**Machine Intelligence Project Report**

**Alzheimer Disease Classification**

**BY**

**Mostafa Ahmed 19105301**

**Marwan Khaled 19104544**

**Marawan Yasser 19105432**

**Nour Hany 19104477**

**Rawan Ramadan 19106199**

**Under the supervision of Dr. Ghada Khoriba**

**Problem Statement:**

In 2020, as many as 5.8 million Americans were living with Alzheimer’s disease (AD). The number of people living with the disease doubles every 5 years beyond age 65. There's no simple and reliable test for diagnosing Alzheimer's disease. It may take several appointments and tests over many months before a diagnosis of Alzheimer's disease can be confirmed. A treatment given at an early stage of AD is more effective, and it causes less minor damage than a treatment done at a later stage. Medical technology advancements have made it possible to identify hidden or complex patterns in diagnostic data to detect diseases earlier and improve treatments. Machine learning approaches have been extensively used in attempts to develop algorithms for the diagnosis of AD. Our goal is to divide MRI scans into four classifications: mild dementia, moderate dementia, non-dementia, and very mild dementia using different machine-learning approaches.

**Related Work:**

The exploration of network architectures has been a part of neural network research since their initial discovery. The recent resurgence in the popularity of neural networks has also revived this research domain. The increasing number of layers in modern networks amplifies the differences between architectures and motivates the exploration of different connectivity patterns and the revisiting of old research ideas. Before discussing our results, we will review some of the related work architectures.

A. DenseNet

DenseNet has made a breakthrough in medical image analysis tasks; Li and Liu [1] used DenseNet to learn local block features of MRI brain image clusters and achieved a better Alzheimer’s disease classification accuracy of 89.5%; Zhang et al. [2] used 3D densely connected convolutional neural network (CAM-CNN) to extract brain MRI multilevel features for classification of Alzheimer’s disease and mild cognitive impairment, densely connected difference at different unit levels; 3D dense unit introduces attention mechanism to generate attention maps and sum transformed MRI hierarchical data into more compact high-level; model has high classification prediction accuracy, and classification performance is at the highest level.

B. MobileNet

C. ResNet

ResNet-50 was used to classify CDR based solely on MRI imagery data (Fulton et. Al, 2019). ResNet-50 predicted the clinical dementia rating (CDR) presence and severity from MRI’s (multi-class classification) using image generation techniques based on an 80% training set resulted in 98.99% three class prediction accuracy on 4139 images (20% validation set) at Epoch 133 and nearly perfect multi-class prediction accuracy on the training set (99.34%). Another study by (Waleed Al Shehri, 2022) stated that the accuracy values for ResNet-50 were 0.8870 and 0.8192.

D. VGG16

Sahrma et al. early diagnosis of AD with two MRI datasets by a VGG16 feature extractor which results in the outcome in the form of accuracy, precision, recall, AUC and F1-score as (90.4%, 0.905, 0.904, 0.969, and 0.904), and (71.1%, 0.71, 0.711, 0.85, and 0.71) for dataset 1 and dataset 2, respectively [5].

Jain et al. used VGG-16 trained on the ImageNet dataset the accuracy of the 3-way classification (AD, CN, and MCI) using the described method is 95.73% for the validation set [6].

E. VGG19

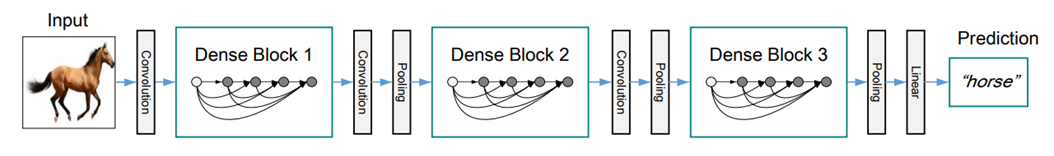
Sasi et al. used VGG19 architecture which provides an AUC of 0.97 and an accuracy of 92% [7]. Helaly et al. used a deep learning approach, specifically VGG19 is used in this work and multi-classified four stages of the AD spectrum. The VGG19 pre-trained model is fine-tuned and achieved an accuracy of 97% for multi-class AD stage classifications [8]. Bhatele et al. paper presents an automated system for the classification of Alzheimer's and Parkinson’s, with the aid of deep transfer learning model VGG19 that outperforms the other three popular deep transfer learning models and delivers an average accuracy of 90% with 70/30 (training/validation split) and 93% with 80/20 (training/validation split) for the multiclass classification as Alzheimer disease, Healthy Control ADNI, Healthy Control PPMI, and Parkinson disease.

