A Parallel Deep Convolutional Neural Network for Alzheimer's disease classification on PET/CT brain images

Husnu Baris Baydargil[1], Jangsik Park[1*], Do-Young Kang[2], Hyun Kang[3] and Kook Cho[4]
[1] Department of Electric Electronic and Communication Engineering, Kyungsung University
309 Suyeong-ro, Daeyeon-dong, Nam-gu, Busan, Korea
[e-mail: baris.baydargil@gmail.com, jsipark@ks.ac.kr]
[2] Department of Nuclear Medicine, Dong-a University College of Medicine, Dong-A University Hospital
#26, Daesingongwon-ro, Seo-gu, Busan, Korea
[e-mail: dykang@dau.ac.kr]
[3]Institute of Convergence Bio-Health, Dong-A University, #26, Daesingongwon-ro, Seo-gu, 49201 , Busan, Korea
[e-mail: kang2000h@gmail.com]
[4]College of General Education, Dong-A University, Busan, Korea
[e-mail: kcho@dau.ac.kr
*Corresponding author: Jangsik Park

Received October 23, 2019; revised January 3, 2020; accepted July 29, 2020; published September 30, 2020

Abstract

In this paper, a parallel deep learning model using a convolutional neural network and a dilated convolutional neural network is proposed to classify Alzheimer's disease with high accuracy in PET/CT images. The developed model consists of two pipelines, a conventional CNN pipeline, and a dilated convolution pipeline. An input image is sent through both pipelines, and at the end of both pipelines, extracted features are concatenated and used for classifying Alzheimer's disease. Complimentary abilities of both networks provide better overall accuracy than single conventional CNNs in the dataset. Moreover, instead of performing binary classification, the proposed model performs three-class classification being Alzheimer's disease, mild cognitive impairment, and normal control. Using the data received from Dong-a University, the model performs classification detecting Alzheimer's disease with an accuracy of up to 95.51%.

Keywords: Computer vision, Deep learning, Convolutional neural networks, Parallel model, Image classification, Alzheimer's disease

This research was supported by the National Research Foundation (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2018R1A2B2008178) and Brain Busan 21+.

1. Introduction

Alzheimer's disease (AD) is statistically the most common type of dementia generally seen in elderly people; it makes up about 70% of the cases in individuals who are 70 years of age or higher [1]. In most people, AD late-onset symptoms begin to appear in their mid-60s, while the early-onset occurs between 30s and mid-60s although the latter case is quite rare. It is also estimated that around 640 million people will be diagnosed by 2050 due to the increase in the aging population [2]. During the progress of the disorder, the brain structure progressively]changes; while the initial damage occurs in the hippocampus. Mild cognitive impairment (MCI) can be considered the early sign of AD, however not everyone with MCI may develop into the AD. As the disease progresses, hard plaques and tangles occur in the brain, causing debilitating issues such as progressive memory loss and at later stages, inability to move. AD also shrinks the hippocampus and cerebral cortex, while enlarging ventricles. Since the hippocampus is the part responsible for episodic and spatial memory and also works as a relay between the brain and the rest of the body, neurons cannot communicate through synapses. This causes issues with remembering, thinking, planning, and judgment. The effects of this progress can be seen as a low-intensity brain cell in medical imaging. As of writing this paper, there is no definite cure for AD, and current treatments help only with prolonging life and alleviating symptoms.

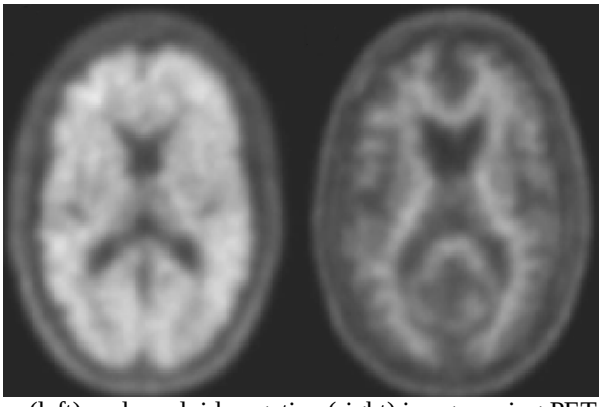

Fig. 1. Amyloid-positive (left) and amyloid-negative (right) images using PET scans are used in the diagnosis of Alzheimer's disease.

As the AD progresses, protein plaques in the brain named amyloid-β (Aβ) and hyperphosphorylated tau and lead to progressive neuron and axon damage. The changes generally follow affecting the early medial temporal lobe (entorhinal cortex and hippocampus), followed by progressive neocortical damage [3]. These changes start to appear years before AD becomes prevalent. It appears that the symptoms of AD start to develop due to the toxic effects of the plaques in the brain achieve a certain clinical threshold. For that reason, this paper's work focuses on PET/CT (Positron Emission Tomography/Computed Tomography) images where the prevalence of these plaques is highlighted in patients.

Since the early 1970s, researchers developed several computer-aided diagnostics systems that were based on manually subtracted feature vectors [4]. Afterward, these vectors were trained in a supervised manner to a machine learning model such as a support vector machine to perform classification. However, it was soon realized that there were serious limitations and shortcomings to such systems [5]. Recognizing these limitations, researchers turned towards data mining approaches in the 1980s and 1990s in hopes of developing more advanced and flexible systems.

Nowadays, the performance of deep learning models is being compared with human experts] in computer-assisted-diagnosis (CAD). Due to deep learning methodologies such as convolutional neural networks (CNN) were able to extract meaningful features from a given image automatically through supervised training and produce highly accurate results, the possibility of human experts being replaced with automated systems in near future is being discussed in the medical field due to the reduced cost and relatively similar performance [6]. Deep learning models show promising results for other medical image analysis tasks such as medical segmentation, disease, and tumor detection, and classification, breast, bone, etc. using microscopy, and ultrasound images as well.

