Final project

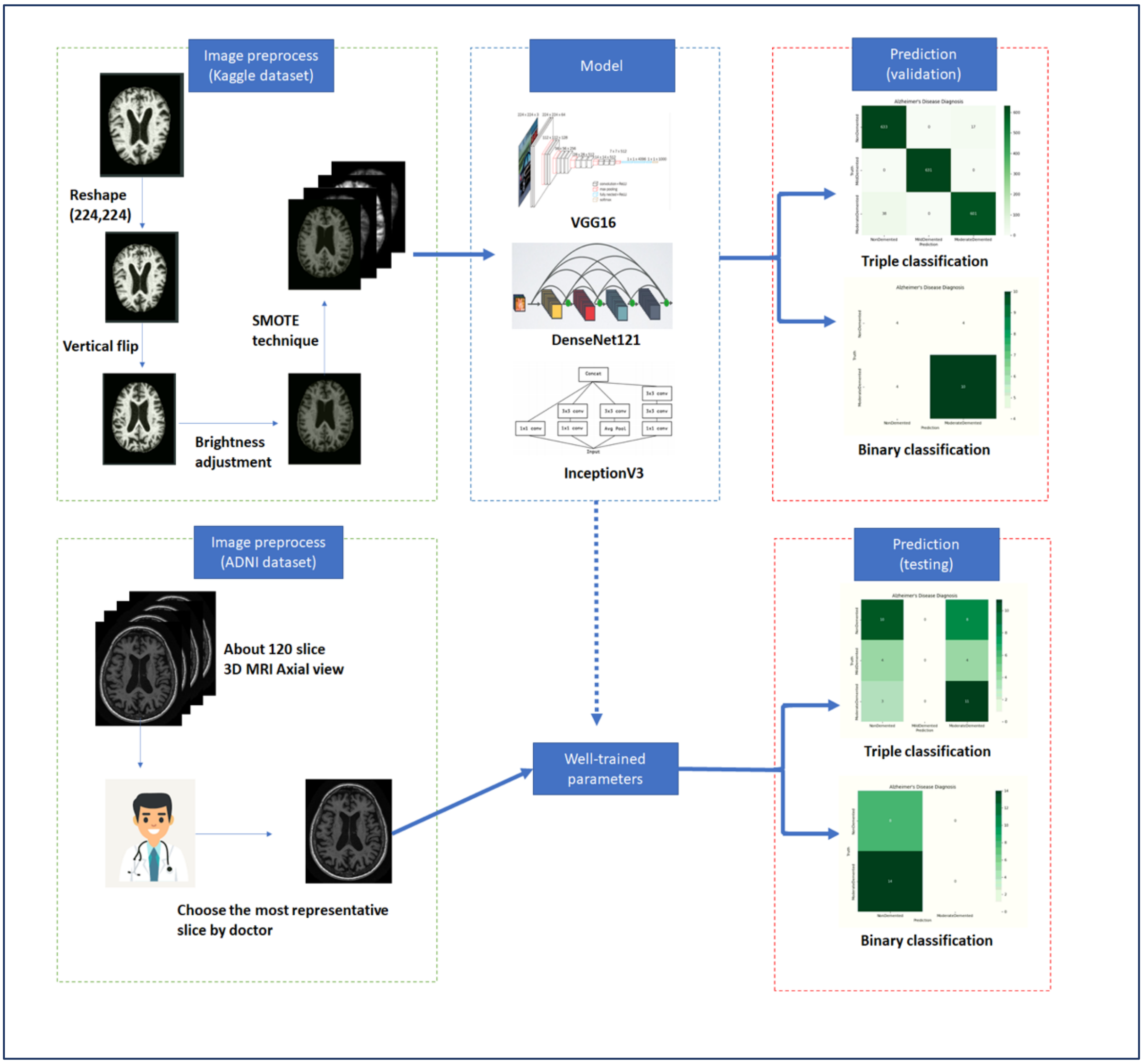
Hsuan-Ju,Chen & Tzu-I,Chuang

R09458004/ R10458007

GitHub: https://github.com/9xEzreaL/ML_projects_ADNI

1. Abstract

In this project, we aimed to classify non-demented, mild-demented and severe-demented three phases of Alzheimer's disease (AD) with using three different models. Then, we build a classifier based on DenseNet121, VGG16 and InceptionV3 that is able to accurately categorize images. Our dataset 1 is an open source from website-Kaggle used as training and validation data which has about 6000 MRI T1 images. Dataset 2 is from Alzheimer's Disease Neuroimaging Initiative (ADNI) used as testing data which has 14 control people, 18 patients of mild cognitive impairment (MCI) and 8 patients of Alzheimer's disease. Finally, we explore the possibility of creating binary and triple classification for different models, though we might be ultimately limited by image brightness, MRI sequence and model bias. The figure below is the flowchart of this project.



