Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

An Extensive Survey of Various Deep Learning Approaches for Predicting Alzheimer's Disease

Nithya V. P, *Mohanasundaram. N,
Research Scholar, Professor & Head, Department of Computer Science and Engineering,
Karpagam Academy of Higher Education,
Coimbatore-641 021

Abstract- The mild cognitive impairments or Alzheimer's disease (AD) prediction has captured the attention of various researchers due to the rapid growth of the disease and the necessity to treat AD in the earlier phase. However, due to the lack of high dimensionality neural data and the availability of lesser number of samples leads to the need for the better predictor model. The advancements in Artificial Intelligence (AI) paved the way for substantial growth of machine learning and deep learning approaches. The significance of deep learning approaches is comparatively higher than that of machine learning approaches in terms of prediction accuracy. Recently, deep learning is considered as the essential tool for predicting AD. Initially, the essential data have to be acquired from the proper source to perform the validation process. The features related with the disease progression needs to be analyzed and learned to reduce the dimensionality and computational overhead. Finally, the classification is performed with learning approaches to improve the prediction accuracy and works efficiently than the prevailing approaches. The anticipated model diminishes the need for human effort and provides easier way for intellectually diagnose the disease. Various experimentations are carried out to show the significance of diverse algorithm. The advantages and the disadvantages related to the prediction model is analyzed in an extensive manner.

Keywords- Alzheimer disease, machine learning, deep learning, dimensionality reduction, computational overhead

1. Introduction

Alzheimer disease (AD) is an irreversible neuro-degenerative disorder eliminates memory and causes complexity over communication gradually and performs some regular activities like walking and speaking [1]. Eventually, it is fatal and the common dementia type which is composed of 70% to 80% of dementia cases. Typically, it is identified for the middle or old-aged people which are started with the protein accumulation and leads to constant memory deterioration (related with cell death, brain shrinkage, and synaptic dysfunction) [2]. The preliminary brain variations are occurred during the initiation of the cognitive decline, and certain biomarkers are considered to be abnormal in the earlier stages [3]. Various researchers recommend some solutions for the brain related AD and initiate the consequences before the appearance of the symptoms.

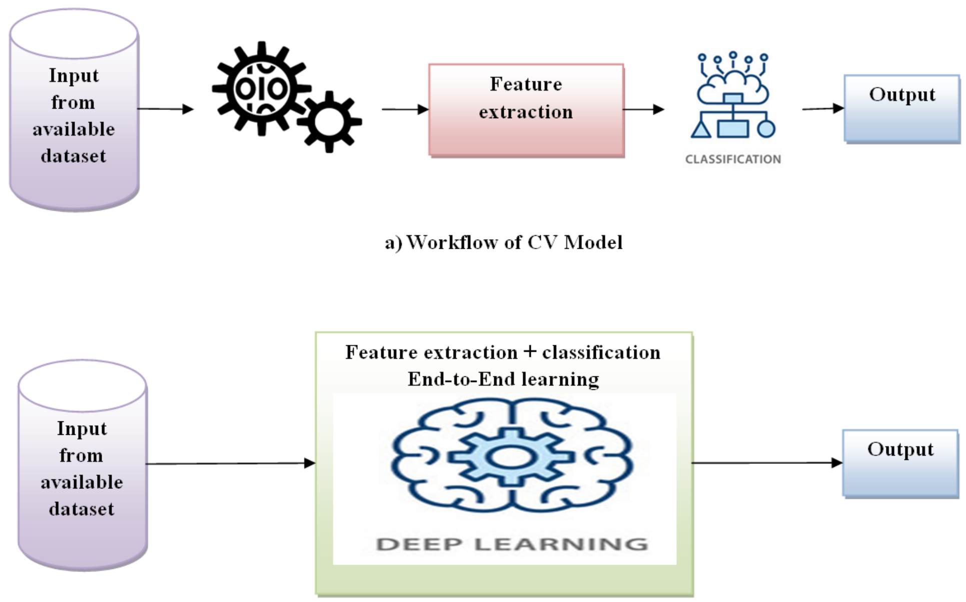
The initial stages of the disease symptoms of AD are categorized with the assumption of Mild Cognitive Impairment (MCI); however, not all the symptoms identified leads to the development of AD [4]. MCI is depicted as the transitional phase from normal to AD, where the individuals have mild variations in cognitive functionalities are identified over the person influenced and related persons. However, it is competent to carry out everyday activities. Roughly, 20% to 25% of people are aged with 65 or older than MCI. Similarly, 40%-45% of persons with MCI generate AD with 5 years [5]. The time range conversion from 5 to 35 months; however, it is roughly 18 months. Then, the patients are classified as MCI non-

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

converters or convertors that show patients and patients' without conversion of AD with 18 months. Also, there are some other MCI sub-types that are mentioned in some cases like late/earlier MCI [6].