In this paper, a parallel pipeline CNN model that uses both convolution operation and dilated convolution operation is proposed. The model extracts two distinct, but complementary image features that are spatially complementary to each other for each given brain image. One pipeline consists of conventional layers, whereas the other pipeline consists of dilated convolutional layers. The extracted features coming from two pipelines are concatenated to be used in the classification stage The proposed model is trained with the dataset that is obtained from Dong-A University Department of Nuclear Medicine, using PET/CT images.

Paper is written in a format such that section 2 discusses related work on AD detection in brain imaging. Section 3 explains the proposed model. Section 4 shows the experimental results and considerations. Section 5 concludes the paper and discusses future work.

2. Alzheimer's Disease Detection with Computer Vision

There has been a massive effort in the field of medical imaging to better observe the changes in the brain the AD causes through time. There have been machine learning applications proposed to classify different stages of AD using imaging data [7, 9]. Structural differences in the brain were identified through such research, to show the difference between a healthy brain and a brain that's affected by AD. There is also a strong correlation between brain connectivity and the AD patient's behavior [10]. The degeneration of the brain cells caused by AD can be seen using a variety of imaging techniques, such as PET/CT, structural and functional magnetic resonance imaging (sMRI, fMRI), diffusion tensor imaging (DTI), etc. These imaging techniques were used in AD diagnosis research such as PET [11], sMRI [12], fMRI [13]. It's been also shown that the features combined from multiple modalities improve the accuracy. According to the results of one study, the accuracy of classification in the AD case has increased from 72% to 93.2% when a multi-modal is used instead of an individual modal [14].

18F-florbetaben (FBB) PET imaging is useful in providing information about Aβ deposition of the brain of a patient with AD. Even though a variety of medical imaging techniques are widely used for AD research, clinical testing is still the default tool in AD diagnosis. However, regarding the prodromal AD and MCI, accurate diagnosis is challenging from the current existing clinical examination, which prompted including biomarkers such as Aβ and/or tau into AD diagnostics. The development of radiotracers to visualize Aβ plaques in the brain has become an active area of AD research. In this work, FBB brain PET imaging data was used to develop and validate the proposed model.

PET/CT is a combination of the cross-sectional anatomic information obtained through CT and the metabolic information obtained through PET and CT, and the sequential images from both devices are combined into a single superposed image. This offers advantages over single CT or PET alone, such as its ability to accurately localize the increased FBB activity in

abnormal locations which may be impossible to localize with PET alone with the complement of lack of anatomical information and CT-based attenuation map. The work so far in the literature shows that PET images provide better classification accuracy [15] than MRI images [16, 17] alone.

Deep learning models are used in so many areas such as object recognition, natural language processing, object tracking, segmentation, etc. due to their extraordinary feature extraction capabilities from input data automatically. Their hierarchical and deep structures allow them to learn high-level features, and further down the pipeline, more abstract features are obtained. These models are also capable of learning certain disease-caused features in the brain from images in their hidden or latent representations. Therefore, it didn't take much time to apply and develop such models to the field of medical imaging for a more accurate diagnosis of disorders and diseases. Researchers used sparse autoencoder for multi-class classification using 2D CNN [18], and 3D CNN [19]. Brosh et. al. [20] developed a deep belief network (DBN) using manifold learning to detect AD. Suk et. al. [21 – 23] developed an autoencoder that incorporated SVM kernels using magnetic current imaging (MCI) for classification. Cárdenas-Peña et. al. [24] developed a model using kernel alignment and showed that supervised pre-training of stacked autoencoder provides higher classification accuracy than unsupervised pre-training with plain autoencoders and principal component analysis (PCA).

Unfortunately, AD is only classified at its later stage, and earlier detection may only help slow the progression of cognitive decline, so it's vital to be able to detect AD at its earlier stages. In this paper, a new proposed very deep parallel CNN which takes in three classes AD, MCI, and normal control (NC), and is capable of extracting different features of the same input data, concatenate these features and provide higher classification accuracy. The dataset used was provided with the collaborating Dong-a University Department of Nuclear Medicine.

3. The Proposed Parallel Model for Classification

3.1 Deep Learning Model

The proposed model is a parallel deep CNN that is designed to extract different abstract color-based spatial information using convolutional and dilated convolutional layers, and concatenate these extracted abstract information to achieve higher accuracy. While one pipeline is a standard 8-layer-deep CNN with normal convolutions, the other pipeline is an 8-layer-deep CNN with dilated convolutions [25]. Both parallel pipelines consist of operations such as Convolution or dilated convolution, Batch Normalization [26], Rectified Linear Units (ReLU) activation, max-pooling, and dropout [27].

Both first four layers of the network perform the first four operations respectively, while the eighth layer also performs dropout. After the computations are done, the features are flattened in the form of a dense layer similar to Huang et. al.'s work [28] and produce a dense block concatenating both of the layers from both pipelines. An example of respective operations can be seen in **Fig. 2**.

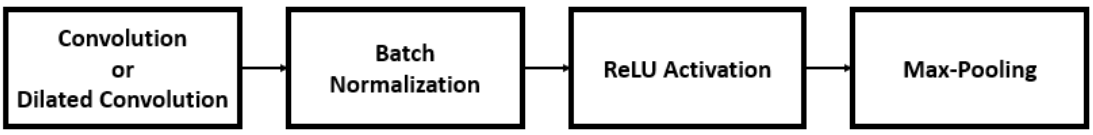

Fig. 2. Standard operations performed by the model down both pipelines.

