

A Gender-neutral Approach to Detect Early Alzheimer's Disease Applying a Three-layer NN

Shithi Maitra¹, Tonmoy Hossain², Abdullah Al-Sakin³,
Sheikh Inzamamuzzaman⁴, Md. Mamun Or Rashid⁵, Syeda Shabnam Hasan⁶
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh

Abstract—Early diagnosis of the neurodegenerative, irreversible disease Alzheimer's is crucial for effective disease management. Dementia from Alzheimer's is an agglomerated result of complex criteria taking roots at both medical, social, educational backgrounds. There being multiple predictive features for the mental state of a subject, machine learning methodologies are ideal for classification due to their extremely powerful feature-learning capabilities. This study primarily attempts to classify subjects as having or not having the early symptoms of the disease and on the sidelines, endeavors to detect if a subject has already transformed towards Alzheimer's. The research utilizes the OASIS (Open Access Series of Imaging Studies) longitudinal dataset which has a uniform distribution of demented, non-demented subjects and establishes the use of novel features such as socio-economic status and educational background for early detection of dementia, proven by performing exploratory data analysis. This research exploits three data-engineered versions of the OASIS dataset with one eliminating the incomplete cases, another one with synthetically imputed data and lastly, one that eliminates gender as a feature—eventually producing the best results and making the model a gender-neutral unique piece. The neural network applied is of three layers with two ReLU hidden layers and a third softmax classification layer. The best accuracy of 86.49% obtained on cross-validation set upon trained parameters is greater than traditional learning algorithms applied previously on the same data. Drilling down to two classes namely demented and non-demented, 100% accuracy has been remarkably achieved. Additionally, perfect recall and a precision of 0.8696 for the 'demented' class have been achieved. The significance of this work consists in endorsing educational, socio-economic factors as useful features and eliminating the gender-bias using a simple neural network model without the need for complete MRI tuples that can be compensated for using specialized imputation methods.

Keywords—Alzheimer's disease; dementia; exploratory data analysis; synthetically imputed data; socio-economic factors; specialized imputation

I. INTRODUCTION

Alzheimer's disease is a growing concern among the world's retired population. It is an irreversible, progressive, neurodegenerative brain disorder that gradually dismantles memory and reasoning skills and eventually, the ability to carry out the simplest of tasks. In 2015, there were approximately 29.8 million [1] people worldwide who had been diagnosed with Alzheimer's disease and the number is increasing day by day. This number is expected to be over 100 million by 2050 [2]. It most often affects about 6% of people 65 years' or older [3]. Furthermore, Alzheimer's disease, historically not

thought to be a normal part of aging, is now considered the most common form of dementia among elderly people which resulted in about 1.9 million deaths in 2015 [4]. Therefore, its socio-economic implications are enormous, carrying a major negative influence upon society and caregivers.

The National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA), now known as the Alzheimer's Association, have defined the most commonly used NINCDS-ADRDA Alzheimer's Criteria as definite, probable, possible and unlikely for diagnosis in 1984 [5]. Clinicians have long advocated early diagnosis i.e., at the possible and probable stages provided medications are frequently more effectual at the onset of the disease and drug-free interventions are also available to decelerate the atrophy of cerebral tissue. Furthermore, a demented state may well represent treatable and reversible medical conditions, other than early Alzheimer's in which case the earlier the actions, the better the results. Moreover, an early diagnosis allows the patient to carve out practicable medical and financial decisions while also potentially allowing caregivers to develop better support system for the affected [6]. More justifications endorsing the early detection include amplified opportunities to participate in clinical trials, additional time to record memories, improved safety etc.

Prognostically, Alzheimer's disease is diagnosed based on a person's medical history, narratives by relatives and behavioral observations. Neuropsychologically, tests such as the Mini-Mental State Examination (MMSE) are recognized to evaluate cognitive impairments indicative of a positive diagnosis of the disease [7]. Radiographically, Alzheimer's disease is characterized by the loss of neurons and synapses in the cerebral cortex and in certain subcortical regions [8]. The hippocampal atrophy, ventricle enlargement and cortex shrinkage are sensitive features of Alzheimer's disease. Therefore, doctors perform scans like Computed Tomography (CT), Magnetic Resonance Imaging (MRI), or Positron Emission Tomography (PET), to rule out other possible causes for the symptom. Analytically, this study shows years of education (EDUC) and Socio-Economic Status (SES) measured on a scale ranging from 1 to 5 as significant features in early detection of the disease. These perspectives justify MMSE, Clinical Dementia Rating (CDR), Estimated Total Intracranial Volume (eITV), Normalized Whole Brain Volume (nWBV), EDUC, SES as distinctive features characterizing Alzheimer's.

Due to the availability of OASIS longitudinal MRI data and machine-learning methodologies, one can measure the similarity of an individual's cortical atrophy with that of a representative Alzheimer's disease patient cohort. The recent bio-medical researches are gaining momentum using Neural Networks (NN). Neural Network models, consisting of multiple hidden layers having different activation units, are transcending traditional learning algorithms like Logistic Regression, Support Vector Machine etc in performance. Neural Networks are able to fit very complex functions with numerous independent variables as features. Such models can discover patterns spread across multiple dimensions upon refinement of its initial parameters through many epochs using the back-propagation algorithm. The derivatives calculated at each step of backpropagation indicates in which direction the parameters should be refined and a learning rate defines its magnitude. The weights represent the mapping from one layer to the other creating a layered, hierarchical architecture. The early hidden layers learn comparatively simpler features while the latter ones learn sophisticated features upon the previously learned simpler features.

