Journal Pre-proof

A deep learning model for Alzheimer's disease diagnosis based on patient clinical records

J.L. Ávila-Jiménez, Vanesa Cantón-Habas, María del Pilar Carrera-González, Manuel Rich-Ruiz, Sebastián Ventura

PII: S0010-4825(23)01279-9
DOI: https://doi.org/10.1016/j.compbiomed.2023.107814
Reference: CBM 107814

To appear in: *Computers in Biology and Medicine*

Received date : 6 May 2023
Revised date : 19 November 2023
Accepted date : 3 December 2023

Please cite this article as: J.L. Ávila-Jiménez, V. Cantón-Habas, M. del Pilar Carrera-González et al., A deep learning model for Alzheimer's disease diagnosis based on patient clinical records, *Computers in Biology and Medicine* (2023), doi: https://doi.org/10.1016/j.compbiomed.2023.107814.

A deep learning model for Alzheimer's disease diagnosis based on patient clinical records

Ávila-Jiménez, J.L.
Cantón-Habas,Vanesa
Carrera-González, María del Pilar
Rich-Ruiz, Manuel
Ventura, Sebastián

November 18, 2023

Abstract

Background: Dementia, with Alzheimer's disease (AD) being the most common type of this neurodegenerative disease, is an under-diagnosed health problem in older people. The creation of classification models based on AD risk factors using Deep Learning is a promising tool to minimize the impact of under-diagnosis.

Objective: To develop a Deep Learning model that uses clinical data from patients with dementia to classify whether they have AD.

Methods: A Deep Learning model to identify AD in clinical records is proposed. In addition, several rebalancing methods have been used to preprocess the dataset and several studies have been carried out to tune up the model.

Results: Model has been tested against other well-established machine learning techniques, having better results than these in terms of AUC with alpha less than 0.05.

Conclusions: The developed Neural Network Model has a good performance and can be an accurate assisting tool for AD diagnosis.

Keywords: Dementia; Alzheimer's Disease; Deep learning; Machine learning; Prediction.

1 Introduction

Dementia, due to its clear link with the process of population ageing, currently represents a major problem, both in terms of prevalence and social repercussions, although the cause of this neurodegenerative pathology remains unknown [1, 2]. Specifically, Alzheimer's disease (AD) is the most prevalent type, representing $60 - 70\%$ of all elderly people with dementia [3, 4].

At the same time, AD under-diagnosis is a prominent reality, especially in this population group, affecting more than 60 % of the senior population,

distributed unevenly across low and high-income countries [5, 6, 7]. This underdiagnosis is not only due to a lack of knowledge of its aetiology, given that several modifiable and non-modifiable factors have been identified as contributing to its onset, but especially to the trivialisation of the first symptoms, which are often incorrectly attributed to the ageing process, leading to delays in seeking health care. In addition, the inherent challenges associated with distinguishing this neurodegenerative pathology from other diseases, such as depression, coupled with the limited training and reluctance of healthcare professionals, can significantly contribute to delays in the diagnostic process [8, 9].

Consequently, the implementation of early diagnosis strategies is essential. Given that a number of modifiable and non-modifiable factors have been identified as contributing to its onset, like diagnosis of depression, lifestyles, age, sex, etc, most of them collected in patient's clinical history, makes these records useful tool for predicting patients susceptible to develop dementia and, specifically, AD [10].

In this sense, the use of Machine Learning techniques, specifically, Deep Learning, presents itself as a promising tool, as they can process large amounts of data efficiently and learn complex patterns that may go unnoticed by healthcare professionals. In fact, Machine learning techniques have been previously used to detect this type of dementia, although mainly based on the recognition of neuroimaging scans. Specifically Support Vector Machines (SVM) have been used to classify medical brain images [11] and SVM with convolutional neural networks have been also used to classify brain images and predict AD in the general population [12].

Furthermore, previous studies have utilized various patient features to develop predictive models through machine learning techniques for early dementia detection [13]. Likewise, in other domains of medical science, Deep Learning has been employed to handle clinical data. For instance, in 2021, a model was developed to predict mortality in COVID-affected patients in[14]. It has also been used to study medical images [15].

In Alzheimer's disease (AD), Deep Learning manifests in two primary ways: models tailored for image recognition and those designed for the exploration of biomarkers [16]. While some studies have sought to predict dementia development in the general population [17], an exploration of the clinical records of patients using this methodology has been notably absent.

This study fills this gap by aiming to construct a Deep Learning model capable of processing clinical data from dementia patients, facilitating the accurate classification of the presence or absence of AD. Through this approach, we attempt to contribute to the early diagnosis and management of this pervasive neurodegenerative ailment.

The main contribution of this study is to propose a deep learning model for the diagnosis of AD that only uses clinical data from the patient as input.

2 Methods

2.1 Dataset description ad preprocess

The study subjects were elderly people, aged 65 years or older, with a medical diagnosis of dementia or AD with a score between 5 and 7 on the Global Deterioration Scale (GDS) and who had been in the dementia care process for at least 3 months.

The study collected a comprehensive set of variables from the patients' clinical records, encompassing sociodemographic information such as gender and age, details regarding the diagnosis and type of neurodegenerative disease, scores from the Global Deterioration Scale (GDS), extensive medical history including conditions like type 2 diabetes mellitus, dyslipidemia, depression, anxiety, and more, prescribed treatments, and the assessment of their level of independence in basic daily activities, as measured by the Barthel Index. Additionally, the diagnosis and classification of neurodegenerative diseases, as well as other medical conditions such as type 2 diabetes mellitus, dyslipidemia, depression, and anxiety, were carried out in accordance with the definitions provided by the International Classification of Diseases, 11th Revision (ICD-11).A detailed description of the variables can be found in Appendix I.

The dataset used in the current work has 100 patterns. Each pattern corresponds with the clinical data of one patient diagnosed with dementia and it is composed of 140 attributes corresponding to different items of their clinical report. The clinical report attributes serve as the input for the initial layer of the Neural Network, as illustrated in Figure 1.

Attributes of the datasets belong to several domains: there are 53 positive integer attributes, 87 Boolean attributes and one real attribute. Boolean attributes have been binarized, with 0 being equivalent to false and 1 to true.

