Alzheimer's Disease Stage Classification in Magnetic Resonance Images Using Deep Learning

Earlbert M. Mercado
College of Engineering Architecture and Fine Arts
Batangas State University
Batangas City, Philippines
earlbert.mercado@g.batstate-u.edu.ph

***Abstract*—Alzheimer's Disease is one of the most common type of dementia and commonly occurs in older people. So far, scientists still don't fully understand its cause and how to cure it. The only possible thing to do is to reduce its progression by detecting it as early as possible. In this paper, an Alzheimer's Disease Stage Classification in Magnetic Resonance Image Using Deep Learning is developed. A pre-trained model of the YOLOv3 algorithm is used to train a new deep learning model. The new model obtained a mean average precision (mAP) of 95.48% and got a 44.44% accuracy score during testing.**

Keywords— Alzheimer's Disease, Magnetic Resonance Image, Deep Learning, YOLOv3

I. INTRODUCTION

As humans get older, the chance of having different kinds of illness gets higher. Having illness as a senior citizen is difficult since the body and the organs do not function well as it should, especially brain-related illness because the brain is the most complex and sensitive organ in the body. It usually has expensive treatment costs for medicines and therapies. One of these is Alzheimer's disease. Alzheimer's disease is the most common type of dementia that slowly destroys memory and thinking skills and, eventually, the ability to perform simple tasks. It causes the brain to shrink and the brain cells to die. Its symptoms include memory loss, confusion, inability to learn new things, difficulty in language, and other cognitive related skills. The percentage of people with Alzheimer's rises with age: 3% of people age 65-74, 17% of people age 75-84 and 32% age 85 and older [1] . It is believed that there are 50 million people living with Alzheimer's all over the world and if breakthroughs are not discovered, numbers could reach 152 million by the year 2050 [2].

There are medicines available that can reduce Alzheimer's symptoms. Few medications approved by the Food and Drug Administration (FDA) are listed in [3]. Aducanumab is a drug that may delay clinical decline with benefits to both cognition and function in people living with Alzheimer's. Another drug that treats cognitive symptoms (memory and thinking) includes Donepezil, Galantamine, Rivastigmine, Memantine, and the combination of Memantine and Donepezil. Unfortunately, these medicines have side effects such as nausea, vomiting, loss of appetite, increased frequency of bowel, headache, confusion, and dizziness. Unfortunately, these medications have side effects such as nausea, vomiting, loss of appetite, increased frequency of bowel, headache, confusion, and dizziness [3].

There are several factors in Alzheimer's progression that can affect each person differently. Biggest factor that increases the chance of having Alzheimer's is age. As stated earlier, starting in the age of 65, the number of people with Alzheimer's doubles every 5 years. Another factor is family history. Researchers believe that genetics may play a role in developing Alzheimer's. Other comorbid chronic conditions such as heart disease, stroke, etc. may also worsen a person's Alzheimer's disease [4] .

Although there are medicines for reducing Alzheimer's symptoms, there is no cure for it yet. With the advanced technologies that can help to treat different medical conditions, scientists still don't fully understand the cause of Alzheimer's and how to cure it. The only thing that is possible is to reduce its progression by detecting it as early as possible.

There are numerous studies that aim to detect Alzheimer's disease. In the study of [5], the authors introduce a computer-aided diagnosis system for Alzheimer's disease detection using machine learning techniques. They got an accuracy of 99.43% in logistic regression and 99.10% on SVM using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset while 84.33% and 83.92% in logistic regression and random forest, respectively using Open Access Series of Imaging Studies (OASIS) brain datasets. Another researchers presents a framework for constructing predictive models of cognitive decline from longitudinal MRI examinations based on mixed effects model and machine learning in [6]. They apply the framework to detect conversion from cognitive normal (CN) to mild cognitive normal (MCI) and from MCI to Alzheimer's. They obtained an average accuracy, precision, and recall score of 69, 73, and 60% respectively for conversion to MCI and 75, 74, and 77% for conversion to Alzheimer's. In [7], a new method is proposed using first-order statistical features in 3d brain MRI images and the experimental results show that they obtained an accuracy score of 90.9%. Two problems were discussed in [8]: first, a classification method is developed to classify MRI as either normal with Alzheimer's and the second one is for the identification and classification between normal subjects, MCI patients, and AD patients. Their method includes wavelet feature extraction from MRIs, dimensionality

