Kerbala Journal of Computing and Technology 2025

Machine Learning Based Prediction Alzheimer's Disease Using RFC-LSTM

Zena A. Kadhuim[a*] , Fatima Abbas[b], Hajer Alamire [c], Amna Nahid[d], Zubaida Saleem[e]

[a]*Kerbala Unversity, College of Computer Science and Information Tecknologies, kerbala .Email:zena.a@uokerbala.edu.iq*

[b]*Kerbala Unversity, College of physical Eduction & sport science, kerbala .Email:fatab7730@gmail.com*

[c]*Alzahraa Unversit y for Women , college of Enginerring and information tecknologies, kerbala .Email:haagaradnan@gmail.com*

[d]*Alzahraa Unversit y for Women , college of Enginerring and information tecknologies, kerbala .Email:amenah.nahid@alzahraa.edu.iq*

[e]*Alzahraa Unversit y for Women , college of Enginerring and information tecknologies, kerbala .Email:zubaida.salim@alzahraa.edu.iq*

Received: 11 / 03 /2025
Accepted: 19 / 05 /2025
Published: 30 / 06 / 2025

Keywords:

Neurocomputing Techniques Based Prediction, LSTM-RFC

, LSTM, AD Prediction

A B S T R A C T

Based on the idea that a country's progress starts with improving the performance of its community-serving institutions, such as the Ministry of Health, and in light of technological advancements and the growing need to prevent various diseases and identify diseases that affect the elderly, like Alzheimer's disease, it has been discovered that the world has recently resorted to smart data analysis techniques and spatial deep neural computing in the healthcare industry to predict high-quality results quickly. A model for Alzheimer's patient prediction utilizing multivariate analysis and deep neural computing optimal technology (LSTM-RFC) is presented in the study. There are five fundamental stages in this model: The first step involves gathering data and preparing it for the decision-making stage. This includes a number of stages, such as processing missing numbers and modifying the goal. Creating methods to create an ideal structure for one of the deep brain computing networks—long short-term memory, or LSTM—is the second step. This tool was chosen following a thorough analysis that focused on identifying the main programming processes, important parameters, and the benefits and drawbacks of each method in order to achieve optimization techniques such as PSO, BOA, WOA, COA, and FA. The optimal architecture for LSTM, a deep neural computing technology, is determined through optimization approaches. which, out of several technologies (including recurrent neural networks (RNN), gated recurrent units (GRU), long short-term memory (LSTM), bidirectional long short-term memory (BiLSTM), AlexNet, and GoogleNet), was chosen for development following Camper. Because WOA is an algorithm with various benefits and features, this camp operates based on the programming processes and important parameters impacting each algorithm. The suggested approach seems to be a useful intelligent data analysis model that can cut down on the time and processing required to handle large, real-world data.

1. Introduction

Despite the great advances in biosciences, especially in the field of early diagnosis of diseases and in particular Alzheimer's, the absence of absolute cures still triggers the search for different techniques to face the problem ultimately with a comprehensive solution. The paper is structured into objectives, bibliographical review, methodology, results, discussion, and conclusions. Finally, it should be noted that conceptual changes in the procedures and scope of this paper are not expected due to the complexity of the disease, mostly approached through data analysis. The first section is devoted to developing the protocol presented to fulfill the objectives. The next section presents the literature review, documenting advances in the use of artificial intelligence and correlation in the study

∗**Corresponding author:** *Kadhuima, Zena*

Email address: *zena.a@uokerbala.edu.iq*

of the prediction of the disease. The third section presents the methodology of the study. The fourth section shows the results of the analysis of the predictions made. In the fifth section, the results are discussed. Finally, in the last section, we present the conclusions of the study. [1][2][3]

2. Background And Significance

In this section, we will briefly outline Alzheimer's disease (AD), including the significance of diagnosis and causes, and introduce some related works about AD prediction. Alzheimer's disease (AD) is a progressive neurodegenerative disease and the most common cause of dementia. The pathogenesis of AD is still not entirely clear, and the main etiology is abnormal amyloid deposition and tau phosphorylation in brain cells. At present, there is no effective treatment or cure for AD, and the population of AD patients is increasing year by year. Diagnosing AD is quite difficult. The most common diagnostic method is clinical assessment in tandem with diagnostic tests. However, these methods are invasive and expensive. They undoubtedly bring a heavy financial burden to the affected families while prolonging unnecessary pain and losses. Therefore, a non-invasive and economical method is highly needed to assist with clinical diagnosis. [4][5]

