Exploration of Machine Learning and Deep Learning Architectures for Dementia Risk Prediction Based on ATN Framework

Samuel O. Danso
School of Computer Science
University of Sunderland, United Kingdom.
Centre for Clinical Brain Sciences, University of Edinburgh. UK
sam.danso@sunderland.ac.uk

Sindhu Prattipati
School of Informatics
University of Edinburgh, United Kingdom
sindhuprattipati@yahoo.com

Ibrahim Alqatawneh
School of Computer Science
University of Sunderland, United Kingdom.

Georgios Ntailianis
Centre for Clinical Brain Sciences
University of Edinburgh, United Kingdom

***Abstract*—Despite the high incidence of Alzheimer's disease (AD), there is no cure for AD yet. Therefore, early identification of individuals at higher risk of developing AD becomes critical, as this may provide a window of opportunity to adopt lifestyle changes to prevent or delay the onset of the disease. We propose a novel approach to developing prediction models using Feed forward Deep Neural Networks. Our models are built using the EPAD LCS v.IMI dataset. We extract a combination of brain imaging, genetics, cognitive and lifestyle features from the dataset to build the prediction models. The prediction is based on the ATN classification framework, with prediction categories of Healthy, Suspected Non-Alzheimer's Pathology (SNAP), and Dementia due to AD continuum. We built a total of 6 prediction models, of which 4 are based on classic Machine Learning (ML) and 2 are Deep Learning (DP) approaches. The best DP model outperforms the classic ML model by F1 score of 14% and AUC score of 13%. We have demonstrated that our Deep Learning-based model has the potential to be deployed as a screening model to predict dementia risk at early stage of the disease.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease and is the most common form of dementia. It is clinically characterized by cognitive impairment and memory loss. AD causes individuals to gradually lose their cognitive abilities, and thus affects their regular life and their relationships. AD brings about behavioural changes and causes cognitive decline to the extent where the individual becomes completely dependent on others for performing regular tasks. Additionally, AD affects the wider society because people with dementia also require health and social care and this has high costs associated with it [1].

AD remains the most prevalent of all types of dementia, accounting for over 60 % of dementia cases [2]. Therefore, finding disease-modifying therapies for AD remains an international priority. The most effective approach may be to slow or prevent AD progression prior to dementia. Secondary prevention strategies aim to identify individuals with evidence of AD pathology but who have not (yet) developed symptoms, that is, at a preclinical stage. To provide a reference framework, an update on the biological definition of AD has been published, describing the disease solely in terms of biomarkers [3].

In 2011, the National Institute on Ageing and Alzheimer's Association created separate diagnostic recommendations for the preclinical, mild cognitive impairment, and dementia stages of Alzheimer's disease. [3] This recommendation led to the development of a framework that allows Alzheimer's disease (AD) to be defined by its underlying pathological processes, which can be documented by examination or in vivo by biomarkers. The diagnosis is therefore based on the biological construct (i.e., biomarkers) instead of clinical manifestations (i.e., symptoms/signs), and this is important for early detection. Biomarkers are grouped into those of Aβ amyloid deposition (A), pathological tau (T), and neurodegeneration(N), which is therefore known as the ATN framework for classification of dementia. The ATN classification system seeks to create a unified framework with which researchers can generate and test hypotheses about the interactions among different biomarkers and cognitive symptoms [13]. The framework enabled us to define three groups of participants as:

- AD continuum - suggesting positive amyloid with positive tau pathology and negative neurodegeneration.
- Suspected non-Alzheimer's Pathology (SNAP) - suggesting a negative amyloid with positive tau pathology and/or positive neurodegeneration.
- Healthy - suggesting negative biomarkers.

In the last few years, improvements in medical imaging,

exponential increase in computational power of affordable computing platforms, and greater availability of medical data sets have increased opportunities to develop machine learning approaches to automate detection,classification, and quantification of diseases. [8] Machine Learning (ML) and Deep learning (DL) are being successfully employed in Healthcare for accurate diagnosis of diseases. There is a lot of scope to apply ML for AD prediction as well. Numerous attempts have been made to develop ML models to predict dementia risk [5] [6]. Similarly, DL learning approaches have shown to be capable of predicting dementia risk with reasonable accuracy [7] [8]. In particular, DL approaches such as Deep Neural Networks(DNNs) are increasingly being employed to develop risk prediction models for dementia, but have primarily focused on imaging data and other biomarkers frequently [7] [8] [11].

