Predicting Alzheimer's Disease at Low Cost Using Machine Learning

Musfiquer Rhman[1,2], Farjana Rahman[1,3], Md. Mintu Hossain [1,4], Umma Habiba Emu[1,5], Khadija Akter[1,6], M. F. Mridha[1,7]
[1] Department of Computer Science and Engineering
Bangladesh University of Business and Technology
Emails: [2] musfiquerrhman@gmail.com, [3] frtithi13@gmail.com, [4] mintulikhon79@gmail.com, [5] apurnakhan14@gmail.com, [6] khadijaakter1768@gmail.com, [7] firoz@bubt.edu.bd

Abstract— Modeling and predicting Alzheimer's disease (AD) is a noteworthy field of study for many years. However, diagnosing and treating AD for poor people in many developing and underdeveloped countries is a luxury. Many countries lack the proper tools. Many people do not have the capability of bearing the cost of MRI tests. In this paper, we analyzed a way of identifying AD as cheaply as possible. We investigated the difference in efficiency between the Machine learning models using Magnetic Resonance Imaging (MRI) data versus models without MRI data. We had around a 3% difference between the best models of these tests. We used the OASIS dataset. We used patients' longitudinal lifestyle data like age, gender, education, income, MMSC score, and other features for Non-MRI models. For MRI models, we also analyzed MRI data along with the lifestyle data. We trained different Machine learning models like Random Forest Classifier, GaussianNB, LinearSVC, Logistic Regression, KNeighbors Classifier, Adaboost Classifier, and several other models. After that, we combined our best models and created a hybrid model. Our best result was 96.07% accuracy with MRI data and 93.37% accuracy without MRI data. The non-MRI result may not be as efficient as the models with the MRI dataset and not crucially inefficient. Our result can be summarized as the Non-MRI data can be used for starting the diagnosis for those poor people.

Keywords—Machine learning, Alzheimer's disease. MRI data, Non-MRI data, cheap.

I. INTRODUCTION

Alzheimer's disease is a progressive neurologic disorder that shrinks the brain and kills the brain cells. Someone is diagnosed with Alzheimer's disease every four seconds. Alzheimer's disease is one of the most common types of dementia affecting millions of older people worldwide.

Dementia is a group of disorders characterized by memory loss and other cognitive impairments. Damage to nerve cells in the brain causes it. It becomes worse with time, and there is no way to cure it.

Every country has elderly with dementia. However, the number is alarmingly increasing in the least developing and developing nations. Research [1] has shown that by 2040, 71% of dementia patients will be from developing nations, and the number will double every 20 years. There is low awareness about dementia in those countries. Many people in those countries think the loss of memory is a typical aging problem. The lack of diagnostic assessment [2], cost of diagnosis deters people from diagnosis of dementia. The average cost of MRI in the USA is 1100 USD, the MMSE is 1.23 USD [3]. In Bangladesh, the price of a brain MRI is about 46 USD. In India, it is about 136 USD. So, Brain MRI is significantly more expensive than MMSE, which many families cannot bear.

There are several factors in dementia. Age and gender are also essential factors. Females are more like to be affected by AD than Male [4]. Low education level [5] is also a crucial factor. Low social class, income, depression, head trauma, epilepsy, diabetes, and stroke are relevant factors [6]. All these factors can be used with Mini-Mental State Exam (MMSE) to early diagnosis of AD. The lower cost of diagnosis will open doors for many poor people and people living in rural areas for AD diagnosis.

Early confirmation of Alzheimer's disease may be helpful to start proper treatment. If the disease is predicted earlier, the progression of the symptoms of the disease can be slow down and save lives.

Several Machine learning models were tested for MRI data and nor-MRI data to measure the gap between those two approaches. Analyzing MRI data gives more accurate results but is expensive. Analyzing non-MRI data is less efficient but is cheap. In this research, we utilize the effectiveness of the popular Machine learning algorithms to prove that analyzing non-MRI data is a viable option for starting the AD treatment. Our research can be summarized as,

- We investigated the existing Machine learning algorithms to produce the best result possible for both of our tests, with and without MRI data. Our research is the first research to perform this benchmark.
- We experimented with several Machine learning algorithms and found AdaBoost gives the best result in MRI data analysis, and our Hybrid model gives the best result for non-MRI data analysis..

The rest of this paper is written as follows: In section II, we conducted a literature review. In section III, we described the detailed description of our study, finally, in section IV, the conclusion of our research.