**Model Architecture:**

A. DenseNet

It introduces direct connections between any two layers with the same feature-map size. It is proved that DenseNet scale naturally to hundreds of layers while exhibiting no optimization difficulties.

Consider a single image that is passed through a convolutional network. The network comprises a number of layers, each of which implements a non-linear transformation that can be a composite function of operations such as Batch Normalization (BN), rectified linear units (ReLU), Pooling, or Convolution. Figure 1 shows the architecture.



*Figure 1: A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature-map sizes via convolution and pooling*

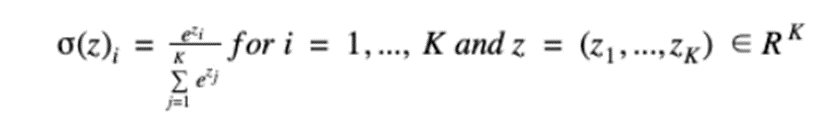
DenseNet tend to yield consistent improvement in accuracy with the growing number of parameters, without any signs of performance degradation or overfitting. Under multiple settings, it achieved state-of-the-art results across several highly competitive datasets. Moreover, DenseNet requires substantially fewer parameters and less computation to achieve state-of-the-art performances. It is believed that further gains in the accuracy of DenseNet may be obtained by more detailed tuning of hyperparameters and learning rate schedules.

B. EfficientNet

EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compoundcoefficient. Unlike conventional practice that arbitrarily scales these factors, the EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients.

D.VGG16

We used the VGG16 model and applied it to the Alzheimer’s disease MRI data. VGG16 is widely known as it was one of the first neural networks to perform well at the Large-Scale Visual Recognition Challenge using a very deep network14. The VGG16 is a standard architecture that contains 13 convolutional and 3 fully connected layers, with 3 × 3 kernels for the convolutional layers and 2 × 2 parameters for the pooling layers. The two convolutional layers in block 1 each use 16 kernels for feature extraction, with image size subsequently reduced in the pooling layer.



*Figure 3*

This formula in figure 3 computes the conditional probability of each class given the input (image) to the model. There is a conditional probability calculated for each of the classifications. The model predicts the image classification by picking the class with the highest conditional probability. Figure 2 shows the architecture.

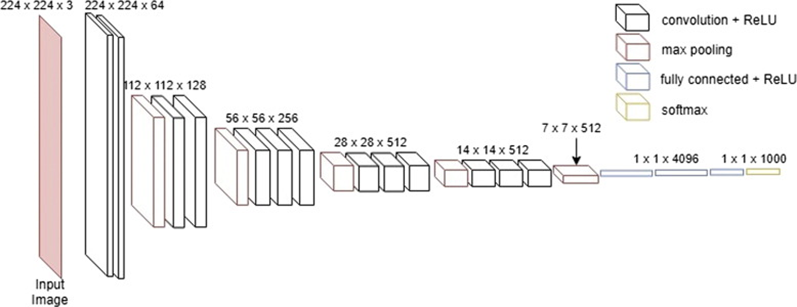
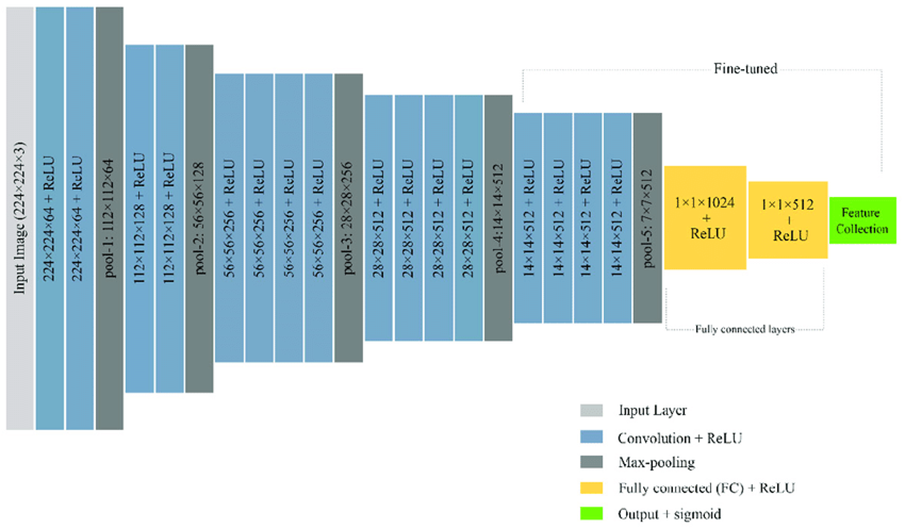


