BANGLADESH UNIVERSITY OF BUSINESS AND TECHNOLOGY

Proposal Summary

Proposal Title:	Predicting Alzheimer's disease using machine learning
Supervisor:	**Dr. M Firoz Mridha**
Affiliation:	**Bangladesh University of Business and Technology**
Department:	**Department of Computer Science and Engineering**
Area of Specialization:	Artificial intelligence, Machine Learning

Team Members

Musfiquer Rhman 17181103098

Md. Mintu Hossain 17181103110

Farjana Rahaman 16173103013 (36)

Umma Habiba Emu 17181103106

Khadija Akter 17181103089

Abstract:

Modeling and prediction of chronic diseases like Alzheimer's disease have become a major concern in recent years. Traditionally this field mostly relies on MRI images, demographics, collecting statistical data, etc. Recent progression in machine learning has started a new era of disease prediction and modeling. This leads to exploring and designing machine learning techniques for processing medical datasets to predict occurrences of Alzheimer's disease. This prediction includes identifying and classifying Alzheimer's disease. Many machine learning algorithms like Decision tree classifiers, Independent Component Analysis, Support Vector Machine, Linear Discriminant Analysis were used to predict the disease, but the precision is not good enough. In this paper, we proposed an ensemble architecture-based Machine Learning model to predict Alzheimer's disease.

Table of Contents

3. Introduction

Alzheimer's disease is a progressive neurologic disorder, it causes the brain to shrink and brain cells to die. Someone is diagnosed with Alzheimer’s disease every four seconds. Alzheimer’s disease is one of the most common forms of dementia affecting millions of old people worldwide. Dementia refers to diseases that can be characterized by a loss of memory or other cognitive impairments. It is caused by damage to nerve cells within the brain. It becomes more severe over time and it’s not curable.

The damage begins in the region of the brain which is responsible for controlling memory. The process starts several years before the first symptom. The loss of the neuron spreads to other regions and later the brain shrinks significantly. Alzheimer’s disease has three stages. Mild, moderate, and severe stages. In the mild stage of Alzheimer’s disease, a person can function normally but he will face problems like coming up with the right name or word, difficulty performing social tasks, forgetting things just after reading it.

The longest stage of Alzheimer’s disease is the middle stage. At this stage, the person needs extra care to function properly. Symptoms at this stage are like, forgetting personal history, name, address, etc. Confusion about where they are, change in sleep pattern, etc. At the severe stage, people lose the ability to respond to their environment, forget about their recent experience and lose awareness about their surroundings, have difficulty communicating, etc.

Early detection of Alzheimer's disease is very helpful to start proper treatment. If the disease is predicted earlier, the progression of the symptoms of the disease can be slow down and can save lives. Alzheimer’s disease can be detected early by magnetic resonance imaging (MRI), it uses a magnetic field and radio waves to create a 3D image of the brain. Or, by analyzing statistical data and Mini-Mental State Exam. Many machine learning methods have been tested for this task in recent years, including support vector machines, independent component analysis, and penalized regression. Studies indicate that computer-based diagnosis is on par or better than diagnosis by radiologists [1].

In this paper, we proposed a new architecture based on Machine Learning. We will implement the model in the Cloud environment for reducing computational complexity.

4. Literature Review

Predicting Alzheimer's disease is a popular research field for researchers. A great amount of work has been done in this field.

XIN HONG et. al [2] focused on identify time relative biomarkers associated with disease status, these biomarkers are cortical Thickness Average (TA), WMParcellation Volume (SV), Surface Area (SA), cortical Thickness Standard deviation (TS), and Cortical parcellation Volume (CV). They found that the Cortical Thickness Average (TA) is significant in predicting Alzheimer's disease progression. They propose a predicting model based on Long short-term memory (LSTM), which might be able to connect previous information to the present task. They noticed that the temporal data for a patient are important for predicting the development of the disease.

Simon F. Eskildsen et al. [3] tried to predict AD using patterns of cortical thickness measurements in subjects with mild cognitive impairment (MCI). They were able to identify individuals with MCI who progress to AD and individuals with MCI who do not progress to AD. They also identified Specific patterns of atrophy and selected features as regions of interest from these patterns. Their prediction accuracies were artificially inflated to a range of 73% to 81%.

