Brain CT image Hemorrhage Prediction

MATH7243 XN Project

Hsiao-Yu Peng
Northeastern University

December 2023

Outline

- Introduction
- Dataset
- Model Training & Evaluation
- Discussion

Introduction

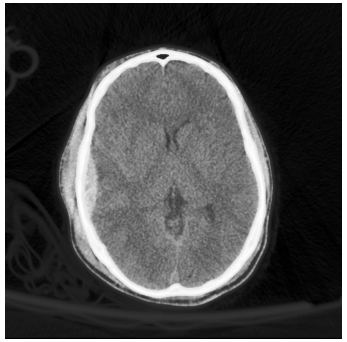
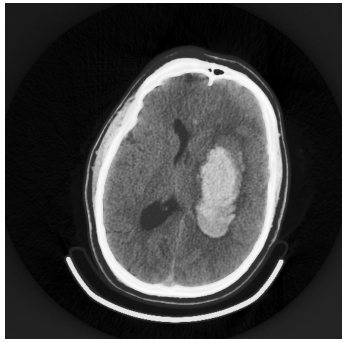
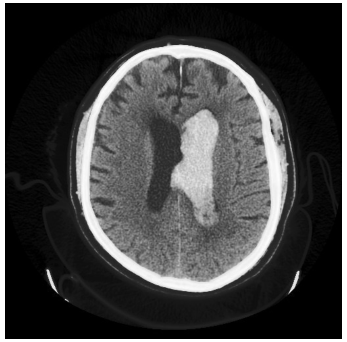
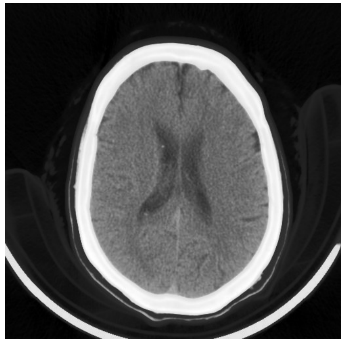
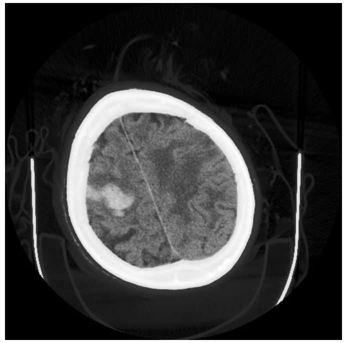
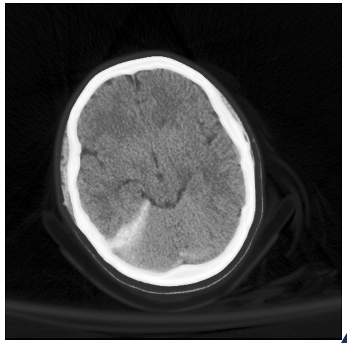
- The early detection of brain hemorrhages is crucial for timely and effective medical intervention.

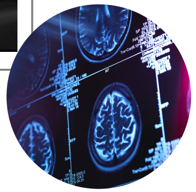
- Challenges include limited access to fast, accurate image guidance and constraints in utilizing it directly at the point-of-care, impacting treatments for brain hemorrhages.

- This project focuses on leveraging the power of machine learning to develop models for classification, regression, and segmentation of hemorrhages in brain CT images.