Flowchart

2. Introduction

2.1 Alzheimer's disease

Alzheimer's is the most common cause of dementia, a general term for memory loss and other cognitive abilities serious enough to interfere with daily life. Alzheimer's disease accounts for 60-80% of dementia cases. The early signs of the disease include forgetting recent events or conversations. As the disease progresses, a person with Alzheimer's disease will develop severe memory impairment and lose the ability to carry out everyday tasks. Alzheimer's disease typically progresses slowly in three general stages: preclinical, mild cognitive impairment and dementia[1]. (fig.1)

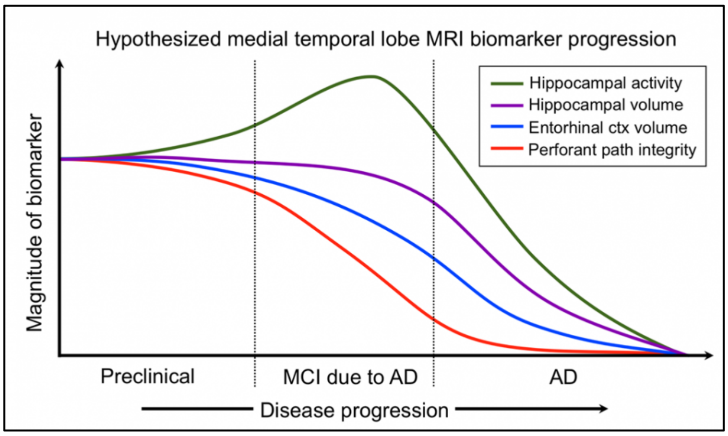


Figure. 1

2.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) uses a powerful magnetic field, radio frequency pulses and a computer for non-invasive in-vivo imaging to produce detailed pictures of organs, soft tissues, bone and virtually all other internal body structures. MRI can detect brain abnormalities associated with mild cognitive impairment (MCI) and can be used to predict which patients with MCI may eventually develop Alzheimer's disease. In the early stages of Alzheimer's disease, an MRI scan of the brain may be normal. In later stages, MRI may show an increase in the size of ventricle and the cortical fold atrophy of the brain shown as the figure 2 below in axial view[2].

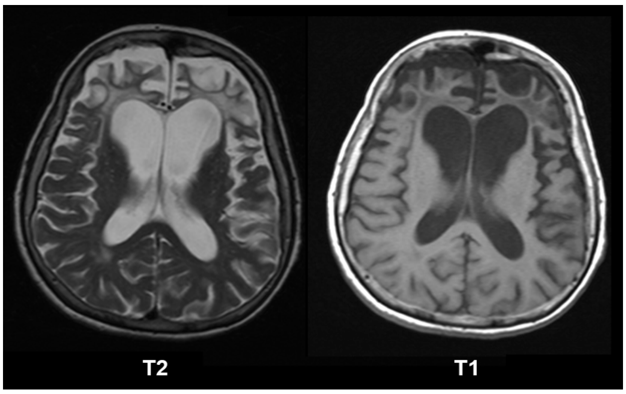
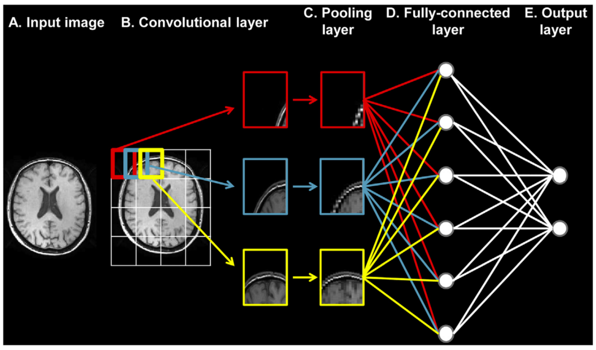


Figure. 2

2.3 Related work

In the previous studies[3,4], they use the architecture of a typical convolutional neural network (CNNs) with deep learning technology (fig.3) to predict Alzheimer disease. Binary classification model is used in the first study. There are four levels: health control (HC), converted-MCI(c-MCI), stable- MCI(s-MCI) and Alzheimer's disease (AD) of staging as the table 1 shown. They do the comparison between different two groups. With the closer group (take c-MCI and s-MCI for example), the accuracy will also drop.

