

Article

A Long Short-Term Memory Biomarker-Based Prediction Framework for Alzheimer's Disease

Anza Aqeel ¹, Ali Hassan ¹ , Muhammad Attique Khan ² , Saad Rehman ², Usman Tariq ³ , Seifedine Kadry ⁴ , Arnab Majumdar ⁵ and Orawit Thinnukool ^{6,*} 

¹ Department of Computer & Software Engineering, CEME, NUST, Islamabad 44800, Pakistan; anzaaqeel@hotmail.com (A.A.); alihasan@ce.ceme.edu.pk (A.H.)

² Department of Computer Engineering, HITEC University, Taxila 47080, Pakistan; attique@ciitwah.edu.pk (M.A.K.); saad.rehman@hitecuni.edu.pk (S.R.)

³ College of Computer Engineering and Science, Prince Sattam Bin Abdulaziz University, Al-Kharaj 16242, Saudi Arabia; u.tariq@psau.edu.sa

⁴ Department of Applied Data Science, Noroff University College, 4608 Kristiansand, Norway; skadry@gmail.com

⁵ Department of Civil Engineering, Imperial College London, London SW7 2AZ, UK; a.majumdar@imperial.ac.uk

⁶ College of Arts, Media and Technology, Chiang Mai University, Chiang Mai 50200, Thailand

* Correspondence: orawit.t@cmu.ac.th

Abstract: The early prediction of Alzheimer's disease (AD) can be vital for the endurance of patients and establishes as an accommodating and facilitative factor for specialists. The proposed work presents a robotized predictive structure, dependent on machine learning (ML) methods for the forecast of AD. Neuropsychological measures (NM) and magnetic resonance imaging (MRI) biomarkers are deduced and passed on to a recurrent neural network (RNN). In the RNN, we have used long short-term memory (LSTM), and the proposed model will predict the biomarkers (feature vectors) of patients after 6, 12, 21, 18, 24, and 36 months. These predicted biomarkers will go through fully connected neural network layers. The NN layers will then predict whether these RNN-predicted biomarkers belong to an AD patient or a patient with a mild cognitive impairment (MCI). The developed methodology has been tried on an openly available informational dataset (ADNI) and accomplished an accuracy of 88.24%, which is superior to the next-best available algorithms.

Keywords: Alzheimer's; long short-term memory; artificial neural network; machine learning



Citation: Aqeel, A.; Hassan, A.; Khan, M.A.; Rehman, S.; Tariq, U.; Kadry, S.; Majumdar, A.; Thinnukool, O. A Long Short-Term Memory Biomarker-Based Prediction Framework for Alzheimer's Disease. *Sensors* **2022**, *22*, 1475. <https://doi.org/10.3390/s22041475>

Academic Editor:
Christophoros Nikou

Received: 23 December 2021

Accepted: 13 February 2022

Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Alzheimer's disease (AD) is a developing and unchangeable degenerative cerebral illness defined by recognition failure and psychological disability. According to [1] the Alzheimer's Association (2019), in recent times, approximately, 90 million people are admitted to hospitals due to AD worldwide. According to the study by [2,3], the estimated number of deaths due to Alzheimer's disease will reach 300 million by the year 2050. There is still no specific treatment that has been developed which can cure or stop this disease and save the patient's life. A few treatments are used by medical practitioners that can delay the rate of AD when detected at the initial stages.

Mild cognitive impairment (MCI) [4] impacts memory, language, thinking, and judgment in a way that is more noticeable than conventional age-related changes. If an individual has minor intellectual impairment, he/she may realize that their memory or intellectual ability has "slipped." People close to the person may see more significant change. Regardless, these changes are usually not extensive enough to interfere with day-by-day life and typical activities. An impaired mental capacity may become worse over time, causing dementia or other neurological conditions. In any case, some patients with minor impairments do not become worse, and some even improve in the long term.

Mild cognitive impairment is one of the preliminary signs of AD [5]. Recent neuroimaging technology is widely employed to identify a few essential biomarkers in the brain for the diagnosis of brain tumors [6]. Similarly, this research brings about an automated system in which NM and MRI values are calculated as important biomarkers to detect AD in the human brain.

Although there is no proven biomarker for the correct or accurate prediction of whether a patient will progress from MCI to AD, multiple approaches have been adapted in different studies to predict the growth of disease. Studies that use different prediction models to predict the growth of disease mostly depend on clinical modalities like magnetic resonance imaging scans [7,8], neuropsychological measures, computerized tomography scans [9,10], diffusion tensor imaging, CSF biomarkers, and positron emission tomography [11,12]. For example, in a study conducted by [13], the authors used a combination of cerebrospinal fluid, fluorodeoxyglucose–positron emission tomography, and MRI biomarkers for the classification of patients who will progress to AD from MCI (stable state). Nevertheless, studies about reducing the growth of disease over time are limited. In the medical domain, most devices are employed for public use, and it is easy to access medical data from several sources [14,15]. These sources can be MRI and CT scans that are longitudinally obtained [16,17]. Conventional methods are mostly used for the analysis of knowledge extraction among several variables (quantities). ML algorithms have provided much help in the prediction of AD [18]. By using algorithmic techniques, it has become possible to gather relevant input data, and to perform some algorithms and generate output, which can be a prediction of the highlighted disease. These techniques can also be beneficial for finding the relationships within the input data, and can help in the early prediction of diseases.

MCI is an early-stage disease which is generally not harmful, and patients who have this disease are likely to function at normal levels. A prediction model for predicting which patients will progress from MCI to AD could help with early diagnosis and an early cure or treatment of the disease. Thus, the significance of the study is also the same as that which will contribute to the question of whether a patient's MCI will progress into AD or not. Various biological markers may be used in the prediction of a patient's disease by their doctor. Similarly, the ML model will forecast the progression of disease using some of the key biomarkers. The NM and MRI values of the patients are biomarkers, like features in ML. Based on these biomarkers, the ML model will create a prediction about whether the patient will progress to AD in the next three years or not. This process is helpful for doctors and clinical staff for speeding up the prediction process. Therefore, a need exists to formulate the question: "To what extent can a classifier predict the progression of subjects from MCI to AD, based on NM and MRI biomarkers?"