The most essential risk factors are family histories and occurrence of genes-associated with individuals' genome. Some AD diagnosis relies on clinical factor along with comprehensive analysis of patients and relatives [7]. However, some ground truth AD diagnosis is performed through autopsy is not clinically essential. Some AD patients are confirmed with autopsy prediction. Devoid of any ground truth data, some patients require certain factor to validate AD [8]. These criteria can be enhanced with better AD understanding and provides possible prediction for living patients. Some essential criteria for AD-based clinical diagnosis is provided by ADRDA 2 and NINCDSI. The criteria are validated based on memory impairment and with the occurrence of added supportive features: Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) based on abnormal cerebrospinal fluid amyloid, neuro-imaging, and tau biomarkers [9]. The association for Alzheimer's' disease like NIA3 have intended to re-model the diagnostic criteria using AD. The anticipated diagnostic criteria are composed of brain amyloid measurement, degeneration, and neuronal injury. Recently, it is concluded that certain criteria is updated with a warranty of 3-5 years to integrate knowledge regarding the disease progression and patho-physiology.

Mini-mental state and Clinical Dementia Rating examination are the two diverse kinds of most frequently adopted test for examining the occurrence of AD [9]. However, it has to be observed that the utilization of ground truth value for the analysis of AD may not be more appropriate. With the extensive analysis over the criteria discussed above, some reports are generated with the clinical prediction model for AD with better accuracy [10]. The prediction accuracy should be higher than the post-mortem diagnosis factors that ranges from 80% to 90% respectively. With these constraints, some clinical prediction models provide superior standard of reference models. Also, it is essential that the accessibility of pronounced bio-markers is extremely limited. Fig 1 depicts the difference among the conventional ML and model DL approaches.



b) Workflow of DL Model

Fig 1 Difference among traditional ML and modern DL approaches

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

With the consideration of the above mentioned limitations, there is some requirements for the multi-class clinical decision support model which is unbiased using variable radiological expertise. It is extensively utilized for differentiating AD and differentiating the stages from Normal Control [11]. Usually, AD patient's categorization from MCI or NC which is not so valuable for predicting the conversion of MCI as AD is obviously apparent devoid of utilizing any sort of expertise when it is not so delayed for treatment purposes. However, various studies are still adopted to handle the issues related to NC and AD problems. It is essential for diverse classification tasks, specifically for better understanding towards the analysis of AD signs [12]. The foremost essential and the major challenge related with AD evaluation are to measure the functionality of MCI and to identify the prediction model for disease development. However, the accessibility of computer-aided systems are not probable for replacing the medical experts. It can only assist in provisioning of essential information for enhancing the prediction accuracy of clinical decision systems. It is observed that, not all the investigations are suitable for predicting AD. Some other stages of disease are earlier or late for of MCI. Usually, prediction of AD using learning approaches are more challenging with the below given factors [13].

- ✓ Errors during pre-processing and acquisition of least quality image acquisition.
- ✓ Unavailability of standard dataset which is composed of huge among of bio-markers and subjects.
- ✓ Some signs are used for distinguishing AD, for instance, shrinkage of brain and it is determined for normal healthy brain of some elder people.
- ✓ Lower differences among the class variance.
- ✓ The ambiguity boundaries ranges from MCI/NC, MCI/AD based AD diagnostic criteria
- ✓ The complexities over medical images are compared to usual images.

This survey provides an extensive analysis over the deep learning classifier models for predicting AD with superior prediction accuracy, reduced computational complexities, and over-fitting issues.

2. Prologue on deep learning approaches

In AI, Machine Learning is considered as a subset of AI that shows complexity relationship among data variables. The superiority of ML commences from the competency to derive predictive model from the huge amount of available data with least cases, i.e. completely devoid of any prior information regarding the data. Some of the most commonly used ML approaches are Artificial Neural Networks (ANN) which draws the attention from biological NN. ANN functionality relies on perceptron which partitions the input data with proper classes [14]. The network model is hierarchically composed of modern deep NN, i.e. multi-layer perceptron classifier model. Some forms of traditional approaches are NB, SVM, LR and LR. These conventional ML approaches are considered for its significance in genomics and medicine [15]. It is observed that the search is significant with linear regression over huge dataset. Even though these tools are efficiently considered for parsing the huge dataset and predicting the relationship among the variables and these conventional learning approaches are considered for manual feature engineering and it suffers from certain overheads and constraints over the scenario that is needed for real-time decision making process.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

This paves the way for the design of deep learning (DL) approaches. DL approaches varies from conventional ML approaches during data representation and it automatically predicts the raw data. In case of shallow feature identification, DL approaches adopts deep and multiple layers for capturing both the higher and lower level data by facilitating the learning of rich input abstraction [16]. The need for manual feature engineering facilitates DL models for uncovering unknown patterns and standardizes the data. Over the DL approaches, Convolutional Neural Networks (CNN) have acquired the attention of the researchers with image-based and computer vision-based medical research. It acquires various data representation from multiple layers where it learns certain image features like visual cortex and arrange it in hierarchical manner [17]. It is composed diverse convolutional layers where the features are learned using pooling layers. It helps in feature reduction and thus requires computational demand by gathering redundant features. The dropout layers are used for selectively turn off the perceptrons to eliminate the reliance of single network components. The output layer collates some features with proper decision, i.e. class 1 or 0. The algorithm needs to be profoundly achieving success with image classification. Fig 2 depicts the flow of diverse DL approaches.