A positive class indicates that the diagnosed dementia is due to AD and a negative one indicates another type of dementia (like vascular, degenerative, mixed dementia, primary degenerative mixed or Lewy dementia).

Positive and negative examples are unbalanced, exhibiting a 76% and 24% percentage in each class. This is common in clinical databases since they are typically focused on patients with one specific disease [18]. It can be considered intrinsic and relative imbalance and may amplify hinder learning of the classifier, according to [19]. Therefore, it is reasonable to preprocess the dataset by applying a rebalancing method to compensate for it.

There are many different rebalancing techniques [19], but we have focused on a few of the most used in classification problems: Random oversampling, NearMiss and SMOTE + TOMEK

Random oversampling generates copy patterns of the sub-represented class to rebalance the entire dataset [19]. In the same way, random undersampling could be used but implies many information losses. Instead of this, NearMiss has been used as an undersampling method. It selects examples based on the distance between majoritarian class examples to minoritarian class examples, trying to identify redundant examples for deletion [19].

SMOTE + TOMEK [20] is a more elaborated rebalance strategy that works in the boundaries between classes. The SMOTE technique interpolates new examples of the minoritarian class among a line that connects two of them. On the other hand, TOMEK is an undersampling method that keeps the examples of the majoritarian class with less Euclidean distance from the minoritarian class. To sum up, SMOTE+TOMEK generates new examples using SMOTE and removes them if the new examples affect the border between classes.

2.2 Neural Network Design

Deep learning is part of machine learning methods, and it is based on using deep neural networks (DNN) to generate models of feature learning [21]. Artificial Neural networks are made of several layers, composed of artificial neurons, that are connected to neurons of the next layer in the same way that synapses work in a human-like nervous system. There is one input layer that receives the input features, one or more hidden layers, and an output layer that generates the output of the model. A fully connected Neural Network has been used where every neuron of the n-th layer receives as input all the outputs of the (i-1)th layer.

In the present work, the neural network has been developed using a pyramidal approximation. One input layer, several hidden layers with fewer units than the previous one, and an output layer with one single output with binary output that indicates the predicted class [22][23]. The idea is that input attributes were combined and selected through the net to finally get a value that indicates the class. Figure 1 shows the network topology.

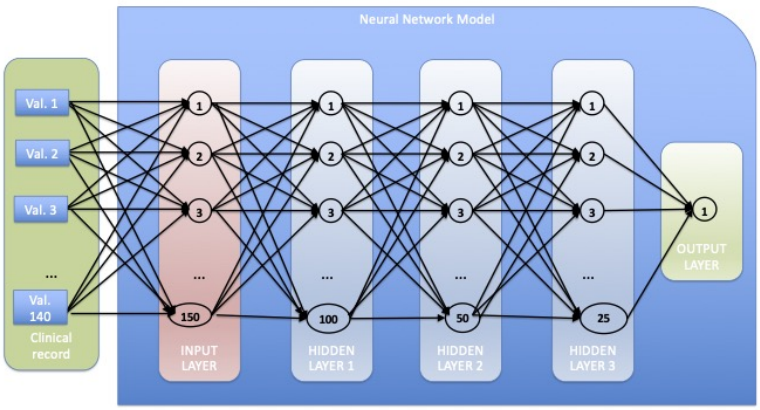

Figure 1: Neural network architecture

The activation function aggregates input and generates the output of a single neuron. Rectified Unit (RELU) has been used as the activation function due

to it mitigates vanishing gradient problems during the training phase [24]. In addition, the output layer is implemented by a single sigmoid that generates the binary output of the classifier.

Since the number of hidden layers and neurons is an open problem that has been dealt with using various strategies [23], we have carried out several tests in order to determine the optimal net topology to deal with our problem, as we describe in section 2.3.

Classification accuracy has been used as a cost function, combined with Adam [25] as an optimization algorithm, an evolution of the Stochastic Gradient Descend Algorithm. Both calculate partial derivatives in order to determine whether the weight contributes to the final error and modify it according to the expected and the real output. To determine the optimal epoch number, an experimental setup has been carried out, as it is shown in section 2.3. On top of that, the training set has not been divided into batches because it is not large enough to be worth it.

Considering that the training dataset is quite small, it is reasonable to use a regularization technique to prevent over-fitting. Dropout [26], based on switching of neurons during the training, has been used and some experiments have been carried on determining the best dropout rate.

2.3 Experimental Design

Metrics based on the confusion matrix accuracy, precision, recall and F-Score have been used to determine the performance of the proposed binary classification model. Considering that it is possible to increase precision by diminishing recall and vice versa, we have focused on the F-Score (harmonic mean between precision and recall) to determine the model performance.

In addition, the Area under ROC curve (AUC) metric has been used ROC curve shows the performance of a classification model plotting two parameters: true positive rate and false-positive rate. AUC shows the goodness of a model in the sense that a perfect separation between classes will have a 1 AUC, and the worst model has a 0 AUC value. Note that random classification models will have a 0.5 AUC value as average.

In every set of tests, 80% of examples have been used for training and 20% for tests. In addition, three executions have been carried out and results have been averaged, to reduce the differences due to the randomness of the training process.

Several experiments have been carried out to determine the best model parameters and to compare it with other Machine Learning proposals.

The dataset is imbalanced as we showed in the previous section, so the first experiments set had the goal of determining the best rebalancing methods to cope with it. Three methods have been tested: Random Oversampling, NearMiss and SMOTE+TOMEK.

Several experiments have been carried out to determine the best model and training parameters, tuning the number of neurons per layer, and testing several Dropout rates and the number of epochs to obtain the best training parameters.

After getting the best parameters, the proposed model has been tested against the K Nearest Neighbor and CART algorithm, a Decision Trees based approach that uses the Gini impurity index to train the model [27]. 8 executions of the three models have been carried out to perform a statistical analysis using the Wilcoxon test [28], [29].

The Wilcoxon signed-rank test, also known as the Wilcoxon rank-sum test, is a non-parametric statistical method used to assess the significance of differences between two related datasets. It does not rely on assumptions of normal data distribution. Instead, this test ranks the absolute differences between paired data points and determines whether the sum of ranks of positive differences significantly differs from the sum of ranks of negative differences. It is particularly valuable when dealing with small sample sizes or data that does not adhere to normal distribution assumptions. In our study, we have employed the Wilcoxon test to compare the performance of the proposed model with that of two other algorithms.