II. LITERATURE REVIEW

Predicting Alzheimer's disease is a popular research field for researchers. A significant amount of work has been done in this field. MRI and PET data analysis is the dominant subfield in predicting Alzheimer's disease. Biomarkers

analysis and longitudinal data analysis for classification and progression are also much popular. A few of them are,

Gopi Battineni et al. [7] used MRI data to construct various machine learning models for predicting dementia in the elderly. Their study included people experiencing brain abnormalities such as mild atrophy, leukoaraiosis, and regular dementia cases of Alzheimer's disease. They trained four Machine Learning models. The combination of all four models with selective features enhanced the Accuracy of dementia prediction. The combination of all four models with selective features increased the Accuracy to 98%

Siqi Liu et al. [8] proposed a stacked auto-encoder deep learning architecture with a SoftMax output and applied it on ADNI MRI images. They could do it with less minimal domain prior knowledge and labeled training samples. They produced an overall Accuracy score of 87.76% in the classification of AD.

Yanyan Lin et al. [9] studied to develop a longitudinal structural magnetic resonance imaging-based prediction system for MCI progression. They collected longitudinal data from 164 MCI patients. To identify MCI patches, they used a discriminative dictionary learning framework instead of segmenting regions of interest. They had a 97 percent accuracy rate.

Courtney Cocherane et al. [10] 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 three tests and four clinical visits with 87% accuracy and 79% recall.

XIN HONG et al. [11] focused on identify time relative biomarkers associated with disease status. 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 connect previous information to the present task.

Garam Lee et al. [12] developed a framework that blends 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 most significant advantage of their approach is that irregular longitudinal data can be used. When they used longitudinal multi-domain data, their model had an accuracy of 81 percent.

Manu Subramoniam et al. [13] proposed classifying 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.

We have reviewed many papers, but none focused on the differences in efficiency between MRI and non-MRI data analysis. In this paper, we did precisely that analysis to determine the efficiency difference

III. Methodology

A. Dataset used

For this study, we used the OASIS [14] longitudinal data. OASIS is the Open Access Series of Imaging Studies (OASIS). This project aims to make neuroimaging data sets of the brain freely available to the scientific community.

OASIS Dataset provides several independent variables including sociodemographic and clinical variables. Sociodemographic variables used in this experiment are:

- Gender: Male = 150, Female = 204.
- Age: Min = 60, Max = 98, Mean = 77.03.
- Years of Education (EDUC): Min = 6, Max =23, Mean = 14.7.
- Socioeconomic Status (SES): Upper = 7, Upper Middle = 74, Middle = 82, Lower Middle = 103, Lower = 88.

Clinical Variables are:

- Mini-Mental State Examination score (MMSE): The MMSE is a 30-point questionnaire widely used to assess cognitive impairment in clinical and research environments [15]. However, MMSE alone cannot be used for the classification of a disease.
- Clinical Dementia Rating (CDR): The CDR is a 5-point scale used to characterizing and tracking a patient's dementia level. CDR is estimated based on a semi-structured interview of the subject and the caregiver (informant) and the clinician's clinical judgment.
- Estimated total intracranial volume (eTIV): The eTIV variable is the estimation of total intracranial brain volume.
- Normalize Whole Brain Volume (nWVB): The nWVB variable measures the volume of the whole brain.
- Atlas Scaling Factor (ASF): Atlas Scaling Factor is a computed scaling factor that transforms native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix). [16]

In our test, we used Gender, Age, Years of education, Socioeconomic Status, Mini-Mental State Examination score, and Clinical Dementia Rating for the non-MRI test. In our Test with MRI data, we used all of these variables.

B. Data processing

First, we removed all the rows with empty data. Initially, we had 373 subjects, but after modifying our dataset, we had 354 subjects. Then we used the One-hot encoder to encoded all categorical variables and performed feature engineering. Finally, we split the data into two parts, the train set and the test set. We used 80% of the data for training our models and 20% for testing purposes.

We conducted our experiment in two steps. In the first step, we included all the MRI data and longitudinal lifestyle data like age, gender, income, education, and other lifestyle data and neuropsychological scores like MMSC. In the second

step, we excluded all the MRI-related data and did the same experiment again. Our dataset has multiple classes like demented, non-demented, and converters.

C. Machine learning algorithms.

After processing our data, we moved onto selecting an efficient Machine learning. Selecting the best machine learning algorithm is very important. So, we experimented with several machine learning algorithms. We used both classification and ensemble learner algorithms. Ensemble learners can sometimes perform better than the classification algorithms by combining the predictions of multiple base estimators.