In this paper, a simple three-layer Neural Network architecture has been trained using the OASIS longitudinal MRI dataset to classify among patients as having early Alzheimer's disease. Imputing data precluded the need for complete tuples making the best use of available data. The study introduces years of education and socio-economic status as two novel features while further nullifying gender as a feature in order to gain better performance measures.

The organization of this paper dictates the second section as presentations of related work, the third section as a narration of methodology, the fourth section as a tabulation of results and the final section as concluding remarks.

II. RELATED WORK

Artificially intelligent former and recent researches undertaken on detection of early or matured stage Alzheimer's can be categorized along three paradigms: detection applying machine learning on structured data, detection using convolutional neural networks on radiographs and detection using hybrid methods combining the former two. Detection of Early Alzheimer's poses a supervised classification problem, some literature on which are reviewed below.

A. Traditional Machine Learning Algorithms

Datta et al. [9] explored ML for classifying dementia using the University of California, Irvine's Alzheimer's Research Center's data. Six ML methods were applied to a database of 578 patients and controls. The neuropsychologists applied the Diagnostic and Statistical Manual of Mental Disorders-04 criteria to classify dementia status. The researchers extracted age, sex, job, education and responses of patients to questions from the Alzheimer's database as features. Using the Frequently Asked Questions (FAQ > 8) and Blessed Orientation-Memory-Concentration (BOMC > 10) tests recommended by Agency for Health Care Policy and Research, accuracies were 69% and 63% respectively, which were 14% to 20% worse than results obtained by ML methods. Combining the two tests (FAQ & BOMC) resulted

in a 60% accuracy. Experiments showed that ML methods can detect dementia 15% to 20% more accurately than applying either the FAQ, BOMC or their combined cut-off criteria.

In another endeavor [10], the researchers utilized MRI related data generated by the Open Access Series of Imaging Studies (OASIS) project. There was an emphasis on exploring the relationship between each feature of MRI tests and dementia of the patient. They conducted exploratory data analysis to state the relationship among data explicitly through visualizations so as to discover the correlations before feature extraction or prediction. Missing values were handled in two ways: dropping of tuples having missing values and replacing corresponding values exploiting off-the-shelf inference libraries. Subjects were classified applying traditional Logistic Regression, SVM, Decision Tree, Random Forest Classifier and AdaBoost the results of which are depicted in Fig. 1.

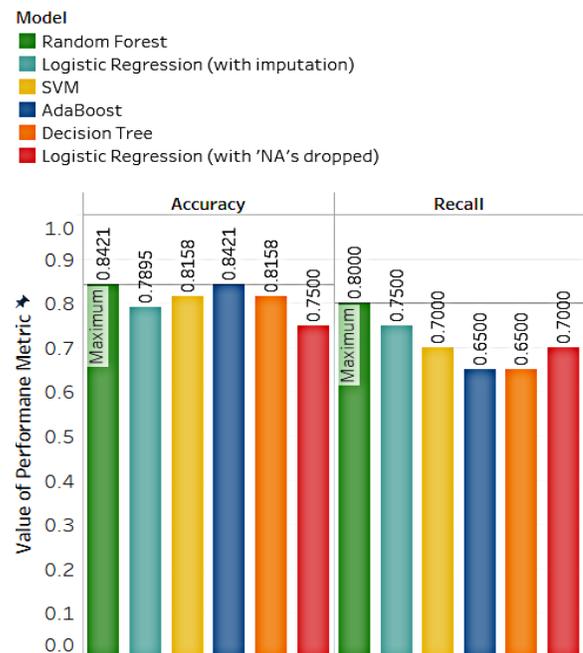


Fig. 1. Results obtained upon applying traditional ML algorithms

Alvarez et al. [11] presented a computer-assisted diagnostic tool based on Principal Component Analysis and Support Vector Machine (SVM) for improving the Alzheimer's diagnosis accuracy by means of SPECT (Single Photon Emission Computed Tomography) images. This process reduced the dimensionality of the feature space from ~500000 to ~100, thus facing the small sample size problem. The application of SVM to high dimensional and small sample size problems still remained a challenge and improving the accuracy SVM-based approaches is a field in development.

B. Convolutional Neural Networks (CNN) on Radiographs

Sarraf et al. [12] classified Alzheimer's data by using CNN deep learning LeNet architecture. For this study, Alzheimer's inflicted patients (24 female and 19 male) and 15 elderly normal control subjects with a mean age of 74.95 years were selected from Alzheimer's Disease Neuroimaging Initiative

dataset. The pre-processing steps for the anatomical data involved the removal of non-brain tissue from T1 anatomical images using Brain Extraction Tool. The product of preprocessing was 45x54x45x300 images in which the first 10 slices of each image were removed for containing no functional information. The researchers adjusted LeNet-5 for functional Magnetic Resonance Imaging (fMRI) data. LeNet differentiated Alzheimer’s from normal control and the average accuracy reached 96.8588%.