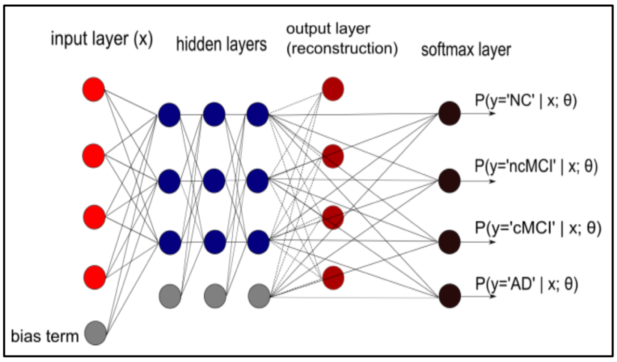


Binary classification results on testing datasets.

		Accuracy	Sensitivity	Specificity
AD *vs* HC	ADNI dataset	99.2%	98.9%	99.5%
	ADNI + Milan dataset	98.2%	98.1%	98.3%
c-MCI *vs* HC	ADNI dataset	87.1%	87.8%	86.5%
	ADNI + Milan dataset	87.7%	87.3%	88.1%
s-MCI *vs* HC	ADNI dataset	76.1%	75.1%	77.1%
	ADNI + Milan dataset	76.4%	75.1%	77.8%
AD *vs* c-MCI	ADNI dataset	75.4%	74.5%	76.4%
	ADNI + Milan dataset	75.8%	74.8%	77.1%
AD *vs* s-MCI	ADNI dataset	85.9%	83.6%	88.3%
	ADNI + Milan dataset	86.3%	84.0%	88.7%
c-MCI *vs* s-MCI	ADNI dataset	75.1%	74.8%	75.3%
	ADNI + Milan dataset	74.9%	75.8%	74.1%

Figure.3 **Table.1**

In the second study, the input layer of the network architecture they designed is the vector of the image, and the hidden layer in the middle is the fully connected layer. Finally, the output layer after the SoftMax output is the probability of four classifications：Normal control(NC), non-converted MCI (nc-MCI), converted MCI(c-MCI) and Alzheimer's disease (AD) as figure 4 shows. Table 2 displays the comparison the accuracy of support vector machine (SVM) and deep learning neural network to distinguish AD/NC and MCI/NC. Table 2 below shows the performance under four categories. As more sub-categories increase, the accuracy will be lower.



Methods	AD vs. NC			MCI vs. NC		
	ACC	SEN	SPE	ACC	SEN	SPE
SK-SVM	84.40	84.64	84.31	76.81	56.14	86.24
MK-SVM	86.42	84.98	87.83	77.25	55.48	87.10
Proposed	87.76	88.57	87.22	76.92	74.29	78.13

Methods	NC	cMCI	cMCI	AD	ACC	SEN	SPE
SK-SVM	40.00	40.00	n 41.67	49.25	43.82	61.96	71.71
MK-SVM	45.00	38.57	52.63	48.36	45.28	70.54	74.03
Proposed	55.43	43.02	31.37	51.96	47.42	65.71	83.75

Figure. 4 **Table.2**

3. Methods

3.1 Data sources

Our training and validation data which has about 6000 MRI T1 images is the open source-dataset from Kaggle. Testing data is from the dataset of Alzheimer's Disease Neuroimaging Initiative (ADNI) which has 14 control people, 18 patients of mild cognitive impairment and 8 patients of Alzheimer's disease. Images from each dataset is shown in figure 5.

