Semantic Image Segmentation using UNET

Presented By

1. Prakhar Agarwal 20135092
2. Samriddhi Jain 20135107
3. Reva Tamaskar 20135127
4. Tejas Tomar 20135129
5. Tushar Raghav 20135130
6. Chhavi Bhardwaj 19135030

Guided by Dr. Prabhas Bhardwaj

Introduction

According to the 2018 data from the International Association of Cancer Registries, more than 28,000 cases of brain tumours are reported in India every year, and more than 24,000 people die due to brain tumours annually.

Out of the various processes involved in diagnosing a brain tumour, imaging tests like Magnetic Resonance Imaging (MRI) and Computerised Tomography (CT) scan are commonly used.

Here, in this project, we use Convolutional Neural Networks (CNN) and Deep Learning using U-Net architecture to segment out the brain tumour from a dataset of MRI images.

What is Artificial Intelligence?

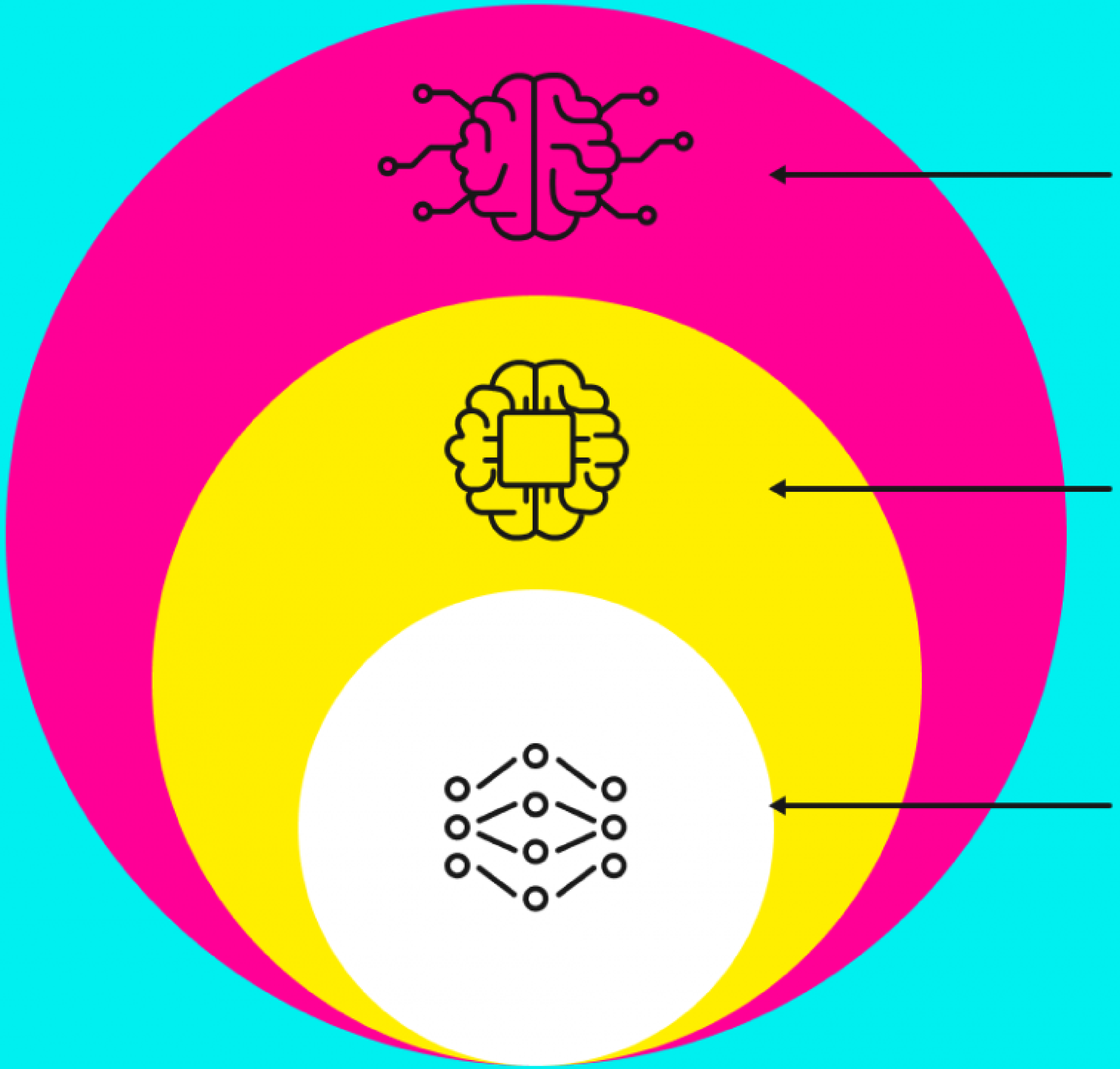
AI is the broader scope of the entire system. Simply put, AI is the process where we use computers to mimic the cognitive abilities and functions of humans in order for machines to carry out tasks in an intelligent way. We do this by creating sets of algorithms by which computers navigate and complete these processes.

What is Machine Learning?

Machine learning is a subset of AI which focuses primarily on the ability of machines to receive data and then learn for themselves. Computers can be trained to automate tasks that would be impossible and time-intensive for a human to complete.

What is Deep Learning?

As machine learning is a subset of artificial intelligence, so deep learning is, in turn, a subset of machine learning. Think of it as the last, tiny piece of the Russian doll of AI. Deep learning can sometimes be referred to as ‘deep neural networks’ as they have many layers involved. Where a ‘normal’ neural network has only one single layer of data to process, deep learning can have multiple layers, hence the name.



Artificial Intelligence (AI)

Any process where we use computers to mimic the cognitive abilities and functions of humans.

