(IJISE) 2022, Vol. No. 15, Jan-Jun **e-ISSN: 2454-6402, p-ISSN: 2454-812X**

EMPLOYABILITY OF MACHINE LEARNING TOOLS AND TECHNIQUES IS EFFICACIOUS IN EARLY DETECTION AND DIAGNOSIS OF ALZHEIMER'S DISEASE

Updesh Sachdeva

Mount Olympus School, Gurugram, Haryana, India

ABSTRACT

Memory and other cognitive parts are significantly affected by Alzheimer's disease (AD). As there is no fix, early disease movement location and deferral are basic for overseeing Promotion. Verbal familiarity is one of the most widely recognized and delicate neuropsychological techniques utilized for identifying and assessing mental decreases in Promotion, in which a subject should name many things as conceivable in 30 or 60 seconds that have a place within a certain classification. A verb fluency (VF) task, a specific subset of verbal fluency that examines the subjects' list of verbs during a given period, is used in this study to develop a method for detecting AD. To identify the possibility of AD, we make use of natural language processing (NLP), random forest (RF), neural network (NN), and recurrent neural network (RNN) machine learning strategies. However, preprocessing the data is required for the developed models to stratify subjects into the appropriate AD and control groups with up to 76% accuracy using RF. This exactness is marginally lower, however not fundamentally, at 67% utilizing RNN and NLP, which includes practically no manual information preprocessing. Utilizing specific VF tasks for early AD detection is now possible, thanks to this study's results.

INTRODUCTION

Alzheimer's disease (AD) is the leading cause of dementia, accounting for 60-80 per cent of cases [1]. Dementia generally refers to a patient's decline in memory and cognitive skills, such as their ability to reason, think, or speak clearly. AD is a degenerative brain disease that originates from damage to brain cells. While no cure for AD exists, earlier disease detection means earlier intervention and more effective care. Despite the growing number of cases of Only a quarter of AD patients are typically diagnosed [2]. Worse yet, the mortality rate of AD in the United States has significantly increased between 2000 and 2018 from 17.6 to 37.3 deaths per 100,000 population [3]. A large research body is dedicated to studying the utilization of language tasks in improving the early detection of AD [4]. This research has shown promise, as AD results in cognitive impairment and typically negatively impacts how patients produce or use language. Past studies have generally covered recording a patient's speech over time and analysing the number and types of words they produce to detect AD [5]. Such an approach to AD detection is promising because there is generally no need for expensive equipment or invasive procedures, and the data collection and analysis can be done remotely. However, existing works on detecting AD from recorded speech data generally use time-intensive tasks, such as open-ended interviews with clinicians [6]. Although verbal fluency, e.g., semantic and phonemic, has been commonly used to detect AD, analyzing the listing of verbs is much less explored. This task of listing verbs has the potential to simplify the evaluation process and can be more readily transferable and generalizable to a large array of languages. Therefore, in this study, we aim to leverage machine learning (ML), natural language processing (NLP), and VF for early AD detection. ML is a branch of artificial intelligence that allows for eliciting patterns from the data. It can draw associations between a set of input variables (e.g., the choice of verbs, the pattern with which they are produced, etc.) and output (response) variables (e.g., at risk for AD or not). NLP is a field at the intersection of artificial intelligence and linguistics concerning interactions between human (natural) language and computers. Both ML and NLP, separately or jointly, have been used in healthcare applications to much success, e.g., in detecting or predict various outcomes or risks for patients using electronic medical records [8]. Using data from a 30-second VF task, we develop a method to detect AD in this study. First, we develop ML models that detect AD using psycholinguistic features of the input verbs extracted by experts from the VF task data. We specifically develop random forest (RF) and neural networks (NN) models. Next, we leverage NLP and ML jointly to develop an end-to-end ML pipeline. That is, we use NLP on the concatenated text string of verbs from subjects to elicit information. We then use this elicited information and the (raw) sequence of verbs produced in a recurrent neural network (RNN) model to detect AD.

