Comparing the Architecture and Performance of AlexNet, Faster R-CNN, and YOLOv4 in the Multiclass Classification of Alzheimer Brain MRI Scans

Ria Mirchandani[1,*], Caroline Yoon[1,*], Sonica Prakash[1], Archita Khaire[1], Alyssia Naran[1], Anupama Nair[1], and Supraja Ganti[1]

[1]AI4ALL
[*]e-mail: ria.d.mirchandani@gmail.com
[*]e-mail: carolineyoon@live.com

ABSTRACT

Although advancements in diagnostic imaging such as magnetic resonance imaging (MRIs) have led to a greater understanding of the diagnosis and treatment of Alzheimer's Disease (AD), medical professionals are still required to analyze the images, which is a time-consuming and error-prone process. With the help of neural network models, diagnoses can be reached more accurately and efficiently. In this study, we compared the performance of three well-known CNN-based algorithms — AlexNet, Faster R-CNN, and YOLOv4 — to determine which one was most accurate at performing multiclass classification of brain MRI scans of AD patients. The dataset utilized was obtained from Kaggle and contained 6400 training and testing MRI images divided into four classes (NonDemented, VeryMildDemented, MildDemented, and ModerateDemented). The ModerateDemented class was extremely underrepresented. To obtain more accurate results, images were added to that class through data augmentation. Experiments were conducted using Google Colab's Tesla P100 GPU. Transfer learning was applied to all three pre-trained models and the datasets were adjusted according to their respective parameters. Post-augmentation, AlexNet had the highest mAP (Mean Average Precision), detecting the object of interest 100% of the time, while YOLOv4 and Faster R-CNN had an mAP of 84% and 99% respectively. However, YOLOv4 had the best performance on the confusion matrix, especially for the ModerateDemented images. As revealed in our experiment, one-stage detectors like YOLOv4 were faster and more accurate than two-stage detectors like Faster R-CNN. Our study successfully implemented these models and made valuable contributions to medical image diagnosis, opening avenues for future research and development.

Introduction

*A*lzheimer's Disease (AD) is a common, yet terminal illness that is characterized by progressive memory loss, disorientation, and pathological markers, including senile plaques and neurofibrillary tangles in the brain[1]. It is the sixth leading cause of death worldwide and the most common form of dementia, accounting for up to 80% of dementia diagnoses. According to the Center for Disease Control and Prevention (CDC) in 2014 more than five million people were living with AD, and this is expected to almost triple to fourteen million people by 2060. Despite its prevalence, there is currently no cure to this degenerative disease. However, advancement in diagnostic imaging such as magnetic resonance imaging (MRIs) have led to a greater understanding of the diagnosis and treatment of AD. Magnetic resonance imaging is a non-invasive imaging modality that allows high spatial resolution and contrast in evaluating brain anatomy, function, and pathology[2]. Multiplanar and three dimensional imaging allows high sensitivity and specificity in detecting and displaying abnormalities in the brain, including the cerebrum, cerebellum, and brainstem.

AD diagnosis in brain MRI scans is primarily based on cerebral atrophy, which is a decrease in brain volume due to loss of gray and white matter in the mesial temporal and temporoparietal cortical lobes. Additionally, the segmentation of brain MRIs taken at different stages can also be used to measure structural changes in the brain during AD. However, analyzing large and complex MRI datasets and extracting information manually can be challenging for clinicians. Moreover, manual assessment of brain MRIs can be especially time-consuming and vulnerable due to errors due to inter- or intra-operator variability issues.[3]. One solution is to automate the MRI analysis method using computerized techniques for MRI segmentation, visualization, and registration to produce accurate results with high confidence[2]. Several past studies have shown that images that were run through and analyzed by either an automated algorithm or neural network model were able to diagnose scans of abnormal tumors more accurately and increasingly efficiently than was possible by the medical professionals[4].

Brain MRI segmentation is considered an essential task in multiple clinical applications because it influences the outcome of the entire analysis. MRI segmentation is most commonly used for measuring and visualizing brain structures, detecting lesions, and for image-guided interventions and surgeries. The main objective of brain MRI segmentation is to divide an image into well-defined regions, where each region is made up of pixels that share the same range of intensities and textures.

