Topik : Alzheimer’s Disease Diagnosis Using Machine Learning

Nomor Kelompok : 1

Kelas : LK01

NIM - Nama Lengkap Ketua Kelompok : 2602055281 - Angel Priscilla Salim

NIM - Nama Lengkap Anggota : 2602117866 - Winanda Hartadi

Kontribusi dalam Project

|  |  |  |  |
| --- | --- | --- | --- |
| **NIM** | **Nama Lengkap** | **Kontribusi** | **Keterangan** |
| 2602055281 | Angel Priscilla Salim | 1. Melakukan review artikel sejumlah 10 | Artikel yang saya review dapat dilihat pada link di samping artikel nomor 1 s/d 10 (<https://docs.google.com/spreadsheets/d/1O6ZF3xcBRtELEhl6faPQ1Xr4SUZW56BW/edit#gid=1503167262> ) |
| 2. Mengerjakan progress week 2 | Menentukan state-of-the-art method of the topic, top 2 dataset yang paling banyak digunakan, dan evaluation matrix |
| 3. Mengerjakan progess week 3 | Membuat workflow method |
| 4. Mengerjakan progress week 4 | Menentukan parameter yang digunakan dalam code |
| 5. Mengerjakan progress week 7 | Membuat paper, PPT, record video presentasi |
| 6. Mengerjakan laporan akhir | Semua bagian kecuali result and discussion, dan conclusion |
| 2602117866 | Winanda Hartadi | 1. Melakukan review artikel sejumlah 10 | Artikel yang saya review dapat dilihat pada link di samping artikel nomor 11 s/d 20 (<https://docs.google.com/spreadsheets/d/1O6ZF3xcBRtELEhl6faPQ1Xr4SUZW56BW/edit#gid=1503167262> ) |
| 2. Mengerjakan progress week 2 | Menentukan state-of-the-art method of the topic, top 2 dataset yang paling banyak digunakan, dan evaluation matrix |
| 3. Mengerjakan progess week 3 | membuat latar belakang untuk keperluan research paper |
| 4. Mengerjakan progress week 4 | memulai membuat code dalam vscode dan google colabs, reserach keperluan code seperti library untuk deep learning dan lain-lain |
| 5. Mengerjakan progress week 5 | melanjutkan membuat code |
| 6. menegerjakan progress week 6 | merevisi code |
| 7. mengerjakan progress week 7 | Membuat paper, PPT, record video presentasi |
| 8. mengerjakan code akhir finalisasi, laporan | bagian result discussion dan conclussion |

Alzheimer’s Disease Diagnosis Using Machine Learning

**Introduction**

Alzheimer’s Disease (AD) is a progressive mental deterioration and incurable neurodegenerative disease that predominantly affects the elderly population, characterised by language and behaviour problems, disorientation, mood swings, loss of motivation, inefficiency in managing self-care, cognitive decline, memory loss, and impaired daily functioning. It is the most common cause of dementia, responsible for approximately 60-70% of all cases globally.[2] As the global population ages, the prevalence of Alzheimer’s Disease is expected to rise significantly, posing a considerable burden on healthcare systems and society as a whole.

Dementia has emerged as a priority in health and social care, particularly in ageing societies. Predominantly affecting the elderly, approximately 2% of dementia cases occur in individuals under the age of 65 (Alickovic & Subasi, 2020). World Health Organization stated that around 50 million people are diagnosed with dementia and approximately 10 million new diagnosed cases every year (World Health Organization, 2020b). According to projections, it is predicted that in 2023, there will be over 75 million people living with dementia. Dementia is currently a more critical medical concert than other illnesses (World Health Organization, 2020a). The major challenge in diagnosing dementia is due to the lack of a standarized detection test (Stamate et al., 2020).[9] Until these days, a cure for dementia diseases are yet to be found, making it one of the most life-threatening diseases. In 2015, a number of 1.9 million of death were recorded due to dementia. [2]

The use of machine learning methods to increase the precision and efficiency of Alzheimer’s Disease diagnosis has gained attention in the last few years. Machine learning algorithms have the potential to analyze vast amounts of heterogeneous data, including clinical assessments, neuroimaging data (such as magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF) biomarkers), genetic information, and even wearable sensor data, to identify patterns and biomarkers indicative of Alzheimer’s Disease.

Nowadays, clinical evaluation which contains cognitive testing, medical records evaluation, and neuroimaging scans is a major component in the diagnosis of Alzheimer’s Disease. However, these methods can be costly, time consuming, and subjective, which often leading to delayed diagnosis and treatment initiation. Furthermore, with misdiagnosis rates up to 30-40%, clinical diagnostic accuracy remains as a major difficulty, especially in the early stages of the illness.

The capability of machine learning to combine several data types and identify complex patterns that could be invisible to human eyes is one of the main benefits of utilizing it to diagnose Alzheimer’s Disease. Machine learning models can determine the differences between pathological changes related to Alzheimer’s Disease and normal ageing by training them on huge datasets that include both healthy subjects and subjects with Alzheimer’s Disease.