Regular convolution operation, while it's very effective, generally struggle to integrate global context to its activations. If one considers a pure CNN with $k \times k$ convolutions without pooling, then the receptive field of each unit is which affects the activation is $l \times (k-1) + k$ with l being the layer index. Therefore, it can be seen that the effective receptive field grows only linearly through the layers. This is limiting when it comes to higher resolution input images. Dilated convolution, however, works different in a way where between signal f and kernel k with dilation factor l is defined as,

$$(k_{*l}f)_t = \sum_{\tau=-\infty}^{\infty} k_\tau \cdot f_{t-l\tau} \tag{1}$$

where $f_{t-l\tau}$ differs from normal convolution where it would be $f_{t-\tau}$. In the dilated convolution operation, the kernel comes into contact with the signal every l^{th} entry, which can also be extended for 2D convolutions. This makes dilated convolution operation computationally more efficient than convolution operation. In the proposed 4-layer-deep model, the conventional convolution pipeline has roughly 60 million trainable parameters, and the dilated convolution pipeline has around 27 million trainable parameters. This equals to around 25% more computationally lightweight pipeline for the dilated convolution pipeline compared with a pipeline that has the same perceptive field using 5×5 convolutional filters.

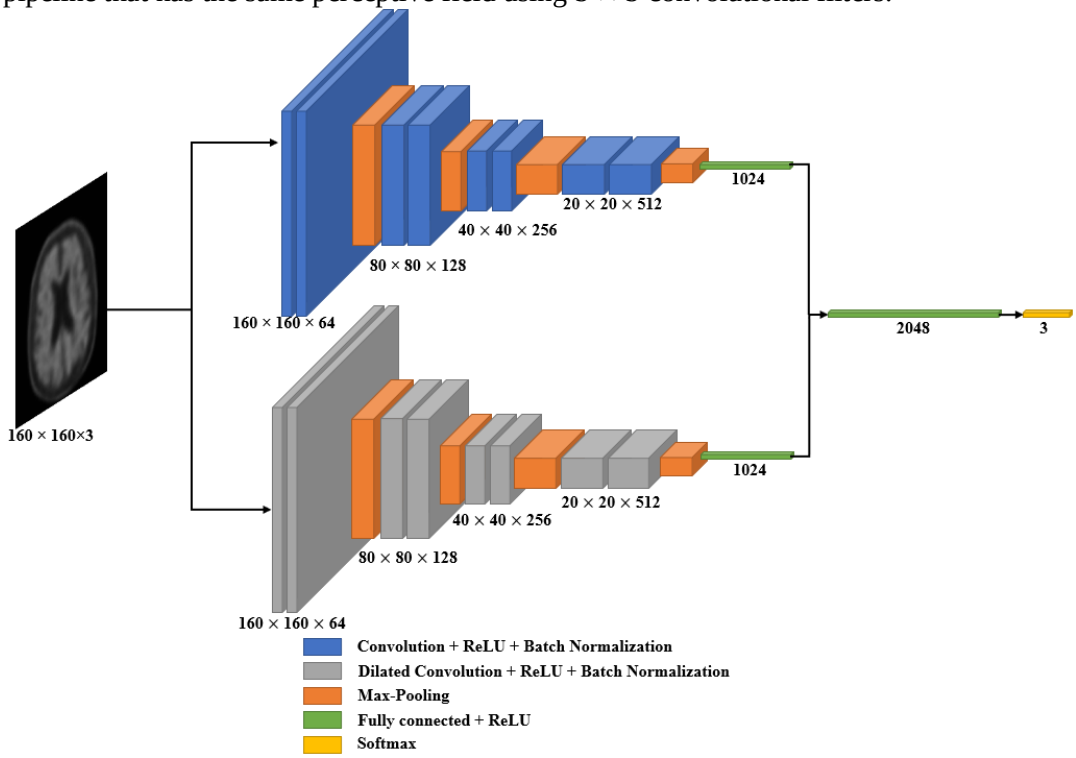

Fig. 3. The architecture of the proposed model. The top pipeline is a regular CNN with eight layers with each specific layer operations highlighted in **Fig. 2**. The bottom pipeline is dilated CNN with eight identical layers except that in this pipeline, dilation operation is used. In both pipelines, fully-connected layers are obtained, concatenated together, and this layer is connected to the classification layer with softmax operation.

Both convolutional and dilated convolutional filters used in pipelines have a size of 3×3, while the dilated convolutional filter also has a dilation size of 2. As the weights and other parameters are constantly changed during training, sometimes the data is affected by being too big or too small, which is a phenomenon called the internal covariance shift. Batch normalization is used to decrease the training time and fixes this issue. Max-pooling with a 2×2 size is used for both pipelines.

Due to its nature, dilated convolution increases the receptive view of the network exponentially depending on the dilation size. In this work, this fact is taken advantage of, as this makes the network lighter and captures different features than plain convolution operation of similar kernel size compared to 5×5 convolutional layer.

The proposed model takes in the initial input with size $160 \times 160 \times 3$, and after each max-pooling operation, the image size decreases by half while the depth doubles. This operation is done four times until there is a 1024-unit fully-connected layer for each pipeline. Concatenating these fully-connected layers gives us a new fully-connected layer with a unit size of 2048, with both pipelines' extracted features into one. This layer is connected to the classification layer with softmax operation, where it's defined as,

$$p_i = \frac{\exp(f_i)}{\sum_i \exp(f_i)}, i = 1, \dots, m \tag{2}$$

where f_i is the representation, p_i is the probability score, and m is the number of classes, which in this case is 3 (AD, MCI, and NC). With this information, L is obtained such that,

$$\mathrm{L} = -\sum_i t_i \, log(p_i) \tag{3}$$

where L is the cross-entropy loss of the model. This value is used during the backpropagation, where gradients are calculated. Say, the PET/CT image data is called t_i, then,

$$\frac{\partial \mathrm{L}}{\partial \mathrm{f}_i} = p_i - t_i \tag{4}$$

Stochastic gradient descent with momentum optimization algorithm was used for the proposed model. The training data is also split with training and cross-validation in 8:2 proportions. Nesterov momentum was also used for speeding up the training and improve convergence. It works as, given the objective function $f(\alpha)$ is to be minimized, Nesterov momentum is given as,

$$v_t = \mu v_{t-1} - \varepsilon \nabla f(\alpha_{t-1} + \mu v_{t-1}) \tag{5}$$

$$\alpha_t = \alpha_{t-1} + v_t \tag{6}$$

where v_t is the velocity, $\varepsilon > 0$ is the learning rate, $\mu \in [0, 1]$ is the momentum coefficient, and $\nabla f(\alpha_t)$ is the gradient at α_t.

