Transaction on Biomedical Engineering Applications and Healthcare
Vol-1 Issue-1, June 2020, PP.18-25

Improving the Predicting Rate of Alzheimer's disease through Neuroimaging Data using Deep Learning Approaches

Subhrajyoti Ranjan Sahu[1], S. Swetha [2]

[1] B. Tech.,ECE, Nalanda Institute of Technology, Bhubaneswar.
[2]Assistant Professor, EEE Dept., Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi
[1]subhrajyotiranjan@hotmail.com, [2]swethasrinivasan4@gmail.com

Abstract - **Recently deep learning has shown a improved performance than machine learning in many of the areas like pattern recognition, image classification computer vision, video segmentation and many more. But out of all these areas, disease classification is one of the major area in which deep learning has shown a remarkable performance than the traditional machine learning algorithms especially in the area of image recognition. Machine learning algorithms are not enough capable to handle the image so in this work we will apply the deep learning approach on the Alzheimer's disease dataset for performing the early detection and classification of the disease and this has done through using neuroimaging data. Previous work done in this area was based on traditional machine learning algorithm and they have used stacked auto encoder (SAC) for dimensionality reduction and they have achieved a classification accuracy of 83.7% during the prediction from initial symptom to final development of Alzheimer's disease. The deep learning algorithm ResNet which is implemented in this paper has shown a classification accuracy of 93% and this is also achieved without applying any dimensionality reduction approach and this has been considered as the best predictive rate on the neuroimaging data till now. The applied ResNet is the improved ResNet and the comparison of both the Resnet models are shown in this work. This deep learning application will also be useful for other types of disease classification like cancer, diabetics, etc.**

Keywords— ResNet, mild cognitive impairments (MCI), ADNI, ReLU, Residual Block, Convolutions

1. Introduction

One of neuropsychologists' major problems in the last 50 years is the perception of the cognitive and behavioral symptoms of dementia and the association between it and underlying brain pathology. Due to increase in the age of dementia-related neurodegenerative disorders, there has been significantly increased in the risk over the years. Although the idea of dementia existed for several thousand years, the main clinical and related neurodegenerative diseases were not discovered until the beginning of the past century. In 1907 Alzheimer's Aloysius has analysed the symptoms during the treatment of a 51 year old woman and his description has become the first neuropsychological characterization of the disorder. Alzheimer's disease (AD), the most prevalent type of dementia, is a significant health concern in the 21st century and it is estimated 5.5 million people aged 65 and over live with AD. Lots of effort is already done by many researchers for the early detection of the disease especially in the presymptomatic stages, in order to delay or prevent the development of diseases [1].

It happens that Alzheimer disease memory get affected slowly, unresponsive, everyday living skills may degrade slowly, and unexpected changes in personality and compliance may occur. Even they did not remember the friends and their loved ones. The burden on themselves and their families is also great. The disease of Alzheimer is more of an irreversible and psychic blow. All treatments are limited to slow the degradation cycle and cause a rift between patients and their families.

The research discussed in [2] that person who is taking care of AD patient also suffers from higher depression incidence. Early prediction, identification and identification are necessary. Although this disease is irreversible, the disease can be recognized early and we can take environmental and drug intervention measures to slow the disease progression. In this work we are incorporating technology for the prediction of AD through deep learning approaches but before discussing about the deep learning implementation, this work will also focus some of the previous work done in this area. Already there are some advanced neuroimaging technology like MRI and PET which are already doing good job in the area of identification of structural and molecular markers of AD [3]. Despite of all these development in the area of neuro imaging there has been rapid developments in neuroimaging techniques which have made incorporation of high-scale multimodal neuroimaging data a challenge.

Consequently, interest in computer-aided learning approaches for integrative analysis has rapidly increased. Some method are already developed for early detection and prediction of AD [4] which can be considered as well-known pattern analysis methods, such as Linear Discriminant analysis (LDA), logistic regression (LR), support vector machine (SVM), linear program boosting method (LPBM) and Support Vector Machine – Recursive Feature Elimination (SVM – RFE).