Lately, a type of deep learning algorithms knows as convolutional neural networks (CNN) has been gaining popularity. CNNs have shown potential in the analysis of medical imaging data and the diagnosis of a number of diseases, which includes Alzheimer’s Disease. Since CNN is able to learn hierarchical features from raw data, it is particularly suitable for tasks involving picture recognition and classification. CNN designs have been effectively used in medical imaging applications including ResNet50, ResNet152, DenseNet, and EfficientNetB7.

ResNet50 is well-known for its depth and performance, especially in image classification tasks. The residual connection is an architecture that allows training deeper networks without the problem of vanishing gradients. A number of residual blocks are stacked together in the ResNet50 model. Each residual block consists of convolutional layers followed by shortcut connections. The shortcut connections allow the network to learn residual functions, making it easier to optimize and thus improve accuracy.

ResNet152 is the extended version of the original ResNet50 architecture, introduced by Microsoft Research in 2015. Therefore, it has 152 layers and is designed to deal with even deeper networks effectively, hence addressing the degradation problem encountered in the training of very deep neural networks. The degradation problem means the phenomenon where the accuracy of a deep neural network saturates and degrades rapidly as the network’s depth increases. It occurs due to optimization difficulties associated with training deep networks. Resnet152 solves this problem through skip connections or shortcut connections, which allow the network to learn residual functions. These shortcuts allow the network to skip over a few layers so that it could learn the identity function more easily and solve the degradation problem.

A deep learning architecture named DenseNet (Dense Convolutional Network) was proposed in 2017. It provides extensive connections between layers, differentiating it from conventional CNN. Unlike ResNet architecture that uses residual connections, every layer in DenseNet is connected to every other layer in a feed-forward fashion. This architecture pattern provides every layer with direct access to features from any earlier layer, allowing feature reuse and improving gradient flow through the network.

EfficientNetB7 was developed by researches at Google AI in 2019 as part of the EfficientNet family of CNN. In essence, EfficientNetB7 provides a scalable and efficient method for constructing deep neural networks that attain state-of-the-art performance in many image classification tasks while keeping computational efficiency. More specifically, it was designed to balance the model size, accuracy, and computational cost, thereby being particularly suited for resource-constrained scenarios, like mobile and edge devices. This architecture is highly focused on computational efficiency and scalability and is thus able to achieve competitive performance with fewer parameters compared to Resnet and DenseNet.

Several studies have proposed Alzheimer’s Disease diagnosis and detection system that utilize various classification techniques and machine learning methods. Some of the previous studies on Alzheimer’s Disease diagnosis are focused on developing models to analyze the anatomical or structural brain images such as MRI and brain functionality to detect any defect or disorder. Furthermore, segmentation tasks were viewed as classification issues, and the voxel, region, or patch-based approaches mostly relied on manually constructed features and feature representations.

In 2021, Xiaomu Tang and Jie Liu proposed Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM) algorithms to classify and predict the different disease progress of Alzheimer’s Disease into four groups, such as Cognitive Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer’s Disease (AD). The study was conducted on 560 subjects aged between 60 and 90 who had received an education or participated in work for at least 6 years and met the diagnostic criteria, which was obtained from ADNI (Alzheimer’s Disease Neuroimaging Initiative). The simulation result shows that the proposed RF algorithm had the best prediction effect on different disease processes of Alzheimer’s Disease.[1]

Another study for predicting Alzheimer's Disease was conducted in 2021 by P. C. Muhammed Raees and Vinu Thomas. The study was conducted to predict and classify subjects into three groups, such as Mild Cognitive Impairment (MCI), Alzheimer’s Disease (AD), and Normal classes. A total of 200 subjects diagnosed with early AD, 400 subjects with MCI, and 200 elderly normal control subject were obtained from ADNI dataset of the University of Southern California. The experiments were conducted by utilizing SVM and different models of Deep Neural Network (DNN) algorithms and resulting the highest prediction accuracy of 80-90% for DNN models.[2]

In 2022, Duaa Alsaaed and Samar Fouad Omar proposed convolutional neural network (CNN) model ResNet50 with conventional Softmax, SVM and RF algorithms to detect AD using brain imaging, including MRI and show the parietal atrophy of AD cases. The experiment was conducted on a MRI dataset obtained from ADNI and MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease) containing 549 subjects including subjects with AD and subjects with normal control. The result shows that the model ResNet50 combined with Softmax achieved a higher accuracy than most of the state-of-the-art models.[3]

In 2020, Loris Nanni, Matteo Interlenghi, Sheryl Brahnam, Christian Salvatore, Sergio Papa, Raffaello Nemni, Isabella Castiglioni, and the Alzheimer’s Disease Neuroimaging Initiative (ADNI) proposed CNN, SVM, Transfer-Learning, and Deep Learning approach to compare deep or transfer learning and conventional machine learning when applied to the same neuroimaging studies for the early diagnosis and prognosis of AD. The study was conducted on MRI brain images obtained from ADNI dataset containing more than 600 subjects, including criteria for different diagnostic classes of patients, such as CN subjects, MCI patients, AD patients, Salvatore-509, and Moradi-264. The study shows that a deep learning network trained from scratch on few hundreds of MRI volumes obtained lower performance than either the fusion of conventional machine learning (ML) systems and the ensemble of 2D pre-trained CNNs, due to the limited sample of images used for training.[4]