Gopi Battineni et. al [4] also used MRI data to develop different ML models to predict dementia in the elderly. Subjects going through brain changes, like leukoaraiosis, mild atrophy, and regular dementia cases of AD, were included in this study. They analyzed fourteen features related to AD and 373 MRI tests of 150 subjects. They trained four Machine Learning models, naive Bayes (NB), support vector machines (SVM), K-nearest neighbor (KNN), and artificial neural networks (ANN). The combination of all four models with selective features enhanced the accuracy of dementia prediction. They got 91.32% accuracy with manual feature selection with 1NN and 96.12% accuracy by SVM with automatic feature selection. The combination of all four models with selective features increased the accuracy to 98%.

Yanyan Lin et. al. [5] studied to develop a longitudinal structural magnetic resonance imaging-based prediction system for MCI (mild cognitive impairment) progression. They collected longitudinal data from 164 MCI patients. They used a discriminative dictionary learning framework to distinguish MCI patches, avoiding the segmentation of regions of interest. They predicted a new subject with fourfold cross-validation (CV), and the area under the receiver operating characteristic curve (AUC) was determined. They achieved an accuracy of 97% and AUC values after fourfold CV was 98.4%.

Siqi Liu et. al. [6] designed a deep learning architecture of stacked auto-encoders and a SoftMax output layer and applied them on MRI images from ADNI. They analyzed multiple classes like demented, non-demented, and converters in one setting. They could do it with less minimal domain prior knowledge and labeled training samples. They produced a better overall accuracy (87.76%) in the binary classification of AD.

Anees Abrola et. al. [7] studied the progression from mild cognitive impairment (MCI) to Alzheimer's disease (AD) by modifying deep residual neural networks. They trained the deep learning models using mild cognitive impairment individuals only, then used a domain transfer learning version and trained additionally on AD and controls. They achieved a test classification accuracy of 83.01 %.

Garam Lee et. al [8] developed a framework that combines cross-sectional neuroimaging biomarkers at baseline, longitudinal cerebrospinal fluid (CSF), and cognitive performance biomarkers obtained from ADNI. They took advantage of the longitudinal and multi-modal nature of available data for discovering nonlinear patterns associated with MCI progression. The proposed framework integrates longitudinal multi-domain data. The biggest advantage of their approach is that irregular longitudinal data can be used. Their model achieved 81% accuracy when incorporating longitudinal multi-domain data.

Minh Nguyena et. al. [9] showed that the minimalRNN model was better than other baseline algorithms for the longitudinal prediction of multimodal AD biomarkers and clinical diagnosis of participants up to 6 years into the future. They explored three different strategies to handle the missing data issue widespread in longitudinal data. they found that the RNN model can itself be used to fill in the missing data, thus providing an integrative strategy to handle the missing data issue., They found that after training the RNN model can perform reasonably well using one input time point with longitudinal data, suggesting the approach might also work for cross-sectional data.

Baiying Lei [10] et. al introduce deep and joint learning along with a two scenarios regression model for AD scores prediction. Different from the commonly used scores prediction methods which focus on machine learning or deep learning based on baseline dataset, they obtain the predicted scores at the next time point from previous time points datasets. Specifically, they integrate the feature selection with fused smoothness term and employ the correntropy to construct the joint learning model. Meanwhile, the DPN algorithm is proposed to further improve feature representation, and then SVR is applied to predict four types of clinical scores. Despite the good performance of the proposed model, there are also several limitations of their current study. The relevant clinical details like gender, age, education level, and other physiological factors of AD were not taken into account in the experiments.

Courtney Cocherane et. al. [11] analyzed different pre-processing methods, machine learning models, and feature selection techniques. Instead of using MRI data, they used longitudinal lifestyle interventions. They achieved more than 90% accuracy and recall in predicting Alzheimer’s disease. They produced a "lean" diagnostic protocol that can predict AD development in someone with only 3 tests and 4 clinical visits with 87% accuracy and 79% recall.

Manu Subramoniam et. al. [12] proposed the classification of Alzheimer’s disease based on a deep neural-network using Magnetic Resonance Images (MRI) as input for the classification task. They have proved that among the VGG architectures, the VGG-16 performed better than VGG-19. Among the residual neural network architectures, Resnet-18 was more accurate than Resnet-101.

Adrien Payan et. al. [13] was able to discriminate between a healthy brain and a diseased brain by analyzing magnetic resonance imaging images as input. They used 3D convolutions on the whole MRI image which yield better performance than 2D convolutions on slices in our experiments. They also found that a 3D approach may boost the classification performance by a small margin. The accuracy was 89.47%.