Machine Learning (ML)

A subset of AI. It focuses on the ability of machines to receive data and improve at a task with experience.

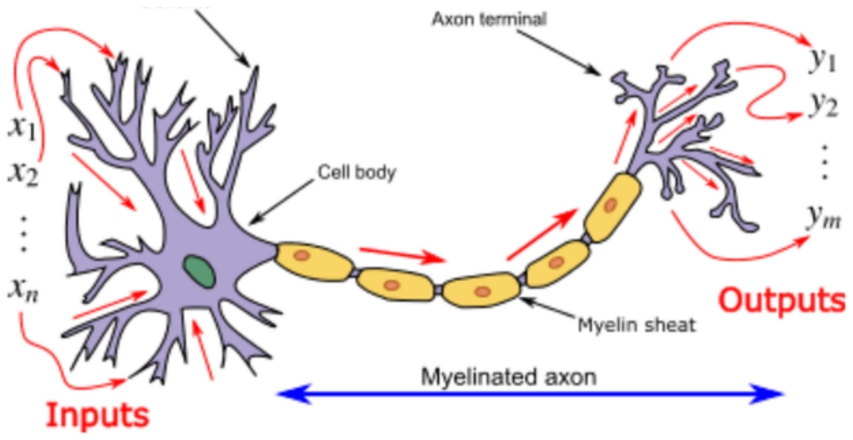
Deep Learning (DL)

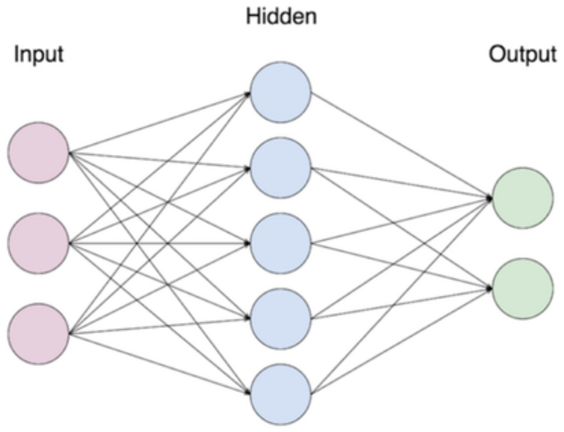
A subset of ML. It works with multi-layered neural networks that can learn by themselves from vast amounts of data.

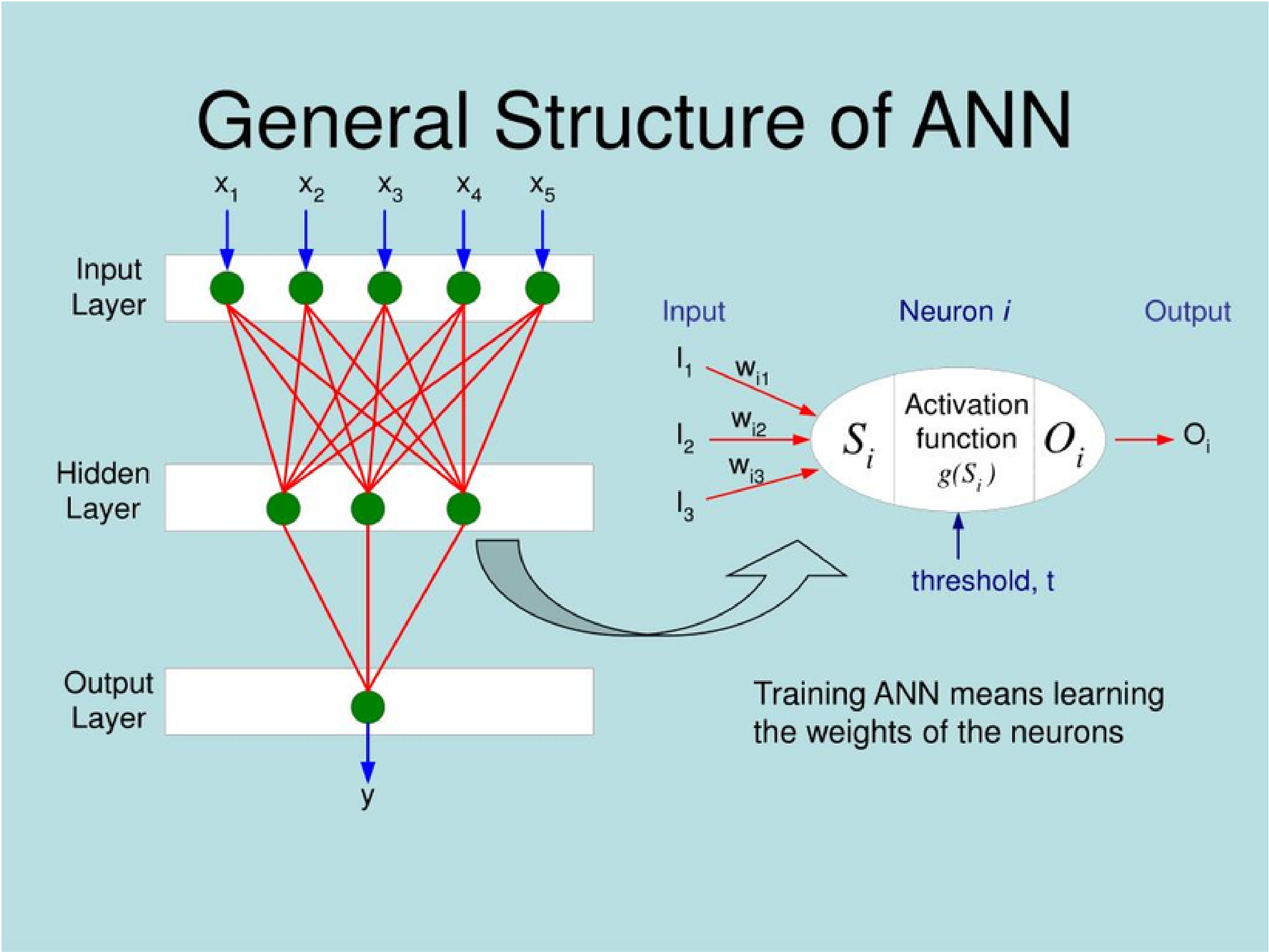
WEAREBRAIN.COM

What is ANN?

