Interpretable Deep Learning Model for Dementia Prediction

Gaurav Shivhare
School of Computing
Dublin City University
Dublin, Ireland
gaurav.shivhare2@mail.dcu.ie

Prof. Tomas Ward
School of Computing
Dublin City University
Dublin, Ireland
tomas.ward@mail.dcu.ie

Abstract* — Public health has always been a big concern in the world, the issue of mental health always grabs the attention of many organizations and educational firms. Dementia is one of the biggest mental illnesses among the elders, which not only affect the patient but the whole family. This research aims to address the prediction of Dementia of a person in his initial stage which can ease the diagnosis of the disease and might help in the reduction of the number of patients at a wider scale. Alzheimer's Disease Neuroimaging Initiative(ADNI) contains the data from longitudinal studies having their early life history, health status, sociodemographic and other different health-related constraints which can be used to develop patterns describing the development of dementia in an individual. Machine learning is capable of identifying these patterns and predict Dementia at an early stage. In Data Science, Deep Learning algorithms are advancing and able to produce better results than traditional machine learning algorithms but deficient in interpretability, this study proposes an approach to overcome this and successfully interprets deep learning model. Also, introducing an efficient model for dementia prediction achieving accuracy up to 86 %.

Keywords: data science, deep learning, public health, dementia, machine learning, classification, prediction

I. INTRODUCTION

Dementia is considered to be the most leading cause of death. As per the World Health Organization(WHO), dementia is 5th most cause of death around the globe while in 2010 dementia was less popular having the 14th place, legged big causes of death like road injury, cancer, and HIV/AIDS by doubling the deaths in these years. Dementia does have a high impact on the high-income and upper-middle-income countries like the United States, Ireland, and Finland while other causes of death are much popular in lower-middle and lower-income countries like India, Sri Lanka, and Nepal. In high-income countries, Dementia is considered to be the 3rd killer with a rate of around 60 people per 100,000 [1]. There are around 50 million people who have dementia. It is projected as 82 million in 2030 and 152 million in 2050 [2]. Further, Asia has the largest population suffering from dementia counting 22.9 million while Europe stands on second with 10.5 million people [3].

Dementia causes irreversible declination in the physical, social and intellectual functioning of a person and also impacts the quality of life of the patients as well as their families. The curers also face many problems like stress, strain, and psychological illness, in fact, depression has been noted in the 15-30% of curers [4].

The prediction of dementia in the early stages will help the patient as well as the curers for the reduction of risk, though there are approved medicines for the treatment, none of them, found helpful in curing fundamental causes of dementia [6]. In some cases cognition is irreversible but treatment in the early stages of dementia can decline the rate [5]. As per World Alzheimer Report the medicines are most effective and accurate when provided in the early stage of dementia [7].

Early prediction of dementia requires different feature patterns of the subjects to train the model. Cross-sectional data or the data at a certain time can be helpful for predicting dementia at a definite level while adding to this longitudinal data is collected over different time can be helpful for identifying a pattern of disease at a wider level. The Alzheimer's Disease Neuroimaging Initiative (ADNI), launched in 2004, provides longitudinal data as a public-private partnership. This longitudinal study aims to measure the progression of early Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) [20]. Three longitudinal studies has been taken place on approximately 1900 participants ageing from 55 to 90 with status Cognitive Normal (CN), MCI and AD. ADNI collects the demographics, MRI, PET, biological markers and cognitive assessments data in the regular intervals of six months (in general).

The ADNI data can be used to predict dementia in initial stage using machine learning algorithms. The task of classification and regression had been very well dealt with the support vector machine (SVM), decision trees, random forest, and other traditional algorithms [16],[17],[18],[19] but deep learning models have achieved state-of-the-art performance on several health prediction tasks [22].

Recurrent Neural Network (RNN) is a popular class of deep learning methods especially used for the sequential data [23]. With the absolute use of data, RNN performs the same task for each element output depending on the previous results to identify the hidden sequences [24]. The RNN keeps on learning from previous outputs without resetting the state. So, the continual input streams cause the network to breakdown, which is improved by the Gated LSTM, helps the network to reset the state when required [25]. The targeted problem can be considered as the sequential classification; hence these gated algorithms give better results [22]. Moreover, deep learning combined with artificial intelligence can perform miracles, but these have the drawback of interpretability, as they learn features in black box [21]. The traditional model elucidates the process very well by defining the feature importance while it is hard to achieve the same for deep models. In this paper, it is targeted to interpret the deep learning model using out-of-the-bag-estimation [26],[31],[39].

The intuition of this study is to provide the probability of having dementia, and expected time of perceiving dementia at the time of subject enrollment using state-of-the-art techniques. In this study, Multimodal data from demographics, MRI and PET is used to compare the results and interpretability of the features governing prediction with traditional algorithms. This research is organized as follows: Section II discusses the similar studies carried out in the field of dementia prediction, followed by the Experimental settings. Finally, results are evaluated and discussed, and the future line of action is outlined.