5. Problem Statement

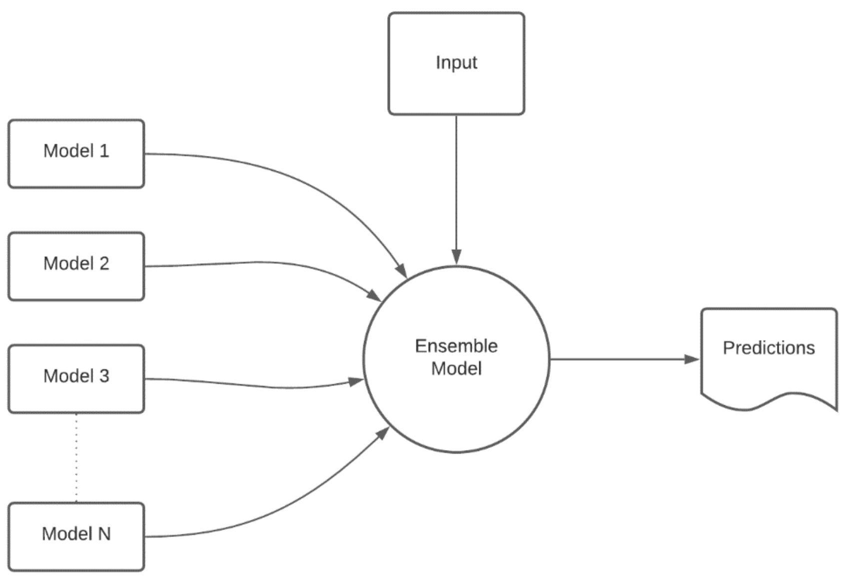
Predicting Alzheimer’s disease early has a significant effect on the treatment and minimizing the damage in patients. But most of the studies in this field focus on the classification of the current stage rather than predicting future effects. And others rely heavily on MRI methods. Which are costly. In our research, we will use machine learning to predict Alzheimer’s disease early in a patient in a cheaper way.

6. Research Objectives

The Following research objectives can be achieved from this research.

- Early prediction of Alzheimer's disease in a person.
- Minimize the necessary diagnostic test and clinical visits need for a patient while maximizing accuracy.
- A new Ensemble Architecture based Machine Learning architecture will be given.

7. Methodology



In this study, a Machine Learning based Ensemble Architecture will be made for the early prediction of Alzheimer's disease. Ensemble Method is a machine learning technique where several Machine learning models are used to produce an optimal model. Instead of depending on any single model, we will propose a hybrid model to make an optimal model. ADNI dataset will be used. Alzheimer's Disease Neuroimaging Initiative (ADNI) is a multisite study that aims to improve clinical trials for the prevention and treatment of Alzheimer's disease (AD) [17]. The main part of the study is to find out the best, more accurate architecture that performs exceptionally great on the dataset. Finally, the architecture will be given to the cloud environment for reducing computational complexity and to find out its best result.

8. Proposed model:

Most of the research in this field focuses on analyzing Magnetic Resonance Imaging (MRI) and PET data. Which are costly and people in many poor countries don’t have adequate diagnostic tools. So, instead of using MRI data, we are going to focus on longitudinal data. Because the cost of medical imaging deters people from testing Alzheimer’s disease.

There are several factors in dementia, Age is the strongest factor. Gender is also an important factor. Females are more like to be affected by AD than Male [14]. Low education level [15] is also a crucial factor. Low social class, income, depression, head trauma, epilepsy, diabetes, and stroke are all relevant factors [16]. All these factors can be used with Mini-Mental State Exam (MMSE) to early diagnosis of AD.

For preprocessing we will first drop all the non-related and incomplete data. Then we will use ‘One hot encoder’ for converting and categorizing our data. Finally, we will use several machine learning models to make an ensemble model that will predict AD in a subject with high accuracy.

Then we will use different machine learning algorithms like Random Forest Classifier, Support Vector Machine, GaussianNB, etc. to find the optimal result.

9. Conclusion

In this study, an Ensemble Architecture based approach will be proposed to predict Alzheimer’s disease. The proposed architecture will have been compared with other Machine learning based architecture. The traditional methods heavily depend on MRI imaging, but prediction can be done without MRI images. We will use Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset for the study. In this research we will predict AD with lifestyle and longitudinal data. Our experiment will show that our approach outperforms the baseline.

