group  Visit  MR Delay M/F Hand  Age
0  OAS2_0001  OAS2_0001_MR1  Nondemented      1         0   M    R   87
1  OAS2_0001  OAS2_0001_MR2  Nondemented      2       457   M    R   88
2  OAS2_0002  OAS2_0002_MR1     Demented      1         0   M    R   75
3  OAS2_0002  OAS2_0002_MR2     Demented      2       560   M    R   76
4  OAS2_0002  OAS2_0002_MR3     Demented      3      1895   M    R   80

   SES  MMSE  CDR  eTIV   nWBV    ASF
0  2.0  27.0  0.0  1987  0.696  0.883
1  2.0  30.0  0.0  2004  0.681  0.876
2  NaN  23.0  0.5  1678  0.736  1.046
3  NaN  28.0  0.5  1738  0.713  1.010
4  NaN  22.0  0.5  1698  0.701  1.034
```

8)CHALENGES FACED

1. **Excel file format not standard and required many engines (openpyxl / xlrd)**
2. **Missing values in some rows; solved by dropna**
3. **Aligning new patient input with model's one-hot encoded features**

9)KEY TAKEAWAY

1. **ML can predict dementia using tabular clinical data.**
2. **RandomForest provides interpretable information.**
3. **Proper preprocessing is crucial for model performance.**
4. **Data format consistency is important for pipeline automation.**

10)REFERENCES

scikit-learn documentation: https://scikit-learn.org/stable/

3. Pandas documentation: https://pandas.pydata.org/

4. Matplotlib documentation: https://matplotlib.org/

5. Openpyxl documentation: https://openpyxl.readthedocs.io/