II. BACKGROUND

A. Data and Health

The studies of many investigators in the prediction of diseases comprise the use of data from social network, medicals, and hospitals. Researchers from the United States developed a supervised classification model for distinguishing the health data from the general data available on Twitter [8]. An expert system had been created using parallel distributed processing based on a multi-layer network for the diagnosis of the diseases created by Japanese fellows following the training by 300 patients [9]. The prediction model for the heart disease created by Raipur Institute of Technology, using Naïve Bayes, Decision Tree, and Clustering which concluded 15 attributes for the prediction of the heart attack [10]. Researchers from the United Kingdom collected data of over 11.3 million patients from the general practitioners across the UK for the observation researches recording Demographics, diagnosis, symptoms, signs, prescriptions, referrals, immunizations, behavioural factors, tests of patients [11]. The vital role of the Big Data analytics in healthcare is covered by the University of Michigan showing how the unstructured or structured data is going to help in the practice of healthcare [12].

B. Dementia and Researched Factors

Studies had revealed the impact of the medical history and other factors for the investigation of dementia. The impact of previous medical history is studied by some researchers claiming that diabetes is a risk factor for dementia after the study from the primary research articles and the data from some electronic databases, they state that during diabetes, glycaemic fluctuations, glycated haemoglobin and hypoglycaemia increase the risk of cognitive decline [14], this describes the fact about the impact of diabetes on dementia growth.

The research found the relationship between the education and cognition power of a person, investigators experimented professional women of age 70-79 years and found women with advanced studies have less cognitive mean than the less educated women [13]. This research had only used the data for female educated women but indicates the impact of education on the cognition power, in this research this variable will be taken on general data and impact will be studied. The other factors like alcohol also contribute to the cognition damage stated in a study that alcohol consumption up to 10g/day improves the cognition ability while more deteriorates it. This study was done on the 13191 weekly drinkers found to have an undergraduate degree and had lower socioeconomic deprivation [15]. Alcohol is a potential variable for dementia prediction while combining it with education and other variables can give effective results.

C. Various Databases Used

A research based on The Health Improved Network [THIN] database is successfully able to predict dementia for aged 60 – 79 years on the basis of the routinely collected data over a period of five years containing lifestyle, cardiovascular and variables of mental health using an algorithm [16]. Moreover, Investigators found the effect of polygenic variation on the risk of dementia, the study used the genotype data from the dataset containing 37154 controls states that the polygenic score is highly significant for the prediction of the Alzheimer disease revealing genetic architecture of Alzheimer have common variants of effects which reflect biological pathway to disease [17]. The above studies had proved the impacts of different variables like lifestyle and genetics in the prediction of dementia. Dementia can be developed by many different orders distinguishing person to person but consolidating different potential variable can give an effective hint for the risk of dementia in the early stages.

One similar work had been done by Tim Wilkinson from the University of Edinburg, attempted to identify dementia on the basis of positive predictive value and sensitivity, his team identified data from Web of Science (Thomson Reuters), MEDLINE (Ovid), PsycINFO (Ovid), EMBASE (Ovid) and CENTRAL (Cochrane Library) published during 1/1/1990 and 14/09/2017, after extraction and selection of the useful data they created a 95% confidence interval by Clopper Pearson method. Later they identified that primary data can give a more accurate value of PPV and the hospital and death data can help in getting good value of the sensitivity and stated sensitivity data is not much efficient for identifying dementia comparing to other diseases maybe because of the less diagnosis and data available in hospitals [18]. This study was more focused on primary care data, while good quality results can be achieved using other potential variables.

One more similar study reveals facts about dementia prediction done by Korean researchers on The Korean National Health Insurance Service Senior Cohort Database consisting more than half a million senior data using the Support Vector Machine with 10-fold cross-validation analysis, this study sampled 11,443 cases and dividing two groups, 850 with dementia and 850 normal cases selected randomly and studied different features, concluding the fact that last two year medical history has a powerful impact on the prediction of dementia [19]. This study was more focused on finding the variable instead of working on different algorithms while researching on algorithms and adjusting them gives better results which the researchers had not focused, though this study has good results they were on the Korean database and investigators had not mentioned its feasibility on other parts of the world. Other databases may give a different result and this research can be used as a comparison which can reveal the sustainability of the impact of variables in the different demographics which further can be used as a variable.