In this paper, we examine the advantage of employing ML and DL in dementia risk prediction based on the ATN framework. To the best of our knowledge this is the first paper to explore and compare the performance of ML and DL models developed using a combination of on a combination of brain imaging, cognitive and genetics as well as lifestyle factors to predict AD risk based on the ATN framework for dementia classification.

II. Materials and Methods

A. Dataset Description

We used the 'EPAD LCS v.IMI' dataset, which is the final dataset in the European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS) [4]. The EPAD LCS, remains one of the largest multi-site dementia studies in the world, with participants from 39 research sites across Europe to serve as platform for research into Alzheimer's disease and to facilitate discovery of new preventative treatments. It contained a total of 2,096 participants, collecting a wide range of cognitive, clinical, neuroimaging and biomarker data to help understand the early stages of Alzheimer's disease. It is a longitional study with visits for the participants we scheduled for follow-up visits in 6 months ,1 year, 2 years, 3 years, as shown in Table 1. Data collection included: cognition, neuroimaging, fluid biomarkers, genetics, lifestyle, clinical and psychiatric assessment, neuropsychiatric symptoms, function and basic demography.

TABLE I
Summary of number of records per visit.

Visit	Frequency
Baseline (Visit 1)	2096
6 months (Visit 2)	1596
1 year (Visit 3)	1225
2 years (Visit 4)	421
3 years (Visit 5)	121

B. ATN Framework-based Class Definition

Earlier work by Calvin et al. [5] proposed cut-off values based on the EPAD LCS dataset. The proposed cut-off values range from 880 to 1100 pg/mL for cerebrospinal fluid (CSF) Aβ1-42, and from 19 to 27 pg/mL for CSF phosphorylated tau (p-tau). Average Medial Temporal Atrophy (MTA) was used as a neurodegenerative factor (N) with values > 1 considered to be positive otherwise it is negative as summarised in Table 1 below. It was therefore logical to generate class labels based on these cut-offs with their respective classes as shown in Table 2: Healthy, AD Continuum 3 and SNAP (suspected Non-Alzheimer's pathology)

TABLE II
Defining Biomarkers using cut-offs

Biomarker	Positive	Negative
Amyloid Aβ1-42(A)	A <= 1000 pg/mL	A > 1000 pg/mL
P-tau (T)	T > 27 pg/mL	T <= 27 pg/mL
Average Medial Temporal Atrophy MTA (N)	N > 1	N <= 1

TABLE III
Defining class labels based on above cut-offs

Amyloid status (A)	**Tau values (T)**	**Neurodegenerative values (N)**	**Class**
Negative	Negative	Negative	**Healthy**
Positive	Positive	Negative	**AD Continuum**
Negative	Either one or both Positive		**SNAP(suspected non-Alzheimer's pathology)**

We then applied these categories to the dataset and obtained class distribution as contained in the data. Figure 1 shows the distribution with the Healthy group representing the majority of 53.31%, followed by the AD Continuum group which have 33.45%. The SNAP is the least represented containing 13.24% of the data.

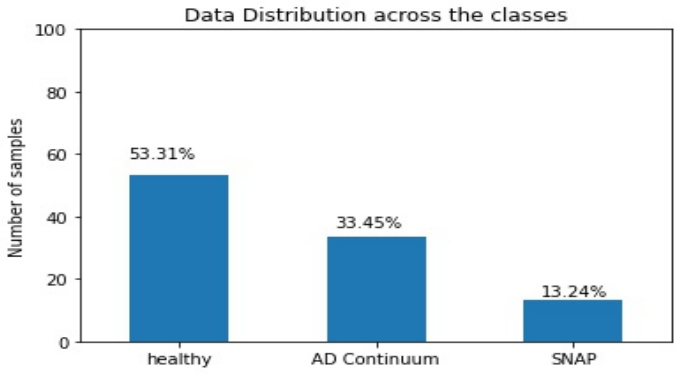

Fig. 1. Percentage of each class in the dataset

C. Model Development Pipeline

Fig. 2 shows the development pipeline of how we built the models. We began with exploratory analysis of the dataset to identify and selected the relevant variables of interest from the dataset. These variables were pre-processed into feature vectors employing various techniques. We then moved to the model development stage where various algorithms were selected and trained to build the models. Models were further optimised by turning the parameters to obtain the hyper-parameters that produced the optimal performance. We discuss the various steps in detail subsequently.

