



Faculty of Pure and Applied Mathematics

Field of study: Applied Mathematics

Specialty: Data Engineering

Master's Thesis

**DEEP CONVOLUTIONAL NEURAL NETWORK
BASED MEDICAL IMAGE CLASSIFICATION
FOR DISEASE DIAGNOSIS**

Aleksandra Patrycja Winiarska

keywords:

Deep Convolutional Neural Network,
U-Net architecture, medical imaging,
brain tumor, encoder-decoder architecture

short summary:

This thesis explores Deep Convolutional Neural Networks (DCNNs), focusing on the U-Net architecture, for brain tumor detection in medical imaging. It evaluates DCNN models, addresses preprocessing challenges, and enhances interpretability. Both binary and multi-class segmentation tasks are addressed, utilizing original, normalized, and augmented datasets. The thesis highlights the importance of data augmentation in improving model performance. By integrating advanced neural network techniques, this research aims to facilitate early and accurate tumor detection, ultimately contributing to more effective treatment plans and better patient outcomes.

Supervisor	dr Daniel Kucharczyk
	Title/degree/name and surname	grade	signature

*For the purposes of archival thesis qualified to:**

- a) category A (perpetual files)
- b) category BE 50 (subject to expertise after 50 years)

* delete as appropriate

stamp of the faculty

Wrocław, 2024

Contents

Aim and scope of the thesis	7
1 Introduction	9
1.1 Basic knowledge of brain tumors	9
1.2 Types and characteristics of brain tumors	9
1.3 Medical imaging methods for brain tumor detection	10
1.4 Fundamentals of image classification	10
1.5 Role of deep learning in medical image analysis	11
1.6 Challenges and considerations in medical image classification	12
2 The U-Net architecture	15
2.1 Overview of Convolutional Neural Networks	15
2.2 Sliding-window architecture	17
2.3 Introduction to U-Net architecture	18
2.4 Architecture	18
2.5 Advantages of U-Net	19
3 Metrics and methods used for evaluation	21
3.1 Confusion matrix	21
3.2 Precision and recall	22
3.3 Accuracy	22
3.4 Dice coefficient	23
3.5 Cross-entropy loss function	23
4 Analysis of real-world data	25
4.1 Data description	25
4.2 Reference model architecture	27
4.3 Binary classification	29
4.4 Performance comparison table	43
4.5 Multi-class classification	45
4.6 Performance comparison table	56
Conclusion	59
Bibliography	61

List of Figures

2.1	U-Net architecture (example for 32×32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations [40].	19
4.1	Bar plot of the class distribution.	26
4.2	Sample images from the data for each tumor type with masks.	26
4.3	Bar plot of the class distribution in training, test, and validation set.	27
4.4	Encoder-decoder architecture (example for the input image of resolution 128×128 pixels). Each box corresponds to a multi-channel feature map. The architecture consists of sequential layers of Conv $3 \times 3 \rightarrow$ BatchNorm \rightarrow ReLU (orange), followed by MaxPool 2×2 (red), and then UpSampling 2×2 (blue), with a final Softmax layer (violet). The number of channels is denoted on the top of the box. The resolution is denoted on the bottom of the box.	28
4.5	U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for original dataset.	29
4.6	Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for original dataset.	30
4.7	U-Net: Confusion matrix for original dataset (for test dataset).	31
4.8	Encoder-decoder: Confusion matrix for original dataset (for test dataset).	31
4.9	U-Net: Prediction visualization for sample images in each tumor type class for original dataset.	32
4.10	Encoder-decoder: Prediction visualization for sample images in each tumor type class for original dataset.	33
4.11	U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.	34
4.12	Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.	35
4.13	U-Net: Confusion matrix for normalized dataset (for test dataset).	36
4.14	Encoder-decoder: Confusion matrix for normalized dataset (for test dataset).	36
4.15	U-Net: Prediction visualization for sample images in each tumor type class for normalized dataset.	37
4.16	Encoder-decoder: Prediction visualization for sample images in each tumor type class for normalized dataset.	38
4.17	U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.	39

4.18 Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.	40
4.19 U-Net: Confusion matrix for augmented dataset (for test dataset).	41
4.20 Encoder-decoder: Confusion matrix for augmented dataset (for test dataset).	41
4.21 U-Net: Prediction visualization for sample images in each tumor type class for augmented dataset.	42
4.22 Encoder-decoder: Prediction visualization for sample images in each tumor type class for augmented dataset.	43
4.23 U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.	45
4.24 Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.	47
4.25 U-Net: Confusion matrix for normalized dataset (for test dataset).	48
4.26 Encoder-decoder: Confusion matrix for normalized dataset (for test dataset).	48
4.27 U-Net: Prediction visualization for sample images in each tumor type class for normalized dataset.	49
4.28 Encoder-decoder: Prediction visualization for sample images in each tumor type class for normalized dataset.	50
4.29 U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.	51
4.30 Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.	52
4.31 U-Net: Confusion matrix for augmented dataset (for test dataset).	53
4.32 Encoder-decoder: Confusion matrix for augmented dataset (for test dataset).	53
4.33 U-Net: Prediction visualization for sample images in each tumor type class for augmented dataset.	54
4.34 Encoder-decoder: Prediction visualization for sample images in each tumor type class for augmented dataset.	55

List of Tables

3.1	Confusion matrix (contingency table).	21
4.1	Table of metrics for U-Net and encoder-decoder for all considered datasets in binary segmentation.	44
4.2	Table of metrics for U-Net and encoder-decoder for all considered datasets in multi-class segmentation.	56

Aim and scope of the thesis

In recent years, the intersection of machine learning and medical image classification has showcased remarkable potential in revolutionizing disease diagnosis and treatment. Among the myriad visual recognition techniques, Deep Convolutional Neural Networks (DCNNs) have emerged as a powerful tool for analyzing medical images with unprecedented accuracy and efficiency. Within this realm, the detection of brain tumors, a critical area in neurology, stands as a prominent focus due to its implications for patient care and prognosis.

This master's thesis embarks on a comprehensive exploration of DCNNs, particularly focusing on the application of the U-Net architecture, in the detection of brain tumors through medical imaging using MRI images. The overarching goal is to advance the field of medical image classification by developing and optimizing DCNN models tailored to accurately identify and locate brain tumors.

The research scope takes into consideration several crucial dimensions. Firstly, the thesis will undertake a comparative study of various DCNN architectures, evaluating their efficacy in terms of accuracy and reliability within a clinical setting. Special attention will be given to the U-Net architecture, which is renowned for its exceptional performance in semantic segmentation tasks. This attribute makes the architecture perfectly suited for medical imaging analysis.

Moreover, the thesis will dive into preprocessing techniques tailored to enhance the performance of DCNN models, tackling challenges such as contrast enhancement inherent in medical images. Additionally, the handling of imbalanced datasets, a common predicament in medical imaging, will be addressed through innovative strategies to ensure the robustness and generalizability of the models.

Furthermore, the research will encompass the implementation of techniques aimed at understanding the decision-making processes of DCNN models, fostering interpretability and trust in their diagnostic outcomes. This includes methods for visualizing and explaining model predictions, essential for clinical application and acceptance. The research aims to pave the way for the responsible integration of AI technologies into clinical practice, ultimately enhancing patient care and outcomes.

Ultimately, this thesis aspires to contribute significantly to the advancement of diagnostic processes, potentially enabling earlier and more accurate detection of brain tumors. Through in-depth experimentation and analysis, the research aims to harness the revolutionary potential of DCNNs, particularly the U-Net architecture, in the realm of medical imaging for the enhanced quality of healthcare delivery.

Chapter 1 provides essential background information on brain tumors, their types and characteristics, and the basic concepts of medical imaging in their detection. It also covers the fundamentals of image classification, the role of deep learning in medical image analysis, and the challenges faced in medical image classification. Chapter 2 provides an insight into the core architecture used in this study. It starts with an overview of Convolutional Neural Networks (CNNs), explaining their essential components. The chapter further

introduces the sliding-window architecture, followed by an in-depth exploration of the U-Net architecture, highlighting its structure and advantages. Chapter 3 describes the various metrics and methods used to evaluate the performance of the models. It includes discussions on the confusion matrix, precision and recall, accuracy (including balanced accuracy), the Dice coefficient, and the cross-entropy loss function. Chapter 4 presents the practical application of the discussed models. It starts with a detailed description of the data, including the train, validation, and test split, as well as data augmentation techniques. Additionally, the chapter describes the reference model architecture. It also covers the results of binary classification using the original, normalized, and augmented datasets, followed by a performance comparison table. The section on multi-class classification evaluates the performance on normalized and augmented datasets, with another performance comparison table provided. The thesis ends with the Conclusion chapter, which summarizes the findings, highlights the effectiveness of the U-Net model, and suggests directions for further research.

Chapter 1

Introduction

1.1 Basic knowledge of brain tumors

Cancer denotes a group of illnesses characterized by the unregulated growth and spread of abnormal cells within the body. These cells have the potential to form tumors or disrupt the normal functioning of organs and tissues. Cancer can originate in any bodily region and can affect individuals of all age groups, though it becomes more prevalent with advancing age.

A brain tumor refers to an abnormal growth or cluster of irregular brain cells. Brain tumors come in various types, with some being noncancerous (benign) and others malignant (cancerous). Magnetic Resonance Imaging (MRI) has emerged as the predominant non-invasive technique for diagnosing brain tumors in recent decades, owing to its superior soft tissue contrast. It is said that every year around 150 000 patients with cancer are affected by brain tumor. The available treatments for cancer differ depending on the specific type and stage of the illness. These include, for example, surgical procedures, chemotherapy, and radiation therapy. In recent times, remarkable advancements in cancer research and treatment have emerged, resulting in enhanced survival rates and quality of life for numerous individuals diagnosed with cancer.

The symptoms associated with brain cancer can fluctuate based on the tumor's size and location. Among the most common symptoms are: headaches, problems with memory, vision or hearing, difficulties with balance and coordination. The prognosis for brain cancer is largely dependent on various factors, including the type of tumor, its positioning within the brain, and the overall health condition of the patient. Certain brain tumors pose treatment challenges due to their specific location and the potential impact on essential brain functions [5, 10].

1.2 Types and characteristics of brain tumors

There are over 120 distinct types of brain tumors, distinguished by their location and the cell types comprising them. Brain cancer may originate from brain tissue itself, termed primary brain tumors, or may stem from cancer spreading from other body regions, known as metastatic brain tumors.

Primary brain tumors are categorized according to the cells of origin and their specific location within the brain. The secondary brain tumors are much more common than primary tumors [5]. They are categorized by where they originate, later spreading in the brain. Some types of secondary brain tumors are:

- Gliomas
 - tumors that develop from glial cells, which support and nourish nerve cells [5, 49].
- Meningiomas
 - tumors that form in the meninges, the protective membranes surrounding the brain and spinal cord [5, 34].
- Pituitary adenomas
 - tumors that develop in the pituitary gland, which controls various hormonal functions in the body [5, 30].

1.3 Medical imaging methods for brain tumor detection

The diagnosis of brain cancer typically requires a comprehensive series of medical assessments and tests conducted by a multidisciplinary team of healthcare professionals, including neurologists, neurosurgeons, oncologists, and radiologists.

In the process of diagnosis, imaging tests are extremely important. They allow to accurately determine the location and often even the type of tumor.

Common imaging tests employed include:

- Magnetic Resonance Imaging (MRI)
 - this diagnostic tool generates highly detailed images of the brain, enabling visualization of the tumor's size, location, and characteristics [1, 5, 25, 41].
- Computed Tomography (CT) Scan
 - utilizing X-rays, a CT scan produces cross-sectional images of the brain, facilitating the detection of abnormalities [5, 17].

After diagnosing the patient with brain cancer, the medical team embarks on identifying the tumor's type, assessing its grade, and determining its stage. This data is essential in formulating an optimal treatment plan [5].

1.4 Fundamentals of image classification

Image classification is a fundamental task in computer vision (the academic field dedicated to providing computers with the ability to interpret and understand visual information from digital images or videos), aiming to categorize images into predefined classes or categories based on their visual content. Image categorization serves a valuable purpose in numerous domains. Among the wide range of tasks within computer vision, image categorization serves as a crucial component for addressing a broad spectrum of challenges, including:

- Object detection
 - classify picture features and identify their locations using bounding boxes [26, 29, 45, 48].
- Object (semantic) segmentation
 - identify individual objects in an image by analyzing the pixels that represent them, rather than simply outline them with bounding boxes as in object detection [26, 29, 45, 48].
- Instance segmentation
 - differentiate between instances of the same class [26, 29, 45, 48].

1.5 Role of deep learning in medical image analysis

Machine learning algorithms have the potential to revolutionize various aspects of medicine such as clinical decision-making, thereby revolutionizing the field of healthcare. The recent success of machine learning algorithms and artificial intelligence in computer vision tasks is particularly relevant today, given the increasing digitization of medical records [21, 38, 46, 50]. AI and machine learning techniques, have shown promising applications in various tasks related to medical image analysis e.g., in brain [7, 12, 35], chest [22, 37, 42], breast [14, 20], cardiac [3, 47] or musculoskeletal [8, 27, 44] imaging. Currently, medical images have become a crucial component of a patient's Electronic Health Record (EHR) and are typically interpreted by human radiologists. However, radiologists face limitations such as speed, fatigue, and varying levels of experience. Training a qualified radiologist is a time-consuming and highly expensive process, which leads some healthcare systems to transfer radiology results to countries like India through tele-radiology. Delays or errors in diagnosis can have catastrophic effects on patients, emphasizing the importance of accurate and efficient automated medical image analysis through machine learning algorithms. It is anticipated that medical image classification will serve as the initial platform for patient interaction with practical artificial intelligence systems.

Medical imaging encompasses various forms of imaging, with their usage steadily increasing over time. The modalities for digital medical images include ultrasound, X-ray, CT scans, MRI scans, PET scans, retinal photography, histology slides, and dermoscopy images, each with its unique characteristics and data generation costs. Factors like processor and memory limitations affect the technical aspects of handling such information and designing algorithms [26, 28, 29].

The application of machine learning in medical image analysis involves various architectures, with Convolutional Neural Networks (CNNs) being the most common.

Machine learning methods can be broadly classified into:

1. Supervised learning – the model is trained on labeled data, where each input image is associated with a corresponding label indicating the class or category [26, 28, 29, 32].

Examples of techniques under supervised learning are:

- Convolution Neural Networks,
- Transfer Learning with CNNs,
- Recurrent Neural Networks (RNNs).

2. Unsupervised learning – the model is trained on unlabeled data, without specified input-output pairs. Instead, the model identifies patterns and structures within the data to uncover meaningful information [26, 28, 29, 32]. Examples of techniques under unsupervised learning are:

- autoencoders,
- clustering,
- dimensionality reduction.

While semi-supervised and reinforcement learning methods also hold potential in medical image analysis, supervised and unsupervised approaches currently dominate the field due to their effectiveness in leveraging labeled and unlabeled data, respectively [26, 28, 29].

1.6 Challenges and considerations in medical image classification

Medical image classification presents several challenges and considerations that need to be addressed for effective analysis and interpretation:

1. Data quality and variability

- Medical images can vary significantly in quality, resolution, and noise levels due to differences in imaging techniques, equipment, and patient conditions. Ensuring data consistency and quality is crucial for training accurate classification models [4, 39].

2. Class imbalance

- In medical datasets, certain classes or categories of diseases may be underrepresented compared to others, leading to class imbalance. This imbalance can affect the performance of classification models, making it challenging to accurately classify minority classes [4, 33].

3. Interpretability and explainability

- Medical image classification models often operate as black-box systems, making it difficult to understand how they arrive at their predictions. Ensuring interpretability and explainability of classification results is essential for gaining trust from healthcare professionals and improving clinical decision-making [15, 43].

4. Robustness to variability

- Medical images may vary in patient demographics, anatomical differences, imaging methodologies, and disease characteristics. Classification models need to be able to adapt to such variability to generalize well across diverse patient populations and imaging conditions [4, 39].

5. Ethical and legal considerations

- Medical image classification models may impact patient outcomes and treatment decisions, raising ethical considerations regarding patient privacy, consent, and algorithm bias. Ensuring compliance with regulations and ethical guidelines is extremely important [33].

6. Clinical validation and integration

- The clinical utility of medical image classification models needs to be consistently validated through clinical studies and trials. Integrating these models into clinical procedures and decision support systems requires collaboration between computer scientists, clinicians, and healthcare administrators [2, 39].

Addressing these challenges and considerations is essential for the successful implementation of medical image classification models in real-world clinical conditions, leading to improved patient care and outcomes [2, 4, 15, 26, 33, 39].

Chapter 2

The U-Net architecture

2.1 Overview of Convolutional Neural Networks

In recent years, Deep Convolutional Networks have demonstrated remarkable advancements, surpassing the state-of-the-art performance in various visual recognition tasks. The current focus of extensive research in medical image analysis is the ability of CNNs to preserve spatial relationships during image processing. Spatial relationships hold significant importance in radiology, such as delineating the interface between bone and muscle or distinguishing between normal and cancerous tissue within the lungs. A CNN begins by taking an input image consisting of raw pixel values and applies a series of operations, including convolutional layers, Rectified Linear Unit (ReLU) layers, and pooling layers. These operations collectively transform the input image, ultimately leading to a fully connected layer that outputs class scores or probabilities, enabling the classification of the input into the class with the highest probability [26, 28, 29].

Convolution layer

Convolution, as defined in image analysis, involves the operation of two functions: the input values (e.g., pixel values) at a position in the image and a filter (or kernel), both represented as arrays of numbers. By computing the dot product between these functions, an output is generated. The filter is then systematically shifted across the image based on the specified stride length, repeating the computation process until the entire image is covered. This process results in the creation of a feature map, indicating areas where the filter detects specific features, such as lines, edges, or curves. In a CNN, these low-level features gradually build up to higher-level features, as they become inputs for subsequent layers in the CNN architecture.

Convolution relies on three key concepts for efficient machine learning:

- sparse connections,
- parameter sharing,
- equivariant representation.

Unlike some neural networks where every input neuron is connected to every output neuron in the subsequent layer, CNN neurons have sparse connections, meaning only certain inputs are linked to the next layer. This localized receptive field allows meaningful

features to be learned gradually, significantly reducing the number of weights to be calculated and improving the algorithm's efficiency. Additionally, by sharing each filter's fixed weights across different positions of the entire image, CNNs reduce memory storage requirements — a concept known as parameter sharing. Parameter sharing leads to equivariant representation, where input translations correspond to feature map translations [26, 28, 29].

According to [26], the convolution operation is represented by the $*$ symbol, with the output (or feature map) denoted as $s(t)$ when the input ($I(t)$) is convolved with a filter or kernel ($K(a)$):

$$s(t) = (I * K)(t). \quad (2.1)$$

When t is restricted to integer values, the discretized convolution is expressed as:

$$s(t) = \sum_a I(a)K(t - a). \quad (2.2)$$

This formula is applicable for one-dimensional convolutional operations. For a two-dimensional convolution operation involving an input $I(m, n)$ and a kernel $K(a, b)$, the expression becomes:

$$\begin{aligned} s(t) &= \sum_a \sum_b I(a, b)K(m - a, n - b) \\ &= \sum_a \sum_b I(m - a, n - b)K(a, b). \end{aligned} \quad (2.3)$$

Softmax

Softmax is a widely used activation function in neural networks, particularly in the output layer for classification tasks. It transforms the raw output scores of a network into probability distributions over multiple classes, guaranteeing that the output values range between 0 and 1 and sum up to 1. This allows for intuitive interpretation of the network's predictions as probabilities. Softmax is especially beneficial in multi-class classification scenarios, where it enables the identification of the most likely class among several options. Softmax is defined as

$$f(\mathbf{x})_i = \frac{\exp x_i}{\sum_{j=1}^n \exp x_j}, \quad (2.4)$$

where x_i values are the elements of the input vector \mathbf{x} [23].

Rectified Linear Unit

The Rectified Linear Unit (ReLU) serves as an activation function within neural networks, effectively setting negative input values to zero. This simplification not only speeds up calculations and training but also reduces the vanishing gradient problem. ReLU, is defined as:

$$f(x) = \max(0, x), \quad (2.5)$$

where x represents the input to the neuron [6, 26, 31].

Pooling layer

The pooling layer is used to efficiently reduce the number of parameters to be computed, as well as the dimensions of the image (only width and height). Most people use max-pooling, but there are other options like average pooling and L2-normalization pooling. Max-pooling works by selecting the maximum input value within a filter while neglecting the remaining values, effectively summarizing the most significant activations within a neighborhood. This approach is grounded in the idea that the relative positioning of strongly activated features holds greater significance than their precise locations [6, 26, 31].

Fully connected layer

The final component in a CNN architecture is the fully connected layer, where each neuron from the previous layer is connected with every neuron in this layer. Taking the result from the previous layer, it computes probability scores for classification into distinct classes. Essentially, it examines the composition of the most intensely activated features to determine the likelihood of an image belonging to a specific class. The number of fully connected layers may vary based on the desired level of feature abstraction [6, 26, 31].

2.2 Sliding-window architecture

Despite Convolutional Networks being in existence for some time, their effectiveness was previously constrained by the limited size of available training datasets and the scale of the networks under consideration. CNNs are most commonly used for classification tasks, where an image yields a single class label as output. But many visual tasks, especially in medical images, require marking each pixel with a label.

To address these challenges, Ciresan et al. [13] presented a novel approach in which they trained a network in a sliding-window architecture to predict the class label of each pixel by providing a local region (patch) surrounding that pixel as input.

Advantages and limitations

This methodology offers dual benefits:

- the network can effectively localize points of interest,
- the training dataset (in terms of patches) substantially exceeds the number of training images.

This innovative approach led to the network's resounding victory in the EM segmentation challenge at ISBI 2012, underscoring its superiority in biomedical image processing tasks.

There are two significant drawbacks to this approach:

1. Slow operation:
 - network runs separately for each patch,
 - overlapping patches lead to redundancy.

2. Trade-off between localization accuracy and context utilization:

- larger patches necessitate a higher number of max-pooling layers, decreasing localization accuracy,
- smaller patches offer limited contextual information to the network [40].