Artificial neural network. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another.



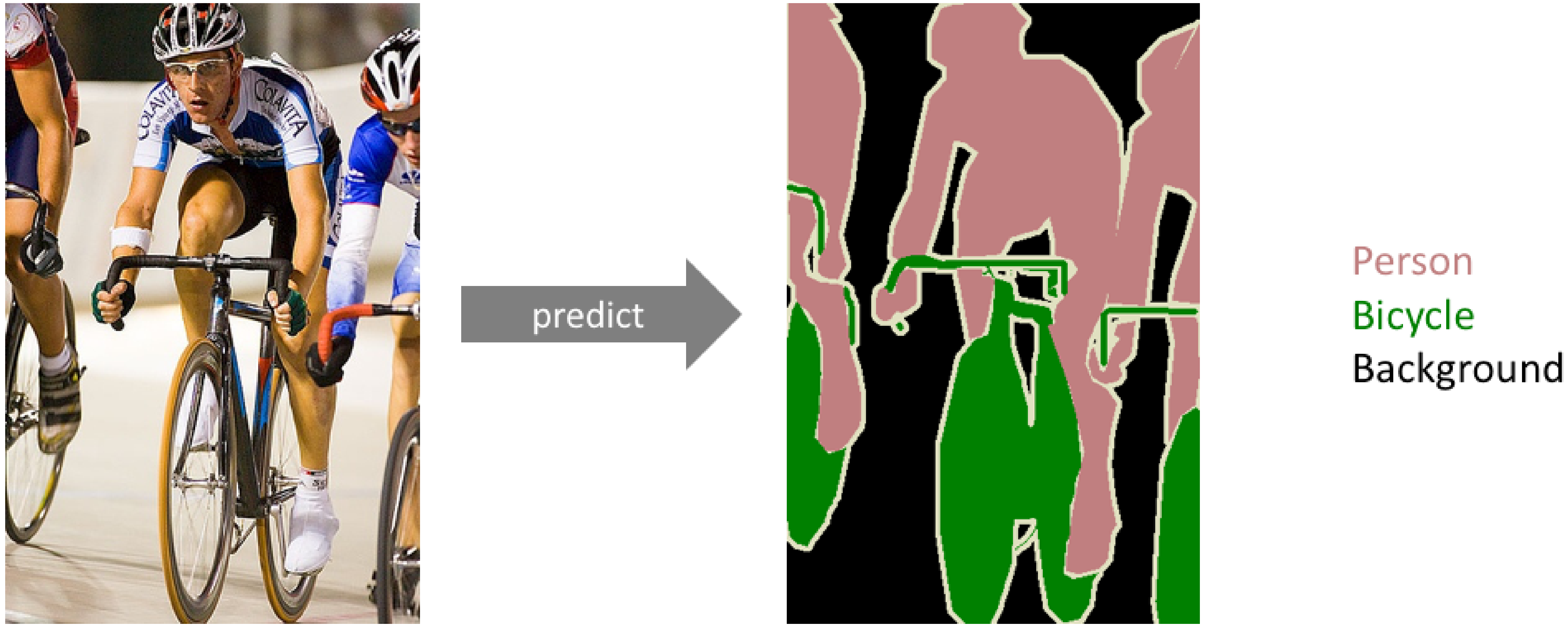




Semantic Image Segmentation

Semantic Image Segmentation is the process of assigning a label to every pixel in the image. It is different from simply classifying or detecting an object in the image as it determines the boundary and edge of every individual label (group).

In the image shown, the labels(groups) are person, bicycle and the background.



Semantic Image Segmentation

Input

segmented

1: Person
2: Purse
3: Plants/Grass
4: Sidewalk
5: Building/Structures

3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	1	1	1	1	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	5	5	5	5	5	5	5
5	5	3	3	3	3	1	1	3	3	5	5	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	4	4	5	5	5
4	4	4	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4

Semantic Labels

This is an example of how the labels are assigned to every pixel of the image.

Understanding Convolutions, Max Pooling, Padding, and Transposed Convolutions

Convolutions:-

Convolution is the simple application of a filter or feature detector to an input that results in an activation. Repeated use of the filter results in a feature map, which indicates the location and strength of the feature on the input, such as an image. In our project, the input is a nin x nin x (channels) matrix whereas the filter map is (f x f x channels) sized matrix.

The relation between nin and nout is as follows:

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1$$

n_{in}: number of input features
n_{out}: number of output features
k: convolution kernel size
p: convolution padding size
s: convolution stride size

Max Pooling:-

It is a pooling operation in which we select the maximum value in the region covered by the filter map. The matrix after the max pooling operation will contain the most prominent features of the previous feature map.

Padding:-

In padding, we extend the area on which the Convolutional Neural Network processes an image. Padding allows the filter to cover more space on the image, thereby giving more accurate results.

Transposed Convolutions

Transposed Convolutions are used to upsample the input feature map to a desired output feature map using some learnable parameters.