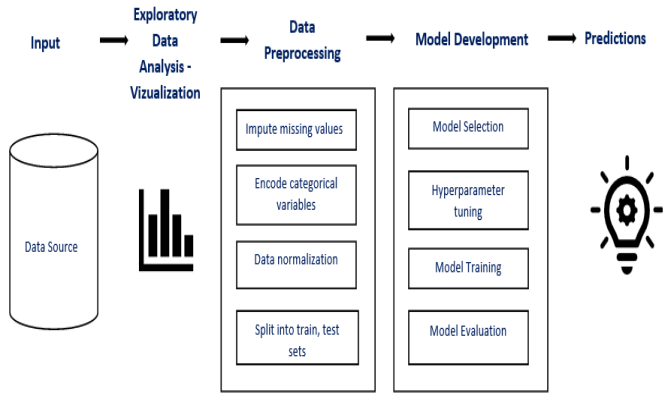

Fig. 2. Machine Learning Pipeline

1) Pre-processing and Feature Engineering: Earlier research identified risks factors of Dementia [14]. We aligned our feature selection scheme with this research and selected these risk factors from the EPAD LCS v.IMI dataset to build our model. We cleaned the data of outliers and nulls before encoding the non-numeric features. All categorical features were converted to numeric [**?**]. We outline how each feature was pre-processed and engineered prior to model training:

- Age: age at last visit was computed
- Ethnicity: binary encoded into 'Caucasian / not' to avoid problems arising due to high dimensionality (95%, majority of the participants were Caucasian)
- Gender: binary encoded this feature (data has 2 genders - male, female)
- Marital status: divided the different categories in this column (which are 'married/ cohabiting', 'single', 'widowed', 'divorced') into 'married/ not' by grouping the latter 3 categories into not married and binary encoding the resultant column.
- Family dementia history: grouped this feature as 'Parental diagnosis of AD/ not'. We focused on parental diagnosis [16] as it is more relevant to the 'early diagnosis and prevention' of Alzheimer's.
- Medical History of a patient: For our project we considered the medical conditions that are known risk factors for Alzheimer's. These are: 1) aural problems, 2) mental health conditions, 3) Hypertension, 4) Diabetes, 5) neurological ailments. [14] We used binary encoding on these columns based on whether/ not each patient has these condition(s). We did not opt for ordinal encoding to avoid the problem of label bias in the deep neural network, which could arise due to incorrect ordering of data [17].
- Physical activity: The data for each patient had different categories for physical activity, which were 'exercises daily', '2-3 times a week',' once a week', '2-3 times a month', 'few times a year', 'never'. We grouped the latter 3 categories into 'not physically active' and rest into 'physically active' and binary encoded the result into physically active/ not.
- Smoking and alcohol: grouped 'never', 'past ', 'current' as "smoker" otherwise "not" and binary encoded this column to smoker/ not. We similarly grouped categories on alcohol consumption and binary encoded the column to alcoholic/ not.
- Drug addiction: binary encoded this column into drug addict/not.
- Systolic blood pressure and diastolic blood pressure: computed blood pressure values as (systolic blood pressure/ diastolic blood pressure) [18] over the 5 visits for each patient and selected the maximum reading as the final blood pressure. The reason for giving importance to the maximum reading is due to the relation between hypertension and AD [1] and also because we wanted to handle worst-case scenarios and to investigate any unusual spikes in BP in our framework. We compared the recorded high blood pressure values with self-reported hypertension of the participants to generate a binary variable indicating hypertension.
- BMI: computed for each patient using the formula, BMI = [weight / (height squared in mts.)] and further grouped values into 3 categories of BMIs- underweight, normal and overweight [19] and encoded these categories using a label encoder, based on our understanding that lower BMI increases the risk of Dementia in middle age [20]
- Apoe e4: binary encoded to indicate the presence or absence of apoe e4 allele for each patient as the presence of this genotype increases the risk of AD [21]

- GDS, PSQI, STAI: removed any outliers and kept the values intact as they are already numeric.
- Four mountains score- We calculated a cumulative score based on the participant's correct responses to the four mountains questionnaire.