Another study for detecting AD based on MRI findings using ML models was conducted in 2020 by Gopi Battineni, Nalini Chintalapudi, Francesco Amenta, and Enea Traini. The study was conducted on MRI images from 150 subjects aged 60 and above with fourteen distinct features related to standard AD diagnosis, obtained from Alzheimer's Disease Research Center (ADRC) of Washington University. The study proposed Naive Bayes (NB), Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), and SVM algorithms and the Receiver Operating Characteristic (ROC) curve metric were used to validate the model performance. This study conducted three independent experiments to evaluate each model. In the first experiment, ANN generated the highest accuracy in terms of ROC at 0.812. NB achieved the highest ROC of 0.942 in the second experiment. Last experiment were conducted to combine the four models creating an esemble or hybrid modelling and resulted an improved accuracy of ROC at 0.991.[5]

Gopio Battineni, Mohmmad Amran Hossain, Nalini Chintalapudi, Enea Traini, Venkata rao Dhulipalla, Mariappan Ramasamy, and Francesco Amenta proposed RF, SVM, NB, Logistic Regression (LR), and ensemble learning models, such as Gradient Boosting and AdaBoost to propose a framework based on supervised learning classifiers in the dementia subject categorization as either AD or non-AD based on longitudinal brain MRI features. A total sample of 150 subjects aged between 60 and 96 years of age including demented and non demented subjects were obtained from OASIS (Open Access Series of Imaging Studies) of neurology and patient database was acquired from the longitudinal pool of ADRC at Washington University. The study shows that three classifiers (RF, NB, and Gradient Boosting) achieved the highest average AUC score of 0.98. However, by considering both classification accuracy metric and AUC, the gradient boosting technique can seem a better potential classifier than others.[6]

In 2022, a study to present an AD detection framework consisting of image denoising of an MRI input data set using an adaptive mean filter, preprocessing using histogram equalization, and feature extraction by Haar wavelet transform was conducted by Mustafa Kamar, A. Raghuvira Pratap, Mohd Naved, Abu Sarwar Zamani, P. Nancy, Mahyudin Ritonga, Surendra Kumar Shukla, and F. Sammy. The study proposed LS-SVM-RBF (Least Squares – Support Vector Machine – Radial Basis), SVM, KNN, and RF algorithms for classification conducted on 416 samples obtained from OASIS. The study shows that LS-SVM-RBF algorithm had a higher specificity and accuracy than the other classifiers. The KNN algorithm outperforms the other classifiers in terms of sensitivity and recall.[7]

A study on investigating the usefulness of rule extraction in the assessment of AD using DT and FR algorithms and integrating the extracted rules within an argumentation-based reasoning framework was conducted in 2020 by K. G. Achilleos, S. Leandrou, N. Prentzas, P. A. Kyriacou, A. C. Kakas, and C. S. Pattichis. The study aims to classify subjects into two groups, such as AD subjects and NC subjects. Dataset contains of brain MRI images from a total of 237 subjects aged between 55 and 90 years of age was obtained from ADNI. The study shows that DT classifier had the accuracy of 77%, sensitivity of 66%, and specificity of 88%. RF classifier had the accuracy of 74%, sensitivity of 56%, and specificity of 91%. Argumentation rules had the accuracy of 91%, sensitivity of 87%, and specificity of 95%. Hence, it was proved that augmentation models can be successfully applied.[8]

In 2021, Roobaea Alroobaea, Seifeddine Mechti, Mariem Haoues, Saaed Rubaiee,

Anas Ahmed, Murad Andejany, Nicola Luigi Bragazzi, Dilip Kumar Sharma, Bhanu Prakash

Kolla, and Sudhakar Sengan conducted common supervised machine learning techniques

for automatic AD detection. The experiment was conducted on a total of 2665 CN subjects,

3924 MCI subjects, and 1731 AD subjects which data was obtained from ADNI and OASIS

brain dataset. The study shows that the best accuracy values provided by the machine

learning classifiers are 99.43% and 99.10% given by respectively, logistic regression and

SVM using ADNI dataset, whereas for the OASIS dataset achieved 84.33% and 83.92%

given by respectively logistic regression and random forest.[9]