Recurrent Neural Networks (RNN) and the corresponding variants like GRU and LSTM are used for analyzing the time-series data. These are analyzed sequentially using the input data and adopts gated method for demonstrating and discarding the information with prior elements by generating proper output. The long-term dependencies are adopted over speech processing, machine translation and text analysis.

Auto-encoder adopts the class of unsupervised learning approaches that predicts the data representation by learning the lower-level dimensional mapping from input to output. It is composed of encoder that discusses the latent representation with input and decoder. The input is reconstructed using the latent representation. It provides the lower-level dimensionality than the input where AE learns the compressed data representation which holds the necessary features with conjunction with learning approaches [18]. Generative Adversarial Network is a newer form of algorithms that aims at providing a novel data with statistical input data using latent data distribution approximately. The algorithms are composed of two GAN: generator and discriminator. The former model is composed of synthetic data from noise by sampling from distribution approximated and the later model intends to differentiate the synthetic and real data instances [19]. The provided network models are engaged with adversarial process for provisioning the data fidelity by improving the accuracy. The outcomes are used for augmenting the prevailing dataset.

The advancements in DL approaches are extensively owing to the penetration of computational ability and resourcing the open-source data. The DL-based application adopts GPU for accelerating the large sized data and to get rid of algorithm complexity along with the constant reduction of training time [20]. The outcomes are measured with the higher-throughput with faster experimentation and facilitating better prediction efficacy. Some open-source framework includes Kera, TensorFlow, Caffe, PyTorch and other those are provided with increased accessibility and facilitate the data sharing over domains. The DL model-based research provides surprising outcomes and provides variations in medical research landscape. Table 1 depicts the comparison of various DL approaches.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

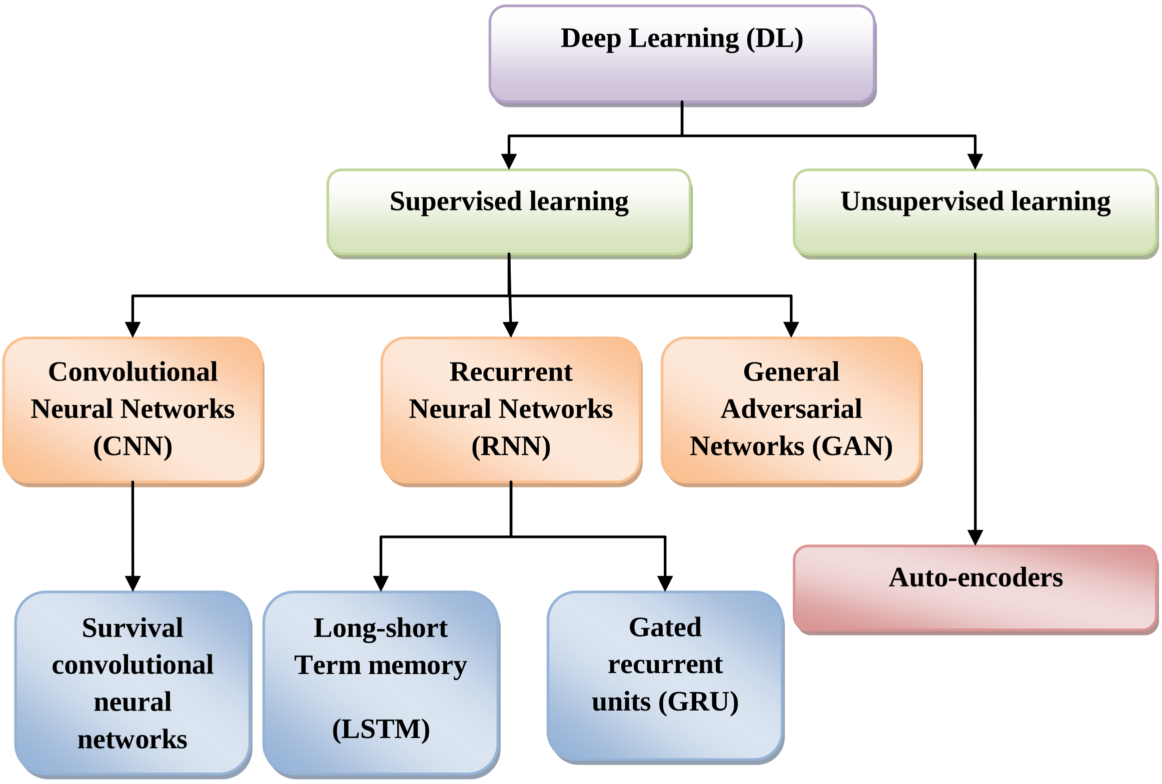


Fig 3 Flow diagram of various DL approaches

Table 1 Various DL algorithms [20]

S. No.	Algorithms	Description
1	**Deep Neural Networks (DNN)**	1. It is a simple algorithm with more than two hidden layers. 2. Adopted for classification and regression-based applications.
2	**Convolutional Neural Networks (CNN)**	1. It works effectually for image-based applications. 2. This network model is extremely efficient for 2D data.
3	**Recurrent Neural Networks (RNN)**	1. This network model is applied for sequence format data 2. Network weights are shared among the network nodes
4	**Deep Belief Networks (DBN)**	1. It is significantly needed for unsupervised and supervised learning models. 2. Hidden layer of sub-networks are available for next sub-network
5	**Deep Auto-encoders (DA)**	1. It is applied for image-based dimensionality reduction 2. Input and output size is the same. 3. It is a supervised learning algorithm
6	**Deep Boltzmann Machine (DBM)**	1. It works in a uni-directional manner like Boltzmann's family.

