T.C.

ISTANBUL AYDIN UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

U-NET BACKBONE ARCHITECTURES FOR 3D SEGMENTATION OF HIPPOCAMPUS VOLUME IN BRAIN MRI

MASTER'S THESIS

Fella Achouri

Department of Software Engineering,

Artificial Intelligence and Data Science Program

June, 2022

T.C.
ISTANBUL AYDIN UNIVERSITY
INSTITUTE OF GRADUATE STUDIES

U-NET BACKBONE ARCHITECTURES FOR 3D SEGMENTATION OF HIPPOCAMPUS VOLUME IN BRAIN MRI

MASTER'S THESIS

Fella Achouri

Department of Software Engineering,
Artificial Intelligence and Data Science Program

Thesis Advisor: Prof. Dr. Ali Okatan .

June, 2022

ONAY FORMU

DECLARATION

I hereby declare with respect that the study "U-Net Backbone Architectures for 3D Segmentation of Hippocampus volume in Brain MRI", which I submitted as a Master thesis, is written without any assistance in violation of scientific ethics and traditions in all the processes from the project phase to the conclusion of the thesis and that the works I have benefited are from those shown in the Bibliography. (01/06/2022)

Fella Achouri

FOREWORD

Foremost, I want to express my gratitude to my advisor, Professor Dr. Ali Okatan. Without his constant encouragement, valuable insights, motivation, and guidance, I would not have been able to complete this project. A special thanks to IAU Graduate school staff for all their support and encouragement during the execution of this work. We could not complete this project without everybody encouragement and support. Even though we could not put all names, we would especially like to thank all IAU teaching staff for the valuable information and feedback received during this work. Last but not least, we would like to thank our parents and friends for their support and encouragements.

June, 2022 Fella Achouri

U-NET BACKBONE ARCHITECTURES FOR 3D SEGMENTATION OF HIPPOCAMPUS VOLUME IN BRAIN MRI

ABSTRACT

The volume of the hippocampus is important in the progression of Alzheimer's disease. To determine Hippocampus volume from 3D Brain MRI data, semantic segmentation approaches have been frequently employed. This research compares the usage of different models as backbones with U-Net architecture for semantic segmentation of the Hippocampus from 3D Brain MRI images. The neural network designs ResNet, SE-ResNet, ResNext, SE-ResNext, and DenseNet were employed as the backbone for the U-Net model. For semantic segmentation of the Hippocampus, the Decathlon 3D Brain MRI images dataset was employed. The data collection contains 260 3D brain MRI pictures, the majority of which were created using 35 MRI slices. An exploratory data analysis was performed, which yielded helpful insights from the dataset; the pixel intensity histogram revealed that the majority of pixels in picture slices are bright, making the segmentation process easier for the model. Images were preprocessed and separated into training and testing sets before using the semantic segmentation model. Because most MRI pictures had variable width, height, and slices, each slice of MRI image was transformed into 64x64 pixels of width and height. The image's empty region has been assigned a 0-pixel value, indicating that the pixel would be black in color. Data augmentation has been performed on the preprocessed dataset to increase the samples in the dataset and to help model generalize in a better way. As loss functions, a combination of dice loss and categorical focused loss was utilized. With a 0.7 threshold, the IOU and F-score metrics were utilized. The U-Net model was trained utilizing many backbones, each of which has an influence on the model's performance and efficiency. SE-ResNet-50 performed best in terms of IOU and F-score metrics for both training and validation sets. Other than that SE-ResNext-50 architecture also provided some good results.

.

Keywords: Semantic Segmentation, U-Net, Hippocampus, Brain MRI, IOU, F-Score, Data Augumentation.

BEYİN MRI'DEKİ HİPOKAMPUS HACMİNİN 3D BÖLÜTLEMESİ İÇİN U-NET OMURGA MİMARİSİ

ÖZET

Hippocampus hacminin Alzheimer hastalığının ilerlemesinde önemli bir rol oynadığı belirtilmektedir. 3B beyin MRI verilerinden Hippocampus hacmini belirlemek için semantik segmentasyon yaklaşımları sıklıkla kullanılmıştır. Bu araştırma, Hippocampus'un 3B beyin MRI görüntülerinden semantik segmentasyonunda U-Net mimarisi ile farklı modellere dayalı olarak kullanımını karşılaştırır. Sinir ağı tasarımları ResNet, SE-ResNet, ResNext, SE-ResNext ve DenseNet, U-Net modelinin omurgası olarak kullanılmıştır. Hippocampus'un semantik segmentasyonu için Decathlon 3B Beyin MRI görüntüleri veri kümesi kullanılmıştır. Veri koleksiyonu, 260 adet 3B beyin MRI resmi içermektedir ve bunların çoğu 35 MRI kalıbı kullanılarak oluşturulmuştur. Keşifsel veri analizi yapılmış ve veri kümesinden yararlı sonuçlar elde edilmiştir; piksel yoğunluk histogramı, resim kalıplarındaki piksellerin çoğunluğunun aydınlık olduğunu göstermiştir ve bu da model için segmentasyon sürecini daha kolay hale getirir. Semantik segmentasyon modelini kullanmadan önce görüntüler ön işleme tabi tutulmuş ve eğitim ve test kümelerine ayrılmıştır. Çoğu MRI resmi değişken genişlik, yükseklik ve kalıplar içermiştir, bu nedenle her bir MRI resminin kalıbı 64x64 piksel genişlik ve yükseklikte dönüştürülmüştür. Görüntünün boş bölgesi 0 piksel değeri ile atanmıştır, bu da pikselin siyah renkli olacağını gösterir. Ön işlem görmüş veri kümesine veri artırma yapılmış ve modelin daha iyi şekilde genelleşmesini sağlamak için yardımcı olması amaçlanmıştır. Kayıp fonksiyonları için dize kaybı ve kategorik odaklı kayıp kombinasyonu kullanılmıştır. IOU ve F-skor metrikleri 0.7 eşik değeri kullanılarak hesaplandı. U-Net modeli, her birinin performans ve verimlilik üzerinde bir etkisi olan çeşitli omurga tasarımları kullanılarak eğitildi. Eğitim ve doğrulama kümeleri için IOU ve F-skor metrikleri açısından en iyi sonuç SE-ResNet-50 tarafından verildi. Bunun yanı sıra SE-ResNext-50 mimarisi de iyi sonuçlar sunmuştur.

Anahtar kelimeler: Semantik Bölütleme, U-Net, Hipokampus, Beyin MRI, IOU, F-Skor, Veri Artırma.

TABLE OF CONTENT

ABBREVIATIONS

AD : Alzheimer Disease

MRI : Magnetic Resonance Imaging

ResNet : Residual Network

ResNext : Aggregated Residual Network

SE : Sequeeze and Excitation

CNN : Convolutional Neural Network

DL : Deep Learning

LIST OF FIGURES

I. INTRODUCTION

Overview

The hippocampus is a region of the brain related with memory function. The hippocampus has the form and structure of a seahorse and is found in the inner part of the temporal lobe. It is also a component of the limbic system, which is responsible for emotional reactions. The hippocampus is important for long-term memory storage, as well as spatial processing and navigation. Alzheimer's disease (AD) is characterized by a gradual neurodegeneration that begins in the hippocampus and progresses throughout the brain. It is distinguished by deficits in recent memory, executive functioning, spatial and temporal orientation. The patient's cognitive ability and autonomy decrease gradually. At the moment, an MRI radiological examination is one of the most advanced methods for measuring the disease. The measurement of hippocampal volume, in particular, has been useful in identifying and tracking the progression of a number of brain disorders. The hippocampus is an important part of the human brain because it helps to consolidate information from short-term to long-term memory. Humans have two hippocampi, one in each hemisphere of the brain. According to studies, the volume of the hippocampus varies among a population, and a "normal" range may be found when age, gender, and brain hemisphere are taken into account. There is one disadvantage to using MRI scans to calculate the volume of the hippocampus: the procedure is time-consuming since each slice of the 3D volume must be analyzed and the structure's shape tracked. The fact that the hippocampus does not have a regular form adds to the challenge.

Several studies have been conducted on the segmentation of Hippocampus volume using MRI images, the majority of which employed 2D images and the majority of which used state-of-the-art CNN models with U-Net architecture. The 2D picture segmentation for the Hippocampus does not produce very accurate findings since the brain has a complicated structure and machine learning methods often struggle to generalize over those images. We can now train and use considerably more complicated deep learning models thanks to recent technological advancements. In this paper, 3D Brain MRI images were trained using a 3D segmentation model. The model was trained on 3D MRI images using the U-Net architecture. The U-Net Convolutional Neural Network architecture was developed primarily for biomedical picture segmentation. The notion of Encoders and Decoders has impacted the U-net design. Encoders are used to down sample the picture in U-net architecture, whereas decoders are used to up sample the image and extract features from it. Backbone is a U-net architectural element that is used to initialize the encoder layers. And after the encoder layers are started, the decoder will be the inverse of the encoding layers. The U-Net backbone enables the use of different CNN architectures as encoders, which improves the performance and efficiency of U-Net models. Previously, state-of-the-art basic CNNs and complicated network designs like VGG16 were chosen as the backbone for the U-Net model for general segmentation tasks, and their performance was evaluated on the ImageNet dataset.

