FACULTY OF COMPUTING AND INFORMATION SYSTEMS

CSDS 336 - FUNDAMENTALS OF MACHINE LEARNING

BRAIN TUMOR DETECTION MODEL ANALYSIS REPORT

REPORTED BY

JESSE AARON QUARTEY 1702306956

OF PROGRAMME BSC DATA SCIENCE AND ANALYTICS

LEVEL 300

GROUP D

TABLE OF CONTENTS

1 INTRODUCTION

Brain tumors are abnormal growths of cells in the brain which can be life-threatening if not diagnosed early. Automated detection using medical imaging can assist radiologists in faster diagnosis. This project explores using convolutional neural networks (CNNs) to classify brain MRI images into pituitary tumor and no-tumor categories**.**

1.1 BACKGROUND OF STUDY

Brain MRI images are commonly used in diagnosing brain tumors. Traditional diagnosis requires manual inspection by radiologists, which, based on precedence, is time-consuming and subject to human error. Machine learning, especially deep learning with CNNs, has shown potential in accurately classifying medical images. This new found technology, along with its methodologies and problem-solving algorithms, has proven to boost efficiency and accuracy of early detection of brain tumors.

1.2 PROBLEM STATEMENT

Early detection of brain tumors is critical. Manual inspection of MRI scans is time-intensive and prone to errors. The challenge is to build a reliable machine learning model that can automatically classify brain MRI images with high accuracy and generalization.

1.3 AIM

The aim is to develop and compare multiple CNN models from scratch for classifying brain MRI images into pituitary tumor and no-tumor classes.

1.4 OBJECTIVES

The objectives of this project are;

1. Build three CNN models from scratch and compare their performance.
2. Evaluate models using AUC, precision, recall and accuracy. As well as a validation curve.

1.5 SIGNIFICANCE OF PROJECT

This project demonstrates the use of deep learning in medical image classification by highlighting the effectiveness of CNNs in assisting radiologists. It also provides insight into the trade-offs between model complexity, regularization, and dataset size, which is useful for future ML projects in healthcare.

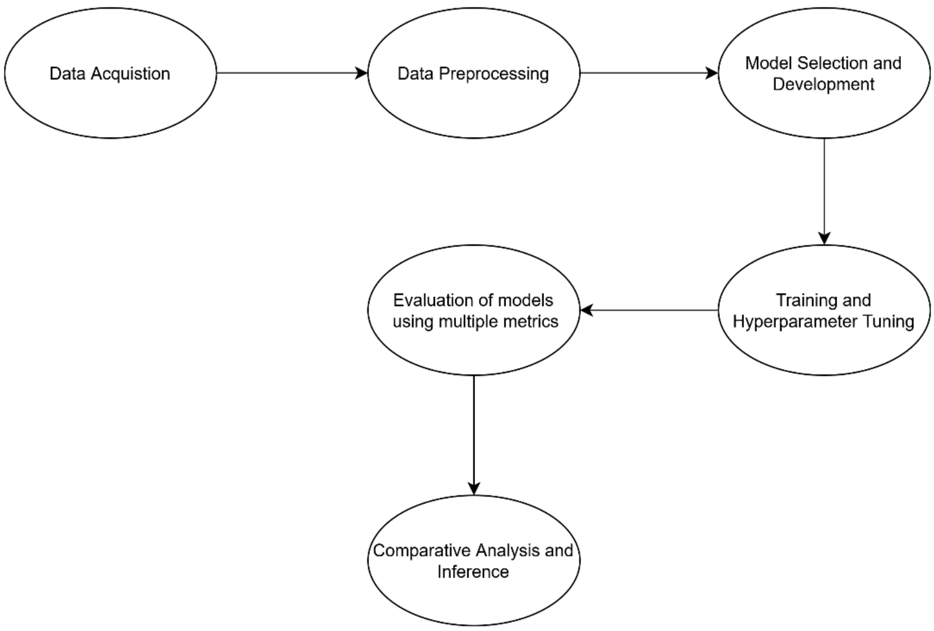
1.6 PRELIMINARY LITERATURE REVIEW

Previous studies have shown that CNNs achieve high accuracy in medical image classification tasks such as breast cancer detection, lung nodule detection, and brain tumor detection. Data augmentation and hyperparameter tuning improve generalization, but small datasets can limit the effectiveness of deeper models.