		2. It is an extended version of RNN
7	**Deep Convolutional Extreme Learning Machine**	1. This network model is used for sampling local connections and it is applied with a Gaussian probability function.

3. Discussion on various other DL architectures

In DL modelling, overfitting plays a pre-dominant role in injecting model complexities which has to be resolved using various architectural levels. Author in [20] discusses a model known as Restricted Boltzmann Machine (RBM) to deal with the over-fitting issues. The stacking of RBM outcomes in superior structural model know as Deep Boltzmann Machine. Similarly, a supervised model known as DBN is adopted to connect diverse unsupervised features by data extraction from all the stacked layers. It is considered to provide superior performance than other approaches which is one among the primary cause of using DL approach to gain popularity. It handles over-fitting issues by diminishing the initialization of weight with RBM. Subsequently, CNN model diminishes the number of parameters efficiently with the insertion of pooling and convolution layers and leads to the complexity reduction. Due to the significance, CNN is extensively adopted over visual recognition field. Author in [20] discusses a unsupervised learning model known as Auto-Encoders (AE) to produce the valuable outcomes with the approximation of input values using SGD and back-propagation. It deals with dimensionality reduction; however, it is extremely complex to train the samples owing to the disappearing nature of gradient issues. Similarly, sparse AE is used for handling these issues by facilitating smaller amount of hidden nodes. The stacked AE is stacks over the sparse AE alike of DBN. Some models like stacked AE, sparse AE, AE, DBM, RBM, and DNN-based deep learning approaches are used for predicting AD. These models are utilized for classifying AD patients from MCI or NC which is considered as the prodromal stages of AD. These approaches are utilized to identify the MCI transformation to AD with multi-model neuro-imaging data. Here, various DL approaches are utilized along with conventional ML approaches.

a. Feed Forward Neural Network (FFNN)

FFNN comprises of various hidden layers among the input and output layers. Every layer is composed of multiple units with fully connected to neighbourhood layers [21], however there are no relationship among the units over the layer. The input vector $'x'$,, the output unit with composition function $'y_k'$, and it is mathematically expressed as in Eq. (1):

$$y_k(x,\theta) = f^{(2)}\left(\sum_{j=1}^{M} w_{k,j}^{(2)} f^{(1)}\left(\sum_{i=1}^{N} w_{j,i}^{(1)} x_i + b_j^{(1)}\right) + b_k^{(2)}\right) \quad (1)$$

Here, $'M'$ specifies number of hidden units, $'b_j'$ and $'b_k'$ specifies input and hidden layer bias, $f^{(1)}(.)$ and $f^{(2)}(.)$ specifies non-linear activation function and $\theta = \left\{w_j^{(1)}, w_k^{(2)}, b_j^{(1)}, b_k^{(2)}\right\}$. It is an efficient approach to compute the FFNN gradients (See Fig 4). It is used to proliferate the error values from output to input layer with updated θ.

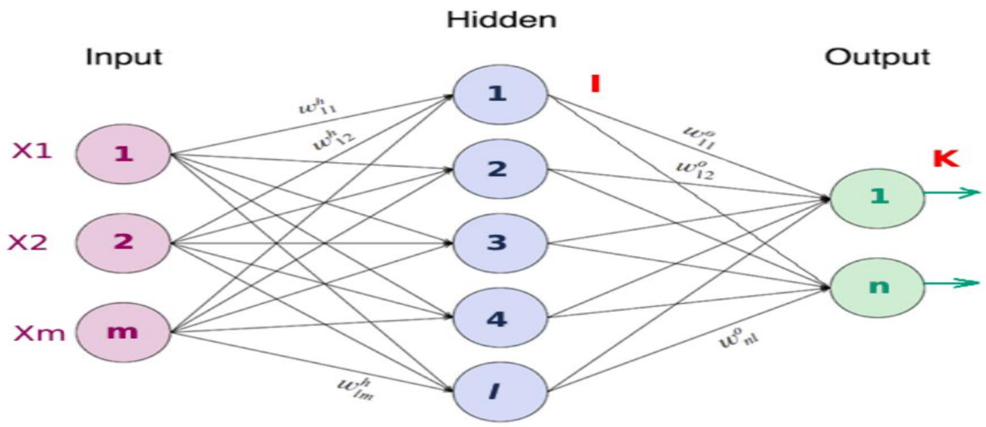


Fig 4 FFNN model

b. Stacked Auto-encoders

Auto-encoder is termed as auto-associator for learning latent representation with input data and utilizes certain representation for output data reconstruction. With simple structural model, the stacked AE representation is completely restrictive [22]. Moreover, when diverse AEs are stacked to generate DN model, it is known as stacked auto –encoders. Due to the structural features, the proposed model is capable of learning and discovering complex patterns with inherent input data. The lower-level layers are simpler with data patterns while the higher-level layers are competent to haul-out complex data patterns. There are diverse AE models like variational AE, sparse AE, denoising AE are anticipated and stacked over SAE. The improved version of AE not only provides essential latent representation; however, it enhances the robustness (See Fig 5). To handle the disadvantages of prevailing approaches are the falling of gradient values to poor local optimum. The significant benefits of pre-training are competent to improve the training dataset size over the unlabelled samples.