The use of several complicated neural network designs to segment hippocampal volume from 3D brain MRI images is examined in this paper. ResNet , SE-ResNet, ResNext, SE-ResNext, and DenseNet are the Convolutional Neural Network designs that were compared in the research. For each model, a combination of Dice and Focal loss was utilized as the loss function. The models' outcomes were assessed using the IOU and F-Score assessment criteria. The results demonstrate that SE-ResNet is the better solution for segmenting Hippocampus volume using 3D Brain MRI images, since this network design allowed the U-Net model to train quicker while using fewer resources. In comparison to previous networks, the SE-ResNet design offers higher performance and efficiency. In this paper, I have presented thorough results as well as a brief explanation of the methods used to train the models, and I have concluded the paper with a conclusion.

Challenges

Much research has been done on the segmentation of Hippocampus Volume and combining the volume of Hippocampus with other cognitive functionalities to better predict the progression of Alzheimer disease. There are multiple challenges associated with the segmentation of hippocampus Volume using Brain MRI, some of them are listed below:

1) Lack of available data to train deep learning models for image segmentation.

2) Image segmentation with Deep learning requires a lot of computational resources.

Goals

This research aims to provide a CNN architecture that will be useful for the segmentation of Hippocampus volume using 3D Brain MRI. This research will also be focused to provide resource efficient solution, and the solution that can be integrated in a real time clinical environment. The main objectives of this research are given below:

1. Making use of different CNN architectures combined with 3D U-Net image segmentation model to provide a best solution for Hippocampus volume segmentation.
2. Making use of combined Dice Loss and Focal Loss function in the semantic segmentation model to improve the learning of the model.

Research Questions & Methods

New research studies are mainly based on the problems that have been arise in previous work or to resolve some questions have been raised at later stages. In this study research questions may include

- Which CNN architecture will provide the better 3D segmented Hippocampus volume using Brain MRI?
- Does the use of combined loss functions will impact the learning of the deep learning model?
- Is the final solution being resource efficient, which will make use of less resources but provide with better results?

Dissertation Structure

This dissertation is splitted into 7 chapters, the chapters include:

Chapter 1 – Introduction, in this chapter, the problem and challenges are introduced.

Chapter 2 – Literature Review: a complete and comprehensive description of Alzheimer Disease and Hippocampus volume with all previous related research is discussed in this chapter.

Chapter 3 – 3D Brain MRI Data: this chapter explains the 3D Brain MRI data, the analysis and all the preprocessing that have been done on the data.

Chapter 4 – CNN Architectures: this chapter explains the basics of CNN architecture and it also provides the detailed analysis of modern CNN architectures that will be used in this research.

Chapter 5 – 3D Semantic Segmentation: this chapter explains the basics of semantic segmentation, it will also explain the U-Net segmentation model which has been used in this research.

Chapter 6 – Experimental Tests, Result and Discussions: this chapter focuses on the experimentations that have been conducted, it will also explain the results which have been obtained through experimentations.

Chapter 7 – Conclusion: this chapter provides a brief conclusion on the research that have been taken, the methodology that have been used with results and how it can be used with other cognitive functionalities.

II. LITERATURE REVIEW

This section will describe the previous research on the segmentation of Hippocampus volume using 3D MRI images, but first I will provide a brief description about Alzheimer disease and why hippocampus volume plays an important role in Alzheimer disease.

Alzheimer Disease And Hippocampus Volume

Among old adults Alzheimer Disease is the common cause of loss of cognitive functionalities. Alzheimer is a brain disease that affect memory, thinking ability and cognitive functionality of a human being, The early symptoms of the disease are memory problems and when the disease gets severe the patient is not even able to do daily common tasks, the patients will forget what they are doing and can also ask the same question again and again. The disease also affects the behavioral changes, in some cases patients become angry, violent or depressed. The disease has been named after Dr. Alois Alzheimer. In 1906, while analyzing the brain tissues of a dead woman who died because of mental illness, Dr. Alzheimer finds some weird changes in the tissues of the brain, which includes abnormal clumps and tangled bundles of fibers. These abnormalities in the tissues of the brain are still considered as the features of Alzheimer Disease. Along with those features in some cases the connections between neurons also known as nerve cells have been lost, and by this loss of connections neurons are not able to transmit messages to different parts of the brain. Alzheimer disease initially starts from those parts of the brain which are responsible for memory such as the entorhinal cortex and hippocampus, after the progression of the disease it also damages cerebral cortex which is responsible for other common tasks such as language, social behavior and reasoning.

The hippocampus is a region of the brain related to memory function. The hippocampus has the form and structure of a seahorse and is found in the inner part of the temporal lobe. It is also a component of the limbic system, which is responsible for emotional reactions. The hippocampus is important for long-term memory storage, as well as spatial processing and navigation.

Alzheimer's disease (AD) is characterized by a gradual neurodegeneration that begins in the hippocampus and progresses throughout the brain. At the moment, an MRI radiological examination is one of the most advanced methods for measuring the disease. The measurement of hippocampal volume, in particular, has been useful in identifying and tracking the progression of a number of brain disorders. The hippocampus is an important part of the human brain because it helps to consolidate information from short-term to long-term memory. Humans have two hippocampi, one in each hemisphere of the brain. According to studies, the volume of the hippocampus varies among a population, and a "normal" range may be found when age, gender, and brain hemisphere are taken into account. There are many searches that have been done on the effect of hippocampus volume on Alzheimer Disease (AD).

3D Semantic Segmentation

3D Brain MRI images have been used to analyze the volume of hippocampus and how it is affecting Alzheimer disease predictions. With the advancements in technology allowed us to perform much more complex computational tasks quickly, they opened a new area of research. 3D semantic segmentation is the area where most of the research are going on. Especially in the medical field, With the help of AI we can change the world. 3D semantic segmentation allowed us to segment the volume of the hippocampus from 3D Brain MRI Images.

III. 3D BRAIN MRI DATA

Magnetic Resonance Imaging (MRI) is a test that helps doctors to analyze the condition of the specific part of the body, this test is performed without injecting any instrument through the skin. Without the use of radiation (x-rays), MRI makes use of powerful magnetic field radiofrequency pulses. The MRI machine passes the data to a high-end computer that will generate detailed high-resolution images. MRI is the most sensitive brain imaging test, and the qualities of an MRI image can be degraded even with a slight movement of the patient body. MRI provides detailed high-resolution images of brain, spinal cord and vascular anatomy, by the help of MRI we are able to visualize the brain in all three planes: axial, coronal and sagittal. In the XYZ coordinate system the axial plane refers to the X-Y plane, the coronal plane refers to the X-Z plane and the sagittal plane refers to the Y-Z plane. With the help of all those planes in MRI we can analyze the image in a 3-Dimensional view.

While doing MRI scans the tissues can be characterized into two different relaxation times known as T1 and T2. These relaxation times provide us with the most common MRI sequences T1-weighted and T2-weighted scans. T1-weighted images are produced by using short TE (Time to Echo) and TR (Repetition Time). In T1-weighted images the contrast and brightness of the image are determined by T1 properties of tissues. T2-weighted images are produced by using long TE (Time to Echo) and TR (Repetition Time). In T2-weighted images the contrast and brightness of the image are determined by T2 properties of tissues. In the figure we can visualize T1 weighted and T2 weighted MRI scans, and in the table the values of TE and TR are given according to the T1 and T2 MRI sequence.

In MRI images we have a concept of image slices, in simple terms slices can be referred as the MRI scans, these scans are sent to the high-end computer to generate the final MRI image. Slices play an important role while constructing MRI images If the thickness of the slices is large then it will result in less noise but less detailed image, and if the slice thickness is small then it will result in a noisier image but with greater details. The anatomical planes also play an important role, for example the thick slices can be good if we want to analyze the data from the Axial Plane and the thin slices can be useful if we want to analyze the data from the Coronal Plane it totally depends on the analysis of the plane and the organ which we are working with.

Data Analysis

Medical Segmentation Decathlon provided the 3D Brain MRI pictures of the Hippocampus. It is a biomedical image analysis challenge (MSD) that offers many datasets for various organs of the body, the datasets consisted of MRI and CT scans according to the body organ, all of these datasets which were provided by MSD are under a permissive copyright license, which means we can use it and share it publicly. The collection has 2633 pictures in total, and all of these images were taken during real world clinical trials, the pictures were collected from multiple institutions and labs. After the collection of the Images these were converted to the Neuroimaging Informatics Technology Initiative (NIFTI) format. NIFTI is basically used to store medical images, these images are MRI, CT scans etc. NIFTI files provide us with a variety of functionalities which we can perform after storing the initial MRI data, for example, we can extend the files to include some more data information. NIFTI stores metadata and the data of an MRI image in a file. The 3D MRI image data is stored in a form of single volume with multiple slices and when we read those data files it will provide us a series of slices. To maintain data matrix direction x, y, and z consistency, all pictures were transposed without resampling.

The Hippocampus dataset included 394 3D Brain MRI pictures, 263 of which were annotated labels and were used for training and validation, while the remaining 131 were unannotated and could be utilized for testing. The MRI scans were collected from healthy people as well as those with non-affective psychotic disorders. The imaging sequence for this dataset was T1-weighted MPRAGE. The actual anatomy of the hippocampus and sections of the subiculum, also known as the anterior and posterior of the hippocampus, were the major focus of this dataset. The dataset was built with the complexity of the brain in mind, because the hippocampus is a small region of the brain and the surrounding area of the hippocampus has extremely complex structures, for this reason the area around the hippocampus has been cropped to remove the unwanted complexities of the brain structure and only the hippocampus area has been retained with a small surrounding area with complex structure, with the help of this dataset the hippocampus can be precisely segmented.