In 2020, a study on Deep Convolutional Neural Network (DCNN) and VGG-16

inspired CNN (VCNN) models implementation to classify the different stages of AD from MRI

images was conducted by Ravi Chandaran Suganthe, Rukmani Sevalaiappan Latha,

Muthusamy Geetha, and Gobichettipalayam Ramakrishnan Sreekanth. The experiments

were carried out on an ADNI dataset containing 1000 sample brain images, which include

the 250 images from CN subjects, 250 images from EMCI subjects, 250 images from LMCI

subjects, and 250 images from AD subjects. The study shows that more than 90% of

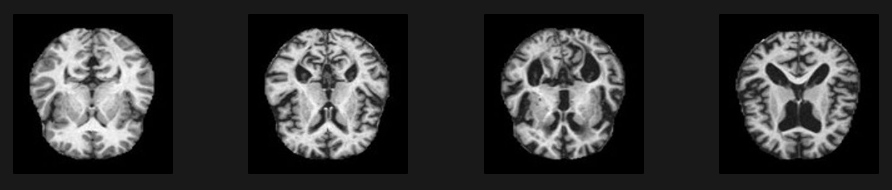
accuracy was achieved in both the models, and VCNN produces higher accuracy.

Our research will compare the accuracy rates of these CNN architectures for the early prediction of Alzheimer’s Disease. These CNN architectures (ResNet50, ResNet152, DenseNet, and EfficientNetB7) performances will be compared using benchmark datasets and clinical data to find the most effective methods of diagnosing Alzheimer’s Disease. The long-term goal, then, is to provide a reliable and efficient tool for the early detection of Alzheimer’s Disease, which enables the implementation of timely interventions and individualized therapeutic strategies. This research adds to the advancement of machine learning-based diagnosis of Alzheimer’s Disease and improves patient outcomes in the clinical realm.

**Methodology**

*Dataset*

The dataset used in this research study was obtained from Kaggle, an online community of data scientists and machine learning practitioners under Google LLC. Consisting a total of 6400 preprocessed magnetic resonance imaging (MRI) images, the dataset has four classes of images: Mild Demented (896 images), Moderate Demented (64 images), Non Demented (3200 images), and Very Mild Demented (2240 images). The MRI images visualize the axial planes of brain anatomy. These images were collected from several websites, hospitals, and public repositories, including ADNI (Alzheimer’s Disease Neuroimaging Initiative), a non-profit organization launched in 2003 by the National Institute of Biomedical Imaging and Bioengineering. Examples of MRI dataset can be seen as below.



*Picture 1.1 Non-Demented MRI, Very Mild Demented MRI, Mild Demented MRI, and Moderate Demented MRI*

*Method*

The main purpose of this research is to compare the performance of various CNN-based architectures to predict Alzheimer’s Disease from a MRI dataset of normal and diseased subject. Pre-trained CNN deep learning models ResNet50, ResNet152, DenseNet, and EfficientNetB7 were utilitized as automated techniques for extracting features from MRI images to diagnose Alzheimer’s Disease. The performance of these CNN models were evaluated using metric evaluation.

*2.1 Data Collection*

The dataset used in this research study was obtained from Kaggle, and online community of data scientists and machine learning practitioners under Google LLC. Collected from several websites, hospitals, and public, repositories, the dataset consists of preprocessed magnetic resonance imaging (MRI) images. Consisting a total of 6400 MRI images, the dataset has four classes of images: Mild Demented (896 images), Moderate Demented (64 images), Non Demented (3200 images), and Very Mild Demented (2240 images). 80% of the dataset were used as training dataset, and the remaining 20% images were used as testing dataset.

*2.2 Implementation of Machine Learning Method*

After collecting data, machine learning methods are implemented. Data pre-processing phase of the MRI is not executed because the dataset contains pre-processed MRI images. The data pre-processing phase objective is to transform the data into a more optimal representation to match the pre-trained CNN’s input size requirements. In this case, the skull from the MRI images was removed to extract the brain and noise was eliminated to improve the model performance. Then, smoothing techniques were used to reduce noise within an image and produce a less pixelated image.

The method implemented for image recognition and processing is Convolutional Neural Network (CNN). CNN methodology contains several types of layers that are commonly used to extract features and make predictions, such as Input Layer, Convolutional Layer, Activation Layer, Pooling Layer, Batch Normalization Layer, Dropout Layer, and Fully Connected Layer. Input layer is used to receive the input data, which in this case the inputted data is the pixel values of image inputted.

Convolutional layer is the essential part and the core building block of deep learning CNN. Its task is to perform feature extraction process, resulting output sets of 2D matrices called feature maps. Each convolutional layer consists of a fixed number of filters that act as feature detectors and extract the features by convolving inputted image with these filters. Each filter acquires the ability to detect the analyzed image’s low level features such as colors, edges, blobs, and corners during training process.

Activation layer introduce non-linearity to the network, which allows it to learn complex patterns and relationships in the data. Pooling layer has sub-sampling layer, responsible for decreasing the size of the feature maps that produced by the convolutional layers. Max pooling is the most used pooling operation to reduce the feature maps by reducing the small region in the image with the maximum value in the region. Max pooling process is performed to avoid overfitting by providing an abstract of the image representation regions and minimizes the computational cost by decreasing the number of parameters. Overfitting is an undesirable machine learning behavior that occurs when the machine learning model gives accurate predictions for training data but now for new data.