Let x=$\{x_i, i=1,\dots,N\}$ be a set of PET/CT images with $x_i \in [0,1,2, \dots, L-1]^{(h \times w \times l)}$, a 3D image with L grayscale values, $h \times w \times l$ voxels and y $\in \{0, 1, 2\}$ the classes in the dataset (AD, MCI, and NC). Next, construct a classifier *f*,

$$f: X \rightarrow Y;\ \mathrm{x} \rightarrow \mathrm{y} \tag{7}$$

that predicts a label y according to an input x with a minimum error rate. We would like to determine this classifier f through an optimal parameter set $\mathrm{w} \in \mathbb{R}^P$ (where P could easily be millions) through minimizing the error rate during the prediction. Therefore, training is an iterative process of finding the set of parameters w, while minimizing the classifier loss,

$$L(w, X) = \frac{1}{n}\sum_{i=1}^{n} l(f(x_i, w), \hat{c}_i) \tag{8}$$

where x_i is the ith image of X, $f(x_i, w)$ is the classifier function predicting the class c_i of x_i in a given w, $\hat{c}_i$ the ground truth class for the specific image, and $l(c_i, \hat{c}_i)$ is the loss function for the wrong prediction ($\hat{c}_i$ instead of c_i). Then we set L for the cross-entropy loss.

$$L = \sum_{i} \hat{c}_i \ log\ c_i \tag{9}$$

3.2 Dataset

In this paper, PET/CT images dataset is the property of Dong-a University Hospital Department of Nuclear Medicine. There are three separate classes, AD, MCI, and NC. The detailed information about the dataset is given in **Table 1**.
In a given PET/CT scan of a person's brain, there were 110 axial images. However, the top 30, and bottom 30 slices provided no useful information about plaque formations in the brain information about classification, so these images were discarded. Therefore, 50 images from each patient's brain were used for training and testing. Each image size is $160 \times 160 \times 3$.

Table 1. The dataset information with the total number of images and the total number of patients for each designated class.

Classes	Total Number of Images	Total Number of Patients
AD	7,050	141
MCI	5,250	105
NC	3,500	70
Dataset	15,800	316

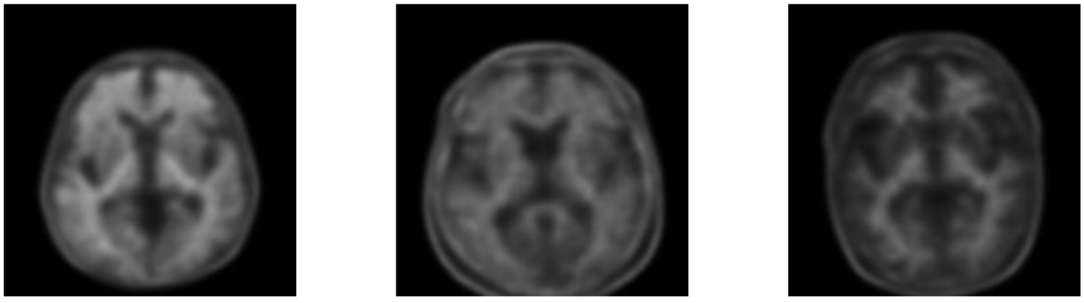

Fig. 4. Sample images from the dataset. Left image AD, middle one MCI, and the right image is the NC case. Image slices were taken from the same slice number of different patient's images (60th slice out of 110).

The human brain in real life is a 3-dimensional organ, however, the medical images only represent one plane, which is the axial plane. Other axes, namely sagittal and coronal axes also carry spatial information in their relevant planes about plaque formations in the brain, which can be used to train the model. Thus, in this work, the proposed model was also tested to see how it would perform with other axes of the brain, sagittal, and coronal. Using the bicubic interpolation method, sagittal and coronal axes were created. For these axes, initial 15 and the last 15 images were discarded for these images did not carry any useful information about. This means for one patient, out of 110 created images, 80 training images contain useful information about AD for both sagittal and coronal axes. The flowchart of the followed method can be seen in **Fig. 5**. The number of images obtained through this process is shown in **Table 2**. An example of created coronal and sagittal images can be seen in **Fig. 6** and **Fig. 7**, respectively.

For a specific set of data points $(x_k, y_k), k = 0{:}N$ cubic spline consists of N cubic polynomial $s_k(x)$, s assigned to each subinterval classifying given constraints. The cubic spline is defined as a function $S(x) = s_j(x)$ on an interval $[x_j, x_{j+1}]$ for $n = 0, 1, \dots, n-1$ defined as,

$$S_j(x) = a_J + b_J(x - x_J) + c_J(x - x_J)^2 + d_J(x - x_J)^3 \tag{10}$$

In deep learning, data augmentation is defined as enlarging the dataset through different means without affecting possible obtainable features and preventing overfitting [29]. Since adding more data in the medical imaging domain would prove costly, augmentation is the most reliable and easiest way to increase the image number in the dataset. Horizontal flip, width shift, height shift operations were used to increase the dataset size. Each image is randomly augmented through one or more of the mentioned methods to create a second image, doubling the total number of images in the dataset. The final number of total images after the data augmentation used in the training is given in Table 3. 20% of the data was used for testing and 10% of the data for validation. The augmentation process was applied separately for training, testing, and validation sets.