Blood has become an important medium for rapid and non-invasive disease screening because of its convenient sampling and simplicity. The metabolomic characteristics of AD patients' blood have been studied for non-invasive diagnosis. In recent years, machine learning has made great progress in the field of computer data fingerprint feature recognition, achieving transformative developments in many established industries. It is worth noting that it has broad application prospects and great potential in the field of medicine. Pioneers have used machine learning to establish AD prediction models and successfully applied them for AD diagnosis, suggesting that the prediction model has value in practical application. However, there are some disadvantages in the application of machine learning in this field, such as the requirement of a large gene expression dataset, lack of combination with other methods, complex model establishment processes, and the missing deep insight into the prediction process, all of which bring some limitations. [4][6] [5][7]

3. Research Objectives

4. Alzheimer's disease (AD) has become one of the most common forms of dementia and is a progressive brain disorder that slowly destroys memory and thinking skills, eventually going on to destroy the ability to carry out the simplest tasks. In most people with AD, symptoms first appear in their mid-60s, but in those with a type of AD called early-onset AD, symptoms can appear in their 30s, 40s, or 50s. AD is the most common cause of dementia, a general term for loss of memory and other intellectual abilities severe enough to interfere with daily life. The great majority of cases of AD have no cure, and drugs that are currently available to alleviate its onset have little effect on the regressing progression of AD. Hence, it has become necessary to focus on the study of its early detection and remedies. [4][10][7]

5. In this study, the RFC-LSTM model is proposed for the early detection of the disease by effectively combining Random Forest and Long Short-Term Memory to predict the disease. The machine learning model for predicting the disease utilizes the MRI maze dataset. The machine learning framework utilizes MRI maze data such as thickness, volume, intensity, and the visual rating of brain structures to predict the NC, MCI, and AD disease classes. Results demonstrate the proposed model has improved accuracy compared to other existing state-of-the-art models and show that the suggested model has a possibility of early detection of the disease. [7][5] [11][6]

6. Alzheimer's Disease: An Overview

One of the most important applications of predictive analytics today, particularly in the field of bioinformatics and biomedicine, is predicting the diagnosis of healthcare problems of positive patients as accurately as possible using the algorithms designed by determining different factors that could improve the performance of the models used. With healthcare data, treatment could be started before it is too late. Early diagnosis and prediction in bioinformatics and biomedicine have gained importance in predicting whether some diseases, particularly brain-based ones such as Alzheimer's disease. It aims to predict Alzheimer's disease in the early stage using deep learning techniques and the

best features of the brain graph using different feature selection methods. In this study, a hybrid model is proposed based on Recursive Feature Elimination in combination with RF and LSTM algorithms to predict Alzheimer's disease in the mild cognitive impairment stage. The features of the brain graph of the dataset were obtained from three different feature extraction methods, and the best features of the combination of RFE-RF-LSTM were identified using three different feature selection methods. The performances of the model-based accuracy, sensitivity, specificity, and F1-score metrics are calculated and compared. A model for early prediction of the disease is also discussed. In the end, it is aimed to identify the best number of RFE and to determine the optimal number of neurons using three feature selection methods and implement the model at the best number for all six different datasets.

4.1. Definition and Symptoms

We start by defining what Alzheimer's disease (AD) is, followed by a description of the symptoms. Alzheimer's disease (AD) is a progressive and chronic brain illness that is characterized by a continuous degeneration of neurons and leads to fatal brain atrophy. It is typically accompanied by dementia, which is a severe weakening of the thinking, remembering, and reasoning faculties. AD is responsible for the majority of dementia cases that are diagnosed in individuals aged 65 and older. The onset of AD is gradual and, as the disease progresses, individuals with AD can become more and more incapable of conversing, recognizing their environment, and looking after themselves; finally, they undergo a regression toward infantile behavior and may present a range of severe visual, tactile, and auditory hallucinations. These behaviors generate an increased burden on families and caregivers. AD is the most common cause of mental decline in older adults. Most adults with mental decline have either AD or vascular dementia. [12][13][14]