Figure 2: VGG16 neural network architecture

VGG 19

The VGG is the abbreviation for Visual Geometry Group Net was used in CNN that has approximately 143 million parameters, these parameters are learned using the ImageNet dataset comprising of 1. 2 million images which contain thousands of classes for training. The VGG-19 Neural Network which is shown in figure 4 consists of 19 layers of deep neural network and has more weight. [10]. VGG CNN has six main structures, each of which is mainly composed of multiple connected convolutional layers and full-connected layers. The size of the convolutional kernel is 3\*3, and the input size is 224\*224\*3. The number of layers is generally concentrated at 16~19. It uses an alternating structure of multiple convolutional layers and non-linear activation layers, which is better than a single convolution The layer structure can better extract image features, use Maxpooling for downsampling, and modify the linear unit (ReLU) as the activation function, that is, select the largest value in the image area as the pooled value of the area. The downsampling layer is mainly used to improve the anti-distortion ability of the network to the image while retaining the main features of the sample and reducing the number of parameters [11].

****

ResNet50:

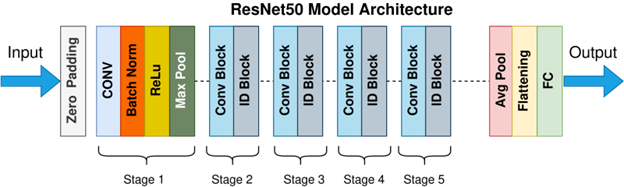
ResNet stands for Residual Network and is a specific type of convolutional neural network (CNN) introduced in the 2015 paper “Deep Residual Learning for Image Recognition”. ResNet-50 is a 50-layer convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer). Residual neural networks are a type of artificial neural network (ANN) that forms networks by stacking residual blocks. The original ResNet architecture was ResNet-34, which comprised 34 weighted layers. It provided a novel way to add more convolutional layers to a CNN, without running into the vanishing gradient problem, using the concept of shortcut connections. The regular network was based on the VGG neural networks (VGG-16 and VGG-19)—each convolutional network had a 3×3 filter. However, a ResNet has fewer filters and is less complex than a VGGNet. A 34-layer ResNet can achieve a performance of 3.6 billion FLOPs, and a smaller 18-layer ResNet can achieve 1.8 billion FLOPs, which is significantly faster than a VGG-19 Network with 19.6 billion FLOPs (read more in the ResNet paper, He et, al, 2015).

The 50-layer ResNet architecture includes the following elements, as shown in the table below:

* **A 7×7 kernel convolution** alongside 64 other kernels with a 2-sized stride.
* **A max pooling layer** with a 2-sized stride.
* **9 more layers**—3×3,64 kernel convolution, another with 1×1,64 kernels, and a third with 1×1,256 kernels. These 3 layers are repeated 3 times.
* **12 more layers** with 1×1,128 kernels, 3×3,128 kernels, and 1×1,512 kernels, iterated 4 times.
* **18 more layers** with 1×1,256 cores, and 2 cores 3×3,256 and 1×1,1024, iterated 6 times.
* **9 more layers** with 1×1,512 cores, 3×3,512 cores, and 1×1,2048 cores iterated 3 times.

(up to this point the network has 50 layers)

* **Average pooling**, followed by a fully connected layer with 1000 nodes, using the softmax activation function.



