Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231107

Classification for Early Imaging of Alzheimer's Disease Based on Deep Learning

Xiuxu LI[a], Lanhua ZHANG[a], Xiaodong LIU[a] and Jin WANG[b,1]
[a] *School of Medical Information and Engineering, Shandong First Medical University and Shandong Academy of Medical Sciences, Tai'an, 271016, China*
[b] *Cybersecurity Industry Development Center (Information Center) of Ministry of Industry and Information Technology, Beijing, 100846, China*

Abstract. Deep learning is important for early warning of Alzheimer's disease in imaging histology studies. In order to realize the early warning of medical images based on deep learning models, traditional machine learning and deep learning models as well as techniques such as confusion matrix and ROC are combined to complete the model classification situation evaluation function, and finally visualize the results of deep learning VGG outperforming the effect of traditional machine learning models. The conclusion of the deep learning VGG model shows that the deep learning model is feasible or meaningful after certain pre-processing and in early Alzheimer's disease image classification, and has an important role in early medical image recognition of certain diseases.

Keywords. Machine Learning, classification, deep learning, confusion matrix, Alzheimer's disease

1. Introduction

Alzheimer's disease is a common progressive neurodegenerative disease of the central nervous system, with clinical symptoms such as progressive memory loss, cognitive decline, and mental and behavioral abnormalities. Alzheimer's disease is a neurodegenerative disease that is clinically characterized by progressive cognitive and memory impairment. It is a complex progressive neurodegenerative disease affecting approximately 14 million people in Europe and the United States, including almost one-half of the population aged 85 years (43%)[1-3].

It is well known that Alzheimer's disease is currently a neuroglial inflammatory disorder involving association with amyloid plaques. The associated pathomechanism explains that microglia is usually associated with specific plaque types that can be explained by characteristic formation mechanisms [4]. The common machine learning models can be used for classification. Deep learning is the extend function based on

[1] Corresponding Author, Jin WANG, Cybersecurity Industry Development Center (Information Center) of Ministry of Industry and Information Technology, Beijing, 100846, China; E-mail: wangjin@miit.gov.cn.

machine learning. Machine learning allows us to classify types of diseases that are close to each other in terms of a parametric factor [5, 6].

Confusion matrix is a matrix showing the predicted and actual classifications [7]. It is an important evaluation indicators in machine learning, and can be seen quite visually model performance in a certain type of data [8]. The logistic function is originally designed to be used in population growth [9]. Later on, it is gradually used to solve the machine learning approach of correlation binary classification, which is used to estimate the likelihood of a certain thing [10].

Support vector machines are machine learning models which can be applied to continuous, binary, and categorical outcome type problems [11]. Naive Bayes utilizes Bayes' law and a strong assumption that the properties of a given class of data are conditionally independent, and it is widely used in the field of machine learning [12]. The nearest neighbor algorithm uses the K-nearest pattern label in the data space as the local feature of the model [13].

Visual geometry group (VGG) network is one of the commonly used pattern recognition models in computer vision, and it has been verified in the previous years that it has a certain effect on the recognition of objects in static images [14]. The use of a small convolutional filter structure for enhancing the depth of the network shows that improvements to the existing technical configuration can be achieved by adding a depth of 16-19 layers to the weight hierarchy [15].

To address degradation issues, ResNet network introduces a deep residual learning approach to simplify the training of deeper networks [16].

In machine learning, we will classify the classification problem in more than two categories, called multi-classification problem; when we need to compare performance metrics between multiple methods and models, a qualified performance metric is useful, and many metrics can be used when testing the capability of a multi-classification classifier [17].

2. Methods

2.1. Data sources and Algorithm Models

The dataset is from Kaggle group which is consists of MRI images, and all images are resized into 128*128 pixels. The dataset has a total of 6400 images and is divided into four types: Mild Demented, Moderate Demented, Non Demented and Very Mild Demented. Data was collected from several website, hospitals, public repositories. And at last these data are preprocessed in advance to facilitate subsequent processing. The sources of the dataset can be gotten from webs (https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers; https://catalog.data.gov/dataset/alzheimers-disease-and-healthy-aging-data) [18].

In order to test the efficient of the different algorithms for the classification of the imaging, we use a variety of machine learning and deep learning methods to classify early medical images. The algorithms includes support vector machine, random forests, K-nearest neighbors, Naïve Bayes, logistic regression, VGG and ResNet.

Support vector machine can perform nonlinear classification through kernel method [11]. Random forest is the extension of decision tree algorithms. In the feature space, if most of the k nearest samples near a sample belong to a certain category, then the samples also belong to that category, that is K-nearest neighbors [13]. Bayesian classification algorithm is a statistical classification method that utilizes probability and statistical knowledge for classification, Naïve Bayes is an algorithm formed based on this method [12]. Through logistic regression analysis, the weights of the independent variables can be obtained.