Training data

- Non-demented: 2561 images
- Mild-demented: 718 images
- Moderate-demented: 53 images

Validation data

- Non-demented: 641 images
- Mild-demented: 180 images
- Moderate-demented: 13 images

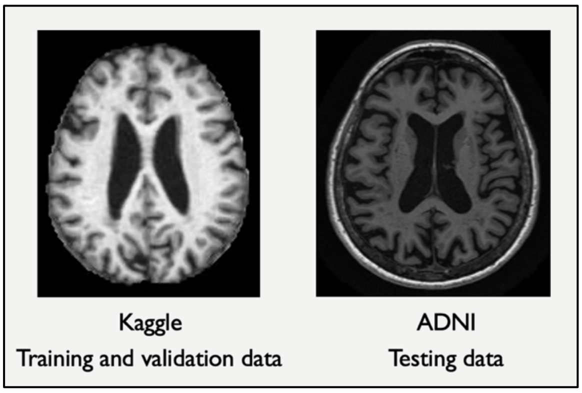


Figure. 5

Testing data

- Normal Control: 14 cases
- Mild Cognitive Impairment: 18 cases
- Alzheimer's Disease: 8 cases

3.2 Imaging processing

For the data from Kaggle as the training and validation data, images are only in axial view. In order to align the information on both two datasets, we also take the axial direction only for testing data from ADNI, and transform the nii file into png file. Because of the data from ADNI has 3-dimensional information and about 160 axial view-images for each case, we ask the doctor to help us select the most representative image.

3.3 Data augmentation

Data Augmentation is a regularization technique that's used to avoid overfitting when training Machine Learning models. In this project, we use the following data augmentation techniques to increase accuracy of training data and testing data.

First, due to the brightness of training data and testing data is inconsistent, we add a random brightness jitter to images. Second, in order to solve the class imbalance problem occurs in our data, we use Synthetic Minority Over-sampling Technique (SMOTE)[5]. SMOTE algorithm is an over-sampling technique focused on generating synthetic data. (fig.6)

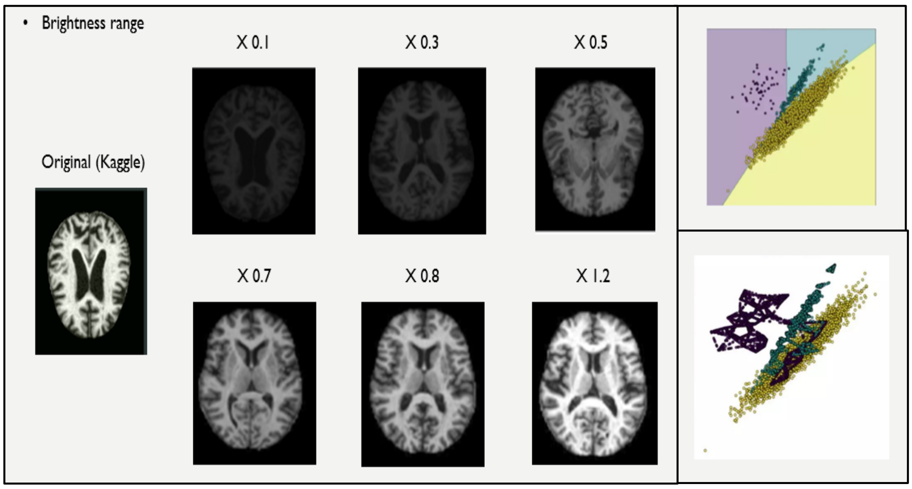


Figure. 6

3.4 Classification

In this project, there will be two main tasks on the classification task, one is binary classification and one is triple categories. To achieve this idea, we correspond non-demented image from Kaggle to normal control from ADNI, mild-demented image to mild cognitive impairment and moderate-demented image to Alzheimer's disease.

We train a 2-classification by using three different models with images of non-demented and moderate-demented to predict the results between normal and AD. For 3-classification, we using images of non-demented, mild-demented and moderate-demented to predict normal, MCI and AD.

4. Models

4.1 DenseNet121

Its basic idea is the same as ResNet, but it establishes a dense connection between all the front layers and the back layers, and its name is derived from this. Another major feature of DenseNet is to realize feature reuse through the connection of features on the channel[6].