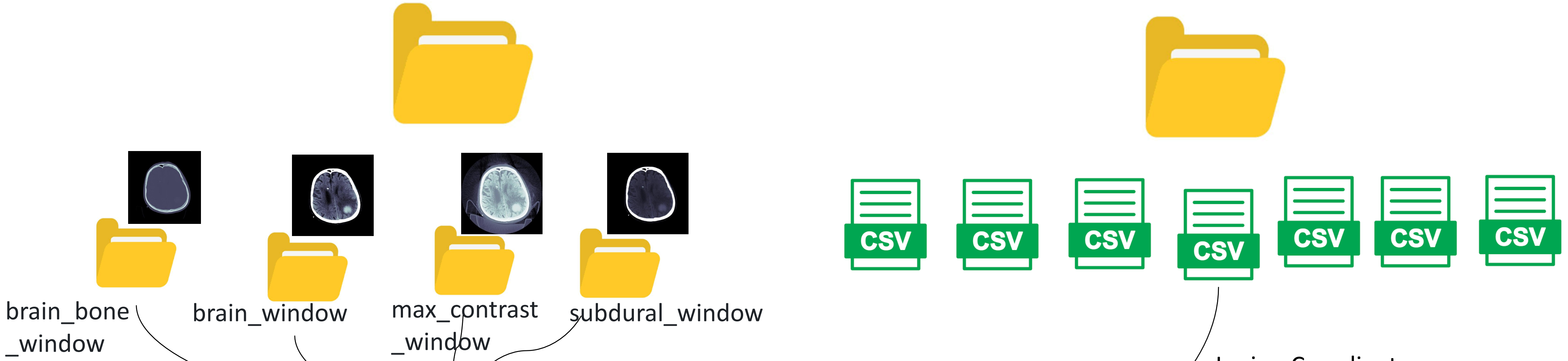
Dataset

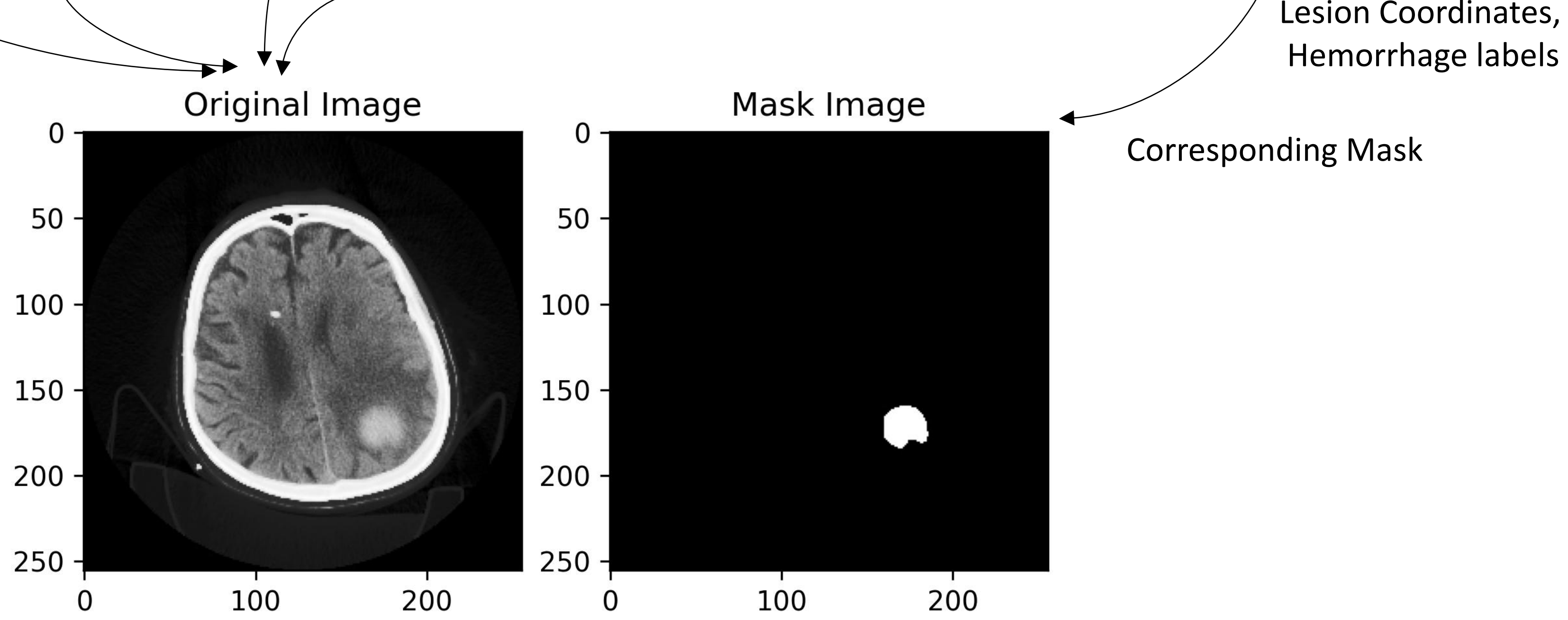
- Experimental Network (XN) Project on MATH 7243
- Partner Company: Zeta Surgical
- Brain Hemorrhage Type:

epidural: 0	Intra-parenchymal: 1	Intra-ventricular: 2	normal: 3	subarachnoid: 4	subdural: 5
					



Data Preparation

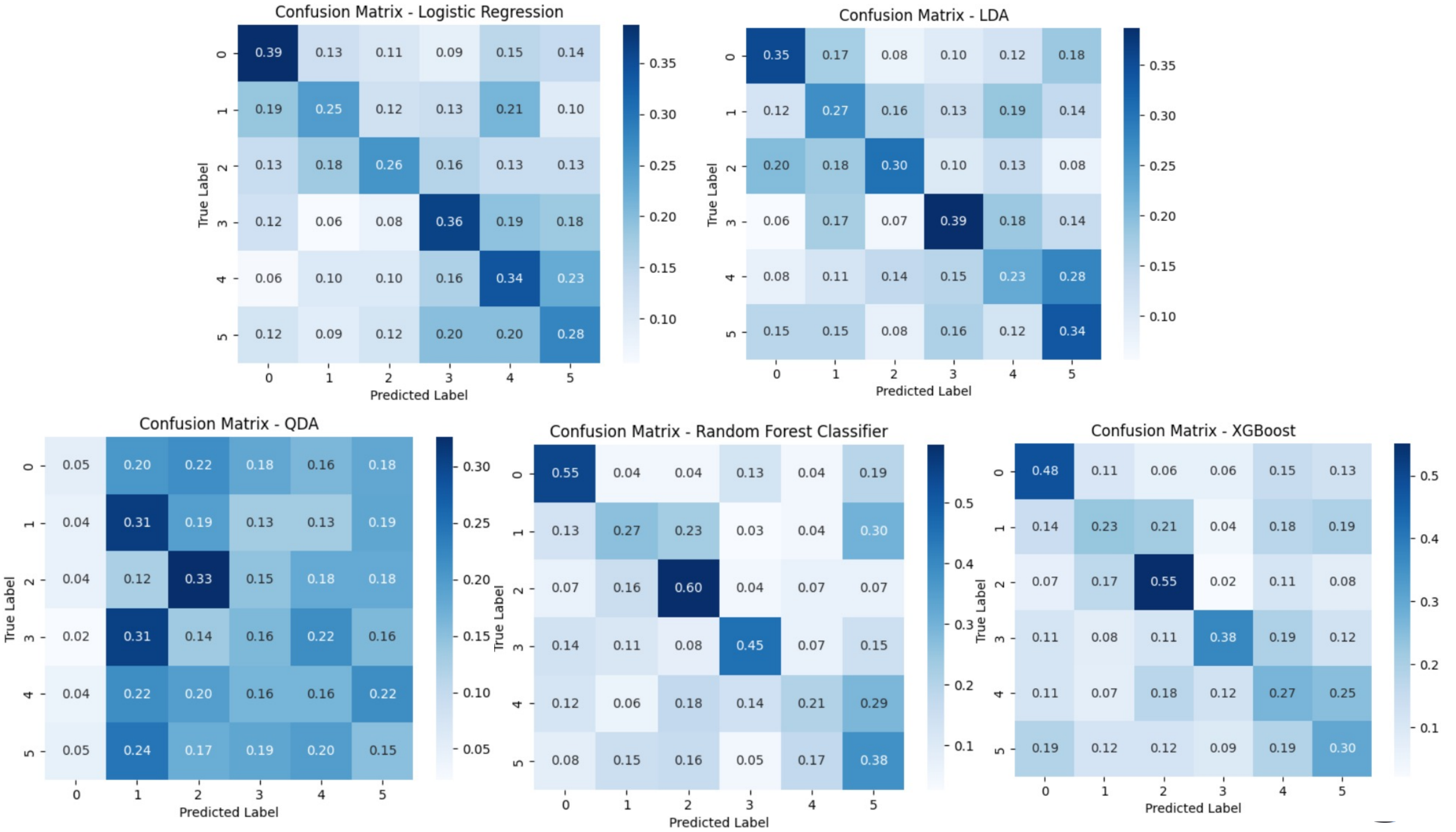




* The raw dicom files have been converted to jpg format by the company.