Many techniques have been introduced within the research of computer vision in the medical domain, such as fuzzy clustering [19], prostate zonal segmentation [20], and a few more [21]. Recently, deep learning has had a great impact in the area of medical diagnoses [22], such as for skin cancer [23–25], brain tumors [26], stomach [27], COVID-19 [28], person re-identification [29] and a few more [30]. The projected work proposes a ML model for the prediction of progression to Alzheimer's disease using long short-term memory in a recurrent neural network. Deep learning has revolutionized the area of image and video processing and computer vision [31]. In the proposed model, the NM and MRI biomarkers (feature vectors) are computed and passed to the RNN. In the RNN, we use LSTM, which has never been used in the context of AD before. This is the gap that as researchers, we are trying to fill in order to understand the more intricate and complex particularities of AD. This model will predict the biomarkers (feature vectors) of patients after 6, 12, 18, 24, and 36 months. Then, these predicted biomarkers be passed through a fully connected neural network model, a multi-layer perceptron, also known as a convolutional neural network. It will predict whether these RNN-predicted biomarkers belong to an AD patient or an MCI patient. If the predicted values provided by RNN lay within the range of values depicting whether a person will have AD in the future, then the

patient's record can be updated, and whether they are more likely to evolve from MCI to AD in the future will also be predicted by CNN.

Contributions of this research are as follows:

- Projecting future clinical variations in biomarker values, only utilizing initial/benchmark information/data.
- RNN is performed to predict biomarker values and then rankings, followed by a fully connected neural network model (multi-layer perceptron) for classification, in which an accuracy of 88.24% is achieved.
- Identifying the strongest indicators of transformation in unimodal and multimodal settings.
- This study is significant for medical practitioners and health care workers in the early prediction and detection of Alzheimer's disease; moreover, future researchers can adopt this model as a basis for their studies to further contribute to the development of algorithms for predicting Alzheimer's disease in the future.
- This study also serves as a training tool for medical institutions to educate and train their students regarding the early prediction and development of Alzheimer's disease.

The paper is structured as follows: Section 2 involves the related literature, Section 3 portrays the adapted methodology, Section 4 evaluates the outcomes and provides a short comparison with recently distributed work. Section 5 concludes the paper.

2. Literature Review

Recently, deep learning has shown improved performance for medical applications [32,33] such as breast cancer [34], retinopathy [35], COVID-19 [36], and many more [37]. The need to foresee the advancement to AD from MCI is consistently important to help treat this illness in its initial phase. It is critical to understand how this illness develops after some time, and for a better understanding it is imperative to know of related irregularities that happen in the cerebrum. Familiarity with these irregularities is important to choose the attributes that predict progression to AD. Recently, numerous ML methods have been proposed for AD forecasting. The vast majority of strategies rely upon the quantity of features, while not many of them are dependent on clinical features. In [3], a hybrid approach for the analysis of the hippocampus using MRI in AD patients was presented. They used a pre-trained DenseNet architecture for the intensity and shape of features. Then, they computed and trained high-level features by combining RNN, and then performed a final classification. The "Alzheimer's Disease Neuroimaging Initiative" Database (www.adni.loni.edu (accessed on 22 December 2021)) was used for experimental analysis and showed improved results compared to existing techniques. Basheera & Ram [38] proposed a deep learning-based methodology for Alzheimer's disease classification. They used MRI types such as T2 weighted volumes, which included 635 MRIs of AD patients and 548 MCI patients. They extracted gray matter features from MRI voxels and passed these to a convolutional neural network. Then, they enhanced brain voxels using a Gaussian filter, and removed the irrelevant masses by skull stripping. Later, segmentation was performed by component analysis, and they passed the output image to CNN for the final prediction. Overall, 90% clinical accuracy was attained, which was better than existing techniques. Basheera & Ram [39] presented a hybrid clustering and CNN model for the prediction of AD, MCI, and CN. They performed skull stripping at the initial stage and improved enhancement via a Gaussian filter. Then, they combined K-means and expectation maximization (EM) methods, and performed segmentation. The fragmented pictures were passed to the CNN model for feature extraction and the final prediction. The authors of [40] presented a multitasking ML model for AD prediction. The regression model defined each task separately and predicted a cognitive score. This interaction was produced for all tests, and towards the end, a relationship was elucidated among them. Finally, the multiple task scores performed were passed to a slope boosting piece for better forecasting. Thus, this interaction assisted in eliminating the irrelevant features for a better prediction. Lei [41] presented an AD prediction framework using longitudinal data. In this study, they calculated the clinical score as a feature value. Then, they performed

feature selection via a corr-entropy approach. Next, the selected features were encoded in a deep polynomial network. Finally, the prediction was performed through a support vector machine. The ADNI dataset was used for the experiment and attained an impressive performance. Furthermore, a biomarker-based approach was also presented by [42], which shows its performance for the correct prediction of patients' development towards AD from normal levels. In this work, NM and MRI biomarkers are extracted as feature vectors, and perform learning through autoregressive modeling.

3. Proposed Methodology

The proposed work comprised of deep learning and biomarker techniques for AD prediction. The proposed framework consists of few primary steps, as shown in Figure 1, such as original baseline data, biomarker extraction as feature vectors, learning features through RNN type name LSTM, and prediction via MLP. Before prediction via CNN, the recurrent neural network features were updated based on the monthly patient record. The details of each listed step in Figure 1 are shown below.

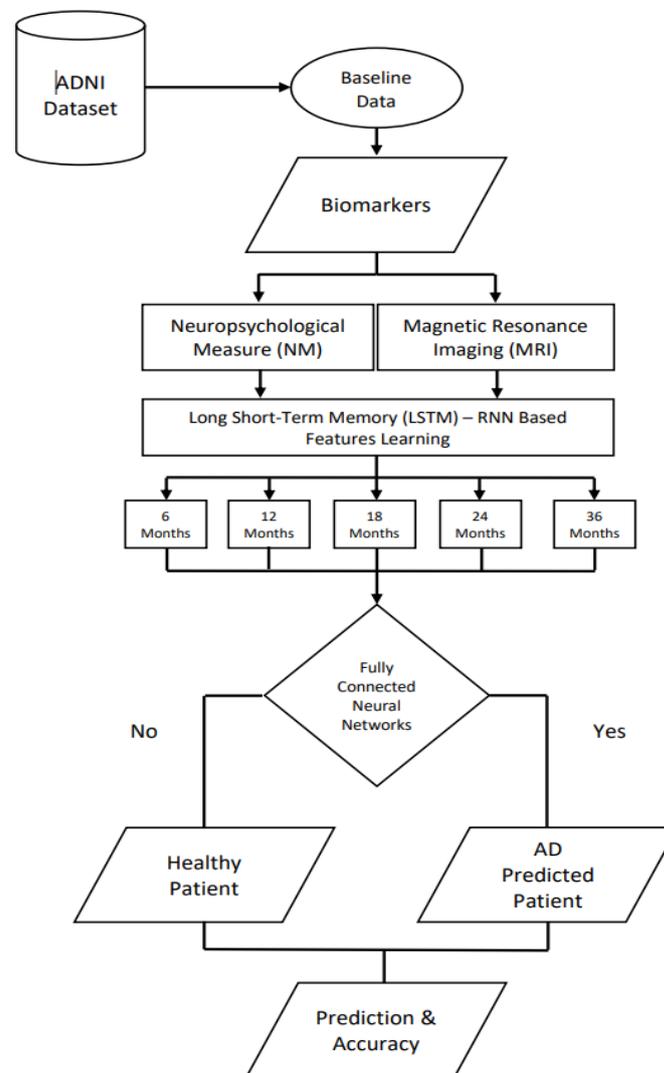


Figure 1. The proposed architecture of Alzheimer's disease's prediction.