Table 4 below shows a summary of the risk factors represented as features used to build the models along with the employed encoding strategy.

TABLE IV
FEATURE SELECTION & PREPROCESSING

Category	**Features**	**Binary Encoded Features**
Sociodemographic features	age, ethnicity, gender, marital status, years of education	ethnicity, gender, marital status
Physical exam results	medical history - diabetes, hypertension, hearing disorders, neurological disorders, mental health conditions family history of dementia,reported clinical diagnosis	diabetes, hypertension, hearing disorders, neurological disorders, mental health conditions
lifestyle factors	physical activity, smoking, drug abuse, alcohol consumption	physical activity, smoking, drug abuse, alcohol consumption
Vital signs	blood pressure, BMI	Blood pressure (Hypertension/Normal), BMI (Obese/normal).
Cognitive assessment	GDS (Geriartic Depression Scale),PSQI (Pittsburgh Sleep Quality Index), STAI-40 (State-Trait Anxiety Inventory), Four mountains test score	Not encoded
Biomarkers	APOE e4	APOE e4
Imaging	MTA (Medial Temporal Atrophy)	Not encoded

2) Model Development: For all the models, we divide the original dataset into training and test sets. Training set has 80% of data and the test set has 20% of data. We encoded and scaled based on techniques employed in [15]- [16] to ensure fast processing, resulting in 20 input feature-set. We then built a feed-forward, fully connected Deep Neural Networks for our prediction problem. In a fully connected network, every node in each layer is connected to every node in the next layer. The number of input nodes is equal to the number of input features (equal to 20 in our case) and the number of output nodes is equal to the number of output classes (3 for our prediction problem). A neural network is a set of interconnected layers. The inputs are the first layer, and are connected to an output layer by an acyclic graph composed of weighted edges and nodes. The hidden layers are between the input and output layers. As this is a multiclass classification problem, the outputs are probabilities of the output class. Therefore, we used SoftMax [23] as the activation for the output layer. For all other layers we used 'tanh' activation [23]. For the optimizer, we chose Adam optimizer [24], a common gradient descent optimization algorithm. As the network makes predictions, it aims to reduce the validation loss, which is our objective. A gradient simply measures the change in all weights with regard to the change in error. In every iteration, the error is propagated backwards, from which the model tweaks the weights of different nodes to try and minimize the error further. Gradient graphs are convex and the step size along the gradient graph is called learning rate. Learning rate is another hyperparameter of the neural network. The loss function we used is categorical cross-entropy loss.

