Machine Learning Approaches To Diagnose Alzheimer's Disease

Vishruth Puttu

ABSTRACT

This study uses machine learning (ML) approaches to model the disease severity of patients with Alzheimer's Disease (AD). Magnetic Resonance Imaging (MRI) images of AD patients were labeled as non-dementia, very mild dementia, mild dementia, and moderate dementia by a physician and were used to train, validate, and test 3 different models. The models included a convolutional neural network (CNN), VGG-16, and VGG-16 + SMOTE. The CNN model was a basic neural network. The VGG-16 model was pre-trained on image data, and the VGG-16 + SMOTE employed a data preprocessing technique that balanced the number of images in each class before training on the VGG-16 pre-trained CNN. The CNN model was the most effective, with the highest accuracy of the three models: 0.98. The VGG-16 + SMOTE model is the second most effective, with an accuracy of 0.97. The least effective model is the one with only VGG-16, with an accuracy of 0.87. The data suggests that the CNN model accurately diagnoses the level of dementia in AD patients and has the potential to be used in medical practices. While using SMOTE significantly improved the accuracy of the least accurate model, the VGG-16 + SMOTE model is still slightly less effective than a regular CNN and requires much more time to run. Thus, a regular CNN is more accurate and efficient.

Keywords

1. Magnetic Resonance Imaging (MRI)
2. Alzheimer's Disease (AD)
3. Machine Learning (ML)
4. Convolutional Neural Network (CNN)
5. VGG-16
6. Synthetic Minority Oversampling Technique (SMOTE)

INTRODUCTION

Background and Context

Alzheimer's Disease (AD), a disease that destroys memories and reduces mental capabilities, stands as the leading cause of dementia[1], robbing millions of their cognitive faculties and placing an immense strain on healthcare systems and societies. Treating people with AD cost the world $1313.4 billion in 2019[2], with few of the treated showing signs of improvement and the rest continuing to suffer. Nearly 1 in every 3 AD patients experience moderate to severe pain[3] in addition to their dementia.

Several different types of medical images could be used to diagnose AD. This paper, however, focuses on the most common one: magnetic resonance imaging (MRI). MRIs detect changes in the brain and identify AD.

Problem Statement and Rationale

This paper aims to investigate the accuracy of different ML models in predicting dementia severity in AD patients. Current research on ML in MRIs for AD includes various models with varying levels of effectiveness for diagnosing AD[4]. Even though several models have been employed, only some effectively process grid-like data, such as images[9] studies have compared to determine which is more effective in diagnosing AD and its severity.

Significance and Purpose

This study would significantly reduce the time researchers spend diagnosing the severity of a patient's AD and identifying suitable models to make this diagnosis. AD symptoms are different in each of the three stages of the disease: the early stage, the middle stage, and the late stage. In the early stage, symptoms include forgetting about recent conversations or events, misplacing items, forgetting the names of places and objects, having trouble thinking of the right word you intend to use, asking questions repetitively, showing poor judgment, or finding it more challenging to make decisions, and becoming less flexible and more hesitant to try new things[5]. In the middle stage, symptoms include increasing confusion and disorientation – for example, getting lost or wandering and not knowing what time of day it is; obsessive, repetitive, or impulsive behavior; believing untrue things (delusions) or feeling paranoid and suspicious about carers or family members; problems with speech or language (aphasia); disturbed sleep; changes in mood, such as frequent mood swings, depression, and feeling increasingly anxious, frustrated, or agitated; difficulty performing spatial tasks, such as judging distances; and seeing or hearing things that other people do not (hallucinations)[5]. In the later stage, symptoms include difficulty eating and swallowing (dysphagia), difficulty changing position or moving around without assistance, weight loss – sometimes severe unintentional passing of urine (urinary incontinence) or stools (bowel incontinence), gradual loss of speech, and significant problems with short- and long-term memory[5].

Treating AD varies depending on the severity. If the patient has mild to moderate AD, they may be prescribed cholinesterase inhibitors or immunotherapy drugs[6]. Cholinesterase inhibitors, which decrease the breakdown of acetylcholine[7], include galantamine (prevents the breakdown of acetylcholine and stimulates nicotinic receptors to release more acetylcholine in the brain), rivastigmine (prevents the breakdown of acetylcholine and butyrylcholine (a chemical similar to acetylcholine) in the brain), and donepezil (a rivastigmine patch that prevents the breakdown of acetylcholine in the brain)[6].