VGG network is a convolutional neural network, and it is often used to extract image features [14, 15]. In order to compare, ResNet50 networks execute the same operation.

AdaDelta, Adam, AdaMax, and Nadam are four adaptive learning rate algorithms [19]. These algorithms are commonly used in neural networks to address the issue of vanishing learning rates and improve training performance. AdaDelta is a modified version of the AdaGrad method. Adam, short for adaptive moment estimation, combines gradient descent with momentum and RMSProp. AdaMax is a variant of Adam that uses the $L\infty$ norm instead of the L2 norm to calculate the gt term. It also includes the same update equations as Adam, with slight modifications. Nadam, or Nesterov accelerated adaptive moment estimation, incorporates Nesterov accelerated gradient into Adam. These algorithms have different hyperparameters, such as learning rate (η), momentum term ($\beta 1$), RMSProp term ($\beta 2$), and ε. The information also provides details on the training time, memory utilization, and accuracy of these algorithms on a test set.

2.2. Work flow of our works

First, we collected fMRI images from open source dataset about early-stage Alzheimer's disease. The data has four classes of images both in training as well as a testing set: Non Demented, Mild Demented, Very Mild Demented, and Moderate Demented. The dataset contains a total of 6400 images. The dataset is listed above, and we adopt some of them as the test samples.

Then, about the computing and programming, we used the Python as surrounding, except the basic libraries, we use the sklearn-kit to normalization the images, and then structured some basic models by logistic, support vector machine, Naïve Bayes, K-nearest neighbor, VGG and ResNet. Both machine learning and deep learning are utilized to ensure that the results are convincing.

Finally, we continuously adjusted and optimized some parameters of the models. By the basic results we conclude that deep learning models work better than machine learning models on early Alzheimer's classification problems, even some machine learning models even make mistakes.

In the final results, we can observe each model through the graphs of confusion matrix and ROC curve, and we can roughly estimate the classification effect of the relevant models.

3. Results

In detailed of the data processing, the image data was first divided into four types according to the early, middle, late and later stages. We resized them into 126*126 sizes, and divided them into test sets and training sets according to the ratio of 3:7 after entering each model method.

About the model building, we used Pytorch architecture [20] to build VGG16, VGG19 networks and used scikit-learn library [21] to build machine learning methods.

All the results can be finally visualized by matplotlib function library. By the programming, the final confusion matrix of every model was normalized. The corresponding python packages contain functions for pre-processing the data and methods for related operations.

Finally we got the confusion matrix for each model and the ROC curve for each model predicting each class. The results of the confusion matrix and the ROC curve were listed below after the normalization process (Fig. 1, Fig. 2). The worst performance of the different models in the test sample is the Naïve Bayes.

The confusion matrices of VGG16 and VGG19 were different (Fig. 3), but they had the similar accuracy and loss (Fig. 4). Meanwhile, ResNet50 was introduced to compare with VGG. We list the score in all the models (Tab. 1).