1.7 METHODOLOGY

1. Dataset Acquisition
2. Data Preprocessing
3. Model Selection and Development
4. Training and Hyperparameter Tuning
5. Evaluation
6. Comparative Analysis and Inference

<u>Conceptual Framework</u>



1.7.1 DATASET ACQUISITION

The dataset used for this project was obtained from Kaggle, a public repository of pre-gathered and structured datasets.

Specifically, the Brain Tumor Classification MRI dataset was sourced from Kaggle at: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri/

The dataset has four classes namely:

- glioma_tumor
- meningioma_tumor
- pituitary_tumor
- no_tumor

As indicated by their names, the classes are self-descriptive. However, in this experiment, only the **pituitary_tumor** and **no_tumor** classes were selected. This decision was made due to the limited size of the dataset, which may not be sufficient for effective generalization in a multi-class classification setting. By reducing the task to binary classification, the models are able to learn more discriminative features and achieve more reliable performance.

1.7.2 DATA PREPROCESSING

The preprocessing technique used here was image augmentation for increasing the size of the dataset for better generalization. This included:

- Rescaling
- Flipping
- Resizing
- Zooming
- Rotation

The intensity of image augmentation was done differently for each model to better comparative analysis.

1.7.3 MODEL SELECTION AND DEVELOPMENT

Three models were made in this project using the Convolutional Neural Networks (CNN) algorithm. The following table shows the models with their architectural layers:

Model	**Architecture**	**Regularization / Augmentation**
Model 1 (Baseline)	Shallow CNN with single convolutional block [Conv2D(16)]	None (rescaling only)
Model 2 (Tuned CNN)	Moderately deep CNN with two convolutional blocks and increased feature capacity [Conv2D(16→32)]	Dropout + heavy augmentation
Model 3 (Deeper CNN)	Deeper CNN with three convolutional layers and higher representational capacity [Conv2D(16→32→32)]	Dropout + mild augmentation (resizing and rescaling only)

1.7.4 HYPERPARAMETER TUNING

Hyperparameters for Model 1

- Learning rate = 0.001
- Batch size = 8
- Epochs = 15
- Number of filters on 1st convolutional layer = 16

Hyperparameters for Model 2

- Learning rate = 0.001
- Batch size = 8
- Epochs = 15
- Number of filters on 1st convolutional layer = 16
- Number of filters on 2nd convolutional layer = 32
- Dropout for 1st convolutional layer = 0.25
- Dropout for 2nd convolutional layer = 0.50

Hyperparameters for Model 3

- Learning rate = 0.001
- Batch size = 8
- Epochs = 15
- Number of filters on 1st convolutional layer = 16
- Number of filters on 2nd convolutional layer = 32
- Number of filters on 3rd convolutional layer = 64
- Dropout for 2nd convolutional layer = 0.2
- Dropout for 3rd convolutional layer = 0.3