C. Hybridized Approach using Neural Networks

Gulhare et al. [13] proposed a Deep Neural Network (DNN) classification method to diagnose Alzheimer’s from MRI. The resulting attributes were respectively the area of the extracted region, the perimeter, mean, standard deviation, 28 horizontal distances (D_1, D_2, \dots, D_{28}), the height and the coordinates of the center of gravity of the region (G_x, G_y). The database included a longitudinal collection of 150 (88 female and 62 male) subjects aged 60 to 96. 72 of the subjects were characterized as non-demented while 78 as demented. The DNN consisted of multiple hidden layers and a softmax layer. The classifier rendered a maximum accuracy of 96.6% with different pairs of attributes. It classified with an accuracy of 90.3% retaining all attributes. The DNN approach showed better performance compared to SVM.

III. PROPOSED METHODOLOGY

The OASIS longitudinal MRI dataset went through intensive preprocessing and exploratory analysis which resulted in three reproductions of the dataset with significant features that were eventually modeled to produce phenomenal results (Fig. 2).

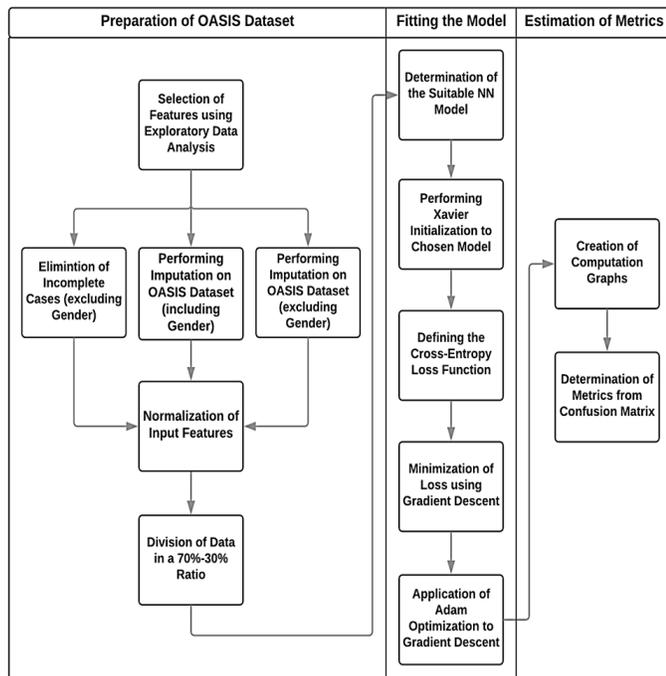


Fig. 2. Workflow for the proposed detection of early Alzheimer’s disease.

A. Preparation of OASIS Dataset

1) Selection of Features using Exploratory Data Analysis (EDA): Exploratory data analysis is a statistical process to summarize tendencies within data, aided by visualizations. EDA was primarily applied on OASIS dataset for extracting insights beyond formal modeling to hypothesize features reinforced by data.

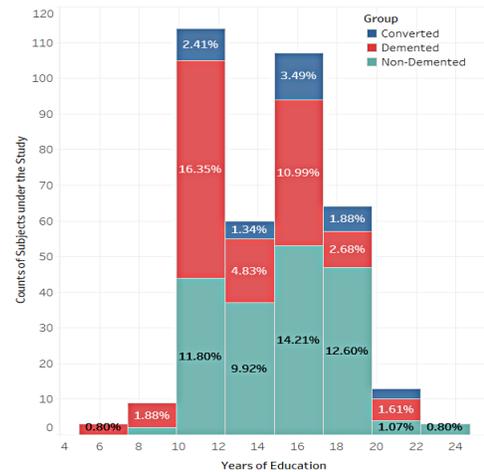


Fig. 3. Percentage of total subjects having respective years of education

An exploratory visualization shows that demented (and converted) subjects have experienced comparatively fewer years of education. Concretely, the subjects who received 10 to 12 years’ schooling, are mostly demented while we see an opposite scenario as learning prolongs—qualifying years of education as a feature to detect early Alzheimer’s (Fig. 3).

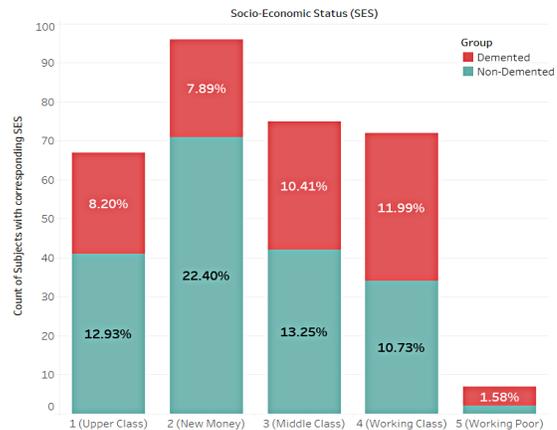


Fig. 4. Percentage of subjects belonging to respective socio-economic status.

Another summarization shows that as a transition is made towards working class from the upper class, the rate of dementia increases as a general trend justifying socio-economic status as another feature characterizing early Alzheimer’s (Fig. 4).

In another discovery, it is explored as a trend that demented patients tend to have fewer years of education’s with a hum-

ble socio-economic background, the opposite being true for healthy subjects (Fig. 5).