MobileNet

Another CNN pre-trained model used in the suggested work is MobileNetV2. It is based on an enduring structure that has been turned upside down, with bottleneck layers connected by lingering connections. The intermediate expansion layer filters have a source of non-linearity that uses simple depth-wise convolutions. It comprises two layers and three blocks Fig. 2 depicts the architecture of MobileNetV2.

MobileNetV2 is founded on a residual structure that has been reversed, with the shortcut connections occurring between the thin bottleneck layers. Lightweight depthwise convolutions are used in the intermediate expansion layer as a source of non-linearity to filter features. We also discover that to retain representational power, non-linearities in the thin layers must be eliminated. We show how this enhances performance and share the insight that inspired this design. The inverted residual block obtains features with 3 3 convolution, then expands the input channels with 1 1 convolution to gain more features, and finally compresses the number of channels with 1 1 pointwise convolution. [12] Expansion-convolution-compression is used throughout, which is more efficient and parametric-free than using a 3 3 convolutional network directly.

The linear bottleneck layer fixes the issue that the ReLU6 loses low-latitude information after the inverse residual block by swapping out the ReLU6 activation function in the penultimate layer with a linear function and removing it between the last high dimension and the low dimension.

The linear bottleneck layer and the inverted residual block combine to generate the bottleneck residual block, which is the fundamental component of MobileNetV2. The shortcut is used when s is 1; it is not applied when s is 2. According to Fig. 4,

**Transformers**

Transformer architectures based on self-attention were first presented in [13], where they demonstrated enhanced machine translation performance. Since then, they have been successfully used for a variety of NLP tasks. Importantly, it has been demonstrated that when these models are paired with pre-training, they obtain almost human performance on a variety of NLP tasks (14; 15). Transformer models receive their input as a series of vectors, often input token embeddings, which are then processed by a stack of Transformer blocks. Each block is made up of two layers: a token wise feed-forward (MLP) layer and a multi-head self-attention layer that uses dot-product attention to aggregate information across tokens. Both employ residual connections and layer normalization.

**Vision Transformers**

The Transformer architecture mentioned previously is also used by ViT [16]. The image pre-processing layer is where the main distinction is found. The image is divided into a series of non-overlapping patches in this layer, which is followed by a learnt linear projection. For instance, breaking a 384x384 image into 16x16 patches would yield a series length of 162. The number of filters used in the 2D convolution used to achieve this defines the concealed size of the sequence input to the Transformer. In addition, ViT adds a unique CLS token to the input after [14], whose representation is used for categorization.