In the 1900s, nuclear magnetic resonance was developed into the MRIs that are now commonly used for medical diagnosis. The first MRI images were produced in the early 1970s, and the first live human subject was imaged in 1977. MRI machines became commercially available in the 1980s and are now commonly used for imaging internal body structures, especially soft tissues like the brain, making them very useful in diagnosing AD[8].

Objectives

This research aims to assess different algorithms and their effectiveness in diagnosing the severity of AD. The machine learning (ML) approaches include CNNs, VGG-16, and VGG-16 with Synthetic Minority Oversampling Technique (SMOTE). A CNN is a specialized class of neural networks designed to effectively process grid-like data, such as images[9]. VGG-16 and SMOTE are a bit more complex than this; VGG-16 is a CNN that is 16 layers deep[10] and has been previously trained on a large set of images CNN, and SMOTE is a technique used in ML to address imbalanced datasets[11], which is helpful in this scenario as many dementia levels have significantly lower amounts of images than other levels.

Scope and Limitations

Limitations of the study include the lack of MRI image datasets of AD patients and the imbalance between the amount of images in each level of dementia.

Methodology Overview

The dataset was downloaded from the machine learning and data science website Kaggle[12,] and the CNN, VGG-16, and VGG-16 + SMOTE models were tested on it. Each model's accuracy and confusion matrices were used to conclude which is the most effective at diagnosing the severity of dementia in AD patients.

METHODS

Research Design

This study compared different machine learning algorithms to find the most effective one in diagnosing dementia severity in AD patients based on MRI images.

Participants/Sample and Data Collection

The dataset included 5,120 training images and 1,280 test images[13]. It included 3,200 MRI images of people without dementia, 2,240 MRI images of people with very mild dementia, 896 images of people with mild dementia, and 64 MRI images of people with moderate dementia[13]. The code can be found at https://github.com/VishruthPuttu/Machine-Learning-Approaches-To-Diagnose-Alzheimer-s-Disease.

Variables and Measurements

Figure 1 shows the CNN model architecture, and Figure 2 shows the VGG-16 model architecture. Using the Adam optimizer and sparse categorical cross-entropy loss function, the CNN model was trained for ten epochs. The VGG-16 models were trained for 75 epochs, with a learning rate of 0.001, and also used the Adam optimizer and sparse categorical cross-entropy loss function. Each batch contained 32 images. The epochs were decided when the loss function reached a stable minimum, and the accuracy plateaued. Hyperparameters were tuned by tweaking the number of layers, the number of nodes per layer, and the learning rate.

Procedure

Figures 1 and 2 show the architecture for the CNN and VGG models, respectively. Each model was implemented using Python and the NumPy, Pandas, matplotlib, and TensorFlow libraries.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	320
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_2 (Conv2D)	(None, 28, 28, 64)	36,928
flatten (Flatten)	(None, 50176)	0
dense (Dense)	(None, 64)	3,211,328
dense_1 (Dense)	(None, 4)	260

Total params: 3,267,332 (12.46 MB)
Trainable params: 3,267,332 (12.46 MB)
Non-trainable params: 0 (0.00 B)

Figure 1: CNN model architecture. Created by student researcher.

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1,792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36,928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
global_max_pooling2d (GlobalMaxPooling2D)	(None, 512)	0
dense (Dense)	(None, 256)	131,328
dense_1 (Dense)	(None, 256)	65,792
dropout (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 4)	1,028

Total params: 14,912,836 (56.89 MB)
Trainable params: 198,148 (774.02 KB)
Non-trainable params: 14,714,688 (56.13 MB)

Figure 2: VGG-16 model architecture (also applies to VGG-16 + SMOTE). Created by student researcher.

Data Analysis

The accuracy and confusion matrices were used for performance assessment to compare the regular CNN, VGG-16, and VGG-16 with SMOTE models.

DISCUSSION

Results

	Accuracy
CNN	0.98
VGG-16	0.87
VGG-16 + SMOTE	0.97

Table 1: Accuracies of each model used. Created by student researcher.