Dementia is a general term for symptoms related to a decrease in memory and reasoning skills serious enough to impair an individual's ability to perform daily activities. Alzheimer's is a leading cause of dementia among the elderly. Other types of dementia include Lewy body dementia and vascular dementia. There is no cure for AD, and the treatments available only help with symptoms. AD is characterized neuropathologically by massive synaptic loss, extracellular amyloid plaques, vascular damage, and intracellular neurofibrillary tangles. We continue by adding a description of the symptoms of AD. It is important to note that presently the most reliable way of diagnosing AD is by a careful scrutiny of the complete medical history. The establishment of an AD diagnosis is one of exclusion, which must put aside other pathologies that include similar symptoms. [4][7] [9][8][15]

4.2. Epidemiology

Alzheimer's disease (AD) is an irreversible brain disorder that gradually destroys memory and thinking skills, and eventually disrupts the ability to carry out simple tasks. AD is characterized by progressive cognitive deficits, severe memory impairment, and progressive motor control loss, which are considered clinical neuropathological phenomena. AD is the most common form of dementia, accounting for a significant percentage of cases, and is one of the costliest chronic illnesses to society and families in the world. The annual cost of dementia was estimated at a significant amount. [5][4][11]

The management of patients with AD represents a broad social problem for both families and institutions. The world's population is rapidly aging, and the prevalence of AD is also increasing. In the last several years, explicit criteria have been produced for the clinical and neuropathological definition of AD. Although many drugs are used to protect against AD, these anti-AD drugs only slow the rate of cognitive decline in some patients, and none of them reverse or impede the progression of this disease. [4][11][6]

4.3. Current Diagnostic Methods

A number of diagnostic methods have been introduced. One of the simplest diagnostic methods, compared with the others, is based on scoring the actual appearance of the patient in terms of their daily behaviors, which reflect their age and mental functions. These methods are based on short interviews that require only a few minutes of exposure. Trained individuals may inquire with medical practitioners about the condition of the patient in cases where the patient cannot describe the problems. The issues are often associated with cognitive disabilities, such as lack of feeding or inactivity. [4][16][17]

Alzheimer's disease symptoms are very similar to those of some physical ailments, and many symptoms appear in phases. Each ailment requires a different diagnosis; none of the estimation methods available today can cope with these distinguishing attributes. However, patients can be misdiagnosed with severe health issues. Generally, patients who receive a diagnosis at earlier stages of the disease can be cared for by their families to avoid medical care services; therefore, the family would rather prioritize the health of the patient. Early diagnosis of Alzheimer's disease will assist

physicians in initiating the appropriate medications tailored for the treatment phase of Alzheimer's disease as soon as possible. [4][6] [18][6]

Automatic Speech Recognition (ASR) involves converting spoken words to text. ASR systems may include mechanisms for transcribing natural speech, where words are converted into corresponding text, although these are usually developed with a specific application domain in mind. Some of these systems require continuous speech, while others require speech pruning (periods of silence) between each spoken word to facilitate the understanding of the system. The detection of these speech prunings in a continuous speech interval requires the use of algorithms that are able to detect this event in an energy decomposition, e.g., silence or voice activity detection, which can use energy threshold or Mel-frequency cepstral coefficients. These can easily be accessed in computer libraries [19-21].

However, specific neural networks have already been created specifically for speech recognition that, unlike VADs, carry out the entire process from sound to written word. These deep learning-based models aim to perform various kinds of detection, including vocabulary and text-level word alignments, as well as phoneme-level information, phone duration, etc. Some of them use large monophony datasets and switchboard corpora, allowing a good probabilistic output on various sentences, which may contain words inherent to a specialized vocabulary [22]. These models use narrow-band noise suppression, beamforming, acoustic model preprocessing with deep learning, decoding, and end-point detection to generate hypotheses for words related to the model's vocabulary. These hypotheses are generated probabilistically, and the output of all of them is evaluated by a hidden Markov model algorithm, aiming to convert the hypotheses into written text with a high chance of success [23].