For our test with MRI data, we started with several Machine learning models. We first run all the models without hyper-parameter tuning. We only selected the models that had more than 50% accuracy and Recall. Some models performed very poorly. For example, KNeighbours Classifier trained with MRI data only got an accuracy score of 45.07% and recall 0.45. So, we did not invest time in tunning hyper-parameters of these models.

We got five models that performed well: Random Forest Classifier, Gaussian NB, Linear SVC, Logistic Regression, and Ada boost classifier. So, we moved on to tunning these models. The supervised learning module performs a stratified ten-fold cross-validation and grid search over selected features for each model. For example, for the Random Forest method, we tunned n_estimators, max_depth, max_features, min_samples_split, min_samples_leaf, and bootstrap for tunning the model. We created a list of different values for each and searched for the best parameters using grid search. After finding the best parameters the first time, for the numerical value parameters like n_estimators, we changed the list of values, this time, we took values close to the previous best and ran again to get the maximum accuracy recall possible. We repeated this process until the Accuracy was not increasing anymore. Then using the parameters that maximized the Accuracy and Recall, we predicted the results for our test dataset.

We repeated the same procedure for our experiment without MRI data.

After that, we used our tunned model to create a hybrid model for both tests. We used soft voting classifier for building the hybrid models. We combined Adaboost classifier, Random Forest, Linear SVC, and Logistic Regression with soft voting classifier for building this hybrid model for MRI data. For non-MRI data, We combined five models. They are Random Forest Classifier, GaussianNB, LinearSVC, Logistic Regression, and AdaBoost Classifier.

D. Comparison

We compared the result of the two experiments. We compared the result in different scales like Accuracy, F1 score, Recall, Precision. Accuracy is the percentage of correct predictions for the test data. The equations are of Accuracy is,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Here, TP stands for True Positive, TN denotes True Negative, FP denotes False Positive, and FN denotes False Negative. However, Accuracy is not always the best scale to measure a model's performance. Precision tells us how many of the predicted positive results are positive. While Recall tells us how many of the positives are captured by our model. The equation of Recall and Precision are,

$$Recall = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

F1 score is used when we need to balance between Precision and Recall. The equation for measuring the F1 score is,

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

We analyzed the Area Under the Receiver Operating Characteristics to determine which model best distinguishes classes. The model with the highest ROC – AUC value has the best separability. F1 score also tells us which model has the optimal false positive and true positive rates.

E. Results

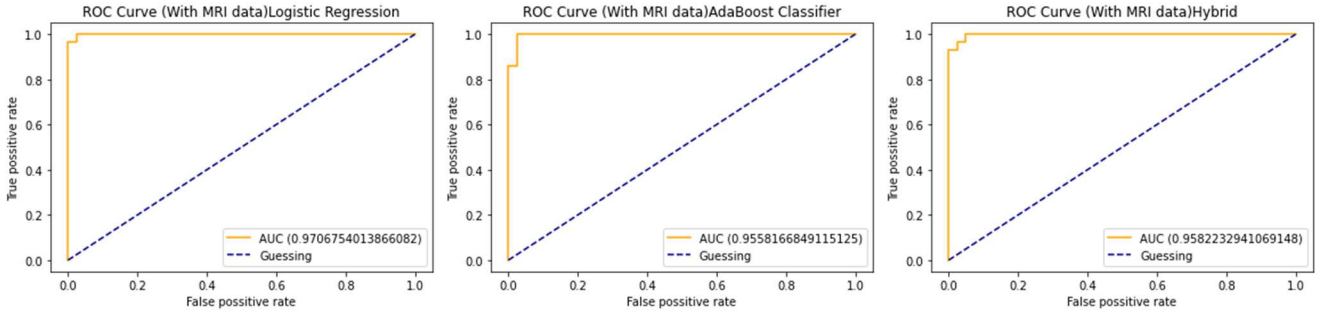
a) MRI data results: We got the best Accuracy and a recall score from AdaBoost Classifier. It had an Accuracy score of 96.07%, a Recall of 0.96, a Precision of 0.96, and an F1 score of 0.96. The Accuracy score of Logistic Regression was 94.37%, a Recall score of 0.94, a Precision score of 0.94, and an F1 score of 0.94.

We got the best AUC score from Logistic Regression, 0.97, AdaBoost Classifier had an AUC score of 0.95. The Area Under the Curve (AUC) equals a classifier's probability of ranking a randomly chosen positive instance higher than a randomly chosen negative example. So the logistic Regression model has the best separability between positive and negative results.