2.4 Ethical aspects

All principles contained in both the Declaration of Helsinki and the Belmont report regarding ethical precepts for biomedical research were respected. Therefore, the relatives or legal representatives of the candidates were informed before their inclusion in the study using a Patient Information Sheet (PIS), where the anonymity and confidentiality of the data were emphasized, and they signed the corresponding written informed consent. In this sense, the study has the authorization of all participating centres and the permission of the Andalusian Research Ethics Committee (Act no. 271, ref. 3672, approved on 5 December 2017).

2.5 Implementation

The experimental study has been executed on one desktop PC with Ubuntu 21.04 LTS, Intel Core i7-11700 Processor, 64 GB RAM and one GPU N-Vidia G-FORCE RTX-3080Ti with 64 Gb RAM DDR5. The experiments have been codified in Phyton 3.8. Deep Neural Network Models tested have been developed using Keras Framework version 2.6.0 [30] and TensorFlow version 2.6.2. [31] Rebalancing algorithms from the imbalanced-learn library version 0.90 [32] have been used. Decision Trees CART algorithm and KNN models have been used via SCK-LEARN Library version 0.24.2 [33]. Source code and datasets are available upon reasonable request.

3 Results and discussion

The first objective of the experimental phase has been to determine which data rebalancing technique is the most suitable for preprocessing the clinical data

used as input by the model. Table 1 shows the results after applying the above-mentioned three rebalancing techniques.

Considering the performance results, Smote TOMEK and random oversampling have a similar one in the whole evaluation metrics as shown in Table 1. Random Oversampling has better values in accuracy and AUC. This result can be explained because random Oversampling generates more patterns than Smote TOMEK. On the other hand, not using a rebalancing method strategy leads to poor performance in all metrics. Near Miss has a humble evaluation performance compared with the others.

Table 2 shows the number and percentages of examples of each class after rebalancing. Random oversampling does better rebalancing work than Smote TOMEK, in the sense that it generates more examples: Random Oversampling generates 152 training examples while Smote TOMEK generates 140. A good number of examples are necessary to correctly train a Neural Network model, so we can conclude that Random Oversampling is the appropriate rebalancing method for our problem.

		Acc.	**Prec.**	**Rec.**	**F-Score**	**AUC**
Not rebalanced	Train	0.810	0.673	0.690	0.698	0.862
	Test	0.583	0.503	0.486	0.463	0.458
Random Oversampling	Train	0.836	0.946	0.813	0.871	0.929
	Test	0.730	0.753	0.731	0.750	0.800
Near Miss	Train	0.893	0.850	0.671	0.777	0.862
	Test	0.667	0.703	0.676	0.650	0.684
Smote TOMEK	Train	0.897	0.860	0.906	0.883	0.897
	Test	0.727	0.800	0.786	0.786	0.865

Table 1: Experimental results for rebalancing methods

	n	**Positive**	**Negative**
Not rebalanced	100	88.89%	11,11%
Random Oversampling	152	66.67%	33,33%
Near Miss	48	66.67%	33,33%
Smote TOMEK	140	86.67%	13,33%

Table 2: Dataset composition after rebalancing

The identification of the optimal neural network architecture is of significant importance within our study . To achieve this, we conducted experiments with varying structures, altering the number of layers and neurons per layer. Results in terms of accuracy and AUC shown in table 3 are plotted in Figure 2.

The graphic shows that the maximum of both metrics is around 42.000 parameters, so we can conclude that the best topology of the tested ones, cor-

Neural Network topology					Trainable Param.	Train		Test	
Input Layer	1^{st} Hidden Layer	2^{nd} Hidden Layer	3^{rd} Hidden Layer	Output Layer		Acc.	AUC	Acc.	AUC
750	350	150	75	1	434151	0.897	0.926	0.667	0.741
500	250	100	50	1	226951	0.901	0.934	0.690	0.848
250	150	75	30	1	87036	0.906	0.935	0.750	0.860
150	100	50	25	1	42901	0.910	0.944	0.797	0.948
100	50	25	10	1	20896	0.883	0.925	0.720	0.824
50	25	10	-	1	8696	0.756	0.913	0.740	0.878
25	10	-	-	1	3846	0.816	0.902	0.740	0.800
25	-	-	-		3601	0.804	0.850	0.733	0.760

Table 3: Tested network architectures

responds with an input layer of 150 units, three hidden layers of 100, 50 and 25 units and the output unit. Figure 1 shows a graphical description of the neural network topology.

After establishing the preprocessing steps and defining the network topology, we proceeded to investigate two critical training parameters. Given the limitation of the training dataset, we explored the impact of including a dropout probability in all layers to mitigate overtraining and potentially enhance performance. Additionally, it is recognized the significance of optimizing the model without falling into overtraining pitfalls, making the number of training epochs a crucial parameter to consider.

Table 4 shows that AUC is hardly affected by dropout rates, but both Accuracy and F-Score have their best at about 0.2 dropout rate. In addition, Table 5 shows results compared with the number of epochs, which is graphic in Figure 3 where the performance in the tree metric smoothly decreased from a maximum of around 1000. Those results make us conclude that a dropout rate of 0.2 and 1000 epochs of training are the best training parameters for our model.