METHODOLOGY

A. Data Collection and Preprocessing

The subject cohort includes 20 AD patients (mean age = 77.85 years) and 25 age-matched controls (mean age = 72.68 years). Each subject is asked to say as many verbs as possible in a 30-second block. The responses are recorded verbatim. The study protocol was reviewed and approved by The University of Tennessee Health Science Center.

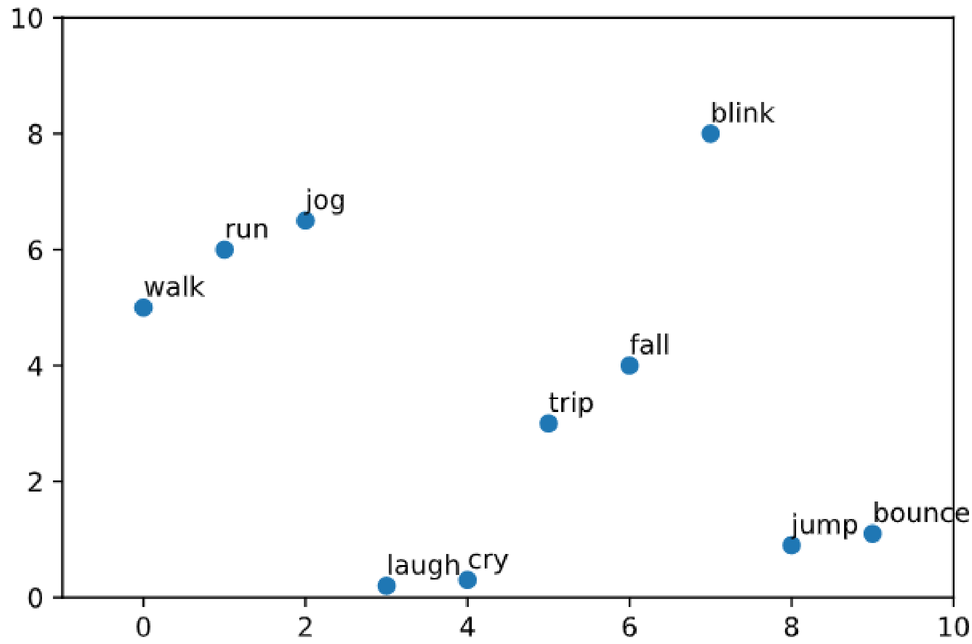

Fig. 1. Example 2D word embedding space, where similar words are closer together

Subject matter experts analyze the data to extract psycholinguistic properties. The analysis is performed to elicit properties about VF responses of individuals with amnestic AD and cognitively healthy older adults.

Specifically, The English Lexicon project, a multi-university the effort to provide a standardized behavioural and descriptive data set for 40,481 words and 40,481 non-words [13] is used for the psycholinguistic analysis. To extract psycholinguistic properties, the root forms of the verbs are used. The properties extracted include [7]: Total number of correctly produced words, length of the word, the number of phonological neighbours that a word has, the number of orthographic neighbours that a word has, how pleasant the word is, the extent to which the word denotes something weak or strong, number of phonemes in the pronunciation, word frequency, and the age of acquisition of the word.

B. Models

Two types of ML models were developed in this study.

The first type relied on features extracted from the psycholinguistic properties. Specifically, we calculated the average, standard deviation, and range of each psycholinguistic property reported for any given subject. This resulted in 60 initial features. We then used these features to develop two ML models, namely RF and NN. RF is an ensemble classifier that uses many decision trees, each fitted on a randomly selected subset of the data, for classification [14]. RF is generally highly robust against overfitting. For the RF model, based on preliminary results using out-of-bag (OOB) error, 100 trees were included in the model. Moreover, we used NNs, another non-direct learning model, for characterization. NNs output the results after transferring data from an input layer to a hidden layer [15]. For the NN model, one hidden layer with 16 hidden nodes were used. The activation function was set to the rectified linear unit (ReLU). Also, the learning rate was set to 0.001. Adam optimizer was used for model training. Lastly, features were normalized before feeding them into the model. The second type of ML model did not rely on features extracted from the psycholinguistic properties. Specifically, we developed an RNN directly using the recorded verbatim. This involved using the concatenated string of verbal responses for any given subject, plus the corresponding word embeddings obtained from NLP. In particular, we used word embeddings to convert the words into vectors, allowing the RNN to form a relationship between different verbs produced by subjects.