D. ADNI Database based Researchs

A set of risk factors and the prediction of the dementia is shown by the University of the Novi Sad, collection of the different algorithms used and their efficiency in different papers[38]. A study based on Magnetic Resonance Imaging(MRI) reveal, its efficiency on dementia prediction [40]. A report by Garam Lee on ADNI dataset was able to achieve an accuracy of 0.81 using the GRU, classifying between MCI and AD [22] while they had not shown any results between CN and AD. The latest paper published in June 2019 shows the use of multimodal longitudinal data integration framework for multiple forms of ADNI numeric data for extracting the complimentary features, this task was carried out for the MRI classification, the framework besides work on the deep learning techniques and shows the efficiency of the state-of-the-art algorithms in Alzheimer's prediction [37].

E. Deep Learning Interpretability Studies

A probabilistic approach is proposed by Fisher for the feature importance in black-box[26], a recent work by Bokang Zhu shows features can be selected in the RNN using their memorization in network, a test called Flagged-1-Bit is proposed for the sequence learning interpretability [39], this paper gives the idea of testing the same model on multiple features, which is basis for the method used in this study.

These study by the different researchers across the world can help in identifying the different perspective of predicting dementia in the early stage based on the different criteria solved out from the above researches. The proposed research introduces an interpretable deep model to predict dementia at the time of subject enrolment trained on ADNI dataset.

III. Experimental Setup And Results

A. Task Overiew

This study finds the optimized way of predicting dementia using a combination of two classifiers and one regressor on cross-sectional and longitudinal data of ADNI as shown "Fig. 1", Model 1 will be trained on Cross-Sectional data and will classify the subject as Dementia(Yes) or not.

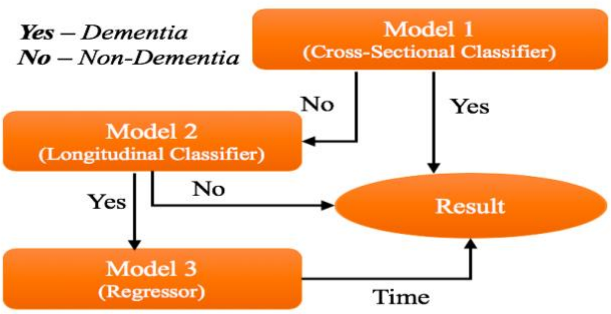

Fig.1. Base Approach for Dementia Prediction

If Yes, it can be said as subject have probability of having dementia in the current stage, if not it will be tested on Model 2 which is an again classifier trained on Longitudinal data depicts that the subject will have the probability of having dementia in future or not, if yes then Model 3 will predict the expected time of getting dementia, which is a longitudinal data trained regressor. The interpretability of deep learning model will be studied on Model 1 compared with other traditional models.

B. Understanding Data

As mentioned earlier, the study had been carried out on the data provided by ADNI. The initial study ADNI1's longitudinal data is used to examine the efficiency of the model, "Fig. 2" shows the number of participants on different visits(VISCODE), distinguishing among DX(Cognitive Status) CN, MCI and AD. On baseline(bl) 193 subjects were enrolled with dementia, 397 with MCI while 229 cognitive normal nature. Considering the fact that among CN and MCI few turned to dementia in second visit(m06) and some in next month and so on.

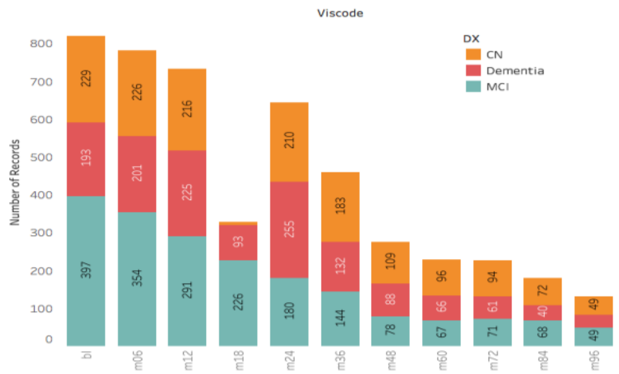

Fig. 2. ADNI 1 Subjects' Visit

Here, the data at a particular visit is used as cross-sectional data while the data in all visits by an individual is used as Longitudinal data refer pre-processing section for better understanding of how data is used. The demographics features used in the study are education, age, gender, ethnicity, race, marital status and the cognitive score obtained by assessment of subject based on study [22], the MRI features are the BSI(Boundary Shift Region) extracted by UCL Institute of Neurology[26] includes whole-brain, ventricles, hippocampus, fusiform, mid-temp. Lastly, the Positron Emission Tomography(PET) features involves mean of FDG-PET within ROI in different Location and Laterality viz Angular Left, Angular Right, Cingulum Post Bilateral, Temporal Left and Temporal Right extracted by UC Berkeley and Lawrence Berkeley National Laboratory for ADNI. For the binary classification, MCI and CN are considered as Non-Dementia while AD is the Dementia patients, so the availability of the data for the ADNI1 in three different modals along with ensemble which contains relative available data in three modals is described in Table I.