Batch normalization layer is used to normalize the convolution layer’s output, features maps, by setting the batch’s average to 0 and the variance to 1. This layer speeds up the training process, using higher learning rates. It prevents the gradients of the model from vanishing during backpropagation. Deep learning models with batch normalization layers are more robust against improper weights initialization.

Dropout layer is used to avoid overfitting phenomena. During the training, the mechanism of this technique is removing neurons randomly. The number of removed neurons are controlled by dropout rate parameter (decides the likelihood of neuron removal). The neurons are removed only during the training process.

Fully connected layer is the last layer of convolutional neural network and acts as a classifier. Its function is to connect the layers in the network and give the final result of the classification. Usually, it is followed by the final layer with a normalized exponential function (Softmax).

In this study, CNN architecture used for the pre-trained CNN model are ResNet50, ResNet152, DenseNet, and EfficientNetB7. These CNN architecture models share common layers such as convolutional, pooling, normalization, and fully connected layers. Regardless having common layers, these CNN architecture models have distinct differences such as the type of blocks used and the use of bottleneck and transition layers.

ResNet50 and ResNet152 CNN models utilize residual blocks to mitigate the vanishing gradient problem, allowing the training of very deep networks. DenseNet model utilizes dense blocks, where each layer receives input from all preceding layer. This dense connectivity patterns allows feature reuse and propagation. EfficientNetB7 model utilizes efficient building blocks inspired by MobileNetV2 architecture, a lightweight CNN model introduced in 2019 by Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen.

ResNet152 and EfficientNetB7 models incorporate bottleneck layers within deeper layers in order to improve computational efficiency. DenseNet model also incorporates bottleneck layers within dense blocks to reduce computational complexity. However, ResNet50 model does not incorporate bottleneck layers.

DenseNet incorporates transition layers to control the growth of feature map dimensions and improve model efficiency. However, transition layers are not incorporated in ResNet50 and ResNet152 architectures. EfficientNetB7 incorporates compound scaling to simultaneously scale up the network depth, width, and resolution, achieving improved performance without significantly increasing computational cost.

*2.3 MRI Image Classification*

Classification process can be done by fully connected layer which function is to produce the final classification scores. Classification processes need activation function to transform the raw outputs of the neural network into a vector of probabilities. In this study, all four CNN-based models architecture have built-in activation layer called Softmax Activation layer to obtain the probability distribution over the classes in case of multi-class classification, which in this experiment the dataset will be classified into four classes: Non-Demented, Mild Demented, Very Mild Demented, and Moderate Demented. Another layer called Global Average Pooling also took part in the classification process to reduce the dimensionality of the feature maps and number of parameters.

*2.4 Performance Evaluation*

Performance indicators are used to evaluate ResNet50, ResNet152, DenseNet and EfficientNetB7 models performance. Performance indicators used in this study are accuracy, precision, recall, and F1-score. True positives (TP) refer to the classifier’s positive tuples that were correctly labelled. False positives (FP) are the negative tuples that were incorrectly labelled as positive. True negative (TN) are the negative tuples that the classifier correctly labelled. False negatives (FN) are the positive tuples that were mislabelled as negative.

Accuracy (ACC) is the percentage of the number of records classified correctly versus the total records. The equation is shown as below.

***ACC = (TP + TN) / (TP + TN + FP + FN)***

Precision (PRE) measures the accuracy of positive predictions. The equation is shown as below.

***PRE = TP / (TP + FP)***

Recall (REC) measure the ability of the classifier to find all the positive samples. The equation is shown as below.

***REC = TP / (TP + FN)***

F1-Score (F1) is the harmonic mean of precision and recall. It provides a balance between precision and recall. The equation is shown as below.

***F1 = 2 \* PRE \* REC / (PRE + REC)***

Recall (REC) measure the ability of the classifier to find all the positive samples. The equation is shown as below.