3.1. ADNI Dataset

For AD, the most mainstream sample set is the ADNI, which was used in this work for validation of the projected framework. In this dataset, 805 subjects were included. These cases were gauge MRI T1-weighted (T1w) information. The ADNI dataset (www.adni.loni.edu)

(accessed on 22 December 2021)) incorporates the positive biomarkers of patients after every 6, 12, 18, 24, and 36 months from the standard. The primary objective of utilizing this dataset was to inspect whether MRI, positron emission tomography, organic markers, and clinical evaluations can be combined to encompass the improvement of MCI and early findings of Alzheimer's disease.

3.2. RNN-LSTM

In this work, the NM and MRI biomarkers were extracted as feature vectors. A feature vector is a vector containing multiple elements about an object. The purpose of the LSTM model is to predict the future feature vectors (biomarkers) of the patient. The biomarker value changes as per the patient's condition. This model can predict what condition the patient's brain will be in after 6, 12, 18, 24, and 36 months. A proposed algorithm was trained using 805 patients' data, as mentioned in the above section. We provided baseline (0 months or the first NM + MRI test of the patient), NM, and MRI biomarkers (feature vectors) as inputs in the model. We trained our model on the biomarkers of patients after 6, 12, 18, 24, and 36 months, as provided in the dataset. Mathematically, this model was formulated as follows: the standard engineering of the LSTM network comprises an info layer, a repetitive LSTM layer, and a yield layer. The data layer is linked with the LSTM layer, as demonstrated in Figure 2.

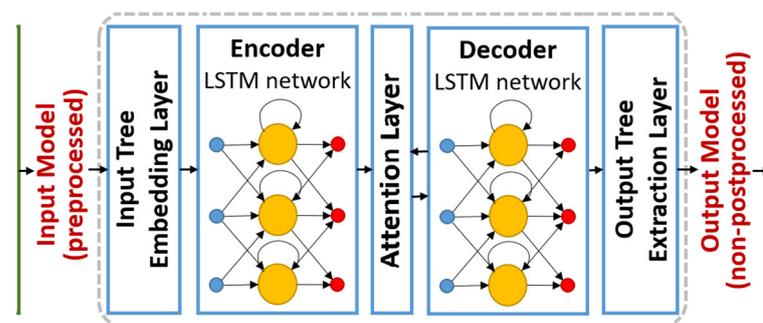


Figure 2. The architecture of the LSTM model (www.modelling-languages.com/lstm-neural-network-model-transformations/ (accessed on 22 December 2021)).

The tedious relations in the LSTM layer are between cell input units and yield units, inputs, yield entryways, and disregard doorways. The cell yield units are related to the yield layer. The number of limits, P , in a standard LSTM network containing one cell in each memory block, can be calculated as:

$$P = l_c \times l_c \times 4 + l_i \times l_c \times 4 + l_c \times l_o + l_c \times 3 \quad (1)$$

where l_c means the memory cells, l_i is the number of information components, and l_o is the quantity of yield units. The time intricacy of the LSTM learning network with a stochastic inclination plummet (SGD) advancement strategy is $O(1)$. Similarly, intricacy per time step is $O(P)$. Learning time is affected by the factor $l_c \times (l_c + l_o)$, when the quantity of information sources is generally few. The LSTM learning model becomes costly in terms of intricacy, when the quantity of yields and memory cells are large. The accompanying conditions charts the organization unit enactments iteratively from $z = 1$ to Z , in order to ascertain predictions from an information arrangement $a = (a_1, \dots, a_Z)$ to $b = (b_1, \dots, b_Z)$:

$$g_z = \sigma(M_{ga}a_z + M_{gk}k_{z-1} + M_{gj}j_{z-1} + bias_g) \quad (2)$$

$$h_z = \sigma(M_{ha}a_z + M_{kh}k_{z-1} + M_{jh}j_{z-1} + bias_h) \quad (3)$$

$$j_z = h_z \odot j_{z-1} + g_z \odot s(M_{ja}a_z + M_{jk}k_{z-1} + bias_j) \quad (4)$$

$$o_z = \sigma(M_{oa}a_z + M_{ok}k_{z-1} + M_{oj}j_{z-1} + bias_o) \quad (5)$$

$$k_z = o_z \odot t(j_z) \quad (6)$$

$$b_z = M_{bk}k_z + bias_b \quad (7)$$

where M is the weight lattices (e.g., M_{ga} is the weight grid from input door to yield entryway), σ means the strategic sigmoid capacity, g , h , j , and o are the information door, neglect door, cell enactment vector, and yield door, k is the cell actuation vector, \odot computes the results of the vectors, and s and t are the initiation capacities (mostly tanh) for cell information and yield. The equations for the last LSTM model with both repetitive and non-intermittent projection layers are depicted beneath:

$$g_z = \sigma(M_{ga}a_z + M_{gx}x_{z-1} + M_{gj}j_{z-1} + bias_g) \quad (8)$$

$$h_z = \sigma(M_{ha}a_z + M_{hx}x_{z-1} + M_{hj}j_{z-1} + bias_h) \quad (9)$$

$$j_z = h_z \odot j_{z-1} + g_z \odot s(M_{ja}a_z + M_{jx}x_{z-1} + bias_j) \quad (10)$$

$$o_z = \sigma(M_{oa}a_z + M_{ox}x_{z-1} + M_{oj}j_{z-1} + bias_o) \quad (11)$$

$$k_z = o_z \odot t(j_z) \quad (12)$$

$$x_z = M_{xk}k_z \quad (13)$$

$$y_z = M_{yk}y_z \quad (14)$$

$$b_z = M_{bx}x_z + M_{by}y_z + bias_b \quad (15)$$

where x denotes the recurrent and y non-recurrent unit activation functions.