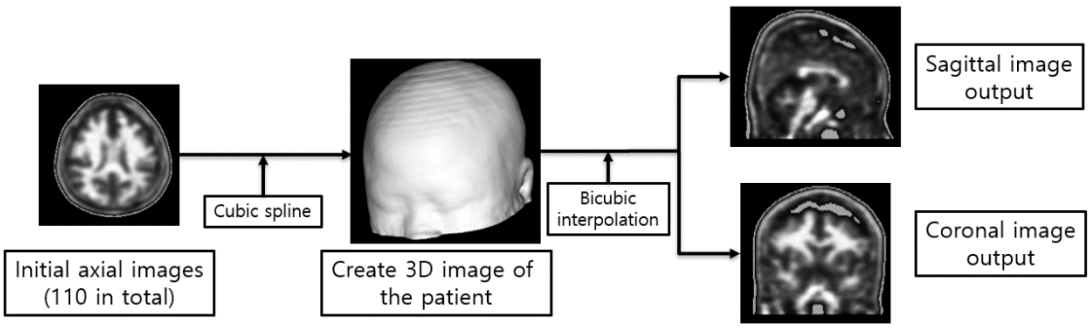

Fig. 5. Flowchart of the method for creating sagittal and coronal axes using the axial images.

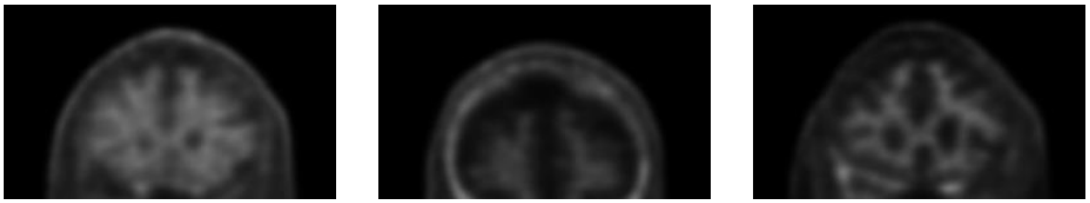

Fig. 6. An example of created coronal images with AD, MCI, and NC, respectively. The slices are taken from the same i^{th} image of a person.

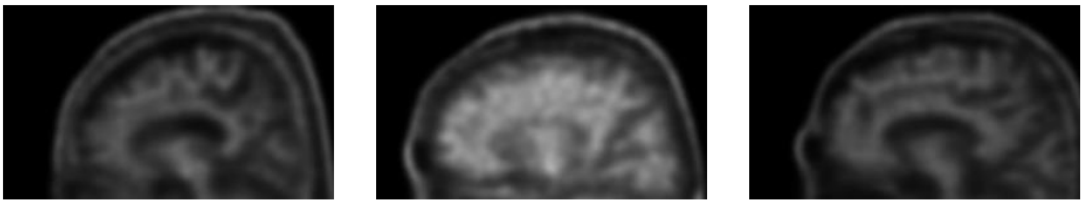

Fig. 7. An example of created sagittal images with AD, MCI, and NC, respectively. The slices are taken from the same i^{th} image of a person.

Table 2. The total number of images that are used for training images after the bicubic interpolation process.

Axes	AD	MCI	NC	TOTAL
Axial	7,050	5,250	3,500	15,800
Coronal	11,280	8,400	5,600	25,280
Sagittal	11,280	8,400	5,600	25,280

Table 3. The total number of images in the dataset after the data augmentation is applied.

Axes	AD	MCI	NC	TOTAL
Axial	14,100	10,500	7,000	31,600
Coronal	22,560	16,800	11,200	50,560
Sagittal	22,560	16,800	11,200	50,560

4. Experimental Results and Considerations

4.1 Experimental Environment and Training

Implementation was done using Tensorflow [30], Keras [31], Python 2.7 in Ubuntu OS 16.04 computer with Intel Xeon E5-2650 v3 CPU, NVIDIA Quadro M5000 dual-GPU, and 64GB RAM. We used SGD with a mini-batch size of 30, a learning rate of 0.001, a weight decay of 0.06, and a momentum of 0.9 with Nesterov optimization. The model's accuracy and loss graph change during the training of sagittal images can be seen in **Fig. 8**.

Small fluctuations that are seen in **Fig. 8** are due to the SGD optimization algorithm trying to reduce the loss value obtained after a forward-propagation pass, by updating the weights in a backward-propagation. It's seen that after 120 epochs, the proposed model was sufficiently

trained with a validation accuracy close to 100%. It can also be seen that both validation and training loss dropped below the value of 0.002 after 120 epochs of training. It would be possible to reduce the loss value even further, however, this also increases the risk of overfitting over the dataset instead of generalizing. Overfitting means the model has memorized the dataset, and it would mean that it would perform less accurately over any other image that it has not been trained with.

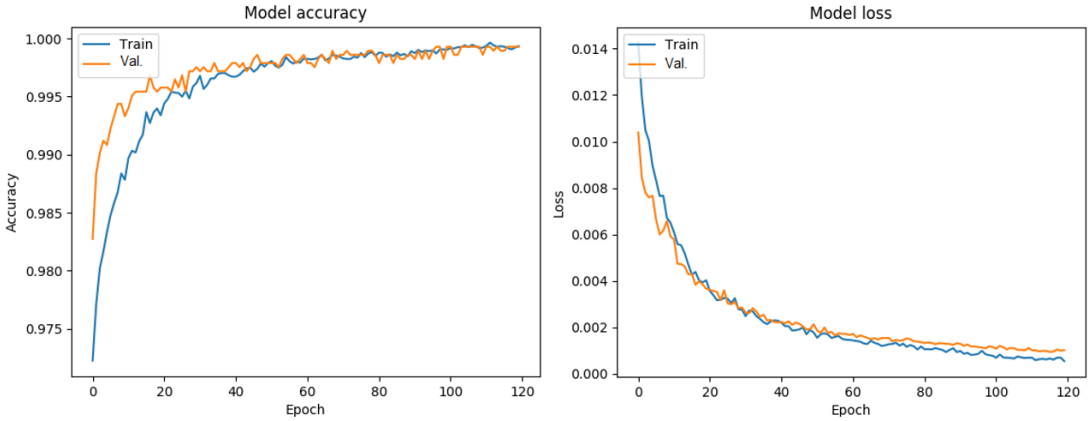

Fig. 8. The model's accuracy and loss value change through the epochs during the training. It can be seen that 120 epochs were sufficient for optimized accuracy and loss value.