DenseNet proposes a more radical dense connection mechanism: that is, all layers are connected to each other. Specifically, each layer will accept all the previous layers as its additional input. The figure below shows the dense connection mechanism of DenseNet. Each layer will be connected with all the previous layers in the channel dimension (the feature map size of each layer is the same here, as will be explained later), and used as the input of the next layer. (fig.7)

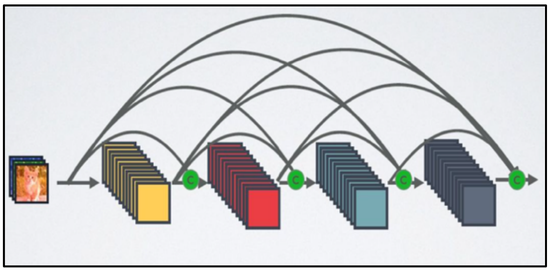


Figure. 7

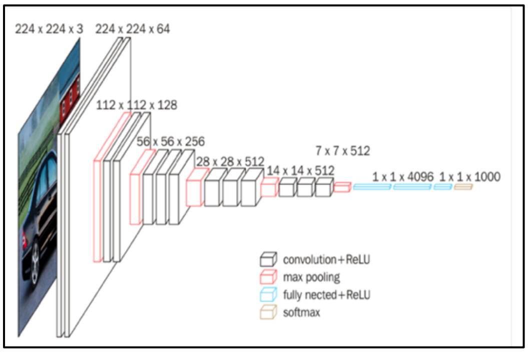


Figure. 8

4.2 VGG16

The VGG Convolutional Neural Network is the network structure of the Oxford University Computer Vision Laboratory participating in the 2014 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) competition, in order to solve the 1000 types of image classification and positioning problems in ImageNet. It can be obtained from the above and right figure that VGG16 has a total of 16 layers, which is a fairly deep convolutional neural network[7]. The structure of various levels of VGG adopts 5 segments of convolution, and each segment has one or more convolutional layers. At the same time, the end of each paragraph is followed by a maximum pooling layer to reduce the image size. (fig.8)

4.3 InceptionV3

One of the most important improvements of Inception V3 is the factorization[8]. The 7x7 convolution is solved into two one-dimensional convolution series (1x7 and 7x1), and the 3x3 convolution is solved into two one-dimensional convolution series (1x3) as figure 9 displays. This can speed up the calculation, and can further increase the depth of the network, increasing the non-linearity of the network. ReLU is required for each additional layer.

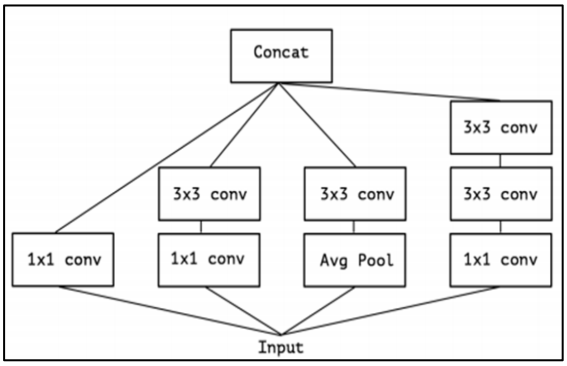


Figure. 9

5. Results

5.1 Training

Our baseline network was InceptionV3. The network was initialized with pre-trained parameters, and RMSprop[9] was used. The hyper-parameters used for our model are as follows: Learning rate = 0.001; Learning rate decay = 0; Loss = Categorical-Cross-entropy; Epoch = 100 and Batch size = 32. Separating the Kaggle dataset into 60% train, 20% validation, 20% testing. We reach accuracy 97.33% on training; 97.01% on validation; 97.14% on testing, which higher about 5% than reference accuracy on Kaggle competitions. (fig.10 and table 3,4)

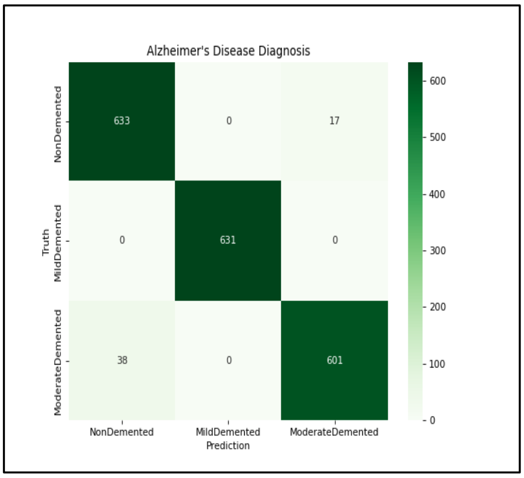


Figure. 10

Parameter	value
Optimizer	RMSprop
Learning rate (Lr)	0.001
Lr decay	0
Loss	CategoricalCrossentropy()
Epoch	100
Batch	32

	Accuracy
Training	97.33%
Validation	97.01%
Testing	97.14%

Table. 3&4

5.2 Binary classification

Three models were used to train and validation the Kaggle dataset and test the ADNI dataset. In binary classification task, these three models and all conditions below can reach accuracy 0.970 to 0.987 in validation data.