3.3. Multi-Layer Perceptron (MLP)

MLP is a type of ANN. The term MLP is used vaguely, and at times freely, to any feedforward ANN, now and again stringently alluding to networks made out of distinctive layers of perceptron. MLP contains 3 layers: a data layer, a hidden layer, and an output layer. As well as the information center points, each center point is a neuron that uses nonlinear activation. MLP utilizes a directed-learning procedure returned to spread for planning. Its various layers and non-direct establishments perceive MLP from a straight perceptron. It can perceive data that is not straightforwardly detachable. In case a multi-layer perceptron has an abrupt authorization work in all neurons, that is, a straight limit that maps the weights to the output of each neuron, direct polynomial numbers show that many layers can be diminished to a 2-layer I/O model. In MLPs, a couple of neurons use nonlinear inception work that was made to show the repeat of movement prospects or ending of neurons.

In later developments, a rectifier linear unit (RELU) is used as one of the likely ways to deal with and conquer the numerical issues determined to have the sigmoid from time to time. The multi-layer perceptron (MLP) contains, in any event, three layers (an information and a yield layer, with one mystery layer) of nonlinear-starting center points. Since multi-layer perceptrons are related, each center point in one layer interfaces with a particular weight W_{ij} to every center point in another layer. The phrase “multi-layer perceptron” does not imply a single perceptron that has different layers. Instead, it contains various perceptrons that are composed into layers. An alternative name is a “multi-layer perceptron network”. Furthermore, MLP “perceptrons” are not perceptrons in the strictest possible sense. Authentic perceptrons are a formally unprecedented instance of fake neurons that use a cut-off order limit, for example, the Heaviside step work. MLP perceptrons can utilize self-assertive actuation capacities. A genuine perceptron performs in a two-fold arrangement, whereas an MLP neuron is free to either perform in order or relapse, contingent on its actuation work (see Figure 3).

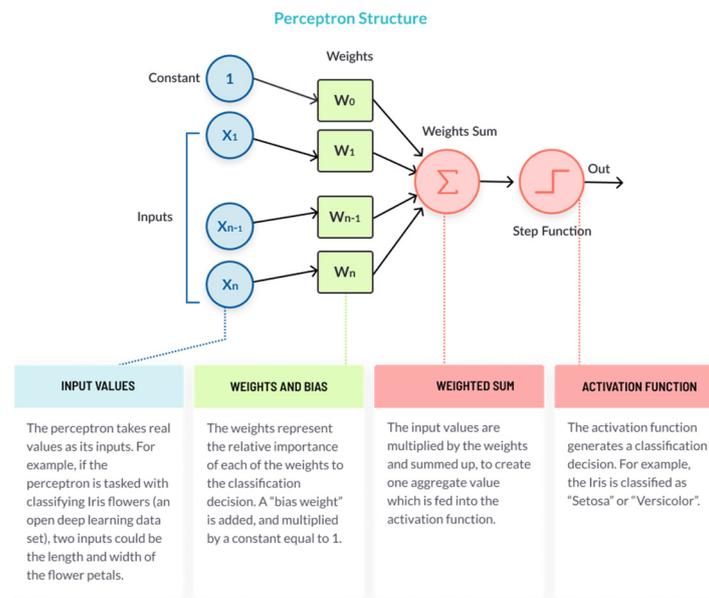


Figure 3. Perceptron structure.

The expression “multi-layer perceptron” was later applied without regard to the nature of the hubs/layers, which can be made out of discretionarily characterized counterfeit neurons and not perceptrons explicitly. This translation dodges the loosening of the meaning of “perceptron” to mean a counterfeit neuron. The perceptron, or neuron in a neural organization, has a basic yet sharp construction. It comprises four sections:

- It takes the data sources, duplicates them by their loads, and calculates their total
- It adds an inclination factor, the number 1 duplicated by a weight
- It feeds the aggregate through the enactment work
- The result is the perceptron yield

Despite the fact that multi-layer perceptrons and neural organizations are basically the same thing, a couple of fixes need to be added before a multi-layer perceptron becomes a full neural organization. These are back spread, hyper boundaries, and advanced constructions.

4. Results and Discussion

The projected system is presented as numerical and tabular results. The validation of the proposed technique was performed using the ADNI dataset as previously explained. The patients’ data were collected after 6, 12, 18, 24, and 36 months. Based on this data, LSTM was trained and output was passed to a multi-layer perceptron network (artificial neural network). This classifier predicts MCI and AD patients based on their monthly values. The performance of this method was calculated by accuracy measures and a false negative rate. Furthermore, the confusion matrix was also produced for verification of the proposed results. Then, the proposed results were compared with previous techniques at the end of this section. All results were computed through a 70:30 approach where cross-validation was 5-Fold. Python was used for implementing this approach on a Corei7 personal computer with 16GB of RAM.

4.1. Results

The proposed forecast results are outlined regarding exactness esteems, root mean square mistakes, and connection coefficients. The outcomes were determined using various mixes of hidden layers. A hidden layer is situated between the information and yield of the calculation, where the capacity applies loads to the data sources, and guides them through initiation work as the yield. Thus, hidden layers perform nonlinear changes to the information sources put into the organization. Hidden layers fluctuate depending on the

capacity of the neural organization, and comparatively, layers may change depending on their related loads.

Hidden layers consider the capacity of a neural organization to be separated into explicit changes in the information. Each hidden layer work is specific to deliver a characterized yield. For instance, hidden layer works that are used to recognize natural eyes and ears might be utilized by resulting related layers to distinguish faces in pictures. While the capacity to distinguish eyes alone is not adequate to freely perceive objects, these layers can work together inside a neural organization.

The accuracy chart (Figure 4) depicts a clear illustration of results when the output model results from the recurrent neural networks (LSTM) were passed through a fully connected neural network model, MLP and ANN. Hidden layers play important roles in determining the accuracy of the predicted results, that is, whether these results are accurate or not. The x-axis shows the number of hidden layers used, while the y-axis states the accuracy results. The result with 88.24% accuracy was chosen to be the best result obtained, and this result was obtained using a 5-Fold cross-validation model. K-fold cross CV is a calculation where a given instructive list is separated into a K number of zones/folds, where each wrinkle is used as a testing set in the long run. We should take the circumstance of 5-Fold cross-validation ($K = 5$).