***REC = TP / (TP + FN)***

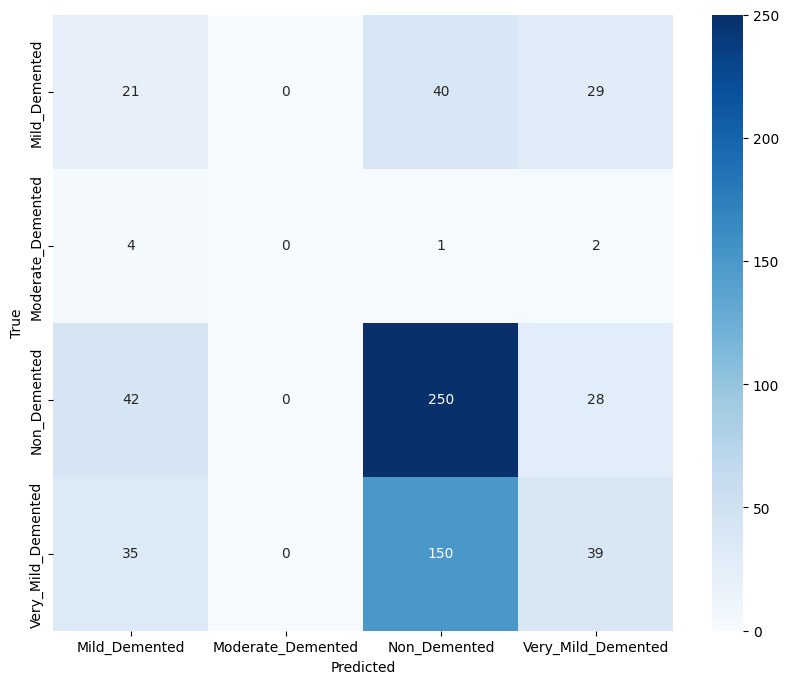
**Result and Discussion**

For this experiment we are using Google Colab environment because it’s ease to use, practicality and free to use. Google Colab provides us the environment for running computational tasks by using CPU or GPU computing units. For this experiment, we are utilising the T4 GPU which is designed for computing tasks particularly tasks that are involving machine learning and deep learning. The T4 GPU supports Tensor Cores which is used in most of Graphical Processing Units that can accelerate the training and neural network processing making it a very good choice for low cost oriented developers. Google Colab is integrated with Jupyter Notebooks making it easier for developers to share and collaborate in projects. Google Colab is advantageous for student developers as it provides very good computational power required for training complex models without having to use the same amount of computing power from our own hardware. Having this much power on computing, we use libraries such as Keras and TensorFlow to enhance the efficiency of developing and deploying deep learning models. Keras is a neural networks API to help building and training neural networks. TensorFLow is an open source platform for machine learning that provides so much functionality of libraries such as layers, models etc..

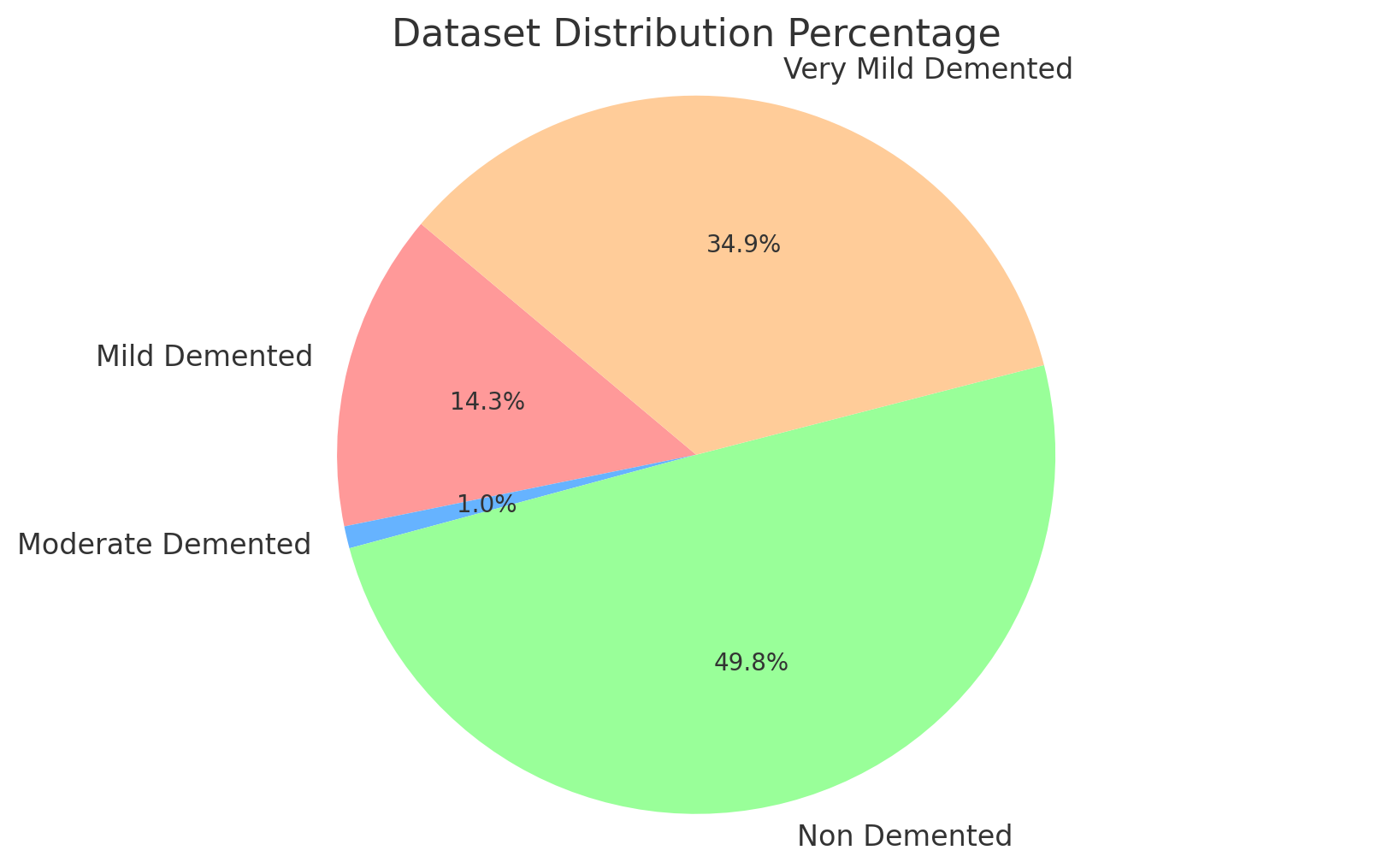
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Models** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| 1 | ResNet50 | 35.25% | 58.75% | 35.25% | 38.50% |
| 2 | ResNet152 | 61.50% | 65.25% | 61.50% | 60.00% |
| 3 | DenseNet | 63.75% | 76.50% | 63.75% | 62.75% |
| 4 | EfficientNetB7 | 64.25% | 78.00% | 64.25% | 64.00% |

