DETECTING BRAIN TUMORS AND MGMT PROMOTER METHYLATION USING A 3D CONVOLUTIONAL NEURAL NETWORK AND THE RSNA-MICCAI RADIOGENOMIC CLASSIFICATION DATASET

#1010905

DATE
09/01/2025

TABLE OF CONTENTS

GROUP MEMBERS

COMPUTER ENGINEERING

Ömer Dursun - 2203452
Barış Başaran - 2200198
Mehmet Cengizhan Kınay - 2102804

Advisor:
Dr. Günet Eroğlu

ARTIFICIAL INTELLIGENCE ENGINEERING

Aleyna Benan Aydı - 2003977
Deniz Arda Yıldız - 2001247
Mert Acar - 2004287

Advisor:
Dr. Fatih Kahraman

The team is composed of five members, each specializing in different aspects of the project:

Ömer Dursun, Barış Başaran, and Mehmet Cengizhan Kınay from Computer Engineering, focusing on algorithm development, system integration, building UI and serving on the web.

Aleyna Benan Aydı, Deniz Arda Yıldız, and Mert Acar from Artificial Intelligence Engineering, concentrating on data processing, neural network design, and performance evaluation.

Each member has specific responsibilities outlined in the Responsibility Matrix (RM) and is aligned with the Work Breakdown Structure (WBS) to ensure a well-organized approach to achieving our goals.

Initially exploring several design alternatives, including a 2D CNN and a traditional image processing method, but decided on a 3D CNN approach after evaluating the need to capture spatial relationships across multiple MRI sequences. This design choice was justified by the ability of 3D CNNs to process volumetric data, enabling the model to capture more comprehensive features that directly correlate with the MGMT methylation status, improving the accuracy and robustness of predictions.

WHAT IS AIM OF PROJECT?

INTRODUCTION

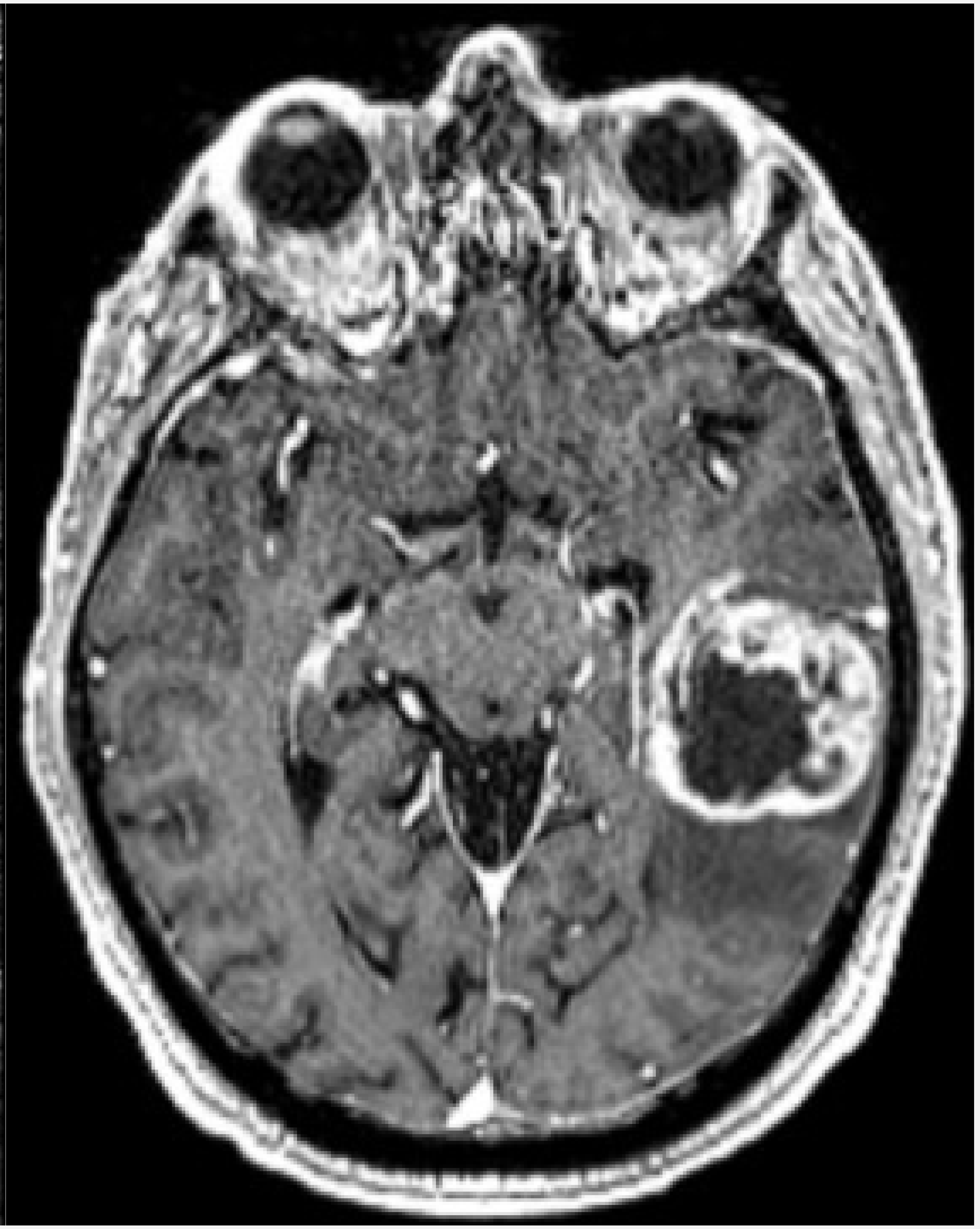
- Our project aims to develop a novel deep learning-based system for detecting brain tumors and predicting MGMT promoter methylation status using MRI scans.