2.3 Introduction to U-Net architecture

To overcome limitations encountered in sliding-window CNNs, a new network was developed. U-Net is a Deep Convolutional Neural Network architecture for segmentation of images. It was first introduced in [40]. This innovative architecture extends a traditional contracting network by integrating successive layers, where pooling operations are replaced by upsampling operators. Consequently, these layers enhance output resolution, crucial for accurate localization. To achieve this, high-resolution features from the contracting path are merged with the upsampled output, facilitating precise localization. This results in a symmetric U-shaped architecture.

To address segmentation challenges in large images, an overlap-tile strategy is employed, allowing seamless segmentation while avoiding GPU memory limitations. Furthermore, extensive data augmentation techniques, such as elastic deformations, can be applied due to limited training data, enhancing the network's ability to learn deformity invariance [40].

2.4 Architecture

Figure 2.1 illustrates the architecture of the network. The network architecture is composed of a contracting path (left side) and an expansive path (right side). The contracting path is constructed according to a standard convolutional network architecture, featuring repetitive application of two 3×3 convolutions (unpadded convolutions), each followed by a Rectified Linear Unit and a 2×2 max pooling operation with a stride of 2 for downsampling. The number of feature channels is doubled at each downsampling step.

On the other hand, the expansive path involves upsampling of the feature map, followed by a 2×2 convolution ("up-convolution") which reduces the number of feature channels, concatenation with the correspondingly cropped feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU function. A 1×1 convolution is applied in the final layer to map each 64-component feature vector to the desired number of classes. In total, the network comprises 23 convolutional layers. It is crucial to select the input tile size, ensuring all 2×2 max-pooling operations are applied to a layer with an even x- and y-size [40].

In the following work, we propose a modification to the network architecture. Following each convolution operation, we incorporate padding to prevent pixel loss at the edges of the image. Additionally, we employ the softmax activation function in the final layer of the network. The initial weights are drawn from a uniform distribution within $[-L, L]$ where $L = \sqrt{6/(n + N)}$, n is the number of input units in the weight tensor and N is the number of output units in the weight tensor. In terms of the loss function, the chosen method is the cross-entropy loss function. During the training process of the network, a batch size equal to 8 was used.

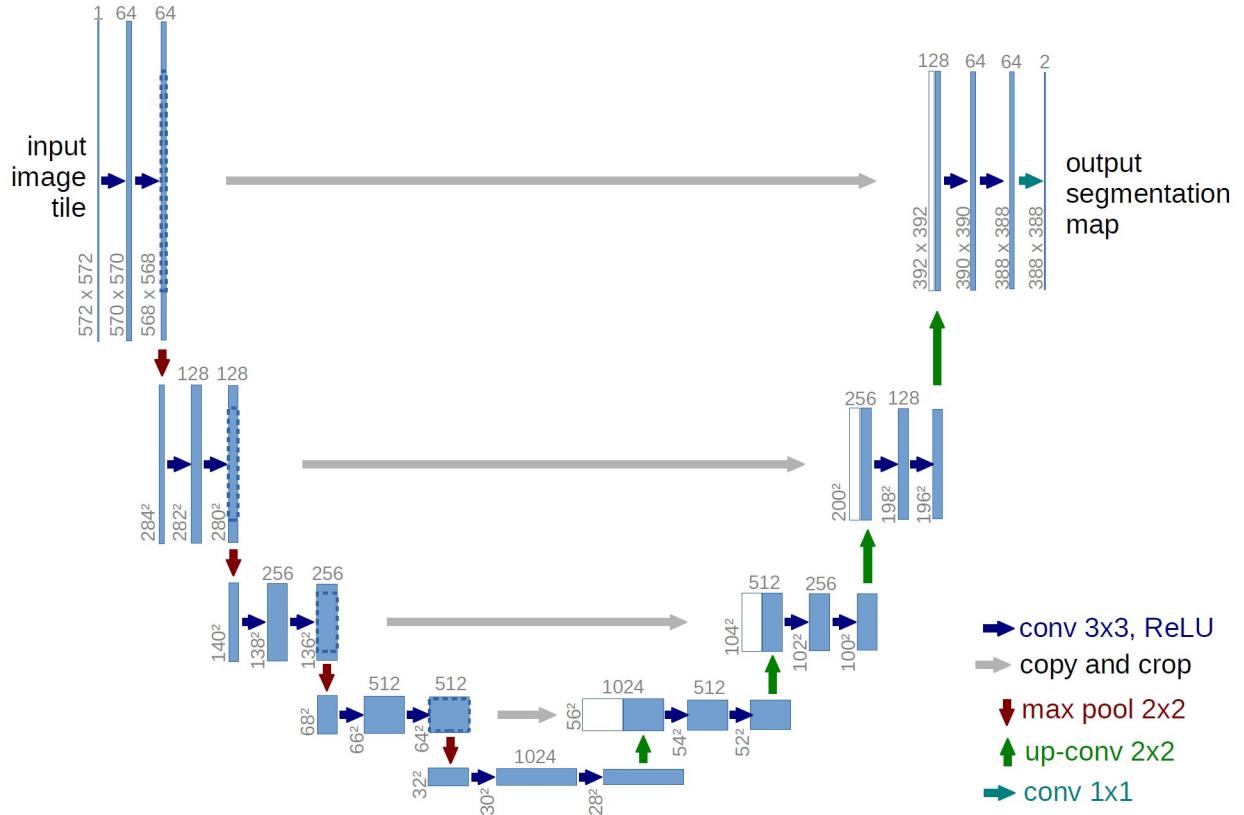


Figure 2.1: U-Net architecture (example for 32×32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations [40].

2.5 Advantages of U-Net

The U-Net architecture offers several key advantages that make it a powerful tool in the field of medical image segmentation. These advantages include:

- Versatility in image inputs:
 - capable of processing various types of image inputs, including grayscale, color and multi-channel images or even 3D volumes,
 - can handle inputs of varying sizes and aspect ratios.
- High precision:
 - provides high precision in segmentation tasks, ensuring accurate identification of objects or regions of interest in images.

- Efficient speed:
 - offers efficient processing speed, enabling rapid analysis and segmentation of images, thereby enhancing workflow in medical imaging tasks.
- Reduced dependence on large datasets:
 - does not require big datasets for training, making it feasible for applications where acquiring large datasets is difficult to achieve [40].

Chapter 3

Metrics and methods used for evaluation

In the realm of classification tasks, the most frequent scenario often involves binary classification, where the objective is to distinguish between two classes. However, there are instances where the classification task extends beyond two classes, leading to what is termed "multi-class classification". In such scenarios, the challenge shifts to accurately categorizing input data into multiple distinct classes, presenting a more complex evaluation structure that demands appropriate methodologies and metrics.

Regardless of the specific approach, the common technique lies in relying on available data (\mathbf{X} variables) to generate optimal predictions ($\hat{\mathbf{Y}}$) for the outcome variable (\mathbf{Y}). In the context of multi-class classification, both the response variable (\mathbf{Y}) and the prediction ($\hat{\mathbf{Y}}$) are discrete random variables, taking on values within a defined set representing different classes. The algorithm calculates the probability of each unit belonging to a particular class, after which a simple classification rule is applied to assign a single class to each observation, often based on the highest probability. While in two-class classification, a threshold is typically used to determine the predicted class for each unit based on the model-assigned probabilities, in multi-class scenarios, various approaches exist [24].

3.1 Confusion matrix

First, we consider the problem of classification into two classes. Each case is mapped to one element from the set P, N (positive and negative class labels, respectively). To distinguish the true class from the predicted class, we use T, F labels (true and false prediction, respectively) for the prediction classes obtained using the model.

Knowing the true and predicted classes, we have four possibilities, which can be written in the form of a confusion matrix (contingency table), presented in the form of a Table 3.1 [18].

Table 3.1: Confusion matrix (contingency table).

		Actual class	
		Positive	Negative
Predicted class	Positive	TP	FP
	Negative	FN	TN

Symbols used in the Table 3.1 are:

- TP – a true positive case; a diseased patient classified correctly,
- FP – a false positive case; a healthy patient classified incorrectly,
- TN – a true negative case; a healthy patient classified correctly,
- FN – a false negative case; a healthy patient classified incorrectly.

3.2 Precision and recall

According to [24], precision is a metric used to evaluate the performance of a classification model, calculated as the ratio of true positive elements to the total number of positively predicted units (the sum of the predicted positives column). The mathematical formula for precision is

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (3.1)$$

Precision measures the accuracy of the model in identifying true positives among all instances predicted as positive. It indicates the reliability of the model's positive predictions, reflecting how much confidence we can place in the model when it labels an individual as positive.

Recall represents the ratio of true positive instances to the total number of actual positive instances (sum of the actual positives). The recall evaluates the model's ability to correctly identify all positive instances within the dataset. In essence, it quantifies how well the model can capture all positive units. The formula for recall is

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (3.2)$$

3.3 Accuracy

The accuracy formula takes into account the total of correctly classified instances (true positives and true negatives) divided by the total number of instances in the dataset (sum of all entries in the confusion matrix). In essence, accuracy represents the likelihood that a randomly selected unit will be correctly predicted by the model. The formula for accuracy is

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3.3)$$

Accuracy provides an overall assessment of how accurately the model predicts across the entire dataset, with each individual unit contributing equally to the metric. However, when considering classes instead of individual instances, highly populated classes carry more weight, potentially skewing the accuracy measure. Therefore, accuracy is suitable when prioritizing overall individual predictions rather than class distribution or specific indicators. However, accuracy can obscure classification errors, especially in imbalanced datasets where one class dominates. Despite its simplicity and intuitiveness, accuracy lacks the ability to identify problematic classes where the model performs poorly [24].

Balanced accuracy

Balanced accuracy, a widely used metric in both binary and multi-class classification, derives from the confusion matrix. It serves as a measure of how likely an individual from any class is to be correctly classified. This metric is particularly valuable when dealing with imbalanced datasets, where class sizes vary significantly. In such cases, balanced accuracy offers a more reliable evaluation compared to standard accuracy, as it accounts for the varying class distributions.

For binary classification tasks, balanced accuracy is defined as

$$\text{Balanced Accuracy} = \frac{\frac{TP}{TP+FN} + \frac{TN}{FP+TN}}{2} = \frac{\text{Recall} + \text{Specificity}}{2}. \quad (3.4)$$

On the other hand, for multi-class classification tasks, balanced accuracy is defined as an average of all recalls

$$\text{Balanced Accuracy} = \frac{\sum_{n=1}^K \text{Recall}_n}{K}, \quad (3.5)$$

where K is the number of classes [24].

3.4 Dice coefficient

The Dice score, also known as the Dice coefficient or Dice similarity coefficient, serves as a common metric for evaluating the agreement between two binary segmentations. It can be formulated as

$$D = \frac{2TP}{2TP + FP + FN}. \quad (3.6)$$

A relatively simple approach to deriving a unified criterion from the binary Dice score for multi-class segmentation is by calculating the mean Dice score. The mathematical formula is as follows

$$D_{\text{mean}}(y, \hat{y}) = \frac{1}{|\mathbf{L}|} \sum_{l \in \mathbf{L}} \frac{2 \sum_i y_l^i \hat{y}_l^i}{\sum_i (y_l^i + \hat{y}_l^i)}, \quad (3.7)$$

where \mathbf{L} is the label space, $\{y\}_{l \in \mathbf{L}, i \in \mathbf{X}}$, $\{\hat{y}\}_{l \in \mathbf{L}, i \in \mathbf{X}}$ are the set label probability vectors for all voxels for the ground truth and the prediction [19]. In order to avoid problems with division by 0, a small constant ε can be added to both nominator and denominator.

3.5 Cross-entropy loss function

To address the vanishing gradient problem caused by the combination of the Mean Squared Error loss function and the sigmoid function, alternative loss functions such as cross-entropy can be employed. The cross-entropy loss function is defined as

$$L = -\frac{1}{m} \sum_{x \in D} \sum_i y_i \ln(\hat{y}_i), \quad (3.8)$$

where m represents the total number of samples in the training set D , i denotes the class and \hat{y}_i is the predicted value for the i^{th} sample [9].

Chapter 4

Analysis of real-world data

Analyzing real-world data is crucial for validating the practical applicability and robustness of machine learning models. By applying the trained models to a real-world datasets, we can assess their performance under realistic conditions, which often involve more variability and noise than controlled experimental setups. This analysis helps in understanding the models' strengths and limitations, and provides insights into how well they can generalize to unseen data. Evaluating real-world data ensures that the models are not only theoretically sound but also effective in practical applications, paving the way for future improvements and research directions.

4.1 Data description

The brain tumor dataset used in this study consists of 3064 MRI images obtained from 233 patients and is available on kaggle.com. The dataset encompasses three distinct types of brain tumors:

- meningioma (708 slices) – represented as class (1),
- glioma (1426 slices) – represented as class (2),
- pituitary tumor (930 slices) – represented as class (3).

The distribution of class division is presented on Figure 4.1. As one can see, the dataset is not balanced in terms of types of brain tumors.

Figure 4.2 presents sample images from the data, for each tumor type, with masks applied. The masks provide visual guidance on the precise boundaries of the tumor regions within the brain anatomy. This visualization is helpful in understanding the characteristics and distribution of tumors across different types, which can be used to further analyze and interpret the medical images.

The Figure also illustrates the inherent class imbalance within the dataset, as each tumor occupies a relatively small area compared to the entire image. Additionally, variations in image brightness and contrast are visible, with some images appearing lighter or darker than others. These factors underscore the complexity and variability present in medical imaging data, highlighting the importance of robust classification algorithms capable of handling such challenges.

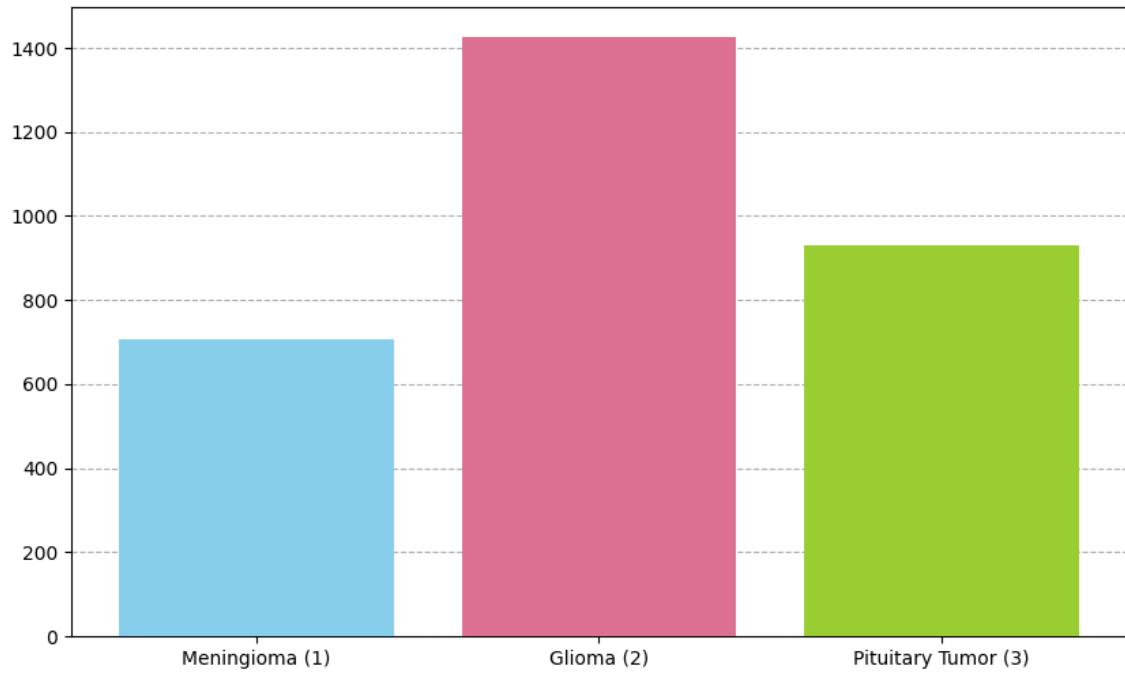


Figure 4.1: Bar plot of the class distribution.

Brain MRI Images for Brain Tumor Detection BrainTumorRetrieval Dataset

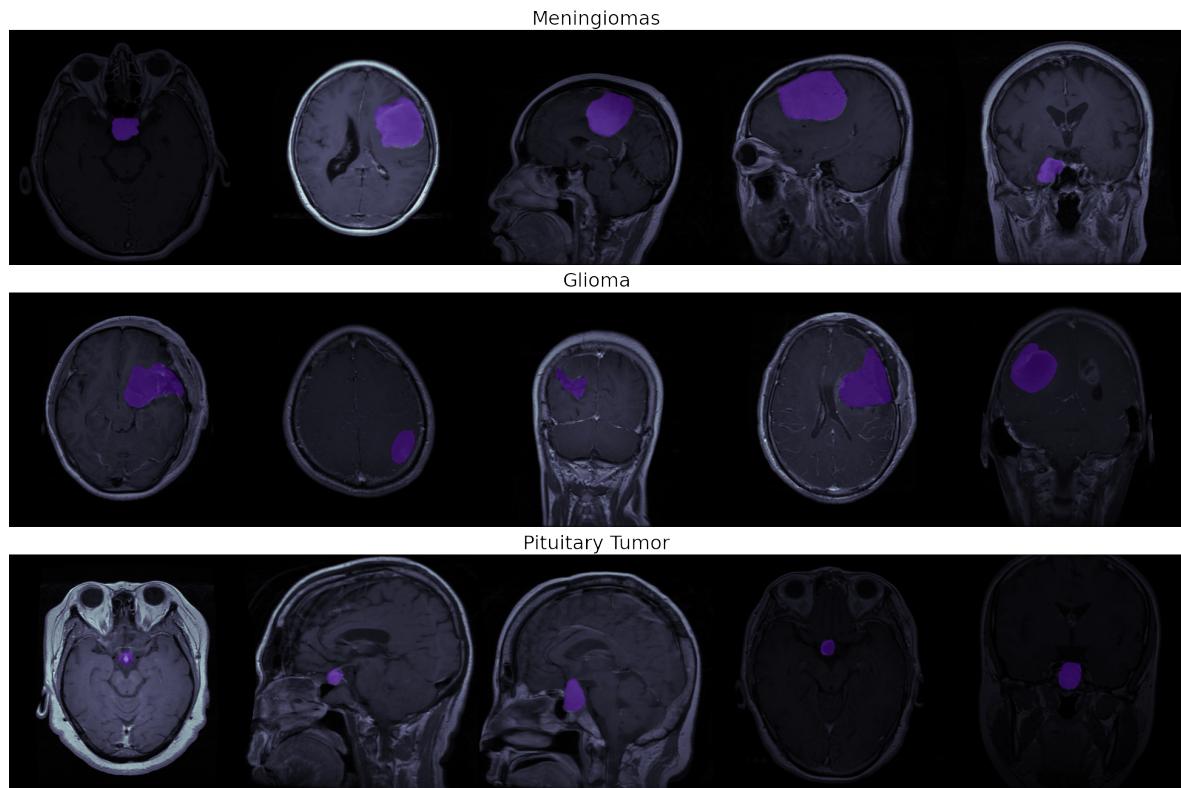


Figure 4.2: Sample images from the data for each tumor type with masks.

Train, validation, and test split

The dataset is divided into three subsets: training, validation, and test sets, maintaining the class distribution across each subset and illustrated on Figure 4.3. Each image has a resolution of 128×128 pixels and is stored in grayscale format. For training, 80% of the dataset is utilized, resulting in 2451 images. From the remaining 20%, 75% is allocated to the test set, accounting for 459 images. The validation set consists of the remaining 25% of the 20%, comprising 154 images. This division ensures adequate representation across the training, validation, and test sets for robust model evaluation.

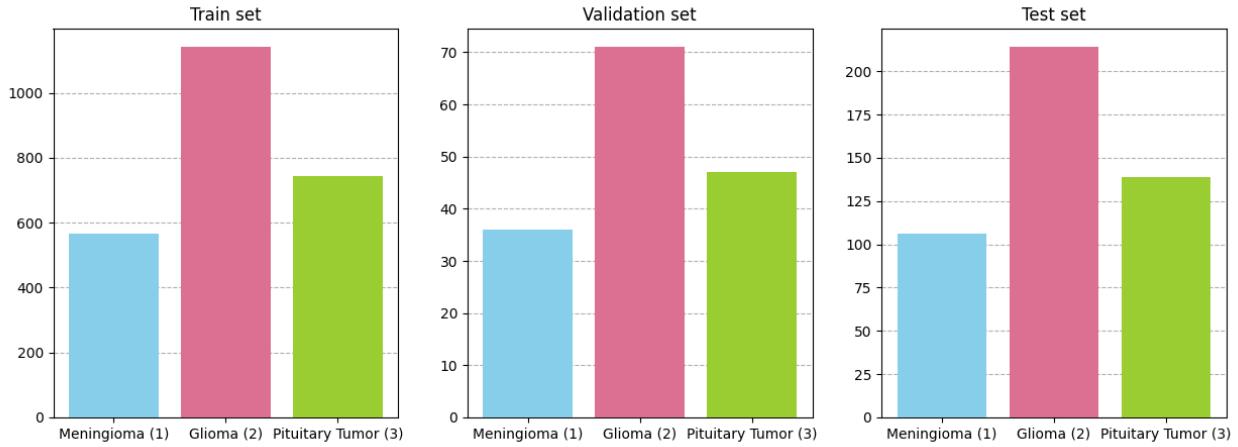


Figure 4.3: Bar plot of the class distribution in training, test, and validation set.

Augmentation

The augmented dataset is created by initially normalizing the original dataset. Subsequently, each image is rotated by multiples of 90 degrees (up to 4 rotations, resulting in the original image) and flipped horizontally, vertically, or both. This process generates twice as many images as in the original dataset. Following this, random adjustments of brightness and contrast are applied to each image from the augmented dataset. Consequently, this yields four times more images than in the original dataset. The percentage of images in the test, train, and validation sets remains consistent.