4.2 Performance of the model

The performance of the model during inference is of utmost importance, and activation maps give clues about what the model has learned. Activation maps in deep learning help understanding where the model's attention is at, in given visualizing data. These fine-grained details in images are crucial in computer vision tasks as the model tries to interpret this extracted information in its task; in this work, classification. As given in **Fig. 9**, **Fig. 10**, and **Fig. 11**, the model's pipeline attention maps can be seen. On each image, the left side is the convolutional pipeline activation map, and on the left side is the dilated convolutional pipeline map. It can be seen in the images that while the convolutional pipeline activations are more focused on narrower areas, dilated convolution pipeline activations are more focused on wider areas, such as the circumference of the brain.

Complementary nature of the pipeline activations points to the fact that more thorough and accurate feature extraction is possible through the proposed model. In all three cases for AD, MCI, NC, convolutional pipeline, and the dilated pipeline extracted features that the other could not obtain. Throughout the pipelines, these features combined led the proposed model to achieve higher accuracy than other models.

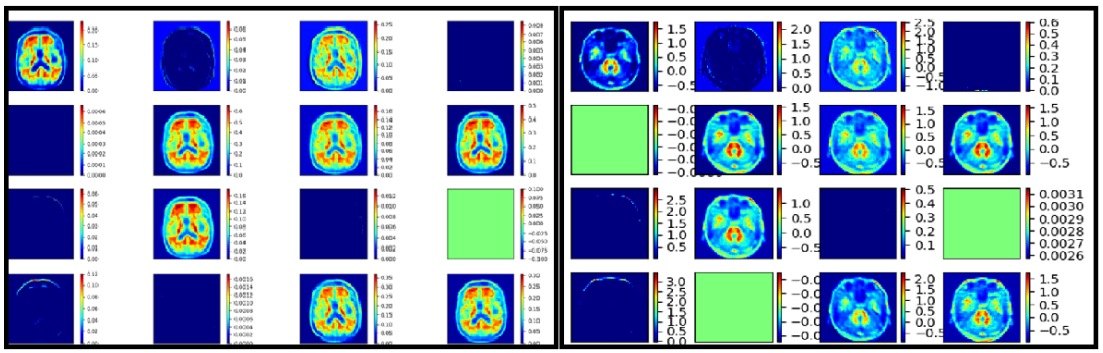

Fig. 9. Activation maps of the model for a test image for the AD case. The left box is the convolutional pipeline, and the right box is the dilated convolution pipeline. It can be seen that the activated areas for the test image complement each other; meaning, the dilated pipeline can capture features not obtained by convolutional pipeline, and vice versa.

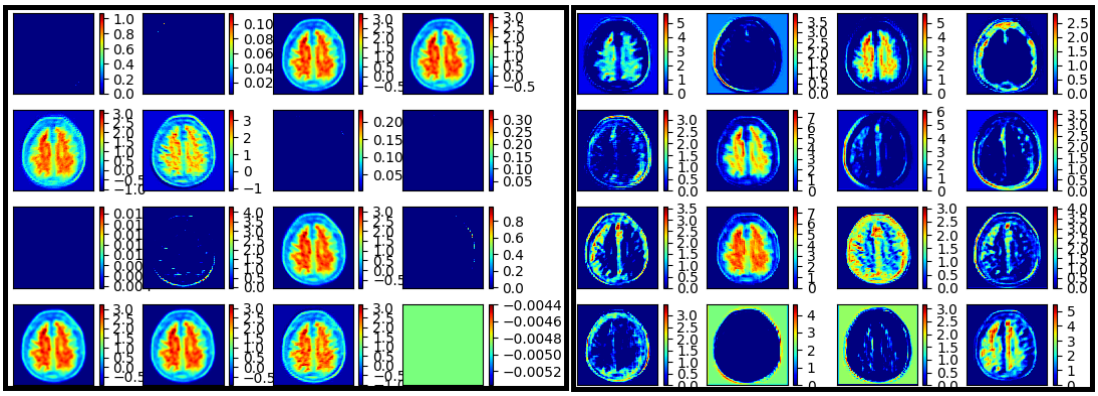

Fig. 10. Activation maps of the model for a test image for the MCI case.

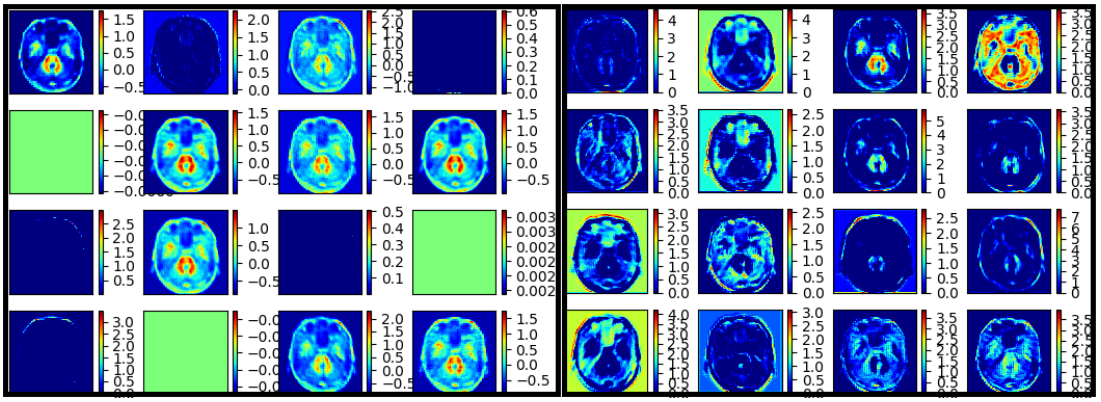

Fig. 11. Activation maps of the model for a test image for the NC case.