- Brain tumors, particularly glioblastomas, are a significant medical challenge, as timely and accurate diagnosis is critical for determining the best course of treatment.

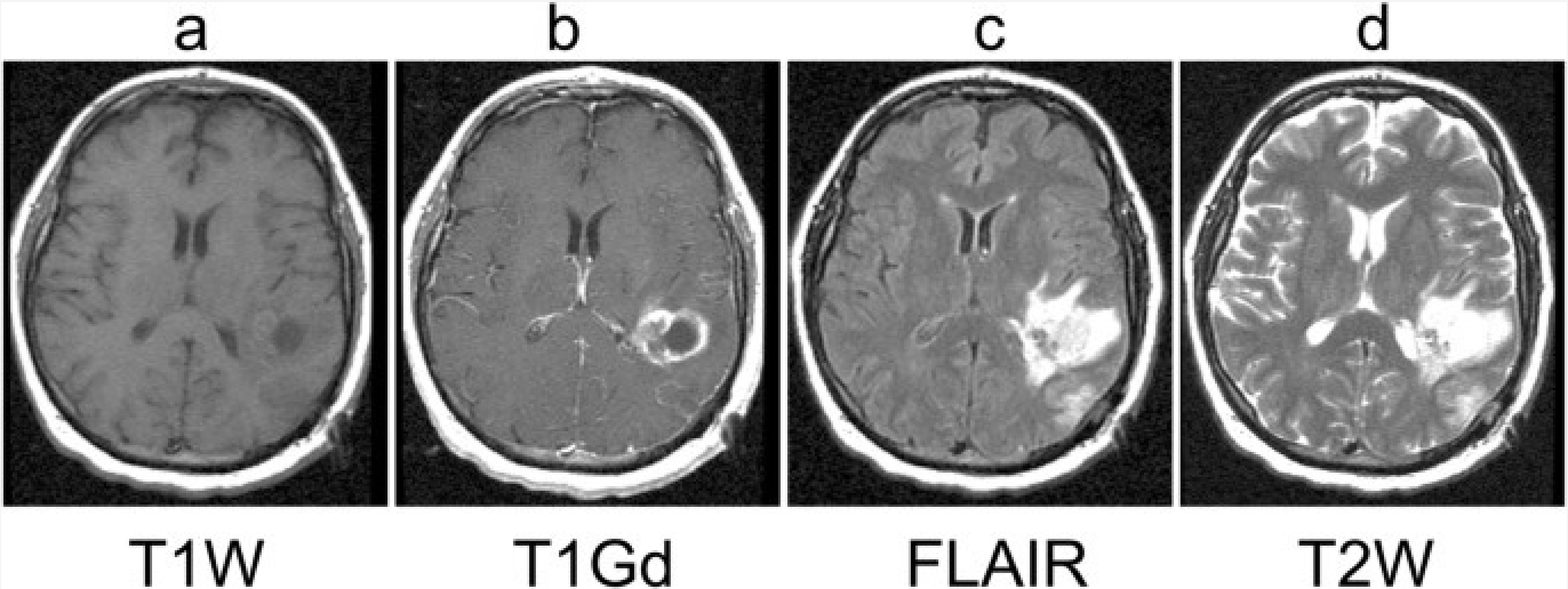
- Traditional methods of determining molecular markers like MGMT methylation require invasive procedures, but our system seeks to automate this process, providing a non-invasive, efficient alternative that leverages the power of 3D convolutional neural networks (CNNs) to analyze MRI data.