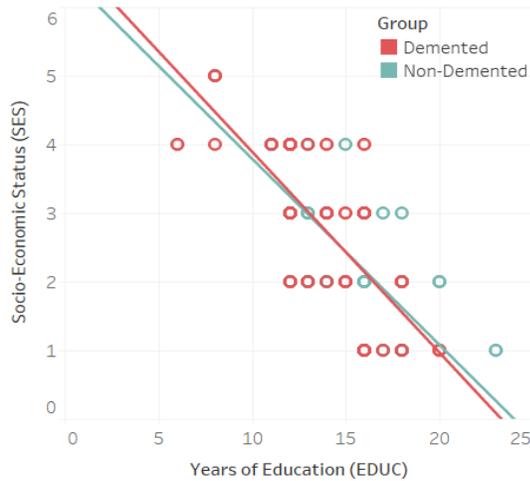


Fig. 5. Trend lines distinguishing subjects with respect to education and socio-economic rank

Subjects found negatively affected often provided reversed result in subsequent visits where CDR raised from ‘questionable’ to ‘mild’. Delay in MRI is expounded by degeneration in tissue. Medically indicative features e.g., MMSE, age, eITV, nWBV, ASF have also been selected, thus assembling ten significant features for early Alzheimer’s diagnosis (Fig. 7).

2) *Elimination of Incomplete Cases (excluding Gender):* Missing values in tuples are treated differently in various linguistic frameworks. Representations such as NaN, garbage values are problematic as they hail from a different distribution causing their derivatives to lead to useless parameters. Thus incomplete tuples have been subsetted out using R, bringing down the number of training examples from 373 to 354.

3) *Performing Data Imputation (including/excluding Gender):* Imputation is a statistical process of assigning a value by inference to a missing field taking into consideration other existing fields and summary of the dataset. In the OASIS dataset, socio-economic status (SES) and Mini-Mental State Examination (MMSE) were missing for some demented patients which were imputed using a tailored version of mean imputation according to the algorithm below (Fig. 6).

This retained all 373 tuples making the best use of available data. However, excluding gender as a feature provided a gender-neutral version of the dataset.

4) *Division of Data into a 70%-30% Ratio:* According to standard machine learning practices, OASIS dataset has been split into a larger training set and a comparatively smaller cross-validation set. 70% of the data have been used for training, assigning 248 tuples for the training purpose while setting aside the rest 30% comprising 106 records for cross-validation. Imputation raised these numbers to 262 and 111, respectively.

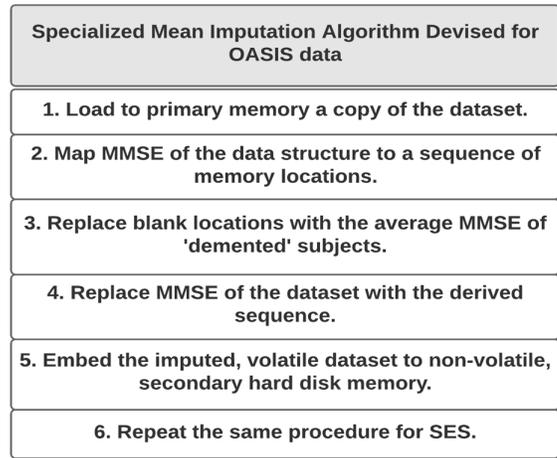


Fig. 6. Specialized Mean Imputation Algorithm.

5) *Normalization of Input Features:* While pre-processing data, it is crucial that parameters belong to the same scale for a fair comparison between them and for the gradient descent to converge following an oriented trajectory. Normalization rescales all numerics in the range [0, 1] using the formula below:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

The OASIS dataset is replete with data hailing from different units and scales. The range of scales for Cognitive Dementia Rating and years of education are not identical, so is the case for any other collection of features—thus justifying normalization.

B. Fitting the Model

1) *Determination of the Suitable Neural Network Model:* Numeric representations of the features constitute the input layer of the model. The weighted inputs propagate through two ReLU-activated hidden layers each containing ten neurons. Finally, a SoftMax layer computes the probabilities for the classes, labeling as the highest (Fig. 7).

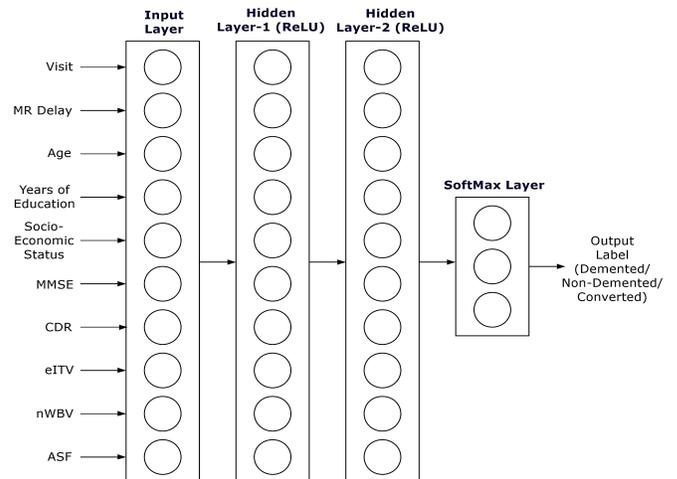


Fig. 7. The proposed three-layer neural network.