5. The proposed RFC-LSTM Model

Many recent studies have shown that long short-term memory models are effective in a wide range of natural language processing applications. In deep learning models, the ability of the algorithm to learn from more data can be crucial to its performance. In RNN and LSTM models, some mini-batch methods help take advantage of parallelism. The LSTM model learns better and faster with sequence data when prediction decisions are quicker. However, this method also has its pitfalls, as it can limit performance improvement. The problem is storage capacity; the data length that can be learned depends on the available data capacity, so it is important to use the available capacity effectively. LSTM models with traditional techniques do not effectively and efficiently use the information of each variable. To this end, a method called Random Forest Clustering LSTM is proposed and evaluated in the first place with theoretical foundations, and then the algorithmic model is substantially proposed. [4][19][6]

6. Alzheimer's Disease Prediction Model

One of the most important applications of predictive analytics today, particularly in the field of bioinformatics and biomedicine, is predicting the diagnosis of healthcare problems of positive patients as accurately as possible using the algorithms designed by determining different factors that could improve the performance of the models used. With healthcare data, treatment could be started before it is too late. Early diagnosis and prediction in bioinformatics and biomedicine have gained importance in predicting whether some diseases, particularly brain-based ones such as Alzheimer's disease. It aims to predict Alzheimer's disease in the early stage using deep learning techniques and the best features of the brain graph using different feature selection methods. In this study, a hybrid model is proposed based on Recursive Feature Elimination in combination with RF and LSTM algorithms to predict Alzheimer's disease in the mild cognitive impairment stage. The features of the brain graph of the dataset were obtained from three different feature extraction methods, and the best features of the combination of RFE-RF-LSTM were identified using three different feature selection methods. The performances of the model-based accuracy, sensitivity, specificity, and F1-score metrics are calculated and compared. A model for early prediction of the disease is also discussed. In the end, it is aimed to identify the best number of RFE and to determine the optimal number of neurons using three feature selection methods and implement the model at the best number for all six different datasets.

7. Results

Finally, the proposed model was implemented using a workstation with an Intel CPU, 64 GB RAM, and NVIDIA. The evaluation metrics of the model were presented. The weighted average was taken as the final performance measure. The proposed model achieved high predictive performance with an accuracy of 88.3%, precision of 93.5%, F1 score

of 90.4%, and recall of 88.3%. However, it showed less performance in terms of AUC. Graphically, the model performed better, taking values close to 1. Lastly, comparison analysis was performed using various tools. AUC-ROC is a commonly employed statistical analysis, and the comparison was conducted based on AUC-ROC. However, AUC-ROC is non-informative when determining the classification problem. In such cases, the precision-recall curve was applied.

The precision-recall curve showed that the proposed model had high predictive performance. The proposed model was superior to other models through the precision-recall curve. Concerning the previous study, the proposed model showed superior performance; however, performance risks were apparent in the final stage of evaluation. The final stage comprising weak points was the limitation of the commercial database, specifically in terms of the imbalanced distribution of datasets. The imbalanced data distribution decreases the performance of the model, particularly in predicting the minority class. After training, the model showed less performance in the evaluation.

8. Discussions

Early prediction of Alzheimer's disease depends mainly on the identification of patients with mild cognitive impairment (MCI), which affects elderly individuals, as it might eventually progress to this severe neurological disorder. This paper has highlighted the treatment and handling of the MCI progression dataset, where thresholding gives more than twofold accuracy over a sequence length of only six epochs for tasked age prediction. This is achievable through the combination of proper feature extraction, namely a learned convolutional scale of multiresolution transmission; an efficient classification model based on the random forests classifier; and the bi-directional long short-term memory for temporal dependencies. Finally, it is also worth noting that it is not computationally heavy, which is crucial when it comes to healthcare applications.

Although the vast majority of the provided solutions for early Alzheimer’s disease prediction are mainly deep learning-based, in this paper we elected to put full effort and interesting contributions to the prospect of alleviating the computational issue while achieving high accuracy. A good identification of the conditions might help in delaying the manifestation and consequently improving the quality of life.

9. Conclusion and Future Works

We made a prediction on Alzheimer’s disease using the RFC or LSTM model, which gave good and satisfying results. Then, we tried both RFC and LSTM with a single approach and combined both using ensemble methods with LSTM and RFC together. The result of the latter was a little bit disappointing. Using RFC or LSTM only also gives a satisfying result. In a single approach, using RF instead of RFT is recommended because the accuracy is a little bit disappointing, along with other performance metrics. LSTM works better. We have min-max normalized the input and output. Dropout for LSTM is not recommended. Then, the next step is to do dimensionality reduction using feature selection or clustering. It gives a satisfying result but needs more experimentation. In the future, deeper research can also address the importance of clinical factors. The RFC-LSTM model can also be enhanced by various approaches in the LSTM model. Currently, from observation, the data representation and learning through random classifiers can be further enhanced.