Because the photos in the collection are in NIFTI format, they are read using the nibabel library. This library supports the reading and writing of several neuroimaging file formats, including NIFTI. As the first stage in the data analysis part, I used multiple planes to show a single slice of a picture. For the purpose of simplicity, we can think of an MRI image slice as a single scan. The slice selection approach in MRI allows us to obtain 3D MRI pictures. Radiologists split MRI pictures into three planes to better interpret them: Axial Plane, Coronal Plane, and Sagittal Plane. The Axial plane (Figure 1) splits the body into top and bottom halves.

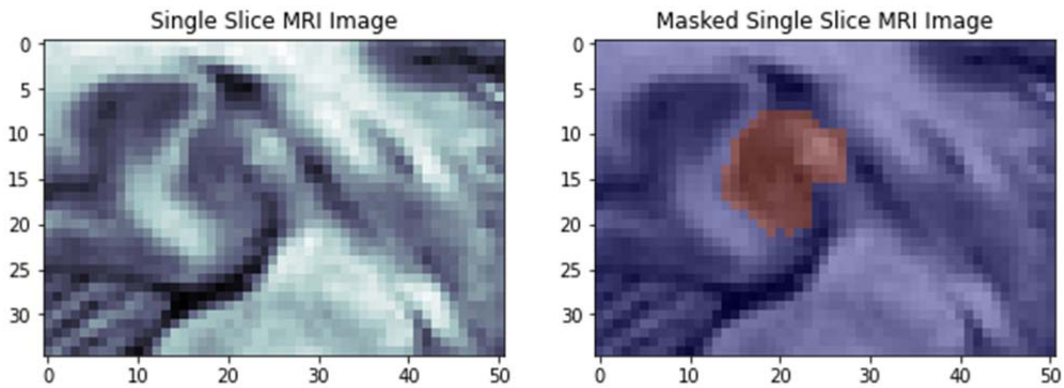

Figure 1 Axial Plane with segmentation mask of hippocampus volume

The Coronal Plane (Figure 2) divides the body into front and back halves and the Sagittal Plane (Figure 3) splits the body into left and right halves.

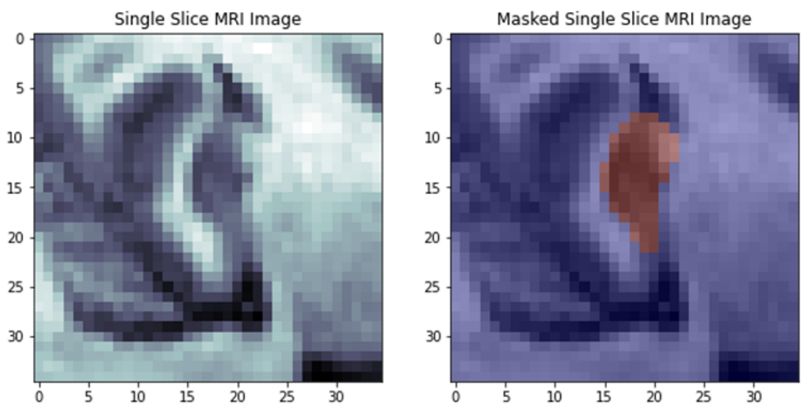

Figure 2 Coronal Plane with segmentation mask of hippocampus volume

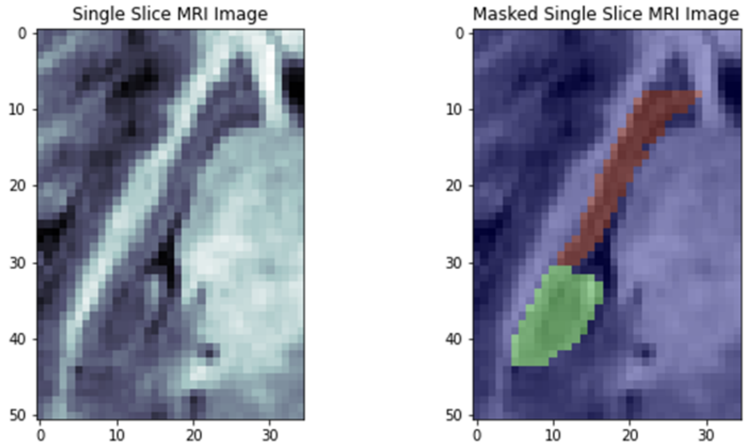

Figure 3 Sagittal Plane with segmentation mask of hippocampus volume.

Each plane provides a distinct aspect of the image, allowing us to grasp the facts more thoroughly. I utilized masked data to emphasize the hippocampus area in the visualizations, but we weren't able to properly depict the hippocampus in a nice way. The Nilearn module in Python enables us to analyze brain MRI pictures. I plotted the MRI picture using the EPI (Echo Planar Imaging) (Figure 4) technique with the aid of this library. EPI is the quickest acquisition method in MRI, allowing us to capture rapid physiological processes in the human body.

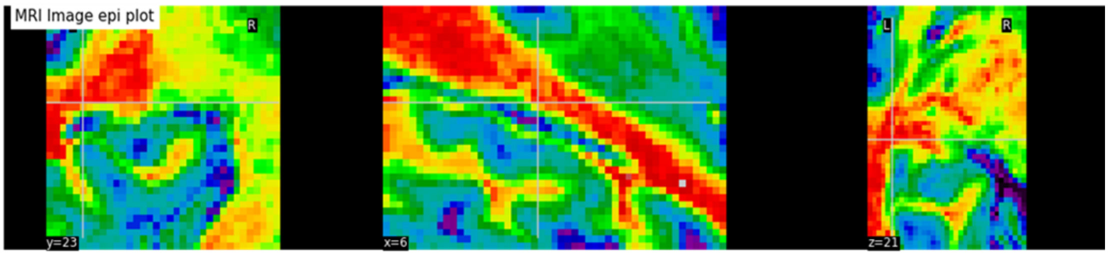

Figure 4 MRI EPI Plot.

The viewing of all the slices of an MRI slices (Figure 5) with annotation (Figure 6) helped me comprehend that the hippocampus is not present in every slice of the MRI image; rather, it is present in only some of the slices of an MRI image.

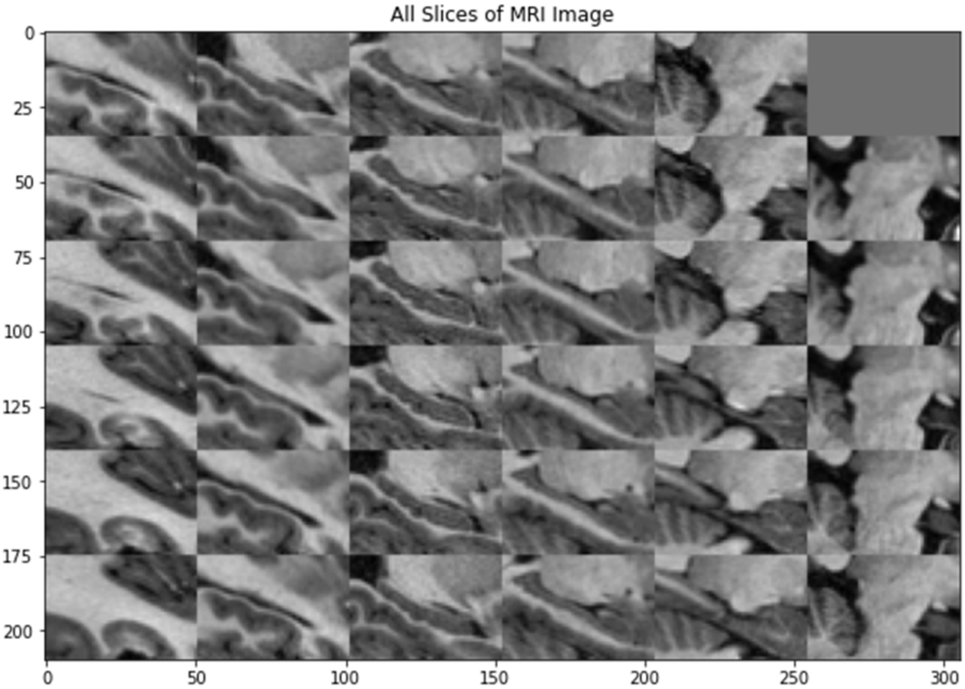

Figure 5 All Slices of Brain MRI.

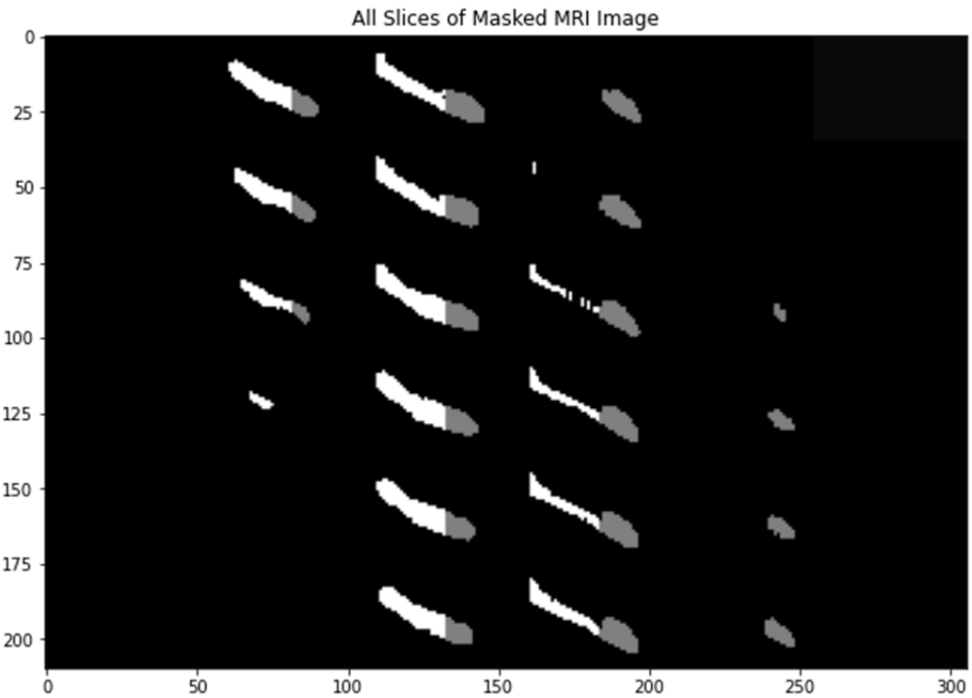

Figure 6 Annotated Slices of Brain MRI.