Dropout rate	Training			Test		
	Acc.	**F-Score**	**AUC**	**Acc.**	**F-Score**	**AUC**
0.1	0.846	0.775	0.923	0.777	0.753	0.853
0.2	0.901	0.870	0.934	0.797	0.793	0.902
0.3	0.796	0.701	0.915	0.727	0.703	0.847
0.4	0.830	0.786	0.806	0.700	0.683	0.845
0.5	0.793	0.739	0.873	0.723	0.703	0.819

Table 4: Performance and Dropout

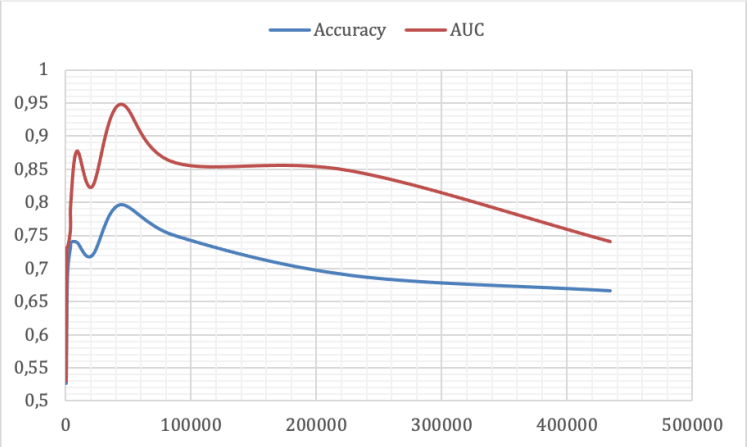

Figure 2: Accuracy and AUC against the number of parameters

Number of epochs	Training			Test		
	Acc.	**F-Score**	**AUC**	**Acc.**	**F-Score**	**AUC**
300	0.786	0.718	0.912	0.697	0.690	0.832
500	0.876	0.784	0.918	0.773	0.757	0.871
1000	0.911	0.871	0.936	0.773	0.763	0.900
3000	0.894	0.872	0.909	0.700	0.730	0.861
5000	0.912	0.908	0.939	0.637	0.630	0.760

Table 5: Performance and Number of epochs

The best parameters to run the model are shown in Table 6.

3.1 Comparison with prior work

The proposed model has been compared with the K-Nearest Neighbors (K-nn), a widely-used instance-based learning algorithm, and a Decision Tree model based on the CART algorithm [27], a common choice in tree-based classification methods frequently employed in machine learning. Comparative results are shown in Table 7.

The NN Model achieved an accuracy of 0.795, which is slightly lower than the CART model but higher than K-nn. This suggests that the NN Model is effective in correctly classifying patients with AD. The slight difference in accuracy between the NN Model and CART might be due to the complexity of the neural network, which can capture intricate patterns in the data. K-nn, on

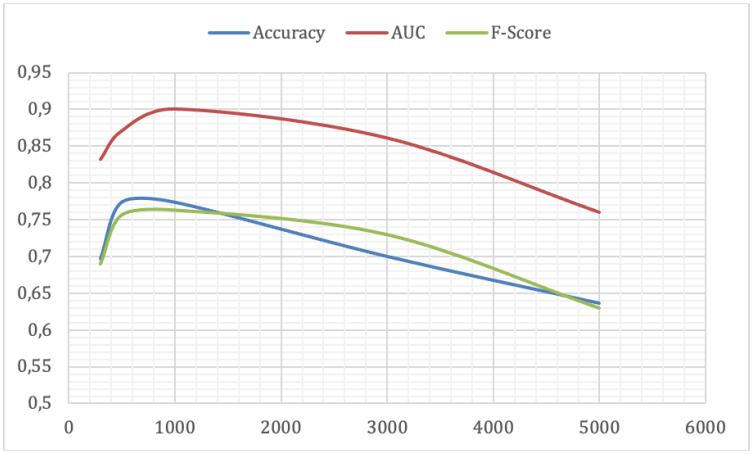

Figure 3: Performance metrics depend on the number of epochs

the other hand, might struggle with high-dimensional data, leading to a lower accuracy.

The F-Score measures the balance between precision and recall, providing a more comprehensive evaluation of the model's performance. The NN Model achieves an F-Score competitive with both CART and KNN. This indicates that the NN Model can effectively identify patients with Alzheimer's disease while maintaining a balance between false positives and false negatives.

AUC assesses the model's ability to distinguish between positive and negative cases. The NN Model outperforms both KNN and CART significantly in terms of AUC. This suggests that the NN Model has a higher discriminatory power and can better separate patients with AD from those without it. This superiority in AUC could be attributed to the deep learning capabilities of the neural network, enabling it to extract and learn intricate features from the clinical data.

It's worth noting that the NN Model exhibits a notably smaller standard deviation compared to KNN and CART across all metrics. This indicates that the NN Model's performance is more consistent and less variable, making it a robust choice for this classification task. The smaller standard deviation suggests that the NN Model is less sensitive to variations in the dataset or training process

The proposed model has been compared using a Wilcoxon test [28] with KNN and CART. Positive and negative ranks obtained in both comparisons are shown in table 8. The test concludes that the developed Neural Network Model is significantly better than KNN in every metric with an alpha less than 0.025. It is also better than CART for AUC with an alpha less than 0.025.

In addition, rank sums indicate that the proposed model has performed

Parameter	Value
Dataset parameter	
Rebalance method	Random oversampling
Neural Network parameter	
Neural Network Topology	5 layers (150, 100, 50,25,1 units)
Activation function	RELU and SIGMOID (output layer)
Training parameters	
Number of epochs	1000
Dropout	0.2
Optimization algorithm	ADAM
Cost Function	Accuracy

Table 6: Optimal parameters

Metric	K-nn	CART	NN Model
Accuracy			
Mean	0.666	0.802	0.795
stDev	0.0700	0.0565	0.0320
F-Score			
Mean	0.645	0.797	0.790
stDev	0.0709	0.05284	0.0312
AUC			
Mean	0.668	0.830	0.913
stDev	0.0688	0.0587	0.0320

Table 7: Comparative between proposed model, and others

better in tests than the CART model but there are no significant differences in Accuracy and f-Score.

3.2 Clinical Relevance

Early diagnosis of dementia in the initial stages of the disease is highly relevant given the current absence of an effective treatment [34]. In this context, the creation and subsequent implementation of Deep Learning models, as proposed in this article, can contribute to the identification of patterns indicative of the presence of the disease and, consequently, to the early detection of AD. Specifically, the generation of Deep Learning models using clinical data available in patients' medical records, not limited solely to those strictly related to cognitive function due to their demonstrated utility in previous studies [35], is posited as a promising tool for both diagnosis and subsequent management of AD.