TABLE I: Data Availability in Modals

Data Modal	Status	
	Non-Dementia	Dementia
Demographic	626	193
MRI	512	140
PET	305	97
Ensemble	258	72

C. Data Preprocessing

The ADNI dataset contains hundred of files available in CSV(Comma Separated Value) format, the primary task is to extract the useful files which can be done by reading the data dictionary provided by ADNI. The useful files are imported using Pandas into data frames which can be further used for the data pre-processing.

1) Missing Data

The missing data in the provided ADNI files are generally coded as -1 and -4 for missing value and not applicable respectively. These missing values are treated as Null values in Pandas, the data for demographics is available while data for MRI and PET is missing when the data is not collected on the visit site or subject not available on the MRI and PET sessions. Hence, the complete data is considered as MCAR (Missing Completely at Random) due to this fact complete data for the visit is eliminated from the training data.

2) Data Selection

ADNI files contain data for all studies, this paper uses ADNI 1 data, due to less complexity while other studies have a complex structure for defining dementia.

3) Data Transformation

Data from multiple files are merged into one data frame based on the RID(Roster ID) and VISCODE of the participants. Multiple transformations needed in different files to get them in the same page, PET data is provided in the long-form which is needed to be converted into the wide form using pivoting. The data for status MCI and CN is aggregated for the binary classification. To evaluate the longitudinal data for model 2 and model 3 in the base approach, the complete time-series of a subject is required to be considered. The status of a subject is considered to be as dementia if they have dementia in future then complete records will be updated with the status dementia for the subject and time will be calculated as the difference of days from the visit exam date, the status of the subject changed as shown in “Fig. 3”.

Fig. 3. Longitudinal Transformation

Moreover, categorical features i.e. Gender, Ethnicity, Race, Marital status and Status are nominal data and needed to be factorized which is done using LabelEncoding function provided by Scikit Library which normalizes the labels into 0 to n-classes-1 [27].

D. Models

As described previously this paper is targeting on the performance and elucidation of models, especially for a deep model. Different classification algorithms will be used on Model 1(Cross-Sectional Data) along with the Deep Model(LSTM).

1) Logistic Regression

The Logistic Regression Model is a classification model for a binary dependent variable. This algorithm use a sigmoid function σ : R-->(0,1) defined as: $\sigma(t) = \frac{1}{1+e^{-t}}$ [28], in this study Logistic regression have been implemented using scikit linear_model library [29]. The coefficients from the model can be used as the feature importance, more the coefficient absolute value, more important the feature.

2) Support Vector Machine(SVM)

A linear SVM kernel is used in the study due to its easy interpretation of nature. Linear SVM finds the maximum margin hyperplane that can divide the group of data into the number of required categories(here: binary). Scikit implementation is used for the experiment [30]. Feature importance in linear SVM is also calculated using the coefficient value.

3) Random Forest Classifier

The combination of many predictive trees is Random forest, depending on the random vector forest select the optimized feature for the classification. A random forest is simply interpretable which uses the out-of-the-bag estimates to determine the feature importance [31].

4) Deep Model

In this study, the Long Short-Term Memory(LSTM) is used as the base cell for training and testing. LSTM is a gated RNN, Recurrent Neural Network(RNN) is a deep learning class specially used for the sequential data. One element of the input sequence is processed at the time, t and its memory state get updated which beholds the history of all elements of the sequence. For N subjects, let each have the sequence $\{x_1, x_2, x_3, \ldots, x_T\}$, having T as the length of the sequence, s state having the hidden units, s_t (“Fig. 4”) and U, V, W the weight metrics extracts the features from the previous memory produces output sequence, o. Here,

$$s_t = tanh(Ws_{t-1} + Ux_t) \text{ and } o_t = \sigma(Vs_t)$$

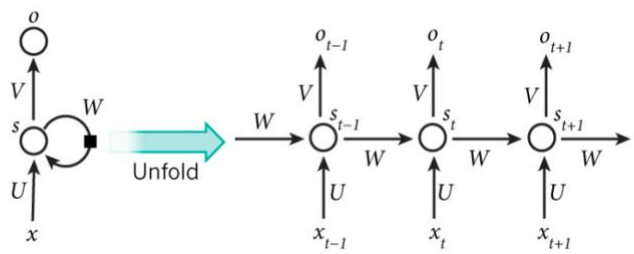

Fig. 4. RNN and Unfolding [32]

Where σ is the SoftMax function. The last output sequence of RNN is treated as the probability vector for classification [32]. RNN is own very effective but training in the RNN get difficult for the long sequences of inputs [33]. LSTM [34] is developed over RNN to fix the problem of Long-term dependency problem using Gated approach, the LSTM cell consists of different layers which can be explained as forget gate layer, input gate layer and output gate layer as shown in “Fig. 5”.