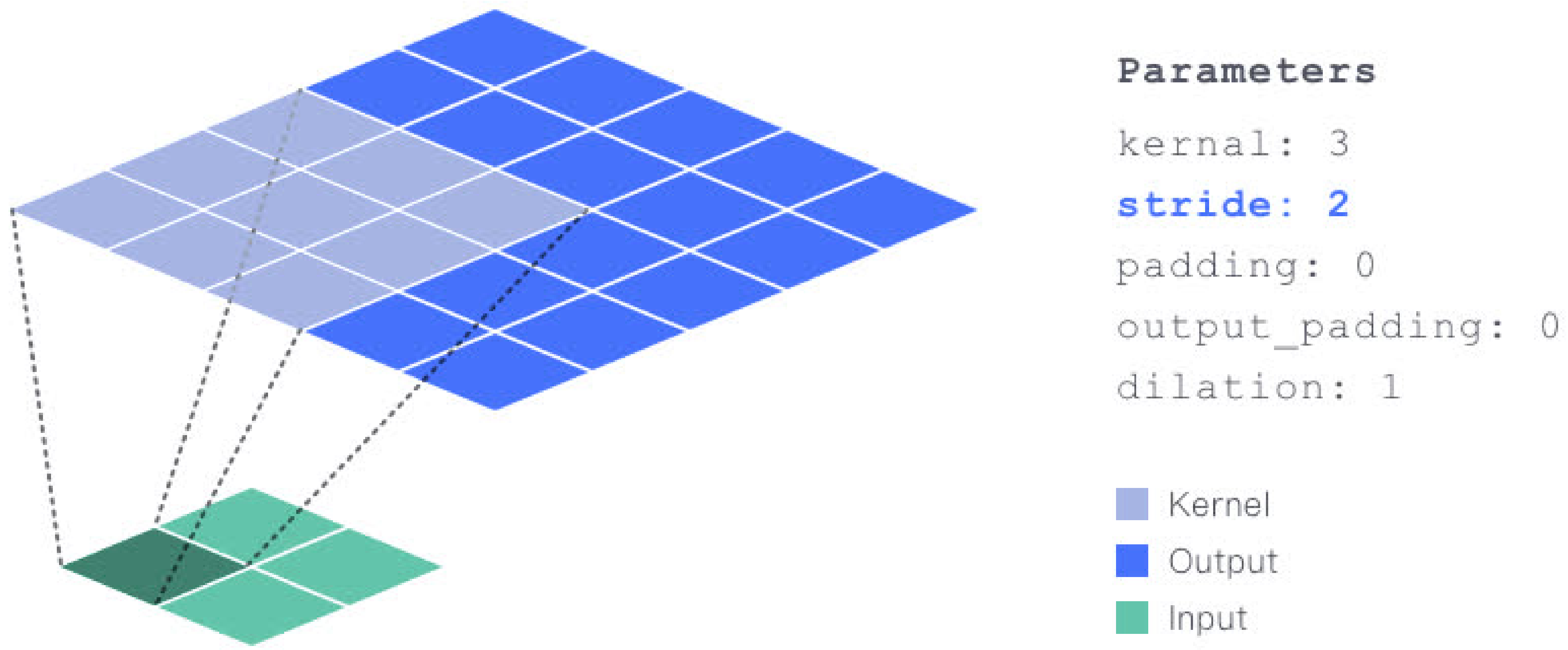
1.Conside a 2x2 encoded feature map which needs to be upsampled to a 5x5 feature map.

2. We take a kernel size of 3x3 with stride 2 and zero padding.

3. Now we take the upper left element of the input feature map and multiply it with every element of the kernel . Similarly, we do it for all the remaining elements of the input feature map.

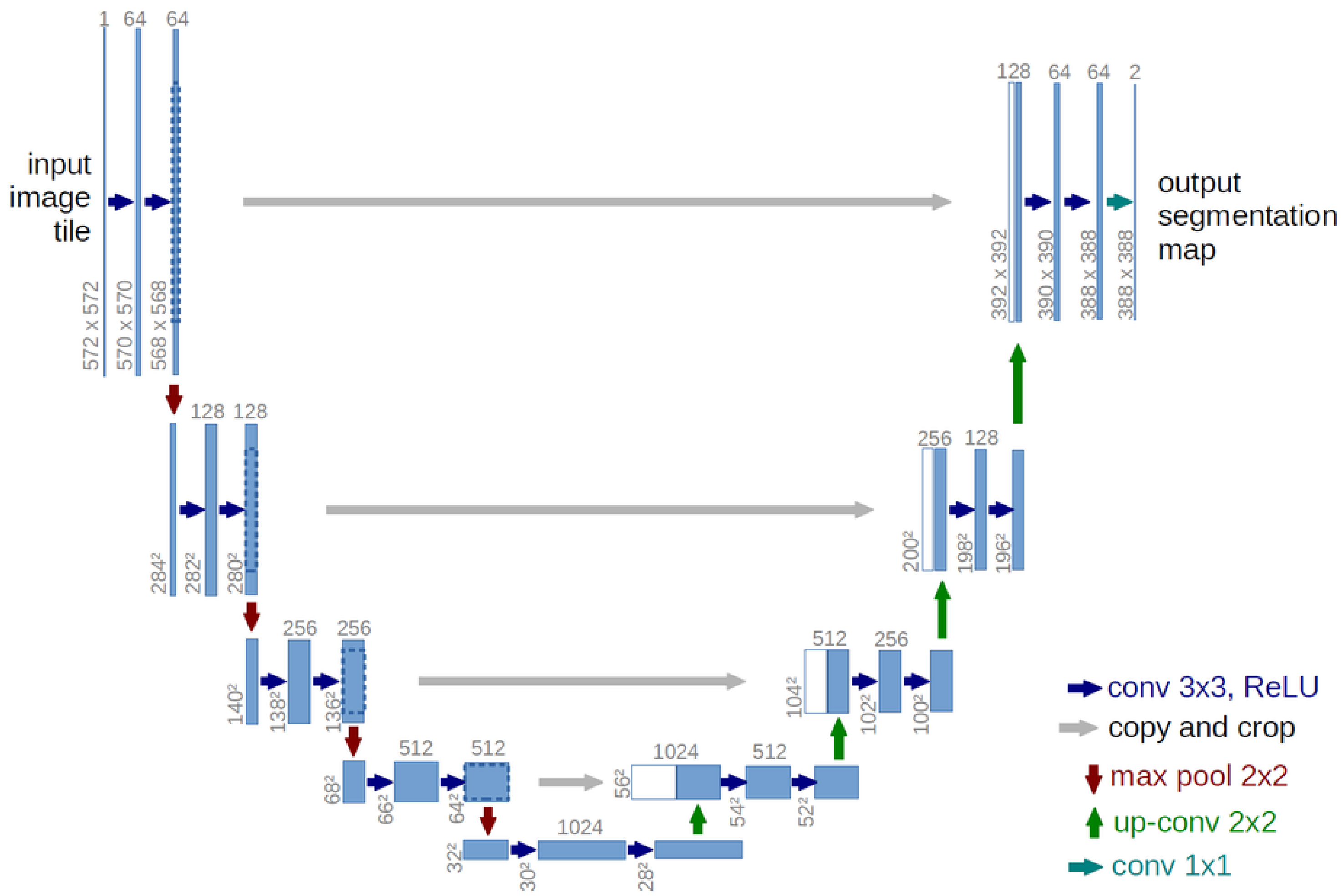
4. We simply add the elements of the over-lapping positions.