In fact, the ability to use the patient's comprehensive medical history and prescribed pharmacological treatment, data available in medical records, for the

	KNN vs Neural Network			CART vs Neural Network		
	R+	R-	Result	R+	R-	Result
Accuracy	0	31	Reject null hypothesis $\alpha = 0.025$	9	12	Accept null hypothesis
F-Score	0	32	Reject null hypothesis $\alpha = 0.025$	15	19	Accept null hypothesis
AUC	0	33	Reject null hypothesis $\alpha = 0.001$	0	33	Reject null hypothesis $\alpha = 0.001$

Table 8: Wilcoxon test results, knn and CART algorithm compared with NN model

construction of these models should be understood as a strength, particularly because they contain a wealth of valuable and detailed information about a person's health over time, and because, in them, early symptoms and health changes of the patient are recorded, usually preceding more costly or invasive medical procedures. So much so that complementary tests inherent to the dementia diagnostic process, such as neuroimaging tests or plasma determinations, used in other studies where Deep Learning models have been created for AD detection [36, 37, 38], may not be available for all patients or may not be performed with the same frequency [39].

In this sense, the use of Deep Learning models based solely on the data from the clinical history can be used as tools to notify and, therefore, alert the healthcare professional of the need to request complementary tests, such as neuroimaging, for confirmation of the diagnosis. Thus, in accordance with scientific literature, Deep Learning models can also be useful in image recognition, even for predicting the progression of patients in the early stages [40].

In this way, all of this will result in an improvement in diagnostic accuracy since Deep Learning models can learn from complex patterns in longitudinal and diverse data, as is the case with the information contained in medical records, whose combination and joint analysis provide a complete picture of the patient's health. At the same time, their implementation in the real healthcare context can lead to a significant improvement in the quality of care [41].

In addition to their significance in the AD diagnostic process, Deep Learning models can be valuable in identifying previously unidentified risk factors for this neurodegenerative pathology [42]. In this way, they contribute to the understanding of the complex underlying mechanisms of this disease, the aetiology of which is still unknown. An example of this is the creation of models for iden-

tifying individuals with AD based on retinal scans, a test initially performed for the diagnosis or monitoring of eye conditions such as diabetic retinopathy or glaucoma [43].

Based on the above, Deep Learning models can serve as support tools in decision-making for healthcare professionals, as they provide additional information that complements their clinical expertise. Therefore, the collaboration between artificial intelligence and professional experience can lead to more accurate diagnoses and, consequently, more effective AD treatment [44].

On the other hand, in terms of cost-effectiveness, the use of Deep Learning models based on data available in medical records can contribute to controlling healthcare expenses [45]. This allows for the optimization of the patient selection process for complementary tests specific to dementia diagnosis, such as blood tests or magnetic resonance imaging (MRI) scans, thereby alleviating the financial burden on both patients and healthcare systems.

In conclusion, Deep Learning models harness the wealth of information contained within patients' medical histories to mitigate the inherent challenges in the AD diagnostic process. They support healthcare professionals and promote the rational use of resources, making them unquestionably essential tools in modern healthcare.

3.3 Future works

Further investigation could be held to apply Deep Learning models to other clinical situations, like pain in patients who are not able to communicate. In addition, it would be interesting to determine the best procedure to use rebalancing methods with clinical data.

4 Conclusions

In this study, we have developed an Artificial Neural Network tailored to analyze clinical data from patients diagnosed with dementia, with the primary goal of automating Alzheimer's disease (AD) diagnosis. Extensive testing was conducted to fine-tune model parameters, including network topology, neuron count, epochs, and dropout rate. Additionally, we have outlined a procedural framework for constructing similar networks, addressing challenges related to imbalanced data commonly encountered in clinical records.

Our results reveal that the proposed Neural Network Model demonstrates a competitive edge over K-Nearest Neighbors (KNN) and CART, exhibiting a significant advantage in terms of Area Under the Curve (AUC) and lower performance variability. The model's deep learning capabilities exhibit remarkable efficacy in capturing intricate clinical data patterns, underscoring its promise as a valuable tool for AD and dementia diagnosis.

In summary, our research represents a significant advancement in the early detection of dementia, particularly Alzheimer's disease, a critical healthcare concern. By leveraging Deep Learning models and comprehensive medical records,

we have demonstrated their potential to discern subtle disease patterns, facilitating early detection. These models not only serve as invaluable aids to healthcare professionals but also contribute to optimizing resource allocation, ultimately enhancing diagnostic accuracy and the quality of care. The application of Deep Learning in medical diagnostics holds transformative potential, offering a path to improved patient outcomes and a more efficient healthcare system.

Acknowledgements

This work has been supported by the Spanish Ministry of Science and Innovation, project PID2020-115832GB-I00. The project was also supported by the European Fund of Regional Development. It has been also supported by the Health Department of the Regional Government of Andalusia (PI-0357?2017).

Declaration of Competing Interest

All the authors declare that they have no competing interests.

References

[1] Emma Nichols, Jaimie D Steinmetz, Stein Emil Vollset, Kai Fukutaki, Julian Chalek, Foad Abd-Allah, Amir Abdoli, Ahmed Abualhasan, Eman Abu-Gharbieh, Tayyaba Tayyaba Akram, Hanadi Al Hamad, Fares Alahdab, Fahad Mashhour Alanezi, Vahid Alipour, Sami Almustanyir, Hubert Amu, Iman Ansari, Jalal Arabloo, Tahira Ashraf, Thomas Astell-Burt, Getinet Ayano, Jose L Ayuso-Mateos, Atif Amin Baig, Anthony Barnett, Amadou Barrow, Bernhard T Baune, Yannick Béjot, Woldesellassie M Mequanint Bezabhe, Yihienew Mequanint Bezabih, Akshaya Srikanth Bhagavathula, Sonu Bhaskar, Krittika Bhattacharyya, Ali Bijani, Atanu Biswas, Srinivasa Rao Bolla, Archith Boloor, Carol Brayne, Hermann Brenner, Katrin Burkart, Richard A Burns, Luis Alberto Cámera, Chao Cao, Felix Carvalho, Luis F S Castro de Araujo, Ferrán Catalá-López, Ester Cerin, Prachi P Chavan, Nicolas Cherbuin, Dinh-Toi Chu, Vera Marisa Costa, Rosa A S Couto, Omid Dadras, Xiaochen Dai, Lalit Dandona, Rakhi Dandona, Vanessa De la Cruz-Góngora, Deepak Dhamnetiya, Diana Dias da Silva, Daniel Diaz, Abdel Douiri, David Edvardsson, Michael Ekholuenetale, Iman El Sayed, Shaimaa I El-Jaafary, Khalil Eskandari, Sharareh Eskandarieh, Saman Esmaeilnejad, Jawad Fares, Andre Faro, Umar Farooque, Valery L Feigin, Xiaoqi Feng, Seyed-Mohammad Fereshtehnejad, Eduarda Fernandes, Pietro Ferrara, Irina Filip, Howard Fillit, Florian Fischer, Shilpa Gaidhane, Lucia Galluzzo, Ahmad Ghashghaee, Nermin Ghith, Alessandro Gialluisi, Syed Amir Gilani, Ionela-Roxana Glavan, Elena V Gnedovskaya, Mahaveer Golechha, Rajeev Gupta, Veer Bala Gupta, Vivek Kumar Gupta, Mohammad Rifat Haider, Brian J Hall, Samer Hamidi, Asif