To improve the detection and segmentation task in MRI analysis, multiple deep learning approaches have been proposed. In this study, three applications of convolutional neural networks (CNN) are introduced: AlexNet, Faster R-CNN, and YOLO. CNNs are a class of neural networks that are most commonly applied to computer vision and represent an analogy to the neuron connectivity pattern in the brain. Specifically, a CNN is made up of one input layer, multiple hidden layers, and an output layer. The hidden layers, which improve the network and control overfitting, are structurally composed of convolutional, activation, pooling, and fully connected layers. In the convolutional layer, a filter slides, or convolves, over an array of pixel values known as the input image and identifies significant features, such as colors, curves, and edges. The output of the first layer is an activation map of numbers, where the numbers indicate the presence of features on the image. The final layer of a CNN, the SoftMax layer, looks at high valued features from the activation map and determines their correlation to a specific classification for the image[5]. During the training phase, the weights and thresholds of layers are adjusted until the training data yields consistent outputs[6].

Despite the relative efficiency of its architecture, CNNs were initially difficult to apply in large scale to high-resolution images, such as the ImageNet dataset. Krizhevsky et. al.[7] developed an 8 layer CNN known as Alexnet, consisting of 5 convolutional layers and 3 fully connected layers with 1376 filters, 60 million parameters, and 650,000 neurons, to classify 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest. Rectified Linear Unit (ReLU) Nonlinearity functions allowed Alexnet to convert negative pixels to zero thus down-sampling the data and allowing it to train quickly. AlexNet is known as one of the leading architectures for object detection tasks and has accelerated the development of many other deep learning models.

Maqsood et. al.[8] evaluated the architecture of AlexNet by training it on the ImageNet dataset and then classifying stages of AD from normal to mildly to moderately demented using transfer learning. The purpose of their study was to propose a transfer-learning based method for multiclass classification of AD and evaluate the effect of various grayscale 3D MRI views on the performance of the classifiers. For fine-tuning, all layers of the AlexNet were extracted as transfer layers and the last three were replaced with modified SoftMax layers, fully connected layers, and an output classification layer, so the algorithm could learn specific features of the dataset. Training was fine-tuned on both segmented and unsegmented 3D views of human brain MRI scans, with overall accuracies of 89.6% and 92.8% for binary and multi-class problems.

The Faster R-CNN object detection algorithm is a two stage object detector: first identifying regions of interest and then passing these regions through a CNN. It is also an enhancement of the R-CNN (Region-Based CNN), which follows a basic object detection pipeline, however it incorporates a CNN after extracting features. In a generic pipeline, region proposals are generated, features are extracted, and regions are classified. The purpose of the region proposals is to suggest objects that may be identifiable in an image. Faster R-CNN improves on the model by implementing Region Proposal Networks (RPN) which uses convolutional layers to generate quicker, more accurate, and customizable region proposals. Upon receiving these proposals, it uses Region of Interest (ROI) Pooling to down-sample the data, with bounding boxes identifying regions objects of interest. The RPN then ranks region boxes and proposes those most likely to contain objects.[9].

First described in 2015, The You Only Look Once (YOLO) algorithm is a machine learning model for image classification and object detection with an open source implementation known as Darknet, developed in the C programming language using CUDA (Compute Unified Device Architecture) technology for general purpose computing on both central processing units (CPUs) and graphics processing units (GPUs). YOLO's architecture performs both tasks in a one-stage process[10], allowing for speedier detection as well as higher mean average precision (mAP) and less background errors. Similar to Faster R-CNN, YOLO is an effective object detection algorithm that applies bounding boxes. However, unlike two-stage algorithms like Faster R-CNN which first generate potential bounding boxes and then run classifiers on the boxes, the YOLO framework utilizes one CNN to simultaneously determine bounding boxes and class probabilities. In other words, "you only look once" at the image to identify objects and their locations. The highly generalized YOLO model outperforms other top detection algorithms on novel input images[11]. For YOLO, we utilized the open-source Darknet implementation of YOLOv4. As described earlier, YOLOv4 is a one-stage detector that does classification and bounding box regression simultaneously. The purpose of the classification stage is to determine the category of the object (e.g. Non, Very Mild, Mild, or Moderate Demented) and the purpose of bounding box regression stage is to determine the location of the object in the image.

Recent works on MRI classification have been heavily dependent on pre-processing techniques and various classification methods in order to obtain good results. Fong et. al.[12] proposed an AD diagnosis solution that bypassed the MRI pre-processing technique by using deep learning detection networks, including Faster R-CNN, SSD, and YOLOv3. Trained on 1000 raw, unprocessed subjects' data from the Alzheimer's Disease NeuroImaging Initiative (ADNI) database, the algorithms obtained a detection accuracy of 0.998 for YOLOv3, 0.982 for SSD, and 0.998 for Faster R-CNN with all surpassing the 0.75 Intersection Over Union (IoU) threshold, an evaluation method which defines the ratio between the overlapping area and the union area of the detected box and its corresponding ground truth box data. Whereas the Faster R-CNN failed to provide any localization and classification input for any inference image size below 600^2, the SSD and the YOLOv3 did not have any limitation on image inference size, achieving a 0.9966 (YOLOv3) and a 0.9418 (SSD) accuracy at inference size of 200^2. This highlights the difference in architecture and performance between two-stage detectors (Faster R-CNN) and one-stage detectors (YOLOv3) when tested on various image sizes.