5.The resulting output will be the final upsampled feature map having the required spatial dimensions of 5x5.



UNET

The architecture contains two paths. The first path is the contraction path (also called as the encoder) which is used to capture the context in the image. The encoder is just a traditional stack of convolutional and max pooling layers. The second path is the symmetric expanding path (also called as the decoder) which is used to enable precise localization using transposed convolutions.



Dataset

The Dataset used in the project is of 3929 Brain MRI images which are used by radiologists to predict the tumour affected region of the brain. In the mask directory, there are 3929 grayscale images which denote whether the MRI image contains a tumour and if so where.

The grayscale images are also known as masks in which the white regions denote the tumour and the rest of the region is black. These will be used for building a supervised learning model. As the dataset is small for the model to get trained, we used Data Augmentation to increase the number of images.

Data Augmentation

In Data Augmentation, we used the following methods:-

- Resizing the image
- Rotating the image
- Flipping the image horizontally and vertically
- Normalizing the image (varying the pixel intensity in the pictures)

Training

Model is compiled with Adam optimizer and we use binary cross-entropy loss function since there are only two classes (tumour and no tumour).
We used PyTorch to implement:

1. Learning rate decay if the validation loss does not improve for 5 continuous epochs.
2. Early stopping if the validation loss does not improve for 10 continuous epochs.
3. Save the weights only if there is an improvement in validation loss.
4. Learning rate = 10^-4
5. Batch size = 64
6. Performed Data Augmentation to counter overfitting

The model is trained on Tesla T4 GPU and takes less than 2 hours to train.