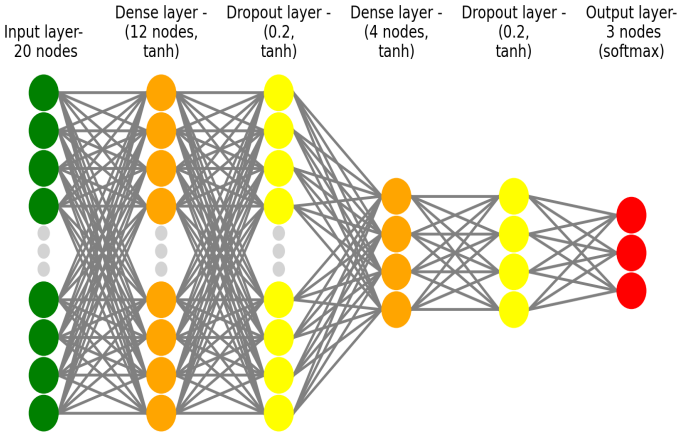

Fig. 3. Architecture of Baseline DNN

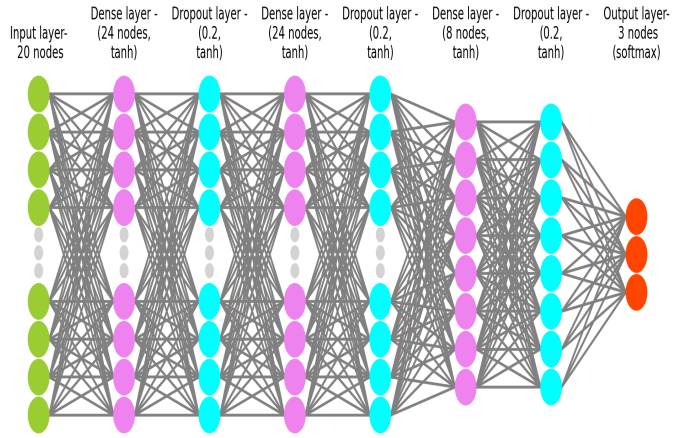

Fig. 4. Architecture of optimized DNN

We fixed the number of epochs to 100 and batch size to 32. We tried optimizing the number of hidden layers (between 1 and 5), the nodes in each layer(between 4 and 32) and the learning rates (in 0.01,0.001,0.0001). Optimizing a neural net architecture aids in finding an efficient model that could give the best performance on the data without overfitting. [26]

Our baseline DNN, shown in Fig. 3, has the following architecture- (12,4,3) nodes, 'tanh' activation in the hidden layer and a softmax activation in the output layer.

Our optimized DNN, shown in Fig. 4, has a 20 node input layer and a 3 node output layer and 3 hidden layers with (24,24,8) nodes, has a 'tanh' activation in the hidden layers and a softmax activation in the output layer.
The DNNs are trained in 100 epochs and use Mini batch gradient descent optimization for better generalization. [25]

TABLE V
SUMMARY OF THE EXPERIMENTAL SETUP FOR THE MODEL TRAINING

Model	Parameters
Baseline Deep Neural Network	Architecture - (12,4,3) nodes in 3 layers Activations - tanh(hidden layers) & softmax(output) Optimizer- Adam Epochs- 100 Batch-size: 64
Optimized DNN	Architecture - (20, 24, 24,8, 3) nodes in 5 layers Activations - relu (hidden layers), softmax (output) Optimizer- Adam Epochs- 100 Batch-size: 64 Dataset : Original

3) Performance evaluation: We evaluated the performance of the Deep Neural Network using the metrics- Accuracy, Precision, Recall, F1 score and visualized using Receiver Operating characteristic (ROC)-Area under the curve(AUC) plot. [**?**]

The formulae for the metrics are as below-

$$Accuracy = (TP + TN)/(FP + TP + FN + TN);$$

$$Precision = TP/(TP + FP);$$

$$Recall = TP/(TP + FN);$$

$$F1 - score = 2 * Precision * Recall/(Precision + Recall);$$

where FP- False positive, TP- True positive, FN- False negative and TN-True Negative

III. RESULTS

Below table summarizes the performance of the 6 models. SVM outperformed all the classic machine learning models, achieving Accuracy (0.59), Precision (0.55), Recall (0.59), F1 score (0.56) and ROC-AUC (0.64). While the baseline deep learning model significantly outperformed SVM achieving Accuracy (0.61), Precision (0.63), Recall (0.71), F1 score (0.67) and ROC-AUC (0.76), the optimised deep learning model achieved overall best performance with Accuracy (0.62), Precision (0.64), Recall (0.74), F1 score (0.69) and ROC-AUC (0.77).

TABLE VI
SUMMARY OF THE MODELS' PERFORMANCE

Model	Accuracy	Precision	Recall	F1 score	ROC-AUC
Naive Bayes	0.13	0.02	0.13	0.03	0.49
KNN	0.58	0.55	0.58	0.54	0.62
Decision Tree	0.53	0.50	0.53	0.50	0.59
SVM	0.59	0.55	0.59	0.56	0.64
Baseline DNN	0.61	0.63	0.71	0.67	0.76
Optimized DNN	0.62	0.64	0.74	0.69	0.77

A closer examination between baseline DNN and optimised show that, while both difference in performance appear to be very similar, the optimised DNN appear to perform much better on the Suspected Non-Alzheimer's pathology (SNAP) - class. Figure 5 and Figure 6 plots show the class wise ROC-AUC curves of baseline DNN and optimised DNN models respectively. Furthermore, to ensure there is no overfitting, Figure 7 shows the loss function curve showing no evidence of overfitting suggesting generalisability potential.