reduction, training-test subdivision and classification using SVM. Results have shown that using feature reduction and feature selection led to worse accuracy while SVM results obtained have shown a better performance. A more sophisticated approach is applied in [9] where the authors used deep learning for early detection of Alzheimer's disease. Authors used two methods to classify the medical images and detect Alzheimer's. The first one uses simple convolutional neural network (CNN) architecture that deals with 2D and 3D scans from ADNI dataset based on 2D and 3D convolution. Another method uses the transfer learning principle to take advantage of the pre-trained models for medical image classification such as the VGG19 model. Results have shown that they achieved 93.61% and 95.17% accuracy for 2D and 3D multi-class AD stage classification. The VGG19 pre-trained model is fined-tuned and achieved 97% accuracy for multi-class AD classification. A novel approach where CNN and Ensemble Learning are combined is proposed in [10] to detect and classify Alzheimer's disease in MRIs. Result revealed an accuracy of 0.84 ± 0.05, 0.79 ± 0.04, and 0.62 ± 0.06, respectively, for classifying Alzheimer's vs. Healthy cognition (HC), MCIc (MCI patient who will converted to AD) vs. HC, and MCIc vs. MCInc (MCI patient who not will converted to AD).

In this paper, the author aimed to develop an Alzheimer's stage classification in magnetic resonance images using Deep Learning, specifically the YOLOv3 algorithm. YOLOv3 is known for having high accuracy and less complexity. This study will benefit hospitals and patients with cognitive decline symptoms by examining their MRIs to detect if the patient has Alzheimer's.

II. METHODOLOGY

The proposed system consists of several steps including dataset gathering, image selection, dataset splitting, image annotation, model training and evaluation, and model deployment and testing.

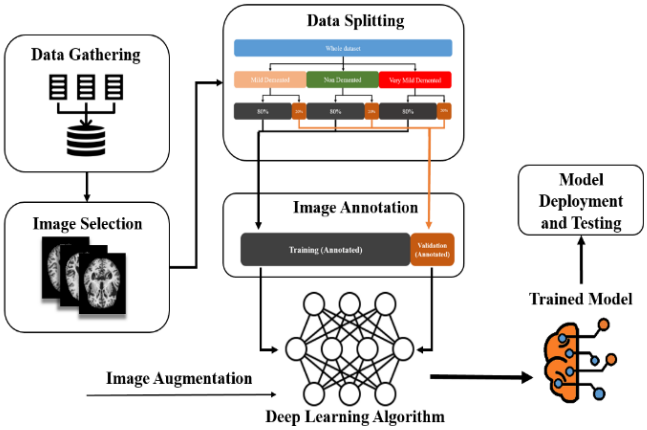

Fig. 1. Workflow Diagram

A. Dataset Gathering

Dataset gathering includes collecting data from various sources compiled into a database. It can be in the form of words, numbers, dates, bool, or in this case, images. It varies on the study what kind of data is needed for training a deep learning model.

B. Image Selection

From the dataset, the data necessary for the study were selected for the training and validation. Any unnecessary data were dropped from the dataset because it might affect the performance of the model.

C. Dataset Splitting

The dataset was split depending on the different classes of the data. Each class is divided into training and validation before annotation.

D. Image Annotation

The dataset is annotated using LabelImg. LabelImg is a graphical image annotation tool written in Python. It can create a bounding box to locate the object needed to detect or classify. The file is saved as an XML file containing the coordinates of the annotated images in PascalVOC format.

E. YOLOv3 Algorithm

YOLOv3 is one of the improved versions of YOLO or You Only Look Once algorithm. YOLO is one of the fastest architectures, it even outperforms other methods [11]. Generally, the YOLOv3 algorithm simply takes an input image, passes it through a neural network (similar to convolutional neural network) to produce an output vector of bounding boxes and class prediction [12]. Some images and videos have a lot of noises, crowds, and other objects in frames that result in the object detection to misinterpret it. YOLOv3 implements a regional method, returning the image region with the greatest probability score that is nearest to the detected object [13]. It is also easy to use and its structure is easy to understand. For this reason, the author chose YOLOv3 as the algorithm for the proposed method.