In recent years, research involving machine learning based identification of MRI patterns in AD has been widespread. Deep learning has been used in the classification of AD with promising results, however certain limitations have been encountered. A 2020 study utilized a CNN-based classification algorithm on MRI images to develop a generalizable, fast, and accurate diagnostic tool[13]. Specifically, the algorithm considered atrophy of the medial temporal lobe, a common clinical indicator of AD related neurodegeneration. The study was the first to successfully use 2-D MRI images for classification, allowing for a reduction in computational power and processing time. However, they faced limitations in that their images, or data, were acquired from only one manufacturer of MRI images. Therefore, having a large and diverse dataset is an important factor for developing an accurate algorithm. Since it is necessary for medical professionals to be the first group to diagnose each scan, finding labeled MRI image datasets is often difficult. This study limitation resulted in varying performance accuracies during testing, indicating that prior data obtained from a variety of MRI scanners is necessary in AD classification.

In this paper, we leverage deep learning architectures including AlexNet and two object detection methods known as Faster R-CNN and YOLOv4 to classify Alzheimer's disease status in patient brain MRI images. We compare the effectiveness and accuracy of these models and address specific limitations and recommendations regarding both our algorithms and dataset, with the hopes of providing valuable insight for future studies related to MRI image classification.

Methods

System Outline

In this experiment, three deep learning computer vision architectures were used: AlexNet, Faster R-CNN, and Yolov4. The experiments were conducted using Google Colab's Tesla P100 GPU. Transfer learning was applied to all three pre-trained models and the models were adjusted according to their respective parameters.

Dataset Acquisition

The data utilized in the completion of this project was obtained from a publicly available database on Kaggle[14]. 6400 segmented MRI images of the brain from an axial view were labeled and divided into four classifications: NonDemented, VeryMildDemented, MildDemented, and ModerateDemented (Fig 1). These images were then divided into the training and test set in a 4:1 ratio, respectively (Table 1). It is important to note that the number of images per class were unbalanced, especially in regards to the ModerateDemented, which had only 52 images in the training and 12 images in the testing sets.

Classes	Training Set	Test Set
NonDemented	717	179
VeryMildDemented	52	12
MildDemented	2560	640
ModerateDemented	1792	448

Table 1. Table for the number of images in each classification for the training and test set. The number of images per class was largely unbalanced, with the ModerateDemented images being vastly underrepresented (total: 64 images) in comparison to the MildDemented (total: 3200).

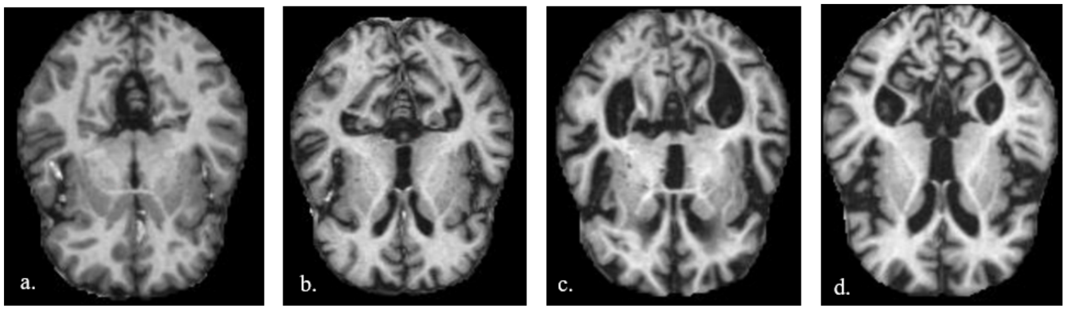

Figure 1. Four different classifications of Kaggle brain MRI dataset (a) Non Demented (b) Very Mild Demented (c) Mild Demented (d) Moderate Demented

Dataset Preparation

LabelImg, a common graphical annotation tool that labels objects in images using bounding boxes, was used for the object detection algorithms (Fig 2). The full set of MRI images were manually labelled in the Darknet .txt format required by the YOLO algorithm for easier classification by the model. Roboflow, an image management platform, was subsequently used to store the images and generate a link for easy accessibility. Since the Faster R-CNN required COCO JSON formatted images, RoboFlow was employed to convert image annotations from .txt files into COCO JSON formatted files. Originally, the models were tested on unaugmented images. However in order to prevent overfitting, Roboflow was then utilized to generate augmented images of the underrepresented ModerateDemented images using features including random flip, crop, and brightness adjustments in order to create 104 new images.