Therefore, words that are closer in meaning (or are related in some way) have more similar vector representations. The RNN includes one hidden layer with 50 hidden nodes. The activation function was set to 'sigmoid'. Adam optimizer was used for model training. The learning rate was again set to 0.001. All models are developed in Python. For ML models, we use Keras [16] with the TensorFlow backend [17]. In addition, we use pre-trained, 300-dimensional word embeddings from the spaCy package are trained on a corpus of web page data [18].

C. Feature Selection for RF and NN

RF allows for ranking feature importance per the total decrease in the Gini measure of node impurities [19]. We use this feature ranking to perform feature pruning. Specifically, the 60 initial features are first ranked based on their importance using RF. The 15 most important features are selected for both RF and NN models.

D. Input Data Tuning for the RNN and NLP Model

Recall that concatenated string of verbal responses for any given subject was used in the RNN and NLP model. To improve model performance, various text string combinations were explored. This included using concatenated strings with and without stumbling (such as "um" and "uh") and with and without repeated verbs if they occurred.

E. Model Evaluation and Metrics

For all models, we employ five-fold cross-validation. In each fold, training is done using balanced sets, i.e., equal numbers of AD patients and healthy controls. We consequently provide mean and standard deviations across the five folds. This helped to avoid favouring the more representative group. Evaluation metrics include accuracy, F1 score, and area under the receiver operating characteristic curve (AUC). AUC is the value that reflects the overall ranking performance of a classifier [20]. Accuracy is the ratio of correct predictions over total predictions.

RESULTS

Fig. 2 displays the importance of the top 15 features included in the RF and NN models. Table I presents the descriptions of these features. As seen in the figure and table, the features are generally drawn from psycholinguistic properties relating to the age of acquisition, number of phonemes, how pleasant the word is, and phonological neighbours a word has, among others. [13].

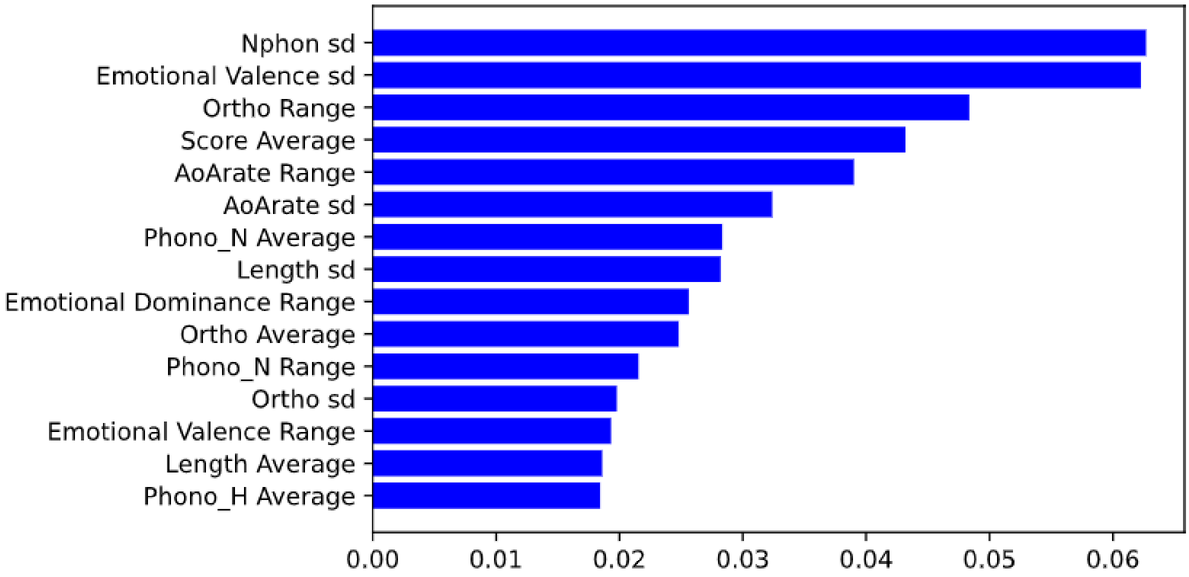

Fig. 2. Feature Importance's.