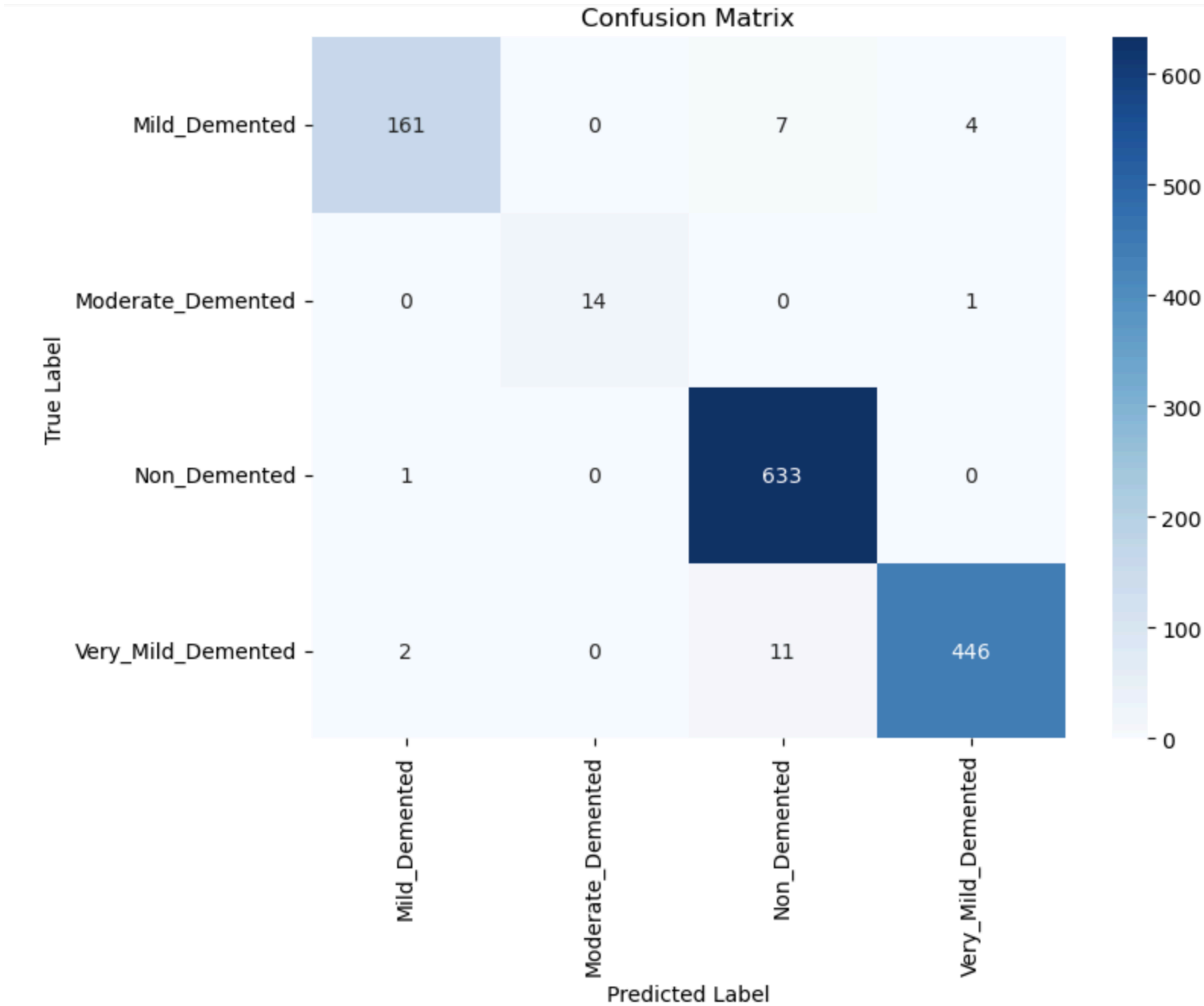

Figure 3: Confusion matrix of the CNN model. Created by student researcher.