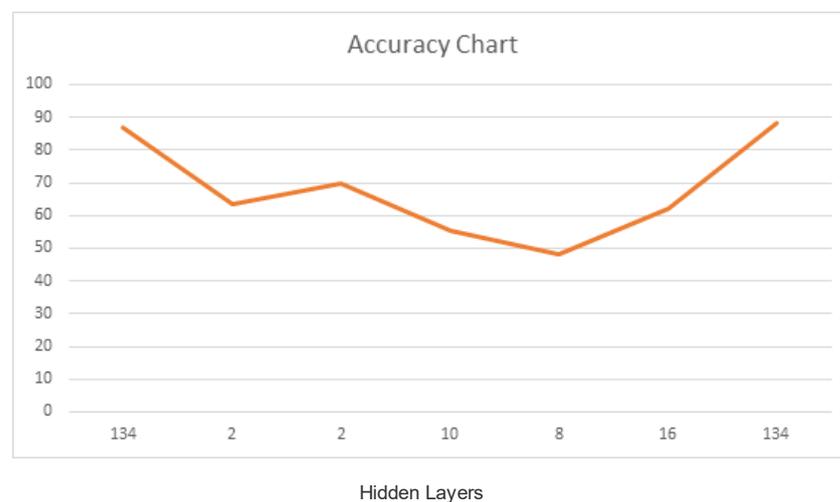


Figure 4. Proposed AD Prediction Results in Terms of Accuracy Value.

Table 1 indisputably depicts the yield of results gained using different mixes of characteristics for hidden layers, learning rates, and energy. The related layers of the artificial neural network was used for gathering the yields from the LSTM model, which contains the expected biomarkers of the patient from check-up to a year and a half. Specifically, the learning rate is a configurable hyper-limit used in the readiness of neural associations that has a particular value, usually between 0.0 and 1.0. The learning rate controls how quickly the model is changed in accordance with the issue. More unassuming learning rates require additional preparation time, especially when more unobtrusive changes are made to the model at each update; however, greater learning rates achieve quick changes and require less preparation time. A learning rate that is too big can cause the model to unnecessarily combine obtrusive changes, although a learning rate that is too small can cause the connection to slow down.

Table 1. Accuracy results.

| Trials | Hidden Layers | Learning Rate | Momentum | Cross-Validation Folds | Correlation Coefficient | RMSError | Accuracy |
|--------|---------------|---------------|----------|------------------------|-------------------------|----------|----------|
| 1 | 134 | 0.3 | 0.2 | 10 | 0.8767 | 0.13 | 86.97 |
| 2 | 2 | 0.1 | 0.1 | 10 | 0.6487 | 0.37 | 63.36 |
| 3 | 2 | 0.3 | 0.2 | 5 | 0.602 | 0.40 | 70 |
| 4 | 10 | 0.3 | 0.2 | 5 | 0.5722 | 0.45 | 55.37 |
| 5 | 8 | 0.3 | 0.2 | 10 | 0.565 | 0.46 | 48 |
| 6 | 2, 4, 8, 16 | 0.3 | 0.2 | 10 | 0.6163 | 0.38 | 61.92 |
| 7 | 134 | 0.3 | 0.2 | 5 | 0.9172 | 0.12 | 88.24 |

Learning rate and force was often changed during various endeavors in order to acquire outcomes that provided the most extreme exactness when 134 hidden layers were added to the model, using a learning pace of 0.3 and energy of 0.2, while keeping cross-validation folds at 5 provided the best outcomes. On sequence 1, there were 134 hidden layers, yet the model used 10-fold cross-validations which provided a precision of 86.97%. When folds were reduced to 5, the model showed a great accuracy of 88.24%.

The root mean square error is a reliably used measure of the degree of the separation between values (with respect to both tests and individuals) expected by a model or validator and the attributes that are observed (see Figure 5). The root mean square deviation focuses on the square base of the second model, indicating the separation between the expected attributes and actual attributes or the quadratic mean of these capabilities. These deviations are called remains/residuals when the computations are performed using a staggering model for examination, and are called goofs (or presumption bungles) when dealt with out-of-test. The RMSD serves to characterize the level of variation in the model.

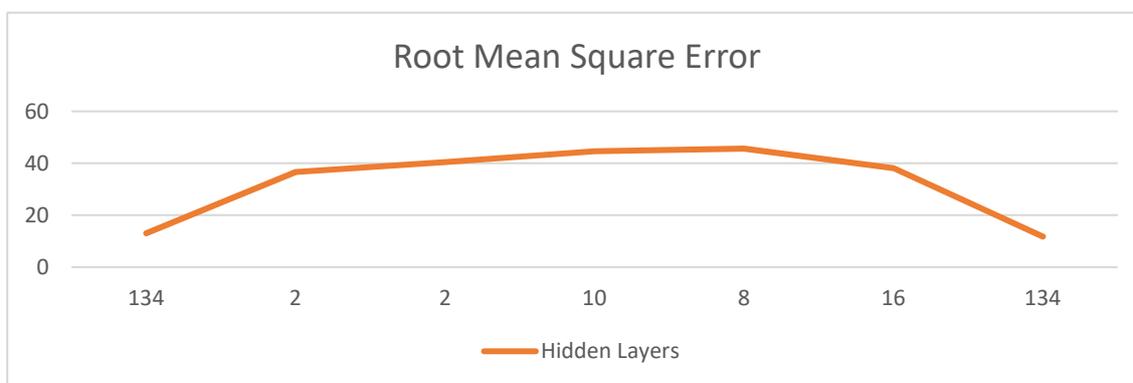
**Figure 5.** The root mean square error.

Table 2 shows the results of the RMSE when using different combinations of hidden layers and other variables. The minimum value of the root mean square error reflects the maximum accuracy and efficiency of our model. Several different values of components were incorporated to achieve the minimum error (that is, maximum accuracy) output for the model for predicting the number of patients who progressed from MCI to AD. The two closest combinations were when the number of hidden layers was 134, with cross-fold validations of 5 and 10. In the case of 10-fold cross-validation, the RMSE was almost 13%, which means the accuracy of the model was 87%; however, when the number of folds used in cross-validation decreased to 5, the RMSE decreased and was computed to be 12%, which gave sufficiently good output with an accuracy of 88.24%. Other combinations of hidden layers proved to be less effective in computing the accuracy of the system in predicting the number of patients who would progress from MCI to AD.