Data Preprocessing

Data preprocessing is a critical step that must be completed before the model can be trained. If the data has been properly preprocessed, the model will be able to produce good output results; otherwise, the model will not be able to converge and will produce poor results. Despite the fact that the Hippocampus dataset was prepared with great care, certain key pre-processing procedures were overlooked. In the first stage, I visualized the size distribution of all MRI pictures in the dataset.

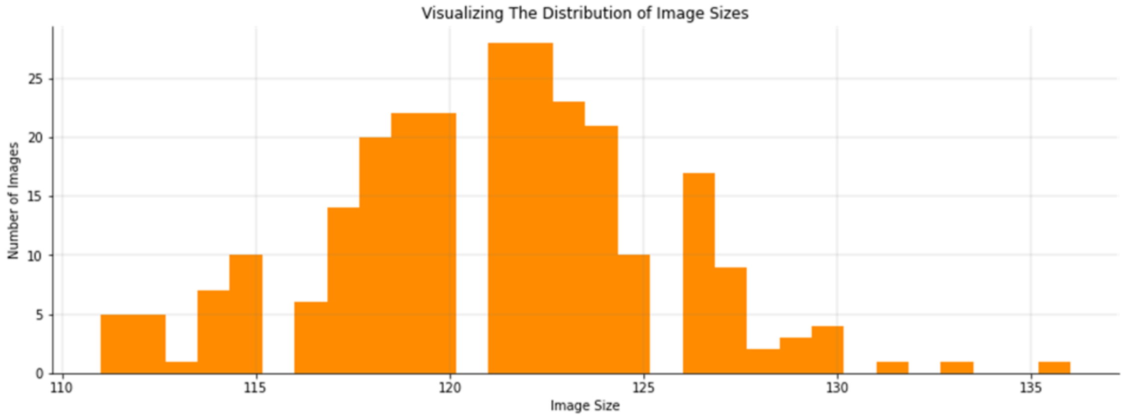

Figure 7 Distribution of all MRI Image Sizes.

When we look at the distribution of MRI picture sizes (Figure 7), we can see that the majority of the images are of various sizes, thus the first stage in the process will be to reshape the image data into a uniform format. The majority of the MRI scans are 35x51 in width and height, with others being 40x50. Because these uncertainties in the data might lead to bias during training, I rearranged the data into 64x64 pixels of width and height. Because the original pictures are smaller, I have appended the matrix with a value of "0," which refers to a black pixel and will not cause a bias in the dataset (Figure 8).

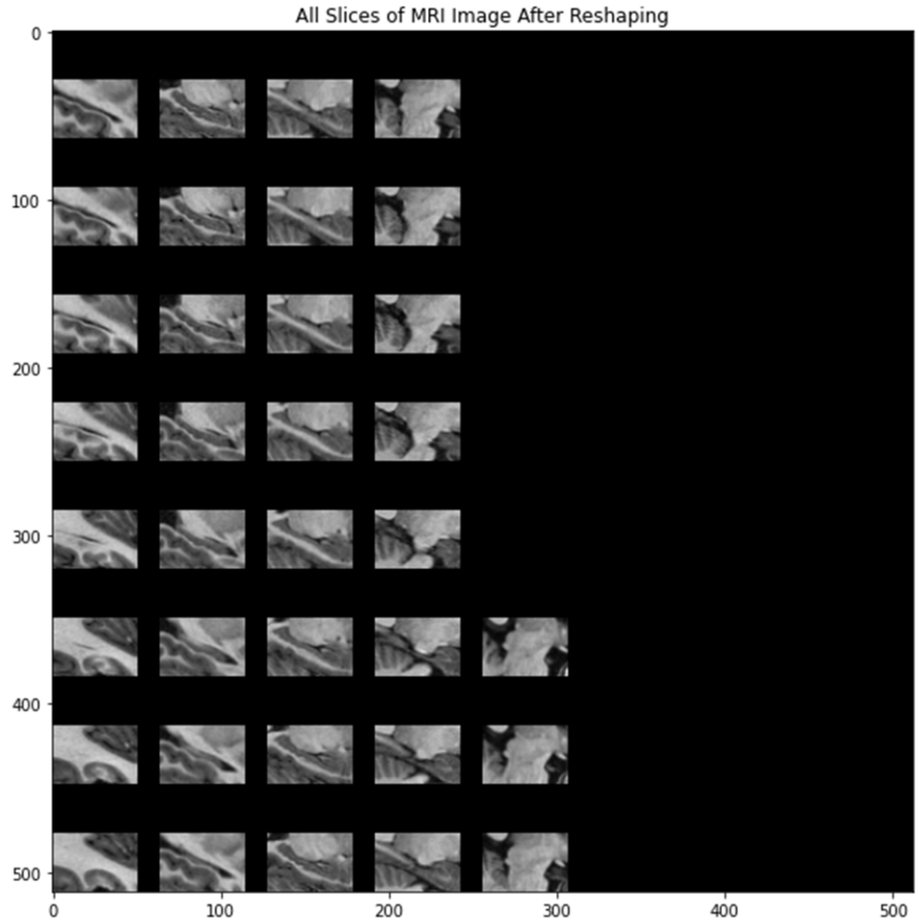

Figure 8 Slices of MRI After Preprocessing.

Data Augmentation

Data augmentation allows us to produce several picture samples. Photographs may be rotated, a zoom effect can be added to images, and the brightness of an image can be raised or lowered as part of data augmentation. Small datasets benefit from data augmentation since it creates several samples of each image. This method not only helps us create additional data, but it also aids the model in overcoming the problem of overfitting.

In our dataset, I'll use on-the-fly data augmentation, which implies that the augmented data for each batch of photographs will be created while training. I randomly rotated the slices of an MRI image for data augmentation; this offered diversity in the dataset and helped the model overcome the overfitting problem. Data augmentation done dynamically during training is a resource-efficient method since it does not require a lot of resources to store enhanced data with the original data prior to model training.

IV. CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES

Convolutional Neural Networks are an extension of Artificial Neural networks which are inspired by neuroscience in which a network of neurons works together to fetch a result. There are different types of Neural networks and each targets a single category. We have Convolutional Neural Networks (CNN) which are widely used for computer vision tasks, these projects involve image classification, image segmentation, object detection and object tracking. Recurrent Neural Networks (RNN) are also a type of Neural networks which target time series problems and are mostly used with Natural Language processing tasks.

The main difference between the ANN and CNN are based upon neurons and their working. The neurons in CNN are organized to work with three dimensions which include height, width, and depth. For working with images Convolutional Neural Networks (CNN) are widely used. CNN first extracts feature from the images and then they work on those extracted features to get a result. CNN has multiple hidden layers, and there are different types of layers used in CNN. The types of layers used in CNN are explained below.

- Convolutional Layer is the first layer used in CNN model which is used to extract the features from the input images. The output of that layer is called a Feature map which gives us the information about edges and corners of the objects in the image. The initial conv layers are used to extract simple features of the objects, and then later conv layers can extract complex object shapes. RELU activation function is used in conv layers.
- Pooling Layer is used to decrease the size of the feature map extracted by the convolutional layer. This helps the model to use less resources as compared to the model in which there is no pooling layer used, it also helps the model to generalize in a better way and overcome overfitting

problems.

- Dropout is used to overcome overfitting problems in the model. Sometimes we have complex models which give high training and low validation accuracy. So, in that case Dropout layers are used, this layer is used to drop the percentage of neurons from the layer, if dropout 0.2 is given then 20% of random neurons will be switched off and their output will not be utilized in the next layer.
- Fully Connected Layers are used before the output layer. They are used when we flatten the output of the last Conv Layer in the model. The flattened vector goes as the input in some Fully connected layers before it will be used by the output layer.

ResNet Architecture

We can create deep neural networks for complicated classification and segmentation tasks thanks to technological improvements, however as the number of layers in the model rises, so does the model's complexity, and a gradient vanishing problem might occur. The deeper network may not function effectively at times because of the gradient vanishing problem, which occurs when the weights are unable to fully converge during backpropagation. The gradient vanishing problem was most of the time addressed with the help of normalization layers, these layers helped the large deep models to converge effectively towards the better solution. The vanishing problem has been addressed by which more and more deep models have been developed. After the gradient vanishing problem, the deep models tend to expose a new degradation problem, this happens when the Deep CNN model starts converging to the solution but suddenly, we see a reduction in accuracy and that reduction happens rapidly. In this situation if we add more layers in the model, it will make the model much deeper/complex and the situation will get worse. One might think that the reduction in accuracy is the cause of overfitting, but the training error also increases higher and higher, so we cannot say that this is an overfitting problem. Deep Residual Network (ResNet) was established in 2015, allowing us to train incredibly deep neural networks without losing information and allowing them to converge in a favorable way. Deep Residual network also resolved the problem of accuracy reduction with very deep models and let us train complex and deep models that can be converged easily with few hyperparameter modifications. The ResNet model employs Residual Blocks, which employ the notion of identity shortcut connection. Identity shortcut connection refers to the technique in which the layer skips some of the layers and then creates a direct connection with the output of that layer. As suggested, the use of shortcut connections does not increase the computational complexity of the model, nor does it add some additional trainable parameters to the model. In the diagram below, we can observe a direct link formed by skipping numerous layers, which is referred to as an Identity shortcut connection.

