American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Machine Learning Approach to Improve Prediction Accuracy of Alzheimer's Disease

Chaiyaporn Mutsalklisana[a]*, Kishore Mohan[b], Akshaya Nagarajan[c], Priyanka Mishra[d]

[a,b,c,d]*Northeastern University, Graduate School of Engineering, Information Systems Program, 130 Snell Engineering Center, 360 Huntington Avenue, Boston 02115, USA*

[a]*Email: c.mutsalklisana@neu.edu*

Abstract

Alzheimer's is a chronic neurodegenerative disease developed due to multiple cognitive deficits that progressively leads to at least one of the following: apraxia, aphasia, agnosia or a disturbance in executive functioning. As of 2012, more than 5.1 million Americans are affected by Alzheimer's. Alzheimer's disease accounts for 60 to 80 percent of dementia cases. Numerous pharmaceutical market leaders attempt on developing a cure for the disease. Significant progress has been made on this field. However, studies have shown that manual assessment of the disease using various parameters including (but not restricted to) Neuroimaging (MRI, PET, etc.), Neuropsychological tests (MMSE, FAQ, GDS, NPI, etc.) and Neurogenetics (TOMM40 gene assessment) yield an accuracy of 96% only. Our study involved integrating all three results and allowing the system to predict whether a patient is suffering from Alzheimer's.

Keywords: Alzheimer's medical care; Machine Learning approach to improve prediction accuracy; Integrating neuroimaging; neuropsychological; neurochemical and neurogenetics results for effective assessment.

1. Introduction

Accurate analysis of Alzheimer's can aid physicians to provide proper healthcare support and medication for patients suffering from Alzheimer's. Alzheimer's is generally characterized by synthesis of large quantities of intraneuronal neurofibrillary tangles containing Tau proteins and extracellular amyloid plaques containing the peptide β amyloid. Drugs like Memantine.

---

* Corresponding author.

Donepezil, Rivastigmine and galantamine are generally prescribed for Alzheimer's patients. Memantine reduces the effects of glutamate, which is produced in larger quantities in patients suffering from Alzheimer's [2]. Prescribing memantine for patients misdiagnosed with Alzheimer's will have severe side-effects. Patients having Mild Cognitive Impairment (MCI) and are above the age of 45 are often misinterpreted with Alzheimer's. Separate studies reveal that 18.18% of the patients misdiagnosed with Alzheimer's are often given wrong medication.

The paper discusses how the machine learning approach can help overcome these problems:

- How do we interpret significant results in terms of numbers from neuropsychological, neurogenetic and neuroimaging tests?
- How do we develop a machine learning model that utilizes the three test results to improve the accuracy of Alzheimer's prediction?