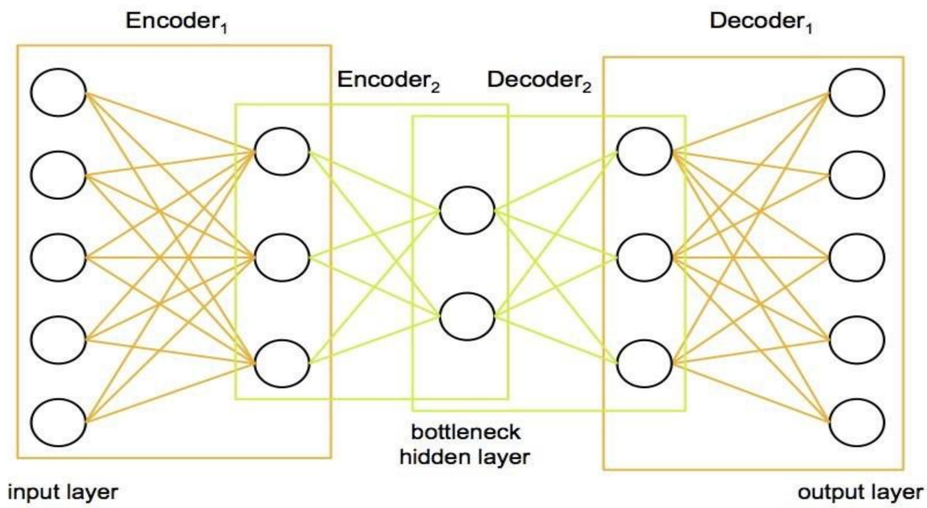


Fig 5 Auto-encoder model

c. Deep Convolutional Neural Networks (DCNN)

DCNN uses kernel function (for filtering), to identify features (image edges) of the given input image. The kernel is a matrix of weights that are trained for predicting certain features [23]. The concept behind CNN is to convolve kernel spatially over the provided input image and validate whether the features are meant for prediction. The convolutional function is performed by evaluating the kernel dot products and input where kernels are overlapped (See

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

Fig 6). The learning process of kernel weight is facilitated using the convolution layer output and summed with both bias term and non-linear activation function. Generally, Activation Functions (AF) are non-linear functions such as ReLU (Rectified Linear Unit), TanH, and Sigmoid. Based on the data nature and classification tasks, AFs are chosen. For instance, ReLU is represented with neurons over the brain. Figure 5.4 depicts the generic view of the CNN model.

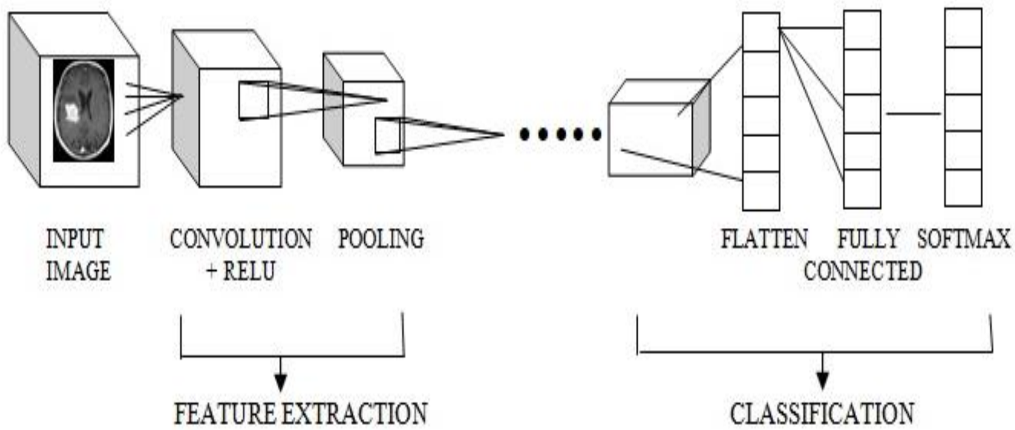


Fig 6 Deep Convolutional Neural Networks (DCNN) architecture

The training process is quickened, and the total memory consumed is reduced over the network. Often, the pooling layer follows the convolutional layer for eliminating redundant data from the input feature. For instance, the max-pooling layer moves the window from input to output to reduce the image pixel's maximal value. Deep CNN possesses various convolutional and pooling layer. Finally, Fully Connected layers (FCL) flattens previous layer's volume, and the output layer evaluates the score (probability and confidence) for output features and classes throughout the network. The output is fed to regression functions like softmax as it maps the vector values. This model provides efficient outcomes with magnificent results for image prediction tasks.