The hyper-parameters have been elected for the following rationale:

- **Number of layers, neurons:** Two hidden layers are chosen to prevent overfitting with ten hidden units, to preclude underfitting.
- **Learning rate:** A small learning rate of 0.0001 has been chosen to prevent overshooting across the minima.
- **Number of epochs:** The model was trained for a large 1500 epochs to quantify an optimum set of parameters.
- **Size of minibatch:** Complementing primary storage, 32 tuples were assigned per minibatch.
- **Adam Optimization parameters:** β_1 : 0.9, β_2 : 0.999, ϵ : 1e-08

2) *Performing Xavier Initialization to Chosen Model:* Xavier initialization ensures delicate initialization of weights in order to keep the signal in a reasonable range of values through multiple layers. This initialized the weights in the network by drawing them from a distribution with zero mean and a specific variance as,

$$Var(W) = \frac{1}{n_{in}} \quad (2)$$

Where W is the initialization distribution for the neuron in question and n_{in} is the number of neurons feeding into it. The distribution used is typically Gaussian or uniform.

3) *Defining the Cross-Entropy Loss Function:* The cross-entropy loss function has been optimized for the three-class classification problem with a view to obtaining the greatest refinement of the parameters. Represented here is precisely the cross-entropy, summed over all training examples [14]:

$$\begin{aligned} & -\log L(\{y^{(n)}\}, \{\hat{y}^{(n)}\}) \\ &= \sum_n [-\sum_i y_i \log \hat{y}_i^{(n)}] = \sum_n H(y^{(n)}, \hat{y}^{(n)}) \end{aligned} \quad (3)$$

where n indicates the number of training examples, y_n denotes the ground-truth value for an individual example, $\hat{y}^{(n)}$ is the prediction of the model and i represents the sequence of activation within a layer.

4) *Minimization of Loss using Gradient Descent:* A set of parameters θ is to be chosen so as to minimize error $J(\theta)$. The gradient descent algorithm starts with some initial θ , then repeatedly performs the update [14]:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad (4)$$

This update is simultaneously performed for all features, i.e., $j = 0, 1, \dots, n$ where α is the learning rate. This is a very natural algorithm that repeatedly takes a step in the direction of the steepest decrease of $J(\theta)$. To implement the algorithm, the partial derivative term has to be computed. If there is only one training example (x, y) , we have [14],

$$\begin{aligned} \frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2 \\ &= 2 \cdot \frac{1}{2} (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (h_\theta(x) - y) \\ &= (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (\sum_{i=0}^n \theta_i x_i - y) \\ &= (h_\theta(x) - y) \cdot x_j \end{aligned}$$

$$\text{Therefore, } \frac{\partial}{\partial \theta_j} J(\theta) = (h_\theta(x) - y) \cdot x_j \quad (5)$$

To modify this method for a training set of more than one examples, it is to be replaced with the following algorithm [14]:

Repeat until convergence {

$$\theta_j := \theta_j + \alpha \sum_{i=1}^m (y^{(i)} - h_\theta(x^{(i)})) x_j^{(i)} \quad (\text{for every } j)$$

}

5) *Application of Adam Optimization to Gradient Descent:* Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The parameters for Adam Optimization are as follows:

- α : The learning rate or step size. Learning rate decay, permissible in Adam, has not been used for Alzheimer's classification.
- β_1 : The exponential decay rate for the first moment estimates (e.g. 0.9).
- β_2 : The exponential decay rate for the second-moment estimates (e.g. 0.999). This value is set close to 1.0 on problems with a sparse gradient.
- ϵ : A very small number to prevent any division by zero in the implementation (e.g. 10E-8).

C. Estimation of Metrics

Single-figure performance measures are customary to measure the proficiency of a learning model. Evaluation metrics such as accuracy, precision, recall are significant within the medical research arena and are computed using confusion matrices created via computation graphs.

1) *Creation of Computation Graphs:* A computational graph is the representation of a collective mathematical function using the frameworks of graph theory. Complying with the ethos of graph theory, a computation graph consists of nodes and edges. The nodes are indicative of either operations (denoted by round shapes) or operands (denoted by rectangular shapes) while the directed edges delineate the sequence of mathematical operations to be performed.

The NN framework of TensorFlow demands a computation graph to be devised before running it as a session to calculate

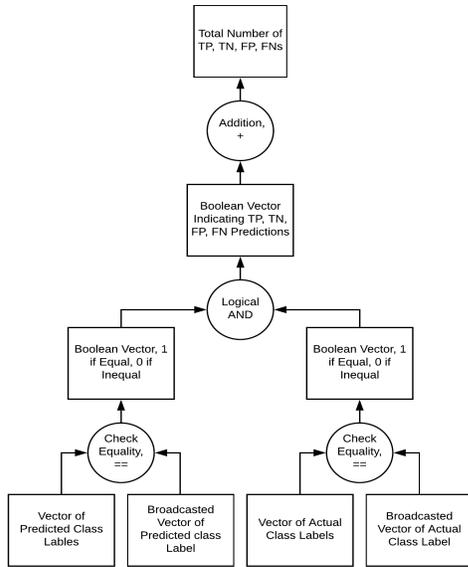


Fig. 8. Generalized computation graph for determining entries of the confusion matrix

the numerics. For this purpose, we manually concoct a computation graph (Fig. 8) in a bottom-up manner to determine the entries associated with the confusion matrix i.e.,

- predicted demented and actually demented, true positives (TPs)
- predicted demented while actually non-demented, false positives (FPs)
- predicted non-demented and actually non-demented, true negatives (TNs)
- predicted non-demented while actually demented, false negatives (FNs)

The popular one-hot boolean representation of class labels has been used for the purpose. We define another computation graph for calculating the accuracy of the implemented model on the cross-validation set (Fig. 9).