Hanif, Graeme J Hankey, Shafiul Haque, Risky Kusuma Hartono, Ahmed I Hasaballah, M Tasdik Hasan, Amr Hassan, Simon I Hay, Khezar Hayat, Mohamed I Hegazy, Golnaz Heidari, Reza Heidari-Soureshjani, Claudiu Herteliu, Mowafa Househ, Rabia Hussain, Bing-Fang Hwang, Licia Iacoviello, Ivo Iavicoli, Olayinka Stephen Ilesanmi, Irena M Ilic, Milena D Ilic, Seyed Sina Naghibi Irvani, Hiroyasu Iso, Masao Iwagami, Roxana Jabbarinejad, Louis Jacob, Vardhmaan Jain, Sathish Kumar Jayapal, Ranil Jayawardena, Ravi Prakash Jha, Jost B Jonas, Nitin Joseph, Rizwan Kalani, Amit Kandel, Himal Kandel, André Karch, Ayele Semachew Kasa, Gizat M Kassie, Pedram Keshavarz, Moien AB Khan, Mahalaqua Nazli Khatib, Tawfik Ahmed Muthafer Khoja, Jagdish Khubchandani, Min Seo Kim, Yun Jin Kim, Adnan Kisa, Sezer Kisa, Mika Kivimäki, Walter J Koroshetz, Ai Koyanagi, G Anil Kumar, Manasi Kumar, Hassan Mehmood Lak, Matilde Leonardi, Bingyu Li, Stephen S Lim, Xuefeng Liu, Yuewei Liu, Giancarlo Logroscino, Stefan Lorkowski, Giancarlo Lucchetti, Ricardo Lutzky Saute, Francesca Giulia Magnani, Ahmad Azam Malik, João Massano, Man Mohan Mehndiratta, Ritesh G Menezes, Atte Meretoja, Bahram Mohajer, Norlinah Mohamed Ibrahim, Yousef Mohammad, Arif Mohammed, Ali H Mokdad, Stefania Mondello, Mohammad Ali Ali Moni, Md Moniruzzaman, Tilahun Belete Mossie, Gabriele Nagel, Muhammad Naveed, Vinod C Nayak, Sandhya Neupane Kandel, Trang Huyen Nguyen, Bogdan Oancea, Nikita Otstavnov, Stanislav S Otstavnov, Mayowa O Owolabi, Songhomitra Panda-Jonas, Fatemeh Pashazadeh Kan, Maja Pasovic, Urvish K Patel, Mona Pathak, Mario F P Peres, Arokiasamy Perianayagam, Carrie B Peterson, Michael R Phillips, Marina Pinheiro, Michael A Piradov, Constance Dimity Pond, Michele H Potashman, Faheem Hyder Pottoo, Sergio I Prada, Amir Radfar, Alberto Raggi, Fakher Rahim, Mosiur Rahman, Pradhum Ram, Priyanga Ranasinghe, David Laith Rawaf, Salman Rawaf, Nima Rezaei, Aziz Rezapour, Stephen R Robinson, Michele Romoli, Gholamreza Roshandel, Ramesh Sahathevan, Amirhossein Sahebkar, Mohammad Ali Sahraian, Brijesh Sathian, Davide Sattin, Monika Sawhney, Mete Saylan, Silvia Schiavolin, Allen Seylani, Feng Sha, Masood Ali Shaikh, KS Shaji, Mohammed Shannawaz, Jeevan K Shetty, Mika Shigematsu, Jae Il Shin, Rahman Shiri, Diego Augusto Santos Silva, João Pedro Silva, Renata Silva, Jasvinder A Singh, Valentin Yurievich Skryabin, Anna Aleksandrovna Skryabina, Amanda E Smith, Sergey Soshnikov, Emma Elizabeth Spurlock, Dan J Stein, Jing Sun, Rafael Tabarés-Seisdedos, Bhaskar Thakur, Binod Timalsina, Marcos Roberto Tovani-Palone, Bach Xuan Tran, Gebiyaw Wudie Tsegaye, Sahel Valadan Tahbaz, Pascual R Valdez, Narayanaswamy Venketasubramanian, Vasily Vlassov, Giang Thu Vu, Linh Gia Vu, Yuan-Pang Wang, Anders Wimo, Andrea Sylvia Winkler, Lalit Yadav, Seyed Hossein Yahyazadeh Jabbari, Kazumasa Yamagishi, Lin Yang, Yuichiro Yano, Naohiro Yonemoto, Chuanhua Yu, Ismaeel Yunusa, Siddhesh Zadey, Mikhail Sergeevich Zastrozhin, Anasthasia Zastrozhina, Zhi-Jiang Zhang, Christopher J L Murray, and Theo Vos. Estimation of

the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. *The Lancet Public Health*, 7(2):e105–e125, 2022.

[2] Frank J. Wolters, Lori B. Chibnik, Reem Waziry, Roy Anderson, Claudine Berr, Alexa Beiser, Joshua C. Bis, Deborah Blacker, Daniel Bos, Carol Brayne, Jean-François Dartigues, Sirwan K.L. Darweesh, Kendra L. Davis-Plourde, Frank de Wolf, Stephanie Debette, Carole Dufouil, Myriam Fornage, Jaap Goudsmit, Leslie Grasset, Vilmundur Gudnason, Christoforos Hadjichrysanthou, Catherine Helmer, M. Arfan Ikram, M. Kamran Ikram, Erik Joas, Silke Kern, Lewis H. Kuller, Lenore Launer, Oscar L. Lopez, Fiona E. Matthews, Kevin McRae-McKee, Osorio Meirelles, Thomas H. Mosley, Matthew P. Pase, Bruce M. Psaty, Claudia L. Satizabal, Sudha Seshadri, Ingmar Skoog, Blossom C.M. Stephan, Hanna Wetterberg, Mei Mei Wong, Anna Zettergren, and Albert Hofman. Twenty-seven-year time trends in dementia incidence in europe and the united states. *Neurology*, 95(5):e519–e531, 2020.