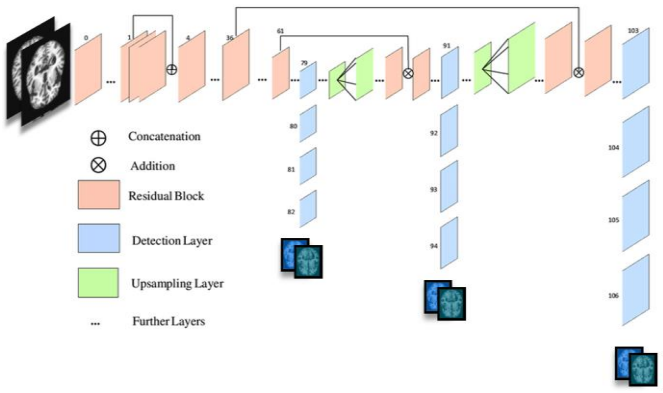

Fig. 2. Network Architecture of YOLOv3

F. Model Training and Environment

Image augmentation is performed on the dataset before feeding it to the algorithm and training the model. The image augmentation process can artificially increase the size or the number of the training dataset by creating modified versions of images based on the dataset. It can scale, crop, flip, and distort the images from the dataset and create new ones. By performing image augmentation, it improves the performance of the model and its ability to generalize [14].

In this paper, each image only goes through the network once to determine the stage of Alzheimer's. YOLOv3 applies a single neural network to the whole image. After that, this network divides the image into regions which provide the bounding boxes and also predict the probabilities for each region and the generated boxes are weighted by the predicted probabilities. The number of experiments was set to 25 and began with a batch size of 4.

G. Model Evaluation

After generating 25 trained models, each one is evaluated and compared to determine which model is trained best using mean average precision (mAP). mAP is the average of Average Precision (AP) values over all classes. Average Precision (AP) is a way to summarize the precision-recall curve into a single value representing the average of all precisions. Using a loop that goes through all precisions/recalls, the difference between the current and next recalls is calculated and then multiplied by the current position. In other words, AP is the weighted sum of precisions at each threshold where the weight is the increase in recall [15]. Mathematically, it is defined as:

$$AP = \sum_{k=0}^{k=n-1}[R(k) - R(k+1)] * P(k) \qquad (1)$$

R = Recalls, P = Precisions
R(n) = 0, P (n) = 1
n = Number of thresholds

Precisely, mean average precision (mAP) is the average of AP values over all classes. AP is calculated individually for each class. This means that there are as many AP values as the number of classes. mAP is defined as:

$$mAP = \frac{\sum_{k=1}^{k=n} AP_k}{n} \qquad (2)$$

AP_k = the AP of class k
n = the number of classes

H. Model Deployment and Testing

The author created a graphical user interface (GUI) for the deployment of the model. The process where done in Anaconda IDE and uses PyQt5 for building the GUI application and the ImageAI for the object detection and classification.

In terms of testing, the model is tested by calculating its classification accuracy score. The number of correct classification is divided to the total number of images tested. The images used in testing are not from the dataset used in training and validation to avoid biases.

$$Accuracy = \frac{\#\ of\ correct\ classification}{total\ \#\ of\ testing\ images} * 100 \qquad (3)$$

III. Results and Discussions

This chapter discusses the dataset preparation result, training and validation loss result, training and validation layer loss result, model evaluation result, and model deployment and testing result.

A. Dataset Preparation Result

The dataset contains magnetic resonance images of the brain obtained from various websites and compiled in kaggle.com. It has four classes: mild demented, moderate demented, non-demented, and very mild demented.

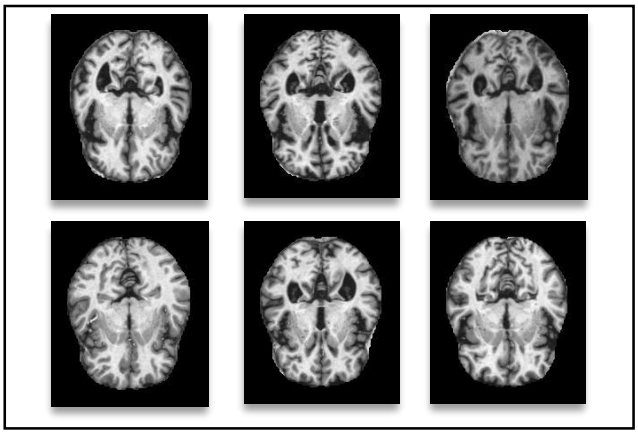

Fig. 3. Sample Images from the Dataset

Out of 6000 magnetic resonance images, 1200 were selected for training and validation of the model. The moderate demented class was dropped from the dataset because it has significant less number of images and can affect the performance of the model. The dataset was split into 3 classes: mild demented, non-demented, and very mild demented. Each classes contains 400 images. Each class was divided into 80% and 20% validation which is equal to 320 and 80 images, respectively.