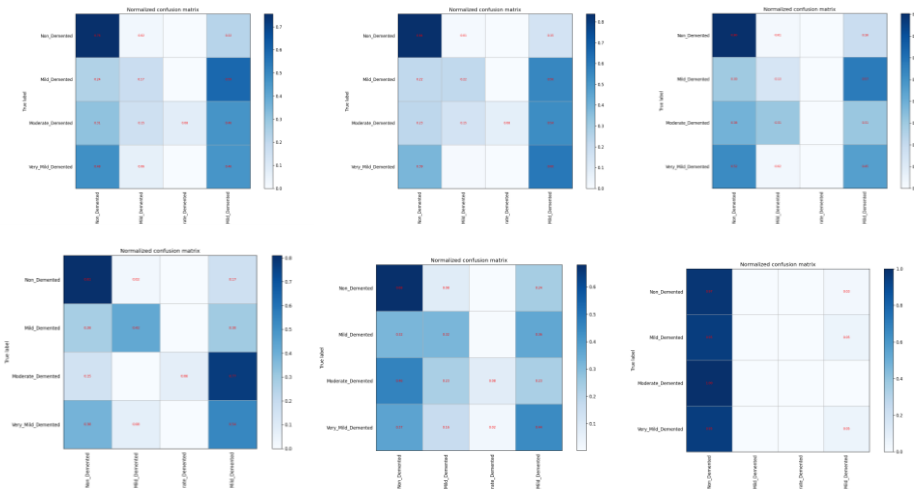

Figure 1. Confusion matrix results in models

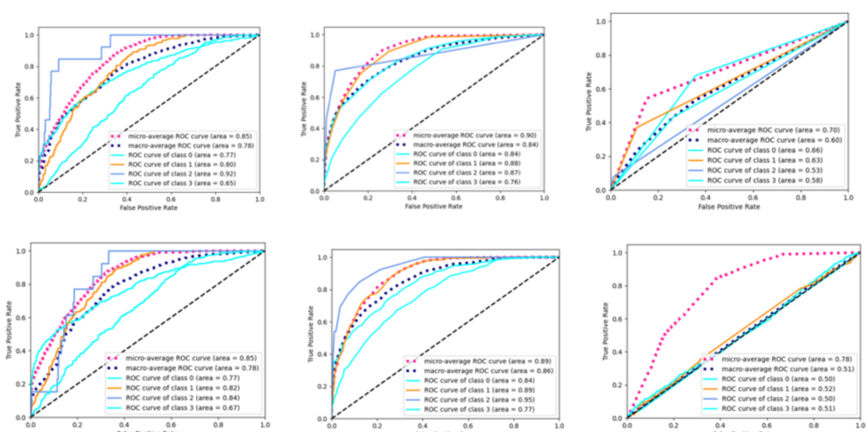

Figure 2. ROC curve results in models

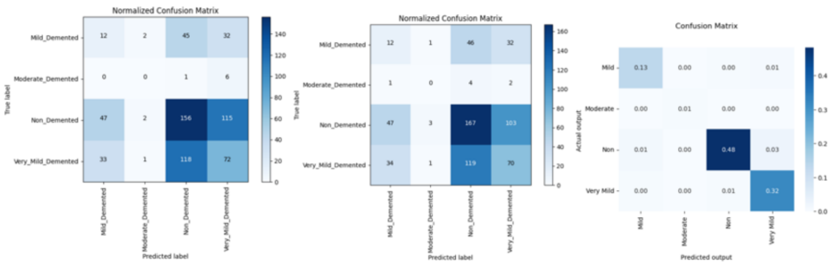

Figure 3. Confusion matrix results of VGG16 and VGG19

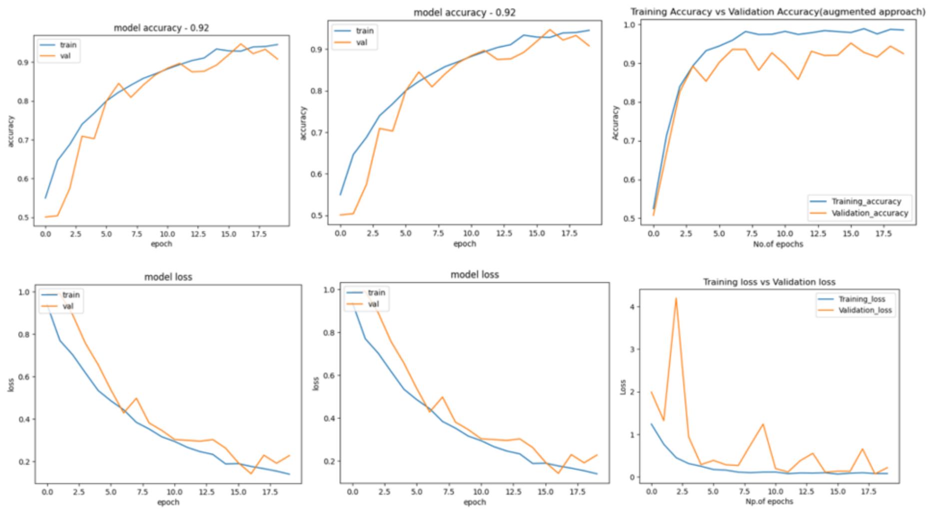

Figure 4. Accuracy and loss results of VGG16, VGG19 and ResNet50

Table 1. Score in Models

	SVM	KNN	RF	Naïve Bayes	Logistic	DT	VGG	ResNet50
Accuracy	0.57	0.65	0.66	0.50	0.59	0.57	0.90	0.92
F1-Score	0.55	0.65	0.64	0.37	0.56	0.57	0.89	0.93
Recall	0.57	0.50	0.66	0.50	0.59	0.57	0.89	0.92
Precision	0.56	0.72	0.67	0.40	0.57	0.57	0.90	0.93

4. Discussion

In the processing of our models, we find that the classification effect of the models is not very good when the data size is small. With the increasing of the size of the data, the classification of our models can achieve better results. Therefore, it can be proposed that machine learning and deep learning are all data-based training methods, and a certain effect can only be seen when a certain amount of data is available. In our models, we didn't do the strengthening of data which can also bring good effect for the models.

From the results of the confusion matrix, the models of support vector machine, logistic, K-nearest neighbors, decision tree and random forest finished the classification of the imaging data although the values of different models are different which implies that the algorithm express the basic functions. But the results of plain Bayes are worse than those of several other methods and models.