Traditional Machine Learning Classification Report

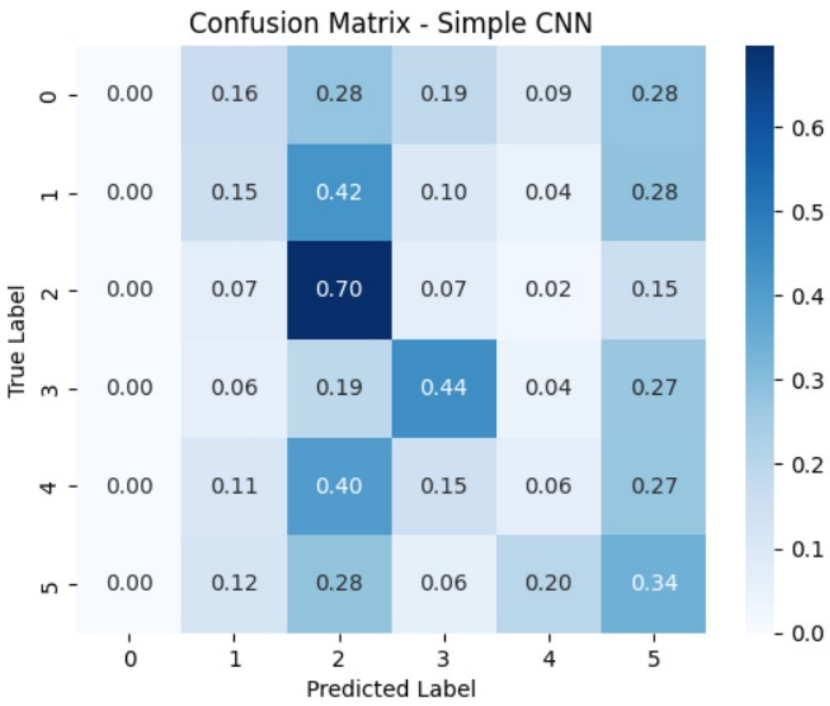
	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.31	0.31	0.31	0.31
LDA	0.31	0.31	0.31	0.31
QDA	0.20	0.19	0.19	0.18
Random Forest Classification	0.40	0.41	0.41	0.40
XGBoost	0.36	0.37	0.37	0.36

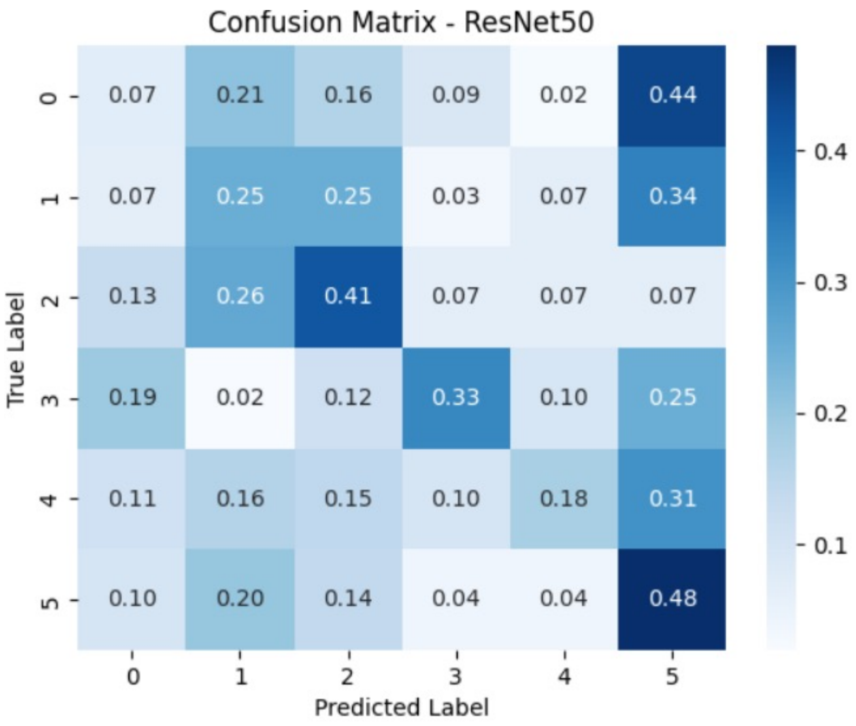

Confusion Matrix - Logistic Regression
0.39 0.13 0.11 0.09 0.15 0.14
0.19 0.25 0.12 0.13 0.21 0.10
0.13 0.18 0.26 0.16 0.13 0.13
0.12 0.06 0.08 0.36 0.19 0.18
0.06 0.10 0.10 0.16 0.34 0.23
0.12 0.09 0.12 0.20 0.20 0.28
True Label
Predicted Label
Confusion Matrix - LDA
0.35 0.17 0.08 0.10 0.12 0.18
0.12 0.27 0.16 0.13 0.19 0.14
0.20 0.18 0.30 0.10 0.13 0.08
0.06 0.17 0.07 0.39 0.18 0.14
0.08 0.11 0.14 0.15 0.23 0.28
0.15 0.15 0.08 0.16 0.12 0.34
True Label
Predicted Label
Confusion Matrix - QDA
0.05 0.20 0.22 0.18 0.16 0.18
0.04 0.31 0.19 0.13 0.13 0.19
0.04 0.12 0.33 0.15 0.18 0.18
0.02 0.31 0.14 0.16 0.22 0.16
0.04 0.22 0.20 0.16 0.16 0.22
0.05 0.24 0.17 0.19 0.20 0.15
True Label
Predicted Label
Confusion Matrix - Random Forest Classifier
0.55 0.04 0.04 0.13 0.04 0.19
0.13 0.27 0.23 0.03 0.04 0.30
0.07 0.16 0.60 0.04 0.07 0.07
0.14 0.11 0.08 0.45 0.07 0.15
0.12 0.06 0.18 0.14 0.21 0.29
0.08 0.15 0.16 0.05 0.17 0.38
True Label
Predicted Label
Confusion Matrix - XGBoost
0.48 0.11 0.06 0.06 0.15 0.13
0.14 0.23 0.21 0.04 0.18 0.19
0.07 0.17 0.55 0.02 0.11 0.08
0.11 0.08 0.11 0.38 0.19 0.12
0.11 0.07 0.18 0.12 0.27 0.25
0.19 0.12 0.12 0.09 0.19 0.30
True Label
Predicted Label