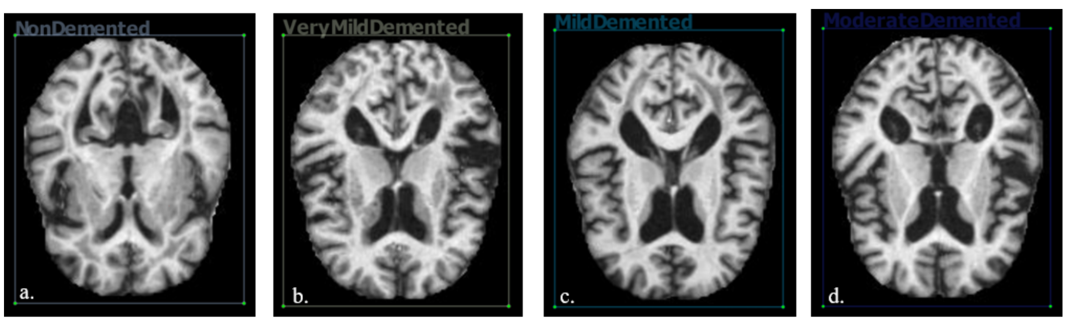

Figure 2. Bounding boxes using LabelImg (a) Non Demented (b) Very Mild Demented (c) Mild Demented (d) Moderate Demented

Experimental Setup

AlexNet

The pre-trained AlexNet model and its parameter values were imported from the PyTorch library and trained with the Alzheimer brain MRI dataset. The input data consisted of a batch size of 32 with a height of 227, a width of 227, and depth of 3, with the AlexNet running through 170 epochs of the training data.

Faster R-CNN

The Detectron2 implementation of Faster R-CNN (X101-FPN) was employed in this study. Detectron2 is an open-source object detection platform based on the PyTorch library that includes implementations of Faster R-CNN and other computer vision algorithms. The Detectron2 API provides the model architecture as well as its pre-trained weights which are based on the COCO (Common Objects in Context) dataset. The Alzheimer brain MRI images were then converted into COCO JSON file format and the model was trained through 8000 epochs.

YOLOv4

The Darknet framework, a custom framework written by Joseph Redmon, was used to implement YOLOv4. The initial weights of the Darknet YOLOv4 algorithm were based on the COCO dataset. The model was subsequently trained with the Alzheimer brain MRI dataset through 8000 epochs.

Results

Initial Study

AlexNet

The AlexNet model reached 100% Mean Average Precision in 170 epochs. Initially, the AlexNet algorithm performed poorly in classifying the test dataset, specifically the MildDemented and ModerateDemented Images, achieving a recall score (true positive) of 0.0% (0/179 images for MildDemented and 0/12 images for ModerateDemented) in both cases (Fig.3). However, it performed slightly better on the VeryMildDemented with a recall score of 3.3% (15/448 images) and significantly better on NonDemented Images with a recall score of 96.9% (620/640 images). Interestingly, AlexNet misclassified the 96.1% of the MildDemented and 96.7% of the VeryMildDemented as NonDemented (Fig.3). This revealed potential overfitting due to the unbalanced nature of the dataset. AlexNet's final accuracy was 99% with a loss value of 0.02 (Table 2).

Faster R-CNN

With the original 5121 unaugmented training images, the Faster R-CNN achieved a 41% Mean Average Precision within 8000 epochs. However, it performed poorly on the MildDemented, ModerateDemented, and VeryMildDemented test images, with a recall score of 0.0% (0/74 images), 0.0% (0/1 image), and 0.7% (2/304 images) respectively (Fig.3). Similar to AlexNet, the Faster R-CNN misclassified 87.8% of the MildDemented images and 99.3% of the VeryMildDemented images as NonDemented images, most likely due to overfitting. However, it performed better with the NonDemented images, achieving an recall score of 99.8% (536/537 images) (Fig.3). Faster R-CNN's final accuracy was 41% with a loss value of 0.30 (Table 2).

YOLOv4

YOLOv4 reached 86% Mean Average Precision in 8000 epochs. Additionally, it classified the test data the best out of all three algorithms. The MildDemented had a recall score of 44.1% (79/179 images), the ModerateDemented had a recall of 16.7% (2/12 images), the NonDemented had a recall of 69.8% (447/640 images), and the VeryMildDemnted had a recall of 83.9% (376/448 images). 29.7% of the NonDemented images were misclassified as VeryMildDemented (Fig.3). Furthermore, after training YOLOv4 achieved a final accuracy of 86% with a loss value of 0.70 (Table 2).