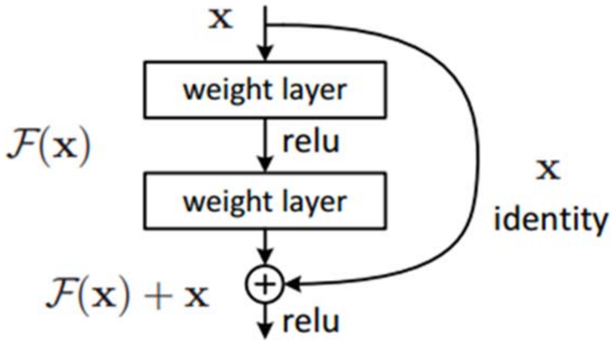

Figure 9 ResNet Architecture.

ResNet provides us with multiple deep architecture options such as: ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152. The number after the ResNet name shows the number of layers in that architecture. We can see that ResNet-152 is the deepest model with 152 trainable layers. ResNet provided us the ability so we can train that much deeper models without worrying about the gradient vanishing or overfitting problems. In the figure we can visualize the ResNet-34 architecture, in the architecture we can see a lot of identity/shortcut connections, these connections help the model to converge quickly in a good way and it also provided the ability to the model so it can overcome overfitting and other problems.

ResNext Architecture

The CNN models are well known due to their complexities and the difficulties which researchers face while developing a new model architecture. Designing a new CNN architecture is a difficult task because of the number of available hyper-parameters which needs to be carefully used, a slight change in a hyper-parameter can lead the model to perform in a worst way. With the development of ResNet architectures we can create a very deep neural network model, with the increment in layers it makes it much more complex to fine tune the hyper-parameters to get good final results from the model. As we know, ResNet architecture is developed by stacking layers with an identity connection link which makes the model deeper and can perform in a good way without losing information. A different CNN architecture which is also known as Inception Network introduced a strategy which is called as split-transform-merge. This strategy splits the input into lower-dimensional embeddings and then after applying the convolutional filters these can again be merged by concatenation technique. It introduced a new paradigm in which the depth of the model was not considered an important thing, but a single layer can be much larger in size because of the split-transform-merge strategy. This behavior of inception models has been merged with the ResNet architecture and developed ResNext architecture. The ResNext architecture stacks the layers like ResNet architecture but also focuses on the split-transform-merge strategy. The main focus of ResNext architecture is on the technique called Cardinality. Cardinality refers to the size of the set of transformations that need to be applied. In the ResNext design, the Residual block is replaced by a block that concentrates on the Split-Transform-Merge technique. The input is distributed over a sequence of convolutional blocks in these ResNext blocks, with the filters applied to each block before combining the outputs. In ResNext architecture, this concept is referred to as Cardinality and is also known as group convolution. The ResNext model's architecture is depicted in the diagram below.

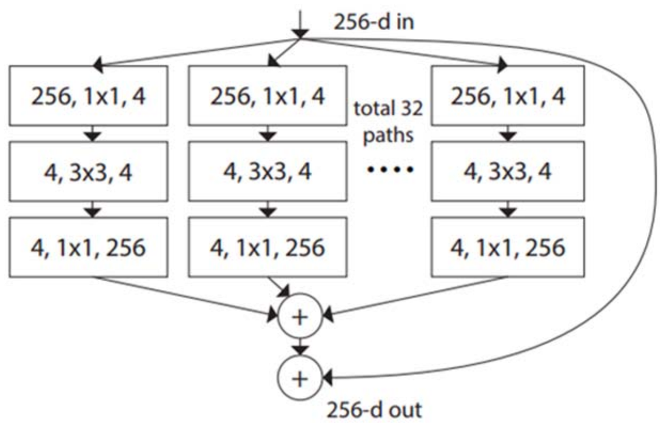

Figure 10 ResNext Architecture.

SE-ResNet and SE-ResNext Architecture

Squeeze and Excitation Networks are the foundation of the SE-ResNet paradigm (SE-Net). The SE blocks in the network are integrated with other CNN model architecture to improve the model's performance for a small additional computational cost. If we look into the traditional CNN model architectures the convolutional layers are responsible for feature extraction, they learn about the connectivity of spatial patterns along with the input channel and then they fuse this spatial pattern information and channel-wise information to learn about the complex patterns inside the given image. Keeping this theory in mind, more powerful feature extraction and pattern representation techniques have been in research which will enable the model to focus more on the required properties or features of the image, this will provide us with effective models with less error rates. Research have shown that the development of new and innovative architectures can improve the performance of CNN models. The SE-Net Squeeze and Excitation network provides a different approach to develop a CNN architecture. The SE-Net focuses more on the relationship between the channels of the input image, it provides the ability by which the model focuses more on the interdependencies between the channels of its convolutional features. With the help of SE-Net architecture the model performs feature recalibration, and by this ability the model gives preference to informative features, and it suppresses those features which are not useful for the particular task. SE-Net makes use of SE-Blocks which are responsible for the feature recalibration and other tasks. An SE-Net architecture can be developed just by stacking SE-Blocks or the SE-Blocks can also be used with the traditional deep neural network architecture to improve the performance of the existing model architecture. SE blocks which are coupled with ResNet architecture provided SE-ResNet architecture. The deep neural network can conduct dynamic channel-wise feature fine tuning using SE blocks in conjunction with ResNet blocks. The SE-ResNet architecture is seen below; we can see that the SE block is utilized after the residual block, and then the Identity shortcut connection or skip connection is used after that. SE-ResNet outperforms the standard ResNet network in terms of performance, dataset generalization, and the ability to utilize more deeper models without worrying about overfitting.

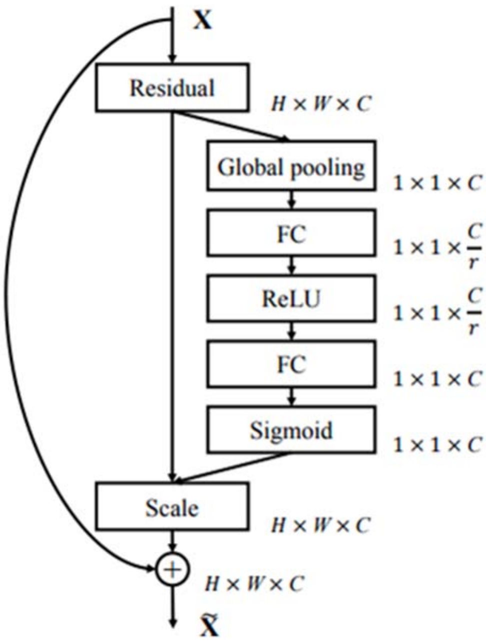

Figure 11 SE-ResNet Architecture.

SE-Blocks have also been integrated with the ResNext network architecture and developed a new SE-ResNext model. The ResNext network is built on the concept of cardinality, and the Squeeze and Excitation Blocks combine with the ResNext design in the SE-ResNext architecture, allowing the network to focus more on dynamic channel-wise feature fine tuning.

DenseNet Architecture

Due to the advancement of technologies, we are able to develop, train and test more and more complex and deep neural network models. Deep Neural Network models are complex and also lose information on some of the layers, if we look at the network architecture of ResNet-152, we can see that this architecture is based on 152 layers and with the help of identity connections the network is able to pass the weights throughout the network without losing information. This direct connection sometimes also loses some information from some layers as it skips multiple layers to connect so the chances are some of the information may be lost. DenseNet provides a new network architecture which connects all the layers directly with each other, this ensures the maximum flow of information between all the layers. In the DenseNet architecture Dense Blocks have been used alongside convolutional layers and pooling layers. In Dense Blocks the parameters that have been passed through layers are relatively lower than some other traditional CNN architectures, this is because the Dense Block connects all layer feature maps together and in that way, there is no point in learning something which has been learned in the previous layers. So, in that way the DenseNet architecture also helps in parameter reduction. DenseNet layers are relatively narrower than the traditional CNN architectures, as for example: it takes 12 filters per layer, and it adds a small set of feature maps to extract useful information from the layer. This approach allows the network to learn variances in feature maps, increasing the network's efficiency. The network architecture of dense blocks is depicted in the diagram below. One of the main differences between DenseNet and other traditional CNN architectures is the ability to work with narrow layers, as it was mentioned before that for complete flow of information all the preceding layers have been connected with the later ones, so in that case layers do not require a large amount of feature parameters. For setting the parameters of layers, the growth rate of the network is used. Growth rate is basically the number of parameters that a layer can have, and experiments have shown that a relatively small growth rate can also be enough for DenseNet in certain cases.