Table 2. The root mean square error (RMSE).

| Trials | Hidden Layers | Root Mean Square Error |
|--------|---------------------------------|------------------------|
| 1 | 134 (10-Fold Cross-Validations) | 0.13 |
| 2 | 2 | 0.37 |
| 3 | 2 | 0.40 |
| 4 | 10 | 0.45 |
| 5 | 8 | 0.46 |
| 6 | 2, 4, 8, 16 | 0.38 |
| 7 | 134 (5-Fold Cross-Validations) | 0.12 |

Correlation coefficients represent an authentic level of the strength of the link between the overall advancements of two variables (see Figure 6). Its characteristics range between -1.0 and 1.0 . A computed number higher than 1.0 or less than -1.0 suggests that there was an error in the assessment of a relationship. An association of -1.0 shows a negative relationship, while an association of 1.0 shows a positive relationship.

**Figure 6.** The correlation coefficient.

4.2. Analysis

A comprehensive analysis is formulated in this section which explains changes in the performance of each biomarker combination with different hidden layers. The proposed AD prediction architecture used two biomarkers, NM and MRI. Then, RNN-type LSTM-based features are learned and passed to a fully connected neural network layer using Python framework. The findings are tabulated in Table 1 in the form of accuracy/precision values, and attained the best accuracy of 88.24% with biomarker combinations of 4 NM and 35 MRI. Furthermore, the RMSE values are also shown in Figure 5 and Table 2, followed by the values of the obtained correlation coefficients while using different numbers of hidden layers.

In Table 1, the best-noted accuracy was 88.24%; other computed measures such as recall rate was 88.16, the prevision rate was 88.64, the F1-Score was 88.39, and the FNR value was 11.84%. The AUC was also computed with a value of 0.92. These values are plotted in Figure 7.

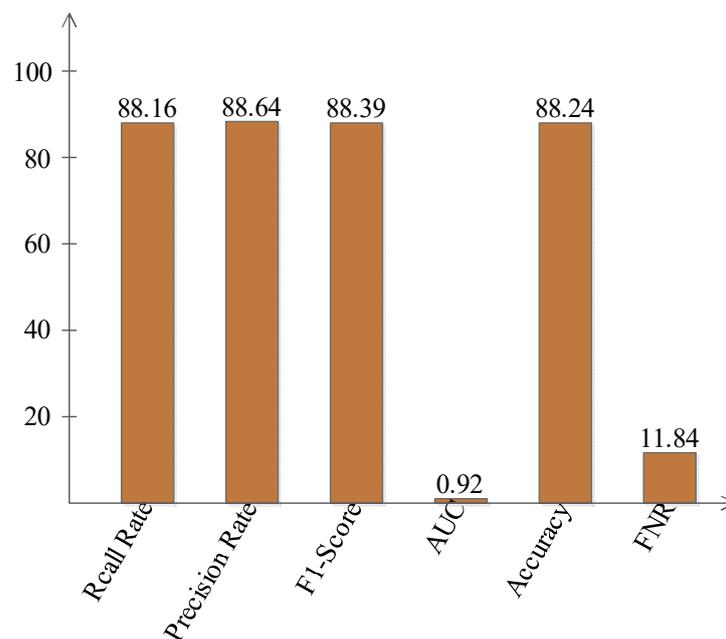


Figure 7. The performance measures computed for the best accuracy value.

Table 3 shows a clear comparison of the results computed through different algorithms and techniques by other authors by using mild cognitive impairment (MCI) samples. In [43], the authors presented a method for AD prediction and achieved an accuracy of 79%. In [44], the researchers achieved an accuracy of 73.95% using an MRI + NM biomarker approach. In [45], the authors used both NM and MRI biomarkers for predictions in their model and achieved an accuracy of 86.6%.

Table 3. A comparison of results between existing techniques.

| Results Comparison | | | | |
|-----------------------------|------------|---------------------|------------------|------------------------|
| Author | Biomarkers | Sample Size | Duration (Years) | Accuracy/Precision (%) |
| Minhas et al. (2017) [42] | NM & MRI | 54 MCIp & 65 MCIs | 2 | 84.29 |
| Minhas et al. (2017) [42] | NM | 37 MCIp & 65 MCIs | 3 | 83.26 |
| Arco et al. (2016) [44] | MRI & NM | 73 MCIp & 61 MCIs | 1 | 73.95 |
| Albright et al. (2019) [45] | NM & MRI | 110 Patients | 2 | 86.6 |
| Our Results | NM & MRI | 167 MCIp & 100 MCIs | 3 | 88.24 |

In [46], the authors created a model using a dataset of 320 patients with 2-year follow-ups and achieved an accuracy of 80.1%. In the proposed approach, we utilize MRI and NM biomarkers, and the model learned using RNN-LSTM, then evaluated the data via a fully connected NN layer for classification. The quantity of hidden layers used for this model was 134, with a learning rate of 0.3 and momentum set as 0.2, with 5-fold cross-validation and a correlation coefficient of 0.9172, and achieved an accuracy of 88.24% with the root mean square error as 0.117, which is an improvement compared to these previous techniques.

5. Conclusions

The current experiment aimed to research the degree to which it is feasible to predict a patient's progression from MCI to AD, and this current models comparison with previous models. The current investigation has demonstrated that the recurrent neural network (LSTM) performed better than other previously utilized classifiers. The most fascinating finding in the current investigation was that the model can anticipate illness progression. However, the model could be improved if more extensive follow-up times indicating

progression were used. Despite the fact that the model lacks down-to-earth applications, the current investigation makes a few essential commitments to the field of AD expectation. This is a key experiment that has used a multi-layer perceptron to predict illness progression. The current examination offers some knowledge into the significance of choosing the correct follow-ups, which adds to progress in discovering better models to use in the future. Improved exactness rates accomplished by using an RNN in this experiment demonstrates that these are truly outstanding and precise predictions of progression from MCI to AD. In the future, dynamic graph convolution [47] and multi-view feature learning [48] techniques shall be consider.