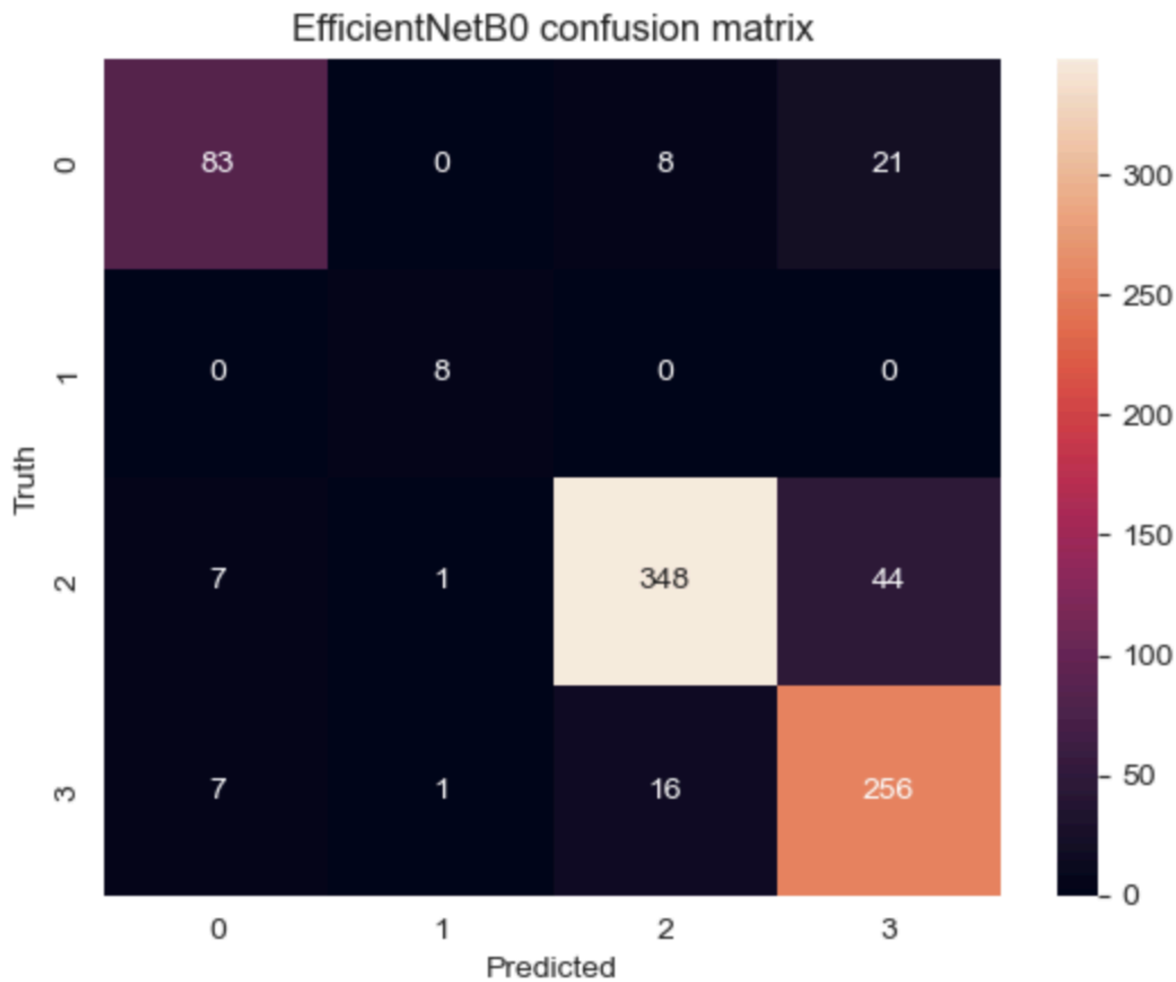

Figure 4: Confusion matrix of VGG-16 model. 0: Mildly Demented, 1: Moderately Demented, 2: Non-Demented, 3: Very Mildly Demented. Created by student researcher.