4.2 Reference model architecture

In this study, we will use an encoder-decoder network for comparison with U-Net. This network is structured to handle image segmentation tasks, transforming input images into segmented output masks that classify each pixel into one of several classes [51, 11, 16, 36].

Figure 4.4 illustrates the architecture of the encoder-decoder network that will serve as reference network in this paper. The network begins with an input layer that receives images of a specified shape. In the first encoding block, a 2D convolutional layer with 64 filters of size 3×3 is applied to the input, followed by batch normalization and a ReLU activation function. This is followed by a max pooling layer with a pool size of 2×2 , which downsamples the feature map. The second encoding block employs similar operations but with 128 filters in the convolutional layer. This pattern continues with the third and fourth

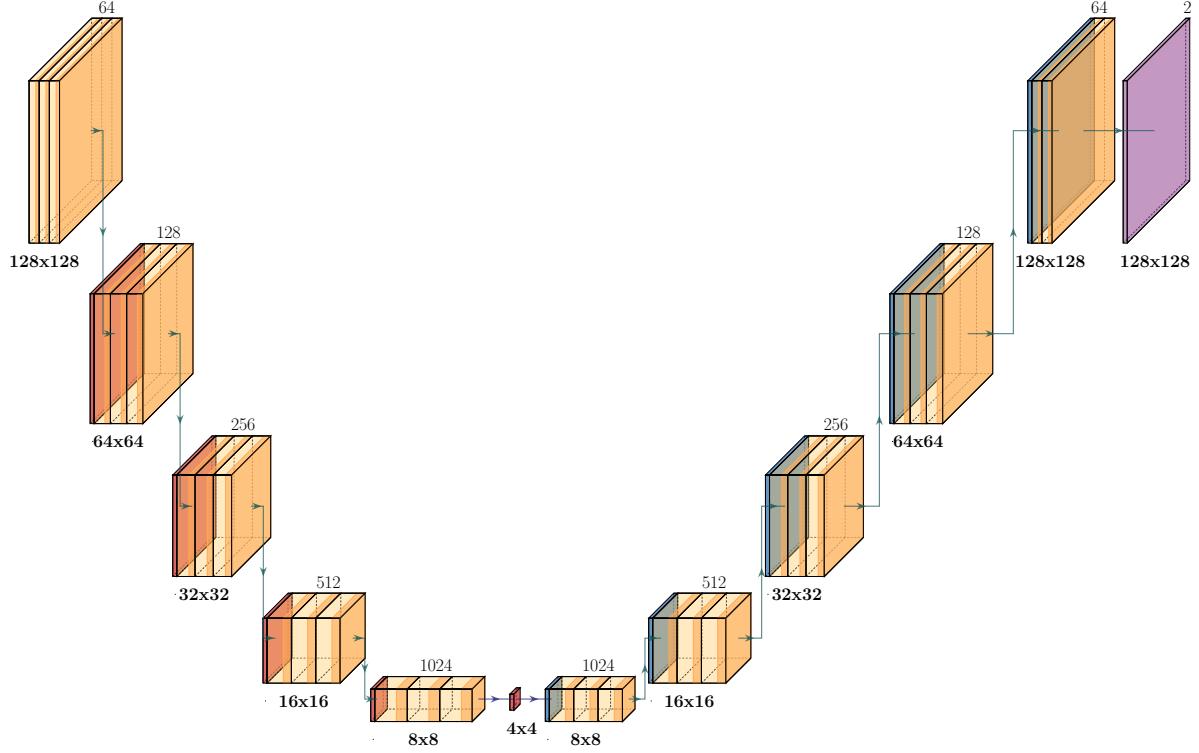


Figure 4.4: Encoder-decoder architecture (example for the input image of resolution 128×128 pixels). Each box corresponds to a multi-channel feature map. The architecture consists of sequential layers of $\text{Conv } 3 \times 3 \rightarrow \text{BatchNorm} \rightarrow \text{ReLU}$ (orange), followed by $\text{MaxPool } 2 \times 2$ (red), and then $\text{UpSampling } 2 \times 2$ (blue), with a final Softmax layer (violet). The number of channels is denoted on the top of the box. The resolution is denoted on the bottom of the box.

encoding blocks, using 256 and 512 filters, respectively. In the fifth encoding block, the number of filters increases to 1024, further deepening the network.

At the network's bottleneck, a max pooling layer with a 2×2 pool size further reduces the feature map's dimensions. The decoder mirrors the encoder's structure, but instead of downsampling, it upsamples the feature maps to reconstruct the spatial dimensions. The first decoding block starts with a 2D convolutional layer with 1024 filters, followed by batch normalization and a ReLU activation function, and then an upsampling layer with a 2×2 size. This process is repeated through the subsequent decoding blocks with 512, 256, 128, and finally 64 filters.

In the output layer, a 2D convolutional layer with a 1×1 kernel size is used. The number of filters in this layer corresponds to the number of classes, and a softmax activation function is applied to produce the final segmented output, assigning class probabilities to each pixel.

4.3 Binary classification

Getting started with the analysis, our focus is on identifying tumors and assigning labels, separating tumor (class 1) and non-tumor (class 0) areas, utilizing the earlier-mentioned data set. Each training cycle lasts 20 epochs, allowing for a comprehensive evaluation of model performance. We will use the original dataset first, then the normalized version, and finally the augmented one.

Original dataset

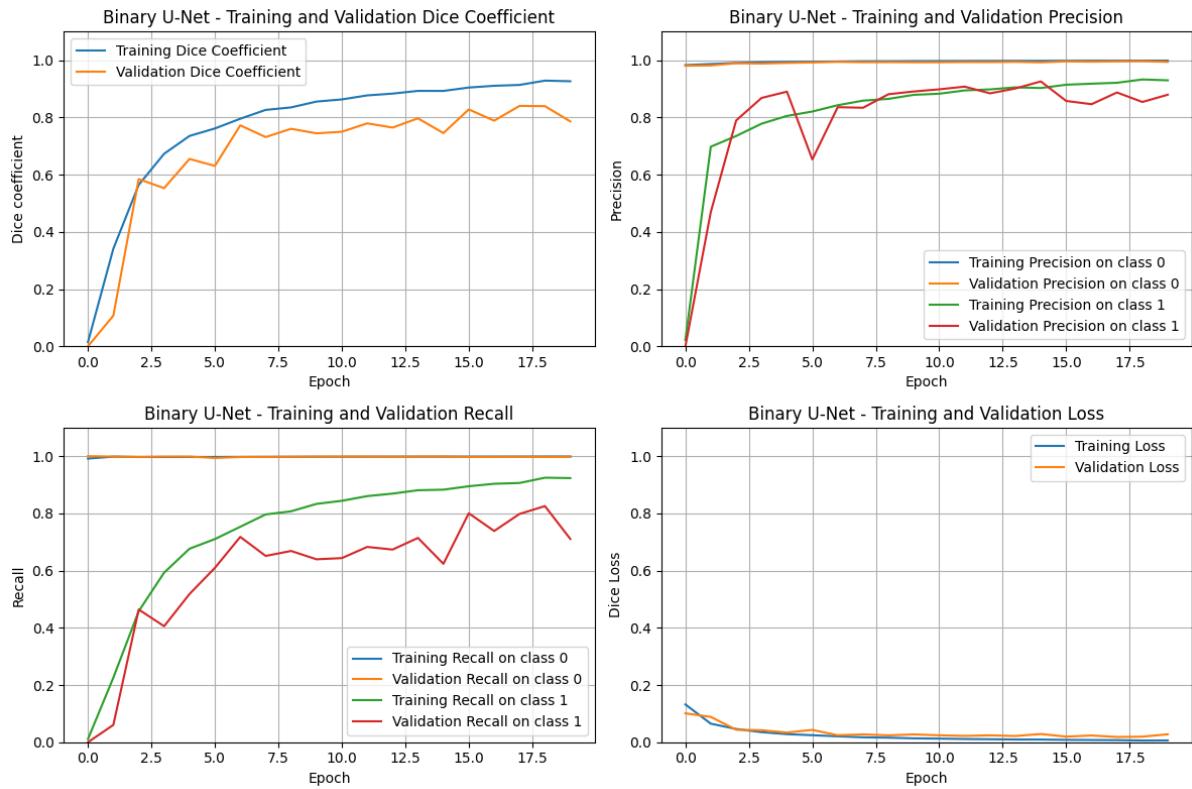


Figure 4.5: U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for original dataset.

Figure 4.5 displays the training and validation history for the original dataset using U-Net, showcasing the Dice coefficient (top left), precision (top right), recall (bottom left), and loss function (bottom right). The plots indicate the effective learning process of the network, as evidenced by the steadily increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

For the Dice coefficient, the training curve outperformed the validation curve, ending at around 0.9. However, the validation Dice coefficient, while improving rapidly in the initial epochs to around 0.65, fluctuates and stabilizes around 0.75 towards the end. This fluctuation suggests some overfitting, where the model performs better on the training data compared to the validation data. Nevertheless, the training process appears to be very efficient, with values increasing rapidly. This suggests that the learning process could potentially be stopped before reaching 20 epochs.

Precision and recall for the “no tumor” class were consistently around 1 for both training and validation sets, indicating that the model rarely misclassifies the background. The training precision and recall for class 1 starts low but increases rapidly to about 0.8–0.9, showing that the model learns to identify the “tumor” class effectively. The validation precision for class 1 (tumor) follows a similar trend but exhibits more fluctuations, which reflects variability in the model’s performance on the validation set. The validation recall for class 1 is more varied, stabilizing at around 0.7, which again suggests occasional underperformance on the validation set.

The loss function reveals the model’s overall performance in minimizing prediction errors. The training loss decreases steadily and stabilizes at a low value, indicating effective learning from the training data. The validation loss also decreases initially and flattens similarly to the training loss but shows slight increases towards the end. These low overall loss values suggest good performance, although the slight increase in validation loss towards the later epochs might indicate the development of overfitting.

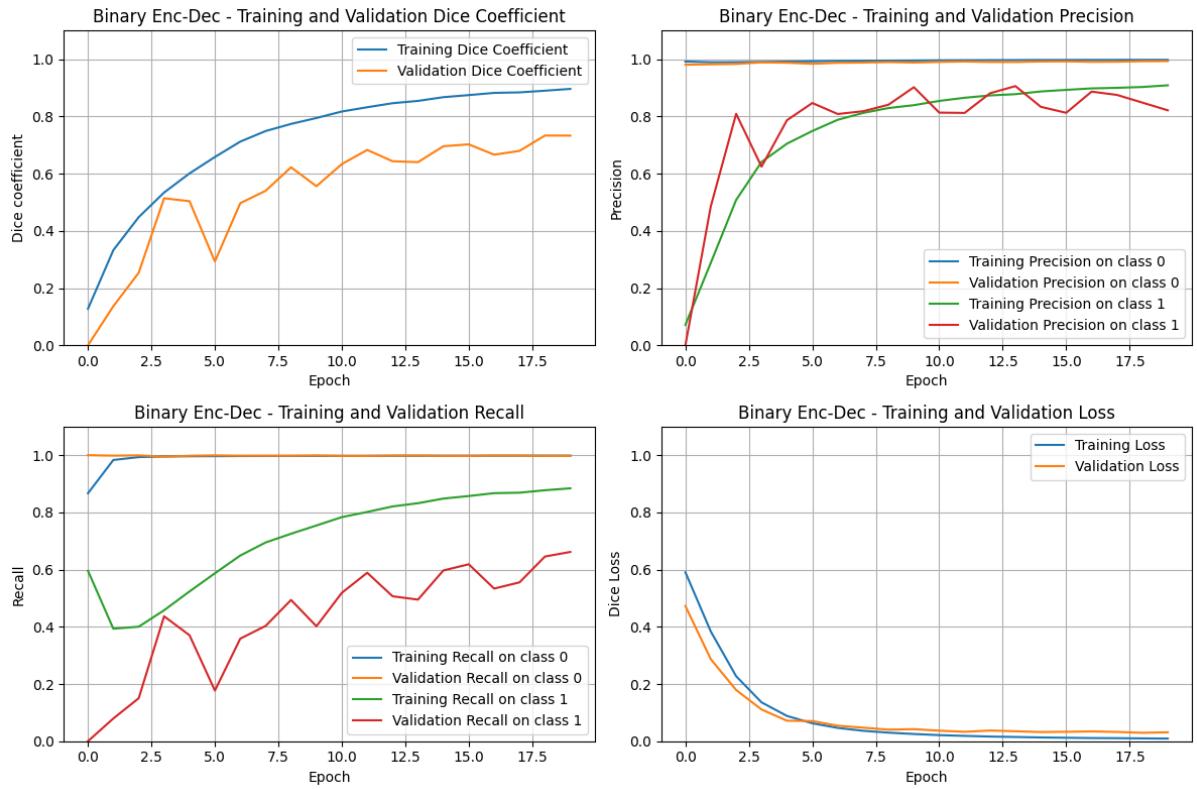


Figure 4.6: Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for original dataset.

Figure 4.6 is analogous to the previously described one, but it represents the training and validation history for the encoder-decoder network. Once again, the plots provide evidence that the network was learning effectively, indicated by the increasing Dice coefficient, precision, and recall over epochs, and a decreasing loss function. However, the learning process for the encoder-decoder architecture is slower compared to U-Net, with values not rising as quickly as before.

The training curve for the Dice coefficient outperformed the validation one, ending at around 90%. The validation Dice coefficient fluctuates between 0.5 – 0.7 and stabilizes around 0.75 towards the end. This fluctuation suggests some overfitting, where the model

performs better on the training data compared to the validation data. Nevertheless, the training process appears to be efficient, but less successful in comparison with U-Net.

Precision and recall for the “no tumor” class remained close to 1 for both training and validation sets throughout the training process. This indicates that the model rarely misclassifies the background. The training precision and recall for class 1 starts low but increases rapidly to about 0.8 – 0.9, showing that the model learns to identify the “tumor” class effectively. It is crucial to mention that the increase is not as rapid as in case of the U-Net. The validation recall for class 1 is more varied, stabilizing at around 0.55, which again suggests underperformance on the validation set.

The training loss decreases steadily and stabilizes at a low value, indicating effective learning from the training data. The validation loss also decreases initially and flattens similarly to the training loss but shows slight increases, relative to training loss, towards the end. These low overall loss values suggest good performance, although the slight increase in validation loss towards the later epochs might indicate the development of overfitting. Once again, it is worth mentioning that the decrease is not as rapid as in case of the U-Net.

In summary, it is important to note that the encoder-decoder architecture resulted in smaller values of Dice coefficient, precision, and recall with higher loss for each epoch compared to U-Net. This suggests that the encoder-decoder network was learning at a slower pace.

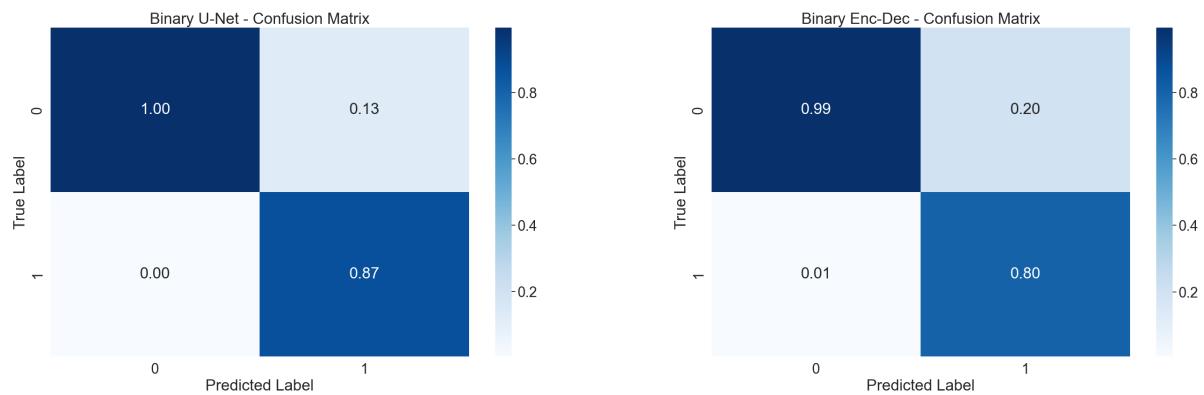


Figure 4.7: U-Net: Confusion matrix for original dataset (for test dataset).

Figure 4.8: Encoder-decoder: Confusion matrix for original dataset (for test dataset).

Figure 4.7 shows the confusion matrix for original dataset for U-Net. A remarkably high true negative rate of 1.00 indicates that the model correctly identifies areas without tumors almost perfectly. The false negative rate is relatively low at 0.13, suggesting that the model occasionally fails to identify tumors when they are present. However, there are no false positives, implying that the model rarely misclassifies non-tumor areas as tumor. The true positive rate is high at 0.87, indicating that the model is generally effective at identifying tumor regions.

Figure 4.8 shows the confusion matrix for original dataset for encoder-decoder network. In this confusion matrix, the true negative rate remains high at 0.99, indicating strong performance in identifying areas without tumors. However, the false negative rate is slightly higher at 0.20 compared to U-Net, suggesting a slightly higher rate of failing to identify tumors when present. There are very few false positives with a value of 0.01, indicating

a low rate of misclassifying non-tumor areas as tumor. The true positive rate is lower with a value of 0.80 compared to U-Net, indicating a slightly lower effectiveness in identifying tumor regions.

Comparing the two models, U-Net demonstrates a slightly higher true negative rate and a lower false negative rate compared to the encoder-decoder. Additionally, U-Net has no false positives, while the encoder-decoder has a very low rate. However, U-Net also exhibits a higher true positive rate compared to the encoder-decoder. Overall, U-Net appears to outperform the encoder-decoder in terms of correctly identifying both tumor and non-tumor regions.

Binary U-Net - Prediction Visualization

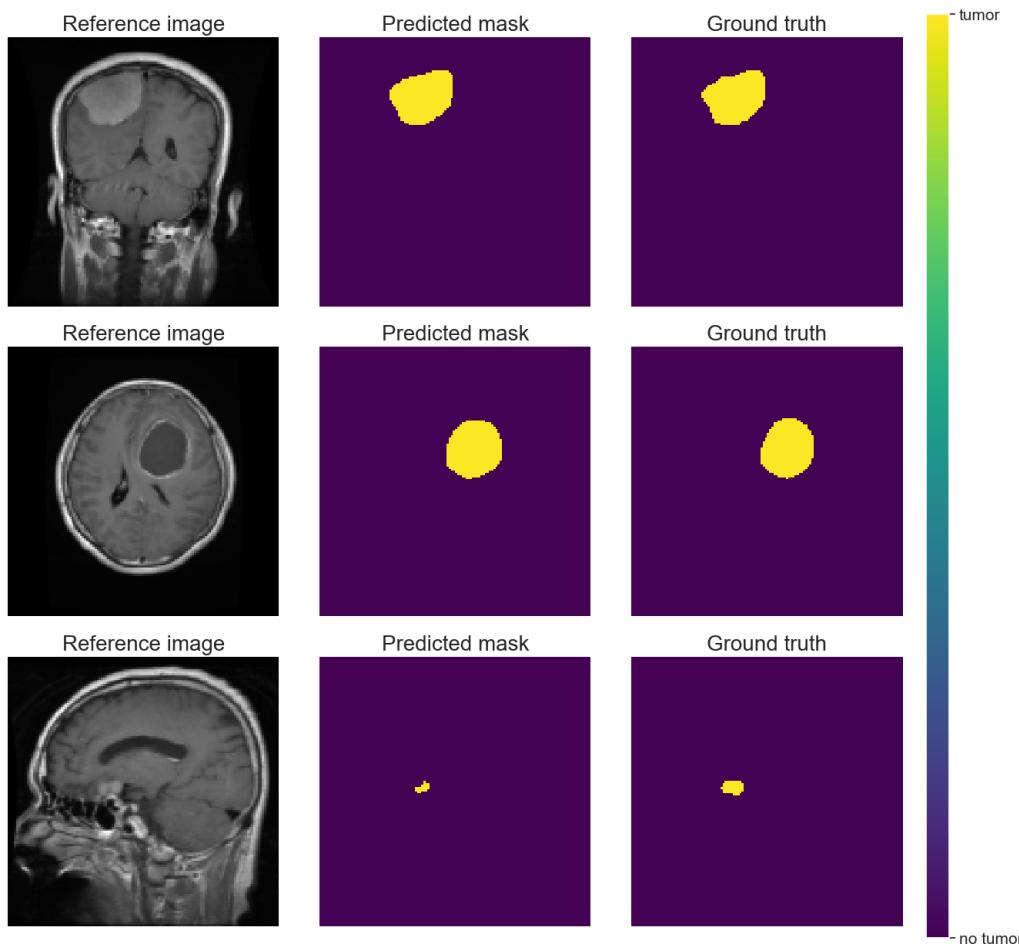


Figure 4.9: U-Net: Prediction visualization for sample images in each tumor type class for original dataset.

Prediction visualization using U-Net for sample images for original dataset is presented in Figure 4.9. It exemplifies its performance in tumor segmentation across three different types of tumor classes, with each row representing a different class. In each row, the reference image, predicted mask, and ground truth are displayed for a particular type of tumor.

In each case, the predicted mask shows a close match to the ground truth, indicating that the model accurately segmented the tumor region. The yellow areas in both the predicted mask and ground truth align well, signifying effective detection and segmentation for this tumor class. In the third row, despite the smaller size of the tumor, the model successfully identifies and segments it with sufficient accuracy. The alignment of the yellow areas in both the predicted mask and ground truth indicates that the model can handle different tumor sizes and still provide acceptable segmentation. This suggests that the U-Net network is effective and reliable for tumor segmentation in the given dataset.