Author Contributions: This work was carried out in collaboration with all authors. A.A., A.H., M.A.K. and O.T. conceived the main idea and contributions for this study, and supervised the work. Methodology, A.A., A.H., A.M. and O.T.; validation, A.A., M.A.K., S.R., U.T., S.K. and A.M.; writing—review & editing, A.A., A.H., M.A.K. and O.T.; data correction, S.R. and U.T.; writing—first draft preparation, A.A., A.H., S.K. and O.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research work was partially supported by Chiang Mai University.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

| | |
|------|------------------------------|
| ML | Machine Learning |
| AD | Alzheimer’s Disease |
| NM | Neuropsychological Measures |
| MRI | Magnetic Resonance Imaging |
| RNN | Recurrent Neural Network |
| LSTM | Long Short-Term Memory |
| MCI | Mild Cognitive Impairment |
| CNN | Convolutional Neural Network |
| MLP | Multi-Layer Perceptron |
| ANN | Artificial Neural Network |

References

1. Alzheimer’s Association. 2019 Alzheimer’s disease facts and figures. *Alzheimer’s Dement.* **2019**, *15*, 321–387. [[CrossRef](#)]
2. Zhan, L.; Zhou, J.; Wang, Y.; Jin, Y.; Jahanshad, N.; Prasad, G.; Nir, T.M.; Leonardo, C.D.; Ye, J.; Thompson, P.M. Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer’s disease. *Front. Aging Neurosci.* **2015**, *7*, 48. [[CrossRef](#)] [[PubMed](#)]
3. Li, F.; Liu, M.; Alzheimer’s Disease Neuroimaging Initiative. A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer’s disease. *J. Neurosci. Methods* **2019**, *323*, 108–118. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, X.; Yao, L.; Wang, X.; Monaghan, J.; Mcalpine, D.; Zhang, Y. A survey on deep learning-based non-invasive brain signals: Recent advances and new frontiers. *J. Neural Eng.* **2021**, *18*, 031002. [[CrossRef](#)] [[PubMed](#)]
5. Wee, C.-Y.; Yap, P.-T.; Zhang, D.; Denny, K.; Browndyke, J.N.; Potter, G.G.; Welsh-Bohmer, K.A.; Wang, L.; Shen, D. Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* **2012**, *59*, 2045–2056. [[CrossRef](#)] [[PubMed](#)]
6. Sharif, M.I.; Alhussein, M.; Aurangzeb, K.; Raza, M. A decision support system for multimodal brain tumor classification using deep learning. *Complex Intell. Syst.* **2021**, 2420. [[CrossRef](#)]
7. Hussain, U.N.; Khan, M.A.; Lali, I.U.; Javed, K.; Ashraf, I.; Tariq, J.; Ali, H.; Din, A. A Unified design of ACO and skewness based brain tumor segmentation and classification from MRI scans. *J. Control. Eng. Appl. Inform.* **2020**, *22*, 43–55.
8. Sharif, M.I.; Li, J.P.; Khan, M.A.; Saleem, M.A. Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images. *Pattern Recognit. Lett.* **2020**, *129*, 181–189. [[CrossRef](#)]

9. Nazar, U.; Khan, M.A.; Lali, I.U.; Lin, H.; Ali, H.; Ashraf, I.; Tariq, J. Review of automated computerized methods for brain tumor segmentation and classification. *Curr. Med. Imaging* **2020**, *16*, 823–834. [[CrossRef](#)]
10. Khan, M.A.; Rubab, S.; Kashif, A.; Sharif, M.I.; Muhammad, N.; Shah, J.H.; Zhang, Y.-D.; Satapathy, S.C. Lungs cancer classification from CT images: An integrated design of contrast based classical features fusion and selection. *Pattern Recognit. Lett.* **2020**, *129*, 77–85. [[CrossRef](#)]
11. Stebbins, G.; Murphy, C. Diffusion tensor imaging in Alzheimer’s disease and mild cognitive impairment. *Behav. Neurol.* **2009**, *21*, 39–49. [[CrossRef](#)]
12. Khan, M.A.; Ashraf, I.; Alhaisoni, M.; Damaševičius, R.; Scherer, R.; Rehman, A.; Bukhari, S.A.C. Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. *Diagnostics* **2020**, *10*, 565. [[CrossRef](#)]
13. Zhang, D.; Shen, D.; Alzheimer’s Disease Neuroimaging Initiative. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease. *NeuroImage* **2012**, *59*, 895–907. [[CrossRef](#)]
14. Khan, M.A.; Hussain, N.; Majid, A.; Alhaisoni, M.; Bukhari, S.A.C.; Kadry, S.; Nam, Y.; Zhang, Y.D. Classification of positive COVID-19 CT scans using deep learning. *Comput. Mater. Contin.* **2021**, *66*, 2923–2938. [[CrossRef](#)]
15. Khan, M.A.; Kadry, S.; Zhang, Y.-D.; Akram, T.; Sharif, M.; Rehman, A.; Saba, T. Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine. *Comput. Electr. Eng.* **2021**, *90*, 106960. [[CrossRef](#)]
16. Lawrence, E.; Vegvari, C.; Ower, A.; Hadjichrysanthou, C.; De Wolf, F.; Anderson, R.M. A systematic review of longitudinal studies which measure Alzheimer’s disease biomarkers. *J. Alzheimer’s Dis.* **2017**, *59*, 1359–1379. [[CrossRef](#)]
17. Rehman, A.; Khan, M.A.; Saba, T.; Mehmood, Z.; Tariq, U.; Ayesha, N. Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture. *Microsc. Res. Tech.* **2021**, *84*, 133–149. [[CrossRef](#)]
18. Franzmeier, N.; Koutsouleris, N.; Benzinger, T.; Goate, A.; Karch, C.M.; Fagan, A.M.; McDade, E.; Duering, M.; Dichgans, M.; Levin, J. Predicting sporadic Alzheimer’s disease progression via inherited Alzheimer’s disease-informed machine-learning. *Alzheimer’s Dement.* **2020**, *16*, 501–511. [[CrossRef](#)]
19. Militello, C.; Rundo, L.; Dimarco, M.; Orlando, A.; Conti, V.; Woitek, R.; D’Angelo, I.; Bartolotta, T.V.; Russo, G. Semi-automated and interactive segmentation of contrast-enhancing masses on breast DCE-MRI using spatial fuzzy clustering. *Biomed. Signal Process. Control.* **2022**, *71*, 103113. [[CrossRef](#)]
20. Rundo, L.; Han, C.; Zhang, J.; Hataya, R.; Nagano, Y.; Militello, C.; Ferretti, C.; Nobile, M.S.; Tangherloni, A.; Gilardi, M.C. CNN-based prostate zonal segmentation on T2-weighted MR images: A cross-dataset study. In *Neural Approaches to Dynamics of Signal Exchanges*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 269–280.
21. Rundo, L.; Militello, C.; Vitabile, S.; Russo, G.; Sala, E.; Gilardi, M.C. A survey on nature-inspired medical image analysis: A step further in biomedical data integration. *Fundam. Inform.* **2020**, *171*, 345–365. [[CrossRef](#)]
22. Zemouri, R.; Zerhouni, N.; Racoceanu, D. Deep learning in the biomedical applications: Recent and future status. *Appl. Sci.* **2019**, *9*, 1526. [[CrossRef](#)]
23. Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; Albuquerque, V. Multi-Class Skin Lesion Detection and Classification via Tele dermatology. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 4267–4275. [[CrossRef](#)]
24. Khan, M.A.; Zhang, Y.-D.; Sharif, M.; Akram, T. Pixels to classes: Intelligent learning framework for multiclass skin lesion localization and classification. *Comput. Electr. Eng.* **2021**, *90*, 106956. [[CrossRef](#)]
25. Kawahara, J.; BenTaieb, A.; Hamarneh, G. Deep features to classify skin lesions. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 1397–1400.
26. Khan, M.A.; Lali, I.U.; Rehman, A.; Ishaq, M.; Sharif, M.; Saba, T.; Zahoor, S.; Akram, T. Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. *Microsc. Res. Tech.* **2019**, *82*, 909–922. [[CrossRef](#)]
27. Khan, M.-A.; Majid, A.; Hussain, N.; Alhaisoni, M.; Zhang, Y.-D.; Kadry, S.; Nam, Y. Multiclass Stomach Diseases Classification Using Deep Learning Features Optimization. *Comput. Mater. Contin.* **2021**, *67*, 3381–3399. [[CrossRef](#)]
28. Rauf, H.T.; Lali, M.I.U.; Khan, M.A.; Kadry, S.; Alolaiyan, H.; Razaq, A.; Irfan, R. Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks. *Pers. Ubiquitous Comput.* **2021**, 2737. [[CrossRef](#)]
29. Bai, X.; Yang, M.; Huang, T.; Dou, Z.; Yu, R.; Xu, Y. Deep-person: Learning discriminative deep features for person re-identification. *Pattern Recognit.* **2020**, *98*, 107036. [[CrossRef](#)]
30. Khan, M.A.; Akram, T.; Zhang, Y.-D.; Sharif, M. Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recognit. Lett.* **2021**, *143*, 58–66. [[CrossRef](#)]
31. Varga, D. No-Reference Image Quality Assessment with Convolutional Neural Networks and Decision Fusion. *Appl. Sci.* **2022**, *12*, 101. [[CrossRef](#)]
32. Zemouri, R.; Racoceanu, D. Innovative deep learning approach for biomedical data instantiation and visualization. In *Deep Learning for Biomedical Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 171–196.
33. Baltres, A.; Al Masry, Z.; Zemouri, R.; Valmary-Degano, S.; Arnould, L.; Zerhouni, N.; Devalland, C. Prediction of Oncotype DX recurrence score using deep multi-layer perceptrons in estrogen receptor-positive, HER2-negative breast cancer. *Breast Cancer* **2020**, *27*, 1007–1016. [[CrossRef](#)]