1.7.5 EVALUATION AND INFERENCE

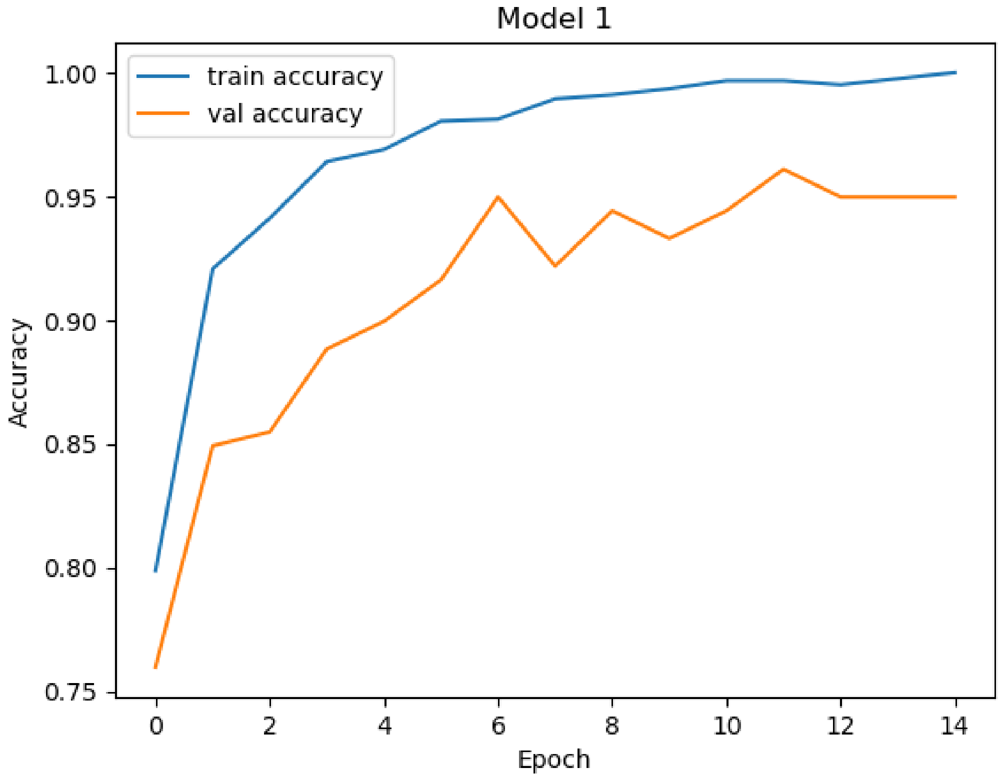
Training Evaluation Metrics

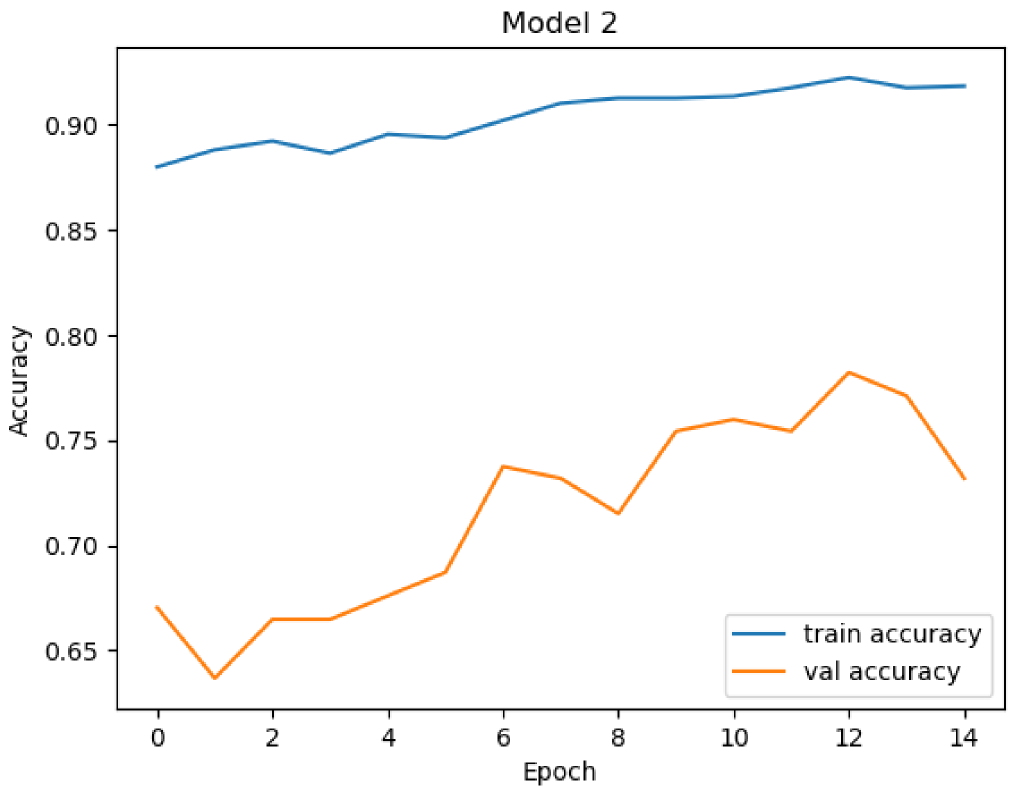
Metric	Model 1	Model 2	Model 3
Training Accuracy	1.0000	0.9182	0.8813
Training AUC	1.0000	0.9098	0.9195
Training Precision	1.0000	0.9758	0.8723
Training Recall	1.0000	0.9647	0.9661