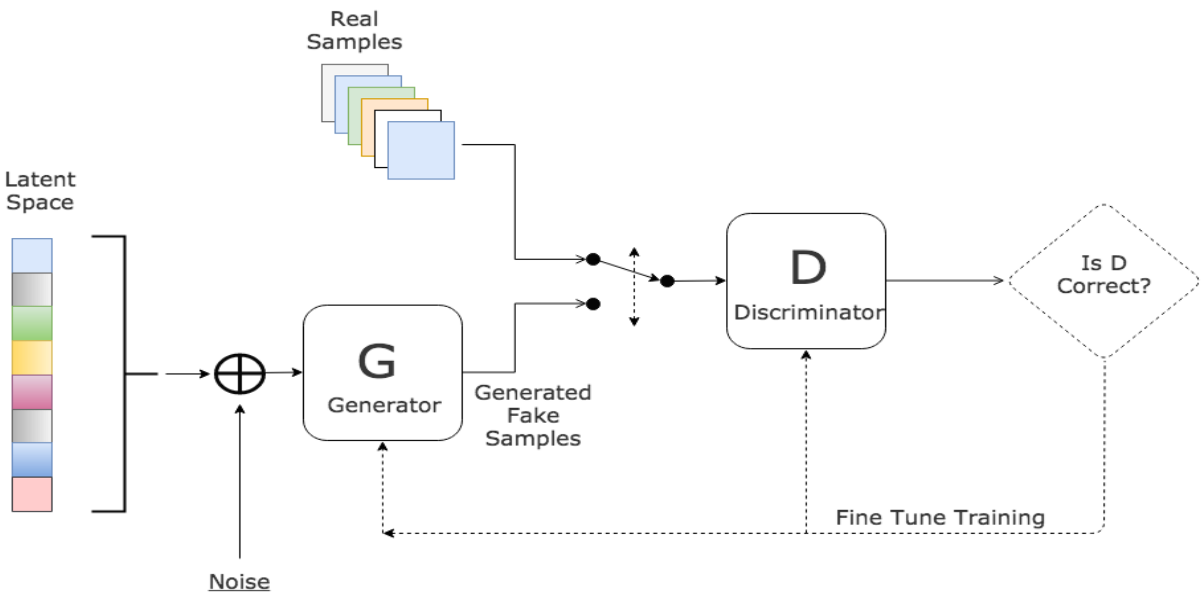


Fig 7 Generative Adversarial Networks (GAN)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

d. Generative Adversarial Networks (GAN)

GAN has attained the interest of various researchers in natural language processing and computer vision. It is composed of two NN models known as generator and discriminators [24]. GAN concurrently trains the discriminator and generator where the generator intends to produce realistic data for fooling the provided discriminator. However, discriminator intends to differentiate fake and original samples. It is mathematically expressed as in Eq. (2):

$$\min_{G}\max_{D} V(G,D) = E_{x\sim p_{data}}(x)[\log D(x)] + E\_(z\sim p_z)\,(z)\,[\log(1 - D(G(z)))] \tag{2}$$

Here, $p_{data}\,(x)$ specifies real data distribution. After simultaneous training, the generator and discriminator possess essential capacity and researches certain point and cannot enhance $p_g = p_{data}$. The discriminator cannot show the difference among the generated and real samples, i.e. $D(x) = 0.5$. there are diverse challenges with related training model and computing the GAN with saddle points and model collapse respectively (See Fig 7). There are diverse GAN variants like Deep convolutional CNN and Wasserstein GAN are proposed to get rid of diverse confronts.

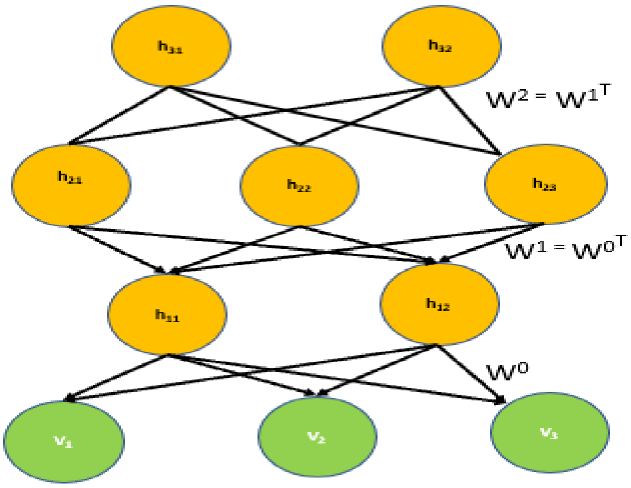


Fig 8 Deep belief Network

e. Deep belief Network (DBN)

DBN stacks over multiple RBM for the construction of deep architecture. It is composed of multiple hidden layers and one visible layer [25]. The lower-level layers produce direct generative models. Moreover, top level layers forms RBM distribution with undirected generative model (See Fig 8). Thus, the provided units $'v'$ and $'L'$ with hidden layers $h^{(1)}, h^{(2)}, .., h^{(L)}$and it is mathematically expressed as in Eq. (3):

$$P(v, h^{(1)}, \dots, h^{(L)}) = P(v|h^{(1)})\left(\prod_{l=1}^{L-2} P(h^{(1)}|h^{(l+1)}))\,P(h^{(L-1)}, h^{(L)}\right) \tag{3}$$

Here, $P(h^{(1)}|h^{(l+1)})$ specifies conditional distribution for hidden layer units and provides unit for hidden layer $l + 1$, and $P(h^{(L-1)}, h^{(L)}$ specifies joint distribution over the top hidden layers $L - 1$ and L respecctively. During the DBN training process, there are two diverse approaches like fine tuning and pre-training. During pre-training process, it is trained with

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

RBM layer for predicting parameter space. The provided layers are trained using RBM [26]. The hidden layers are trained with observation data from other representation. The training process is repeated concurrently. Fine-tuning is performed to search the optimal parameters. For practical applications, the attained parameters from pre-training step are utilized for DNN initialization and the deep model are used for fine-tuning by supervised learning models.