4.3 Results

The proposed model was compared with other well-known state-of-the-art models such as VGG16 [32], GoogLeNet Inception v4 [33], ResNet50 [34], and a sparse autoencoder specifically developed for AD classification in [35]. The results show that the proposed model outperforms all the comparison models by a large margin. The results can be seen in **Table 4**.

The total number of trainable parameters and training time in benchmark models and the proposed model is given in **Table 5**. Even though the model has a high number of trainable parameters, it shows higher accuracy than the benchmark models. In model inference, the accuracy of the model is calculated as,

$$Accuracy = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{113}$$

The proposed model was able to extract more useful information from created sagittal images than the original axial images. This means that the spatial information carried in the sagittal plane is found to be more useful for the proposed model to perform classification more accurately.

Table 4. Accuracy comparison between the developed model and other state-of-the-art classification models.

Models	Axial	Coronal	Sagittal	Average
GoogLeNet (InceptionV4)	83.72%	82.38%	83.77%	83.29%
VGG16	91.33%	91.42%	92.24%	91.66%
Sparse Autoencoder	93.37%	93.04%	93.54%	93.31%
ResNet50	89.41%	90.33%	92.25%	90.66%
Developed Model	**94.72%**	**94.09%**	**95.51%**	**94.77%**

Table 5. Trainable parameter comparison between the proposed model and the benchmark models. The model has a large number of parameters.

Model	Trainable Parameters	Training Time
GoogLeNet (Inception v4)	6M	1h45m
ResNet50	23M	3h32m
Sparse Autoencoder	68M	6h58m
VGG16	109M	8h13m
Proposed Model	87M	7h33m

5. Conclusion

In this paper, a parallel deep learning CNN model is proposed that is both capable of performing accurately in a given dataset, and computationally more efficient than similar very deep models such as VGG16. Convolutional and dilated convolutional pipelines work complementary with each other in extracting features the other pipeline cannot, which provides higher accuracy. Although the majority of the works in the literature focus on binary classification, in this paper multi-class classification was focused on, for the future development of AD diagnosis. Sagittal and coronal plane images were also tested for relevant spatial information, which showed that the sagittal plane provided with the highest accuracy score with 91.51%. Even though the model has been tested for the AD diagnosis only, it's believed that it can be used for other medical imaging areas, as well as modifications for further use given its simplistic yet robust approach to the extraction of image features. The focus of future work will be on increasing the accuracy as well as lowering the computational cost of the model.

References

[1] Alzheimer's Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 15, pp. 321-387, 2019. Article (CrossRef Link).

[2] R. Brookmeyer, E. Johnson, K. Ziegler-Graham and H.M. Arrighi, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186-191, July 2007. Article (CrossRef Link).

[3] G. Frisoni, N.C. Fox, C. Jack, P. Scheltens and P. Thompson, "The clinical use of structural MRI in Alzheimer's disease," *Nature Reviews Neurology*, vol. 6, no.3, pp. 67-77, February 2010. Article (CrossRef Link)

[4] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. Van Ginneken and C.I. Sanchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, December 2017. Article (CrossRef Link)

[5] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Systems with Applications*, vol. 138, pp. 112821, December 2019. Article (CrossRef Link)

[6] A. Tufail, C. Rudisill. C. Egan, V.V. Kapetanakis, S. Vega-Salas, C.G. Owen, A. Lee, V. Louw, J. Anderson, G. Liew, L. Bolter, S. Srivinas, M. Nittala, S. Sadda, P. Taylor, and A.R. Rudnicka, "Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders," *Ophthalmology*, Vol. 124, no. 3, pp. 343-351, March 2017. Article (CrossRef Link)

[7] Y. Fan, S.M. Resnick, X. Wu and C. Davatzikos, "Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study," *Neuroimage*, vol. 41, no. 2, pp. 277-285, June 2008. Article (CrossRef Link)

[8] K. Hu, Y. Wang, K. Chen, L. Hou and X. Zhang, "Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis," *Neurocomputing*, vol. 175, pp. 132-145, October 2016. Article (CrossRef Link)

[9] R. Filiphovych, and C. Davatzikos and Initiative ADN et. al., "Semi-supervised pattern classification of medical images: application to mild cognitive impairment (MCI)," *Neuroimage*, vol. 55, no. 3, pp. 1109-1119, April 2011. Article (CrossRef Link)

[10] A.K. Ambastha, "Neuroanatomical characterization of Alzheimer's disease using deep learning," *Journal of National University of Singapore*, October 2015.