During the implementation of machine learning algorithms there are some predefine method on the basis of which classification has to be done [5]. The focus of machine learning algorithm is mainly on four steps i.e. feature selection, feature extraction, dimensionality reduction and the last one is the classification algorithm. Although the first three steps are considered to be preprocessing steps and all these procedures requires specialized skills and optimization techniques. The problem with these preprocessing techniques is that they require much amount of time during the operations.

Other than the time consumption there are some more problems are there when dealing with machine learning classification algorithm for the prediction of AD. One of the problems occurs with the use of feature selection approach is that during the selection of features from the various neuroimaging modalities to inherit more informative measures based on combinations this may also involve the factors like subcortical volumes, cortical thickness, gray matter densities, brain glucose, etc and this are considered to be a problem [6].

To overcome from all these problems occur due to machine learning algorithms in the area of AD, deep learning algorithm came into existence. The deep learning algorithms are using the raw data from the neuroimaging for generating the essential features which has performed considerable attention during high dimensional medical image analysis.[7] In this work Convolutional Neural Network Algorithm is used for prediction purpose and it has outperformed other state of the art algorithm.

2. Background Study

In this section discussion will be on research which has done earlier in the area of AD through machine learning approaches on the different types of images generated by the neuroimaging devices. Lu et al. [8] has implemented a new deep-learning framework for discriminating people with AD through the use of a deep neural multimodal and multiscale network. This system has computed 82.4 percent specificity to classify individuals with mild cognitive impairment (MCI) that converts to AD at 3 years prior to conversion (86.4 percent conversion accuracy within one- to three years), a clinically diagnosed sensitivity of 94.23% and a specialization of 86.3 percent for non-demented controls Classification.

Ortiz et al. [9] has solved the detection problem through sparsely replicated data, which also allows the combination of specialized classifiers for the classification of multimodal images (PET and MRI). It's a new way of efficiently combining SVC classifications which is, using the distance measured for each class in each classifier for the hyperplane, allowing the most discriminatory image mode to be selected in each case. While functions in diagnosed Alzheimer's patients (AD) are clearly visible as compared to control subjects, behavioral changes that appears in the early stages of the disease and are more significant in the case of patients with Mild Cognitive Impairment (MCI).

Shi et al. [10] Has implemented a system that is used to fuse and learn task representation from Multimodal Neuroimaging Data to diagnose AD with a multimodal stacked DPN algorithm (MM-SDPN) consisting of two step SDPN (MM-SDPN). Specifically, two SDPNs are first used to learn high-level MRI and PET characteristics, which are then given to another SDPN to combine multimodal neuroimaging information. For binary and multi-class classification activities, the proposed MM-SDPN algorithm is used for the ADNI dataset. The result has shown that multimodal, learning-based AD diagnostic algorithms are performing better than the other algorithms which are based on stacked encoder.

Raza et al. [11] has proposed a new machine learning algorithms which is used to diagnose and monitoring of AD-like disease. The diagnosis phase of AD-like diseases is achieved by interpreting deep learning magnetic resonance imaging (MRI) scans, and an active monitoring system is provided to track the everyday behavior of subjects by using inertial sensors used by their bodies. The activity tracking offers a supportive structure for daily tasks and assesses the vulnerability of patients based on the level of activity. In contrast to well-known current methods, this model has shown 82% improvement over other existing model.

Frazer et al. [12] has derived the data are derived from the Dementia Bank corpus, which provides 240 narrative samples to 167 patients diagnosed with "possible" or "probable" AD, and an additional 233 to 97 controls. In order to differentiate between AD-participants and successful checks they measured a number of linguistic variables from transcripts and acoustic variables through the corresponding audio files. They pursue an exploratory factors study on these speech and language interventions with an oblique promax rotation to analyze the degree of heterogeneity of AD linguistic impairments and provide the interpretation of the resulting factors.