From the results of the ROC curves, the models of support vector machine, logistic, K-nearest neighbors and random forest got the efficient results. The decision tree got worse results than others. The Naïve Bayes is the worst same as the confusion matrix which implied that the Naïve Bayes didn't suit for these MRI imaging classification because Naïve Bayes demands individual features to be independent of each other in the calculations. Meanwhile the results implies that too little training data for some categories can also lead to worse classification results in the end.

The reason why such a problem occurs, we suspect that the amount of the model data is the main reason in the process of training in some, which resulted in the results of the training without achieving a satisfactory effect.

And through the final results, it is found that the results of the deep learning model are slightly improved compared to the traditional machine learning methods.

5. Conclusions

In this paper, we try to use traditional machine learning models and deep learning models for early medical image classification of Alzheimer's disease. In order to provide more methods for image classification in the clinical application, we used support vector machine, random forests, K-nearest neighbors, Naïve Bayes, logistic regression, and VGG networks to test the classification function. We computed the confusion matrix and ROC curves for all the models to demonstrate the effects of the models.

By the comparing of different models with confusion matrix and ROC curves, we found that most of the classification algorithm model could achieve the function with different effects. Data factors may lead to these reasons, and in the future, this result can be tested through a larger amount of data.

Acknowledgements

This research was supported by the Students' project for innovation and entrepreneurship training program of Shandong First Medical University and Shandong Academy of Medical Sciences (No. 2022104391174), Youth Science Foundation Cultivation Support Program of Shandong First Medical University and Shandong Academy of Medical Sciences (No. 202201-036) and the Tai'an Science and Technology Innovation Development Project (2022GX025).

References

[1] Alzheimer's Association. 2011 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 2011,7(2):208–244.

[2] Blennow K, Dubois B, Fagan AM, et al. Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease. Alzheimer's & Dementia, 2015,11(1):58–69.
[3] Hort J, O'Brien JT, Gainotti G, et al. EFNS guidelines for the diagnosis and management of Alzheimer's disease: EFNS guidelines for AD. European Journal of Neurology, 2010,17(10):1236–1248.
[4] D'Andrea MR, Cole GM, Ard MD. The microglial phagocytic role with specific plaque types in the Alzheimer disease brain. Neurobiology Aging, 2004,25(5): 675-683.
[5] Rakhimov M, Akhmadjonov R, Javliev S. Artificial Intelligence in Medicine for Chronic Disease Classification Using Machine Learning. 2022 IEEE 16th International Conference on Application of Information and Communication Technologies, 2022:1–6.
[6] Xie S, Guo R, Li N, et al. Brain fMRI processing and classification based on combination of PCA and SVM. 2009 International Joint Conference on Neural Networks, 2009:3384–3389.
[7] Kohavi R, Provost F. Glossary of terms. Machine Learning, 1998,30(2-3):271-274.
[8] Sammut C, Webb GI. Encyclopedia of machine learning and data mining. Springer, 2017.
[9] Cramer JS. The Origins of Logistic Regression. Tinbergen Institute Working Paper, 2002(119):4.
[10] Landwehr N, Hall M, Frank E. Logistic model trees. Machine Learning, 2005,59(1-2):161-205.
[11] Guenther N, Schonlau M. Support vector machines. Stata Journal, 2016, 16(4):917-937.
[12] Webb GI, Keogh E, Miikkulainen R. Naïve Bayes. Encyclopedia of machine learning, 2010,15(1):713-714.
[13] Kramer O, Kramer O. K-nearest neighbors. Dimensionality reduction with unsupervised nearest neighbors, 2013:13-23.
[14] Wang L, Guo S, Huang W, et al. Places205-VGGNet Models for Scene Recognition, Computer science, 2015. DOI:10.48550/arXiv.1508.01667.
[15] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science, 2014. DOI:10.48550/arXiv.1409.1556.
[16] He KM, Zhang XY, Ren SQ, et al. Deep Residual Learning for Image Recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:770-778.
[17] Grandini M, Bagli E, Visani G. Metrics for Multi-Class Classification: an Overview. Preprint, 2020. DOI:10.48550/arXiv.2008.05756.
[18] Sachin Kumar, Sourabh Shastri. Alzheimer MRI Preprocessed Dataset. Kaggle. 2022. https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset.
[19] Iqbal I, Odesanmi GA, Wang J, et al. Comparative Investigation of Learning Algorithms for Image Classification with Small Dataset. Applied Artificial Intelligence, 2021, 35(10):697-716.
[20] 10.34740/kaggle/dsv/3364939Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019. DOI:10.48550/arXiv.1912.01703.
[21] Pedregosa F. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2013, 12(10):2825-2830.