Table I presents the descriptions of these features.

Top 15 most important features for rf and nn models in the order of importance

Feature	**Description** [13]
Nphon Sd	Standard deviation of number of phonemes in the pronunciation.
Emotional Valence sd	Standard deviation of how pleasant a word is.
Ortho Range	Range of the number of orthographic neighbors a word has.
Score Average	Average of the indicator of when a person said a new verb.
AoArate Range	Range of age of acquisition obtained through adults' ratings.
AoArate sd	Standard deviation of age of acquisition obtained through adults' ratings.
Phono_N Average	Average of the number of phonological neighbors that a word has.
Length sd	Standard deviation of the length of the word.
Emotional Dominance Range	Range of the extent to which the word denotes something that is weak or strong.
Ortho Average	Average of the number of orthographic neighbors a word has.
Phono_N Range	Range of the number of phonological neighbors that a word has.
Ortho sd	Standard deviation of the number of orthographic neighbors a word has.
Emotional Valence Range	Range of how pleasant a word is.
Length Average	Average of the length of the word.
Phono_H Average	Average of the number of phonological neighbors that a word has including homophones.

We used this approach for the rest of the study. Table II presents the averages and standard deviations of the evaluation metrics for RF, NN, and RNN models. As seen in the table, RF slightly outperforms NN and RNN models. This model can detect AD participants with an accuracy of 76%. Note that the RF model relies on features extracted from psycholinguistic properties that require big data pre-processing by subject matter experts. However, despite minimal data pre-processing, the RNN model can correctly detect AD at 67%. Of the three models. Table III lists the p-values of these tests. The table shows that the differences between the RF model and the other models are not statistically significant. This concludes that the results from the RF model are similar to the other two models.

CONCLUSIONS

Even when employing an RNN that requires almost no pre-processing of the subjects' VF data, our findings demonstrate that using NLP, we are able to accurately detect AD with an accuracy that is above the average of chance. Our accuracy scores fall within the reported accuracy ranges of several clinical AD detections methods, such as EEG and brain scans that are considerably more costly and time-consuming than VF tasks [21]. Our results thus show promise for detecting AD using data-driven methods without resorting to cost prohibitive, invasive or time-consuming clinical procedures. As our results indicate, RF performs slightly better detecting AD than NN and RNN. However, the differences are not significant. It is worth noting that the RF requires considerable data pre-processing. While the RF requires analysis and computation of psycholinguistic properties, the RNN requires concatenating the subjects' verbs. The latter methodology provides a much more efficient, time and cost-saving means to detect AD with 67% accuracy and can be conducted remotely. Given these benefits and the insights derived here regarding its potential effectiveness, further exploration into using an RNN with an NLP after collecting subjects' verb listings stands out as a worthy venture. More specifically, future work may include further refining and tuning the RNN and NLP while studying the patient-specific covariates, including age and comorbidities more comprehensively.

REFERENCES

[1] A. Association, "Dementia vs. alzheimer's disease: What is the difference?" 2021. [Online]. Available: https://www.alz.org/alzheimersdementia/ difference-between-dementia-and-alzheimer-s

[2] A. N. Today, "Alzheimer's disease statistics." [Online]. Available: https://alzheimersnewstoday.com/alzheimers-disease-statistics

[3] J. Elflein, "Death rate due to alzheimer's disease in the u.s. 2000- 2019," Alzheimer's Association, 2021.