Binary Enc-Dec - Prediction Visualization

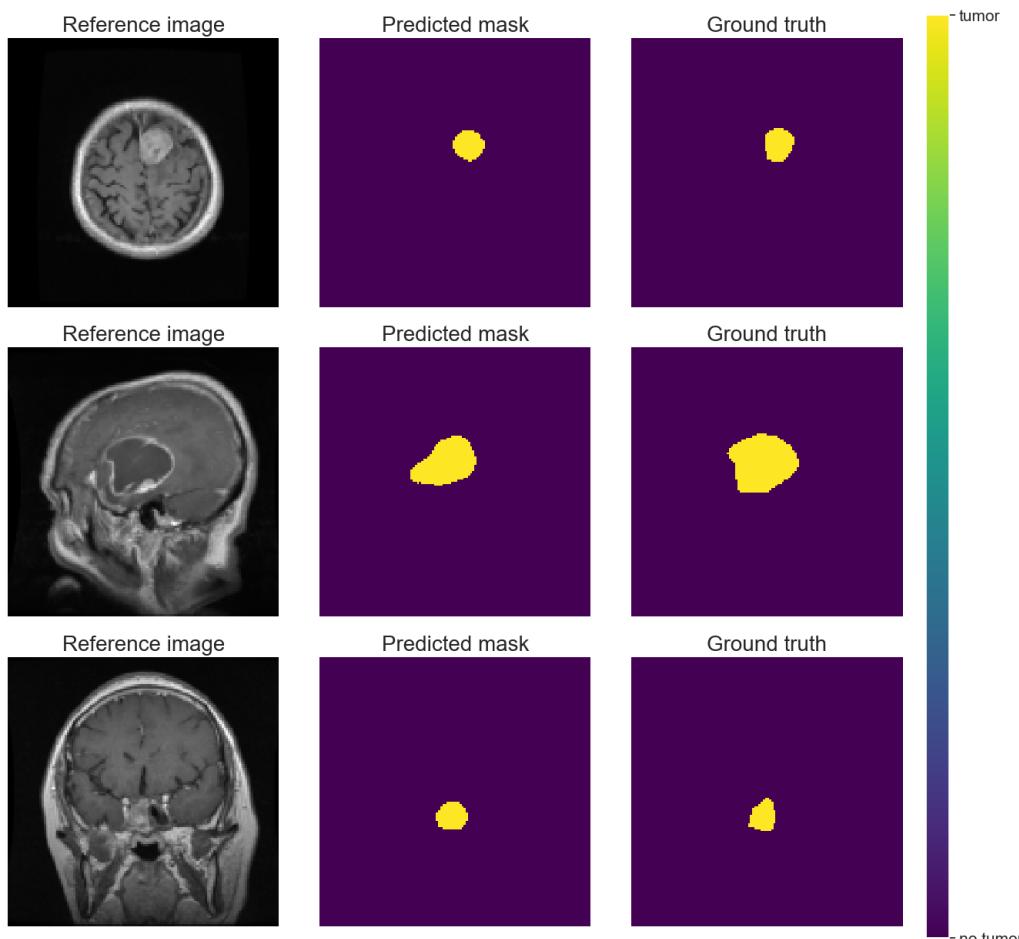


Figure 4.10: Encoder-decoder: Prediction visualization for sample images in each tumor type class for original dataset.

The prediction visualization for the encoder-decoder architecture is presented in Figure 4.10. Similarly, it is again displayed across three different types of tumor classes, with each row representing a distinct class.

In each row, the predicted mask closely matches the ground truth, indicating that the model effectively segmented the tumor region. The yellow areas in both the predicted mask and the ground truth are well-aligned, suggesting accurate detection and segmentation for this tumor class. Even with the smaller size of the tumor in the third row, the model successfully identifies and segments it accurately. The alignment of the yellow areas in both

the predicted mask and ground truth indicates that the model can handle different tumor sizes and still provide accurate segmentation.

Comparing these two networks, they both demonstrate strong performance in tumor segmentation across different tumor types. However, it is crucial to mention that these are only sample images for a single run, therefore variations may appear throughout the whole dataset.

Normalized dataset

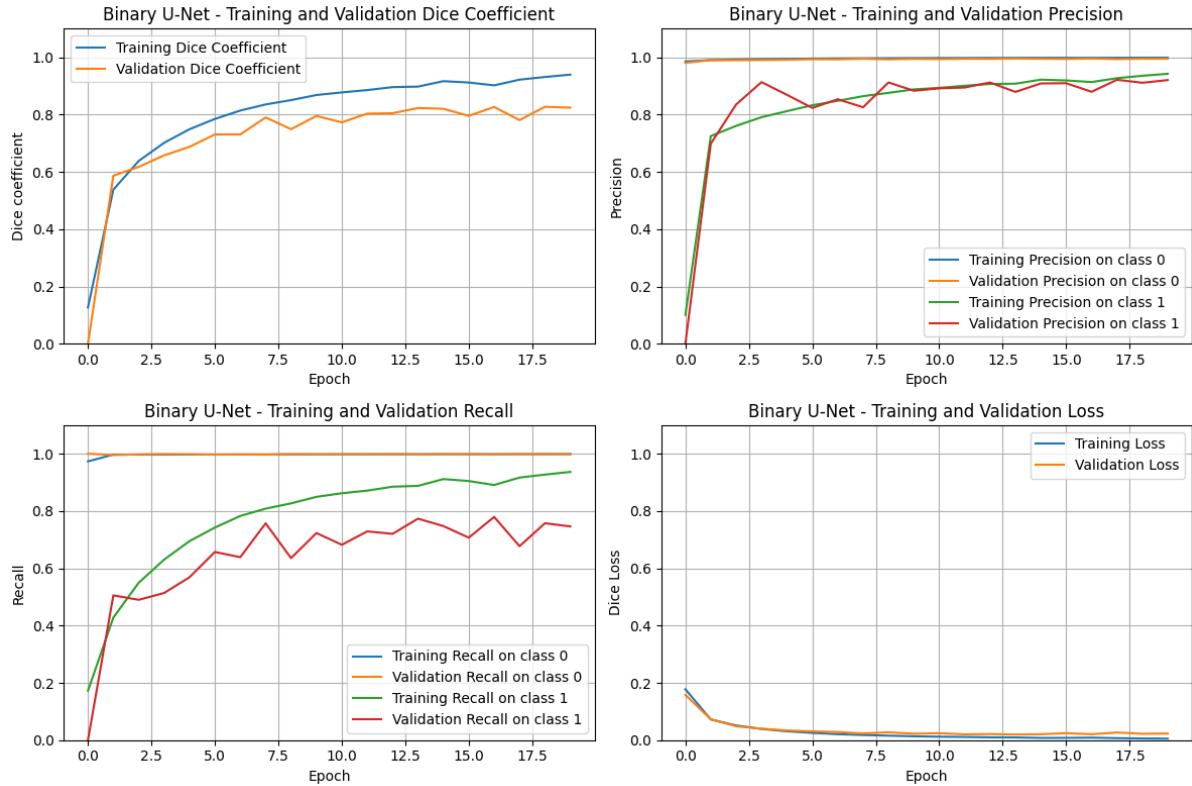


Figure 4.11: U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.

Figure 4.11 shows the training history of U-Net network for the normalized dataset. The tracked metrics include Dice coefficient, precision, recall, and loss for both training and validation datasets. The plots indicate the effective learning process of the network, as evidenced by the steadily increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

The training Dice coefficient increases rapidly initially and then stabilizes around $0.9 - 0.95$, indicating that the model is learning to predict well on the training data. Validation Dice coefficient also increases rapidly but stops at around 0.8, which is lower than the training one. This suggests some overfitting, as the model performs better on the training data compared to the validation data.

Both precision and recall for class 0 (“no tumor”) remain very high and stable for both training and validation datasets, close to 1. This indicates that the model is very good at identifying the background correctly. Precision for class indicating tumor shows some fluctuations, but generally stays above 0.8 for both training and validation sets.

The validation precision is slightly lower than the training precision, again indicating a degree of overfitting. Recall for class 1 starts lower and increases gradually, but shows more variance compared to precision. Training recall for “tumor” class stabilizes around 0.85, while validation recall fluctuates around 0.6 – 0.7. This suggests that the model is less consistent in identifying all relevant pixels including tumors, particularly on the validation dataset.

Training loss decreases rapidly initially and then stabilizes around a relatively small value. This shows that the model is learning effectively. Validation loss follows a similar trend to the training loss, decreasing rapidly and stabilizing. The proximity of training and validation loss suggests that while there is some overfitting, it is not severe.

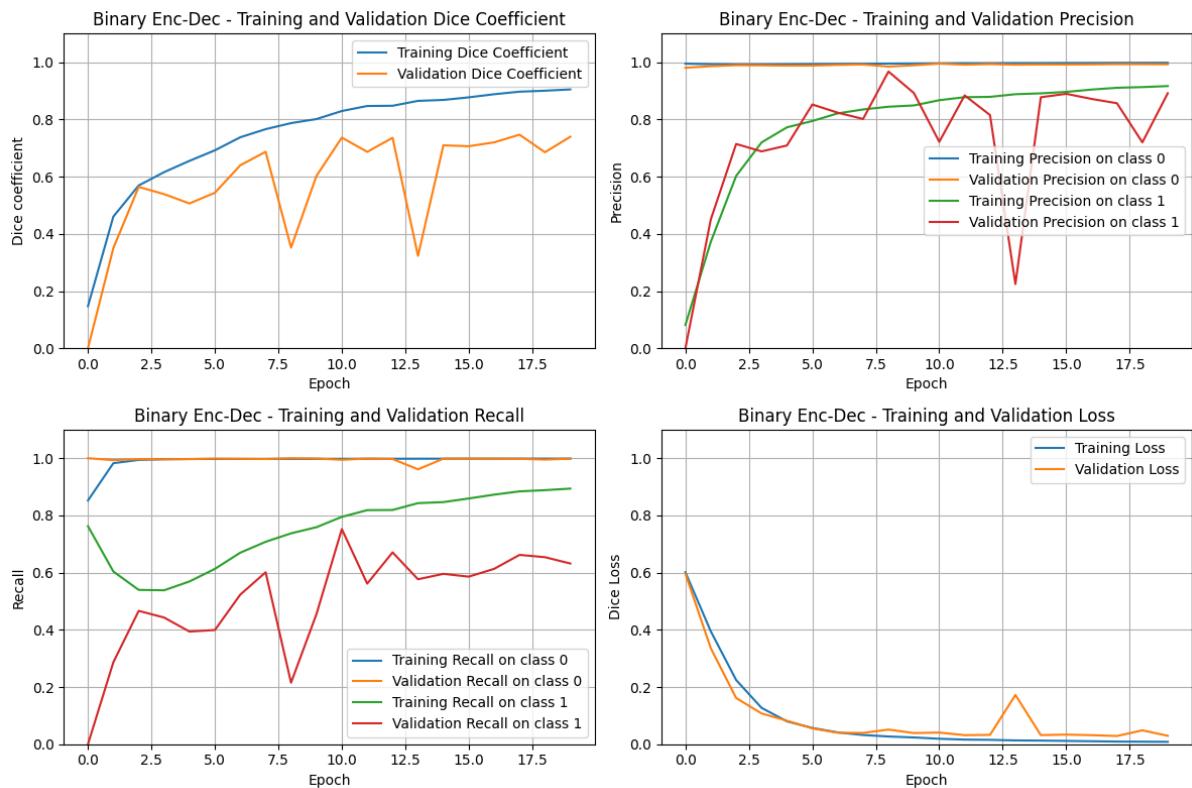


Figure 4.12: Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.

Figure 4.12 shows the training history of a binary encoder-decoder model with tracked metrics including Dice coefficient, precision, recall, and loss for both training and validation datasets. The plots indicate the effective, but fluctuating learning process of the network, as evidenced by the increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

The training Dice coefficient increases rapidly initially and then continues to improve more gradually, stabilizing around 0.8. This indicates effective learning on the training data, but the performance is slightly lower compared to the U-Net’s training Dice coefficient. The validation Dice Coefficient exhibits significant fluctuations, but generally oscillates around 0.6 – 0.7. This is consistently lower than the U-Net’s validation Dice coefficient, suggesting the encoder-decoder model struggles more with generalization.

Precision and recall remain very high and stable for both training and validation sets, similar to the U-Net, indicating effective background identification. However, in class

1 precision shows substantial fluctuations, especially in the validation set. Although it peaks near 0.8 at times, it lacks the stability seen in the U-Net model. This variability indicates the encoder-decoder model has inconsistent performance in identifying pixels containing tumor correctly. Recall for “tumor” class exhibits more variability and lower values than the U-Net, with the validation recall often decreasing below 0.6 and even dropping below 0.4 at some points. This suggests that the encoder-decoder model is less reliable in detecting all relevant pixels with tumors compared to the U-Net.

Training and validation loss plot indicates that training loss decreases rapidly initially and stabilizes at a low value, similar to the U-Net. This shows effective learning during training. However, the validation loss follows a similar trend to training loss but with notable fluctuations around epoch 12. While the general trend mirrors the U-Net, the fluctuations indicate less stable validation performance.

It is crucial to note that, while both models show effective learning, the U-Net outperforms the encoder-decoder model in terms of generalization and stability, particularly in handling tumor detection. This suggests that for tasks requiring precise segmentation, the U-Net may be a more reliable choice.

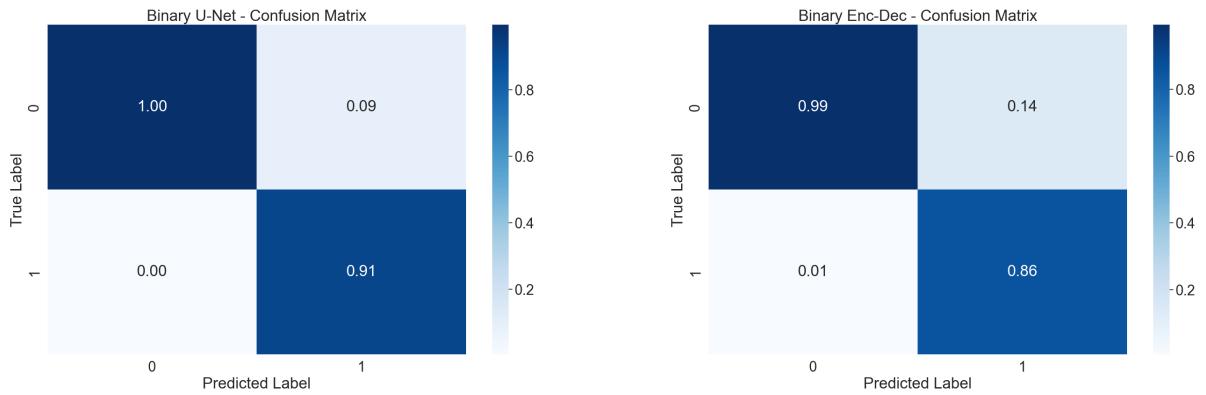


Figure 4.13: U-Net: Confusion matrix for normalized dataset (for test dataset).

Figure 4.14: Encoder-decoder: Confusion matrix for normalized dataset (for test dataset).

Figure 4.13 shows the confusion matrix for normalized test set for U-Net. A remarkably high true negative rate of 1.00 indicates that the model correctly identifies areas without tumors almost perfectly. The false negative rate is relatively low at 0.09, suggesting that the model occasionally fails to identify tumors when they are present. However, there are no false positives, implying that the model rarely misclassifies non-tumor areas as tumor. The true positive rate is high at 0.91, indicating that the model is generally effective at identifying tumor regions.

Figure 4.14 shows the confusion matrix for normalized test set for encoder-decoder network. In this confusion matrix, the true negative rate remains high at 0.99, indicating strong performance in identifying areas without tumors. However, the false negative rate is slightly higher at 0.14 compared to U-Net, suggesting a slightly higher rate of failing to identify tumors when present. There are very few false positives with a value of 0.01, indicating a low rate of misclassifying non-tumor areas as tumor. The true positive rate is lower with a value of 0.86 compared to U-Net, indicating a slightly lower effectiveness in identifying tumor regions.

Comparing the two models, U-Net demonstrates a slightly higher true negative rate and a lower false negative rate compared to the encoder-decoder. Additionally, U-Net

has no false positives, while the encoder-decoder has a very low rate. However, U-Net also exhibits a higher true positive rate compared to the encoder-decoder. Overall, U-Net appears to outperform the encoder-decoder in terms of correctly identifying both tumor and non-tumor regions. It is also essential to point that an overall performance is better for the normalized dataset than for the original one for both networks (Figures 4.7 and 4.8).

Binary U-Net - Prediction Visualization

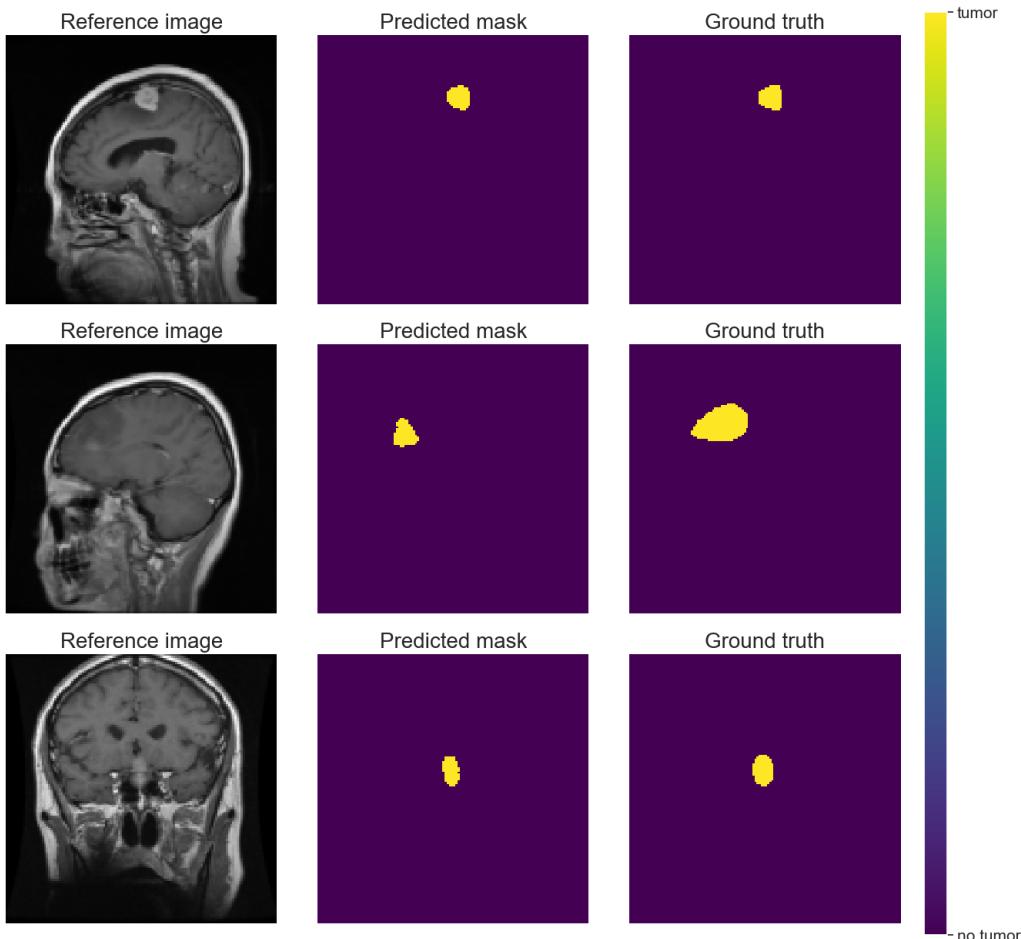


Figure 4.15: U-Net: Prediction visualization for sample images in each tumor type class for normalized dataset.

Prediction visualization using U-Net for sample images for normalized dataset is presented in Figure 4.15. It exemplifies its performance in tumor segmentation across three different types of tumor classes, with each row representing a different class. In each row, the reference image, predicted mask, and ground truth are displayed for a particular type of tumor.

In the first and third row, the predicted mask accurately captures the tumor region, with the yellow area closely matching the ground truth. This indicates that the model effectively segmented the tumor region, demonstrating high accuracy for this tumor classes (despite the size of the tumor). In the second row, the predicted mask shows a good alignment with the ground truth, although the shape of the predicted tumor region is slightly different. Despite these minor discrepancies, the model correctly identifies the tumor region, indicating reliable performance for this tumor class.

Overall, this network demonstrates effective tumor segmentation across different tumor types. The predicted masks closely match the ground truth in most cases, suggesting that the model performs well in identifying and segmenting tumor regions accurately. Despite minor shape discrepancies in some predictions, the model consistently identifies the correct tumor regions, indicating robust performance across various tumor classes.

Binary Enc-Dec - Prediction Visualization

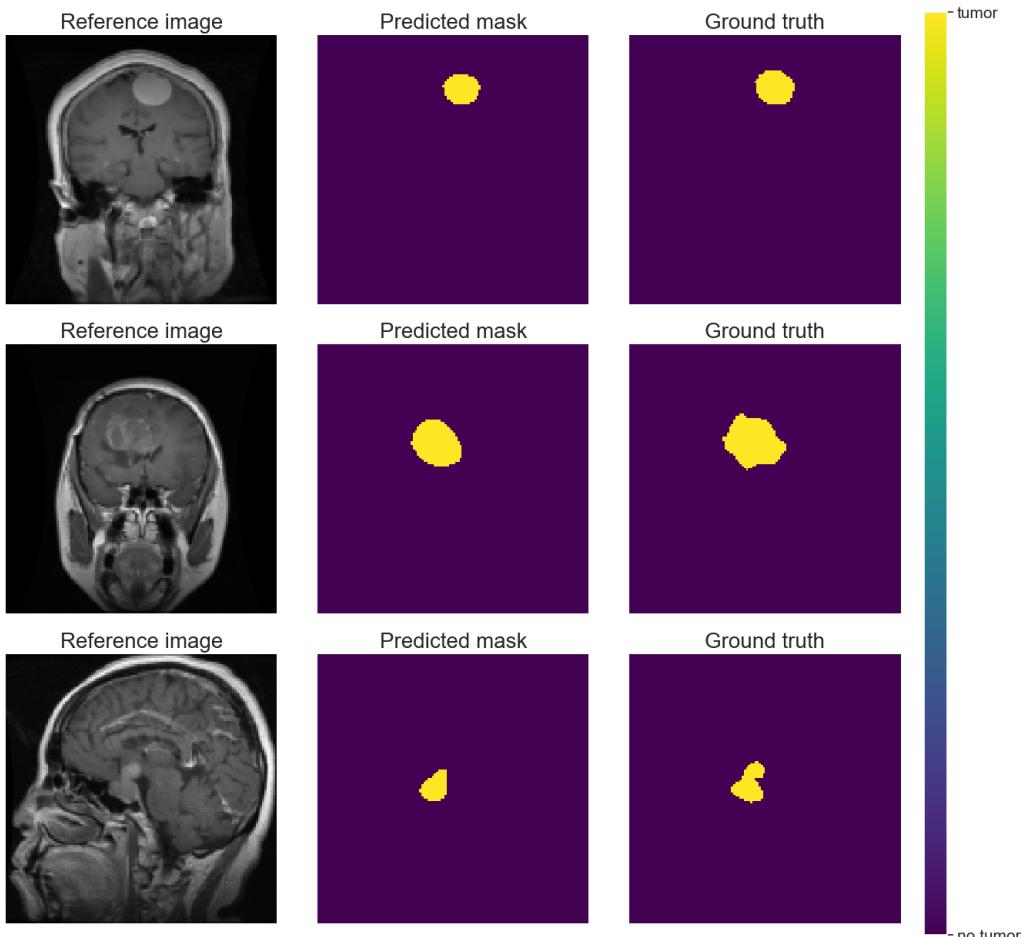


Figure 4.16: Encoder-decoder: Prediction visualization for sample images in each tumor type class for normalized dataset.