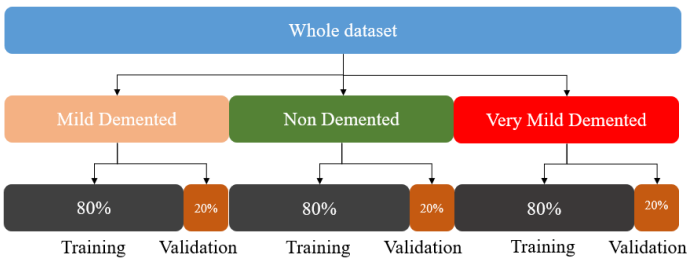

Fig. 4. Data Splitting

After splitting the data according to its classes and to training and validation, it is annotated using LabelImg. Annotations are saved as XML files in PASCAL VOC format. The XML file contains the coordinates of the bounding box in the image.

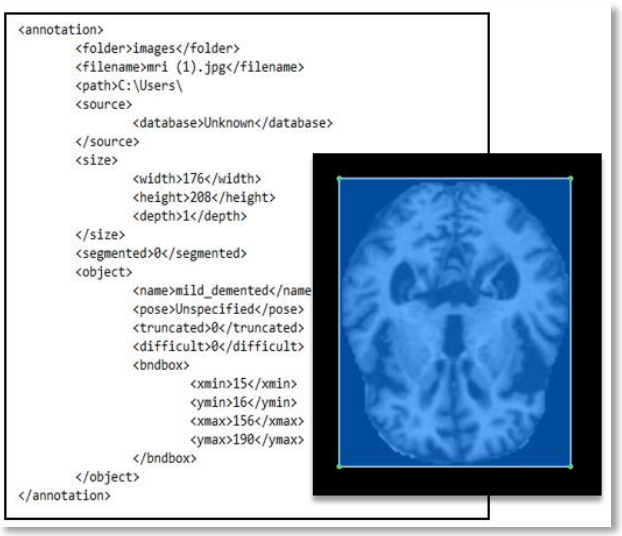

Fig. 5. XML File Content and the Annotated Image

B. Training and Validation Loss Result

Among the 25 epochs, the first trained model got a 10.58% loss and dramatically decreased at the second model that got approximately 1.80% loss. The lowest training loss was obtained in the last epoch. The first epoch got a validation loss of 1.56% loss and got also the lowest loss in the last epoch with 0.1% loss.

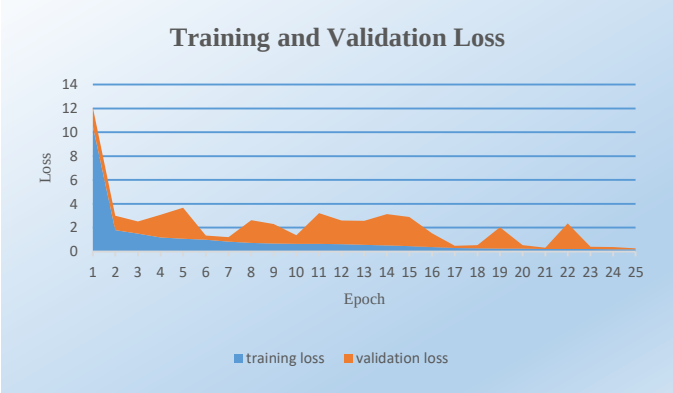

Fig. 6. Training and Validation Loss Result

C. Training and Validation Layer Loss Result

The layer 1 loss decreased from 3.15% loss on the first epoch to 0.17% loss on the last epoch while the layer 2 and 3 loss decreases exponentially from n*10^0 all the way down to n*10^-6. It is noticeable that there was a spike in the 10th epoch. This might be caused by training interruption due to limited runtime of Colab GPU.