Flowchart

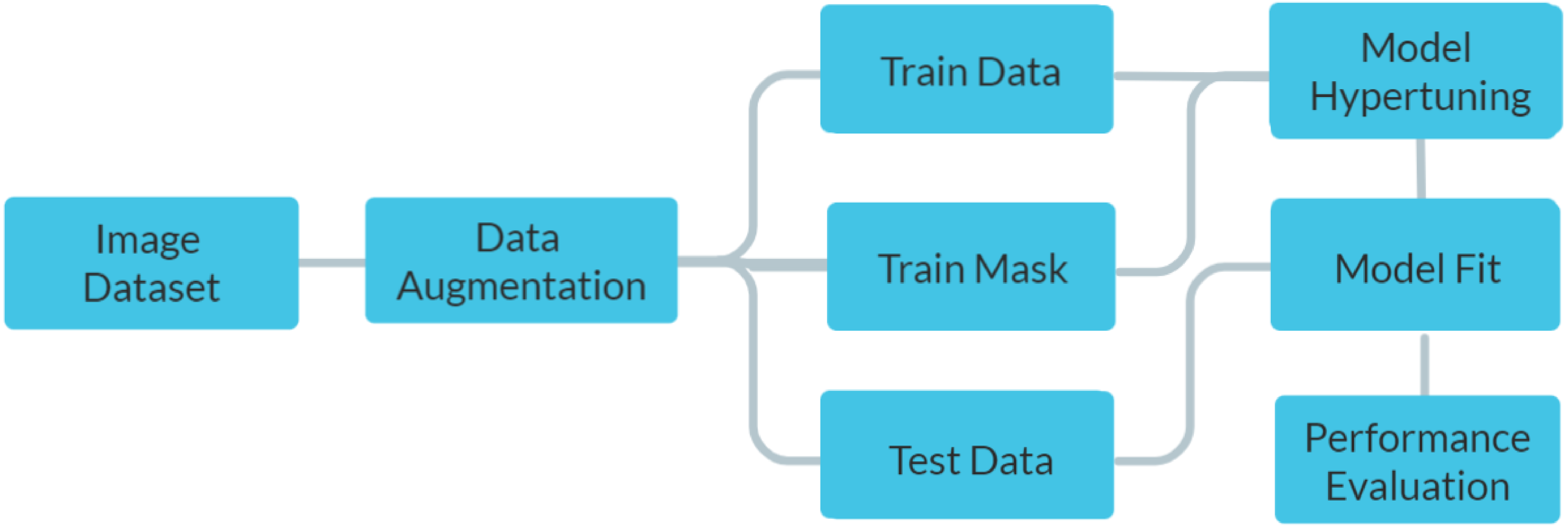