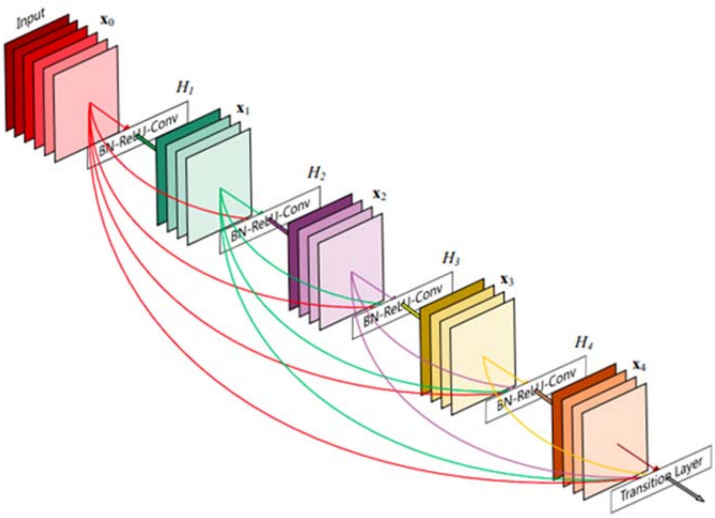

Figure 12 DenseNet Architecture.

V. 3D SEMANTIC SEGMENTATION

Computer Vision is one the vast fields in Machine Learning. In Computer Vision we have multiple types of problems for example, image classification, multi class image classification, object detection, and image segmentation. Each of them deals with a specific category of task.

Semantic Segmentation is a type of problem in computer vision which deals with finding objects and also their locations in an image. If we talk about Image classification problems then in those models we classify the whole image according to specific labels, for example: if the image is of a cat or dog. But in segmentation problems we detect and classify objects inside the image. Detection means, the model will find its location and size inside the image and classify it according to the provided labels, in other words we can say that in segmentation we classify each pixel of an image and in that way, we can detect and localize an object inside the image. There are two types of segmentations in Machine learning: semantic segmentation and instance base segmentation. In Instance base segmentation each object of a class has been treated differently, or in other words the model is instance aware. In semantic segmentation we classify and detect objects based on the classes. Semantic segmentation has a lot of applications and currently it's been widely used in medical applications. In medical applications it has been used with 2D and 3D images, 3D imaging data can be taken from CT scans, MRI, etc. With the help of 3D semantic segmentation, we are able to find and calculate the volume of tumors which helps in the diagnosis of multiple diseases. In this research the 3D-Semantic Segmentation model is used to detect and quantify the hippocampus volume from 3D Brain MRI data. There are many available CNN architectures for semantic segmentation problems but while working with medical image data, U-Net CNN architecture outperforms all of it counter-parts.

3D U-Net Segmentation Model

The U-Net architecture is a deep convolutional neural network design which is mostly used for biomedical image segmentation. The U-Net architecture was developed in 2015 and it has provided a great impact in the field of computer vision. U-Net architecture uses the concept of Fully Convolutional Networks and is used for both classification and localization of the object in the image. Let's now understand the architecture of the U-Net model. As the name suggests the architecture of the model is of U-shape. This architecture is built upon convolutional layers and no dense or flatten layers have been used in this architecture. The left side of the architecture is known as the encoder part and the right side is known as the decoder part. The encoder is responsible to extract features from the image and also down samples the image as the image passes through each layer of the encoder. The encoder uses a max pooling layer with a stride of 2, this helps in the down sampling of the image. On the other hand, the decoder which is the right part of the U-Net architecture is responsible for up sampling the image. If we look closely to the architecture of the network, we will also realize that decoders are making use of skip connections, this connection directly connects the output of the encoder layer with the decoder layer. The skip connections are an important part of U-Net architecture as it preserves the loss of the previous layers and make use of that in the decoder layers so that they will have a strong impact on overall values, this technique helps the model to learn in a good way.

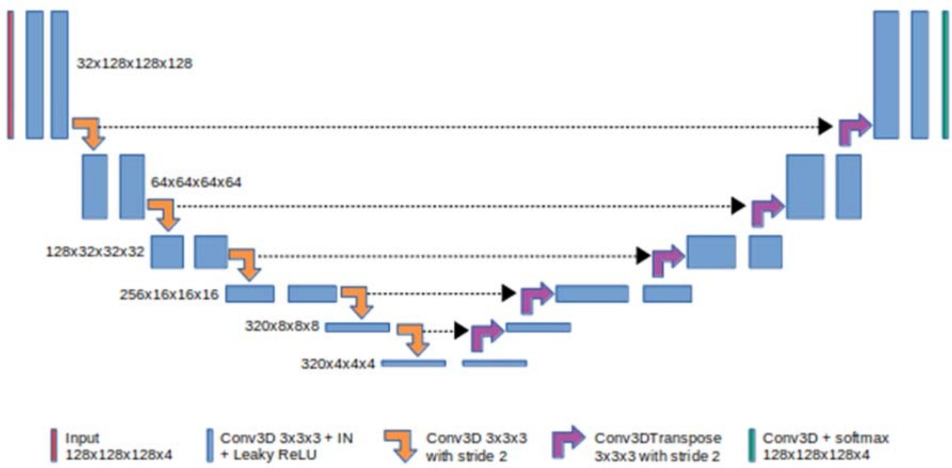

Figure 13 3D U-Net Architecture.

In U-Net architecture we have an architectural element called as Backbone. The backbone is an architectural element that outlines how the layers of encoders will be defined, as well as how the decoder layers will be formed. The backbone of the U-net model can be basic CNN model, Vgg-16 , ResNet architectures, and so on. For example, if a ResNet-34 model is utilized as the backbone of a U-Net model, then the encoder will be of ResNet34 design, while the decoder will be its inverse.

Loss Functions

There are multiple loss functions available for segmentation tasks, each loss function has its own advantages and disadvantages. In segmentation tasks the most used loss functions are Focal Loss and Dice Loss functions. For segmentation tasks, Focal loss is a sort of cross entropy loss that falls under the Distribution based loss category. It works effectively with unbalanced datasets because it emphasizes the contribution of a small number of classes while downplaying the impact of classes with a large number of samples. The cross-entropy formula is updated by adding a modulating factor, as illustrated in the equation below:

If the gamma value is 1, then focal loss is the same as cross entropy, but if the gamma value is increased, it will focus more on the class with less samples. Dice Loss is based on the Dice coefficient measure, which is used to determine picture similarity. The dice coefficient was also utilized as a loss function in 2016. Dice loss is used in semantic segmentation to achieve region-based loss. Dice coefficient can be described in terms of binary data and also for multiclass data. The Dice coefficient which is used for binary data is denoted with (DH), and for multiclass data it is denoted with (DS) . “H” refers to Hard Coefficient and “S” refers to Soft Coefficient. Dice Similarity Coefficient can also be described in terms of True Positives, False Positives and False Negatives.

To make use of both loss functions, combination of Dice and Focal loss is used in this study. These loss functions work well for unbalanced datasets, and in our MRI pictures, the background class has covered a lot of area, making the dataset unbalanced and making it difficult to the model to generalize. In this scenario, combining these loss functions may aid in the model's convergence.

Evaluation Metrics

Evaluation metrics are used to measure the quality of the segmentation model. For this reason, we have used two different Evaluation metrics which are F-Score and IOU score. The model's accuracy was assessed using the intersection over union (IOU) and F-Score metrics. The harmonic mean of precision and memory is provided by the F-Score, which combines precision and recall.

The IOU, commonly known as the Jaccard Index, is a widely used statistic in semantic segmentation. IOU is a useful statistic because it calculates the area of overlap between the predicted segment and the ground truth (Intersection) and divides it by the total area of the predicted segment and ground truth (Union).

VI. EXPERIMENTATION AND RESULTS

The results of testing the U-Net segmentation model with various backbone architectures is presented in this section. The experimental setup is first briefly described, and then the findings from several models were compared for the final analysis.

Experimental Setup

The 3D Brain MRI dataset has been processed and cleaned before training the model, as I have described in the previous section about the various phases of data cleaning, all those steps have been taken to clean and preprocess the data. After that, the dataset was split into two subsets: training and testing. When we have a short dataset, we must employ data augmentation to assist the model overcome overfitting. Because data augmentation creates several samples of the data and takes some resources to retain the enhanced data, runtime data augmentation pipelines with a batch size of four have been designed. While training the model, the images were augmented based on the batch size and then used for training. The loss and evaluation metrics for each epoch were tracked using a tensor board. Learning rate of 0.0001 was used with Adam Optimize. While initializing the model, the threshold of 0.7 was used for both IOU metrics and F-Score metrics, and the pretrained weights of each backbone architecture on the ImageNet dataset were employed. With a batch size of four, each model was trained on 100 epochs.

Results

The training and validation outcomes of the segmentation model with each backbone are given in the table below. The SE-ResNet-50 model produced the greatest results for segmenting the hippocampus in 3D brain MRI data, according to the overall results.

By looking at the table's findings, we can observe that the DenseNet-121 overfitted the data because the training and validation evaluation metric scores are so far off. ResNext-50 and ResNet-50 have also shown positive results, but ResNext-50 exhibits overfitting. Both models with SE blocks performed well in terms of validation and training. In the Figures, we can visualize the evaluation metrics and loss function results for each U-Net model. As for each model the learning rate was 0.0001, this helped the model to learn in small steps and it also helped the model to not miss the global minima. In the loss function visualizations we are not able to find any spikes, this is because the learning rate was not too much high, if the learning rate is not optimal then the model can lose a lot of information or maybe it will going to miss global minima and after a certain point the loss will going to suddenly starts increasing which is a bad sign for a machine learning model, also the spikes in the loss function can indicate problems and suggests that we need to optimize the hyperparameters for better results.