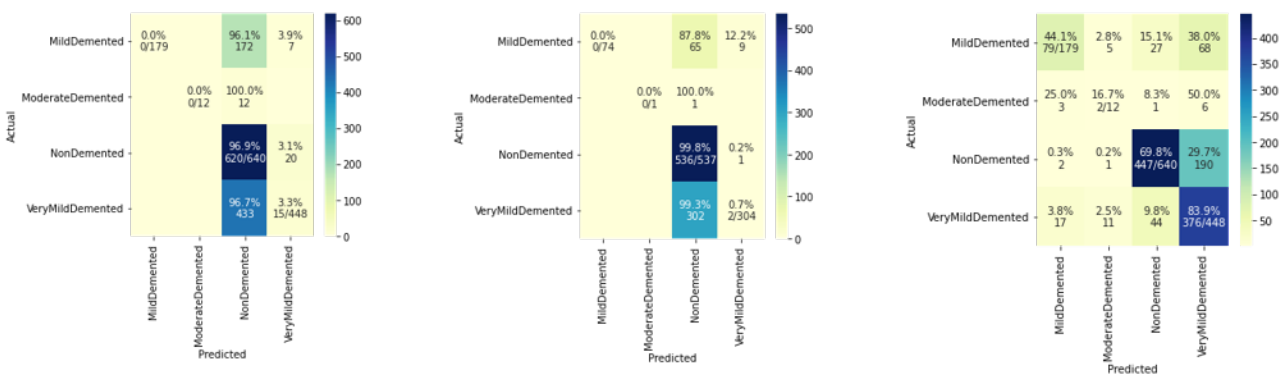

Figure 3. Confusion matrix for AlexNet (left), Confusion Matrix for FASTER R-CNN (middle), Confusion matrix for YOLOv4 (right). All three were trained on the unaugmented training set (5121 images).

Classifiers	AlexNet	Faster R-CNN	YOLOv4
Loss	0.02	0.30	0.70
Accuracy	99%	41%	86%

Table 2. Table displaying accuracy and loss curve statistics for AlexNet, Faster R-CNN, and YOLOv4.

Modified Study

AlexNet

With the addition of 104 new augmented images, AlexNet performed better with a 100% Mean Average Precision in 170 epochs. The recall score for MildDemented had improved to 3.9% (7/179 images), ModerateDemented to 16.7% (2/12 images), 49.2% (315/640 images), and 35.5% (159/448 images) (Fig. 4). The newly generated loss and training accuracy curve for AlexNet was subsequently generated (Fig. 5) with the final accuracy of 99% and a loss value of 0.05 displayed in Table 3.

Faster R-CNN

Similarly, the Faster R-CNN also improved with a 84% Mean Average Precision within 8000 epochs. However, it still did not perform well with the Mild and ModerateDemented images. It was unable to classify any of the MildDemented and ModerateImages images correctly with a recall score of 0.0%. The NonDemented and VeryMildDemented recall scores were 46.4% (252/543 images) and 91.5% (311/340 images) respectively (Fig. 4). Interestingly, Faster R-CNN misclassified 53.6% of the NonDemented as VeryMildDemented. With augmented data, the Loss and Training Accuracy curves over the 8000 epochs were generated (Fig. 5), with a final significantly improved accuracy of 84% and loss of 0.28 (Table 3).

YOLOv4

YOLOv4 reached 99% Mean Average Precision in 8000 epochs, classifying the test data significantly better. The YOLOv4 reached a recall score of 35.2% (63/170 images) for MildDemented, 100% (12/12 images) for ModerateDemented, 62.8% (402/640 images) for NonDemented, and 53.6% (240/448) for VeryMildDemented (Fig. 4). Additionally, the training and loss accuracy curve is displayed below (Fig. 5), with a final accuracy of 99% and loss value of 0.75 (Table 3).