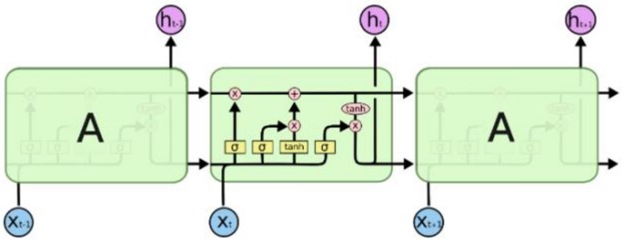

Fig. 5. LSTM Structure
source:https://colah.github.io/posts/2015-08-Understanding-LSTMs/

The forget gate layer consisting of a sigmoidal function which helps in deciding which data to perceive and to eliminate, while the input gate layer decides the values to be updated and create vector using a combination of sigmoidal and tanh functions, now the output layer again consisting of a combination of a sigmoidal and tanh function help in producing the final result of the cell. LSTM updates the state which helps to calculate the cost which is further needed to be optimized can be done using AdamOptimizer [35]. Adam algorithm is highly effective as it achieves results very fast. The algorithm use average of the second moments of the gradient along with the mean(first moment) to update the learning rate, also calculates the exponential moving average of the gradient along with the squared gradient, and the beta1 and beta2 control the decay rate of the moving averages. In this study, Tensorflow implementation is used which initialize the values, learning_rate = 0.001, beta1 = 0.9, beta2 = 0.999 [36].

E. Interpretability

The Model 1(Cross-Sectional Classifier) is used for the study of the interpretation of the deep model hence the same data had been used in different algorithms to get the feature importance and evaluation metrics. The experimental setting is done by fixing the constant weight(initial weight to feed LSTM), bias, batch size(Number of sequences for cost optimization) and epoch. Then, the feature importance for the deep model is calculated by out-of-the-bag estimation i.e. deep model is being evaluated by iteration of the model, each time eliminating one feature and assigning the difference between the actual model F1 Score with a new model after removing a feature given as

$$f_{imp} = f1_s - f1_{s-f}$$

where f_{imp} is the importance of a feature, $f1_s$ is f1 score with all features and $f1_{s-f}$ is the f1 score achieved by removing the feature.

Later, for the comparison of the feature importance in different models, all the values are normalized in the scale from 0 to 1. This can give an insight how a single feature is important in different models which have a different approach for the prediction like the coefficients in Logistic regression and Support Vector Machine or the out-of-the-bag approach of the Random Forest Classifier.

F. Results and Evaluation

The data is randomly split into 80-20% for training and testing of the model. Further, models had been 5-Fold cross-validated accordingly.

The evaluation of the classification models(Model 1 and Model 2 in Figure 1 is done using the f1 score, ROC-AUC(Area under the curve) and accuracy while the regressor model(Model 3) is evaluated by optimizing the root mean squared error. The formula for calculating the above metrics are shown in Table II.

TABLE II: Metrics used for Evaluation

Metric	***Evaluation***
F1 Score	$\frac{2 * (Precison * Recall)}{Precison + Recall}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Root Mean Squared Error	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(O - E)^2}$

where TP is True Positive, TN is True Negative, FP is False Positive, FN is False Negative, N is total items, O is Observed Result and E is Expected Result.

The primary goal of this study was to produce an effective way of dementia prediction using deep learning and interpret it, the initial experiments involve the model performance on different data modal shown in Table III.

TABLE III: Achieved results. Column 2 and 3 is accuracy and F1-Score respectively achieved by LSTM model evaluation on test data of modal.

Modal	**Accuracy**	**F1 Score**
Demographic	0.79	0.719
MRI	0.80	0.445
PET	0.76	0.621
Ensemble	**0.83**	**0.743**

Based on the performance in different modal, the ensemble is selected for the further process. The Model 1(Cross-Sectional Classifier) is used for the study of the interpretation of the deep model hence the same data had been used in different algorithms to get the feature importance and evaluation metrics. The Receiver Operating Characteristics(ROC) curves for the participating models are shown in “Fig. 6”.

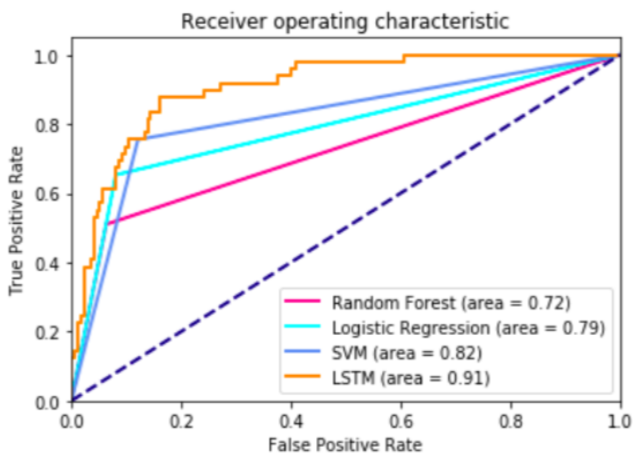

Fig. 6. Cross-Sectional Classifier ROC curves and AUC

TABLE IV : Normalized Feature Importance Score. Column 1 list the features used Models and Column 2-5 list the normalized scores by the mentioned model