34. Zemouri, R.; Omri, N.; Devalland, C.; Arnould, L.; Morello, B.; Zerhouni, N.; Fnaiech, F. Breast cancer diagnosis based on joint variable selection and constructive deep neural network. In Proceedings of the 2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME), Tunis, Tunisia, 28–30 March 2018; pp. 159–164.
35. Nawaz, M.; Nazir, T.; Javed, A.; Tariq, U.; Yong, H.-S.; Khan, M.A.; Cha, J. An Efficient Deep Learning Approach to Automatic Glaucoma Detection Using Optic Disc and Optic Cup Localization. *Sensors* **2022**, *22*, 434. [[CrossRef](#)]
36. Syed, H.H.; Khan, M.A.; Tariq, U.; Armghan, A.; Alenezi, F.; Khan, J.A.; Rho, S.; Kadry, S.; Rajinikanth, V. A Rapid Artificial Intelligence-Based Computer-Aided Diagnosis System for COVID-19 Classification from CT Images. *Behav. Neurol.* **2021**, *2021*, 2560388. [[CrossRef](#)]
37. Zemouri, R.; Omri, N.; Morello, B.; Devalland, C.; Arnould, L.; Zerhouni, N.; Fnaiech, F. Constructive deep neural network for breast cancer diagnosis. *IFAC-PapersOnLine* **2018**, *51*, 98–103. [[CrossRef](#)]
38. Basheera, S.; Ram, M.S.S. Convolution neural network-based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **2019**, *5*, 974–986. [[CrossRef](#)]
39. Basheera, S.; Ram, M.S.S. A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI. *Comput. Med. Imaging Graph.* **2020**, *81*, 101713. [[CrossRef](#)]
40. Tabarestani, S.; Aghili, M.; Eslami, M.; Cabrerizo, M.; Barreto, A.; Rishe, N.; Curiel, R.E.; Loewenstein, D.; Duara, R.; Adjouadi, M. A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. *NeuroImage* **2020**, *206*, 116317. [[CrossRef](#)]
41. Lei, B.; Yang, M.; Yang, P.; Zhou, F.; Hou, W.; Zou, W.; Li, X.; Wang, T.; Xiao, X.; Wang, S. Deep and joint learning of longitudinal data for Alzheimer's disease prediction. *Pattern Recognit.* **2020**, *102*, 107247. [[CrossRef](#)]
42. Minhas, S.; Khanum, A.; Riaz, F.; Khan, S.A.; Alvi, A. Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 818–825. [[CrossRef](#)]
43. Gomar, J.J.; Bobes-Bascaran, M.T.; Conejero-Goldberg, C.; Davies, P.; Goldberg, T.E.; Alzheimer's Disease Neuroimaging Initiative. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. *Arch. Gen. Psychiatry* **2011**, *68*, 961–969. [[CrossRef](#)]
44. Arco, J.E.; Ramírez, J.; Górriz, J.M.; Puntonet, C.G.; Ruz, M. Short-term prediction of MCI to AD conversion based on longitudinal MRI analysis and neuropsychological tests. In *Innovation in Medicine and Healthcare 2015*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 385–394.
45. Albright, J.; Alzheimer's Disease Neuroimaging Initiative. Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **2019**, *5*, 483–491. [[CrossRef](#)]
46. Gomar, J.J.; Conejero-Goldberg, C.; Davies, P.; Goldberg, T.E.; Alzheimer's Disease Neuroimaging Initiative. Extension and refinement of the predictive value of different classes of markers in ADNI: Four-year follow-up data. *Alzheimer's Dement.* **2014**, *10*, 704–712. [[CrossRef](#)]
47. Zhao, K.; Duka, B.; Xie, H.; Oathes, D.J.; Calhoun, V.; Zhang, Y. A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. *NeuroImage* **2022**, *246*, 118774. [[CrossRef](#)]
48. Zhang, Y.; Zhang, H.; Adeli, E.; Chen, X.; Liu, M.; Shen, D. Multiview feature learning with multiatlas-based functional connectivity networks for MCI diagnosis. *IEEE Trans. Cybern.* **2020**, 1–12. [[CrossRef](#)]