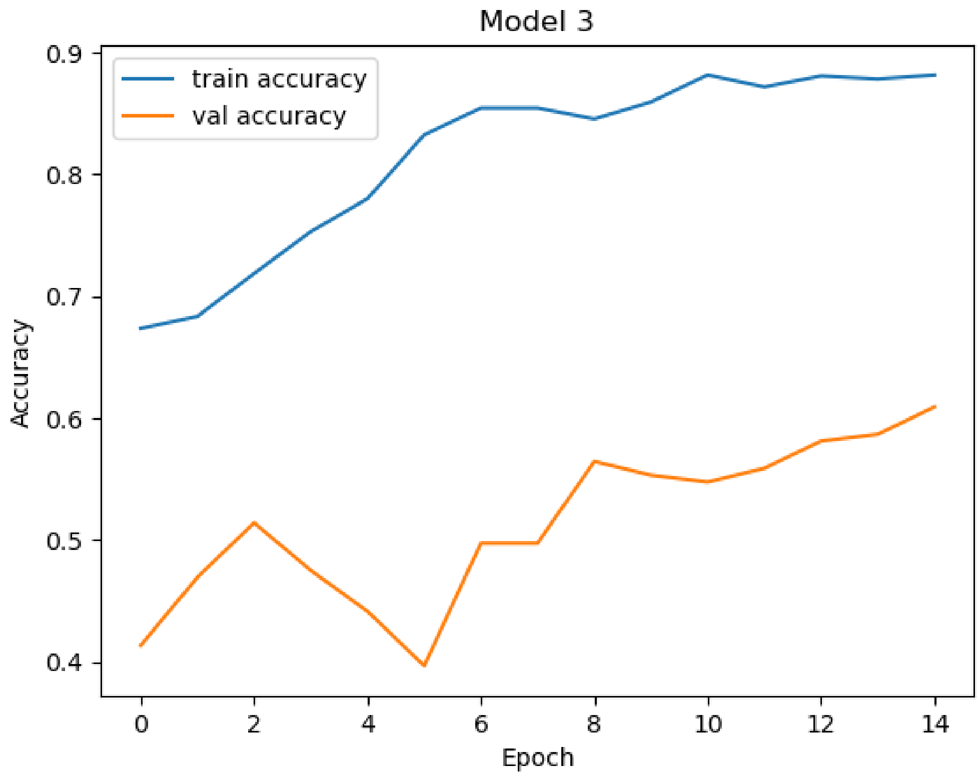
Validation Evaluation Metrics

Metric	Model 1	Model 2	Model 3
Validation Accuracy	0.9497	0.7318	0.6089
Validation AUC	0.9983	0.8246	0.6403
Validation Precision	1.0000	0.9333	0.5370
Validation Recall	0.8784	0.3784	0.3919