References

[1] S. Kloppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, C. C. Roberts, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. Frackowiak, “Accuracy of dementia diagnosis--a direct comparison between radiologists and a computerized method,” Brain, vol. 131, no. 11, pp. 2969–2974, 2008.

[2] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, and J. Yang, “Predicting Alzheimer’s Disease Using LSTM,” IEEE Access, vol. 7, pp. 80893–80901, 2019.

[3] S. F. Eskildsen, P. Coupé, D. García-Lorenzo, V. Fonov, J. C. Pruessner, and D. L. Collins, “Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning,” NeuroImage, vol. 65, pp. 511–521, 2013.

[4] G. Battineni, N. Chintalapudi, F. Amenta, and E. Traini, “A Comprehensive Machine-Learning Model Applied to Magnetic Resonance Imaging (MRI) to Predict Alzheimer’s Disease (AD) in Older Subjects,” Journal of Clinical Medicine, vol. 9, no. 7, p. 2146, 2020.

[5] Y. Lin, K. Huang, H. Xu, Z. Qiao, S. Cai, Y. Wang, and L. Huang, “Predicting the progression of mild cognitive impairment to Alzheimer’s disease by longitudinal magnetic resonance imaging-based dictionary learning,” Clinical Neurophysiology, vol. 131, no. 10, pp. 2429–2439, 2020.

[6] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, “Early diagnosis of Alzheimer's disease with deep learning,” 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), 2014.

[7] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, “Deep residual learning for neuroimaging: An application to predict progression to Alzheimer’s disease,” Journal of Neuroscience Methods, vol. 339, p. 108701, 2020.

[8] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, “Predicting Alzheimer’s disease progression using multi-modal deep learning approach,” Scientific Reports, vol. 9, no. 1, 2019.

[9] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, and B. T. T. Yeo, “Predicting Alzheimer's disease progression using deep recurrent neural networks,” NeuroImage, vol. 222, p. 117203, 2020.

[10] B. Lei, M. Yang, P. Yang, F. Zhou, W. Hou, W. Zou, X. Li, T. Wang, X. Xiao, and S. Wang, “Deep and joint learning of longitudinal data for Alzheimer's disease prediction,” Pattern Recognition, vol. 102, p. 107247, 2020.

[11] Courtney Cochrane, David Castineira1, Nisreen Shiban, Pavlos Protopapas, Application of Machine Learning to Predict the Risk of Alzheimer's Disease: An Accurate and Practical Solution for Early Diagnostics. arXiv:2006.08702v1 [q-bio.QM] 2 Jun 2020.

[12] Manu Subramoniam, Aparna T. R., Anurenjan P. R., and Sreeni K. G. Manu Subramoniam, Aparna T. R., Anurenjan P. R., and Sreeni K. G. Deep learning based prediction of Alzheimer's disease from magnetic resonance images. arXiv:2101.04961v1 [eess.IV] 13 Jan 2021

[13] Adrien Payan1 and Giovanni Montana, Predicting Alzheimer’s disease: a neuroimaging study with 3D convolutional neural networks. arXiv:1502.02506v1 [cs.CV] 9 Feb 2015.8

[14] Raicher, Irina, Marta Maria Shimizu, Daniel Yasumasa Takahashi, Ricardo Nitrini, and Paulo Caramelli. "Alzheimer's disease diagnosis disclosure in Brazil: a survey of specialized physicians' current practice and attitudes." International Psychogeriatrics, Vol 20, no. 3, page 471, 2008.

[15] A. R. Borenstein, C. I. Copenhaver, and J. A. Mortimer, “Early-Life Risk Factors for Alzheimer Disease,” Alzheimer Disease & Associated Disorders, vol. 20, no. 1, pp. 63–72, 2006.

[16] M. A. Lopes, S. R. Hototian, S. E. Bustamante, D. Azevedo, M. Tatsch, M. C. Bazzarella, J. Litvoc, and C. M. Bottino, “Prevalence of cognitive and functional impairment in a community sample in Ribeirão Preto, Brazil,” International Journal of Geriatric Psychiatry, vol. 22, no. 8, pp. 770–776, 2007.

[17] “Alzheimer's Disease Neuroimaging Initiative,” Wikipedia, 15-Mar-2021. [Online]. Available: https://en.wikipedia.org/wiki/Alzheimer%27s_Disease_Neuroimaging_Initiative#cite_note-:2-1. [Accessed: 18-Mar-2021].