We set up four ranges of brightness. First, using VGG16 model, then got all prediction belongings to Non-Demented (Normal). When bright range set up between [0.3, 0.7] using InceptionV3 model, all prediction belongings to Non-Demented (Normal). As we try more wider range of brightness, the prediction belongings to Moderate-Demented (AD) instead. We further try on Densenet121 model, the prediction belongings to Moderate-Demented (AD) in all three bright range. (fig.11)

4.3 Triple classification

In the triple-classification task, the three model network architectures we proposed tend to be classified as Non-Demented (Normal) when the brightness is low. As the brightness range increases, the Densenet121 model and the VGG16 model tend to be classified as Mild-Demented (MCI), while the InceptionV3 model tends to be classified as Moderate-Demented (AD) shown as figure 12.

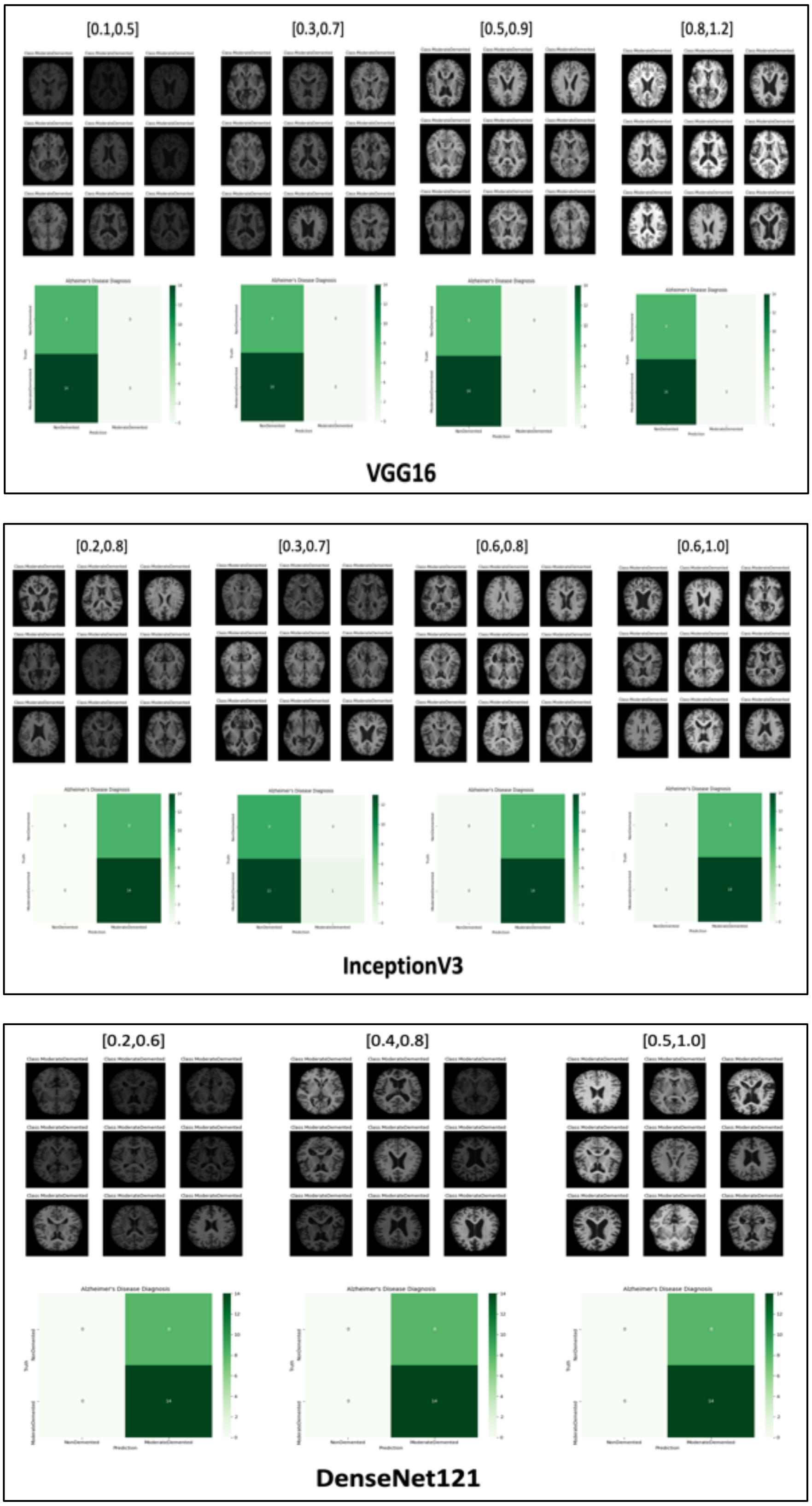


Figure. 11

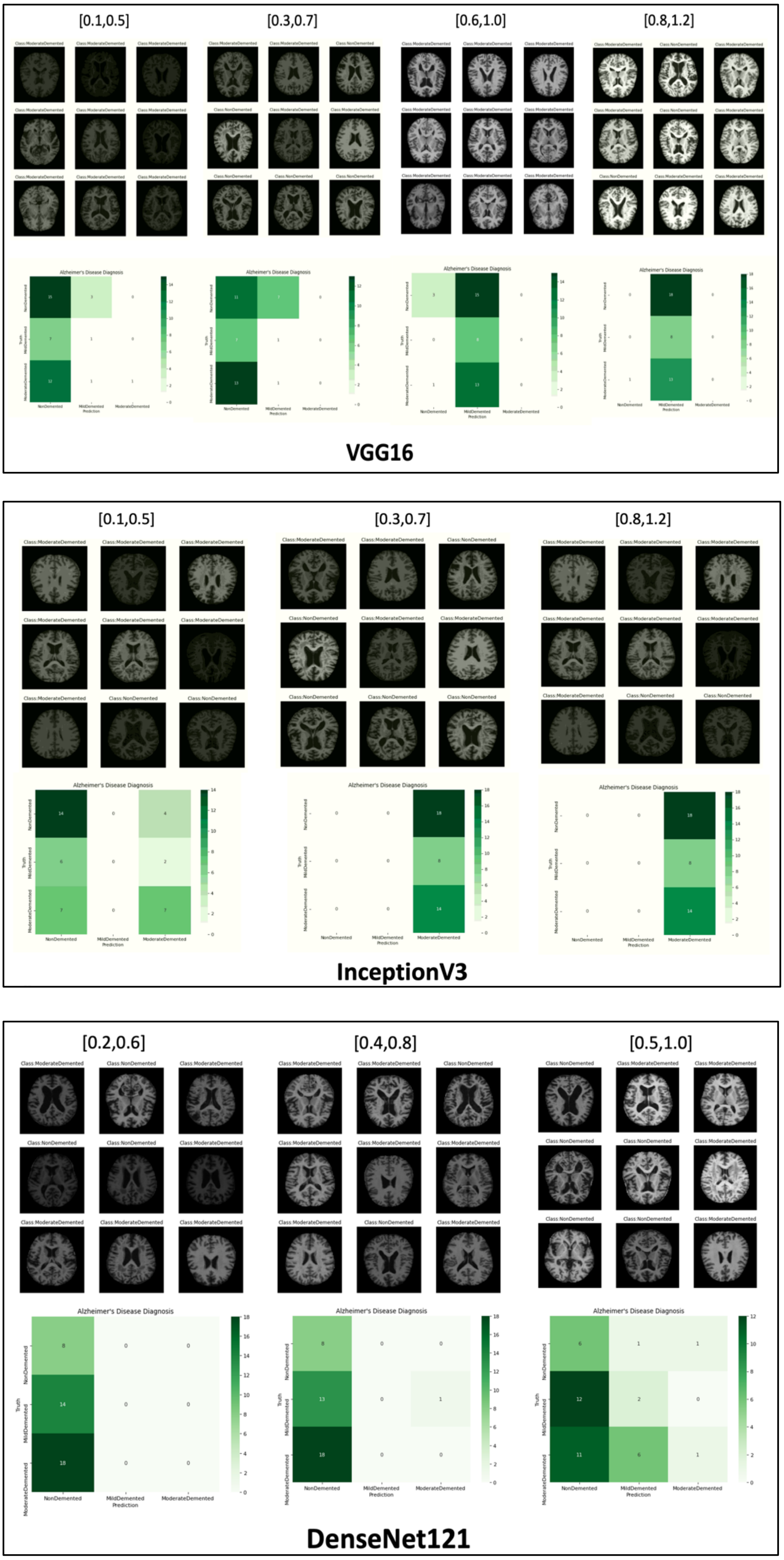

[0.1,0.5]
[0.3,0.7]
[0.6,1.0]
[0.8,1.2]
VGG16
[0.1,0.5]
[0.3,0.7]
[0.8,1.2]
Alzheimer's Disease Diagnosis
InceptionV3
[0.2,0.6]
[0.4,0.8]
[0.5,1.0]
Alzheimer's Disease Diagnosis
NonDemented
MildDemented
ModerateDemented
Prediction
Truth
DenseNet121

Figure. 12

6. Discussion