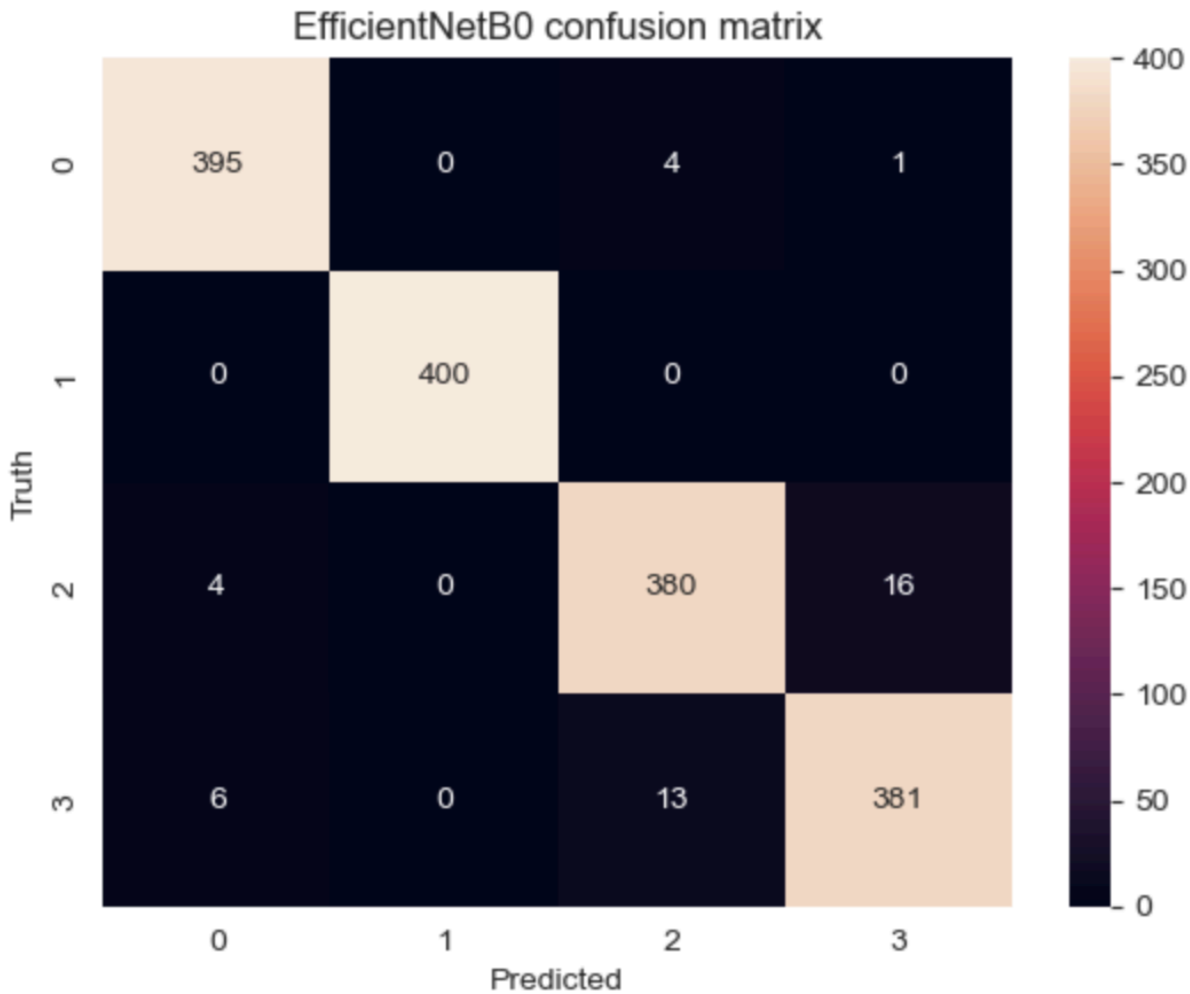

Figure 5: Confusion matrix of VGG-16 + SMOTE model. 0: Mildly Demented, 1: Moderately Demented, 2: Non-Demented, 3: Very Mildly Demented. Created by student researcher.

As seen in Table 1, the CNN model is the most accurate with an accuracy of 0.98, the VGG-16 + SMOTE model is second most accurate with an accuracy of 0.97, and the VGG-16 model is the least accurate with an accuracy of 0.87.

Based on the confusion matrices, we can see that CNN has the highest misidentifications in the very mild dementia category, with 13 out of 459 images being mislabeled (Figure 3). However, this is very reasonable considering the category's number of images. As for the VGG-16 and VGG-16 + SMOTE models, they had the highest misidentifications in the non-dementia category, with 52 out of 400 images being mislabeled in the VGG-16 model (Figure 4) and 20 out of 400 images being mislabeled in the VGG-16 + SMOTE model (Figure 5). Even though both models have mislabeling, this proves how useful SMOTE is because its class balancing led to a decrease in mislabeled images.

Restatement of Key Findings

The most effective model is the CNN, with an accuracy of 0.98. The second most effective model is VGG-16 + SMOTE**,** with an accuracy of 0.97. The least effective model is the one with only VGG-16, with an accuracy of 0.87.

Applying SMOTE to the dataset before training a VGG-16 model significantly improved its performance, making its accuracy comparable to the standard CNN's. With SMOTE, the VGG-16 model showed higher accuracy, likely due to class imbalance in the dataset. However, once SMOTE was applied, the VGG-16 achieved accuracy levels similar to the CNN, demonstrating strong performance. The difference in accuracy between the two models could be attributed to potential mislabeled data or other limitations within the training or testing datasets, such as noise or incomplete representation of the target classes. The pre-trained VGG-16 model may also have been trained to a set of images that does not generalize well to the MRI images used in this study.