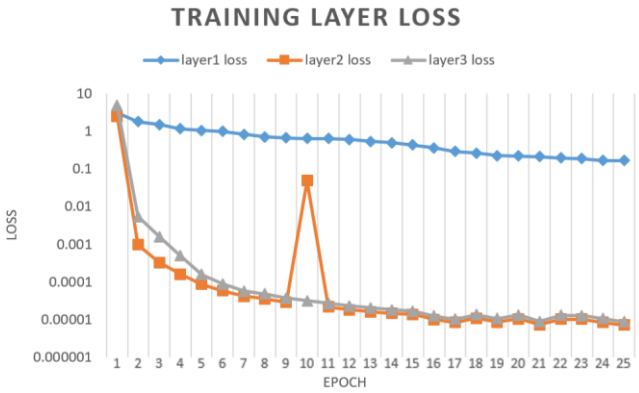

Fig. 7. Training Layer Loss Result

Same with the training layer loss, the layer 1 and layer 2 validation loss decrease exponentially as the number of epoch increases. The validation layer 1 changed only between 0 to 2% as shown in the figure 8.

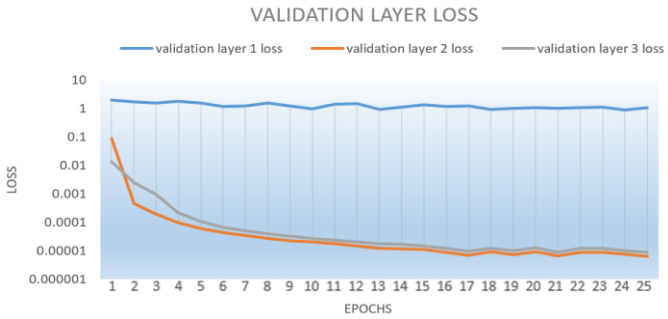

Fig. 8. Validation Layer Loss

D. Model Evaluation Result

The 25 trained models were evaluated using mean average precision (mAP). According to the chart, the first epoch got the lowest mAP of 0.4602 which almost doubled in the second epoch. The 24th epoch obtained the highest mAP, thus this model is used for the deployment and testing. It has a mAP of 0.9578 or 95.78%.

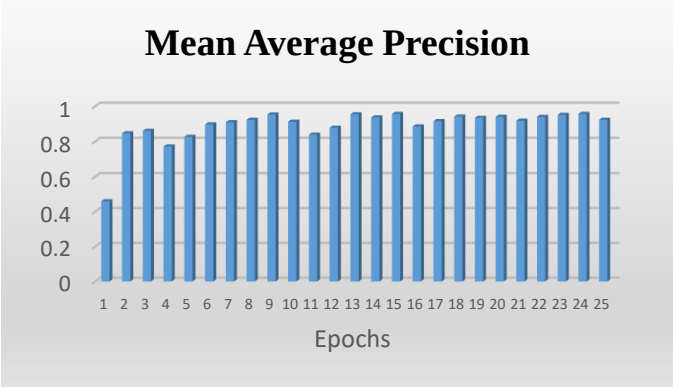

Fig. 9. Mean Average Precision of Each Epoch

E. Model Deployment and Testing

For the model deployment, a GUI application is created with the chosen trained model h5 file along with its JSON configuration file generated after training. It has three functions: detect Alzheimer's stage in an image, in a video, and in live stream via webcam.

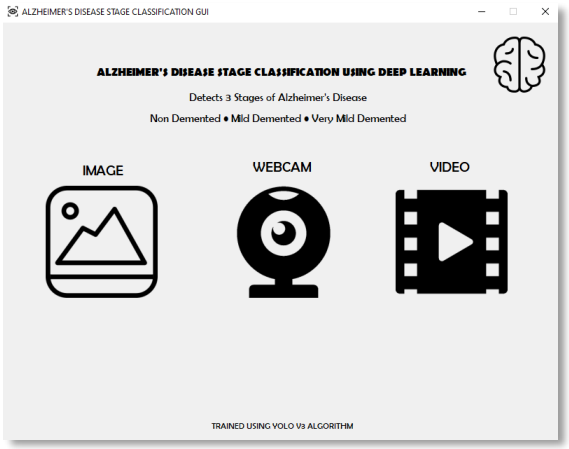

Fig. 10. Alzheimer's Disease Stage Classification GUI

The author took a photo of an actual MRI using a mobile phone and tested the model using the live stream detection via webcam. The model precisely detected the brain in the MRI and correctly identified the Alzheimer's stage. The duration of the live stream took 5 seconds and produced 118 frames and obtained a mAP of 97% and higher.