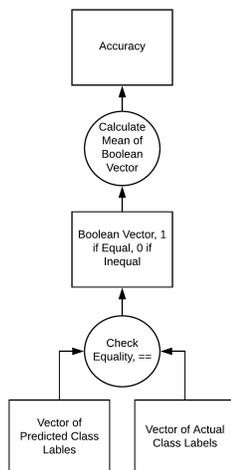


Fig. 9. Computation graph depicting computation of accuracy.

Upon checking for equality, the output boolean vector indicates positively the training examples identified correctly as possessing early Alzheimer's. Thus, the statistical mean of this data structure provides the fraction of correctly identified patients for the classifier.

2) *Determination of Metrics from Confusion Matrix:* In the jargon of machine learning, concretely in the problem of statistical classification, a confusion matrix is a specific tabular layout used to explain the performance of a classification model on a set of cross-validation data for which the true labels are available. Rows of the matrix represent the instances in a predicted class while columns represent the instances in an actual class (or vice versa). The name originates from that it makes viable to see if the system is confusing the classes (i.e. commonly mislabeling one as another).

		predicted class		
		demented	non-demented	converted
actual class	demented	predicted demented and actually demented, TPs	predicted non-demented while actually demented, FNs	predicted converted while actually demented
	non-demented	predicted demented while actually non-demented, FPs	predicted non-demented and actually non-demented, TNs	predicted converted while actually non-demented
	converted	predicted demented while actually converted	predicted non-demented while actually converted	predicted converted and actually converted

Fig. 10. Confusion matrix for early Alzheimer's classification problem with a focus on categories 'demented' and 'non-demented'.

The matrix (Fig. 10) is a special kind of contingency table, with two dimensions and identical sets of classes in both dimensions. For our medical diagnosis problem, we select accuracy, precision and recall as evaluation metrics using the terms calculated in the confusion matrix.

i) Accuracy

- Accuracy attempts to answer the following question: What proportion of predictions (both demented and non-demented) was actually correct?
- Accuracy is mathematically defined as follows:
 $accuracy = (TP+TN)/(P+N)$
- A model that produces no false predictions provides an accuracy of 1.0.

ii) Precision

- Precision attempts to answer the following question: What proportion of 'demented' identifications was actually correct?

- Precision has been calculated as follows:
 $precision = TP/(TP+FP)$
- A model that produces no false positives renders a precision of 1.0.

iii) Recall

- Recall attempts to answer the following question: What proportion of actual ‘demented’ was identified correctly?
- Mathematically, recall has been defined as follows:
 $recall = TP/(TP+FN)$
- A model that produces no false negatives delivers a recall of 1.0.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental results (Fig. 13) of this study encompass three iterations on the three-layer model, each using a different data-engineered revision of the OASIS longitudinal MRI data. Even the lowest accuracy achieved among these three iterations in the research surpasses the accuracies achieved using orthodox machine learning methods [10].

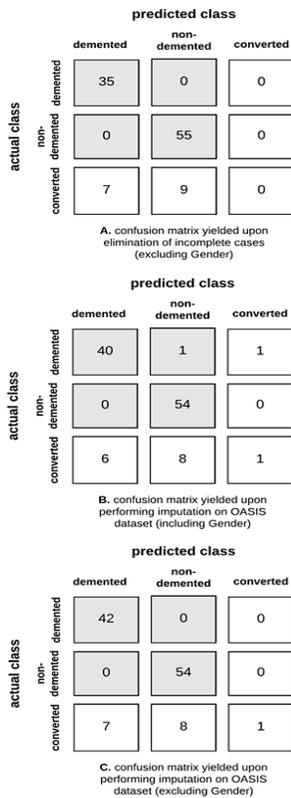


Fig. 11. Confusion matrices resulting from the application of three reproductions of the OASIS data to the proposed model with a focus on categories ‘demented’ and ‘non-demented’.

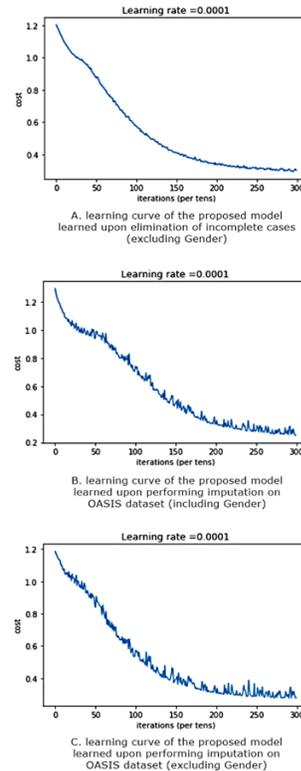


Fig. 12. Learning curves of the proposed neural network learned over three adaptations of the OASIS dataset.