The prediction visualization for the encoder-decoder architecture is presented in Figure 4.16. Similarly, it is again displayed across three different types of tumor classes, with each row representing a distinct class.

In each row, the predicted mask closely matches the ground truth, indicating that the model effectively segmented the tumor region. The yellow areas in both the predicted mask and the ground truth are well-aligned, suggesting accurate detection and segmentation for this tumor class. In the third row, the predicted mask shows a good alignment with the ground truth, although the shape of the predicted tumor region is slightly different – the predicted tumor is more round compared to the true mask. Despite these minor discrepancies, the network correctly identifies the tumor region, indicating reliable performance for this tumor class.

Comparing these two networks, they both demonstrate strong performance in tumor segmentation across different tumor types. In some scenarios, the shapes of the predicted tumor regions slightly differ, but in the vast majority of cases, networks correctly identify the tumor regions. Again, it is crucial to mention that these are only sample images for a single run, therefore variations may appear throughout the whole dataset.

Augmented dataset

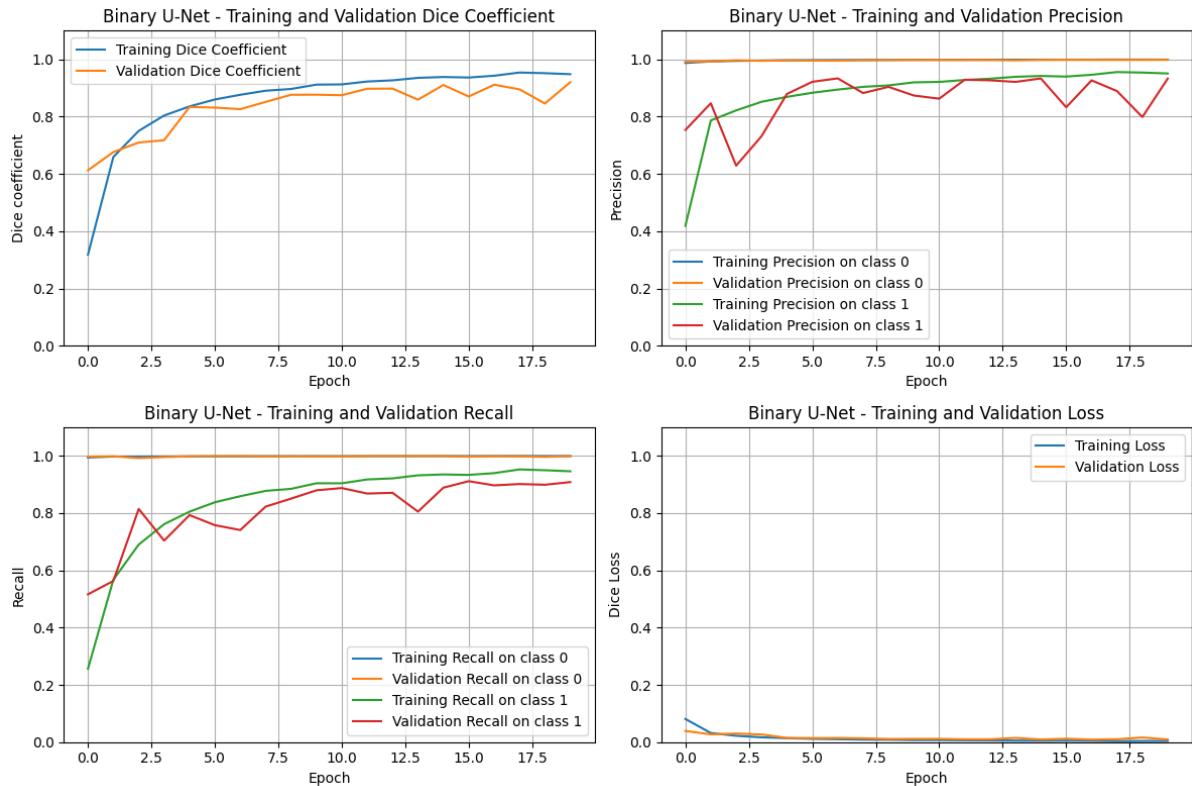


Figure 4.17: U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.

Figure 4.17 displays the training and validation history for the augmented dataset using U-Net, showcasing the Dice coefficient (top left), precision (top right), recall (bottom left), and loss function (bottom right). The plots indicate the effective learning process of the network, as evidenced by the steadily increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

For the Dice coefficient, the training curve outperformed the validation curve, ending at around 0.95. However, the validation Dice coefficient, while improving rapidly in the initial epochs to around 0.8, fluctuates and stabilizes around 0.9 towards the end. This fluctuation suggests some overfitting, where the model performs better on the training data compared to the validation data. Nevertheless, the training process appears to be very efficient, with values increasing rapidly. This suggests that the learning process could potentially be stopped before reaching 20 epochs.

Precision and recall for the “no tumor” class were consistently around 1 for both training and validation sets, indicating that the model rarely misclassifies the background. The training precision and recall for class 1 starts low but increases rapidly to about

0.8 – 0.9, showing that the model learns to identify the “tumor” class effectively. The validation precision and recall for class 1 (tumor) follow a similar trend but exhibits more fluctuations, which reflects variability in the model’s performance on the validation set.

The training and validation loss decreases steadily and stabilizes at a low value, indicating effective learning from the training data. These low overall loss values suggest good performance, although the slight increase in validation loss might indicate the development of overfitting.

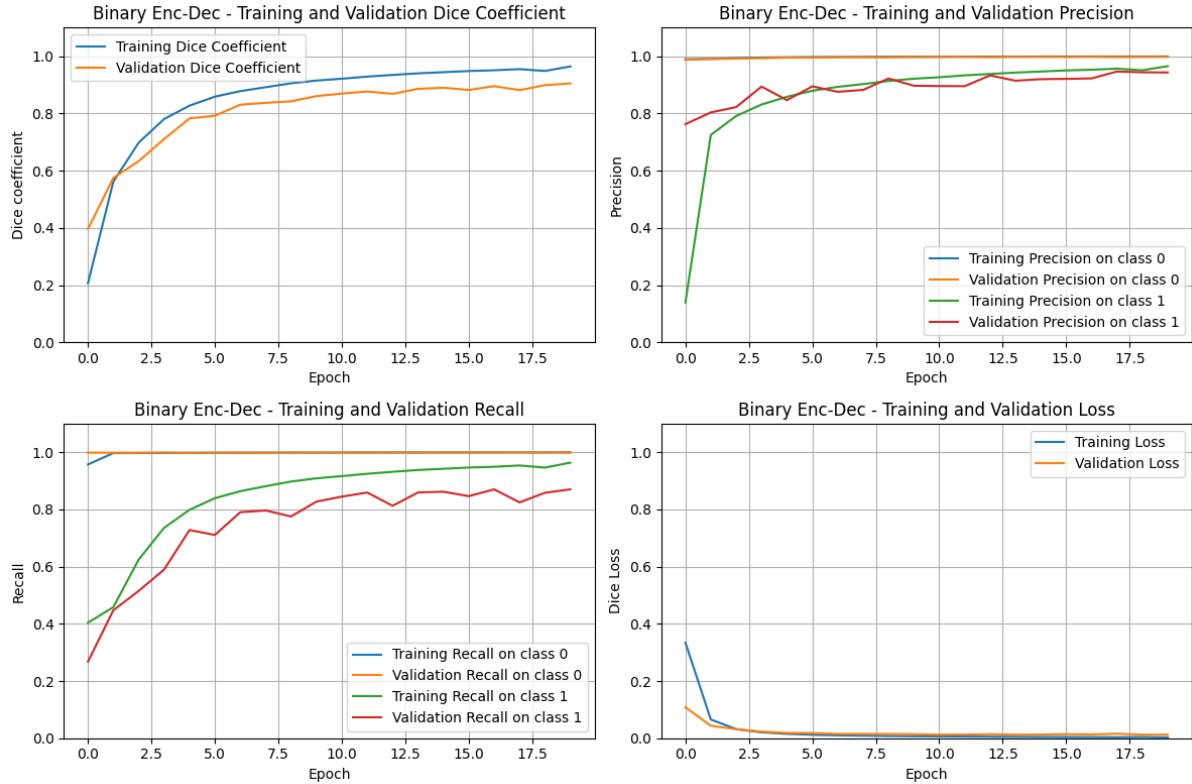


Figure 4.18: Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.

Figure 4.18 shows the training history of a binary encoder-decoder model with tracked metrics including Dice coefficient, precision, recall, and loss for both training and validation datasets. The plots indicate the effective, but slightly fluctuating learning process of the network, as evidenced by the increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

The training Dice coefficient begins low but rapidly increases, stabilizing around 0.9. Its validation Dice coefficient also rises quickly but exhibits less fluctuation than for U-Net, stabilizing around 0.8. Therefore, it can be stated that encoder-decoder demonstrates a slightly higher and more stable Dice coefficient in both training and validation compared to the U-Net.

Precision and recall for class 0 remain very high and stable for both training and validation sets, similar to the U-Net, indicating effective background identification. The training precision and recall for class 1 starts low but increases rapidly to about 0.8 – 0.9, showing that the model learns to identify the “tumor” class effectively. The validation precision for “tumor” class follows a similar trend but exhibits fewer fluctuations, which reflects variability in the model’s performance on the validation set. The validation recall

for class 1 is less varied, stabilizing at around 0.85. This indicates that the encoder-decoder model has slightly better recall stability for class 1 compared to the U-Net, with both models performing similarly well for class 0.

In terms of loss, it begins high and decreases rapidly, stabilizing close to zero. The validation loss follows a similar trend, stabilizing slightly above the training loss. Both models demonstrate effective loss reduction and stabilization, indicating good convergence.

In summary, while both U-Net and encoder-decoder models perform well, the second one shows superior stability and performance, particularly in precision and recall for the more challenging class 1. Both models perform excellently for class 0 and exhibit similar effectiveness in reducing and stabilizing loss. Therefore, the encoder-decoder model is marginally better overall due to its higher stability and slightly superior performance metrics.

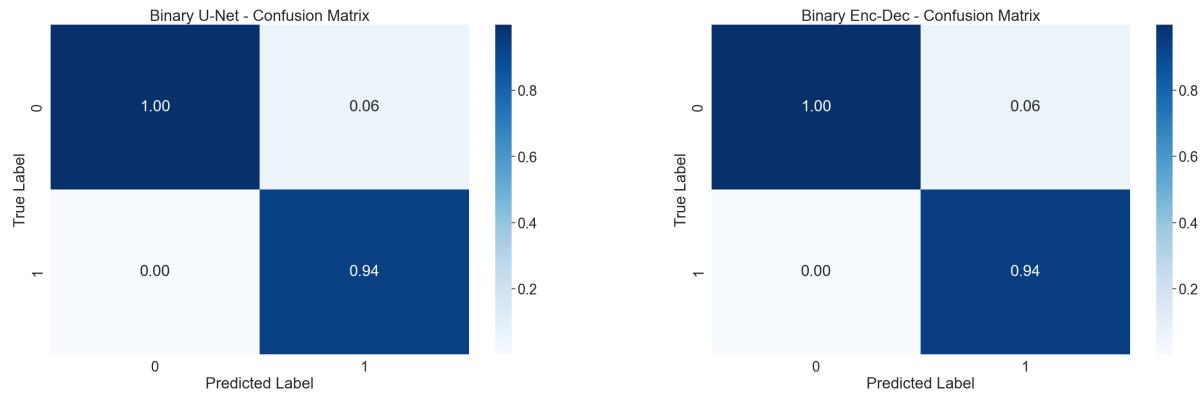


Figure 4.19: U-Net: Confusion matrix for augmented dataset (for test dataset).

Figure 4.20: Encoder-decoder: Confusion matrix for augmented dataset (for test dataset).

Figure 4.19 shows the confusion matrix for augmented test set for U-Net. A remarkably high true negative rate of 1.00 indicates that the model correctly identifies areas without tumors almost perfectly. The false negative rate is relatively low at 0.06, suggesting that the model occasionally fails to identify tumors when they are present. However, there are no false positives, implying that the model rarely misclassifies non-tumor areas as tumor. The true positive rate is high at 0.94, indicating that the model is generally effective at identifying tumor regions.

Figure 4.20 shows the confusion matrix for augmented test set for encoder-decoder network. Comparing the two models, both of them yield the same result. It is also essential to point that an overall performance is better for the augmented dataset than for the normalized one for both networks (Figures 4.13 and 4.14).

Prediction visualization using U-Net for sample images for augmented dataset is presented in Figure 4.21. It exemplifies its performance in tumor segmentation across three different types of tumor classes, with each row representing a different class. In each row, the reference image, predicted mask, and ground truth are displayed for a particular type of tumor.

In each case, the predicted mask shows a close match to the ground truth, indicating that the model accurately segmented the tumor region. The yellow areas in both the predicted mask and ground truth align well, signifying effective detection and segmentation for this tumor class. In the first row, despite the smaller size of the tumor, the model successfully identifies and segments it with remarkable accuracy. The alignment of the yellow areas in both the predicted mask and ground truth indicates that the model can

Binary U-Net - Prediction Visualization

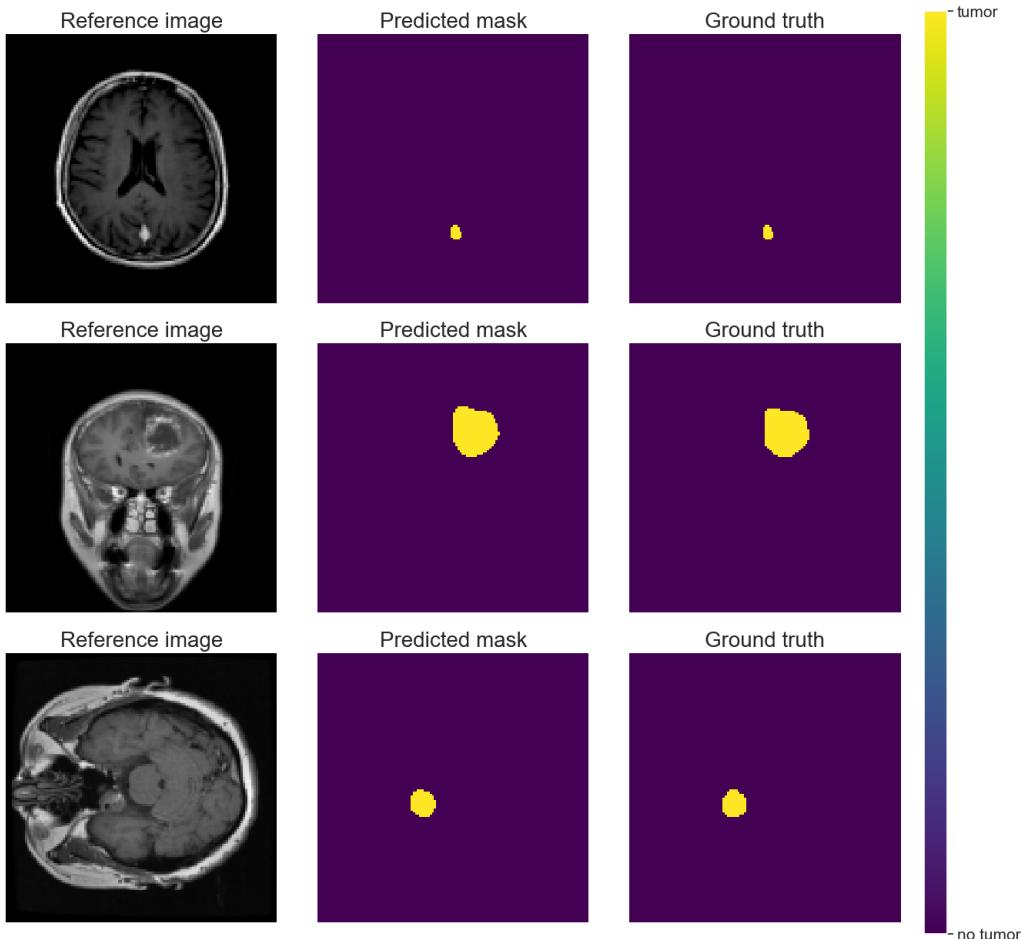


Figure 4.21: U-Net: Prediction visualization for sample images in each tumor type class for augmented dataset.

handle different tumor sizes and still provide acceptable segmentation. This suggests that the U-Net network is effective and reliable for tumor segmentation in the given dataset.

The prediction visualization for the encoder-decoder architecture is presented in Figure 4.22. Similarly, it is again displayed across three different types of tumor classes, with each row representing a distinct class.

In each row, the predicted mask closely matches the ground truth, indicating that the model effectively segmented the tumor region. The yellow areas in both the predicted mask and the ground truth are well-aligned, suggesting accurate detection and segmentation for this tumor class. Even with the smaller size of the tumor in the third row or bigger size in the second row, the model successfully identifies and segments it accurately. The alignment of the yellow areas in both the predicted mask and ground truth indicates that the model can handle different tumor sizes and still provide accurate segmentation.

Comparing these two networks, they both demonstrate strong performance in tumor segmentation across different tumor types. It is essential to mention that these are only sample images for a single run, therefore variations may appear throughout the whole dataset.

Binary Enc-Dec - Prediction Visualization

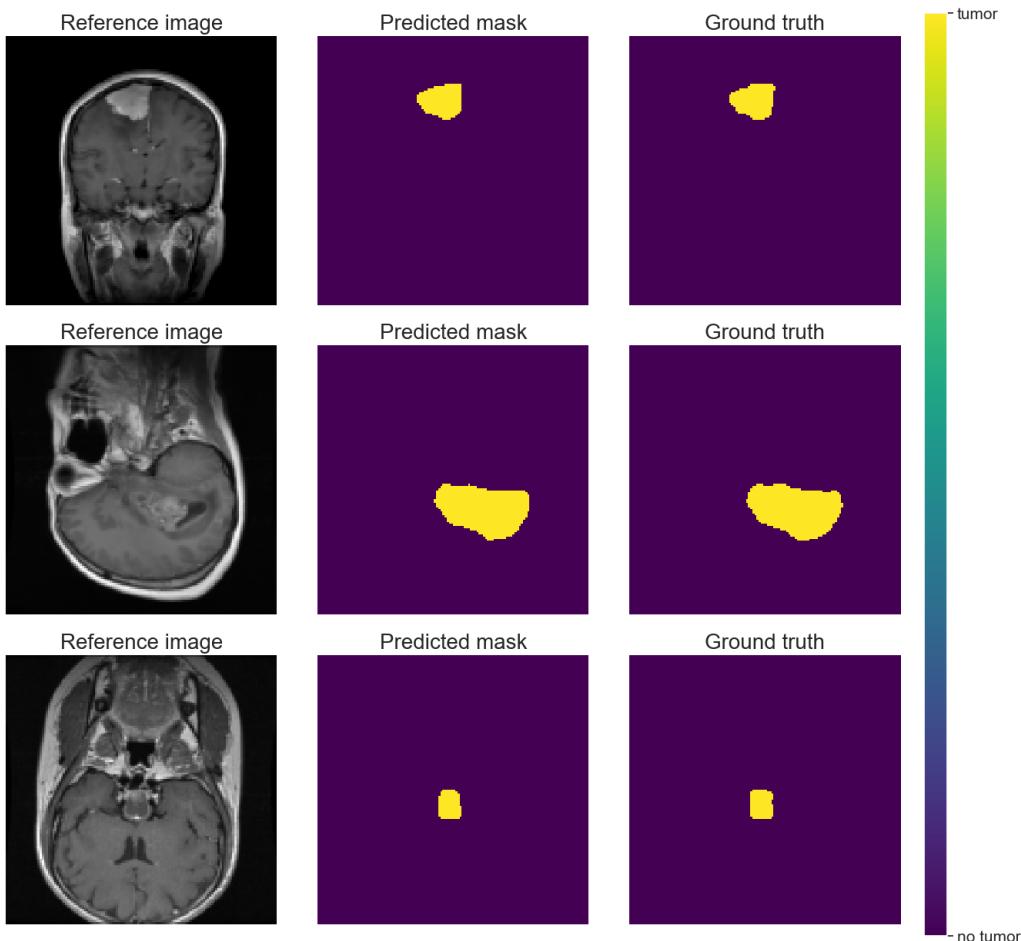


Figure 4.22: Encoder-decoder: Prediction visualization for sample images in each tumor type class for augmented dataset.

4.4 Performance comparison table

Table 4.1 presents metrics for U-Net and encoder-decoder networks across three different datasets: original, normalized, and augmented. Each network is analyzed based on loss, Dice coefficient, balanced accuracy, recall for class 0 (“no tumor”) and class 1 (“tumor”), as well as precision for both classes.