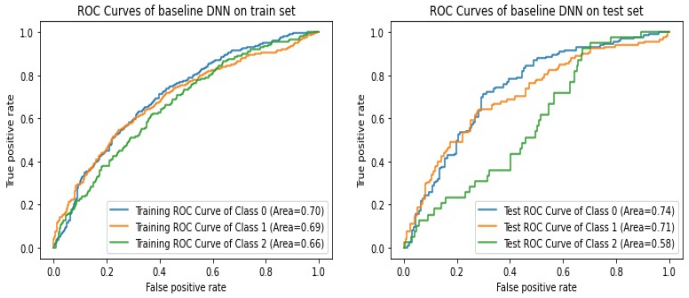

Fig. 5. AUC curves of baseline DNN on train vs. test set

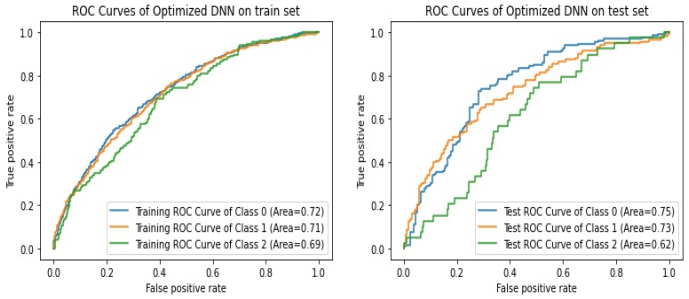

Fig. 6. AUC curves of Optimised DNN on train vs test set

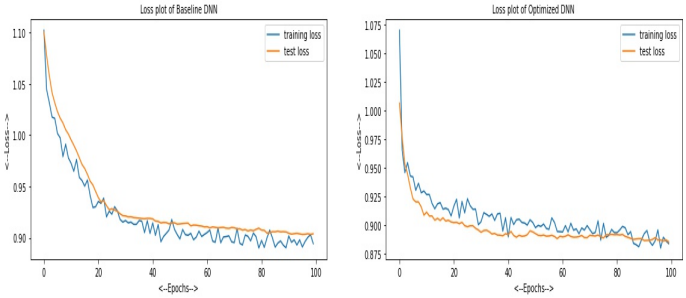

Fig. 7. Loss plot of baseline DNN vs optimized DNN

IV. Discussion

While existing machine-learning based risk prediction models have predominantly focused on various risk profiles such as the combination of genetic factors and family history [16]; neurodegeneration plasma biomarkers [29]; cardiovascular risk factors, aging, and incidence of dementia - CAIDE risk score) [32]; and Framingham risk scores [31], little has been done with regards to machine learning based risk prediction based on ATN framework until recently [32], which focused on classic machine learning methods. Our work extends this as we explored the performance of classic machine learning and deep learning based risk prediction models based on ATN framework. It is worth noting that ATN based risk prediction models has the additional benefit of indicating the stage of disease such as the state of neurodegeneration as well as identifying non amyloid driven pathology, which makes these models ideal for dementia screening in populations.

As observed in other works [9] [10], the deep learning-based models consistently outperformed the classic machine learning models (Naive Bayes, KNN, Decision Tree and SVM). This is because Deep neural networks with non-linear activations can better capture real-world complex interactions between various predictive factors. Further, our proposed deep learning model with an optimized architecture performs better than the baseline deep learning model. This is because complex models can capture complex interactions between data better. [33] Nevertheless, the optimized DNN still has scope for improvement.The optimized DNN had been able to decently predict 'healthy' and 'AD Continuum' groups but struggled to predict 'Suspected Non-Alzheimer's pathology (SNAP)' as we had very few samples of this particular group. Possible steps to overcome this problem would be to employ data augmentation techniques or assign class weights to the data that is fed to the DNN.

Investigating the impact of choosing an appropriate activation function on training the DNN models as a way of improving on detection performance. Also, with the promising performance of DNN, future work will explore the interaction of the underlying features responsible for the performance of the model. For example, we could examine the effect of sleep, depression, anxiety, age and BMI interactions, apoe e4, and gender and classes interactions and evaluate the performance of the models. Therefore, an explainable DNN would not only enhance our ability to carry out feature set analysis of the models, but would also improve on the utility of real world applications of the model.

Acknowledgments

We wish to acknowledge the team at Alzheimer's Disease Data Initiative for their wonderful support. The prompt response to issues by the team made a huge difference carrying out this work.