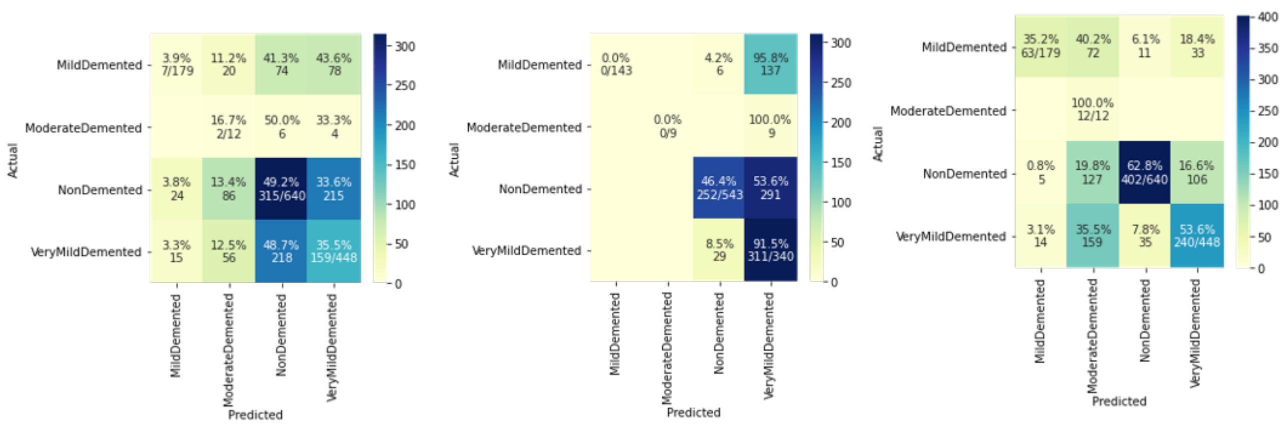

Figure 4. (Left) Confusion matrix for AlexNet, (Middle) Confusion matrix for Faster R-CNN, (Right) Confusion matrix for YOLOv4. All three were trained on the augmented training set (5225 images).

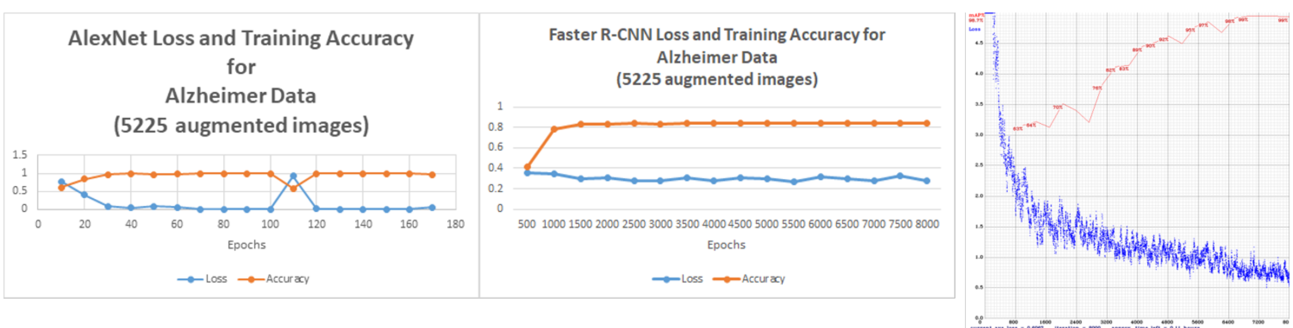

Figure 5. (Left) Accuracy and loss curve for AlexNet, (Middle) Accuracy and loss curve for Faster R-CNN, (Right) Accuracy and loss curve for YOLOv4. All three were trained on the augmented training set (5225 images).

Classifiers	AlexNet	Faster R-CNN	YOLOv4
Loss	0.05	0.28	0.75
Accuracy	99%	84%	99%

Table 3. Table displaying accuracy and loss curve statistics for AlexNet, FASTER R-CNN, and YOLOv4 after data augmentation.

Discussion

Evaluation

Experimentally, we displayed the performance of our classifiers using a confusion matrix (Fig.3, 4), which is a common evaluation metric for binary and ternary classification. We also adopted two other evaluation standards including Mean Average Precision (mAP) and Accuracy and Loss Curves. mAP is a popular metric used to measure the performance of object detection models. It is calculated by first measuring the intersection over union (IoU), which measures the overlap between the predicted bounding box and the ground truth bounding box. The Precision (false negatives) and Recall (false negatives) values are then calculated using IoU values for a given IoU threshold (0.5). The general definition for Average Precision (AP) is the area under the precision-recall curve. Once that is found, the average of all APs over all object classes are calculated to achieve the mAP score (Fong et. al., 2020). The Accuracy and Loss curves are the two of the most-well known metrics in machine learning (Table 2, 3, Fig. 5). Accuracy refers to the count of predictions where the predicted value is equal to the true value and is often monitored throughout the training phase to determine the model's overall performance. The loss function refers to the uncertainty of a prediction based on how much the prediction varies from the true value. It is a summation of errors made for the samples in the training set and helps adjust the weights in the neural network.