Ahmad et al. [13] implemented a novel method for fusing existing color spaces which has shown more effective results in practice than single color spaces. The segmented objects are the mouths, ears, paws, legs, and leaves of the tree. Using multiple databases to reflect these issues, the ANN was trained as an object or non-object to the color of the pixel and its surrounding 8 neighbors. In the testing phase, the trained data was utilized to split the nine pixels of the text image into an object or non-object. To analyze the impacts of the blending done on different types of color information from the various color models of the targeted pixel they have used the vector function for training the model.

3. Proposed Methodology

i. Dataset

In this work we have obtained data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [14]. The patients are selected on the basis of three categories i.e. normal controls (NC), mild cognitive impairments (MCI) and Alzheimer's disease. A total of 8809 data samples are available and for each patient both MRI images and clinical data was captured. To perform the research we have taken structural T1-weighted Magnetic Resonance Imaging (MRI), because it is easily available with all the patients. Although the images retrieved from the ADNI database are already preprocessed through different pipeline stages. To extract the image of the brain, an ANTs Cortical Thickness Pipeline was executed. After this execution an affine transform was applied on all the brain images so that they have the same orientation.

ii. ResNet

In this work the implementation of deep learning will be done in the form of Convolutional Neural Network (CNN) [15] by implementing the ResNet3D or conv residual network. We have used ResNet3D because these neural networks are getting deeper and deeper. Before discussing ResNet3D, first we will discuss about the basic concept of ResNet.

The ResNet neural network trained a very deep 152-layer neural network that has some very interesting tricks and ideas which can effectively work on the image recognition system. By using ResNet we can build effective Convolutional Neural Network. Although it's very difficult to train deep neural networks because of vanishing and exploding gradient types of problems. In this network, activation from one layer and suddenly feed it to another layer even much deeper in the neural network. Through this we can build ResNet which enables you to train very, very deep networks, sometimes even networks of over 100 layers. The given below equation will show how the residual block will be build for ResNets.

:

In the initial processing of neural network assume the two layers in the network with some activations in x[l] layer, then go to x[l+1], and then activation after two layers is x[l+2]. Now let's see the computation of x[l] and then apply the linearity into it which can be performed by using the equation given below:

$$z[l+1] = w[l+1]\, x[l] + b[l+1] \quad (1)$$

In the above equation, the input is multiplied with the weight matrix and adds the bias vector. Later on to get x[l+1], a ReLU activation function is added for the nonlinearity [16]. This is achieved by the equation:

$$x[l+1] = R(z[l+1]). \quad (2)$$

R represents the ReLU activation function The same linear steps will be followed in the next layer,

$$z[l+2] = w[l+2]\, x[l+1] + b[l+2] \quad (3)$$

At last there will be another ReLU operation that will be performed for the non linearity and this is represented by the equation:

$$x[l+2] = R(z[l+2]) \quad (4)$$

Through the above equation it can be easily observe that to get information from x[l] to x[l+2], it needs to go through all these steps and they are considered to be the main path of this set of layers. We will make some changes in the residual net by putting x[l], and just forward it first, copy it and forward it on neural network by directly sending the information from x[l] to x[l+2]. So rather than following a complete path the information from x[l] can now follow a direct path to go much deeper into the neural network. So the last equation will be

$$x[l+2] = R(z[l+2] + x[l]). \quad (5)$$

The inclusion of x[l] will form this residual block.