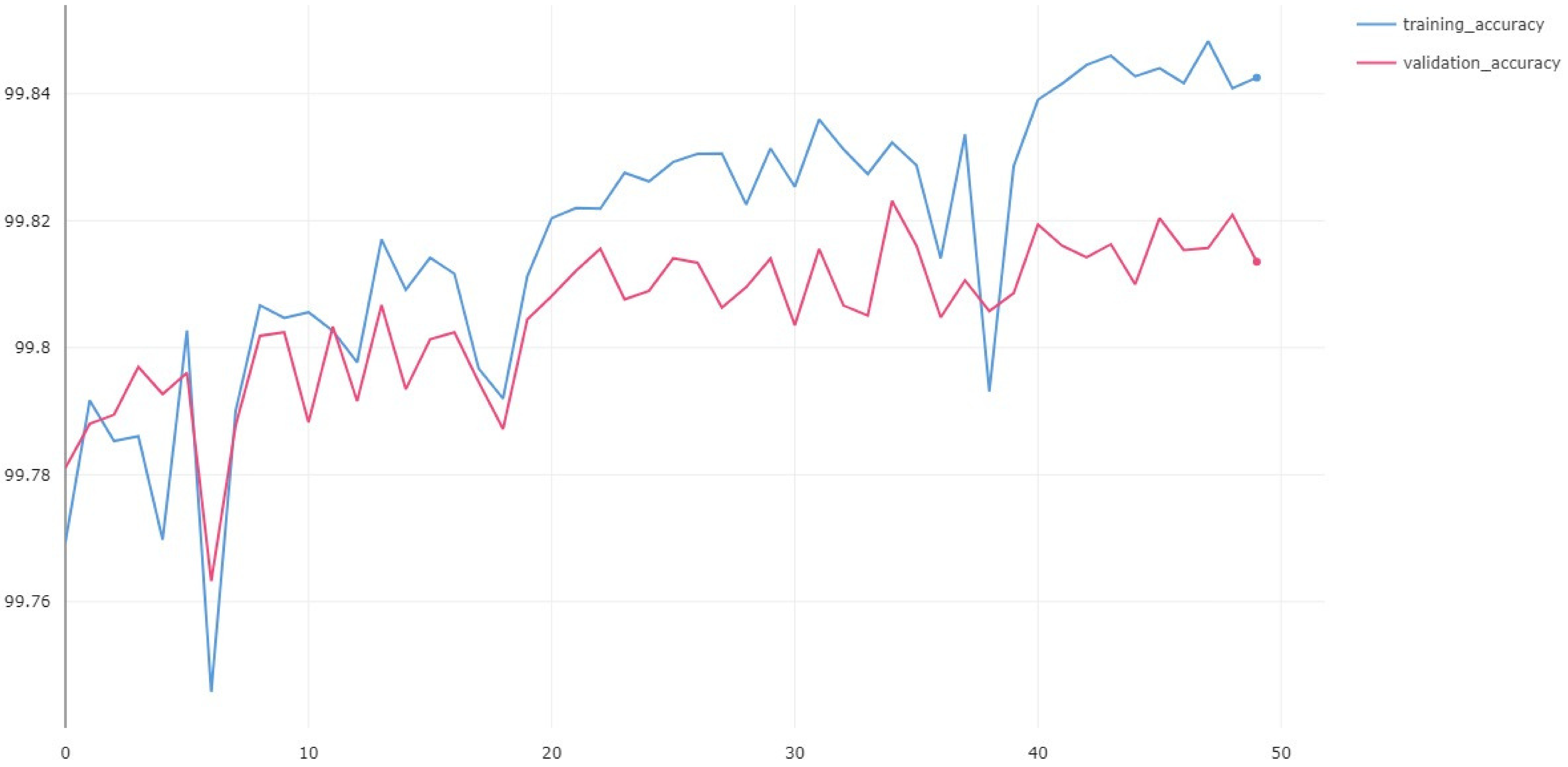
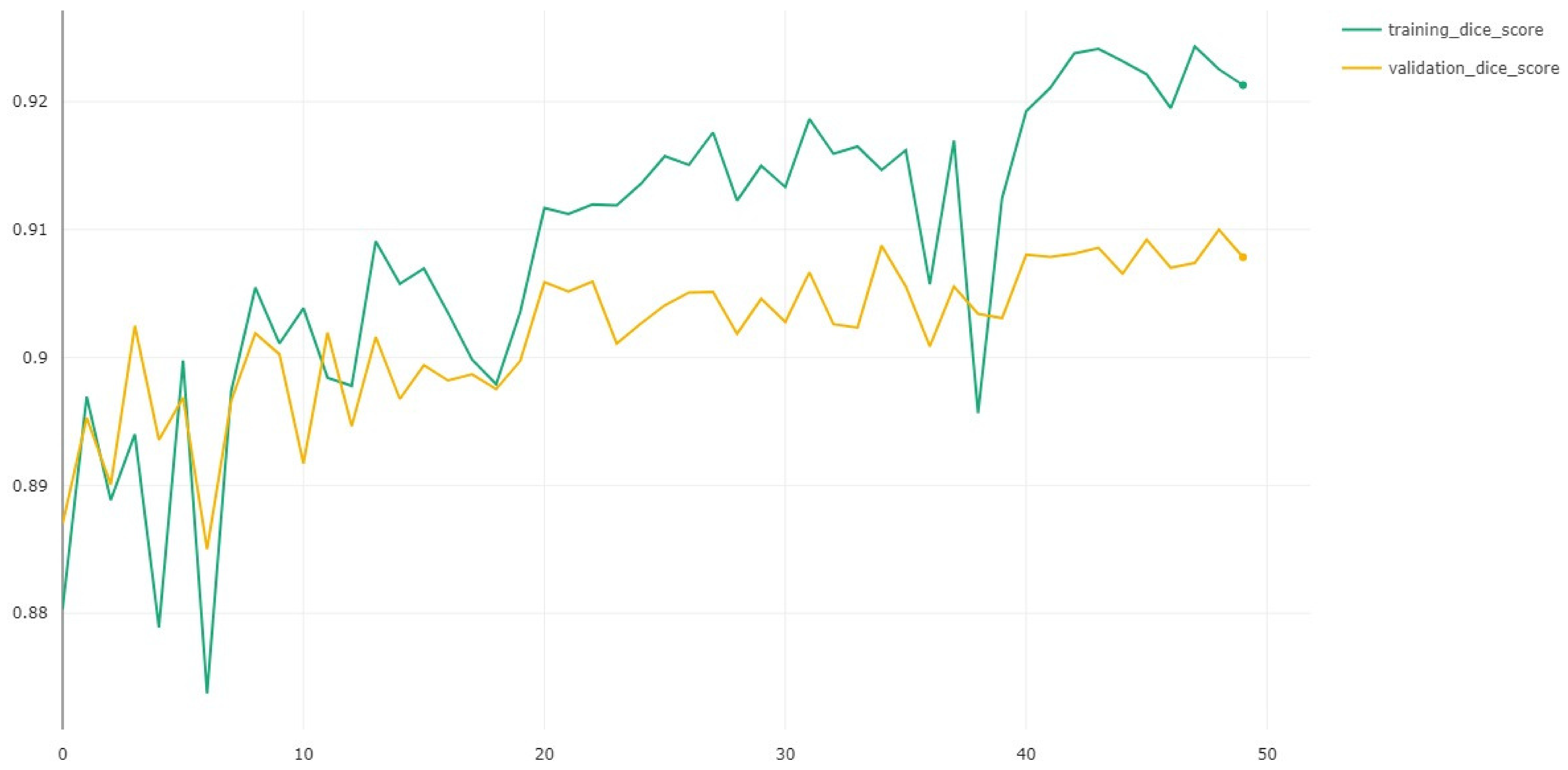