For the U-Net on the original dataset, the loss is 0.02443, and the Dice coefficient is 0.78660, indicating moderate segmentation performance. The balanced accuracy stands at 0.85738, with a high recall for class 0 (0.99828) but lower for class 1 (0.71647), highlighting some difficulty in detecting tumors. Precision for class 0 is 0.99537, and for class 1, it is 0.87195, suggesting reasonably accurate predictions.

On the normalized dataset, U-Net shows improved performance with a loss of 0.01951 and a higher Dice coefficient of 0.82450. The balanced accuracy also increases to 0.87516. Recall for class 0 remains very high at 0.99883, and for class 1, it improves to 0.75150. Precision for class 0 is 0.99594, while for class 1, it increases to 0.91321 compared to the original dataset, indicating better tumor detection and fewer false positives compared to the original dataset.

Table 4.1: **Table of metrics for U-Net and encoder-decoder for all considered datasets in binary segmentation.**

	U-Net			Encoder-decoder		
	Original	Normalized	Augmented	Original	Normalized	Augmented
Loss	0.02443	0.01951	0.00809	0.02594	0.02715	0.01072
Dice coefficient	0.78660	0.82450	0.92866	0.73593	0.73485	0.91687
Balanced accuracy	0.85738	0.87516	0.95986	0.83837	0.82007	0.94689
Recall (0)	0.99828	0.99883	0.99892	0.99726	0.99828	0.99901
Recall (1)	0.71647	0.75150	0.92080	0.67948	0.64186	0.89477
Precision (0)	0.99537	0.99594	0.99862	0.99476	0.99416	0.99817
Precision (1)	0.87195	0.91321	0.93666	0.80260	0.85935	0.94009

The U-Net achieves its best performance on the augmented dataset, with the lowest loss of 0.00809 and the highest Dice coefficient equal to 0.92866. The balanced accuracy is significantly higher at 0.95986. Recall for class 0 is 0.99892, and for class 1, it rises significantly to 0.92080. Precision for class 0 is 0.99862, and for class 1, it is 0.93666, demonstrating highly effective tumor detection and segmentation.

In comparison, the encoder-decoder network on the original dataset shows lower performance, with a slightly higher loss of 0.02594 and a lower Dice coefficient of 0.73593. The balanced accuracy is 0.83837, which is an inferior result compare to U-net. Recall for class 0 is 0.99726, and for class 1, it is 0.67948, indicating significant challenges in detecting tumors. Precision for class 0 is 0.99476, and for class 1, it is 0.80260, reflecting less accurate tumor predictions than U-Net for the same dataset.

When applied to the normalized dataset, the encoder-decoder network does not show significant improvement. The loss is slightly higher at 0.02715, and the Dice coefficient remains low at 0.73485. The balanced accuracy drops to 0.82007. Recall for class 0 is 0.99828, while for class 1, it decreases to 0.64186. Precision for class 0 is 0.99416, and for class 1, it is 0.85935, indicating a slight improvement in precision but poorer recall for tumors. These values are substantially smaller than in case of U-Net for normalized dataset.

On the augmented dataset, the encoder-decoder network's performance improves significantly. The loss decreases to 0.01072, and the Dice coefficient rises to 0.91687. The balanced accuracy is high at 0.94689. Recall for class 0 is 0.99901, and for class 1, it improves to 0.89477. Precision for class 0 is 0.99817, and for class 1, it is 0.94009, showing substantial improvement in tumor detection and segmentation compared to the same network trained and tested on normalized dataset.

Comparing the networks for each dataset shows that U-Net consistently outperforms the encoder-decoder network. On the original dataset, U-Net has a lower loss and higher Dice coefficient, balanced accuracy, recall, and precision for both classes. For the normalized dataset, U-Net again shows superior performance with lower loss, higher Dice coefficient, and better recall and precision for both classes. On the augmented dataset, both networks show their best performance, but U-Net achieves a lower loss and higher Dice coefficient and balanced accuracy. Recall and precision for class 1 are also slightly better or insignificantly worse in U-Net, indicating more accurate tumor segmentation.

In conclusion, the best configuration for tumor segmentation is the U-Net model trained on the augmented dataset. This combination yields the lowest loss, highest Dice coefficient, and highest balanced accuracy. Both recall and precision for the tumor class (class 1) are the highest in this configuration, ensuring accurate and reliable tumor detection and segmentation. The encoder-decoder network also performs well on the augmented dataset, but does not match the performance metrics of the U-Net. Therefore, U-Net with data augmentation is the optimal choice for this binary tumor segmentation task.

It is noteworthy that the training and testing was performed only once for both networks. That being said, variations in values of all metrics may appear. Performing numerous independent experiments and calculating the mean values would be beneficial.

4.5 Multi-class classification

Now we extend our classification task to include multiple tumor types (meningioma, glioma, and pituitary tumor) and the absence of tumor. Using the normalized and augmented datasets employed in the binary classification task, we aim to accurately classify each pixel of the image into one of these four classes. Each training session will consist of 20 epochs to ensure comprehensive evaluation.

Normalized dataset

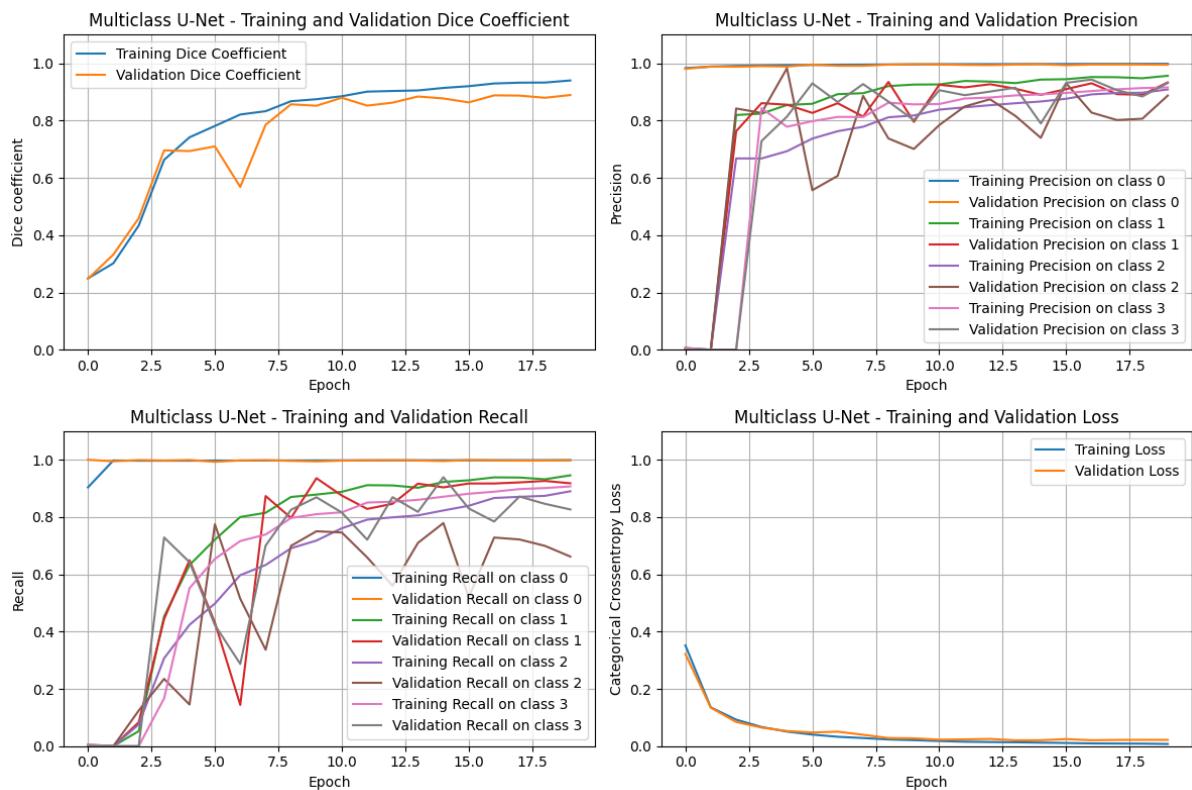


Figure 4.23: U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.

Figure 4.23 displays the training and validation history for the normalized dataset using U-Net, showcasing the Dice coefficient (top left), precision (top right), recall (bottom left), and loss function (bottom right). The plots indicate the effective learning process of the network, as evidenced by the increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

Both the training and validation Dice coefficients improve significantly during the initial epochs, with the training Dice coefficient reaching around 0.95 by the end. However, the validation Dice coefficient, while also improving, is consistently lower than the training Dice coefficient, stabilizing around 0.9. This gap suggests a degree of overfitting.

The precision plot reveals high precision for class 0 (no tumor) throughout the training, while the precision for other classes (meningioma, glioma, pituitary tumor) fluctuates initially but stabilizes at high values. The validation precision for these classes follows a similar trend but with more noticeable fluctuations, indicating variability in the model's ability to accurately classify these tumor types during validation. Recall for class 0 remains at the value of 1.0 throughout, indicating perfect recall for the "no tumor" class in both training and validation. For classes 1, 2, and 3, the recall improves significantly in the early epochs and stabilizes at high values, though there is more variability compared to the precision plot. This suggests that the model is generally good at identifying the presence of tumors, but may occasionally miss some instances during validation.

The loss plot indicates a steady decrease in both training and validation loss, converging to a low value by the end of training. The close alignment between the training and validation loss curves suggests that the model generalizes well to the validation set, despite the observed overfitting in the Dice coefficient plot.

Overall, the model performs reasonably well, especially in terms of precision and recall for the "no tumor" class, with reasonable performance for the tumor classes. The main area of concern is the overfitting indicated by the difference in Dice coefficients and the variability in precision and recall for tumor classes during validation.

Figure 4.24 displays the training and validation history for the normalized dataset using encoder-decoder network, showcasing the previously described metrics. The plots indicate the effective learning process of the network, as evidenced by the increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

The Dice coefficient plot shows a steady increase for both training and validation sets, with the training reaching around 0.85 and the validation reaching about 0.8. While there is a gap between training and validation Dice coefficients, the overall pattern is similar to the U-Net model, indicating some degree of overfitting.

In the precision plot, class 0 achieves high precision early on and remains stable, similar to the U-Net model. For the other classes (meningioma, glioma, pituitary tumor), precision improves significantly after the initial epochs and stabilizes, though with more fluctuations compared to the U-Net. The validation precision generally follows the training precision trends but with noticeable variability, suggesting inconsistent performance in the validation set. Recall for class 0 remains very high throughout the training, similar to the U-Net model. Recall for the other classes increases during the initial epochs and then stabilizes, though it shows more variability than precision. The validation recall is generally lower than the training recall for tumor classes, indicating that the model may be missing some tumor instances during validation, which is consistent with the observed variability.

The loss plot indicates a sharp decrease in both training and validation loss during the initial epochs, converging to low values and staying stable, much like in the U-Net

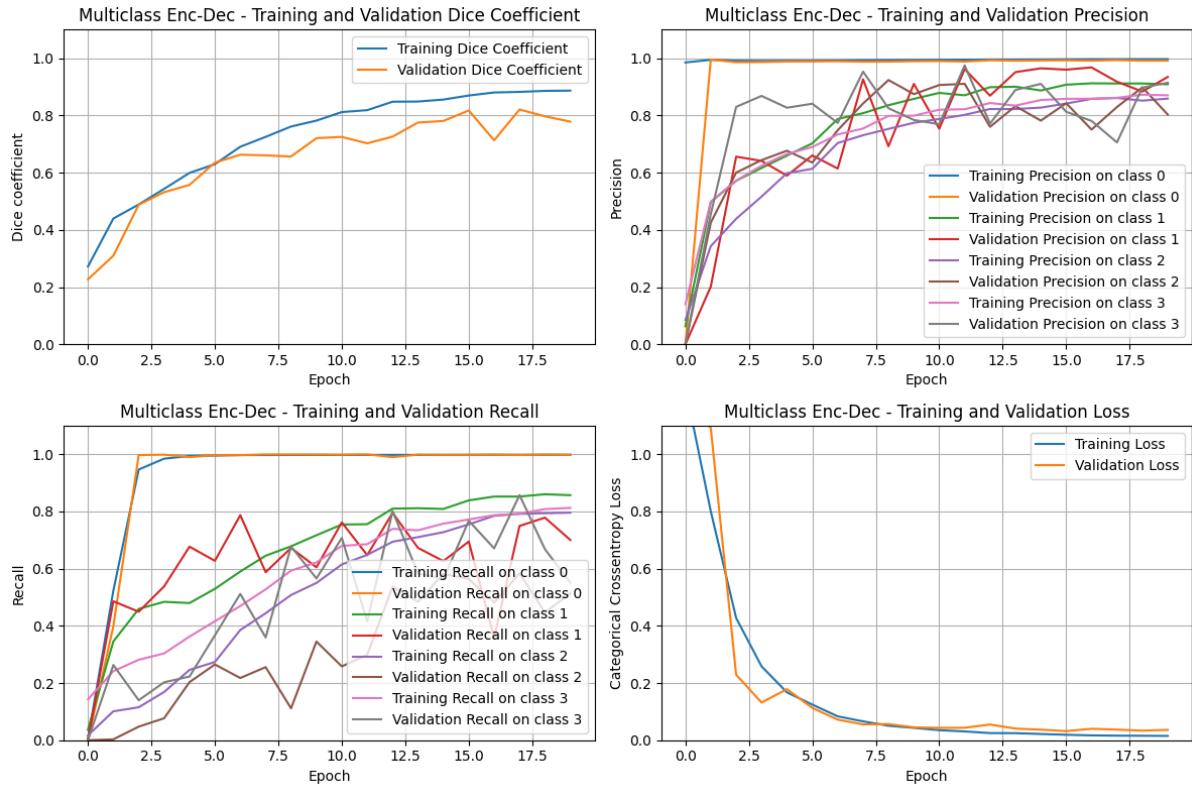


Figure 4.24: Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for normalized dataset.

model. The close alignment between the training and validation loss curves suggests good generalization, despite the fluctuations observed in precision and recall for the tumor classes.

Comparatively, both models exhibit some degree of overfitting, as evidenced by the gap between training and validation metrics. The U-Net model shows a slightly better overall performance in terms of the Dice coefficient and less fluctuation in precision and recall for the tumor classes. However, the encoder-decoder model's performance is comparable, with similarly high precision and recall for the “no tumor” class and reasonable performance for the tumor classes, though with more variability.

Figure 4.25 shows the confusion matrix for normalized test set for U-Net. For this network architecture, the confusion matrix shows that the model achieves a perfect prediction for the “no tumor” class with a probability of 1.00. However, there are minor misclassifications for meningioma, glioma, and pituitary tumor with probabilities of 0.05, 0.16, and 0.09, respectively. For the meningioma class, the U-Net model demonstrates a high accuracy with a correct prediction probability of 0.94, and a minimal misclassification as pituitary tumor with a probability of 0.01. The glioma class is correctly identified with a probability of 0.84, with negligible misclassification as meningioma at 0.01. Finally, the pituitary tumor class is predicted with a probability of 0.90, with no significant misclassification into other categories.

Figure 4.26 shows the confusion matrix for normalized test set for encoder-decoder. This network, in comparison, shows a high accuracy for the “no tumor” class with a correct prediction probability of 0.99, but it also has higher misclassification rates into meningioma, glioma, and pituitary tumor with probabilities of 0.09, 0.26, and 0.11, respectively. For the

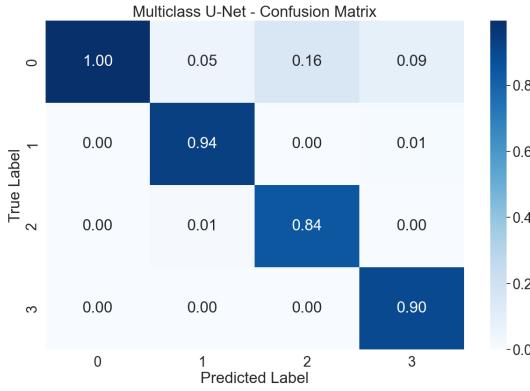


Figure 4.25: U-Net: Confusion matrix for normalized dataset (for test dataset).

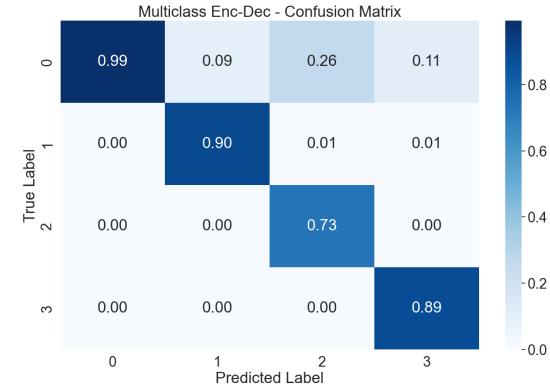


Figure 4.26: Encoder-decoder: Confusion matrix for normalized dataset (for test dataset).

meningioma class, the encoder-decoder network shows a correct prediction probability of 0.90, and minor misclassifications into glioma and pituitary tumor with probabilities of 0.01 each. The glioma class is correctly predicted with a probability of 0.73, indicating a lower accuracy compared to the U-Net architecture, and no significant misclassification into other categories. The pituitary tumor class is predicted correctly with a probability of 0.89, which is slightly lower than the U-Net model's performance.

In general, the U-Net model demonstrates superior performance across all tumor classes compared to the encoder-decoder model. The U-Net achieves higher accuracy in predicting no tumor, meningioma, glioma, and pituitary tumor classes. Additionally, the U-Net model shows fewer misclassifications, suggesting it is more effective in accurately segmenting and identifying different tumor types. These findings indicate that the U-Net model is a more reliable choice for this specific task of tumor detection and segmentation.

Prediction visualization using U-Net for sample images for normalized dataset is presented in Figure 4.27. Each row corresponds to a different tumor type class, with the reference image, predicted mask, and ground truth displayed side by side for comparison.

In the first row, the reference image shows an MRI scan with a visible meningioma tumor. The model's predicted mask identifies a region as a meningioma tumor, which closely matches the ground truth, indicating accurate identification. In the second row, the reference image depicts an MRI scan with a glioma tumor. The model's predicted mask highlights the glioma tumor region, closely aligning with the ground truth and suggesting correct identification. In the third row, the reference image shows an MRI scan with a pituitary tumor. The predicted mask from the model marks the pituitary tumor region accurately, matching the ground truth well. The alignment of the areas recognized as tumors in both the predicted mask and ground truth indicates that the model can handle different tumor sizes and shapes and still provide excellent segmentation.

Overall, the visualization demonstrates that the U-Net network performs effectively in detecting and classifying different types of brain tumors, as the predicted masks align closely with the ground truth for each tumor type.

The prediction visualization for the encoder-decoder architecture is presented in Figure 4.28. Similarly, it is again displayed across three different types of tumor classes, with each row representing a distinct class.

In the first row, the reference image shows an MRI scan with a visible meningioma tumor. The predicted mask from the model identifies a region as a meningioma tumor,

Multiclass U-Net - Prediction Visualization

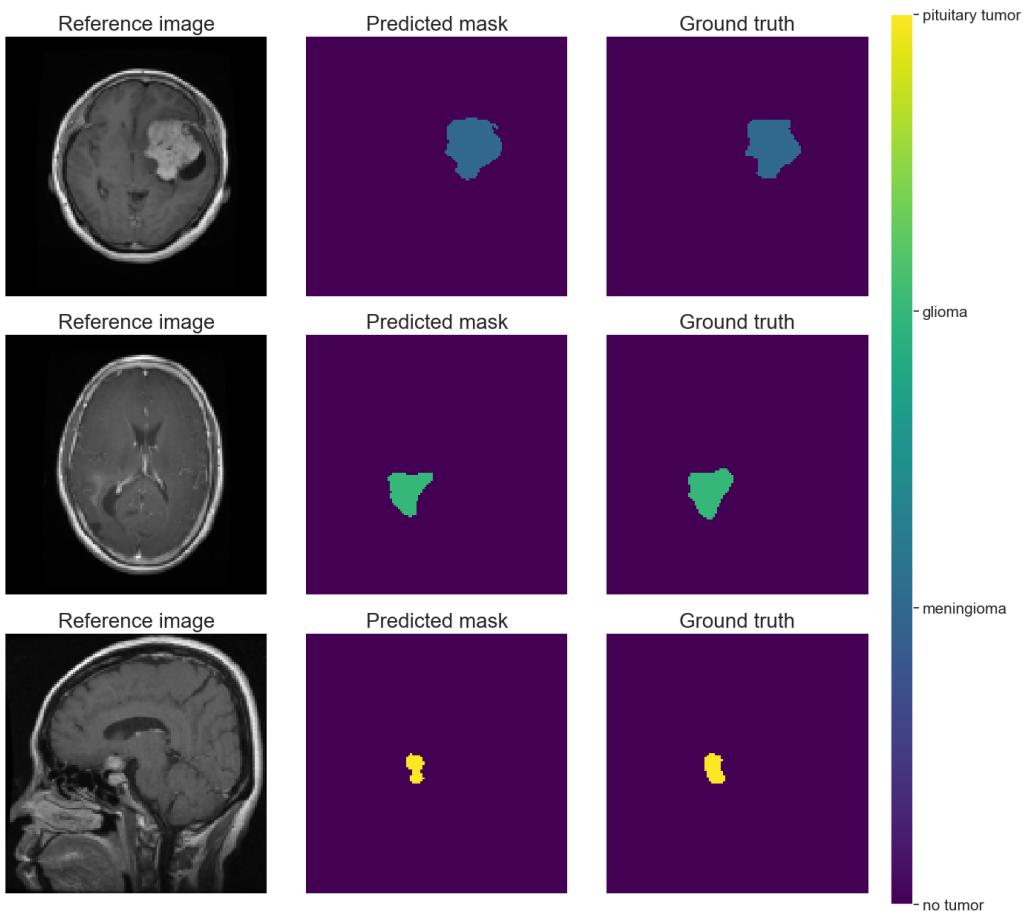


Figure 4.27: U-Net: Prediction visualization for sample images in each tumor type class for normalized dataset.

which partially matches the ground truth, indicating some level of accuracy but with noticeable differences. Compared to the previous U-Net network, the meningioma tumor prediction is less precise, with the encoder-decoder failing to capture the full extent of the tumor as accurately. In the second row, the reference image illustrates an MRI scan with a glioma tumor. The model's predicted mask highlights the glioma tumor region, which closely aligns with the ground truth, suggesting a good level of accuracy. This is similar to the performance of the U-Net model, where the glioma tumor was also accurately identified. In the third row, the reference image shows an MRI scan with a pituitary tumor. The predicted mask from the model marks the pituitary tumor region, which accurately matches the ground truth, indicating effective detection. This performance is comparable to the U-Net, which also demonstrated accurate identification of the pituitary tumor.