[4] W. Jarrold, B. Peintner, D. Wilkins, D. Vergryi, C. Richey, M. L. Gorno-Tempini, and J. Ogar, "Aided diagnosis of dementia type through computer-based analysis of spontaneous speech," in Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Baltimore, Maryland, USA: Association for Computational Linguistics, Jun. 2014, pp. 27–37. [Online]. Available: https://www.aclweb.org/anthology/W14-3204

[5] D. Shibata, S. Wakamiya, A. Kinoshita, and E. Aramaki, "Detecting Japanese patients with Alzheimer's disease based on word category frequencies," in Proceedings of the Clinical Natural Language Processing Workshop (ClinicalNLP). Osaka, Japan: The COLING 2016

Organizing Committee, Dec. 2016, pp. 78–85. [Online]. Available: https://www.aclweb.org/anthology/W16-4211

[6] K. Palmer, L. B¨ackman, B. Winblad, and L. Fratiglioni, “Detection of alzheimer’s disease and dementia in the preclinical phase: population based cohort study,” Bmj, vol. 326, no. 7383, p. 245, 2003.

[7] E. J. Paek and L. L. Murray, “Quantitative and qualitative analysis of verb fluency performance in individuals with probable alzheimer’s disease and healthy older adults,” American journal of speech-language pathology, vol. 30, no. 1S, pp. 481–490, 2021.

[8] Q.-Y. Zhong, E. W. Karlson, B. Gelaye, S. Finan, P. Avillach, J. W. Smoller, T. Cai, and M. A. Williams, “Screening pregnant women for suicidal behavior in electronic medical records: diagnostic codes vs. clinical notes processed by natural language processing,” BMC medical informatics and decision making, vol. 18, no. 1, pp. 1–11, 2018.

[9] V. K¨epuska and G. Bohouta, “Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home),” in 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 99–103.

[10] T. Gadekallu, A. Soni, D. Sarkar, and L. Kuruva, “Application of sentiment analysis in movie reviews,” in Sentiment Analysis and Knowledge Discovery in Contemporary Business. IGI Global, 2019, pp. 77–90.

[11] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, B. T. Yeo, A. D. N. Initiative et al., “Predicting alzheimer’s disease progression using deep recurrent neural networks,” NeuroImage, vol. 222, p. 117203, 2020.

[12] H. Li and Y. Fan, “Early prediction of alzheimer’s disease dementia based on baseline hippocampal mri and 1-year follow-up cognitive measures using deep recurrent neural networks,” in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 368–371.

[13] D. A. Balota, M. J. Yap, K. A. Hutchison, M. J. Cortese, B. Kessler, B. Loftis, J. H. Neely, D. L. Nelson, G. B. Simpson, and R. Treiman, “The english lexicon project,” Behavior research methods, vol. 39, no. 3, pp. 445–459, 2007.

[14] L. Breiman, “Random forests,” Machine learning, vol. 45, no. 1, pp. 5–32, 2001.

[15] B. M¨uller, Neural Networks An Introduction, ser. Physics of Neural Networks. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995.

[16] F. Chollet et al., “Keras,” https://keras.io, 2015.

[17] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Man´e, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke,

(IJISE) 2022, Vol. No. 15, Jan-Jun **e-ISSN: 2454-6402, p-ISSN: 2454-812X**

V. Vasudevan, F. Vi´egas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[18] M. Honnibal and I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," 2017, to appear.

[19] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to data mining. Pearson Education India, 2016.

[20] M. Hossin and M. Sulaiman, "A review on evaluation metrics for data classification evaluations," International Journal of Data Mining & Knowledge Management Process, vol. 5, no. 2, p. 1, 2015.

[21] J. E. Gaugler, R. L. Kane, J. A. Johnston, and K. Sarsour, "Sensitivity and specificity of diagnostic accuracy in alzheimer's disease: a synthesis of existing evidence," American Journal of Alzheimer's Disease & Other Dementias®, vol. 28, no. 4, pp. 337–347, 2013.