[11] K.R. Gray, R. Wolz, R.A. Heckermann, P. Aljabar, A. Hammers, D. Rueckert and Initiative ADN et. al., "Multi-region analysis of longitudinal FDG-PET for classification for the Alzheimer's disease," *Neuroimage*, vol. 60, no. 1, pp. 221-229, March 2012. Article (CrossRef Link)

[12] C. Davatzikos, P. Bhatt, L.M. Shaw, K.N. Batmangehelich and J.Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiology of Aging*, vol. 32 no.12, pp. 2322.e19-2322.e27, December 2011. Article (CrossRef Link)

[13] H, Suk, C.Y. Wee and D. Shen, "Discriminative group sparse representation for mild cognitive impairment classification," in *Proc. of 4th Int. work-shop on machine learning in medical imaging*, vol. 8134, pp.131-138, September 2013. Article (CrossRef Link)

[14] R.J. Perrin, A.M. Fagan and D.M. Holtzman, "Multimodal techniques for diagnosis and prognosis of Alzheimer's disease," *Nature*, vol. 461, no. 7266, pp. 916-922, October 2009. Article (CrossRef Link)

[15] H. Choi and K.H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," *Behavioural Brain Research*, vol. 344, pp. 103-109, February 2018. Article (CrossRef Link)

[16] M. Liu, J. Zhang, E. Adeli and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Medical Image Analysis*, vol. 43, pp. 157-168, October 2017. Article (CrossRef Link)

[17] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an emsemble system of deep convolutional neural networks," *Brain Informatics*, vol. 5, no. 2, May 2018. Article (CrossRef Link)

[18] A. Gupta, M. Ayhan and A. Maida, "Natural image bases to represent neuroimaging data," in *Proc. of the 30th Int. Conf. on machine learning*, vol. 28, pp. 987-994, June 2013. Article (CrossRef Link)
[19] A. Payan and G. Montana, "Predicting Alzheimer's disease; a neuroimaging study with 3D convolutional networks," in *Proc. of 4th Int. Conf. on Pattern Recognition Applications and Methods*, vol. 2, February 2015. Article (CrossRef Link)
[20] T. Brosch, R. Tam and Initiative ADN et. al., "Manifold learning of brain MRIs by deep learning," in *Proc. of the Int. Conf. on Medical Image Computing and Computer-assisted Intervention*, vol. 16, pp. 633-640, 2013. Article (CrossRef Link)
[21] H.I. Suk, S.W. Lee, D. Shen and Initiative ADN et. al., "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *Neuroimage*, vol. 101, pp. 569-582, November 2014. Article (CrossRef Link)
[22] H.I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in *Proc. of Int. Conf. on Medical Image Computing and Computer-assisted Intervention*, vol. 16, pp. 583-590, September 2013. Article (CrossRef Link)
[23] H.I. Suk, S.W. Lee, D. Shen and Initiative ADN et. al., "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Structure and Function*, vol. 220, no. 2, pp. 841-859, March 2013. Article (CrossRef Link)
[24] D. Cárdenas-Peña, D. Collazos-Huertas and G. Castellanos-Dominguez, "Centered kernel alignment enhancing neural network pretraining for MRI-based dementia diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 1-10, January 2016. Article (CrossRef Link)
[25] Y. Fisher and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," in *Proc. of the Int. Conf. on Learning Representations*, May 2016. Article (CrossRef Link)
[26] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proc of the Int. Conf. on Machine Learning*, pp. 448-456, July 2015. Article (CrossRef Link)
[27] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhundinov, "Dropout: a simple way to prevent neural networks from overfitting," *Machine Learning*, vol. 15, no. 1, pp. 1929-1958, June 2014. Article (CrossRef Link)
[28] G. Huang, Z.Liu, L. Van der Mateen and K.Q. Weinberger, "Densely connected convolutional networks," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2261-2269, July 2017. Article (CrossRef Link)
[29] A. Krizhevsky, I. Sutskever and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, no. 2, pp. 1097-1105, December 2012. Article (CrossRef Link)
[30] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghenmawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, "Tensorflow: Large-scale machine learning on heterogeneous systems," 2015. https://www.tensorflow.org
[31] F. Chollet et. al., "Keras," 2015. https://github.com/keras-team/keras
[32] K. Simonyan and A. Sizzerman, "Very deep convolutional networks for large-scale image recognition," in *Proc. of 3rd Int. Conf. on Machine Learning*, May 2015. Article (CrossRef Link)
[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going Deeper with Convolutions," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1-9, June 2015. Article (CrossRef Link)
[34] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.770-778, July 2016. Article (CrossRef Link)
[35] D. Jha and G.R. Kwon, "Alzheimer's Disease Detection Using Sparse Autoencoder, Scale Conjugate Gradient and Softmax Output Layer with Fine Tuning," *Int. Journal of Machine Learning and Computing*, vol. 7, no. 1, pp. 13-17, February 2017. Article (CrossRef Link)

Husnu Baris Baydargil is a Ph.D. student at the Department of Electrics, Electronics and Communication Engineering, Kyungsung University, Busan, Republic of Korea. He obtained his bachelor's degree from Bahcesehir University, Istanbul, Turkey from the Department of Software Engineering. He obtained his M.Sc. degree in the Department of Electrics, Electronics and Communication Engineering, Kyungsung University under the supervision of Dr. Jong-Kwang Song. His research is focused on computer vision, machine learning, and image classification.

Jangsik Park received the B.S., M.S and Ph.D. degrees in Electronics Engineering from Busan National University in 1992, 1994, and 1999, respectively. From 1997 to 2011, he was a professor at Dongeui Institute of Technology. Since 2011, he has been a professor in the Department of Electrical and Electronics Engineering, Kyungsung University. His research interests include video/image processing and understanding, speech and audio signal processing and embedded systems.

Do-Young Kang is the director and chief professor of the department of nuclear medicine in Dong-A university medical center and Dong-A university college of medicine and also the director of the institute of convergence bio-health in Dong-A University, Busan, Korea. His research is focused on image analysis of degenerative brain diseases such as Alzheimer's dementia and Parkinson's disease.

Hyeon Kang is a master's degree student at the Department of Nuclear Medicine, College of Medicine, Dong-A University, Busan, Republic of Korea. He obtained his bachelor's degree from the Department of Computer Engineering, Dongseo University, Busan, Republic of Korea. His current research is focused on the application of machine learning techniques to molecular imaging such as PET or MRI.

Kook Cho graduated with a degree in electrical and electronic engineering from Ritsumeikan University in Japan. He is studying machine learning and deep learning analysis and education of voice and image data.