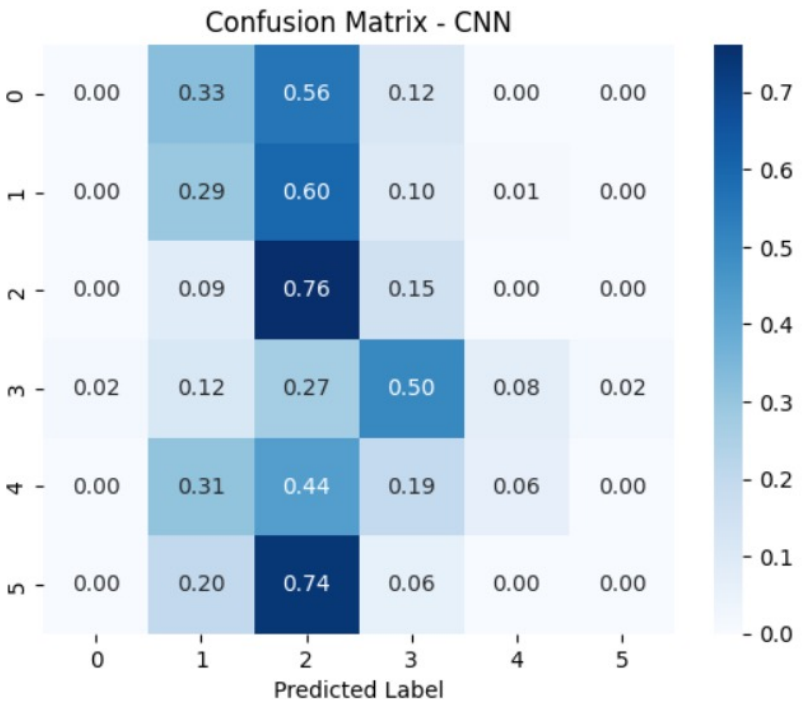
CNNs Classification Report

	Accuracy	Precision	Recall	F1-score
Simple CNN	0.26	0.23	0.28	0.22
ResNet50	0.28	0.30	0.29	0.27
CNN (10 layers)	0.27	0.23	0.27	0.20

CNNs Confusion Matrix







Class Mapping:
epidural: 0, intraparenchymal: 1, intraventricular: 2, normal: 3, subarachnoid: 4, subdural: 5

Image Segmentation

- U-net

- Pilot Run: Utilized 4 types of image data sets. “Epidural, ”Intraparenchymal", ”Subdural", ”Intraventricular”
- Final Run

  a. Incorporated additional images, subarachnoid and normal types, for training the U-net model.

  b. Data Augmentation Strategies

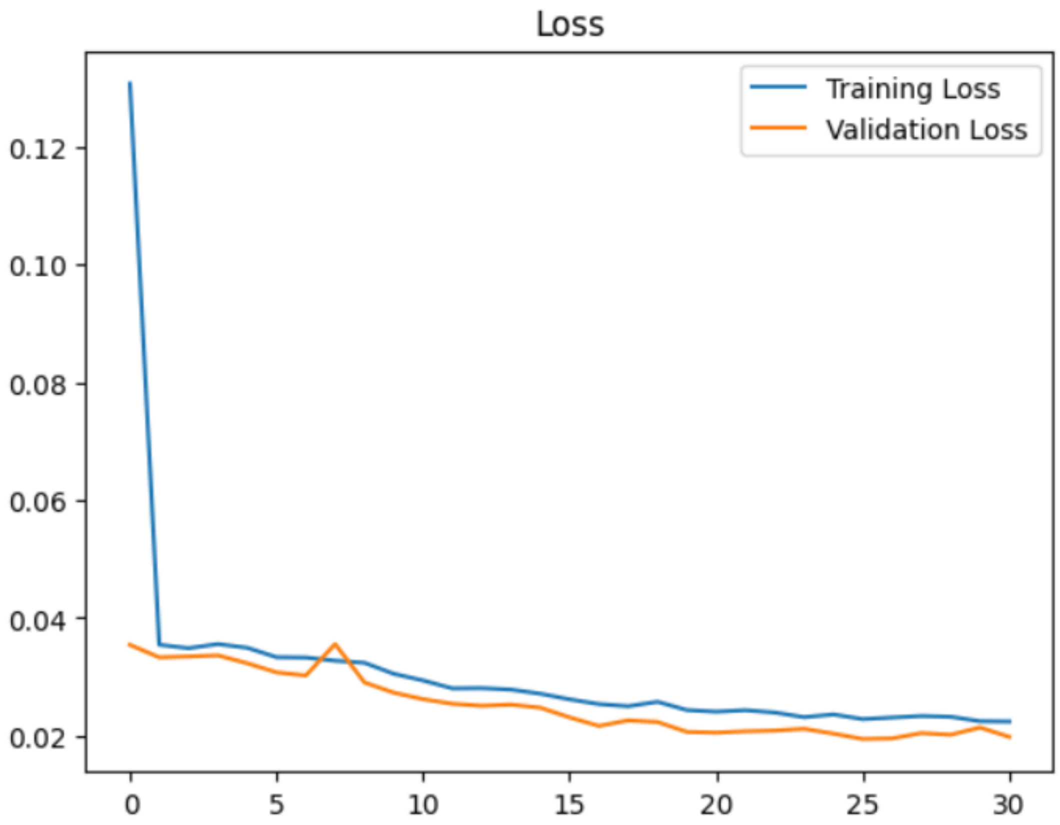
- Dice coefficient (from 0 to 1)