The table above highlights the performance of four neural network models on an image classification task. The metrics are accuracy, precision, recall and F1-Score. ResNet50 shows the lowest performance with accuracy of 35.25%, precision of 58.75%, and F1-Score of 38.50%. This shows that the ResNet50 model struggles to accomplish better score on each metrics. The more advanced model like ResNet152, DenseNet and EfficientNetB7 shows the better scores and performance compared to the ResNet50. For example, the EfficienNetB7 model achieves the highest score among three other models with precision of 78.00% and F1-Score of 64.00%

The incremental improvements from ResNet50 to DenseNet, ResNet152 and EfficientNetB7 shows that network design and depth of layers contribute to make better feature extraction and classification capabilities. These results making us understand the importance of using advanced architectures for complex classification tasks is to achieve higher accuracy.



The confusion matrix above shows that the ResNet50 model has major difficulties in classifying between classes, especially for the underpercentaged data like Moderate Demented images. The high number of missclassifications highlights the impacts of the unbalanced dataset.



We can calculate the percentage distribution of each class within the dataset. The Mild Demented class has approximately 14.36% of the dataset (896 out of 6400 images), the Moderate Demented Class represents about only 1% of the dataset (64 out of 6400 images), the Non Demented Class makes up 50% (3200 out of 6400) and the Very Mild Demented class has 35% of the dataset (2240 out of 6400 images)

This distribution highlights an imbalance within the dataset. With the Non Demented and Very Mild Demented classes being disproportionally larger than the Mild Demented and Moderate Demented classes. This can influence the training process that leads to model to perform major on the more prevalent classes but poorly on the underrepresented ones.

The current performance of the models, particularly ResNet50, indicates that there is still significant room for improvement in accurately classifying the different classes of dementia in this imbalanced dataset. While advanced architectures like EfficientNetB7 and DenseNet show promise with higher accuracy and F1-scores, the overall results suggest that the machine learning approach is not yet optimal for this specific task. The novel aspects of these models, such as EfficientNetB7’s compound scaling and DenseNet’s dense connectivity, provide a strong foundation, but their potential is hindered by the dataset’s imbalance. To enhance model performance, it is crucial to address the dataset’s class imbalance through techniques like data augmentation, oversampling of minority classes, or implementing class-weighted loss functions during training. Additionally, exploring ensemble methods or integrating domain-specific knowledge into the model design could further improve the classification accuracy and robustness, making the machine learning approach more effective for early and accurate detection of dementia.

**Conclusion**

To conclude this research on diagnosing Alzheimer's Disease using machine learning, the pre-trained convolutional neural network (CNN) architectures ResNet50, ResNet152, DenseNet, and EfficientNetB7 were applied for AD diagnosing system. This research was conducted on the dataset obtained from Kaggle. The highest accuracy was achieved by EfficientNetB7 architecture model with the average accuracy of 64.25%, average precision of 78.00 %, average recall of 64.25%, and average F1-score of 64.00%. The lowest accuracy was achieved by ResNet50 with the average accuracy of 35.25%, average precision of 58.75%, average recall of 35.25%, and average F1-score of 38.50%. Hence, it can be concluded that the EfficientNetB7 CNN architecture model is the most effective model to diagnose Alzheimer's Disease.

It can be concluded that EfficientNetB7 emerged as the strongest performer across all metrics (accuracy, precision, recall, and F1-score). This is likely due to its ability to achieve high accuracy with a relatively low number of parameters, potentially making it more efficient to train and deploy compared to deeper models like ResNet152. ResNet152 achieved better accuracy than ResNet50, demonstrating the benefit of increased depth for this image classification task. However, it did not outperform EfficientNetB7 or DenseNet in terms of precision and F1-score. DenseNet offered a competitive performance with accuracy and F1-score close to EfficientNetB7. However, EfficientNetB7 surpassed DenseNet in precision. ResNet50 lagged behind the other models in most metrics. This could be due to its shallower architecture compared to the other models.