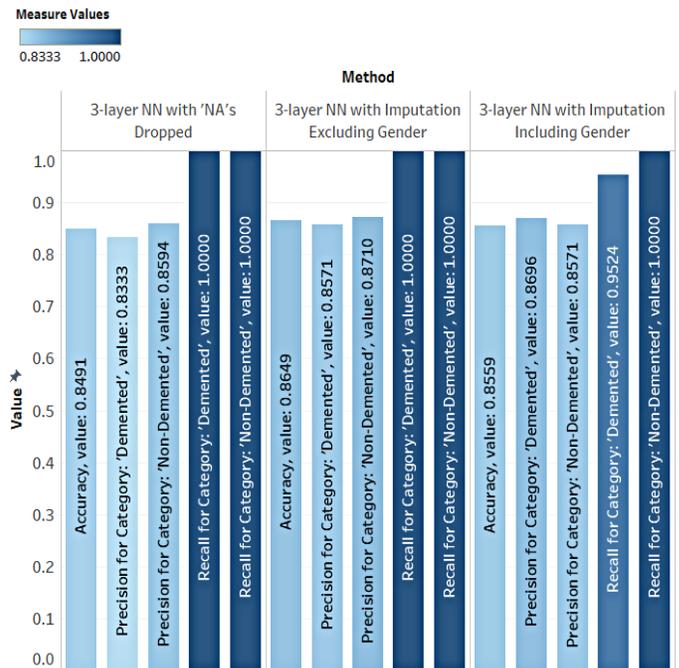


Fig. 13. Metrics yielded by the proposed neural network in general.

A. Metrics Yielded upon Elimination of Incomplete Cases (excluding Gender)

The variation of the dataset which dropped training examples containing at least one absent attribute value, in fact, made a waste of the available entries within the incomplete tuple and came up with the performance measures shown in Fig. 13 in general.

Although this is the variation in the study to yield the least accuracy, yet it outsmarted all typical machine learning techniques on the OASIS data. Drilled down to just two classes (Fig. 11.A) as having (demented) or not having (non-demented) early Alzheimer’s, the classifier produces impressive results as given in Table I according to the formulae given in Section III.C.2.

TABLE I. METRICS EVALUATED AS PER APPROACH A

performance measure	calculated value
accuracy	$(35+55)/(35+0+0+55)=1.0000$
precision	$35/(35+0)=1.0000$
recall	$35/(35+0)=1.0000$

B. Metrics Yielded upon Performing Imputation on OASIS Dataset (including Gender)

The variation of the dataset using customized, statistical mean imputation method made the best use of available data and synthesized artificially inferred data to replace the missing attributes. This statistically analytical approach created realistic data and also included gender as a feature rendering the general metrics as summarized in Fig. 13.

This variation of the dataset outperformed the previous metrics in approach A. Focusing on just two classes (Fig. 11.B) as having (demented) or not having (non-demented) early Alzheimer’s, the model produces successful results as the following (Table II).

TABLE II. METRICS EVALUATED AS PER APPROACH B

performance measure	calculated value
accuracy	$(40+54)/(40+1+0+54)=0.9894$
precision	$40/(40+0)=1.0000$
recall	$40/(40+1)=0.9756$

C. Metrics Yielded upon Performing Imputation on OASIS Dataset (excluding Gender)

This variation of the dataset is statistically imputed with customization as per labels, similar to the approach followed in B except for that gender has been eliminated as an input feature. The adaptation is an imputed version of the dataset described in A. This delivered the metrics as tabulated in Fig. 13 in general. This edition of the dataset outshined approach B and all other learning models (Fig. 1), making the model unbiased towards gender. Concentrating on just two classes (Fig. 11.C) as having (demented) or not having (non-demented) early Alzheimer’s, the model produces remarkable results like

the following (Table III):

TABLE III. METRICS EVALUATED AS PER APPROACH C

performance measure	calculated value
accuracy	$(42+54)/(42+0+0+54)=1.0000$
precision	$42/(42+0)=1.0000$
recall	$42/(42+0)=1.0000$

The research primarily being an attempt to classify the subjects as having or not having early Alzheimer’s, tends to center its attention on just two categories namely, demented and non-demented. Due to an uneven share of the ‘converted’ category in the dataset, the model performed incompetently on this category which is venial given the purpose of the research.

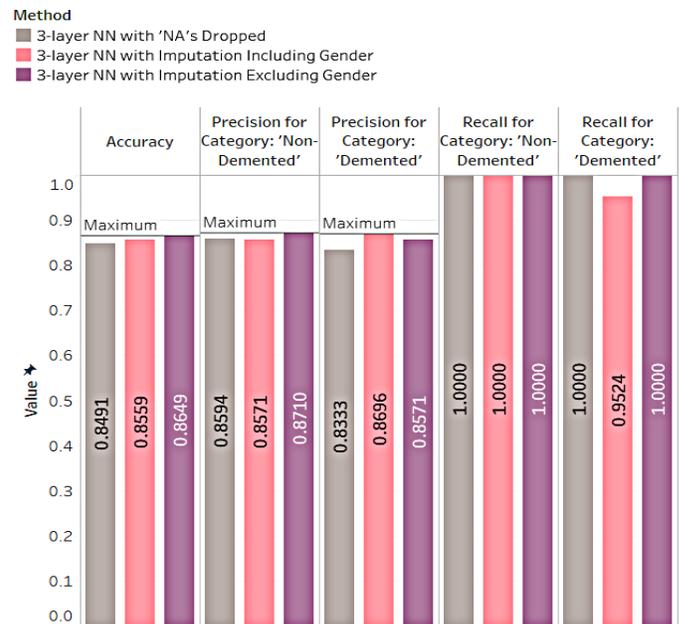


Fig. 14. Comparison of differently learned weights (Fig. 12) on the basis of performance measures.

While deployment of the model, the dataset can be engineered depending on utility as they each excelled in different metrics. Should accuracy be the foremost priority, parameters learned using the gender-exclusive imputed dataset will be an ideal pick. Likewise, parameters derived by training on any imputed dataset should suffice for perfect precision or recall (Fig. 14).