The results of our hybrid model were not as good as the individual models. Our hybrid model scored an Accuracy score of 92.55%, a Recall score of 0.93, a Precision score of 0.93, an F1 score of 0.93, and an AUC score of 0.95.

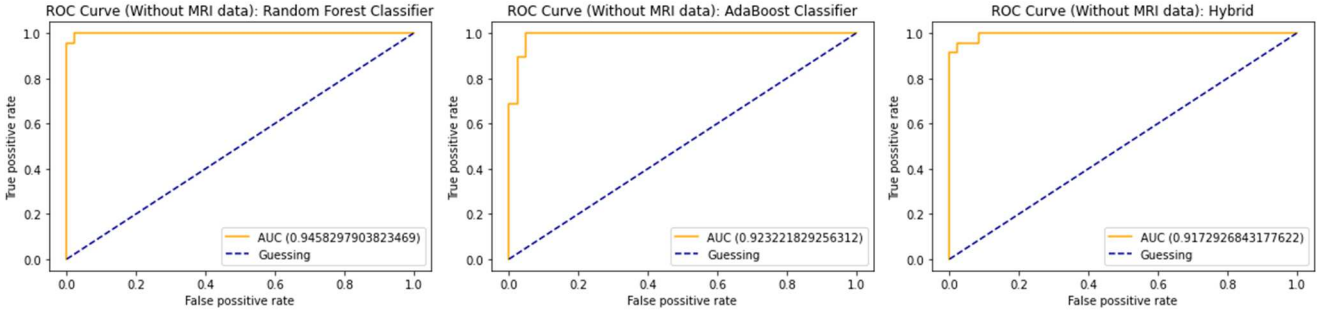
TABLE I. This table shows the accuracy, recall, precision and f1 score of all the models trained using MRI data.

Models	*Accuracy*	*recall*	*precision*	*F1 Score*
Random Forest	*93.14%*	*0.93*	*0.93*	*0.93*
GaussianNB	*86.00%*	*0.86*	*0.86*	*0.86*
Logistic Regression	*94.37%*	*0.94*	*0.94*	*0.94*
AdaBoost Classifier	*96.07%*	*0.96*	*0.96*	*0.96*
Linear SVC	*90.14%*	*0.90*	*0.90*	*0.90*
Hybrid	*92.55%*	*0.92*	*0.92*	*0.92*



(1) (2) (3)

Fig 1. The Receiver Operating Characteristic curve (ROC) and with Area Under the Curve (AUC) score of Logistic Regression (1), AdaBoost Classifier (2), and Our Hybrid model that was created by combining Adaboost classifier, Random Forest, Linear SVC, and Logistic Regression with soft voting classifier in the experiment using MRI data.



(1) (2) (3)

Fig 2. The Receiver Operating Characteristic curve (ROC) and with Area Under the Curve (AUC) score of Random Forest Classifier (1), AdaBoost Classifier (2), and Our Hybrid model that was created by combining Random Forest Classifier, GaussianNB, LinearSVC, Logistic Regression, and AdaBoost Classifier with soft voting classifier in the experiment without using MRI data.

a) Non-MRI data results: Our best performing model was the Random forest classifier when we trained our models without MRI data. This model scored an Accuracy score of 90.14%, a Recall of 0.90, a Precision of 0.90, and an F1 score of 0.90. The second-best performing models were LinearSVC and the AdaBoost classifier. Both models achieved an Accuracy score of 89.14%, a Recall of 0.89, a Precision of 0.89, and an F1 score of 0.89. The AUC scores for these Random forest Classifiers, AdaBoost classifier, and LinearSVC were 0.945 and 0.923 and 0.92, respectively.

In our training with MRI data KNeighbors Classifier performed very poorly. However, in this case, when trained without MRI data, this model performed better than before. It is a perfect example that modifying a dataset can improve the performance of a model.

We got our best result when we combined all the models. Our Hybrid model scored an Accuracy score of 93.37%, a Recall of 0.93, a Precision of 0.93, and an F1 score of 0.93. The AUC score for this hybrid model was 0.917.