Overall, the encoder-decoder architecture shows strong performance in detecting tumors, similar to the U-Net network. However, the U-Net model provides a more precise prediction compared to the encoder-decoder, which shows some inaccuracies in capturing the full extent of the tumor. This suggests that while both models are effective, the U-Net model may have a slight edge in accuracy for certain tumor types. It is essential to mention that these are only sample images for a single run, therefore variations may appear throughout the whole dataset.

Multiclass Enc-Dec - Prediction Visualization

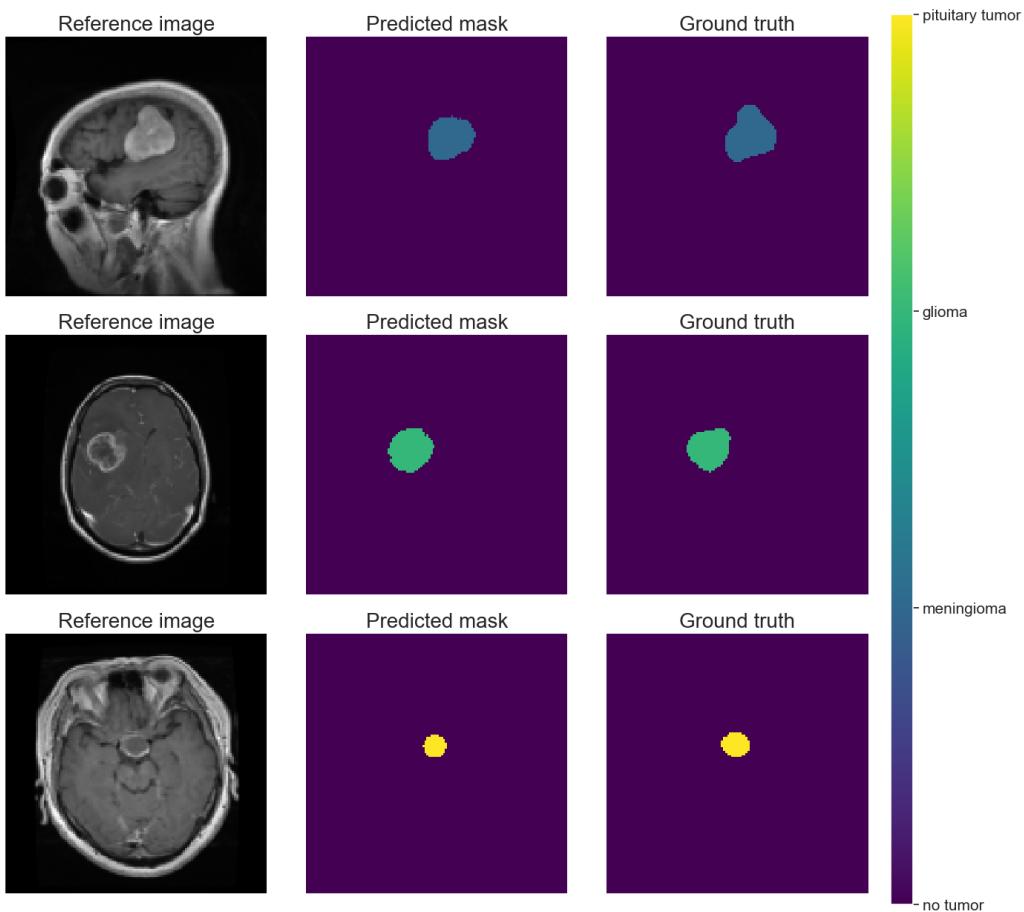


Figure 4.28: Encoder-decoder: Prediction visualization for sample images in each tumor type class for normalized dataset.

Augmented dataset

Figure 4.29 displays the training and validation history for the augmented dataset using U-Net, showcasing the Dice coefficient (top left), precision (top right), recall (bottom left), and loss function (bottom right). The plots indicate the effective learning process of the network, as evidenced by the increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

Both the training and validation Dice coefficients improve significantly during the initial epochs, with the training Dice coefficient reaching around 0.95 by the end. However, the validation Dice coefficient, while also improving, is consistently lower than the training Dice coefficient, stabilizing around 0.9. This gap suggests a relatively minor degree of overfitting.

The precision plot reveals high precision for class 0 (no tumor) throughout the training, while the precision for other classes (meningioma, glioma, pituitary tumor) fluctuates initially but stabilizes at high values. The validation precision for these classes follows a similar trend but with more noticeable fluctuations, indicating variability in the model's ability to accurately classify these tumor types during validation. Recall for class 0 remains at the value of 1.0 throughout, indicating perfect recall for the "no tumor" class in both training and validation. For classes 1, 2, and 3, the recall improves significantly in the early epochs and stabilizes at high values, though there is slightly more variability compared

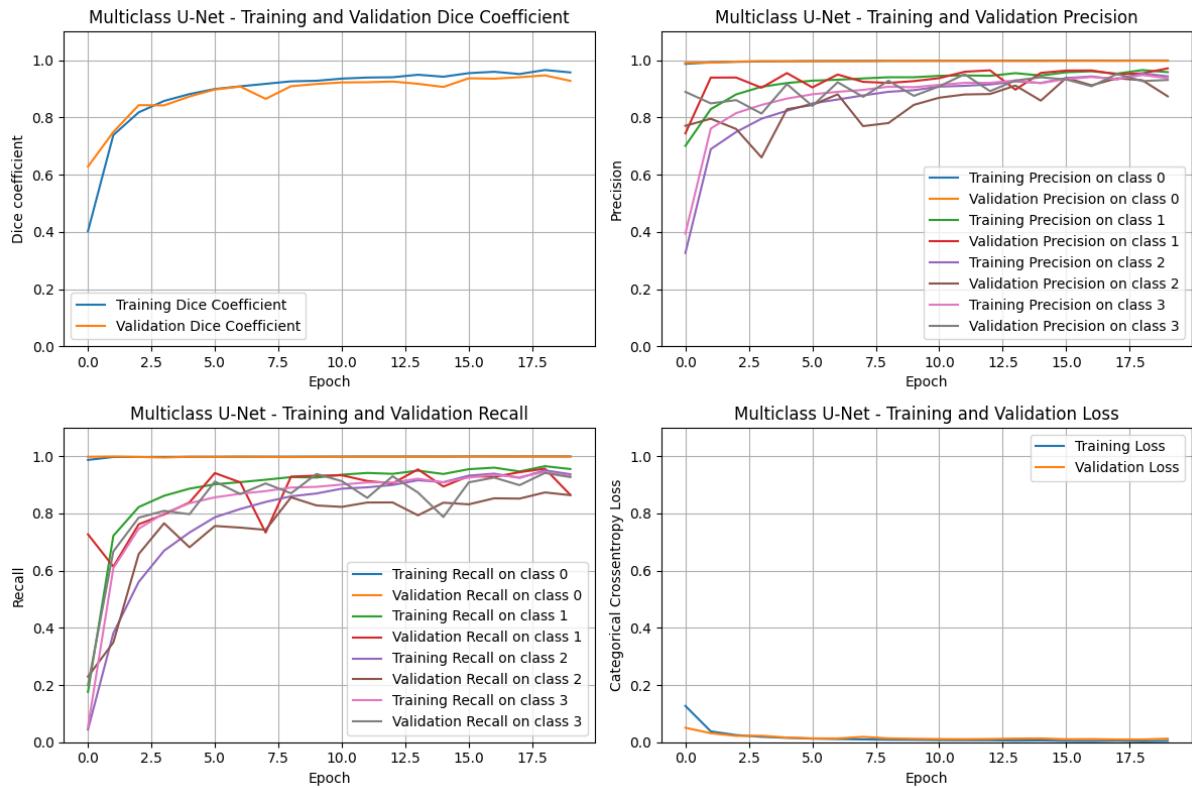


Figure 4.29: U-Net: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.

to the precision plot. This suggests that the model is generally good at identifying the presence of tumors, but may occasionally miss some instances during validation.

The loss plot indicates a steady decrease in both training and validation loss, converging to a low value by the end of training. The close alignment between the training and validation loss curves suggests that the model generalizes well to the validation set.

Overall, the model performs reasonably well, especially in terms of precision and recall for the “no tumor” class, with reasonable performance for the tumor classes. The main area of concern is the overfitting indicated by the difference in Dice coefficients and the variability in precision and recall for tumor classes during validation.

Moreover, it is noteworthy that we observed fewer fluctuations, higher values of metrics and smaller gaps between training and validation curves in the augmented set compared to the normalized version (Figure 4.29). The presence of additional data results in enhanced network stability and superior performance.

Figure 4.30 displays the training and validation history for the augmented dataset using encoder-decoder network, showcasing the previously described metrics. The plots indicate the effective learning process of the network, as evidenced by the increasing Dice coefficient, precision, and recall over epochs, along with a decreasing loss.

The Dice coefficient plot shows a steady increase for both training and validation sets, with the training reaching around 0.95 and the validation reaching about 0.9. While there is a slightly bigger gap between training and validation Dice coefficients, the overall pattern is similar to the U-Net model, indicating some degree of overfitting.

In the precision plot, class 0 achieves high precision early on and remains stable, similar to the U-Net model. For the other classes (meningioma, glioma, pituitary tumor), precision

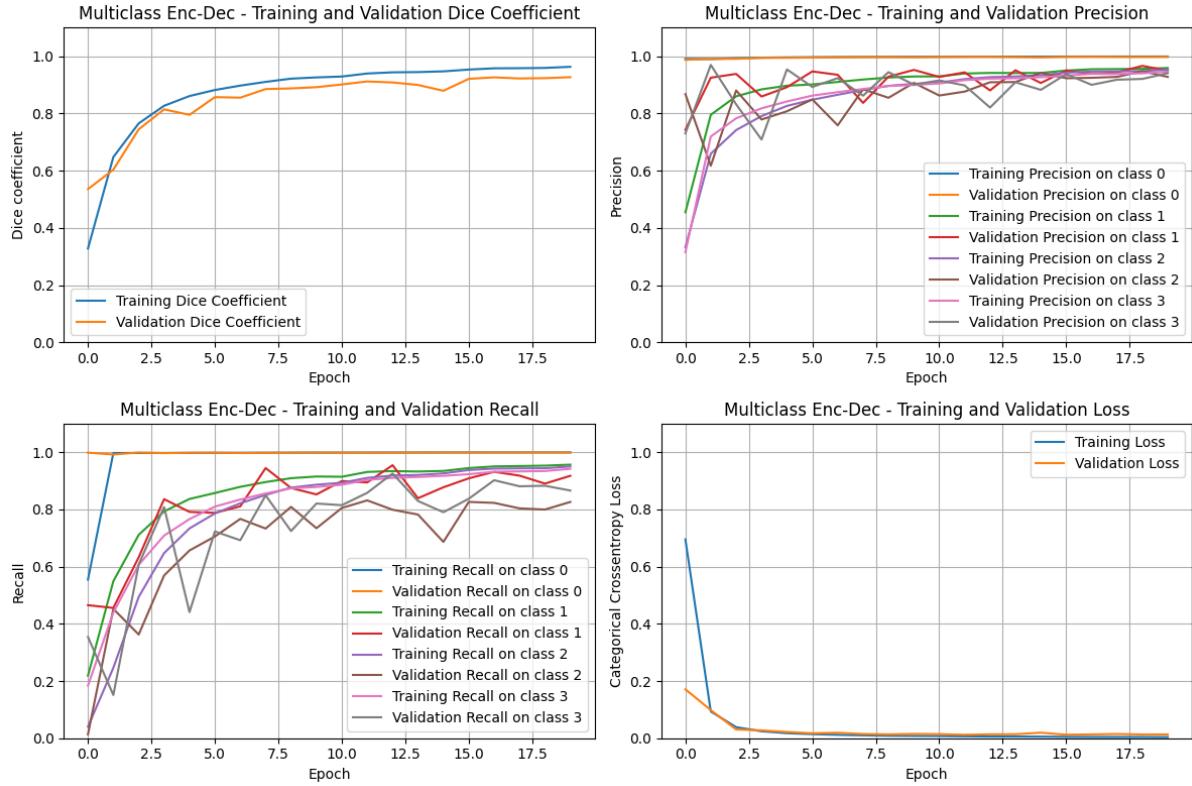


Figure 4.30: Encoder-decoder: Plots of training and validation history; including Dice coefficient, precision, recall, and loss for augmented dataset.

improves significantly after the initial epochs and stabilizes, though with slightly more fluctuations compared to the U-Net. The validation precision generally follows the training precision trends but with noticeable variability, suggesting inconsistent performance in the validation set. Recall for class 0 remains very high throughout the training, similar to the U-Net model. Recall for the other classes increases during the initial epochs and then stabilizes, though it shows more variability than precision. The validation recall is generally lower than the training recall for tumor classes, indicating that the model may be missing some tumor instances during validation, which is consistent with the observed variability.

The loss plot indicates a decrease in both training and validation loss during the initial epochs, converging to low values and staying stable, much like in the U-Net model. The close alignment between the training and validation loss curves suggests good generalization, despite the fluctuations observed in precision and recall for the tumor classes.

Comparatively, both models exhibit some degree of overfitting, as evidenced by the gap between training and validation metrics. The U-Net model shows a slightly better overall performance in terms of the Dice coefficient and less fluctuation in precision and recall for the tumor classes. However, the encoder-decoder model's performance is comparable, with similarly high precision and recall for the “no tumor” class and reasonable performance for the tumor classes, though with more variability.

Figure 4.31 shows the confusion matrix for augmented test set for U-Net. For this architecture a perfect accuracy of 1.00 in predicting the “no tumor” class (class 0) is obtained, with minor misclassifications into meningioma, glioma, and pituitary tumor at probabilities of 0.03, 0.11, and 0.08, respectively. For the meningioma, the network shows high accuracy with a correct prediction probability of 0.97, and slight misclassifications into glioma with

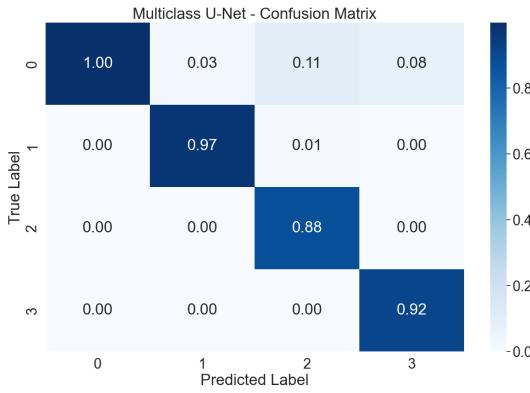


Figure 4.31: U-Net: Confusion matrix for augmented dataset (for test dataset).

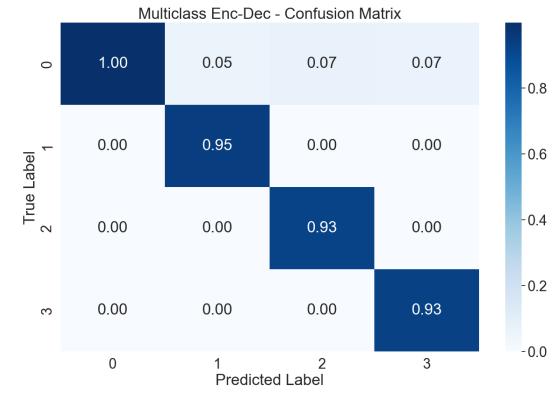


Figure 4.32: Encoder-decoder: Confusion matrix for augmented dataset (for test dataset).

probability equal to 0.01. The glioma is correctly predicted with a probability of 0.88, with no significant misclassification into other classes. Finally, the pituitary tumor has a correct prediction probability of 0.92, indicating robust performance with no misclassification into other tumor classes.

Figure 4.32 shows the confusion matrix for augmented test set for encoder-decoder. This network achieves a perfect accuracy of 1.00 for the “no tumor” class, with minor misclassifications into meningioma, glioma, and pituitary tumor at probabilities of 0.05, 0.07, and 0.07, respectively. For the meningioma class, the encoder-decoder architecture has a correct prediction probability of 0.95, with no significant misclassification into other classes. The glioma class is correctly predicted with a higher probability of 0.93, demonstrating an improved performance compared to the U-Net model for this class. Lastly, the pituitary tumor class shows a correct prediction probability of 0.93, slightly better than the U-Net model’s performance for this class.

In conclusion, both models demonstrate strong performance in tumor detection, with the encoder-decoder showing slightly better accuracy and fewer misclassifications in certain tumor classes compared to the U-Net. While the U-Net performs better in predicting meningioma, the encoder-decoder excels in predicting glioma and pituitary tumors. These results suggest that the encoder-decoder model might be more effective overall in accurately segmenting and identifying different tumor types, although both models have their respective strengths. It is also essential to point that an overall performance is better for the augmented dataset than for the normalized one for both networks (Figures 4.25 and 4.26).

Prediction visualization using U-Net for sample images for augmented dataset is presented in Figure 4.33. Each row corresponds to a different tumor type class, with the reference image, predicted mask, and ground truth displayed side by side for comparison.

In the first row, the reference image shows an MRI scan with a visible meningioma tumor. The model’s predicted mask identifies a region as a meningioma tumor, which closely matches the ground truth, indicating correct identification. In the second row, the reference image depicts an MRI scan with a glioma tumor. The model’s predicted mask highlights the glioma tumor region, closely aligning with the ground truth and indicating correct identification. In the third row, the reference image shows an MRI scan with a pituitary tumor. The predicted mask from the model marks the pituitary tumor region accurately, matching the ground truth well. The alignment of the areas recognized

Multiclass U-Net - Prediction Visualization

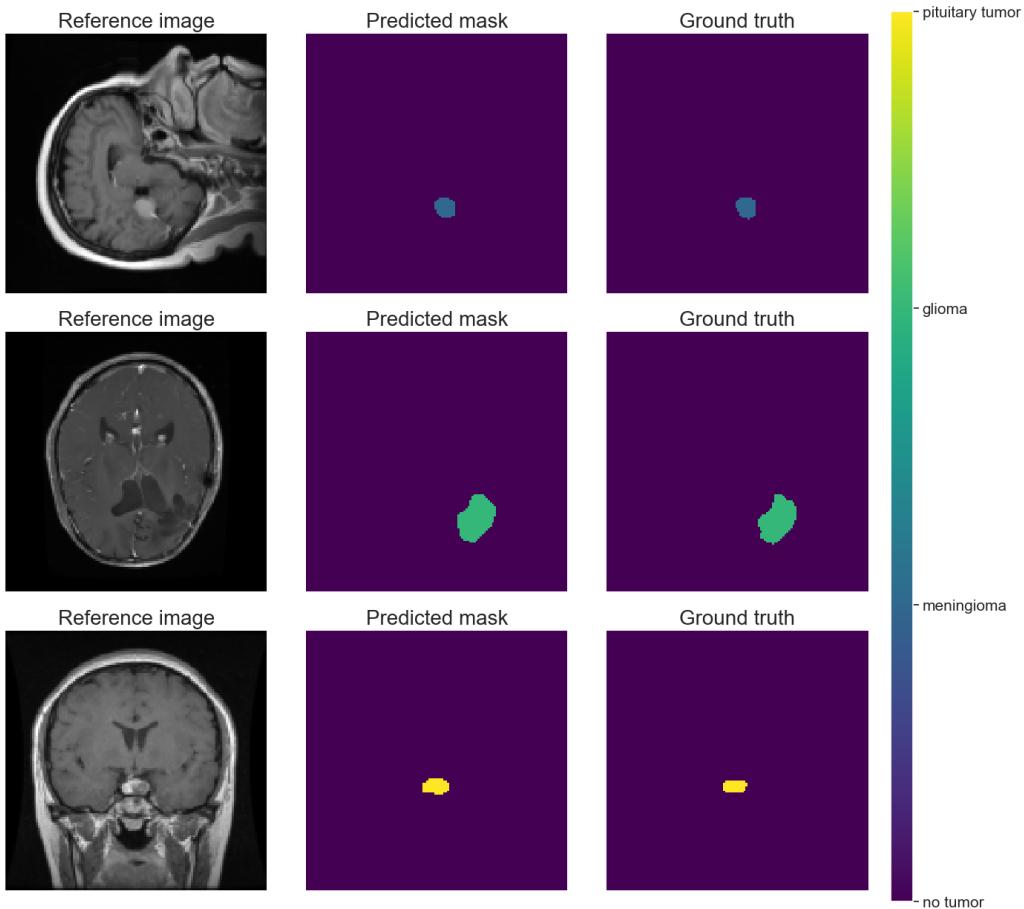


Figure 4.33: U-Net: Prediction visualization for sample images in each tumor type class for augmented dataset.

as tumors in both the predicted mask and ground truth indicates that the model can handle different tumor sizes and shapes and still provide highly accurate segmentation.

Overall, the visualization demonstrates that the U-Net network performs effectively in detecting and classifying different types of brain tumors, as the predicted masks align closely with the ground truth for each tumor type.

The prediction visualization for the encoder-decoder architecture is presented in Figure 4.34. Similarly, it is again displayed across three different types of tumor classes, with each row representing a distinct class.