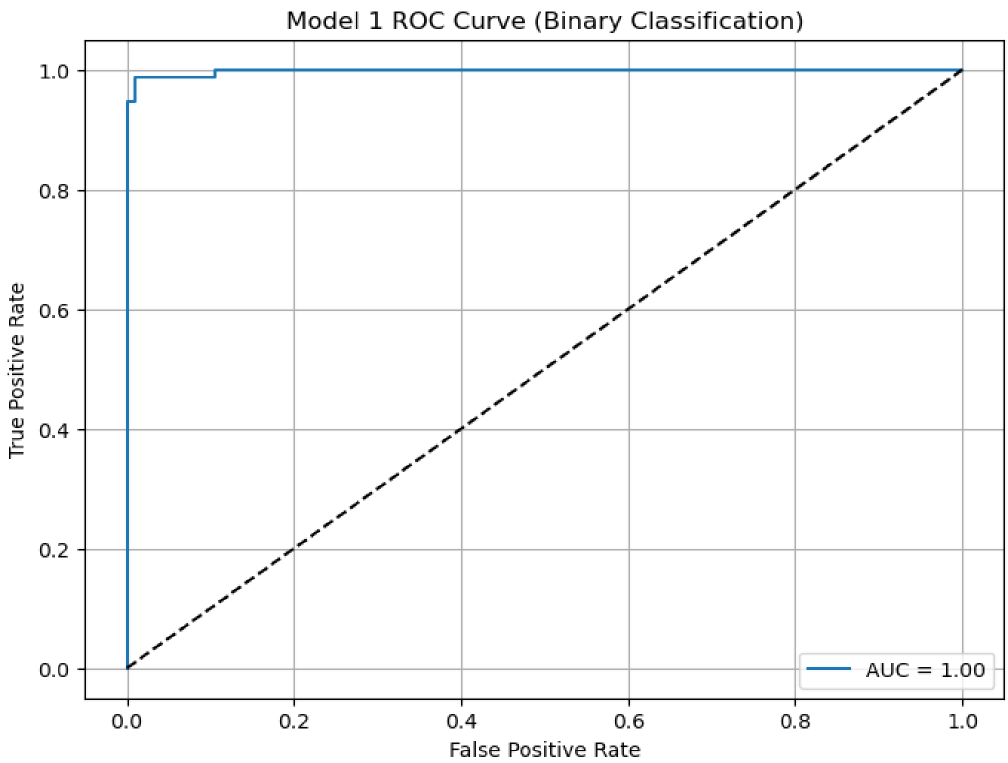
Validation Curves For Each Model

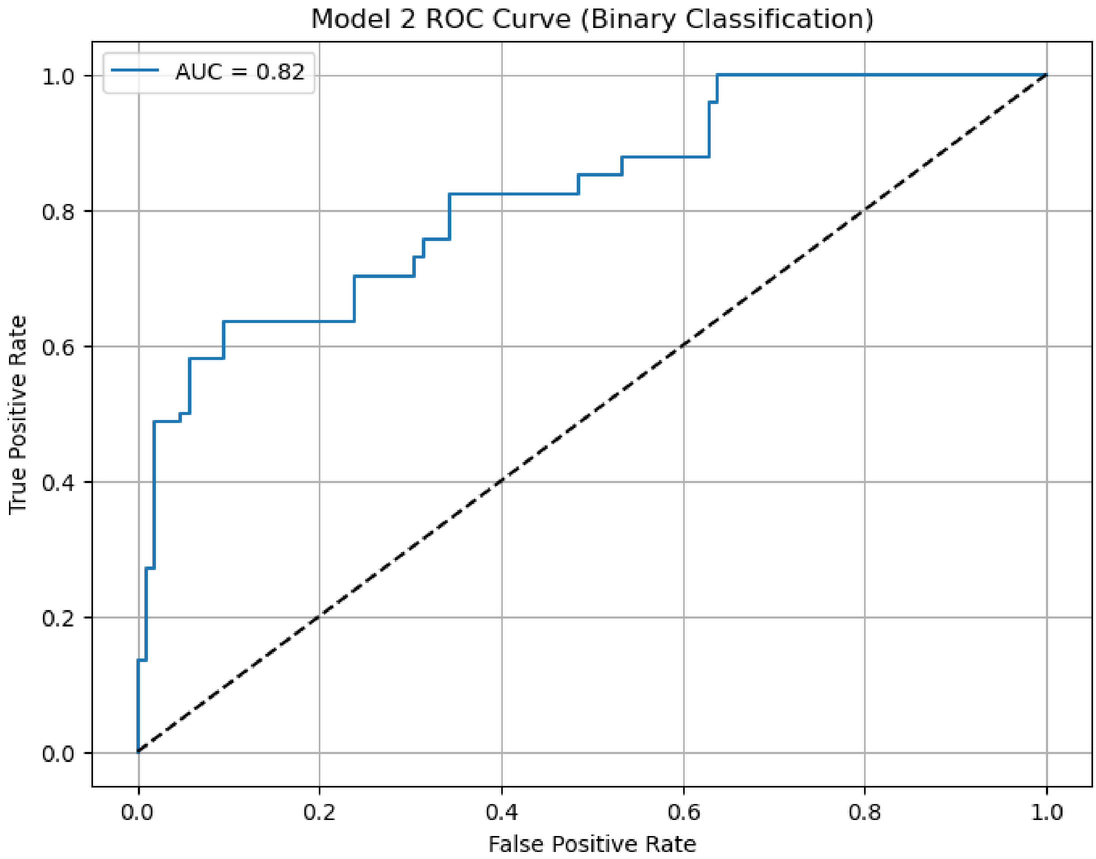


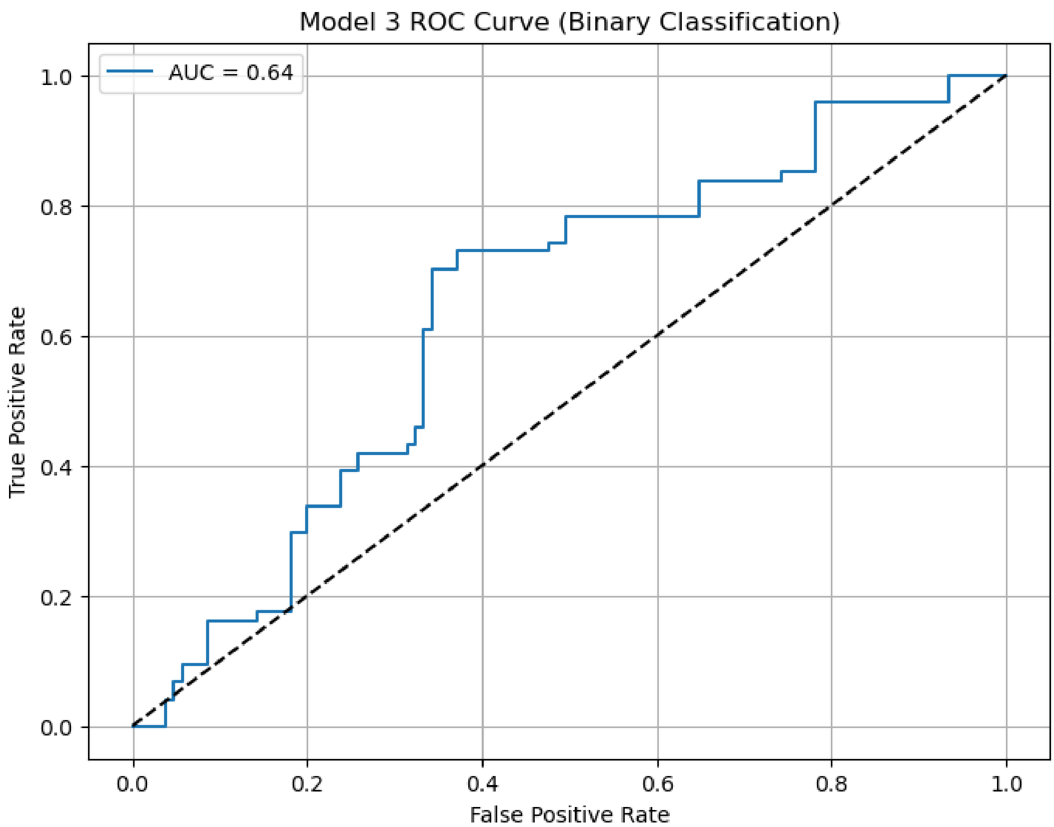




ROC Curve For Each Model




Model 2 ROC Curve (Binary Classification)
AUC = 0.82
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0


Model 3 ROC Curve (Binary Classification)
AUC = 0.64
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

Comparative Analysis

The three models differ primarily in depth and preprocessing. Model 1 is a simple baseline CNN, Model 2 introduces an additional convolutional layer and heavy augmentation, while Model 3 is a deeper architecture with mild augmentation.

Model 1 achieved near-perfect training accuracy, while Model 2 and 3 showed slightly lower training accuracy due to increased complexity and the regularizing effects of intensified augmentation. Model 3, despite being deeper, overfit the limited dataset as evidenced by the gap between training and validation performance.

The validation performance indicated that the simpler baseline CNN (Model 1) generalized best on unseen data, achieving higher validation accuracy and AUC than the deeper models. The heavy augmentation in Model 2 reduced overfitting but slightly lowered discriminative performance, while Model 3's depth led to poor generalization due to insufficient data for effective learning.

Overall, these results demonstrate that for small medical imaging datasets, simpler architectures with minimal augmentation may outperform deeper networks with heavy augmentation.

Inference

The baseline CNN (Model 1) performed best on this dataset, highlighting the effect of dataset size on model complexity. Regularization and augmentation, as shown in Model 2 and Model 3, help prevent overfitting but may reduce performance if the dataset is small. IN conclusion, increasing model complexity without sufficient data can negatively impact generalization, highlighting the importance of balancing architecture depth, data preprocessing, and dataset size.