The simplistic, gender-neutral, three-layer neural network proposed superseded other hyper-parametrically tuned machine learning approaches (Fig. 15) through all three iterations using different variations of the OASIS dataset, performing greatly in terms of medically significant metrics (accuracy, precision, recall). Furthermore, the model has been trained (Fig. 12) on structured data, the business value of which is generally greater while the cost of acquisition is relatively lower than medical imagery.

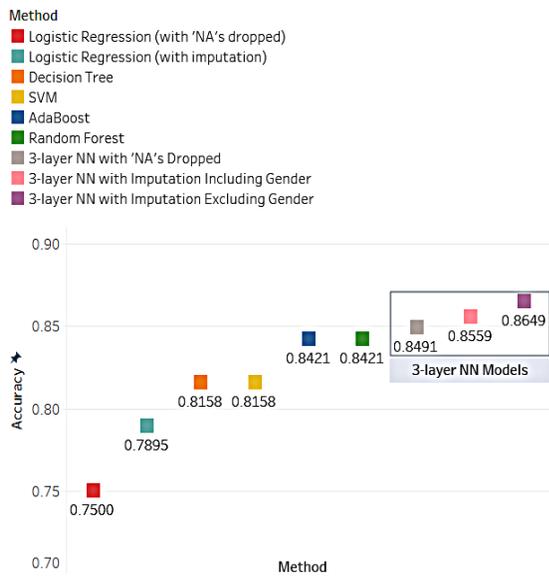


Fig. 15. Comparison among traditional learning models and proposed model.

V. FUTURE WORKS AND CONCLUDING REMARKS

Traditional machine learning methods, being efficient in predicting results upon carefully selected features, have thus far been a convenient approach towards early Alzheimer's detection. This study proves a simplistic neural network approach to be an even better methodology for its remarkable feature-extraction properties leading to best performance metrics (e.g., accuracy, precision and recall) obtained so far on the OASIS dataset.

However, the scope of this study can be broadened to classify among more stages of dementia upon data augmentation. Due to scarcity of data, this experiment performed dismally on the 'converted' category, causing a dent to the overall metrics—although providing tremendous results in the 'demented' and 'non-demented' categories which surpassed all conventional results.

This unique study introduces two novel features namely a person's socio-economic standing and educational background—bringing into question the role of gender in the prediction. This research also precludes the need for complete MRI data for a patient as the missing attributes can be inferred using customized imputation methods. This makes the model feasible, cost-effective. This study has unleashed new dimensions to current researchers intriguing them to look for features in a broader scope.

REFERENCES

- [1] GBD 2015 Disease Injury Incidence Prevalence Collaborators (October 2016). "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015". *Lancet*. 388 (10053):1545602. doi:10.1016/S01406736(16)316786. PMC 5055577. PMID 27733282.
- [2] Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia*. 2007;3(3):186–91. doi:10.1016/j.jalz.2007.04.381. PMID 19595937. World population prospects: the 2006 revision, highlights [PDF]. 2007.
- [3] Mendez MF (November 2012). "Early-onset Alzheimer's disease: non-amnesic subtypes and type 2 AD". *Archives of Medical Research*. 43 (8): 677–85. doi:10.1016/j.arcmed.2012.11.009. PMC 3532551. PMID 23178565.
- [4] GBD 2015 Mortality Causes of Death Collaborators (October 2016). "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015". *Lancet*. 388 (10053): 1459–544. doi:10.1016/S0140-6736(16)31012-1. PMC 5388903. PMID 27733281
- [5] Clinical Diagnosis of Alzheimer's Disease: Report of the NINCDS-ADRDA Work Group under the Auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939–44. doi:10.1212/wnl.34.7.939. PMID 6610841.
- [6] Verywell Health. (2019). 12 Benefits of Early Detection in Alzheimer's Disease. [online] Available at: <https://www.verywellhealth.com/12-benefits-of-early-detection-in-alzheimers-disease-98046> [Accessed 15 Jan. 2019]
- [7] The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*.1992;40(9):92235. doi:10.1111/j.15325415.1992.tb01992.x. PMID 1512391
- [8] Neuropathologic Changes in Alzheimer's Disease. *The Journal of Clinical Psychiatry*. 2003;64 Suppl 9:7–10. PMID 12934968.
- [9] Datta, Piew, W. R. Shankle, and Michael Pazzani. "Applying machine learning to an Alzheimer's database." *Proceedings of the AAAI-96 Spring Symposium AI in Medicine: Applications of Current Technologies*, Stanford, CA, USA. 1996.
- [10] Kaggle.com. (2019). DETECTING EARLY ALZHEIMER'S — Kaggle. [online] Available at: <https://www.kaggle.com/hyunseokc/detecting-early-alzheimer-s> [Accessed 10 Jan. 2019].
- [11] Alvarez, Ignacio, et al. "Alzheimer's diagnosis using eigenbrains and support vector machines." *International Work-Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2009.
- [12] Sarraf, Saman, and Ghassem Tofghi. "Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks." *arXiv preprint arXiv:1603.08631* (2016).
- [13] Gulhare, Kajal Kiran, S. P. Shukla, and L. K. Sharma. "Deep Neural Network Classification method to Alzheimer's Disease Detection." *International Journals of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277-128X (Volume-7, Issue-6), June 2017
- [14] Ng, Andrew. "CS229 Lecture notes." *CS229 Lecture notes 1* (2000) [Accessed 29 Nov. 2018].