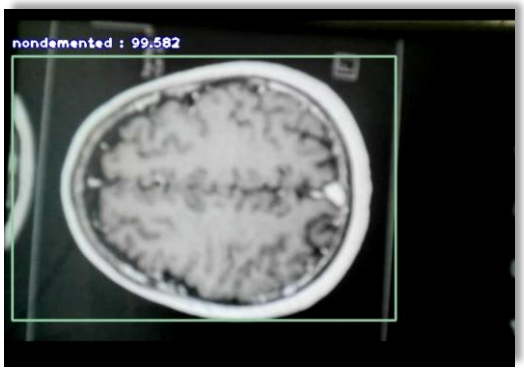

Fig. 11. Screenshot from the Live Stream Detection

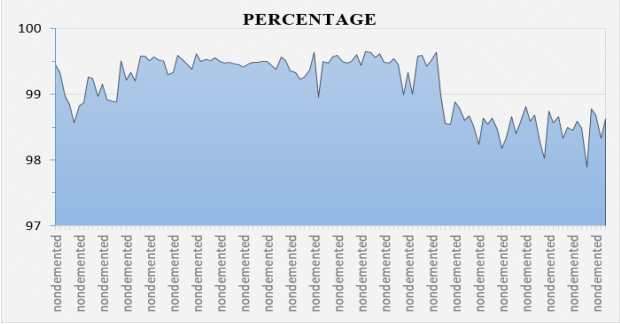

Fig. 12. Live Stream Detection Result

To test the accuracy of the model, 3 random images for each class were selected, 3 non demented, 3 mild demented, and 3 very mild demented. To avoid biases, the images used were obtained from the remaining images on the original source of the dataset which has a total of 6000 images including the 1200 images used for training the model.

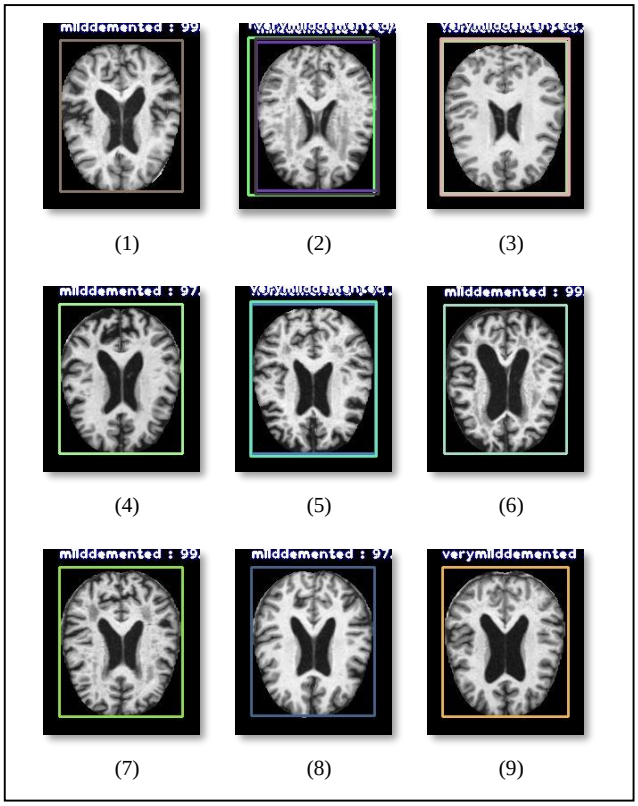

Fig. 13. Sample MRIs for Accuracy Testing

ACTUAL	PREDICTION			RESULT
	Non demented	Mild demented	Very mild demented	
Non demented (1)	0%	99.81%	0%	wrong
Non demented (2)	37.68%	67.08%	31.68%	wrong
Non demented (3)	85.76%	0%	32.71%	correct
Mild demented (4)	0%	97.68%	0%	correct
Mild demented (5)	71.87%	0%	66.02%	wrong
Mild demented (6)	0%	99.87%	0%	correct
Very mild demented (7)	0%	99.42%	0%	wrong
Very mild demented (8)	0%	97.73%	0%	wrong
Very mild demented (9)	0%	0%	99.38%	correct

Table 1. Accuracy Testing Result

The accuracy score of the model is calculated by dividing the number of correct classification to the total number of testing and multiplying it to 100. Out of 9 trials, only 4 were correctly predicted. Although the trained model got a 95.48% mean average precision, it only obtained an accuracy score of 44.44%.