Algorithm (Accuracy) / Features	LSTM (0.86)	Logistic Regression (0.86)	SVM (0.85)	Random Forest (0.84)
Cognitive Score	**1.000**	**1.000**	**1.000**	**1.000**
Temporal Left	**0.423**	0.030	0.057	**0.211**
Race	**0.411**	0.061	0.000	0.000
Cingulum Post Bilateral	**0.408**	0.024	0.056	**0.450**
Angular Left	**0.214**	0.034	0.066	**0.324**
Gender	0.090	0008	0.046	0.005
Marital Status	0.076	**0.103**	0.093	0.021
Angular Right	0.071	0.031	0.058	**0.276**
Hippocampus	0.055	0.000	0.000	**0.260**
Age	0.053	0.027	**0.198**	**0.122**
Ventricles	0.045	0.000	0.000	**0.214**
Mid-Temp	0.043	0.000	0.000	**0.336**
Ethnicity	0.025	0.027	0.021	0.019
Temporal Right	0.007	0.019	0.042	**0.231**
Education	0.000	0.071	**0.241**	**0.118**

ROC curve is a plot between True Positive Rate(tpr or recall) and False Positive Rate(fpr = $\frac{FP}{FP+TN}$) defines the efficiency of a model for distinguishing between classes. Clearly, LSTM has an advantage over the other traditional algorithms acquiring Area Under the Curve(AUC) of 0.91 implies the model has a 91% chance of successfully distinguishing between dementia and non-dementia, making it the best classifier among the four classifiers.

It is required to verify the importance of features that LSTM gathers using the proposed approach with traditional approaches. Table IV shows the normalized scores of the features in the models and accuracy achieved by the model. The accuracy achieved by all the models can be considered as similar. So, the interpretability of the LSTM can be validated by comparing the score of the features with other models, the most important feature in LSTM is Cognitive Score, while all other models have same feature with highest importance, hence the interpretability method used for LSTM can be verified in a sense. Moreover, it is visible from the normalized feature score that LSTM had given higher importance viz greater than 0.1 (bold in Table IV) to 5 features while logistic regression(LR) relies highly on just two features and SVM on three, shows the inability of LR and SVM to identify the sequence and ability of LSTM to relate between the different features. Random Forest Classifier makes the best use of features by providing importance to 11 out of 15 features though only two feature have importance of greater than 40%, while comparing to LSTM which have four features contributing more than 0.4. Moreover, the top feature scores of LSTM looks very similar to the random forest due to the fact that both algorithms prunes the features to reduce the cost. Hence, from the scores and analysis, it can be said that the deep model like LSTM can be interpreted with the proposed approach.

After the successful test of the LSTM for Cross-Sectional Classifier, the same is applied for the longitudinal Classifier giving an accuracy of 0.85 and F1 Score 0.78. Longitudinal Regressor is used to predict the time which gives root mean squared error of 18.62 which is quite well, implies when the time for the subject to perceive dementia will be calculated, there will be a tolerance of 18.62 months must be needed to consider.

"Fig. 7" shows the visual comparison of the predicted and true values(ground truth). The Model is able to regress, the

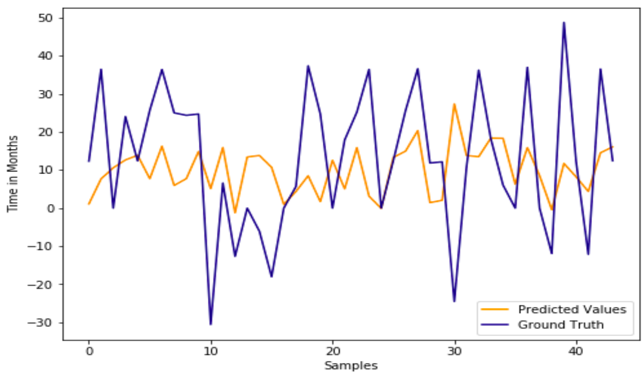

Fig. 7. Time Prediction Ground Truth V/S Predicted Values

Time in months quite close to the actual values. For sample between 0 and 10 predicted value is almost equal to the actual time in months, while there are contrary prediction visible at sample 30 but the figure own says the prediction is almost able to identify the pattern of regression.

All the models experimented and finalized had been cross-validated using 5-Fold Cross-Validation, the results for all deep models are shown in Table V. The classifiers are highly efficient with the average of 83.9% and 86% for the Model 1 and Model 2 respectively, the less value of Standard Deviation(SD) shows the absence of over-fitting in models. The regressor has an average RMSE of 18.52 with a deviation of 0.97, shows no overfitting for the data but there is a scope of improvement in the regression or approach for time prediction.