Through data augmentation in training the Kaggle dataset, we made an improvement that reaching an accuracy of 97%. Therefore, the model can perform well on both the Kaggle dataset and the ADNI dataset at the same time.

Summarize above the experiment, images from two dataset are different from their brightness by visualizing. After we set different range of the brightness of input-images, each model did not follow a same significant trend to predict the classification. These three models predict on their own way. From these results it can be seen that brightness may not be the only reason that affects the prediction. There must exist other effects between Kaggle dataset and ADNI dataset, which is not that simple.

7. Conclusion

In conclusion, despite the good training results from DenseNet121, VGG16 and InceptionV3, the testing results are very different from each. We speculated that these three models learn unequal features (green part in figure 13) to determine the image belong to which classification. In other words, the features they capture may have bias but sufficient to cope with their training process that up to 97% accuracy. When switching to the testing data, they make interpretations based on the results of their respective training and giving totally different testing results.

To sum up, does a model really learn the representative feature during training process?

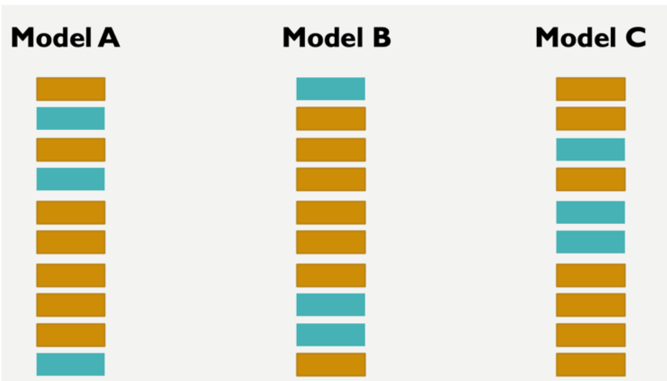