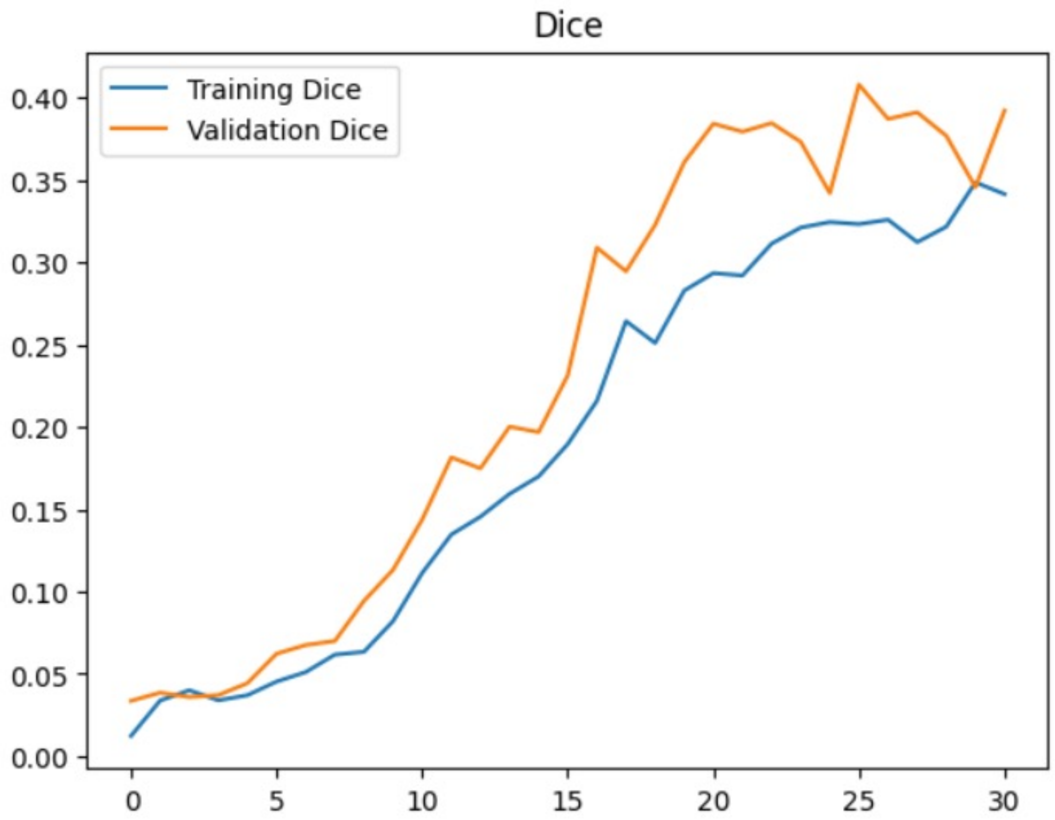
\- measures the similarity between two sets of data

$$\frac{2*|X\cap Y|}{|X|+|Y|}$$

(X is the predicted set of pixels and Y is the ground truth)

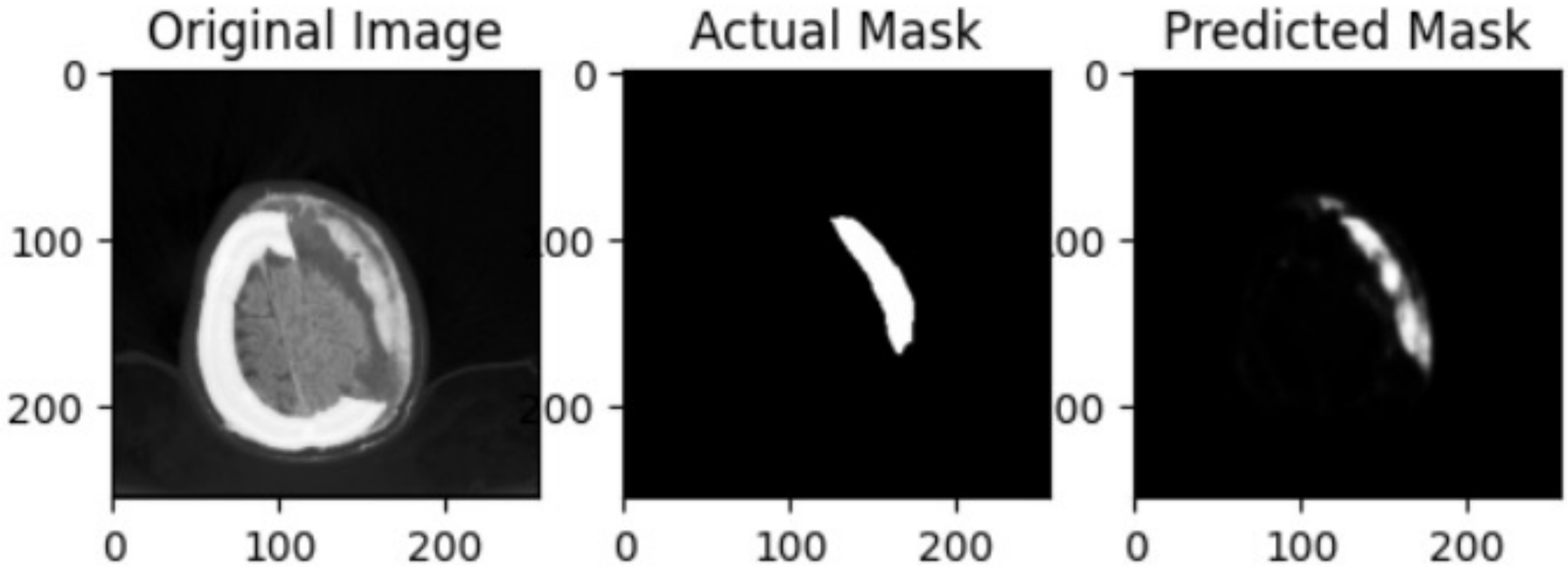
Segmentation (Pilot Run)