[3] Marcos Vinícius Ferreira Silva, Cristina de Mello Gomide Loures, Luan Carlos Vieira Alves, Leonardo Cruz de Souza, Karina Braga Gomes Borges, and Maria das Graças Carvalho. Alzheimer's disease: risk factors and potentially protective measures. *Journal of Biomedical Science*, 26(1):33, 2019.

[4] Daniel Ferreira, Lilisbeth Perestelo-Perez, Eric Westman, Lars-olof Wahlund, Antonio Sarria, and Pedro Serrano-Aguilar. Meta-review of csf core biomarkers in alzheimer's disease: The state-of-the-art after the new revised diagnostic criteria. *Frontiers in Aging Neuroscience*, 6, 2014.

[5] Linda Lang, Angela Clifford, Li Wei, Dongmei Zhang, Daryl Leung, Glenda Augustine, Isaac M Danat, Weiju Zhou, John R Copeland, Kaarin J Anstey, and Ruoling Chen. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. *BMJ Open*, 7(2):e011146–e011146, 2017.

[6] Tilly Eichler, Jochen René Thyrian, Johannes Hertel, Bernhard Michalowsky, Diana Wucherer, Adina Dreier, Ingo Kilimann, Stefan Teipel, and Wolfgang Hoffmann. Rates of formal diagnosis of dementia in primary care: The effect of screening. *Alzheimers Dement (Amst)*, 1(1):87–93, Mar 2015.

[7] George M Savva and Antony Arthur. Who has undiagnosed dementia? a cross-sectional analysis of participants of the aging, demographics and memory study. *Age Ageing*, 44(4):642–647, Jul 2015.

[8] Jill Phillips, Constance Dimity Pond, Nerida Elizabeth Paterson, Cate Howell, Allan Shell, Nigel P Stocks, Susan M Goode, and John E Marley. Difficulties in disclosing the diagnosis of dementia: a qualitative study in general practice. *Br J Gen Pract*, 62(601):e546–53, Aug 2012.

[9] Jennifer Yates, Miriam Stanyon, Rajvinder Samra, and Linda Clare. Challenges in disclosing and receiving a diagnosis of dementia: a systematic review of practice from the perspectives of people with dementia, carers, and healthcare professionals. *Int Psychogeriatr*, 33(11):1161–1192, Nov 2021.

[10] Gill Livingston, Jonathan Huntley, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, Sergi G Costafreda, Amit Dias, Nick Fox, Laura N Gitlin, Robert Howard, Helen C Kales, Mika Kivimäki, Eric B Larson, Adesola Ogunniyi, Vasiliki Orgeta, Karen Ritchie, Kenneth Rockwood, Elizabeth L Sampson, Quincy Samus, Lon S Schneider, Geir Selbæk, Linda Teri, and Naaheed Mukadam. Dementia prevention, intervention, and care: 2020 report of the lancet commission. *Lancet*, 396(10248):413–446, Aug 2020.

[11] Sven Haller, Duy Nguyen, Cristelle Rodriguez, Joan Emch, Gabriel Gold, Andreas Bartsch, Karl O Lovblad, and Panteleimon Giannakopoulos. Individual prediction of cognitive decline in mild cognitive impairment using support vector machine-based analysis of diffusion tensor imaging data. *J Alzheimers Dis*, 22(1):315–327, 2010.

[12] Ting Shen, Jiehui Jiang, Yupeng Li, Ping Wu, Chuantao Zuo, and Zhuangzhi Yan. Decision supporting model for one-year conversion probability from mci to ad using cnn and svm. *Annu Int Conf IEEE Eng Med Biol Soc*, 2018:738–741, Jul 2018.

[13] Mingyue Hu, Xinhui Shu, Gang Yu, Xinyin Wu, Maritta Välimäki, and Hui Feng. A risk prediction model based on machine learning for cognitive impairment among chinese community-dwelling elderly people with normal cognition: Development and validation study. *J Med Internet Res*, 23(2):e20298, Feb 2021.

[14] Saranya Sankaranarayanan, Jagadheshwar Balan, Jesse R Walsh, Yanhong Wu, Sara Minnich, Amy Piazza, Collin Osborne, Gavin R Oliver, Jessica Lesko, Kathy L Bates, Kia Khezeli, Darci R Block, Margaret DiGuardo, Justin Kreuter, John C O'Horo, John Kalantari, Eric W Klee, Mohamed E Salama, Benjamin Kipp, William G Morice, and Garrett Jenkinson. Covid-19 mortality prediction from deep learning in a large multistate electronic health record and laboratory information system data set: Algorithm development and validation. *J Med Internet Res*, 23(9):e30157, Sep 2021.

[15] Tausifa Jan Saleem, Syed Rameem Zahra, Fan Wu, Ahmed Alwakeel, Mohammed Alwakeel, Fathe Jeribi, and Mohammad Hijji. Deep learning-based diagnosis of alzheimer's disease. *J Pers Med*, 12(5), May 2022.

[16] Suhad Al-Shoukry, Taha H. Rassem, and Nasrin M. Makbol. Alzheimer's diseases detection by using deep learning algorithms: A mini-review. *IEEE Access*, 8:77131–77141, 2020.

[17] Woo Jung Kim, Ji Min Sung, David Sung, Myeong-Hun Chae, Suk Kyoon An, Kee Namkoong, Eun Lee, and Hyuk-Jae Chang. Cox proportional hazard regression versus a deep learning algorithm in the prediction of dementia: An analysis based on periodic health examination. *JMIR Med Inform*, 7(3):e13139, Aug 2019.

[18] Sara Belarouci and Mohammed Amine Chikh. Medical imbalanced data classification. *Advances in Science, Technology and Engineering Systems Journal*, 2(3):116–124, 2017.