TABLE V: Results achieved by 5-Fold Cross-Validation

Model	**Result**		
	Metrics	**Mean**	**SD**
Cross-Sectional Classifier	**Accuracy**	0.839	0.031
	F1 Score	0.781	0.045
Longitudinal Classifier	**Accuracy**	0.860	0.012
	F1 Score	0.814	0.016
Regressor	**RMSE**	18.52	0.97

IV. CONCLUSION

As Dementia is a leading problem of the world, researches are required to be done, many previous piece of researches showed that there is a scope of the machine learning prediction for identifying the risk of dementia, and deep learning is considered as a boon but they are still unemployed due to their less interpretability, this paper is successfully able to employ an efficient approach to predict dementia using state-of-the-art techniques with their elucidation at a certain level. This research can help many families for the early diagnosis based on regular data collected from the people.

V. FUTURE WORK

This study was limited and generalized for prediction and elucidation and highly focused on algorithmic implementation, while focusing more on feature selection may result in better accuracy. The work had already been extended to study interpretability with other features on the same database. Further, it is planned to compare the results and validate with other databases in the similar field. The feature importance was tested on the proposed approached only on LSTM, a test on other algorithms may give more insights and better way to interpret the results. Moreover, the regression of time of person getting dementia is less researched field and requires more work to be done.

ACKNOWLEDGEMENT

I thank DCU for providing this opportunity to present my work. I am highly thankful to my supervisor Prof. Tomas Ward for guiding, motivating and availing the required resources throughout the practicum. This work was not possible without the knowledge I had gained from Module Professors who skilled me to overcome the technical and non-technical challenges faced during the practicum.

REFERENCES

[1] World Health Organization. (2018). The top 10 causes of death. [online] Available at: http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [Accessed 22 Nov. 2018].

[2] World Health Organization. (2018). Dementia. [online] Available at: http://www.who.int/news-room/fact-sheets/detail/dementia [Accessed 25 Nov. 2018].

[3] Alz.co.uk. (2018). [online] Available at: https://www.alz.co.uk/sites/default/files/pdfs/global-impact-dementia-infographic.pdf [Accessed 25 Nov. 2018].

[4] Deborah Brooks; Catherine Ross; Elizabeth Beattie, Caring For Someone With Dementia: The Economic, Social, And Health Impacts Of Caring And Evidence Based Supports For Carers, Dementia Care School of Nursing, Queensland University of Technology, Paper 42 October 2015.

[5] Alzwisc.org. (2018). Importance of an early diagnosis. [online] Available at: https://www.alzwisc.org/Importance%20of%20an%20early%20diagnosis.htm [Accessed 25 Nov. 2018].

[6] Anand, P.; Singh, B. A review on cholinesterase inhibitors for Alzheimer's disease. Arch. Pharmacol. Res. 2013, 36, 375–399

[7] Prince, M.; Bryce, R.; Ferri, C. World Alzheimer Report 2011. The Benefits of Early Diagnosis and Intervention; Alzheimer's Disease International: London, UK, 2011.

[8] David A. Broniatowski,; Micheal J. Paul; Mark Dredze, National and local influenza surveillance through Twitter: An analysis of the 2012–2013 influenza epidemic. PLoS ONE 2013, 8, e83672.

[9] K. Saito, R. Nakano, Medical diagnostic; expert system based on PDP model, Proc. IEEE International Conference on Neural Networks, Vol 1 (1988), pp. 255-262.

[10] Jyoti Soni; Ujma Ansari; Dipesh Sharma; Sunita Soni, International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

[11] Byeon, H.; Lee, Y.; Lee, S.Y.; Lee, K.S.; Moon, S.Y.; Kim, H.; Choi, S.H. Association of alcohol drinking with verbal and visuospatial memory impairment in older adults: Clinical Research Center for Dementia of South Korea (CREDOS) study. Int. Psychogeriatr. 2015, 27, 455–461.

[12] Ashwin Belle,[1,2] Raghuram Thiagarajan,[3] S. M. Reza Soroushmehr,[1,2] Fatemeh Navidi,[4] Daniel A. Beard,[2,3] and Kayvan Najarian,, Big Data Analytics in Healthcare, BioMed Research International Volume 2015, Article ID 370194, 16 pages.

[13] Lee, S.; Kawachi, I.; Berkman, L.F.; Grodstein, F. Education, other socioeconomic indicators, and cognitive function. Am. J. Epidemiol. 2003, 157, 712–720.

[14] Ojo, O.; Brooke, J. Evaluating the association between diabetes, cognitive decline and dementia. Int. J. Environ. Res. Public Health 2015, 12, 8281–8294.