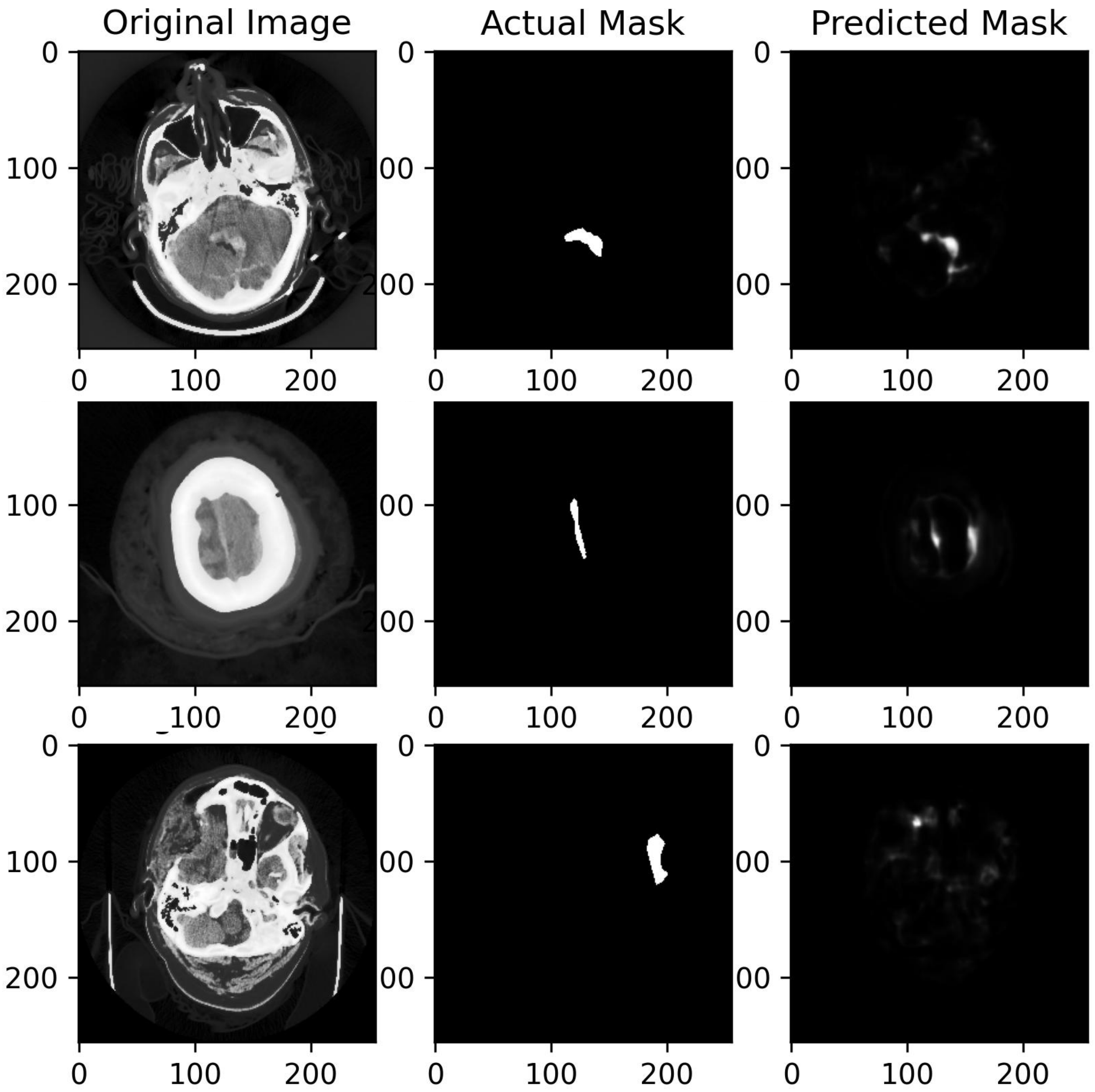


Segmentation (Pilot Run)

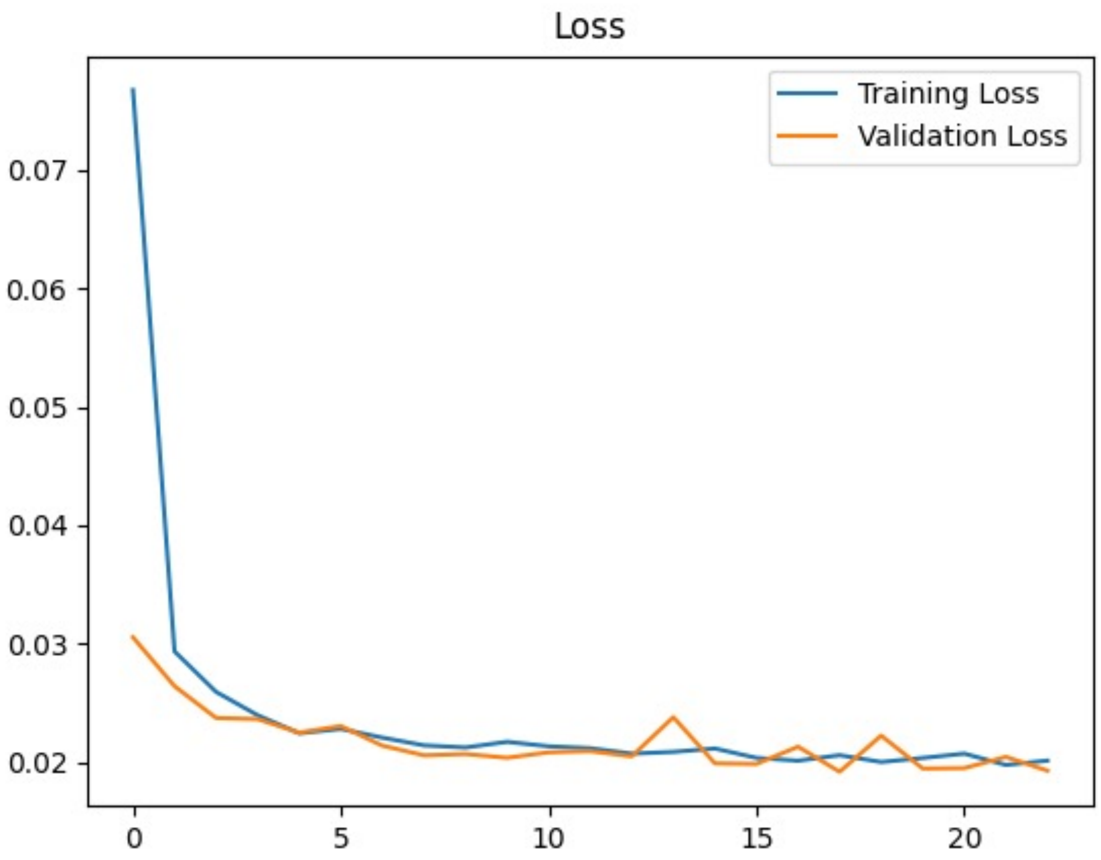
```
Dice Coefficient on Test Set: 0.3565
Dice Loss on Test Set: 0.6435
```