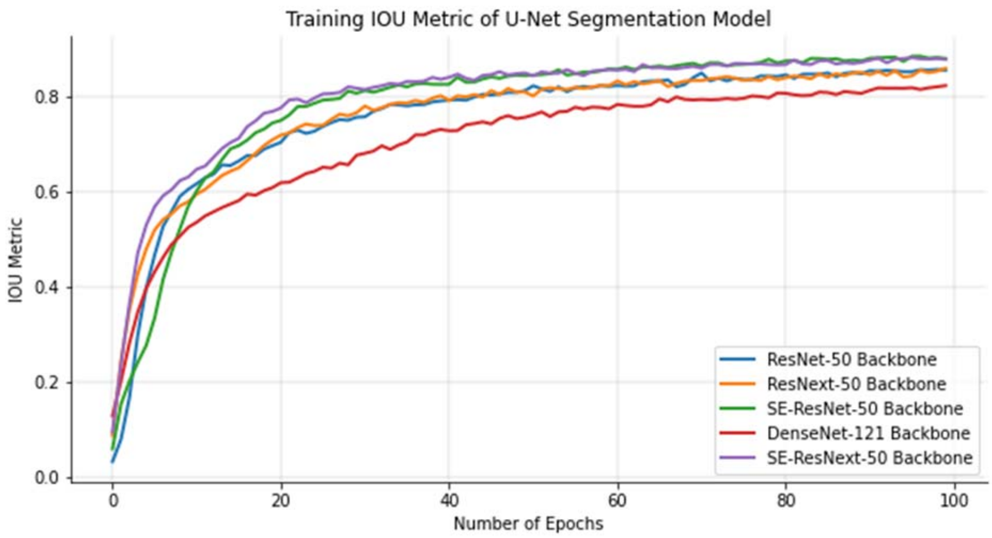

Figure 14 Training IOU Metric.

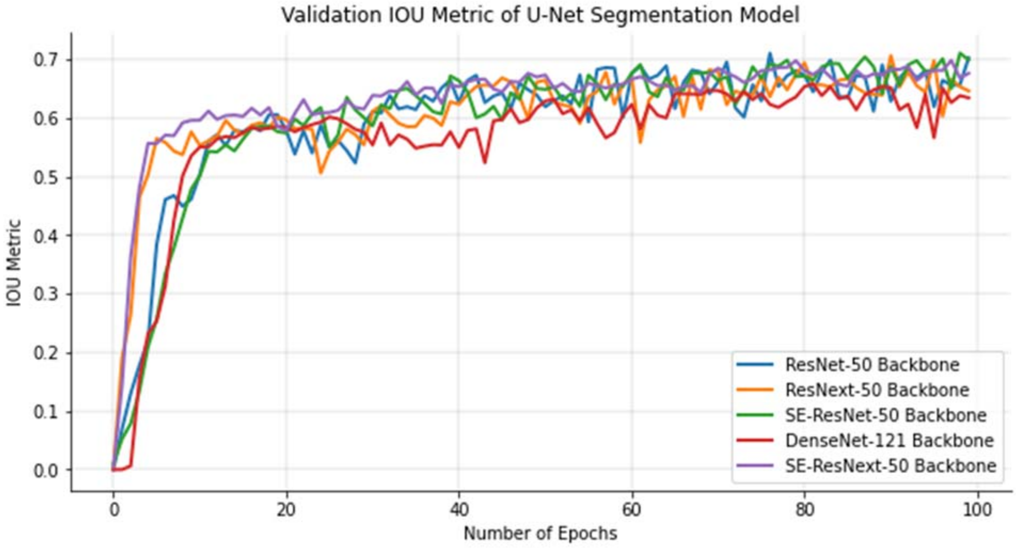

Figure 15 Validation IOU Metric.

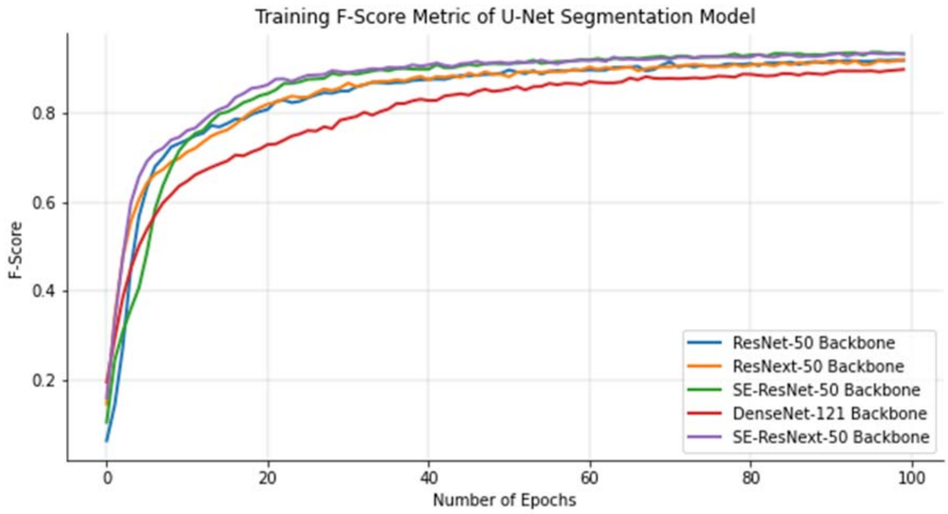

Figure 16 Training F-Score Metric.

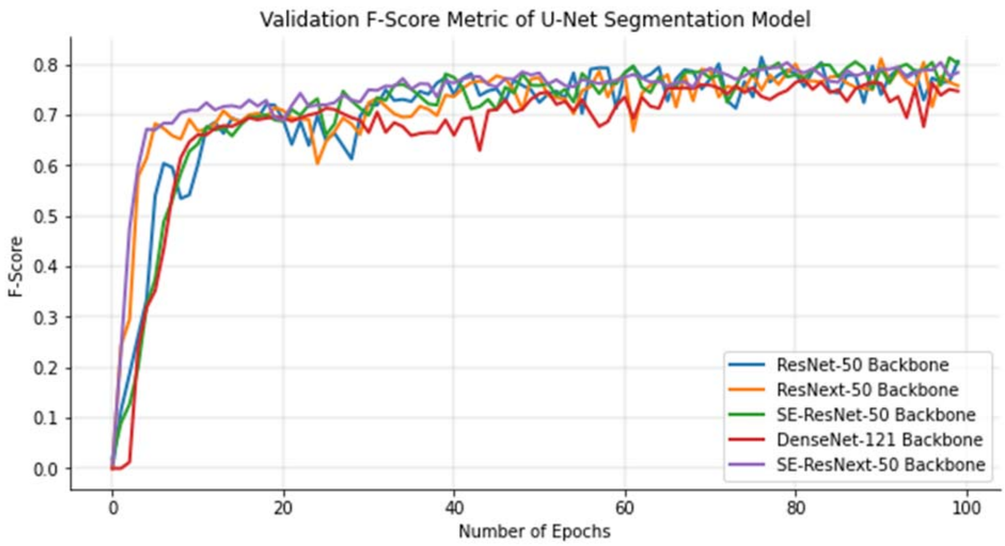

Figure 17 Validation F-Score Metric.

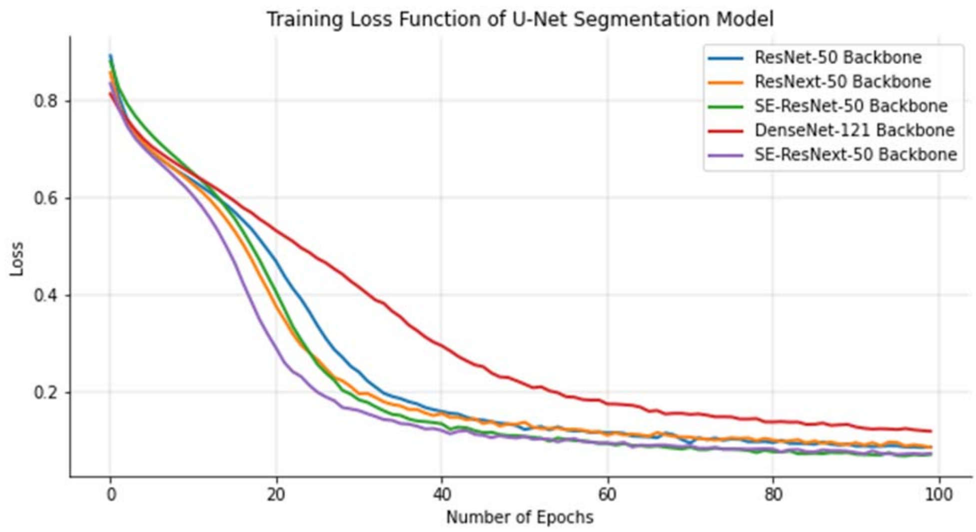

Figure 18 Training Loss.