**Reference**

1. X. Tang and J. Liu, “Comparing different algorithms for the course of Alzheimer’s disease using machine learning," *Annals of Palliative Medicine*, vol. 10, no. 9, pp. 9715–9724, 2021, doi: 10.21037/apm-21-2013.
2. P. C. M. Raees and V. Thomas, "Automated detection of Alzheimer’s Disease using Deep Learning in MRI," *J. Phys. Conf. Ser*., vol. 1921, no. 1, p. 012024, 2021, doi: 10.1088/1742-6596/1921/1/012024.
3. D. AlSaeed and S. F. Omar, "Brain MRI analysis for Alzheimer’s disease diagnosis using CNN-Based feature extraction and machine learning," *Sensors*, vol. 22, no. 8, p. 2911, 2022, doi: 10.3390/s22082911.
4. L. Nanni et al., "Comparison of transfer learning and conventional machine learning applied to structural brain MRI for the early diagnosis and prognosis of Alzheimer’s disease," *Front. Neurol.*, vol. 11, 2020, doi: 10.3389/fneur.2020.576194.
5. G. Battineni et al., "A comprehensive Machine-Learning model applied to magnetic resonance imaging (MRI) to predict Alzheimer’s disease (AD) in older subjects," *J. Clin. Med.*, vol. 9, no. 7, p. 2146, 2020, doi: 10.3390/jcm9072146.
6. G. Battineni et al., "Improved Alzheimer’s disease detection by MRI using multimodal machine learning algorithms," *Diagnostics*, vol. 11, no. 11, p. 2103, 2021, doi: 10.3390/diagnostics11112103.
7. M. Kamal et al., "Machine learning and image processing enabled evolutionary framework for brain MRI analysis for Alzheimer’s disease detection," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–8, 2022, doi: 10.1155/2022/5261942.
8. "Extracting Explainable Assessments of Alzheimer’s disease via Machine Learning on brain MRI imaging data," *IEEE Conference Publication | IEEE Xplore*, Oct. 2020, doi: 10.1109/ICNC48906.2020.9288036.
9. R. Alroobaea et al., "Alzheimer’s disease early detection using machine learning techniques," *Res. Squar*, 2021, doi: 10.21203/rs.3.rs-624520/v1.
10. O. N. Dara et al., "Alzheimer’s Disease diagnosis Using Machine Learning: a survey," *Appl. Sci.*, vol. 13, no. 14, p. 8298, 2023, doi: 10.3390/app13148298.
11. A. D. Arya et al., "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer’s disease," *Brain Informatics*, vol. 10, no. 1, 2023, doi: 10.1186/s40708-023-00195-7.
12. Z. Zhang et al., "Conventional machine learning and deep learning in Alzheimer’s disease diagnosis using neuroimaging: A review," F*ront. Comput. Neurosci.*, vol. 17, 2023, doi: 10.3389/fncom.2023.1038636.
13. M. Sudharsan and G. Thailambal, "Alzheimer’s disease prediction using machine learning techniques and principal component analysis (PCA)," *Mater. Today Proc.*, vol. 81, pp. 182–190, 2021, doi: 10.1016/j.matpr.2021.03.061.
14. C. Kavitha et al., "Early-Stage Alzheimer’s disease prediction using machine learning models," *Front. Public Health*, vol. 10, 2022, doi: 10.3389/fpubh.2022.853294.
15. A. Khan and S. Zubair, "An Improved Multi-Modal based Machine Learning Approach for the Prognosis of Alzheimer’s disease," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2688–2706, 2022, doi: 10.1016/j.jksuci.2020.04.004.
16. V. Diogo et al., "Early diagnosis of Alzheimer’s disease using machine learning: a multi-diagnostic, generalizable approach," *Alzheimer’s Res. Ther.*, vol. 14, no. 1, 2022, doi: 10.1186/s13195-022-01047-y.
17. M. B. Antor et al., "A comparative analysis of machine learning algorithms to predict Alzheimer’s disease," *J. Healthcare Eng.*, vol. 2021, 2021, doi: 10.1155/2021/9917919.
18. H. A. Helaly et al., "Deep learning approach for early detection of Alzheimer’s disease," *Cogn. Comput.*, vol. 14, no. 5, pp. 1711–1727, 2021, doi: 10.1007/s12559-021-09946-2.
19. Suganthe, R. C., Latha, R. S., Geetha, M., & Sreekanth, G. R. (2020). Diagnosis of alzheimer’s disease from brain magnetic resonance imaging images using Deep Learning algorithms. *Advances in Electrical and Computer Engineering*, *20*(3), 57–64. https://doi.org/10.4316/aece.2020.03007
20. Lin, Q., Che, C., Hu, H., Zhao, X., & Li, S. (2023). A comprehensive study on early alzheimer’s disease detection through advanced machine learning techniques on MRI data. *Academic Journal of Science and Technology*, *8*(1), 281–285. https://doi.org/10.54097/ajst.v8i1.14334