These pictures can be directly send to the top by creating a direct path to reach there. The nodes represented here are applied with linearity function and ReLU because the skip connection or the shortcut paths are applied before the non linearity and the captured image is passed through the second layer. After the linear transformation, x[l] is injected just before the ReLU and it jumps directly by skipping over a layer (almost 2 layer) in order to process the information deeper into the neural network. It can easily analysed that residual blocks allows to train much deeper neural networks and the way a ResNet is build by taking away many of these residual blocks, blocks like this and stacking them together to form a deep network. So, if the network is not residual, then we can build the residual net by adding all those skip connections, even if they're short connections like this exist in a model. In this network every two layers ends up with the additional change that transforms each of them into a residual block and the use of regular optimization algorithm, such as a gradient descent, or one of the fancier optimization algorithms on a train or a simple network is done. So without all the extra residuals, without any extra short cuts or skip connections, it has been analysed that the increase in the number of layers will tends to decrease the training error but after a while they tend to go back up again.

In order to deepen the neural network, the performance has to be improved on the training set. But what happens to ResNet is that even as the number of layers gets higher the training error also tends to increase even if we train a network of over a hundred layers. There has been insertion of different types of activations in the intermediate layers that helps to enter into the deeper neural network. Due to this during the training of deeper neural networks, the real loss of performance will get minimized with the help of vanishing and exploding the gradient problems.

To perform well on the training set, let us take X feeding in to some big neural network and just outputs some activation x[l]. Let's say for this example that you are going to modify the neural network to make it a little bit deeper. So, use the same big NN, and this output's x[l], and we're going to add a couple extra layers to this network so let's add one layer there and another layer there. And just for output x[l+2]. Only let's make this a ResNet block, a residual block with that extra short cut.

Given below figure shows the block diagram of ResNet

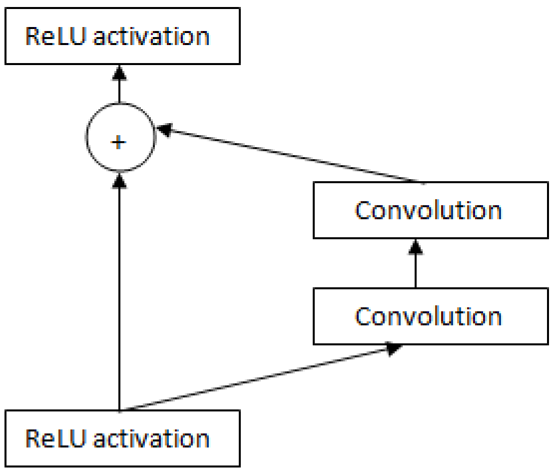

Figure 1. Block diagram of ResNet

Activation function ReLU [17] is defined by the equation:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (6)$$

In the above equation f(x) is the output function and x is the input unit. It is easy to train CNN's because ReLU function alleviates the problems of vanishing gradients through its identity map in the positive quadrants.

Mathematically we can define convolution as function that derived by the integration of two different functions so that it can show how one has made impact on the another. Convolution [18] is defined as:

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \quad (7)$$

f and g are the two different function in this equation and they are integrated to form a convolution.

iii. Improving the ResNet

We use 6 residual blocks each for ResNet like architecture whose identity mapping comprises a minimum of two convolutions with batch normalization and ReLU of 3x3x3 filter size and 64 or 128 filters. In improved ResNet, ConvBlocks are replaced by convolutionary layers from the standard ResNet. The range is reduced by three 2x2x2 phase convolutions until residual blocks and a limit of one 5x5x5 kernel pooling layer in front of full-connected layer. Just after the first fully connected layer, a dropout with p=0.7 and batch normalization has been applied. Lastly there is a second fully connected layer with softmax function [19] to produce the output probabilities.

In figure 1 we have shown the flowchart of the proposed work, the flowchart has shown the various steps which is performed during the classification of neuroimaging data. The data is processed through improved ResNet model and it has computed the different parameters like accuracy, AUC, sensitivity and specificity.