IV. CONCLUSION

Alzheimer's disease is one of the most common types of dementia. It is commonly occurring to elderly people, and there is still no cure for it yet. The only thing that can do is to reduce its progression by detecting it as early as possible. For this reason, a deep learning model is created using YOLOv3 Algorithm to detect the stage of Alzheimer's disease to a magnetic resonance image. Of all the 25 trained models, the 24th epoch is chosen as the model for the deployment and testing. It obtained a mAP of 95.48% but only obtained an accuracy of 44.44%. This suggests that YOLOv3 is a good algorithm for detecting an object in an image but it performs poorly in classifying Alzheimer's disease stage in a magnetic resonance image.

For the future studies, the author suggests increasing the number of dataset and epoch when training deep learning models. The author also suggests using different deep learning algorithm that is good for classification.

ACKNOWLEDGEMENT

The author would like to express his sincere gratitude to the College of Engineering, Architecture, and Fine Arts of Batangas State University, especially to the Electrical and Computer Engineering Department and to Dr. Alvin S. Alon, Head of Digital Transformation Center for his guidance and advice in accomplishing this paper.

REFERENCES

[1] "2020 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 16, no. 3, pp. 391–460, 2020, doi: 10.1002/alz.12068.

[2] "Number of disease reports," *Alzheimer's Dis. Res.*, 2021.

[3] A. Association and Alzheimer's Association, "FDA-approved t reatments for Alzheimer's," no. Mci, pp. 1–5, 2019.

[4] C. A. Taylor, E. D. Bouldin, K. J. Greenlund, and L. C. Mcguire, "Comorbid Chronic Conditions among Older Adults with Subjective Cognitive Decline, United States, 2015-2017," *Innov. Aging*, vol. 4, no. 1, pp. 1–10, 2020, doi: 10.1093/geroni/igz045.

[5] R. Alroobaea *et al.*, "Alzheimer's Disease Early Detection Using Machine Learning Techniques," *Res. Sq.*, pp. 1–16, 2021.

[6] S. A. Mofrad, A. Lundervold, and A. S. Lundervold, "A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease," *Comput. Med. Imaging Graph.*, vol. 90, no. February, 2021, doi: 10.1016/j.compmedimag.2021.101910.

[7] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, "Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI," *Front. Bioeng. Biotechnol.*, vol. 8, no. October, 2020, doi: 10.3389/fbioe.2020.534592.

[8] L. J. Herrera, I. Rojas, H. Pomares, A. Guillén, O. Valenzuela, and O. Baños, "Classification of MRI images for Alzheimer's disease detection," *Proc. - Soc. 2013*, pp. 846–851, 2013, doi: 10.1109/SocialCom.2013.127.

[9] H. A. Helaly, M. Badawy, and A. Y. Haikal, "Deep Learning Approach for Early Detection of Alzheimer's Disease," *Cognit. Comput.*, no. September, 2021, doi: 10.1007/s12559-021-09946-2.

[10] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, "Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning," *Front. Neurosci.*, vol. 14, no. May, pp. 1–19, 2020, doi: 10.3389/fnins.2020.00259.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[12] S. Kumar, D. Yadav, H. Gupta, O. P. Verma, I. A. Ansari, and C. W. Ahn, "A novel yolov3 algorithm-based deep learning approach for waste segregation: Towards smart waste management," *Electron.*, vol. 10, no. 1, pp. 1–20, 2021, doi: 10.3390/electronics10010014.

[13] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Mach. Learn. with Appl.*, vol. 6, no. March, p. 100134, 2021, doi: 10.1016/j.mlwa.2021.100134.

[14] J. Brownlee, "How to Configure Image Data Augmentation in Keras," Apr. 12, 2019. https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/ (accessed Dec. 03, 2021).

[15] A. F. Gad, "Evaluating Object Detection Models Using Mean Average Precision (mAP)," 2020. https://blog.paperspace.com/mean-average-precision/ (accessed Dec. 03, 2021).