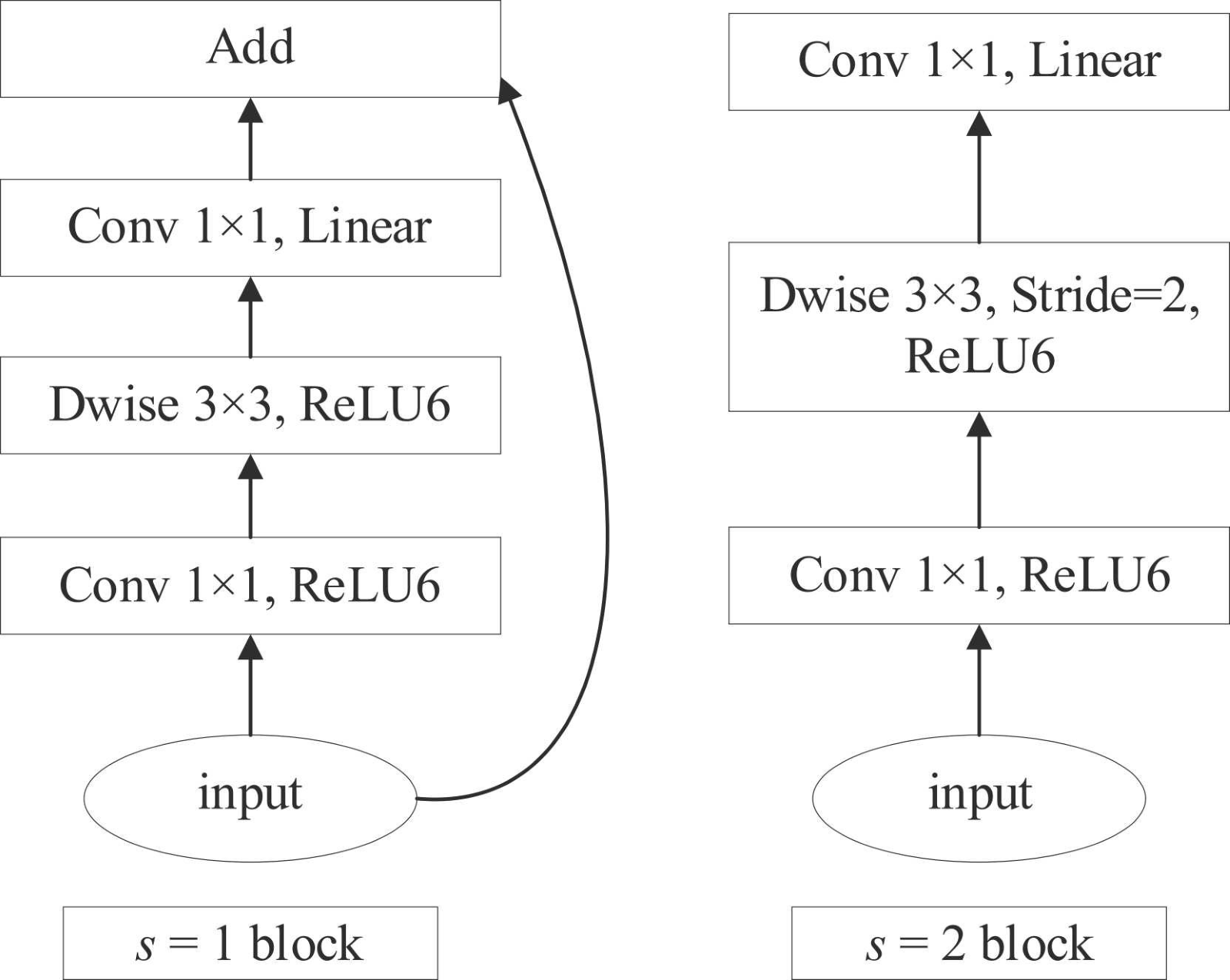


fig4

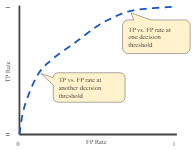
Our dataset is not balanced, so we cannot use accuracy as our metric. For this tutorial, we will be using ROC AUC. Intuitively, ROC AUC gives a score, with higher scores closer to 1 indicating that the different classes can be distinguishable for the model. A lower score closer indicates that the the model cannot distinguish between different classes. A score of 0.5 indicates that the ordering the images is pretty much random. Learn more about ROC AUC.

## **ROC curve**

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

* True Positive Rate
* False Positive Rate

True Positive Rate (TPR) is a synonym for recall and is therefore defined as follows:

****

**Results:**

Ending up with two different approaches: one deep neural network, and another transformer model. The researchers concluded the study with results that showed high accuracy of Classification of AD.

After applying different approaches to the dataset, The following is concluded:

•Vision transformers (VIT) model showed its capability to increase accuracy with a highly pleasing training accuracy result of nearly 82.9% on the first dataset that we trained on

•And 94.8% in the second dataset that we tested on.

The following table shows the Testing Results of the different approaches**:**

| Metrics | DenseNet | EfficientNet | MobileNetV2 | VGG16 | VGG19 | VIT | ConvNext | ResNet50 | ResNet101 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Precision (Test Data 1) | 0.6718 | 0.5790 | 0.75 | 0.68 | 0.71 | 0.8416 | 0.7149 | 0.55 | 0.55 |
| Recall (Test Data 1) | 0.6482 | 0.5098 | 0.421 | 0.65 | 0.67 | 0.8614 | 0.85225 | 0.34 | 0.49 |
| F1\_score (Test Data 1) | 0.6596 | 0.5415 | 0.497 | 0.65 | 0.67 | 0.8514 | 0.778 | 0.80 | 0.83 |
| Precision (Test Data 2) | 0.7048 | 0.6226 | 0.5397 | 0.79 | 0.73 | 0.9594 | 0.9093 | 0.54 | 0.60 |
| Recall (Test Data 2) | 0.6902 | 0.5158 | 0.4302 | 0.77 | 0.70 | 0.9303 | 0.9357 | 0.40 | 0.48 |
| F1\_score (Test Data 2) | 0.6974 | 0.5639 | 0.4779 | 0.77 | 0.70 | 0.94463 | 0.9223 | 0.40 | 0.48 |
| Accuracy on first dataset (test dataset) | 0.8933 | 0.8061 | 0.7845 | 0.6740 | 0.6927 | 0.829554 | 0.7638 | 0.47 | 0.54 |
| Accuracy on second dataset | 0.9057 | 0.8008 | 0.7658 | 0.7850 | 0.7175 | 0.948438 | 0.916250 | 0.49 | 0.56 |
|  |  |  |  |  |  |  |  |  |  |