Image Dataset
Data Augmentation
Train Data
Train Mask
Test Data
Model Hypertuning
Model Fit
Performance Evaluation

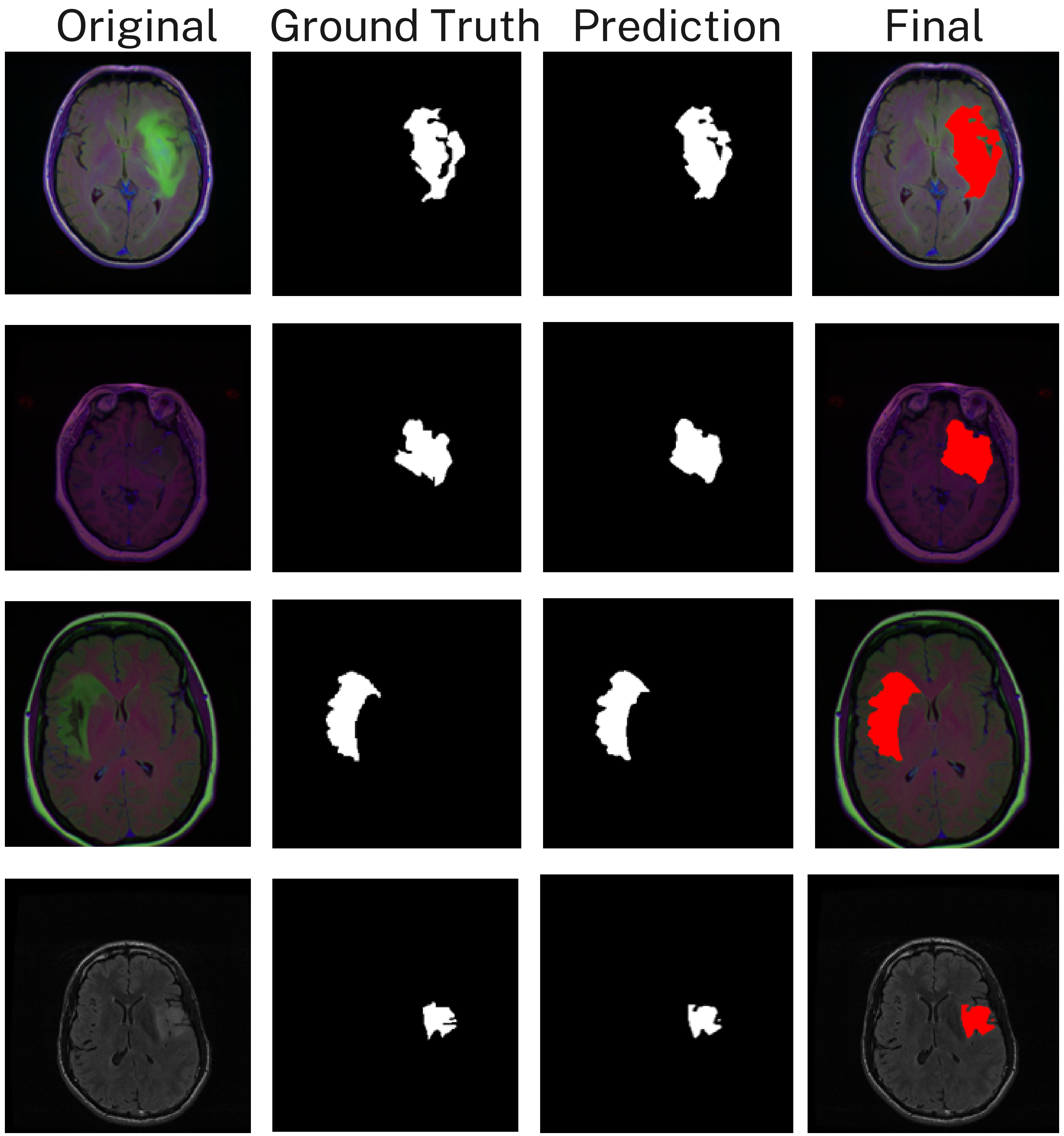


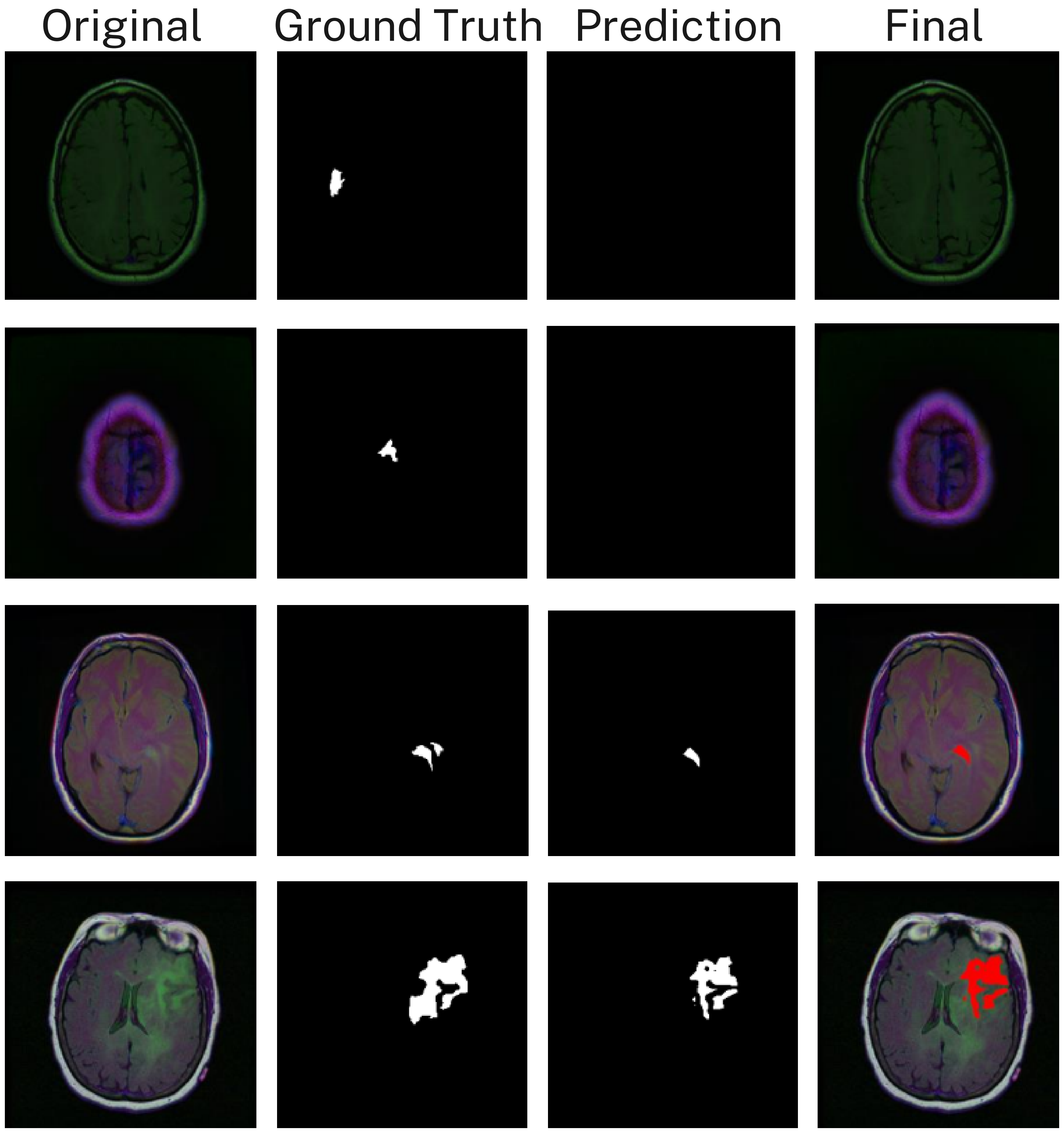
Training and Validation accuracy increases with number of epochs



Training and Validation dice score increases with number of epochs

Results


Original
Ground Truth
Prediction
Final


Original
Ground Truth
Prediction
Final

Conclusion

- In this exploratory project, we have proposed a U-Net model for automatic brain tumor segmentation from MRI. To evaluate the model we have used LGG Segmentation dataset, which contains MRI images of 110 patients.
- The segmentation results under the proposed model tend to approach the ROI target of each brain tumor MRI image quite well.
- Our model have achieved a dice score of **0.91**, which is significantly higher than that of other methods. At the same time, the addition of other modules to the U-Net model can also improve the network performance, such as the double attention U-Net model and residue-intensive module, etc.

References

1. https://www.researchgate.net/publication/342126824_UNet-VGG16_with_transfer_learning_for_MRI-based_brain_tumor_segmentation
2. https://ieeexplore.ieee.org/abstract/document/8964956
3. https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28
4. https://www.emerald.com/insight/content/doi/10.1108/IJICC-02-2020-0013/full/html
5. https://www.sciencedirect.com/science/article/pii/S0925231217318763
6. https://www.narayanahealth.org/brain-tumour
7. https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation
8. Code

Thank you!