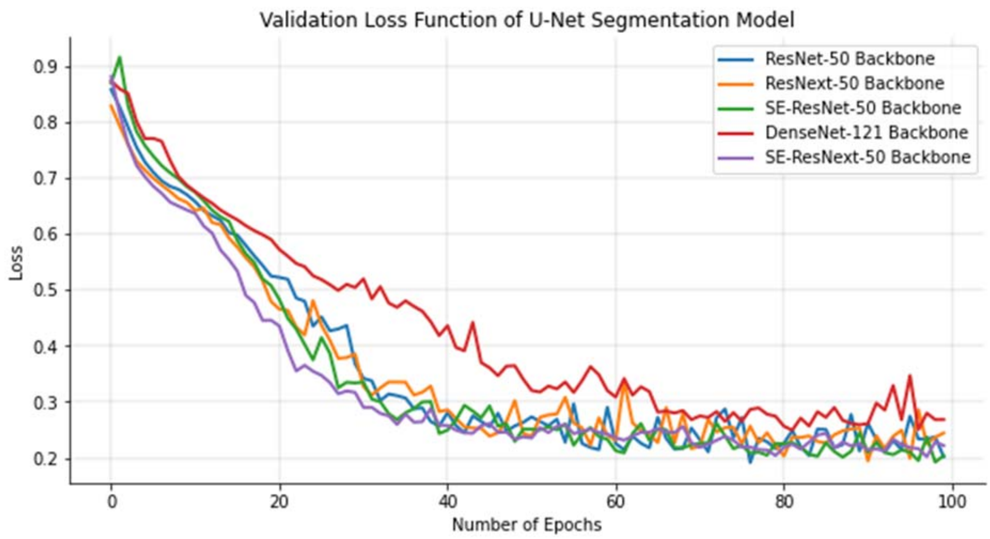

Figure 19 Validation Loss.

VII. CONCLUSION AND FUTURE WORK

The 3D U-Net segmentation model was used to perform semantic segmentation on a 3D MRI dataset . The dataset has been preprocessed using the standard preprocessing techniques. As the dataset was not large data augmentation technique was used to assist the learning of deep learning models. The data augmentation has been applied on the fly, in other words while training the models the data augmentation pipeline has been applied on a batch of images, by the help of this technique I was able to train deep learning models without worrying much about the resource allocation problems because no extra memory resource was used to store the augmented dataset. This study looked at five distinct deep neural network architectures for 3D hippocampal volume segmentation. In terms of training and validation outcomes, all of the neural network architectures when combined with U-Net model performed well, however the SE-ResNet-50 backbone architecture outperformed all of them. The SE-ResNext-50 design functioned admirably; however, it was not as efficient in training as the SE-ResNet-50 architecture. The usage of Combined Dice and Focal Loss Functions also greatly aided the model's generalization. One of the most important stages in predicting Alzheimer's disease development is segmenting hippocampal volume from brain MRI images. The hippocampus volume can be used in conjunction with other cognitive and functional characteristics of patients to accurately anticipate the progression of Alzheimer's disease and to take suitable actions to control it.

VIII. REFERENCES

Books

GROVER, V. P., TOGNARELLI, J. M., CROSSEY, M. M., COX, I. J., TAYLOR-ROBINSON, S. D., & MCPHAIL, M. J. (2015). “Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians”, **J Clin Exp Hepatol**, vol. 5, no. 3, pp. 246-55.

YEUNG, M., SALA, E., SCHÖNLIEB, C. B., & RUNDO, L. (2022). “Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation”, **Computerized Medical Imaging and Graphics**, vol. 95, pp. 102026.

ANAND, K. S., & DHIKAV, V. (2012). “Hippocampus in health and disease: An overview”, **Annals of Indian Academy of Neurology**, vol. 15, no. 4, pp. 239-246.

CARMO, D., SILVA, B., YASUDA, C., RITTNER, L., LOTUFO, R. (2021). “Hippocampus segmentation on epilepsy and Alzheimer's disease studies with multiple convolutional neural networks”, **Heliyon**, vol. 7, no. 2, e06226.

STEHLING, M. K., TURNER, R., & MANSFIELD, P. (1991). “Echo-planar imaging: magnetic resonance imaging in a fraction of a second”, **Science,** vol. 254, no. 5028, pp. 43-50.

Articles

XIE, S., GIRSHICK, R., DOLLÁR, P., TU, Z., & HE, K., "Aggregated Residual Transformations for Deep Neural Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), http://dx.doi.org/10.1109/CVPR.2017.634.

HU, J., SHEN, L., & SUN, G., "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, http://dx.doi.org/10.1109/CVPR.2018.00745.

HE, K., ZHANG, X., REN, S., & SUN, J., "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), http://dx.doi.org/10.1109/CVPR.2016.90.

HUANG, G., LIU, Z., VAN DER MAATEN, L., & WEINBERGER, K. Q., "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), http://dx.doi.org/10.1109/CVPR.2017.243.

REZATOFIGHI, H., TSOI, N., GWAK, J., SADEGHIAN, A., REID, I., & SAVARESE, S., "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), http://dx.doi.org/10.1109/CVPR.2019.00075.

LIN, T. -Y., GOYAL, P., GIRSHICK, R., HE, K., & DOLLÁR, P., "Focal Loss for Dense Object Detection," 2017 IEEE International Conference on Computer Vision (ICCV), http://dx.doi.org/10.1109/ICCV.2017.324.

LIU, S., & DENG, W., "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), http://dx.doi.org/10.1109/ACPR.2015.7486599.

KINGMA, D., & BA, J., "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations, https://arxiv.org/abs/1412.6980.

DENG, J., DONG, W., SOCHER, R., LI, L. -J., LI, K., & FEI-FEI, L., "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, http://dx.doi.org/10.1109/CVPR.2009.5206848.

PEREZ, L., & WANG, J., "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," Technical Report or Preprint, https://arxiv.org/abs/1712.04621.

ANTONELLI, M., REINKE, A., BAKAS, S., FARAHANI, K., ANNETTEKOPP-SCHNEIDER, A., LANDMAN, B., LITJENS, G., MENZE, B., RONNEBERGER, O., SUMMERS, R., GINNEKEN, B., BILELLO, M., BILIC, P., CHRIST, P., DO, R., GOLLUB, M., HECKERS, S., HUISMAN, H., JARNAGIN, W., & CARDOSO, M. J., "The Medical Segmentation Decathlon," Technical Report.

RONNEBERGER, O., FISCHER, P., & BROX, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," Lecture Notes in Computer Science (LNCS), 9351, pp. 234-241, doi: 10.1007/978-3-319-24574-4_28.

SOKOLOVA, M., JAPKOWICZ, N., & SZPAKOWICZ, S., "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science (LNCS), 4304, pp. 1015-1021, doi: 10.1007/11941439_114.

PRENCIPE, B., ALTINI, N., CASCARANO, G.D., BRUNETTI, A., GUERRIERO, A., & BEVILACQUA, V., "Focal Dice Loss-Based V-Net for Liver Segments Classification," Applied Sciences (Appl. Sci.), 12, 3247, doi: 10.3390/app12073247.

RESUME

FELLA ACHOURI

SUMMARY

An experienced Data Scientist having a passion for leveraging my expertise in programming, machine learning, and AI to create intelligent systems that solve real-world problems. My project portfolio showcases my ability to apply these skills in a variety of contexts, from predicting personality traits using NLP to classifying skin diseases with CNN. Currently, I am pushing the boundaries of AI at Atilax, where I am developing a system that uses NLP to extract meaningful insights from unstructured data. I am committed to driving innovation and helping organizations make data-driven decisions with greater speed and accuracy.

SKİLLS

PROGRAMMİNG LANGUAGES: Python, JavaScript, C++, Java, Visual Basic, Assembly Language.

PYTHON LİBRARİES: Numpy, Pandas, Scikit-learn, Keras, TensorFlow.

MACHİNE LEARNİNG AND AI: Deep Learning (ANN, CNN, RNN), Natural Language Processing, Computer Vision.

DATABASES: Oracle, SQL, NoSQL.

DATA VİSUALİZATİON TOOLS: Matplotlib, Seaborn.

DESKTOP APPLİCATİON DEVELOPMENT: Visual Basic.

NETWORKİNG: Cisco 1, Cisco 2.

PROJECTS

Big Five Personality Prediction Using Natural Language Processing

- Developed an NLP model to Predict Big Five Personality Scores from written text. It was used to analyze the personality of different users.
- Achieved MSE of 0.03, the loss error.

3D Segmentation of Hippocampus Volume in Brain MRI

- Developed a hybrid RNN and CNN model to classify 10 different genres of songs.
- Achieved 85% testing accuracy.

Skin Disease Classification with CNN

- Developed a CNN model to classify 3 different types of skin diseases.
- Achieved 88% testing accuracy.
- Integrated it into a web application for end users.

Traffic Sign Classification with CNN

- Developed a CNN model to classify different types of Traffic Signs.
- Achieved 98% testing accuracy.

Hospital Management System for CHE Constantine Hospital

- Developed a Desktop application for hospital management using Visual Basic.
- This system was designed to streamline the management of doctors' schedules, facilitate the transmission of patient test reports to physicians, handle patient data files, and oversee hospital room allocation.
- This system also connected laboratories with the hospital.

EXPERİENCE

Data Scientist **Nov 2022 – present**

Atilax, Remote

As an AI professional specializing in Natural Language Processing (NLP), I am passionate about building intelligent systems that can process and understand human language. My current focus is on developing a smart system that leverages NLP to extract meaningful insights from unstructured data, enabling organizations to make data-driven decisions with greater speed and accuracy. Through my work, I am helping to push the boundaries of what is possible with AI and driving innovation in this exciting field.

EDUCATİON

BS Software Engineering **2015 - 2019**

Université Constantine 2 Abdelhamid Mehri, Constantine, Algeria

Turkish Language **2019 - 2020**

İstanbul Aydin University, Istanbul, Türkiye

Master’s Data Science & AI **2020 - 2023**

İstanbul Aydin University, Istanbul, Türkiye