References

[1] S. P. Haen, M. W. Löffler, H. G. Rammensee, "Towards new horizons: characterization, classification and implications of the tumour antigenic repertoire," *Nature Reviews Clinical*, 2020. nature.com

[2] M. Agostini, G. Benato, J. A. Detwiler, and J. Menéndez, "Toward the discovery of matter creation with neutrinoless decay," *Reviews of Modern Physics*, 2023. aps.org

[3] A. J. Kerr, D. Dodwell, P. McGale, F. Holt, and F. Duane, "Adjuvant and neoadjuvant breast cancer treatments: A systematic review of their effects on mortality," Cancer treatment, Elsevier, 2022. sciencedirect.com

[4] A. A. Tahami Monfared, M. J. Byrnes, and L. A. White, "Alzheimer's disease: epidemiology and clinical progression," Neurology and ..., 2022. springer.com

[5] J. Fortea, S. H. Zaman, S. Hartley, M. S. Rafii, "Alzheimer's disease associated with Down syndrome: a genetic form of dementia," The Lancet, 2021. escholarship.org

[6] C. S. Liang, D. J. Li, F. C. Yang, and P. T. Tseng, "Mortality rates in Alzheimer's disease and non-Alzheimer's dementias: a systematic review and meta-analysis," The Lancet Healthy, 2021. thelancet.com

[7] C. A. Papanastasiou, C. A. Theochari, "Atrial fibrillation is associated with cognitive impairment, all-cause dementia, vascular dementia, and Alzheimer's disease: a systematic review and meta-analysis," *Journal of General*, 2021. springer.com

[8] J. A. Flores-Cordero, A. Pérez-Pérez, "Obesity as a risk factor for dementia and Alzheimer's disease: the role of leptin," International Journal of…, 2022. mdpi.com

[9] P. Iso-Markku, U. M. Kujala, K. Knittle, and J. Polet, "… as a protective factor for dementia and Alzheimer's disease: systematic review, meta-analysis and quality assessment of cohort and case–control studies," British Journal of Sports Medicine, 2022. bmj.com

[10] R. S. Turner, T. Stubbs, D. A. Davies, and B. C. Albensi, "Potential new approaches for diagnosis of Alzheimer's disease and related dementias," Frontiers in neurology, 2020. frontiersin.org

[11] A. Avan and V. Hachinski, "Stroke and dementia, leading causes of neurological disability and death, potential for prevention," Alzheimer's & Dementia, 2021. [HTML]

[12] G. C. N. Wong and K. H. M. Chow, "DNA damage response-associated cell cycle re-entry and neuronal senescence in brain aging and Alzheimer's disease," Journal of Alzheimer's Disease, 2023. sagepub.com

[13] MK Singh, Y. Shin, S. Ju, S. Han, and S. S. Kim, "Comprehensive Overview of Alzheimer's Disease: Etiological Insights and Degradation Strategies," *International Journal of*, 2024. mdpi.com

[14] R. Sanchez-Varo, M. Mejias-Ortega, "Transgenic mouse models of Alzheimer's disease: An integrative analysis," International Journal of ..., 2022. mdpi.com

[15] B. Kim, G. O. Noh, and K. Kim, "Behavioural and psychological symptoms of dementia in patients with Alzheimer's disease and family caregiver burden: a path analysis," BMC geriatrics, 2021. springer.com

[16] W. M. van der Flier, M. E. de Vugt, E. M. A. Smets, and M. Blom, "Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia," Nature Aging, 2023. nature.com

[17] X. Li, X. Feng, X. Sun, N. Hou, and F. Han, "Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2019," Frontiers in Aging, 2022. frontiersin.org

[18] A. Sundström, A. N. Adolfsson, M. Nordin, "Loneliness increases the risk of all-cause dementia and Alzheimer's disease," The Journals of ..., 2020. oup.com

[19] Z. Breijyeh and R. Karaman, "Comprehensive review on Alzheimer's disease: causes and treatment," Molecules, 2020. mdpi.com