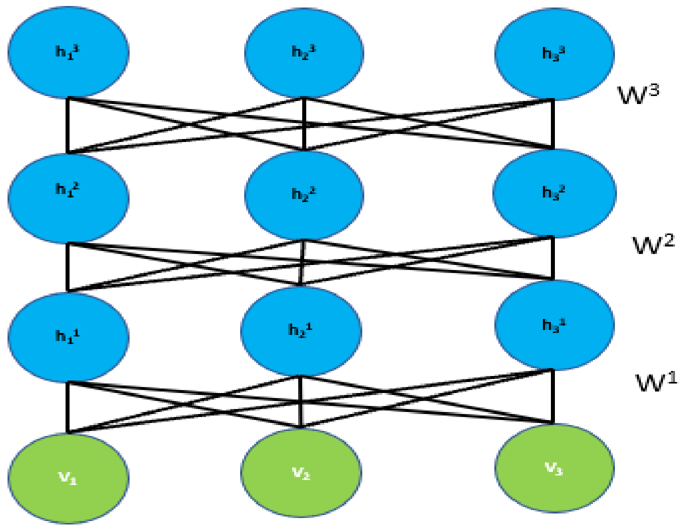


Fig 9 Deep Boltzmann Machine (DBM) model

f. Deep Boltzmann Machine (DBM)

DBM is modelled with multiple stacking RBM. Moreover, the layers of DBM are un-directed and the variables with the hidden layers are independent mutually [27]. Therefore, the hidden layers are connected with corresponding neighbourhood layers like $l-1$ and $l+1$respectively where the probability distribution is based on $P(h^{(1)}|h^{(l-1)}, h^{(l+1)})$ respectively (See Fig 9). When the neighbourhood layers are provided, conditional probabilities are visible and hidden layers are provided with logistic sigmoid functions are provided as in Eq. (4), Eq. (5) and Eq. (6):

$$P(v_i|h^1) = \sigma\left(\sum_j W_{ij}^{(1)} h_j^{(1)}\right) \tag{4}$$

$$P\left(h_k^{(l)}\middle|h^{l-1}, h^{l+1}\right) = \sigma\left(\sum_m W_{mk}^{(l)} h_m^{(l-1)} + \sum_n W_{kn}^{(l+1)} h_n^{(l+1)}\right) \tag{5}$$

$$P\left(h_t^L\middle|h^{(L-1)}\right) = \sigma\left(\sum_s W_{st}^{(L)} h_s^{(L-1)}\right) \tag{6}$$

Here, $h^{(l)}$ specifies the conditional probability of hidden units, the probability integrates both upper hidden layer $h^{(l+1)}$ and lower hidden layer $h^{(l-1)}$. By integrating information of both upper and lower layers are specified using power representation which is completely robust over the observed noisy data. Moreover, the character provides DBM conditional probability $P(h^{(l)}|h^{(l-1)}, h^{(l+1)})$ is extremely complex than that of conventional model.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 848-860
Received 15 April 2021; Accepted 05 May 2021.

4. Advantages of Deep Learning

The following are the foremost advantages of using DL approaches [28]:

- ✓ **Increasing size of datasets,** the investigators want DL to work perfectly and to attain better outcomes with the larger datasets
- ✓ **Increasing model size,** with the availability of huge datasets, the researchers' major hurdle/obstacle is the storage and ability to process the enormous data. However, with Graphics Processing Unit (GPU) and Central Processing Unit (CPU) faster computation, larger memory space paved the way to work using larger datasets.
- ✓ **Increasing accuracy,** DL improves the ability to provide more appropriate outcomes, and it is used extensively in various real-time applications.

5. Research challenges

DL algorithms and medical applications show constant evolution for generating better performance outcomes like image recognition. Specifically, it works efficiently with valid inference, i.e. testing and training environment needs to be similar. The major drawback associated with this model is its complexity to change the potential bias of network when the intricacy is identified to a greater extent in terms of reproducibility and transparency [29]. These sorts of complications are resolved using the scale neuro-imaging data accumulation and analysis with the relationship among DL approaches and the features. The parameter disclosure is used for attaining the outcomes and the mean values are used for experimentation and mitigate the reproducibility issue. DL can resolve certain problem by extracting attributes from the provided input data directly devoid of any pre-processing for feature selection with complexity over various data formats like genetic data and neuro-imaging. The input data based weight adjustment is carried out over the closed network automatically where the addition of input causes ambiguity and confusion. The extensive development with DL approaches intends to resolve all these issues by provisioning problem-specific outcomes [30]. When huge amount of data is acquired, the research methodology has to be adopted with deep learning approaches to give more resourceful outcomes. The 2D CNN model is converted to 3D CNN model which deals with multi-modal neuro-images. Additionally, Generative Adversarial Network model is used for producing synthetic medical images for augmenting data. Moreover, reinforcement learning is used for changing the data and provides better decision and applicable over various medical fields.