[15] Giovanni Piumatti; Simon C Moore; Damon M Berridge; Chinmoy Sarkar; John Gallacher , The relationship between alcohol use and long-term cognitive decline in middle and late life: a longitudinal analysis using UK Biobank, Journal of Public Health, Volume 40, Issue 2, 1 June 2018, Pages 304–311.

[16] Walters, K.; Hardoon, S.; Petersen, I.; Iliffe, S.; Omar, R.Z.; Nazareth, I.; Rait, G. Predicting dementia risk in primary care: Development and validation of dementia Risk Score using routinely collected data. BMC Med. 2016, 14, 6.

[17] Valentina Escott-Price, Common polygenic variation enhances risk prediction for Alzheimer's disease, Brain, Volume 138, Issue 12, 1 December 2015, Pages 3673–3684.

[18] Wilkinson, T., Ly, A., Schnier, C., Rannikmäe, K., Bush, K., Brayne, C., Quinn, T. and Sudlow, C. (2018). Identifying dementia cases with routinely collected health data: A systematic review. Alzheimer's & Dementia, 14(8), pp.1038-1051.

[19] Kim, H., Chun, H., Kim, S., Coh, B., Kwon, O. and Moon, Y. (2017). Longitudinal Study-Based Dementia Prediction for Public Health. International Journal of Environmental Research and Public Health, 14(9), p.983.

[20] Adni.loni.usc.edu. (2019). ADNI | About. [online] Available at: http://adni.loni.usc.edu/about/ [Accessed 24 Nov. 2018].

[21] Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B. and Yang, G. (2017). Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), pp.4-21.

[22] Lee, G., Nho, K., Kang, B., Sohn, K. and Kim, D. (2019). Predicting Alzheimer's disease progression using multi-modal deep learning approach. Scientific Reports, 9(1).

[23] Mikolov T, Karafiat M, Burget L, Cernocky J, Khudanpur S (2010) Recurrent neural network based language model, In Proc of Interspeech 2010, pp. 1045–1048.

[24] Lu, H., Li, Y., Chen, M., Kim, H. and Serikawa, S. (2017). Brain Intelligence: Go beyond Artificial Intelligence. Mobile Networks and Applications, 23(2), pp.368-375.

[25] Gers, F., Schmidhuber, J. and Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. Neural Computation, 12(10), pp.2451-2471.

[26] Fisher, A., Rudin, C., Dominici, F.: All models are wrong but many are useful: Variable importance for black-box, proprietary, or misspecified prediction models, using model class reliance. arXiv preprint arXiv:1801.01489 (2018).

[27] Scikit-learn.org. (2019). 5.9. Transforming the prediction target (y) — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/preprocessing_targets.html#preprocessing-targets [Accessed 2 Mar. 2019].

[28] Hosmer, David W.; Lemeshow, Stanley (2000). Applied Logistic Regression (2nd ed.). Wiley. ISBN 978-0-471-35632-5.

[29] Scikit-learn.org. (2019). sklearn.linear_model.LogisticRegression — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html [Accessed 2 Mar. 2019].

[30] Scikit-learn.org. (2019). 1.4. Support Vector Machines — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/svm.html#svr [Accessed 2 Mar. 2019].

[31] Breiman, L. Machine Learning (2001) 45: 5. https://doi.org/10.1023/A:1010933404324.

[32] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444, https://doi.org/10.1038/nature14539 (2015).

[33] Bengio, Y., Simard, P. & Frasconi, P. Learning long-term dependencies with gradient descent is di cult. IEEE Trans Neural Netw 5, 157–166, https://doi.org/10.1109/72.279181 (1994).

[34] Hochreiter, S. & Schmidhuber, J. Long short-term memory. Neural computation 9, 1735–1780 (1997).

[35] Kingma, Diederik P and Ba, Jimmy Lei. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[36] TensorFlow. (2019). tf.train.AdamOptimizer | TensorFlow Core r1.14 | TensorFlow. [online] Available at: https://www.tensorflow.org/api_docs/python/tf/train/AdamOptimizer [Accessed 1 Aug. 2019].

[37] Lee, G., Kang, B., Nho, K., Sohn, K. and Kim, D. (2019). MildInt: Deep Learning-Based Multimodal Longitudinal Data Integration Framework. Frontiers in Genetics, 10.

[38] Bratić, B., Kurbalija, V., Ivanović, M. et al. J Med Syst (2018) 42: 243. https://doi.org/10.1007/s10916-018-1071-x.

[39] Zhu, B.,Zhang, R., Long, D. and Mao, Y. (2019),Understanding Feature Selection and Feature Memorization in Recurrent Neural Networks, arXiv e-prints, arXiv:1903.00906.

[40] Jack CR Jr, Barnes J, Bernstein MA, et al. Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2. Alzheimers Dement. 2015;11(7):740–756. doi:10.1016/j.jalz.2015.05.002