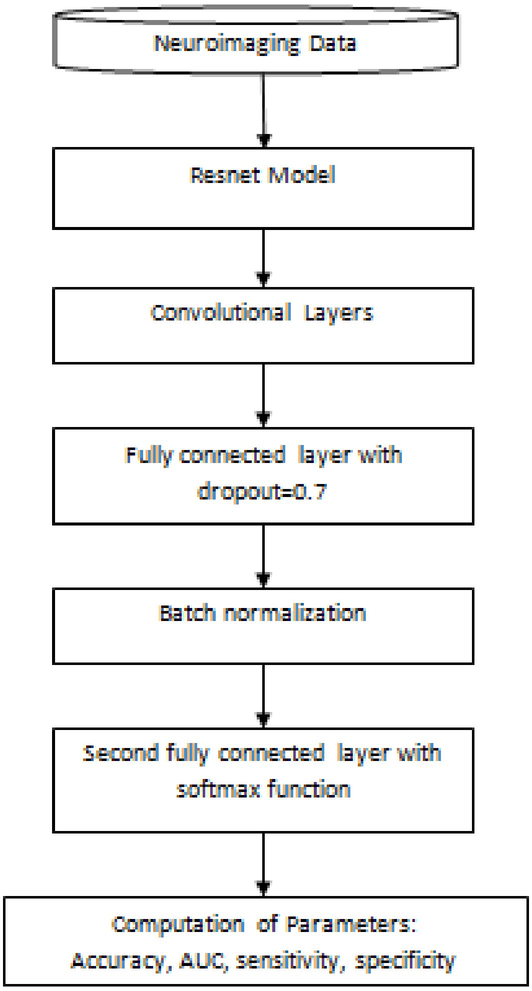

Figure 2. Flowchart of the proposed ResNet convolution Model

4. Experiment and Result Analysis

The work is implemented using Python 3 and the data is based on neuroimaging data with the conversion predictor, we minimize a balanced class binary cross-entropy weighted loss function, with initial learning rate 10−3 using the Nesterov momentum optimizer [20]. There will be a decrement of 10 fold after the epoch of 30 and 50. The batch size is of 128 dimensions, the reason for selecting this dimension is to match the entire batch into the GPU. There are 70 epochs in total.

Execution of 5-fold group cross-validation with five different folds is done to make the classification performance more accurate. When there is different scan of the same patient then we have use the patient id for training and testing purpose. During the cross validation procedure of the neuroimaging data, we have trained a separate neural network on the training dataset and have used the validation set for managing the learning rate.

The metrics used for the computing the prediction performances of the model are the accuracy, AUC, sensitivity and specificity [21]. The classification implemented in this work is binary classification in which class 0 represents the stable MCI and class 1 represents the converged MCI [22]. The dataset is split in training and testing dataset of 75% and 25% of the dataset respectively.

Results

Given below Table 1. will shows the result computed from the neuroimaging dataset with the implementation of improved ResNet Covolutional Neural Network. The results are compared with the other state of algorithms like logistic regression [23] and Xgboost [24].

Table 1. Comparison of the models on the basis of different types of metrics

Models	Accuracy	AUC	Sensitivity	Specificity
Logistic Regression (Clinical dataset)	77%	0.61	0.67	0.76
Xgboost (Clinical dataset)	76%	0.59	0.66	0.78
Inception-v4 [25]	75%	0.587	0.68	0.78
ResNet	82%	0.638	0.73	0.77
Improved ResNet	93%	0.671	0.94	0.93