Implications and Significance

The CNN model is comparable to human physicians and can be used to train medical students and residents and diagnose dementia in AD patients independently. This shows the potential for ML applications in healthcare settings. The accuracy of the VGG-16 + SMOTE model is greater than that of the only VGG-16 model, which highlights how SMOTE can improve the accuracy of ML models.

Connection to Objectives

The objective of this research, to find the most effective ML model to determine dementia levels in AD patients, has been met. We have come to a conclusion that is very well supported by our data and that can be applied in healthcare as we aspired.

Recommendations

Future research needs to focus on using more extensive datasets and different versions of the models we used and others we did not use. For example, researchers could add extra layers onto VGG-16 to see if that makes the model more accurate. They could also incorporate EHR data to predict if a patient is at risk for developing AD/dementia based on their medical history and MRIs. Researchers must also account for overfitting in the VGG-16 models and might want to consider using simpler models that take less time to run to solve this issue. Along with this, researchers might want to also consider using SMOTE with a regular CNN and using transfer learning in future work.

Limitations

Many different factors could have impacted the results of our research. These include the lack of MRI images of people with moderate dementia, the relatively small size of the dataset we use, and overfitting.

All of these could have deflated or inflated the accuracy of the models. However, as these limitations were constant throughout all the models used, their impact could also be minimal.

Closing Thought

The use of ML in healthcare is promising. ML approaches are coming very close to the capabilities of human clinicians and need more push to become fully useful. Using this technology, clinicians could give MRI results to patients in minutes instead of weeks.

For this to happen, however, more people must research different models and parts of existing models to make them more efficient. By doing this, people will enjoy lower levels of human error in interpreting MRIs and quick results. This rapid process will help people determine their dementia faster and treat it faster before it worsens.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to Dr. Daniel Plymire from the University of Texas at Arlington for offering me precious advice and a lot of support with my research study.

CITATIONS

1. World Health Organization. *Dementia*. (2023, March 15). https://www.who.int/news-room/fact-sheets/detail/dementia
2. A. Wimo, K. Seeher, R. Cataldi, E. Cyhlarova, J. L. Dielemann, O. Frisell, M. Guerchet, L. Jönsson, A. K. Malaha, E. Nichols, P. Pedroza, M. Prince, M. Knapp, & T. Dua. The worldwide costs of dementia in 2019. Alzheimer's & dementia : the journal of the Alzheimer's Association, 19(7), 2865–2873. (2023). https://doi.org/10.1002/alz.12901
3. W. Achterberg, S. Lautenbacher, B. Husebo, A. Erdal, & K. Herr. Pain in dementia. Pain reports, 5(1), e803. (2019). https://doi.org/10.1097/PR9.0000000000000803
4. I. Malik, A. Iqbal, Y. H. Gu, & M. A. Al-Antari. Deep Learning for Alzheimer's Disease Prediction: A Comprehensive Review. Diagnostics (Basel, Switzerland), 14(12), 1281. (2024). https://doi.org/10.3390/diagnostics14121281
5. *Symptoms - Alzheimer's Disease*. NHS. (2024, July 4). https://www.nhs.uk/conditions/alzheimers-disease/symptoms/
6. *How is alzheimer's disease treated? | National Institute on Aging*. National Institute on Aging. (2023, September 12). https://www.nia.nih.gov/health/alzheimers-treatment/how-alzheimers-disease-treated
7. R. Singh & N. M. Sadiq. Cholinesterase inhibitors. National Center for Biotechnology Information. (2023, July 17). https://www.ncbi.nlm.nih.gov/books/NBK544336/
8. *MRI uses fundamental physics for clinical diagnosis*. American Physical Society. (2006, July 3). https://www.aps.org/apsnews/2006/07/mri-fundamental-clinical-diagnosis
9. goelaparna1520. *Convolutional Neural Network (CNN) in Machine Learning*. GeeksforGeeks. (2024, March 13). https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-learning/
10. *ImagePretrainedNetwork*. vgg16. (n.d.). https://www.mathworks.com/help/deeplearning/ref/vgg16.html
11. S. Satpathy. *Smote for imbalanced classification with python*. Analytics Vidhya. (2024, August 5). https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/

12. Your machine learning and Data Science Community. Kaggle. (2010, April). https://www.kaggle.com/
13. @dataset{alzheimer_mri_dataset,

author = {Falah.G.Salieh},

title = {Alzheimer MRI Dataset},

year = {2023},

publisher = {Hugging Face},

version = {1.0},

url = {https://huggingface.co/datasets/Falah/Alzheimer_MRI}

}