High level architecture of the system

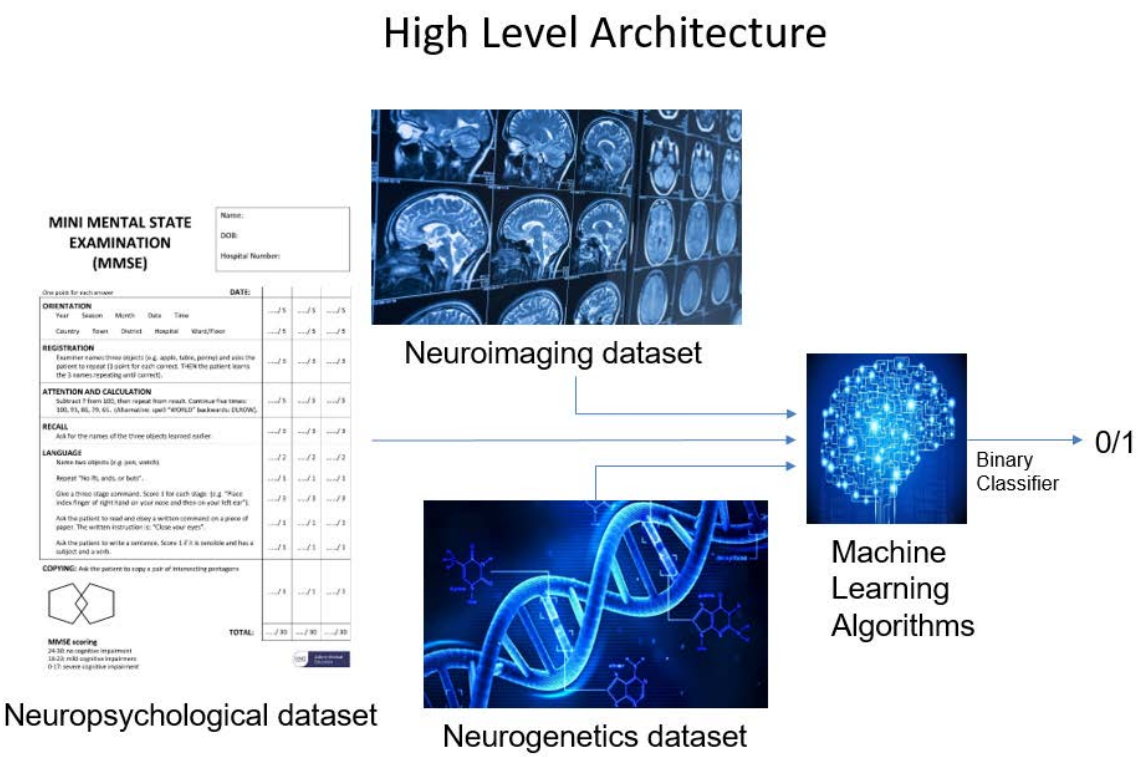

Figure 1: High level architecture of the machine learning model

1.1. Objectives of this scientific paper

- Use algorithms to process the neuroimaging, neuropsychological and neuro-genetic datasets.

- Devise a system to integrate significant features from the three datasets.

- Predict Alzheimer's using the random forest classifier hypothesis.

2. Using algorithms to process the three datasets

The system will consider four primary neuropsychological test scores suggested by Alzheimer's association. The four tests include:

- **MMSE : Mini Mental State Exam**

Scores range from 0 to 30. Less than 26 indicates Alzheimer's

- **FAQ : Functional Activities Questionnaire**

Response for 10 questions are recorded. Each question takes values ranging from 0 to 3 [3].

Scores range from 0 to 30. A cut-point of 9 (dependent in 3 or more activities) is recommended.

• Dependent = 3 • Requires assistance = 2 • Has difficulty, but does by self = 1 • Normal = 0

- **GDS : Geriatric Depression Scale**

Response for 10 questions are recorded. Each question takes values ranging from 0 to 3.

• 1–4 No cause for concern • 5–9 Strong probability of depression • 10+ Indicative of depression

- **NPI : Neuropsychiatric Inventory**

• Hallucinations • Agitation/Aggression • Depression/Dysphoria • Aberrant motor behavior [4]

• Sleep and Nighttime Behavior Disorders • Appetite and Eating Disorders

Gene results from TOMM40 A1, dominant allele and TOMM40 A2, recessive allele of TOMM40 gene are processed to identify the significant gene that contributes to the model. Hippocampus size difference from MRI scans of the brain is taken as significant neuroimaging input parameter to the model.

2.1 Identifying significant test from the neuropsychological dataset

The system performed Principal Component Analysis for dimensionality reduction on the four neuropsychological tests including MMSE, NPI, FAQ and GDS of patients taken over different periods of time.

Inference:

Our studies with 240 patients who took all 4 tests at different time intervals has revealed that FAQ, that determines the functional ability of the patient is more significant than MMSE, that determines the cognitive

ability of the patient. Subsequently, likelihood of a patient suffering from Alzheimer's is higher when functional activities are heavily disturbed.

Input parameter to the system: Normalized FAQ scores given as the first input.

2.2 Identifying significant allele from the TOMM40 gene

Neurogenetics result set for different patients assessed from TOMM40, which was the primary Biomarker [5] for neurogenetic assessment was taken into consideration. TOMM40 Allele-1 and TOMM40 Allele-2 was clustered to identify the greater range distribution.