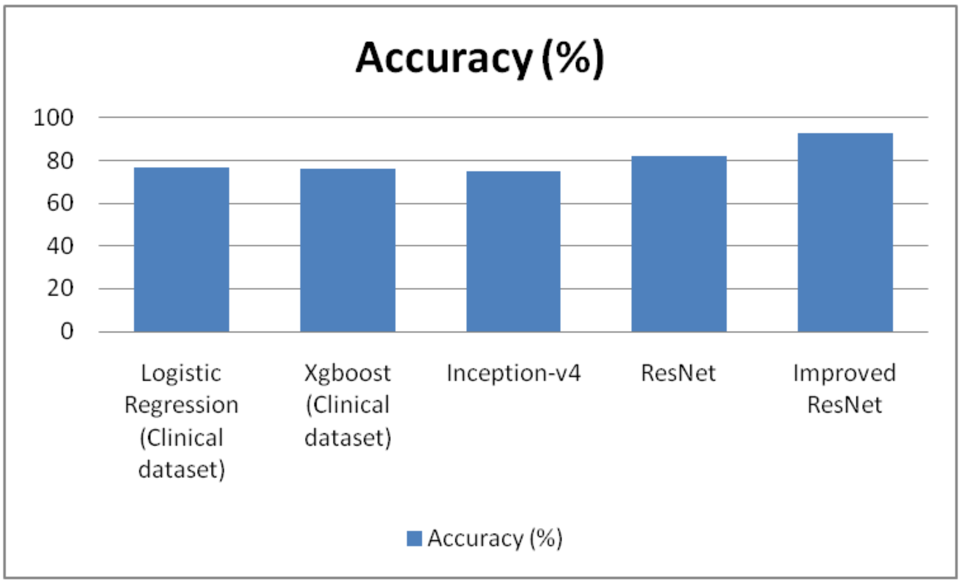

Figure 3. Accuracy comparison of the models

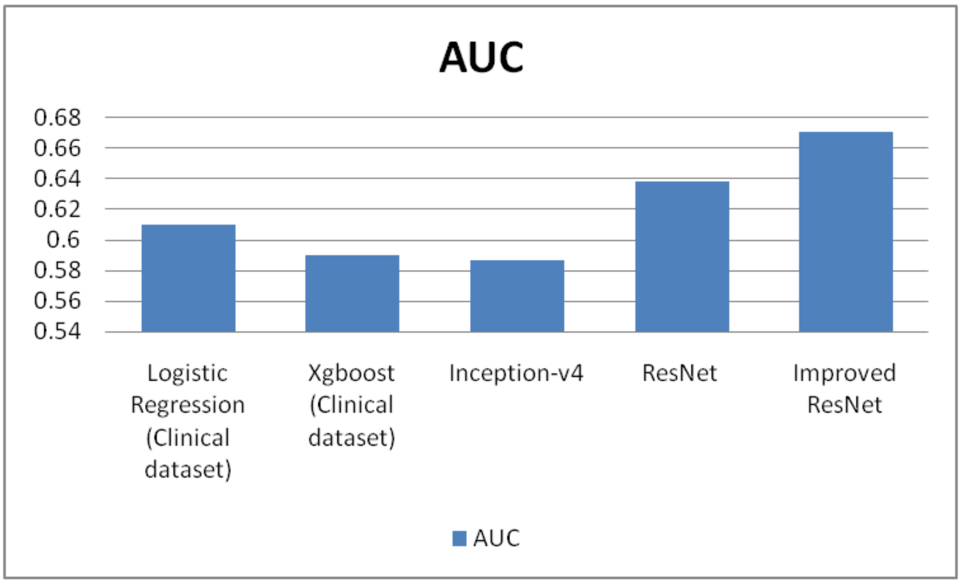

Figure 4. AUC comparison of the models

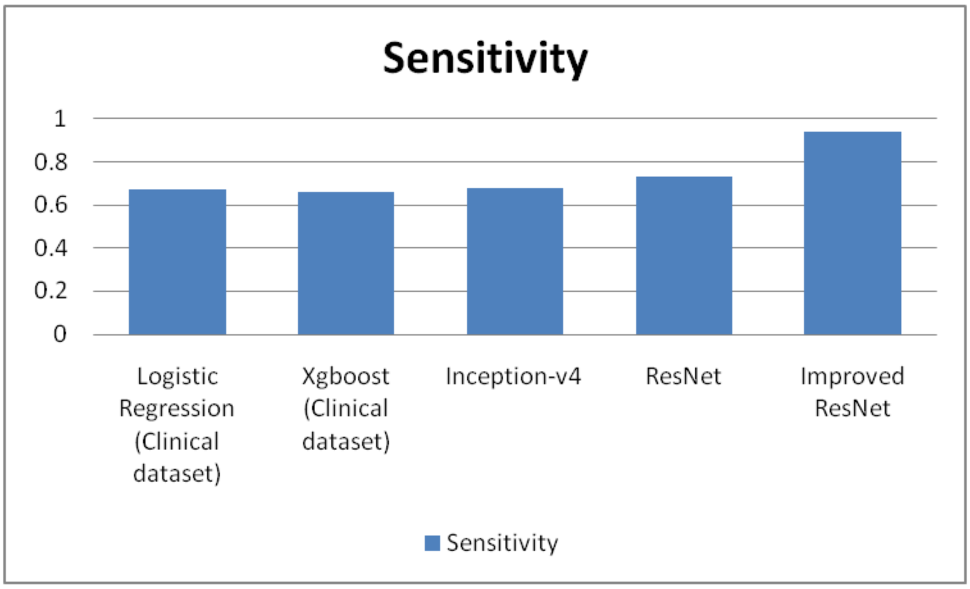

Figure 5. Sensitivity comparison of the models

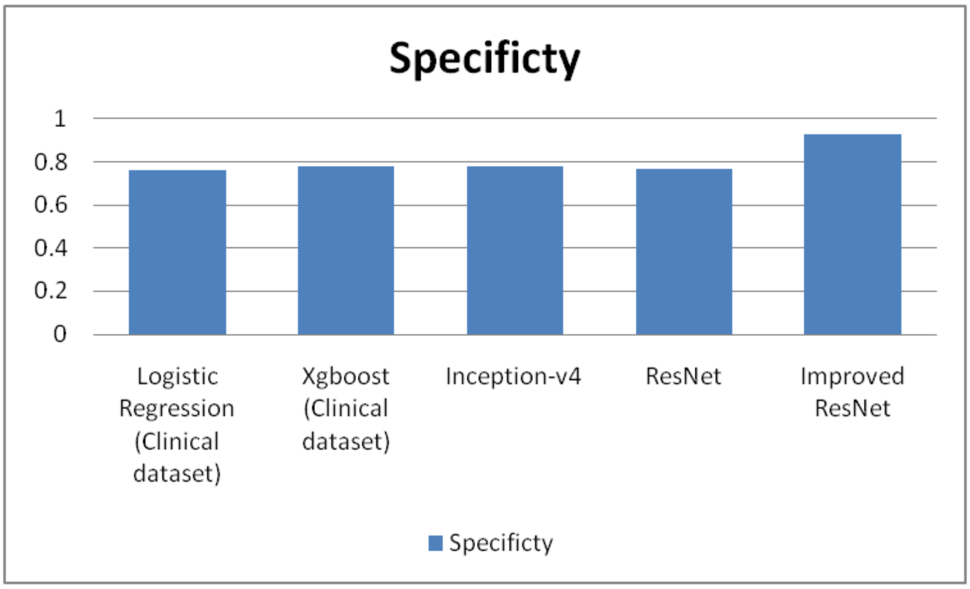

Figure 6. Specificity comparison of the models

Above graphs has shown the result analysis based on ADNI database computed by the improved ResNet model. The model has applied on the ADNI dataset and it has achieved the accuracy of 93%, AUC of0.671, sensitivity of 0.94 and the value of specificity is 0.93. The model has outperform other model due to its improved results.

5. Conclusion

The discussed model in this work has achieved classification accuracy i.e. 93%, better than the other state of the art algorithms. Although from this work it has been observed that deep learning algorithms can handle the complex problems by applying the property of non linearity and this works fine especially in case of single class neuroimaging prediction. But still it is a challenge to provide a good accuracy on the multi-class classification network and this will be the future development in this area which can be applied for diagnostic and prognosis biomarkers for psychiatric and neurological disorders.