**GitHub Link:**

<https://github.com/1NourHany/Alzheimer-s-Disease-Detection->

**Data & References:**

[1] Alzheimer's Dataset (4 class of Images)<https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>

[2] Alzheimer Features<https://www.kaggle.com/datasets/brsdincer/alzheimer-features>

[3] F. Li, M. Liu, and Alzheimer's Disease Neuroimaging Initiative, “Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks,” Computerized Medical Imaging and Graphics, vol. 70, pp. 101–110, 2018.<https://doi.org/10.1016/j.compmedimag.2018.09.009>

[4] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, and X. Long, “A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification,” Magnetic Resonance Imaging, vol. 78, pp. 119–126, 2021<https://doi.org/10.1016/j.mri.2021.02.001>

[5] Shagun Sharma, Kalpna Guleria, Sunita Tiwari, Sushil Kumar, A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans, Measurement: Sensors, Volume 24,2022,100506, ISSN 2665-9174,https://doi.org/10.1016/j.measen.2022.100506.

[6] Jain, R., Jain, N., Aggarwal, A., &amp; Hemanth, D. J. (2019). Convolutional neural network based alzheimer’s disease classification from Magnetic Resonance Brain Images. Cognitive Systems Research, 57, 147–159. https://doi.org/10.1016/j.cogsys.2018.12.015

[7] Sasi, Dr. Smitha and Dayanada Sagar. “Detection Of Alzheimer’s Using Neural Network Models – CNN And VGG19.” (2021).

[8] Helaly, H.A., Badawy, M. & Haikal, A.Y. Deep Learning Approach for Early Detection of Alzheimer’s Disease. Cogn Comput 14, 1711–1727 (2022).<https://doi.org/10.1007/s12559-021-09946-2>

[9] Bhatele, Kirti Raj. “Classification of Neurodegenerative Diseases Based on VGG 19 Deep Transfer Learning Architecture: A Deep Learning Approach.” Bioscience Biotechnology Research Communications 13, no. 4 (2020): 1972–80.<https://doi.org/10.21786/bbrc/13.4/51>.

[10] Sudha, V., and Dr. T. R. Ganeshbabu. “A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning.” Computers, Materials &amp; Continua 66, no. 1 (2020): 827–42.<https://doi.org/10.32604/cmc.2020.012008>.

[11] Sudha, V., and Dr. T. R. Ganeshbabu. “A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning.” Computers, Materials &amp; Continua 66, no. 1 (2020): 827–42. <https://doi.org/10.32604/cmc.2020.012008>.

[12] Zhang, L., Wang, J., Li, B., Liu, Y., Zhang, H., & Duan, Q. (2022). A mobilenetv2-senet-based method for identifying fish school feeding behavior. *Aquacultural Engineering*, *99*, 102288. https://doi.org/10.1016/j.aquaeng.2022.102288

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, èukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 6000±6010, 2017. 1, 2

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171±4186, 2019. 1, 2

[15] Yihe Dong, Jean-Baptiste Cordonnier, and Andreas Loukas. Attention is not all you need: Pure attention loses rank doubly exponentially with depth, 2021.

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020. 1, 2, 3, 8, i

**Team Contribution**

1. DenseNet & EfficientNet Models (Mostafa)

2. Transformers (Marwan Khaled)

3. ResNet50, ResNet101,ResNext (Marawan Yasser)

4. VGG16 and VGG19 (Nour Hany)

5. MobileNetv2 (Rawan Ramadan)

6. Report & Research Paper (All members)