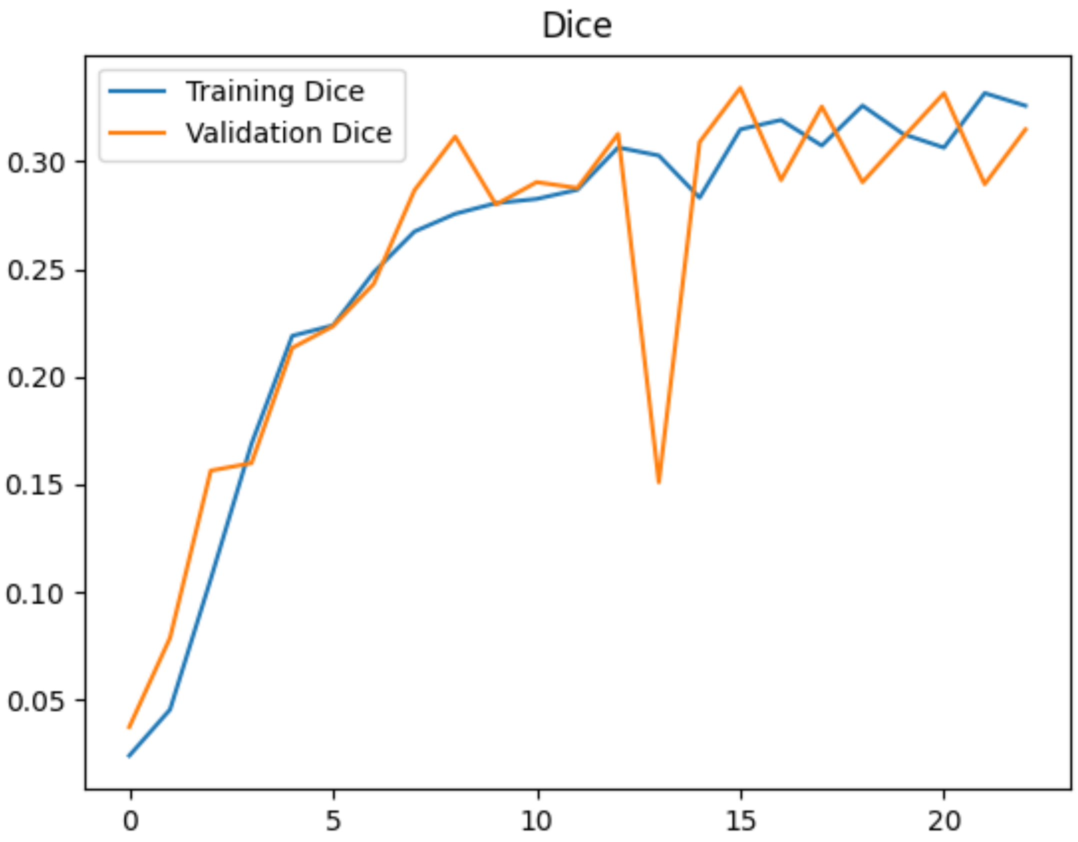


Segmentation (Pilot Run)



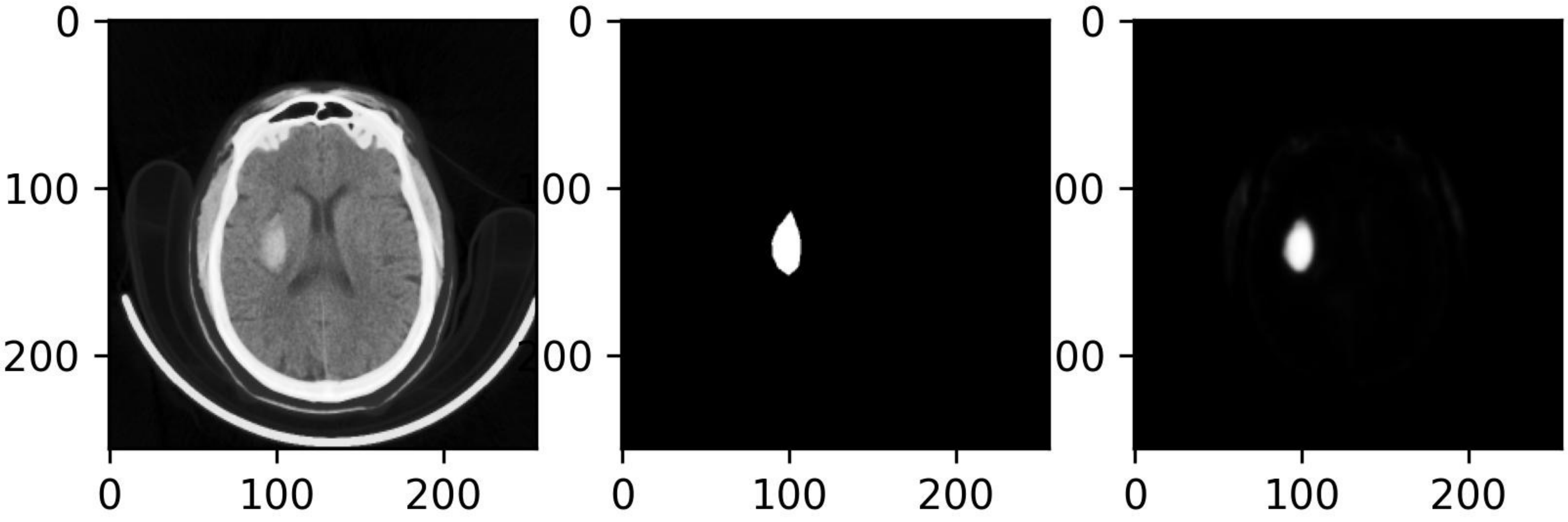
Segmentation (Final Run)