References

1. J. E. Galvin, “Prevention of Alzheimer’s Disease: Lessons Learned and Applied,” J. Am. Geriatr. Soc., 2017.
2. A. S. Alfakhri et al., “Depression among caregivers of patients with dementia,” Inq. (United States), 2018.
3. D. P. Veitch et al., “Understanding disease progression and improving Alzheimer’s disease clinical trials: Recent highlights from the Alzheimer’s Disease Neuroimaging Initiative,” Alzheimer’s and Dementia. 2019.
4. S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, “A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer’s disease and its prodromal stages,” NeuroImage. 2017.
5. N. S. Vassilieva, “Content-based image retrieval methods,” Program. Comput. Softw., 2009.
6. B. C. Riedel et al., “Uncovering Biologically Coherent Peripheral Signatures of Health and Risk for Alzheimer’s Disease in the Aging Brain,” Front. Aging Neurosci., 2018.
7. S. M. Plis et al., “Deep learning for neuroimaging: A validation study,” Front. Neurosci., 2014.
8. D. Lu et al., “Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer’s Disease using structural MR and FDG-PET images,” Sci. Rep., 2018.
9. A. Ortiz, F. Lozano, J. M. Gorriz, J. Ramirez, and F. J. Martinez Murcia, “Discriminative Sparse Features for Alzheimer’s Disease Diagnosis Using Multimodal Image Data,” Curr. Alzheimer Res., 2017.
10. J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, “Multimodal Neuroimaging Feature Learning with Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer’s Disease,” IEEE J. Biomed. Heal. Informatics, 2018.
11. M. Raza, M. Awais, W. Ellahi, N. Aslam, H. X. Nguyen, and H. Le-Minh, “Diagnosis and monitoring of Alzheimer’s patients using classical and deep learning techniques,” Expert Syst. Appl., 2019.
12. K. C. Fraser, J. A. Meltzer, and F. Rudzicz, “Linguistic features identify Alzheimer’s disease in narrative speech,” J. Alzheimer’s Dis., 2015.
13. A. B. A. Hassanat, M. Alkasassbeh, M. Al-awadi, and E. A. A. Alhasanat, “Color-based object segmentation method using artificial neural network,” Simul. Model. Pract. Theory, 2016.
14. S. G. Mueller et al., “The Alzheimer’s disease neuroimaging initiative,” Neuroimaging Clinics of North America. 2005.
15. Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, “Recent advances in convolutional neural network acceleration,” Neurocomputing, 2019.
16. B. Hanin, “Universal Function Approximation by Deep Neural Nets with Bounded Width and ReLU Activations,” Mathematics, 2019.
17. R. Arora, A. Basu, P. Mianjy, and A. Mukherjee, “Understanding deep neural networks with rectified linear units,” in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.
18. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.
19. M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks arXiv:1311.2901v3 [cs.CV] 28 Nov 2013,” Comput. Vision–ECCV 2014, vol. 8689, pp. 818–833, 2014.
20. R. Singari, K. Singla, and G. Chawla, “Deep Learning Framework for Steel Surface Defects Classification,” Int. J. Adv. Prod. Ind. Eng., 2019.
21. A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” Pattern Recognit., 1997.
22. R. C. Petersen, “Mild cognitive impairment as a diagnostic entity,” in Journal of Internal Medicine, 2004.
23. D. Bzdok, M. Eickenberg, O. Grisel, B. Thirion, and G. Varoquaux, “Semi-supervised factored logistic regression for high-dimensional neuroimaging data,” in Advances in Neural Information Processing Systems, 2015.
24. C. Gao et al., “Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson’s disease,” Sci. Rep., 2018.
25. C. Szegedy et al., “Inception v4,” Popul. Health Manag., 2015.