Comparison

In this study, we compared the performances of the models that were trained with the augmented training dataset (5225 images). Overall, the AlexNet and YOLOv4 had the highest mAP, with both being able to detect the object of interest 99% of the time, whereas the Faster R-CNN achieved a mAP of 84%. On the other hand, YOLOv4 had the best performance on the confusion matrix, especially in regards to the ModerateDemented images, which it correctly classified 12 out of 12 times. Overall, YOLOv4 correctly classified 35% of the MildDemented, 100% of the ModerateDemented, 63% of the NonDemented, and 54% of the VeryMildDemented (Fig. 4). The Faster R-CNN seemed to perform the worst in both the confusion matrix and the mAP metric. Even with the augmented data, it still was not able to classify any of the MildDemented and ModerateDemented images correctly (Fig. 4). Lastly, in terms of loss value, the AlexNet had the lowest loss value of 0.05, while the YOLOv4 had a loss value of 0.75 (Table 3), revealing that the AlexNet produced fewer errors than the YOLOv4.

There are mainly two types of state-of-the-art object detection methods. On one hand, there are two-stage detectors, including Faster R-CNN and Mask R-CNN, which (1) uses Region Proposal Networks to generate areas of interest in the first stage (2) inputs the region proposals into the pipeline for classification and object detection in stage 2. On the other hand, there are single-stage detectors, including YOLO and SSD (Single Shot MultiBox Detector), which approach object detection as a simple regression problem by obtaining the input image and learning the class probabilities and bounding box coordinates in one stage. Two-stage detectors like Faster R-CNN tend to reach the highest accuracy rates, but are typically slower and take longer to train, whereas one-stage detectors are typically much faster, but achieve lower accuracy. However, in our study the YOLO seemed to outperform the Faster R-CNN, with an mAP of 99% as compared to Faster R-CNN's 84%. There are a number of possible reasons why Faster R-CNN may have under-performed. First, considering that two-stage detection methods are typically slower, perhaps Faster R-CNN was not trained long enough and required training for more than 8000 epochs. Another postulation is that the image sizes weren't large enough. According to Fong et. al. (2020)[12], one-stage detectors do not have limitations in regards to image inference size and are able to efficiently detect smaller images, whereas two-stage detectors are poor classifiers of smaller images smaller less than 600^2 (In our case, our images were dimension size 176 x 208).

Limitations

The biggest limitation was the lack of representation of the ModerateDemented and MildDemented images in the dataset. These two classes were largely underrepresented, whereas the NonDemented class consisted of 2560 training images. This imbalance resulted in overfitting and the models classifying most of the images as NonDemented. Even with the addition of 104 new augmented ModerateDemened images, there was still an unequal representation of classes in the dataset. Fortunately, with image augmentation, we were able to improve the accuracy of all three models, with the YOLOv4 correctly classifying 100% the ModerateDemented images for the first time. Image augmentation can therefore help improve the performance of the algorithms for image classification and object detection. However, it is still more important that the classes be equally represented in the data.

Future Studies

In the future, we will continue to improve upon our models, adjusting parameters and optimizing the performances of the YOLOv4 and Faster R-CNN. This would include training the models for longer, augmenting the images for optimized image detection, and of course, generating additional images to balance out the unequal representation in the training set. Another step for the study is to perform ensemble classification for all 3 algorithms using a "voting ensemble," in which the prediction receiving the highest number of votes would be utilized. The value of an ensemble classifier is that it is able to correct errors made by an individual classifiers and combine multiple classifiers to achieve the highest accuracy.

Conclusion

The purpose of this study was to utilize three well-known CNN-based algorithms, AlexNet, Faster R-CNN, and YOLOv4 to perform multiclass classification of brain MRI scans of Alzheimer Disease patients. Not only were we able to successfully implement these models, but we have also contributed valuable and meaningful insight to medical image diagnosis. We also hand labelled the training images, and this annotated dataset can be used for further research. Hopefully our study has contributed to understanding of medical image diagnoses using machine learning that future research can build upon. For practicing radiologists, predictions from algorithms such as YOLOv4 can supplement their professional judgement of a patient's Alzheimer condition.