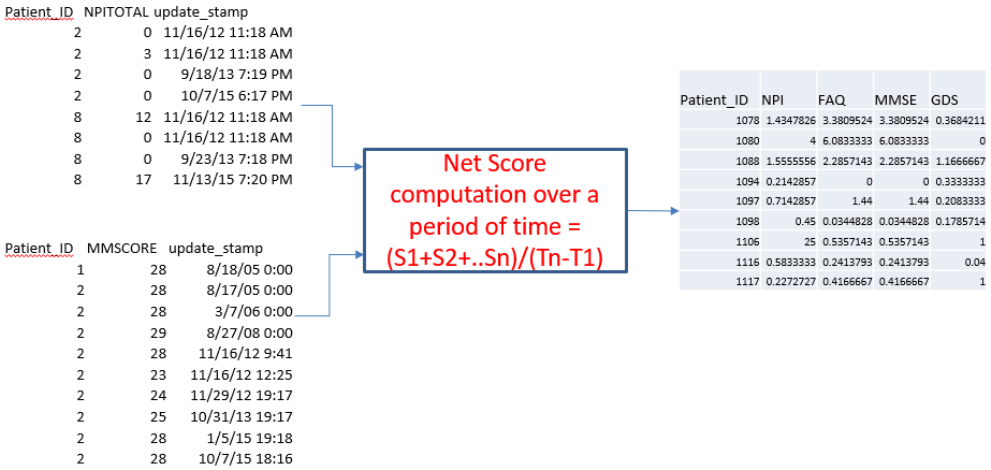

Figure 2: Formula to compute net test scores over a period of time

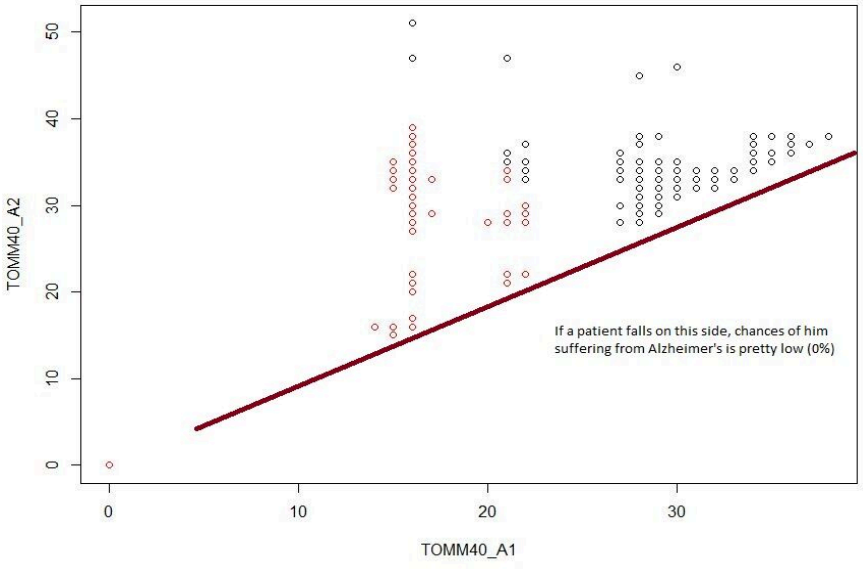

Figure 3: K-Means cluster on TOMM40 dataset

Inference:

- Red cluster indicates patients falling in level 1-3
- Black cluster indicates patients in level 4-7
- TOMM40 A2, recessive gene contributes greater range in protein synthesis

Input parameter to the system: TOMM40 A2 given as the second input.

2.3 Calculating the hippocampus size difference

The neuron degradation results in overall size shrinkage of human brain. The degradation initiates from hippocampus [6], which keeps track of all human memories and progresses slowly towards all parts of the brain. Since hippocampus is primarily affected, the region of MRI scan that holds hippocampus alone is segmented to identify size deviation in an individual over a period of time [7].

The following steps are involved in the process:

1. Neuro-image segmentation using Convolutional Neural Network (CNN) algorithm [8]
2. Laplacian transform to normalize segmented image
3. Detection of edges on normalized image
4. Detection of corners with edges detected
5. Calculation of size difference with corners detected.

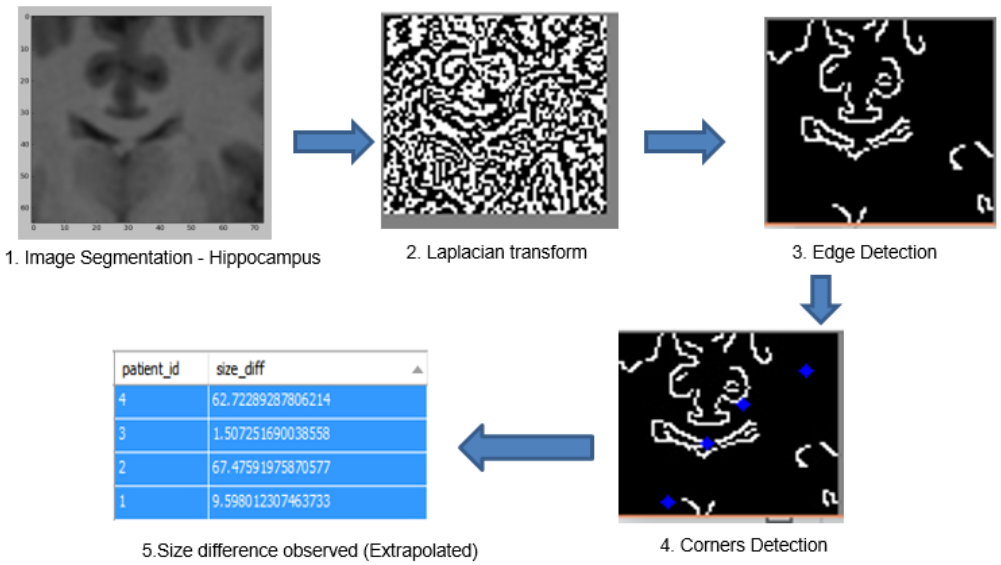

Figure 4: Steps involved in processing MRI images

Input parameter to the system: Size difference observed

3. Final Feed to the model

Problem Type:

Supervised Classification

Inputs: Normalized FAQ scores, TOMM40_A2 results, size difference

Outputs: Binary classifier (0/1)

Independent classifier was highly decorrelated with dependent features and hence algorithms that best fit the decorrelated dataset was chosen. The model was trained with Random Forests algorithms with close to 244 patient records. .

Hypothesis:

$$P(c|v) = 1/T \sum_{t=1}^{T} p_t(c|v)$$

Table 1: Training data summary

Training Patient record count	844
Patients with Alzheimer's	630
Patients without Alzheimer's	214

Table 2: Test data summary

Test Patient record count	200
Actual Patients with Alzheimer's	194
Actual Patients without Alzheimer's	6

Table 3: Machine predicted data summary

Test Patient record count	200
Predicted Patients with Alzheimer's	193
Predicted Patients without Alzheimer's	7

4. Results

A test set of 200 patients with 194 affected and 6 non-affected was fed to the system and the model predicted 193 records accurately.

Error percentage = 0.83%

Prediction Accuracy = 99%

System proposes a new technique of integrating the neuroimaging, neuropsychological and neuro-genetic datasets of a patient and has a successful prediction rate of 99% (z-value: -1.5)

5. Conclusion

The system will assist physicians to predict Alzheimer's accurately.

Overall benefits achieved from the system include:

- Isolating significant features from least significant tests
- Integrating these significant isolated features
- Interpreting Alzheimer's in a patient

Limitations of the system:

- Severity of the disease (Mild or moderate or severe) cannot be assessed by the system
- Limited spatial resolution of MRI scans might result in slight deviation of size differences observed

6. Future

Apart from the three datasets, the system can be trained with neurochemical test results and neurophysiological test results [9]. The system can also be trained with clinical trials performed on various patients. Adding all these parameters to the system will improve the prediction accuracy further.

Acknowledgments

[1] ADNI – Alzheimer's disease neuroimaging initiative for providing us the neuropsychological, neuroimaging and neurogenetics dataset.

References

[1] O. Kosunen, H. Soininen, L. Paljärvi, O. Heinonen, S. Talasniemi, P. J. Riekkinen Sr (1996, February). "Diagnostic accuracy of alzheimer's disease: A neuropathological study" Springer-Verlag, pp.91-185.

[2] Joseph E Gaugler, Haya Ascher-Svanum, David L Roth, Tolulope Fafowora, Andrew Siderowf and Thomas G Beach (2013, December). "Characteristics of patients misdiagnosed with Alzheimer's disease and their medication use: an analysis of the NACC-UDS database" BMC Geriatrics, pp.13-137.

[3] "Alzheimer's Society Assessment and Diagnosis sheet" Internet: https:alzheimers.org.uk

[4] "Neuropsychiatric Inventory Assessment" Internet: https:dementia-assessment.

[5] A D Roses, M W Lutz1, H Amrine-Madsen, A M Saunders, D G Crenshaw, S S Sundseth, M J Huentelman, K A Welsh-Bohmer, and E M Reiman(2009, December). "A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease" The Pharmacogenomics Journal, pp.375-384.

[6] Olfa Ben Ahmed, Jenny Benois-Pineau, Michele Allard, Chokri Ben Amar, Gwenalle Catheline. Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. Multimedia Tools and Applications, Springer Verlag, 2014, pp.35.

[7] S. Lia et al (2007, January), "Hippocampal Shape Analysis of Alzheimer Disease Based on Machine Learning Methods" American Journal of Neuro Radiology. [Electronic] Available: http://www.ajnr.org/content/28/7/1339.full.pdf+html

[8] Jieping Ye et al (), "Machine Learning Approaches for the Neuroimaging Study of Alzheimer's Disease" Al Redux [Electronic] Available: http://www.lifesciences.ieee.org/images/pdf/machinelearning11112011.pdf

[9] Daoqiang Zhang et al (2011, October), "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease" Elsevier. [Electronic] Available: http://www.sciencedirect.com/science/article/pii/S105381191101144X