Comparing this result with our previous results shows that GaussianNB and Hybrid models also performed better when trained without MRI data, like KNeighbors Classifier. Linear SVC performed the same as before. However, our overall Accuracy decreased. Previously our best model had 96.07% Accuracy, but when we removed all the MRI data, our Accuracy was 93.37%. So, we got an Accuracy drop of 2.7%. When it comes to detecting a patient, we always want to avoid False Negatives. So, we can consider the recall score for selecting our optimal model. The Recall decreased by 0.03, from 0.96 to 0.93, which is not a significant drop. The drop in the AUC score was 0.03. It dropped from 0.97 to 0.94.

TABLE II. This table shows the accuracy, recall, precision and f1 score of all the models trained without MRI data.

Model	*Accuracy*	*recall*	*precision*	*F1 Score*
Random Forest	*90.14%*	*0.90*	*0.90*	*0.90*
GaussianNB	*88.73%*	*0.89*	*0.89*	*0.89*
Logistic Regression	*88.73%*	*0.88*	*0.88*	*0.88*
AdaBoost Classifier	*89.14%*	*0.89*	*0.89*	*0.89*
Linear SVC	*89.14%*	*0.89*	*0.89*	*0.89*
KNeighbors	*83.10%*	*0.83*	*0.83*	*0.83*
Hybrid	*93.37%*	*0.93*	*0.93*	*0.93*

IV. Conclusion & Future Scope

From our result, we can see no considerable gap in efficiency between our two experiments, only about a 3% difference in Accuracy. MRI tests have significant importance for monitoring the disease and treatment. Some

papers have found better Accuracy with MRI data than ours. Fulton et al. [17] found a 99.34% correct classification rate using MRI images and 91.3% accuracy in predicting CDR using non-MRI data. So, MRI classification is nearly perfect. So non-MRI methods have about a 7% gap, which is significant. However, the main objective of our research was not to find an optimal way of predicting Alzheimer's disease but to show that the non-MRI methods are not too inefficient comparing to MRI data analysis methods.

Though the Non-MRI method has good results, it is not perfect. Some Demented patients may not be detected with this approach. It is recommended to do an MRI test if possible for better detection and avoid the risk of false-negative prediction. False-negative predictions can be dangerous sometimes when it comes to detecting a disease. However, the non-MRI methods can provide reliable results when there is a lack of diagnostic assessment or cannot afford an MRI test. This will reduce their initial cost by a margin. The non-MRI method can raise awareness in developing nations, as most of them are unaware of Alzheimer's disease.

These results can be further improved by counting more AD-related factors, like depression, head trauma, epilepsy, diabetes, stroke, and similar factors. We firmly believe our research will pave the way to raise awareness and cheap diagnosis of AD with excellent efficiency.

ACKNOWLEDGMENT

Our research work uses the dataset from OASIS. The OASIS project is supported in part by grants P50 AG05681, P01 AG03991, R01 AG021910, P20 MH071616, U24 RR0213.

REFERENCES

[1] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P. R. Menezes, E. Rimmer, and M. Scazufca, "*Global prevalence of dementia: a Delphi consensus study*," The Lancet, vol. 366, no. 9503, pp. 2112–2117, 2005.

[2] I. Raicher, M. M. Shimizu, D. Y. Takahashi, R. Nitrini, and P. Caramelli, "*Alzheimer's disease diagnosis disclosure in Brazil: a survey of specialized physicians' current practice and attitudes*," International Psychogeriatrics, vol. 20, no. 03, 2007.

[3] K. Fiore, "*Copyright Issues Hinder MMSE Use*," MedPage Today, 09-Jun-2015. [Online]. Available: https://www.medpagetoday.org/neurology/dementia/52040?vpass=1. [Accessed: 11-May-2021].

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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

[14] "OASIS Brains," *OASIS Brains - Open Access Series of Imaging Studies*. [Online]. Available: https://www.oasis-brains.org/#data. [Accessed: 11-May-2021].

[15] Pangman, VC; Sloan, J; Guse, L. "*An Examination of Psychometric Properties of the Mini-Mental Status Examination and the Standardized Mini-Mental Status Examination: Implications for Clinical Practice*". Applied Nursing Research. 13 (4): 209–213, 2000.

[16] R. L. Buckner, D. Head, J. Parker, A. F. Fotenos, D. Marcus, J. C. Morris, and A. Z. Snyder, "*A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume*," NeuroImage, vol. 23, no. 2, pp. 724–738, 2004.

[17] Fulton, L., Dolezel, D., Harrop, J., Yan, Y. and Fulton, C., 2019. '*Classification of Alzheimer's Disease with and without Imagery Using Gradient Boosted Machines and ResNet-50'*. Brain Sciences, 9(9), p.212.