Figure. 13

8. <u>Reference</u>

[1] https://faculty.sites.uci.edu/myassa/research/developing-biomarkers-for-preclinical-ad/

[2] https://www.evroclinic.com/lechenie-za-rubezhom/lechenie-v-germanii/oblasti-meditsiny/nevrologiya/vidy-boleznej/bolezn-pika

[3] Silvia Basaiaa, Federica Agostaa, Luca Wagnerc. *Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks.* NeuroImage: Clinical 21 (2019) 101645

[4] Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., & Feng, D. E*arly diagnosis of Alzheimer's disease with deep learning. In 2014 IEEE 11th international symposium on biomedical imaging (ISBI) (*2014, April). (pp. 1015-1018).

[5] Chawla, Nitesh V., et al. "*SMOTE: synthetic minority over-sampling technique." J*ournal of artificial intelligence research 16 (2002): 321-357.

[6] Huang, Gao, et al. "*Densely connected convolutional networks."* Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[7] Simonyan, Karen, and Andrew Zisserman. "*Very deep convolutional networks for large-scale image recognition."* arXiv preprint arXiv:1409.1556 (2014).

[8] Szegedy, Christian, et al. "*Rethinking the inception architecture for computer vision."* Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[9] Ruder, Sebastian. "*An overview of gradient descent optimization algorithms."* arXiv preprint arXiv:1609.04747 (2016).