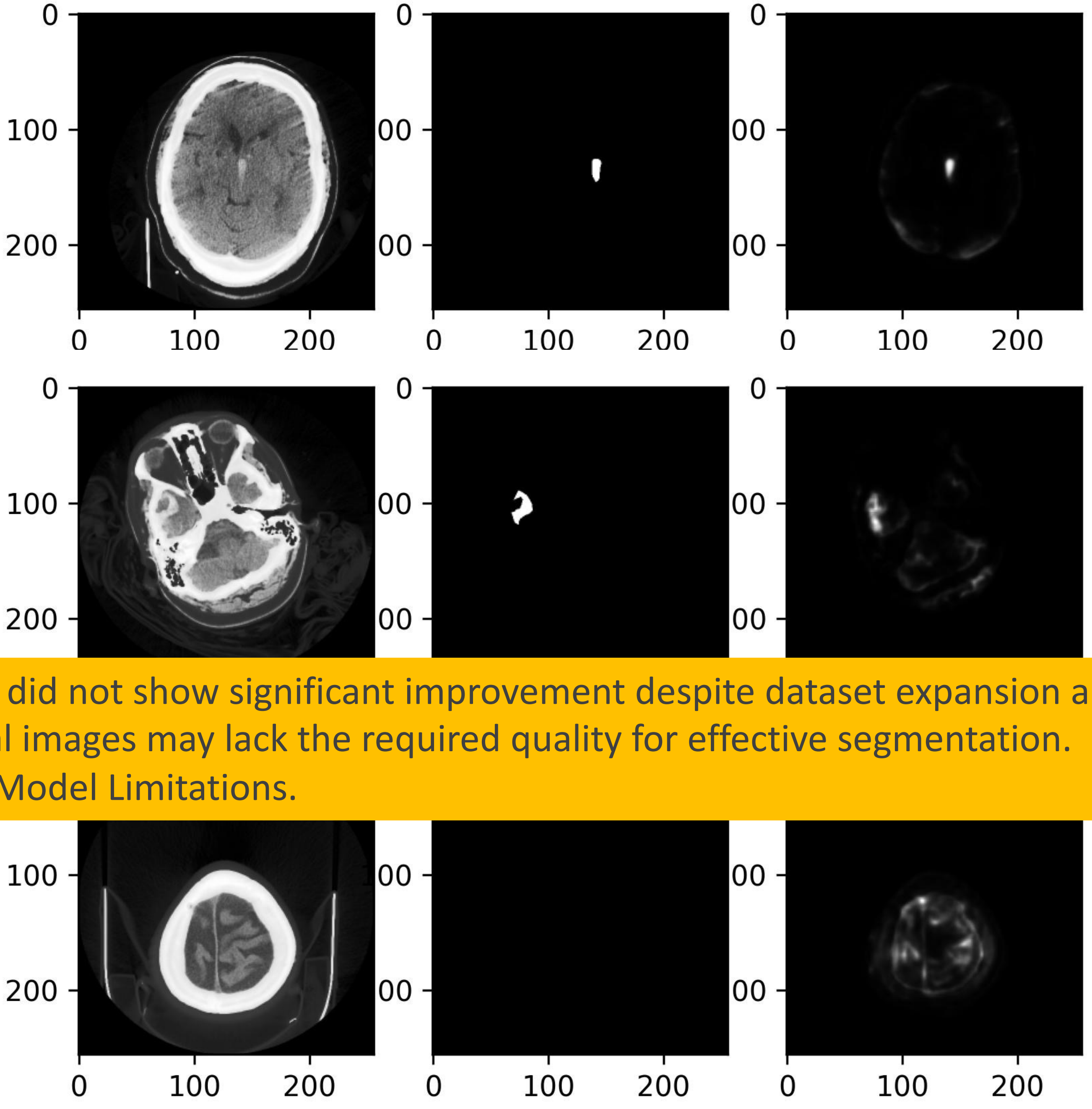


Segmentation (Final Run)

```
Dice Coefficient on Test Set: 0.3578
Dice Loss on Test Set: 0.6422
```



Challenges



→ Dice coefficient did not show significant improvement despite dataset expansion and augmentation.
→ Some additional images may lack the required quality for effective segmentation.
→ Potential Unet Model Limitations.

Discussion

- **Optimize Image Dataset:**
  - Improve dataset quality through enhanced preprocessing and augmentation.
  - Address biases and limitations within the dataset.
- **Explore Alternative Image Segmentation Models:**
  - Implement alternative models, such as Unet + ResNet architectures for precise segmentation.
  - Fine-tune parameters for optimal performance.
- **Simultaneous Classification and Segmentation:**
  - Develop a unified model for simultaneous classification and segmentation.
  - Explore multitasking approaches for efficiency and accuracy.