References

1. Schachter, A. & Kenneth, D. Alzheimer's disease. *Dialogues clinical neuroscience* **2**, 91–100 (2000).
2. Yamanakkanavar, N., Choi, J. Y. & Lee, B. Mri segmentation and classification of human brain using deep learning for diagnosis of alzheimer's disease: a survey. *Sensors* **20**, 3243 (2020).
3. Despotović, I., Goossens, B. & Philips, W. Mri segmentation of the human brain: challenges, methods, and applications. *Comput. mathematical methods medicine* **2015** (2015).
4. Walsh, F. Ai "outperforms" doctors diagnosing breast câncer. *BBC News* (2020).
5. Kandel, I., Castelli, M. & Popovič, A. Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images. *J. Imaging* **6**, 92 (2020).
6. Hardesty, L. Explained: neural networks. *MIT News. Saatavissa: https://news. mit. edu/2017/explained-neural-networks-deeplearning-0414. Hakupäivä* **19**, 2020 (2017).
7. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
8. Maqsood, M. *et al.* Transfer learning assisted classification and detection of alzheimer's disease stages using 3d mri scans. *Sensors* **19**, 2645 (2019).
9. Park, S. H., Yoon, H. S. & Park, K. R. Faster r-cnn and geometric transformation-based detection of driver's eyes using multiple near-infrared camera sensors. *Sensors* **19**, 197 (2019).
10. Peixoto, H. M., Teles, R. S., Luiz, J. V. A., Henriques-Alves, A. M. & Santa Cruz, R. M. Mice tracking using the yolo algorithm. *PeerJ Prepr.* **7**, e27880v1 (2019).
11. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788 (2016).
12. Fong, J. X., Shapiai, M. I., Tiew, Y. Y., Batool, U. & Fauzi, H. Bypassing mri pre-processing in alzheimer's disease diagnosis using deep learning detection network. In *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, 219–224 (IEEE, 2020).
13. Bae, J. B. *et al.* Identification of alzheimer's disease using a convolutional neural network model based on t1-weighted magnetic resonance imaging. *Sci. Reports* **10**, 1–10 (2020).
14. Dubey, S. Alzheimer's dataset (4 class of images). *Kaggle, Dec* (2019).

Supplementary information

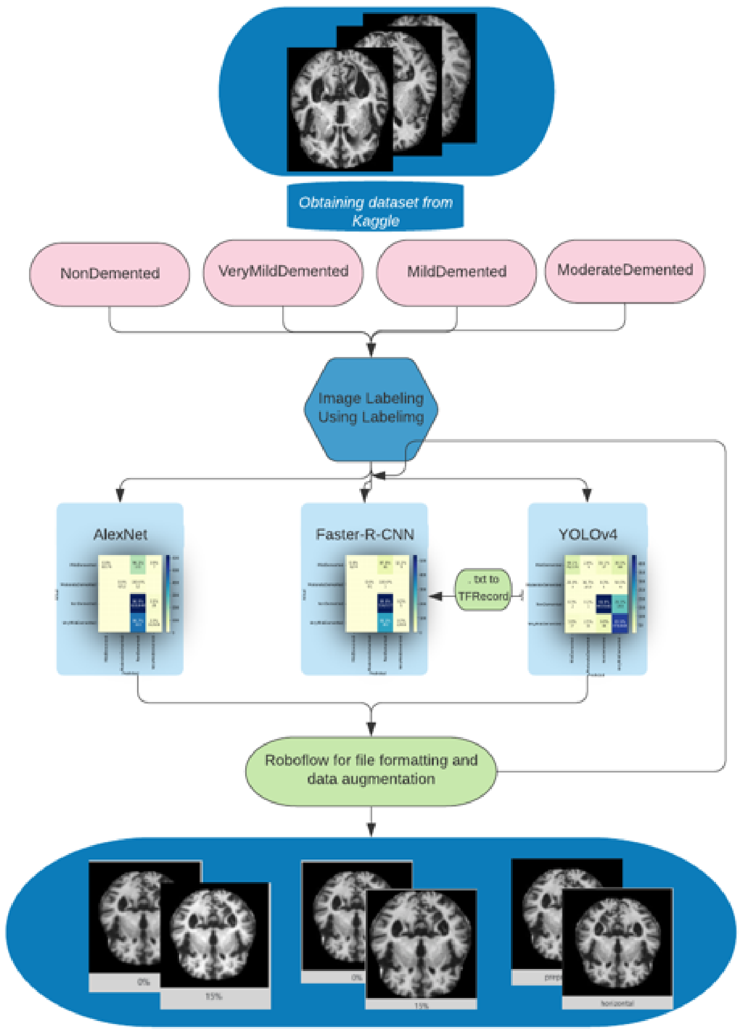

Figure 6. Flowchart for Methodology - (1) Obtained 4600 Alzheimer brain MRI images from Kaggle (2) Divided into four classes: NonDemented, VeryMildDemented, MildDemented, ModerateDemented (3) Manually labeled each image with bounding boxes using LabelIMG (4) Trained dataset on three classifiers: AlexNet, Faster R-CNN, YOLOv4 (5) Employed RoboFlow for file formatting (e.g. Converting between .txt for YOLOv4 to COCO JSON for Faster R-CNN) (6) Data augmentation using random flip, crop, and brightness to generate 104 new images (7) Results were generated using confusion matrices and loss and accuracy curves