The extensive research models on AD using DL approaches are evolving with superior transparency and performance. When the computational resources on multi-modal neuro-imaging data is evolved rapidly, the research based AD classification is transformed to advanced DL model indeed of hybridization approaches. These sorts of approaches need to need to be modelled for integrating complete data format in DL networks.

6. Conclusion

From the extensive analysis of Deep Learning over medical image processing, it is obvious that DL plays a key role in medical application. It is used for disease predictions with improved prediction accuracy. DL acts as a bridge between the learning and visualization process. It works effectually over the enormous amount of data accessible over the dataset with reduced computational cost and time. With this analysis, it is noted that DCNN plays a predominant role in disease classification.

REFERENCES

[1] Alzheimer's Association (2018). 2018 Alzheimer's disease facts and figures. Alzheimer's Dementia 14, 367–429. doi: 10.1016/j.jalz.2018.02.001

[2] Galvin, J. E. (2017). Prevention of Alzheimer's disease: lessons learned and applied. J. Am. Geriatr. Soc. 65, 2128–2133. doi: 10.1111/jgs.14997

[3] De strooper, B., and Karran, E. (2016). The cellular phase of Alzheimer's disease. Cell 164, 603–615. doi: 10.1016/j.cell.2015.12.056

[4] Liu, M., Cheng, D., Yan, W., and Alzheimer's Disease Neuroimaging Initiative. (2018a). Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front. Neuroinform. 12:35. doi: 10.3389/fninf.2018.00035

[5] Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., et al. (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's Disease. IEEE Trans. Biomed. Eng. 62, 1132–1140. doi: 10.1109/TBME.2014.23 72011

[6] Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structuralMR and FDG-PET images. Sci. Rep. 8:5697. doi: 10.1038/s41598-018-22871-z

[7] Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage 155, 530–548. doi: 10.1016/j.neuroimage.2017.03.057

[8] Samper-Gonzalez, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., et al. (2018). Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data. Neuroimage 183, 504–521. doi: 10.1016/j.neuroimage.2018.08.042

[9] Suk, H.-I., Lee, S.-W., Shen, D., and The Alzheimer's Disease Neuroimaging, I. (2015). Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct. Funct. 220, 841–859. doi: 10.1007/s00429-013-0687-3

[10] Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2019). Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's disease neuroimaging initiative. Alzheimers Dement 15, 106–152. doi: 10.1016/j.jalz.2018.08.005

[11] F. Falahati, E. Westman, and A. Simmons, "Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging," Journal of Alzheimer's Disease, vol. 41, no. 3, pp. 685-708, 2014.

[12] Kazemi and S. K. Houghten, "A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data," in Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2018 pp. 1-8.

[13] Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future," in Classification in BioApps: Springer, 2018, pp. 323-350

[14] Kazemi and S. K. Houghten, "A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data," in Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2018 pp. 1-8.

[15] Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future," in Classification in BioApps: Springer, 2018, pp. 323-350.

[16] Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," IEEE Access, vol. 6, pp. 9375-9389, 2018.

[17] Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, 2017.

[18] Suk, S.-W. Lee, D. Shen, and the Alzheimer‘s Disease Neuroimaging Initiative, "Deep sparse multi-task learning for feature selection in Alzheimer‘s disease diagnosis," Brain Structure and Function, vol. 221, no. 5, pp. 2569-2587, 2016.

[19] Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," NeuroImage, vol. 129, pp. 292-307, 2016

[20] Qiao, Y. Lv, C. Cao, Z. Wang, and A. Li, "Multivariate deep learning classification of Alzheimer‘s disease based on hierarchical partner matching independent component analysis," *Frontiers in Aging Neuroscience,* vol. 10, 2018, Art. no. 417.

[21] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. Nature 323:533. doi: 10.1038/323533a0

[22] Shen, D., Wu, G., and Suk, H.-I. (2017). Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221–248. doi: 10.1146/annurev-bioeng-071516-044442

[23] Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv 1502.03167.

[24] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative adversarial nets,” in Advances in Neural Information Processing Systems, 2672–2680. Available online at: https://papers.nips.cc/ paper/5423-generative-adversarial-nets

[25] Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural Comput. 18, 1527–1554. doi: 10.1162/neco.2006.18.7.1527

[26] Hinton, G. E., Dayan, P., Frey, B. J., and Neal, R. M. (1995). The “wakesleep” algorithm for unsupervised neural networks. Science 268, 1158–1161. doi: 10.1126/science.7761831

[27] Salakhutdinov, R. (2015). Learning deep generative models. Annu. Rev. Stat. Appl. 2, 361–385. doi: 10.1146/annurev-statistics-010814-020120

[28] Vieira, S., Pinaya, W. H., and Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci. Biobehav. Rev. 74, 58–75. doi: 10.1016/j.neubiorev.2017.01.002

[29] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S. (2019). A comprehensive survey on graph neural networks. arXiv 1901.00596

[30] Zeng, L. L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., et al. (2018). Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. Ebiomedicine 30, 74–85. doi: 10.1016/j.ebiom.2018.03.017