Author Contributions

Samuel O Danso conceived the idea and designed the experiments.
Sindhu Prattipati and Ibrahim Alqatawneh conducted the experiments.
Georgios Ntailianis was responsible for curating and cleaning the data
All authors contributed equally to the writing of the manuscript.

Funding

This work received no funding support, but was performed as part of the employment of the authors.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this article.

Data Availability

Data used for this work is available on ADDI plateform and available upon submission of research proposal which is reviewed by the EPAD consortium.

References

[1] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, J. Cohen-Mansfield, et al., "Dementia prevention, intervention, and care," The Lancet, vol. 390, no. 10113, pp. 2673–2734, Nov. 2017.

[2] Kalaria R., Maestre G., Arizaga R., Friedland R., Galasko D., Hall K., Luchsinger J., Ogunniyi A., Perry E., Potocnik F., others "Alzheimer's disease and vascular dementia in developing countries: prevalence, management, and risk factors," *The Lancet Neurology*, vol. 7, no. 9, pp. 812–826, 2008.

[3] Jack C., Bennett D., Blennow K., Carrillo M., Dunn B., Haeberlein S., Holtzman D., Jagust W., Jessen F., Karlawish J., Liu E. "NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease," **, vol. , no. , pp. , .

[4] Solomon A., Kivipelto M., Molinuevo J., Tom B., Ritchie C. "European prevention of Alzheimerâ€™s dementia longitudinal cohort study (EPAD LCS): study protocol," *BMJ open*, vol. 8, no. 12, pp. e021017, 2018.

[5] Calvin C., de Boer C., Raymont V., Gallacher J., Koychev I. "Prediction of Alzheimerâ€™s disease biomarker status defined by the â€˜ATN frameworkâ€™among cognitively healthy individuals: results from the EPAD longitudinal cohort study," *Alzheimer's Research & Therapy*, vol. 12, no. 1, pp. 1–16, 2020.

[6] Pellegrini E., Ballerini L., Hernandez M., Chappell F., González-Castro V., Anblagan D., Danso S., Muñoz-Maniega S., Job D., Pernet C., others "Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, no. , pp. 519–535, 2018.

[7] Byeon H. "Screening dementia and predicting high dementia risk groups using machine learning," *World Journal of Psychiatry*, vol. 12, no. 2, pp. 204, 2022.

[8] James C., Ranson J., Everson R., Llewellyn D. "Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients," *JAMA network open*, vol. 4, no. 12, pp. e2136553–e2136553, 2021.

[9] Akhter M., Jiangbin Z., Naqvi I., AbdelMajeed M., Zia T. "Abusive language detection from social media comments using conventional machine learning and deep learning approaches," *Multimedia Systems*, vol. , no. , pp. 1–16, 2021.

[10] Paterakis N., Mocanu E., Gibescu M., Stappers B., van Alst W. "Deep learning versus traditional machine learning methods for aggregated energy demand prediction," **, vol. , no. , pp. 1–6, 2017.

[11] Jo T., Nho K., Saykin A. "Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers in aging neuroscience*, vol. 11, no. , pp. 220, 2019.

[12] Shah D., Kapdi R., Patel J., Patel J. "Three Class Classification Of Alzheimer's Disease Using Deep NEURAL Networks.," *Current Medical Imaging*, vol. , no. , pp. , 2022.

[13] Hamdi M., Bourouis S., Rastislav K., Mohmed F., others "Evaluation of Neuro Images for the Diagnosis of Alzheimer's Disease Using Deep Learning Neural Network," *Frontiers in Public Health*, vol. 10, no. , pp. 35, 2022.

[14] Livingston G., Huntley J., Sommerlad A., Ames D., Ballard C., Banerjee S., Brayne C., Burns A., Cohen-Mansfield J., Cooper C., others "Dementia prevention, intervention, and care: 2020 report of the Lancet Commission," *The Lancet*, vol. 396, no. 10248, pp. 413–446, 2020.

[15] Hancock J., Khoshgoftaar T. "Survey on categorical data for neural networks," *Journal of Big Data*, vol. 7, no. 1, pp. 1–41, 2020.