[19] Haibo He and E.A. Garcia. Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on*, 21:1263 – 1284, 10 2009.

[20] Gustavo E. A. P. A. Batista, Ronaldo C. Prati, and Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.*, 6(1):20–29, jun 2004.

[21] Yann LeCun, Y. Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.

[22] Quynh Nguyen and Marco Mondelli. Global convergence of deep networks with one wide layer followed by pyramidal topology. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[23] Dongyoon Han, Jiwhan Kim, and Junmo Kim. Deep pyramidal residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5927–5935, 2017.

[24] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[25] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *ICLR (Poster)*, 2015.

[26] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[27] Dan H. Moore II. Classification and regression trees, by leo breiman, jerome h. friedman, richard a. olshen, and charles j. stone. brooks/cole publishing, monterey, 1984,358 pages, $27.95. *Cytometry*, 8(5):534–535, 1987.

[28] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[29] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1):1–30, 2006.

[30] François Chollet. keras. https://github.com/fchollet/keras, 2015.

[31] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, pages 265–283, USA, 2016. USENIX Association.

[32] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.

[33] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[34] Jill Rasmussen and Haya Langerman. Alzheimer's disease – why we need early diagnosis. *Degenerative Neurological and Neuromuscular Disease*, 9:123 – 130, 2019.

[35] Muhammad Shahbaz, Shahzad Ali, Aziz Guergachi, Aneeta Niazi, and Amina Umer. Classification of alzheimer's disease using machine learning techniques. In *Proceedings of the 8th International Conference on Data Science, Technology and Applications*. SCITEPRESS - Science and Technology Publications, 2019.

[36] Badiea Abdulkarem Mohammed, Ebrahim Mohammed Senan, Taha H. Rassem, Nasrin M. Makbol, Adwan Alownie Alanazi, Zeyad Ghaleb Al-Mekhlafi, Tariq S. Almurayziq, and Fuad A. Ghaleb. Multi-method analysis of medical records and mri images for early diagnosis of dementia and alzheimer's disease based on deep learning and hybrid methods. *Electronics*, 10(22), 2021.

[37] Manhua Liu, Fan Li, Hao Yan, Kundong Wang, Yixin Ma, Li Shen, and Mingqing Xu. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer's disease. *NeuroImage*, 208:116459, 2020.

[38] Gopi Battineni, Nalini Chintalapudi, Mohammad Amran Hossain, Giuseppe Losco, Ciro Ruocco, Getu Gamo Sagaro, Enea Traini, Giulio Nittari, and Francesco Amenta. Artificial intelligence models in the diagnosis of adult-onset dementia disorders: A review. *Bioengineering*, 9(8), 2022.

[39] Dipanjan Banerjee, Abilash Muralidharan, Abdul Rub Hakim Mohammed, and Bilal Malik. Neuroimaging in dementia: A brief review. *Cureus*, 12, 06 2020.

[40] Sheng Liu, Arjun V. Masurkar, Henry Rusinek, Jingyun Chen, Ben Zhang, Weicheng Zhu, Carlos Fernandez-Granda, and Narges Razavian. Generalizable deep learning model for early alzheimer's disease detection from structural mris. *Scientific Reports*, 12(1):17106, 2022.

[41] Kenji Karako, Peipei Song, and Yu Chen. Recent deep learning models for dementia as point-of-care testing: Potential for early detection. *Intractable and Rare Diseases Research*, 12(1):1–4, 2023.

[42] Francisco Balea-Fernandez, Beatriz Martinez-Vega, Samuel Ortega, Himar Fabelo, Raquel Leon, Gustavo Marrero Callico, and Cristina Bibao-Sieyro. Analysis of risk factors in dementia through machine learning. *Journal of Alzheimer's Disease*, 79:1–17, 12 2020.

[43] Carol Y Cheung, An Ran Ran, Shujun Wang, Victor T T Chan, Kaiser Sham, Saima Hilal, Narayanaswamy Venketasubramanian, Ching-Yu Cheng, Charumathi Sabanayagam, Yih Chung Tham, Leopold Schmetterer, Gareth J McKay, Michael A Williams, Adrian Wong, Lisa W C Au, Zhihui Lu, Jason C Yam, Clement C Tham, John J Chen, Oana M Dumitrascu, Pheng-Ann Heng, Timothy C Y Kwok, Vincent C T Mok, Dan Milea, Christopher Li-Hsian Chen, and Tien Yin Wong. A deep learning model for detection of alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. *The Lancet Digital Health*, 4(11):e806–e815, 2022.

[44] Stephen D Auger, Benjamin M Jacobs, Ruth Dobson, Charles R Marshall, and Alastair J Noyce. Big data, machine learning and artificial intelligence: a neurologist's guide. *Practical Neurology*, 21(1):4–11, 2021.

[45] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J. Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Yi Zhang, Gerardo Flores, Gavin E. Duggan, Jamie Irvine, Quoc Le, Kurt Litsch, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, Dana Ludwig, Samuel L. Volchenboum, Katherine Chou, Michael Pearson, Srinivasan Madabushi, Nigam H. Shah, Atul J. Butte, Michael D. Howell, Claire Cui, Greg S. Corrado, and Jeffrey Dean. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1(1):18, 2018.

Appendix I. Description of variables

Sociodemographic data	
Gender Age Marital status Usual place of residence (rural/urban)	Clinical history
Health variables	
Diagnosis of dementia	
All prescribed pharmacological treatment	Clinical history
Medical history: type 2 diabetes mellitus, high blood pressure, dyslipidemia, anxiety, depression and in case of depression time from diagnosis, osteoporosis, osteoarthritis, atrial fibrillation, pacemakers, chronic anaemia, epilepsy, Parkinson's disease and glaucoma.	
Cognitive impairment	Global Deterioration Scale
Autonomy in activities of daily living	Barthel Index

Table 9: Description of variables

Highlights

- A Deep Learning model to classify Alzheimer patients from clinical data has been developed.
- Model has, better results than other well-established machine learning techniques.
- Several rebalancing methods have been used to preprocess clinical datasets
- The developed Neural Network Model can be an accurate assisting tool for Alzheimer Disease diagnosis

Declaration of Competing Interest

All the authors declare that they have no competing interests.