In the first row, the reference image shows an MRI scan with a visible meningioma tumor. The predicted mask from the model identifies a region as a meningioma tumor, which closely matches the ground truth, indicating a superior level of accuracy. Compared to the previous network, the meningioma tumor prediction is similarly highly accurate. In the second row, the reference image illustrates an MRI scan with a glioma tumor. The model's predicted mask highlights the glioma tumor region, which quite closely aligns with the ground truth, suggesting a good level of accuracy. However, compared to U-Net, this prediction is less precise, with the encoder-decoder failing to capture the full extent of the tumor as accurately. In the third row, the reference image shows an MRI scan with a pituitary tumor. The predicted mask from the model marks the pituitary tumor

Multiclass Enc-Dec - Prediction Visualization

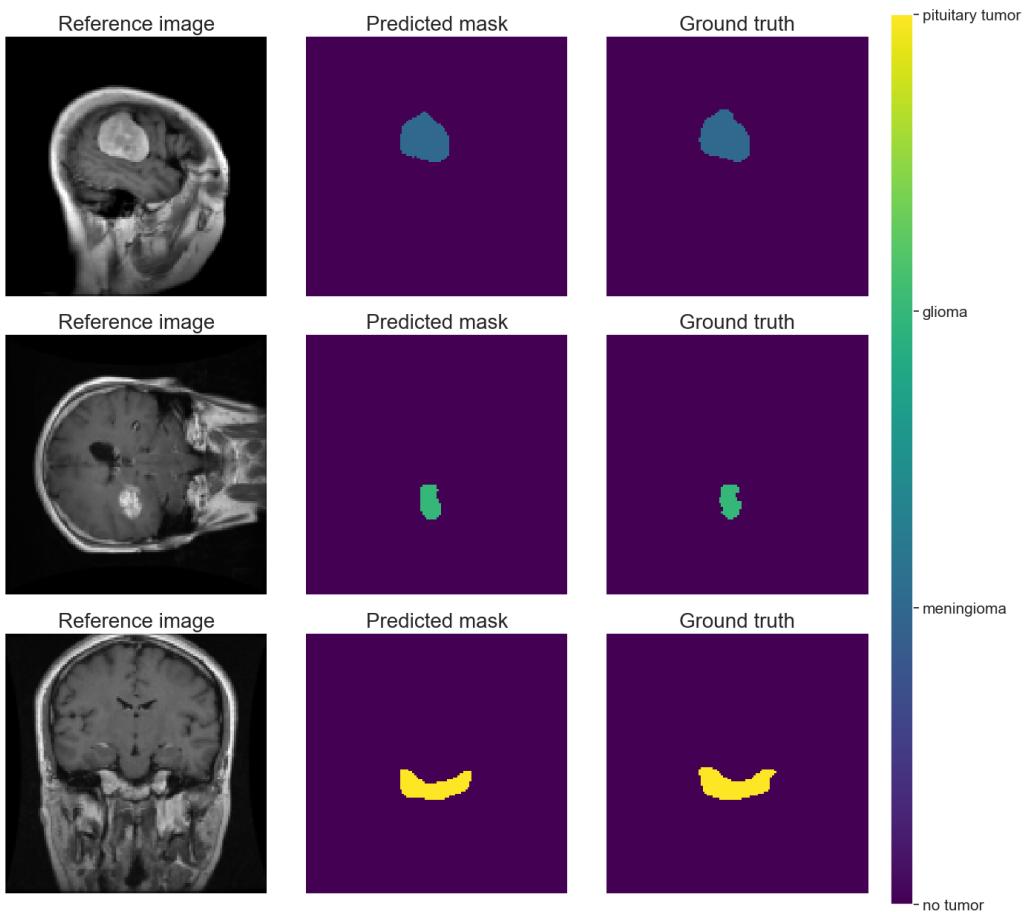


Figure 4.34: Encoder-decoder: Prediction visualization for sample images in each tumor type class for augmented dataset.

region, which accurately matches the ground truth, indicating effective detection. This performance is comparable to the U-Net, which also demonstrated accurate identification of the pituitary tumor.

Overall, the encoder-decoder architecture shows strong performance in detecting tumors, similar to the U-Net network. However, the U-Net model provides a more precise prediction compared to the encoder-decoder, which shows some inaccuracies in capturing the full extent of the tumor. This suggests that while both models are effective, the U-Net network may have a slight edge in accuracy for certain tumor types. It is essential to mention that these are only sample images for a single run, therefore variations may appear throughout the whole dataset.

4.6 Performance comparison table

Table 4.1 presents metrics for U-Net and encoder-decoder networks across two different datasets: normalized and augmented. The analysis of metrics provides insight into their performance for multi-class tumor segmentation, where class 1 represents meningioma, class 2 represents glioma, class 3 represents pituitary tumor, and class 0 represents “no tumor”. Each network is analyzed based on loss, Dice coefficient, balanced accuracy, recall, and precision for all classes.

Table 4.2: **Table of metrics for U-Net and encoder-decoder for all considered datasets in multi-class segmentation.**

	U-Net		Encoder-decoder	
	Normalized	Augmented	Normalized	Augmented
Loss	0.01889	0.01101	0.03582	0.01187
Dice coefficient	0.88607	0.93126	0.76772	0.92905
Balanced accuracy	0.89919	0.94765	0.79125	0.93903
Recall (0)	0.99837	0.99855	0.99776	0.99889
Recall (1)	0.91697	0.88116	0.67986	0.91222
Recall (2)	0.66228	0.88658	0.50258	0.86658
Recall (3)	0.85503	0.91480	0.57273	0.85845
Precision (0)	0.99610	0.99825	0.99300	0.99786
Precision (1)	0.93822	0.97262	0.91008	0.94532
Precision (2)	0.84512	0.88191	0.73309	0.92662
Precision (3)	0.90093	0.91888	0.88929	0.93172

For the U-Net network on the normalized dataset, the loss is 0.01889, with a Dice coefficient of 0.88607, indicating robust segmentation performance. The balanced accuracy is 0.89919, suggesting a high overall recall across all classes (as it is an average of all recalls). The recall for class 0 is exceptionally high at 0.99837, showing that almost all non-tumor regions are correctly identified. For class 1, the recall is 0.91697, reflecting strong performance in detecting this tumor type. However, the recall for class 2 is lower at 0.66228, indicating some challenges in identifying gliomas. Class 3 has a recall of 0.85503, showing good but not perfect detection. Precision values for all classes are high, particularly for class 0 (0.99610) and class 1 (0.93822), indicating accurate predictions.

On the augmented dataset, U-Net shows improved performance with a reduced loss of 0.01101 and a higher Dice coefficient of 0.93126. The balanced accuracy increases to 0.94765, demonstrating enhanced overall performance. Recall for class 0 remains high at 0.99855. Recall for class 1 is slightly lower at 0.88116 compared to the normalized dataset, but recall for class 2 significantly improves to 0.88658, showing better glioma detection. Recall for class 3 also improves to 0.91480. Precision for all classes remains high, with class 1 reaching 0.97262, indicating more accurate detection and fewer false positives.

For the encoder-decoder network on the normalized dataset, the loss is higher at 0.03582, and the Dice coefficient is lower at 0.76772, indicating less effective segmentation compared to U-Net. The balanced accuracy is 0.79125. Recall for class 0 is high at 0.99776, but for class 1 is lower at 0.67986, showing difficulties in detecting meningiomas. Recall for class

2 is very low at 0.50258, highlighting significant challenges in identifying gliomas. Recall for class 3 is also low at 0.57273. Precision values for all classes are lower than those for U-Net, particularly for class 2 (0.73309) and class 3 (0.88929), reflecting a higher FP rate.

On the augmented dataset, the encoder-decoder network shows substantial improvement. The loss decreases to 0.01187, and the Dice coefficient rises to 0.92905. The balanced accuracy is 0.93903, indicating a high overall recall. Recall for class 0 is very high at 0.99889. Recall for class 1 improves significantly to 0.91222, and for class 2 to 0.86658, demonstrating better performance in tumor detection than when a normalized dataset was used. Recall for class 3 also improves to 0.85845. Precision values for all classes increase, with class 2 reaching 0.92662 and class 3 at 0.93172, indicating more accurate predictions with fewer false positives cases.

Comparing the two networks for each dataset, U-Net generally outperforms encoder-decoder. On the normalized dataset, U-Net has a lower loss, higher Dice coefficient, and higher balanced accuracy. Recall and precision for all tumor classes are higher in U-Net, indicating better performance in detecting and accurately segmenting tumors. On the augmented dataset, both networks show improved performance, but U-Net still leads with slightly better (or insignificantly worse) metrics across all categories. U-Net achieves lower loss, higher Dice coefficient, and higher balanced accuracy. Recall and precision for the tumor classes are in general also better or only marginally worse in U-Net network compared to the encoder-decoder.

In conclusion, the best configuration for multi-class tumor segmentation is the U-Net model trained on the augmented dataset. This combination results in the lowest loss, highest Dice coefficient, and highest balanced accuracy. The recall and precision for all tumor classes are also remarkably high in this configuration, ensuring accurate and reliable tumor detection and segmentation. The encoder-decoder network also performs well on the augmented dataset but does not match the performance metrics of U-Net. Therefore, U-Net with data augmentation is the optimal choice for this multi-class tumor segmentation task.

It is noteworthy that the training and testing was performed only once for both networks. That being said, variations in values of all metrics may appear. Performing numerous independent experiments and calculating the mean values would be beneficial.

Conclusion

In this thesis, we explored the effectiveness of Deep Convolutional Neural Networks (DCNNs) for tumor segmentation in brain MRI images. Two architectures were implemented and evaluated: the U-Net and the encoder-decoder network. Using a dataset of brain MRI images, we conducted experiments employing diverse preprocessing techniques, including the utilization of normalization and data augmentation. The networks were trained and evaluated on binary classification (tumor and “no tumor”) and multi-class classification (“no tumor”, meningioma, glioma, and pituitary tumor). Each training session lasted the same number of epochs, with equal hyperparameters.

The analysis of real-world data for tumor segmentation highlights several important observations. In the binary segmentation task, U-Net consistently demonstrated superior performance across a variety of datasets compared to the encoder-decoder network. On the original dataset, U-Net exhibited a steadily increasing Dice coefficient, precision, and recall over epochs, with a decreasing loss function. This indicated an effective learning process. Although the encoder-decoder network also showed improvements, its learning was slower and less robust, resulting in lower performance metrics. When trained on the normalized dataset, both U-Net and encoder-decoder showed improved performance. However, U-Net again outperformed encoder-decoder, particularly in terms of generalization and stability. The normalized dataset yielded higher overall performance for both networks compared to the original dataset, with U-Net maintaining a lead in all key metrics. For the augmented dataset, which provided a larger and more varied training set, both models performed well, but the encoder-decoder showed marginally better stability and performance. Nonetheless, U-Net’s overall performance metrics remained slightly higher, suggesting it as the more reliable model for accurate tumor segmentation.

The multi-class segmentation analysis mirrored the binary results, with U-Net generally outperforming the encoder-decoder network. Both models showed some degree of overfitting, but U-Net’s performance in terms of the Dice coefficient and the stability of precision and recall across tumor classes was superior. On the normalized dataset, U-Net demonstrated higher accuracy in predicting various tumor types, with fewer misclassifications. The augmented dataset further improved the performance of both models, reducing fluctuations and enhancing stability. U-Net’s higher metrics across all tumor classes indicated its effectiveness and reliability.

Overall, the U-Net model trained on the augmented dataset emerged as the best configuration for both binary and multi-class tumor segmentation. This combination resulted in the lowest loss, highest Dice coefficient, and highest balanced accuracy, ensuring accurate and reliable tumor detection and segmentation. The encoder-decoder network performed well, however it was not able to match U-Net’s performance metrics.

Based on these findings, we believe that U-Net is the most appropriate choice for brain tumor segmentation tasks. Its superior accuracy and robustness highlight its potential in clinical applications. We are confident that utilizing Deep Convolutional Neural Networks,

particularly U-Net, can significantly enhance patient care by providing more precise and reliable tumor detection. This advancement could lead to earlier diagnoses, more effective treatment plans, and improved overall outcomes for patients.

For further research, it would be beneficial to perform multiple independent runs to average out any variations in the performance metrics. Additionally, exploring other architectures and advanced data augmentation techniques could provide further insights into improving segmentation accuracy and robustness. Integrating more sophisticated post-processing methods and incorporating additional clinical data could also enhance the model's applicability and performance in real-world scenarios.

Bibliography

- [1] ABDUSALOMOV, A., MUKHIDDINOV, M., WHANGBO, T. K. Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers* 15 (08 2023), 4172.
- [2] ADLUNG, L., COHEN, Y., MOR, U., ELINAV, E. Machine learning in clinical decision making. *Med* 2 (2021), 642–665.
- [3] AVENDI, M., KHERADVAR, A., JAFARKHANI, H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri. *Medical Image Analysis* 30 (12 2015).
- [4] BALACHANDAR, N., CHANG, K., KALPATHY-CRAMER, J., RUBIN, D. Accounting for data variability in multi-institutional distributed deep learning for medical imaging. *Journal of the American Medical Informatics Association : JAMIA* 27 (03 2020).
- [5] BARKADE, G., BHOSALE, P., SHIRSATH, S. Overview of brain cancer, its symptoms, diagnosis and treatment. *IP International Journal of Comprehensive and Advanced Pharmacology* 8 (2023), 159–164.
- [6] BHOSALE, S., PATIL, S. Machine learning applications in medical image analysis.
- [7] BREBISSON, A., MONTANA, G. Deep neural networks for anatomical brain segmentation.
- [8] CAI, Y., LAIDLEY, D., KORNECKI, A., SHMUILOVICH, O., LUM, P. A., LI, S. Multi-modal vertebrae recognition using transformed deep convolution network. *Computerized Medical Imaging and Graphics* 51 (04 2016).
- [9] CHEN, Y., LI, L., LI, W., GUO, Q., DU, Z., XU, Z. Chapter 2 - fundamentals of neural networks. In *AI Computing Systems*, Y. Chen, L. Li, W. Li, Q. Guo, Z. Du, and Z. Xu, Eds. Morgan Kaufmann, 2024, pp. 17–51.
- [10] CHIEFFO, D., LINO, F., FERRARESE, D., BELELLA, D., DELLA PEPA, G., DOGLIETTO, F. Brain tumor at diagnosis: From cognition and behavior to quality of life. *Diagnostics* 13 (2023), 1–12.
- [11] CHO, K., MERRIENBOER, B., BAHDANAU, D., BENGIO, Y. On the properties of neural machine translation: Encoder-decoder approaches.
- [12] CHOI, H., JIN, K. Fast and robust segmentation of the striatum using deep convolutional neural networks. *Journal of Neuroscience Methods* 274 (10 2016).

- [13] CIREAN, D., GIUSTI, A., GAMBARDELLA, L. M., SCHMIDHUBER. Deep neural networks segment neuronal membranes in electron microscopy images. *Proceedings of Neural Information Processing Systems 25* (2012).
- [14] DALMŞ, M., LITJENS, G., HOLLAND, K., SETIO, A., MANN, R., KARSSEMEIJER, N., GUBERN-MÉRIDA, A. Using deep learning to segment breast and fibroglandular tissue in mri volumes. *Medical Physics 44* (12 2016).
- [15] DHAR, T., DEY, N., BORRA, S., SHERRATT, R. Challenges of deep learning in medical image analysis - improving explainability and trust. *IEEE Transactions on Technology and Society PP* (2023), 1–1.
- [16] DHEEPA, G., PL, C. An efficient encoder-decoder cnn for brain tumor segmentation in mri images. *IETE Journal of Research 69* (08 2022), 1–12.
- [17] FAHMI, F., APRIYULIDA, F., KEMALA NASUTION, I., SAWALUDDIN. Automatic detection of brain tumor on computed tomography images for patients in the intensive care unit. *Journal of Healthcare Engineering 2020* (07 2020), 1–13.
- [18] FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters 27*, 8 (2006), 861–874.
- [19] FIDON, L., LI, W., GARCIA-PERAZA-HERRERA, L., EKANAYAKE, J., KITCHEN, N., OURSELIN, S., VERCAUTEREN, T. *Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation Using Holistic Convolutional Networks*. 2018, pp. 64–76.
- [20] FONSECA, P., MENDOZA, J., WAINER, J., FERRER, J., PINTO, J., GUERRERO, J., CASTANEDA, B. Automatic breast density classification using a convolutional neural network architecture search procedure.
- [21] GALIĆ, I., HABIJAN, M., LEVENTIĆ, H., ROMIĆ, K. Machine learning empowering personalized medicine: A comprehensive review of medical image analysis methods. *Electronics 12* (10 2023), 4411.
- [22] GAO, M., LU, L., WU, A., BUTY, M., SHIN, H.-C., ROTH, H., PAPADAKIS, G., DEPEURSINGE, A., SUMMERS, R., XU, Z., MOLLURA, D. Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 6* (06 2016), 1–6.
- [23] GOODFELLOW, I., BENGIO, Y., COURVILLE, A. *Deep Learning*. MIT Press, 2016.
- [24] GRANDINI, M., BAGLI, E., VISANI, G. Metrics for multi-class classification: an overview.
- [25] KATTI, G., ARA, S., SHIREEN, D. Magnetic resonance imaging (mri) - a review. *Intl J Dental Clin 3* (03 2011).
- [26] KER, J., WANG, L., RAO, J., LIM, T. Deep learning applications in medical image analysis. *IEEE Access PP* (2017), 1–1.

- [27] KOREZ, R., LIKAR, B., PERNUŠ, F., VRTOVEC, T. Model-based segmentation of vertebral bodies from mr images with 3d cnns. pp. 433–441.
- [28] LECUN, Y., BENGIO, Y., HINTON, G. Deep learning. *Nature* 521 (2015), 436–44.
- [29] LITJENS, G., KOOI, T., BEJNORDI, B. E., SETIO, A. A. A., CIOMPI, F., GHAFORIAN, M., VAN DER LAAK, J. A., VAN GINNEKEN, B., SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. *Medical Image Analysis* 42 (2017), 60–88.
- [30] MARKOU, M., LAVRENTAKI, A., NTALI, G. Recent advances in understanding and managing pituitary adenomas. *Faculty Reviews* 12 (2023).
- [31] MOHAMMED, F. A., TUNE, K. K., ASSEFA, B. G., JETT, M., MUHIE, S. Medical image classifications using convolutional neural networks: A survey of current methods and statistical modeling of the literature. *Machine Learning and Knowledge Extraction* 6, 1 (2024), 699–735.
- [32] MOHRI, M., ROSTAMIZADEH, A., TALWALKAR, A. *Foundations of Machine Learning*, 2 ed. Adaptive Computation and Machine Learning. MIT Press, 2018.
- [33] NAIK, N., HAMEED, B., SHETTY, D., SWAIN, D., SHAH, M., PAUL, R., AGGARWAL, K., IBRAHIM, S., PATIL, D., SMRITI, K., SHETTY, S., RAI, B., CHLOSTA, P., SOMANI, B. Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? *Frontiers in Surgery* 9 (2022), 862322.
- [34] OGASAWARA, C., PHILBRICK, B., ADAMSON, D. Meningioma: A review of epidemiology, pathology, diagnosis, treatment and future directions. *Biomedicines* 9 (03 2021), 319.
- [35] PAYAN, A., MONTANA, G. Predicting alzheimer’s disease: a neuroimaging study with 3d convolutional neural networks. *ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods, Proceedings* 2 (02 2015).
- [36] PRIETO-ORDAZ, O., RAMIREZ, G., GONZALEZ GURROLA, L. C., LOPEZ-SANTILLAN, J., MONTES, M. Brain tumor segmentation using an encoder-decoder network with a multiscale feature module. pp. 1–6.
- [37] RAJKOMAR, A., LINGAM, S., TAYLOR, A., BLUM, M., MONGAN, J. High-throughput classification of radiographs using deep convolutional neural networks. *Journal of Digital Imaging* 30 (10 2016).
- [38] RANA, M., BHUSHAN, M. Machine learning and deep learning approach for medical image analysis: Diagnosis to detection. *Multimedia Tools and Applications* 82 (12 2022), 1–39.
- [39] RENARD, F., GUEDRIA, S., DE PALMA, N., VUILLERME, N. Variability and reproducibility in deep learning for medical image segmentation. *Scientific Reports* 10 (08 2020).
- [40] RONNEBERGER, O., FISCHER, P., BROX, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer International Publishing, 2015, pp. 234–241.

- [41] SAEEDI, S., REZAYI, S., KESHAVARZ, H., ROSTAM NIAKAN KALHORI, S. Mri-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making* 23 (01 2023).
- [42] SALEHINEJAD, H., VALAEE, S., DOWDELL, T., COLAK, E., BARFETT, J. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks.
- [43] SINGH, A., SENGUPTA, S., LAKSHMINARAYANAN, V. Explainable deep learning models in medical image analysis, 2020.
- [44] SPAMPINATO, C., PALAZZO, S., GIORDANO, D., ALDINUCCI, M., LEONARDI, R. Deep learning for automated skeletal bone age assessment in x-ray images. *Medical image analysis* 36 (02 2017).
- [45] STOCKMAN, G., SHAPIRO, L. G. *Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001.
- [46] VAROQUAUX, G., CHEPLYGINA, V. Machine learning for medical imaging: Methodological failures and recommendations for the future. *npj Digital Medicine* 5 (04 2022), 48.
- [47] WOLTERINK, J., LEINER, T., VIERGEVER, M., ISGUM, I. Automatic coronary calcium scoring in cardiac ct angiography using convolutional neural networks. vol. 9349, pp. 589–596.
- [48] YADAV, S., SAWALE, M. A review on image classification using deep learning. *World Journal of Advanced Research and Reviews* 17 (2023), 480–482.
- [49] YANG, K., WU, Z., ZHANG, H., ET AL. Glioma targeted therapy: Insight into future of molecular approaches. *Molecular Cancer* 21, 1 (2022), 39–39.
- [50] YOON, H., JEONG, Y. J., KANG, H., JEONG, J., KANG, D.-Y. Medical image analysis using artificial intelligence. *Progress in Medical Physics* 30 (01 2019), 49.
- [51] ZHOU, S., NIE, D., ADELI, E., YIN, J., LIAN, J. High-resolution encoder-decoder networks for low-contrast medical image segmentation. *IEEE Transactions on Image Processing PP* (06 2019), 1–1.