[16] Wolters F., van der Lee S., Koudstaal P., van Duijn C., Hofman A., Ikram M., Vernooij M., Ikram M. "Parental family history of dementia in relation to subclinical brain disease and dementia risk," *Neurology*, vol. 88, no. 17, pp. 1642–1649, 2017.

[17] Weinzierl S., Zilker S., Brunk J., Revoredo K., Nguyen A., Matzner M., Becker J., Eskofier B. "An empirical comparison of deep-neural-network architectures for next activity prediction using context-enriched process event logs," *arXiv preprint arXiv:2005.01194*, vol. , no. , pp. , 2020.

[18] Staessen J., Li Y., Hara A., Asayama K., Dolan E., Oâ€™Brien E. "Blood pressure measurement anno 2016," *American journal of hypertension*, vol. 30, no. 5, pp. 453–463, 2017.

[19] Garcia-Ptacek S., Faxen-Irving G., Cermakova P., Eriksdotter M., Religa D. "Body mass index in dementia," *European journal of clinical nutrition*, vol. 68, no. 11, pp. 1204–1209, 2014.

[20] Qizilbash N., Gregson J., Johnson M., Pearce N., Douglas I., Wing K., Evans S., Pocock S. "BMI and risk of dementia in two million people over two decades: a retrospective cohort study," *The lancet Diabetes & endocrinology*, vol. 3, no. 6, pp. 431–436, 2015.

[21] Liu C., Kanekiyo T., Xu H., Bu G. "Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy," *Nature Reviews Neurology*, vol. 9, no. 2, pp. 106–118, 2013.

[22] Thara D., PremaSudha B., Xiong F. "Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques," *Pattern Recognition Letters*, vol. 128, no. , pp. 544–550, 2019.

[23] Sola J., Sevilla J. "Importance of input data normalization for the application of neural networks to complex industrial problems," *IEEE Transactions on nuclear science*, vol. 44, no. 3, pp. 1464–1468, 1997.

[24] Bock S., Weiß M. "A proof of local convergence for the Adam optimizer," **, vol. , no. , pp. 1–8, 2019.

[25] Qian X., Klabjan D. "The impact of the mini-batch size on the variance of gradients in stochastic gradient descent," *arXiv preprint arXiv:2004.13146*, vol. , no. , pp. , 2020.

[26] Hassan Ramchoun, Youssef Ghanou, Mohamed Ettaouil, Mohammed Amine Janati Idrissi "Multilayer Perceptron: Architecture Optimization and Training," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, pp. 26-30, 2016.

[27] Hossin M., Sulaiman M. "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, pp. 1, 2015.

[28] Danso S., Zeng Z., Muniz-Terrera G., Ritchie C. "Developing an Explainable Machine Learning-Based Personalised Dementia Risk Prediction Model: A Transfer Learning Approach With Ensemble Learning Algorithms," *Frontiers in Big Data*, vol. 4, no. , pp. , 2021.

[29] Shen X., Huang Y., Chen S., Guo Y., Tan L., Dong Q., Yu J., Weiner M., Aisen P., Petersen R., Jack C., Jagust W., Trojanowki J., Toga A., Beckett L., Green R., Saykin A., Morris J., Perrin R., Fernandez D. "Plasma phosphorylated-tau181 as a predictive biomarker for Alzheimerâ€™s amyloid, tau and FDG PET status," *Translational Psychiatry*, vol. 11, no. , pp. 585, Nov. 2021.

[30] Kivipelto M., Ngandu T., Laatikainen T., Winblad B., Soininen H., Tuomilehto J. "Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study," *The Lancet Neurology*, vol. 5, no. 9, pp. 735–741, 2006.

[31] Kaffashian S., Dugravot A., Nabi H., Batty G., Brunner E., Kivimäki M., others "Predictive utility of the Framingham general cardiovascular disease risk profile for cognitive function: evidence from the Whitehall II study," *European Heart Journal*, vol. 32, no. 19, pp. 2326–2332, 2011.

[32] Koychev I., Marinov E., Young S., Lazarova S., Grigorova D., Palejev D. "Identification of preclinical dementia according to ATN classification for stratified trial recruitment: A machine learning approach," *PLOS ONE*, vol. 18, no. 10, pp. e0288039, 2023.

[33] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, Fuad E. Alsaadi "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, no. , pp. 11-26, 2017.