- The RSNA-MICCAI Brain Tumor Radiogenomic Classification dataset is used for model development.

WHAT IS GBM? - GBM SUBTYPE?



IMAGING BRAIN ? - IMAGING TECHNICS ?



MGMT PROMTER AND RADIOGENOMİC

Radiogenomics:A field of science that tries to correlate radiomics data with genetic data at the same time.

MGMT Promoter: Using MRI to analyze MGMT methylation status, combined with artificial intelligence (AI) and deep learning, enhances the precision of radiogenomic models, extending their diagnostic capabilities.

WHAT IS MGMT?

Full Name: O6-methylguanine-DNA methyltransferase

Function: Enzyme that repairs DNA damage caused by alkylating agents.

Role in Cancer:

- Protects cells from mutations and cancer by repairing DNA.
- Particularly important in brain tumors like glioblastoma.

MGMT Methylation:

- Methylation of MGMT gene promoter = silencing of the enzyme.
- Tumor cells with methylated MGMT are more sensitive to chemotherapy (e.g., temozolomide).

Clinical Relevance:

- MGMT methylation status is a key biomarker for treatment response.
- Predicts better outcomes with chemotherapy.

SYSTEM OVERVIEW AND DESIGN

Sub-systems:

- **Preprocessing:** Data cleaning, skull stripping, and normalization of MRI sequences.
- **Model Architecture:** 3D CNN that processes volumetric MRI data.
- **Data Integration:** Combining different MRI sequences (T1w, T1wCE, T2w, FLAIR) into a unified 3D input.

Responsibilities: Team members handle various tasks like data preprocessing, model architecture design, and system integration.

Challenges:

- Handling multi-sequence MRI data.
- Addressing dataset imbalance.
- Managing variations in MRI scanner protocols.

DATASET AND PREPROCESSING

RSNA-MICCAI Dataset: Contains multi-modal MRI scans (T1w, T1wCE, T2w, FLAIR) for brain tumor patients, labeled for MGMT methylation status.

Data Organization and Labeling

- **Data Organization:** MRI scans from different modalities are organized into a 4-channel input format.
- **Labeling:** Labels for each MRI scan are associated with MGMT promoter methylation status, indicating whether the tumor is methylated or unmethylated.

Preprocessing Steps:

- **Skull Stripping:** Removal of non-brain tissue from the MRI scans.
- **Normalization:** Standardizing intensity values across sequences to minimize scanner-related variations.
- **Resizing:** Ensuring all volumes have consistent dimensions, which is essential for feeding the data into the network for processing.
- **Modality Stacking:** Multiple MRI sequences (T1w, T1wCE, T2w, and FLAIR) are combined into a single 4-channel volume to provide a richer input for deep learning models

DATASET AND PREPROCESSING

Data Augmentation: designed to increase the variability of the dataset, making the model more robust and preventing overfitting. These techniques are particularly useful for medical imaging tasks where labeled data might be limited.

- **Random Horizontal Flip:** The MRI scans are flipped along one or more axes to create mirror images, which helps the model generalize better to variations in tumor position or orientation.

- **Random Rotation:** Rotating the MRI scans around various axes to simulate different orientations that a scan might have in real-world data.

- **Random Zoom:** This method is useful to simulate different viewing distances.

- **Intensity Shifts:** Adjusting the intensity values of the images to simulate variations in scanning protocols, different scanner settings, or noise patterns.

DISCUSSION

Clinical Relevance: The AI model can serve as a second opinion for radiologists, providing quick and consistent insights into both anatomical and molecular tumor characteristics.

Challenges:

- **Interpretability:** Understanding the model's decision-making process through visualization tools like Grad-CAM.
- **External Validation:** The need for validation on diverse datasets to ensure generalization across different populations and MRI protocols.
- **Integration into Clinical Workflow:** Ensuring the model can be seamlessly incorporated into existing hospital systems like PACS.

Future Work:

- Improving the model’s interpretability and robustness.
- Conducting clinical trials for real-world validation.
- Expanding the model to predict additional biomarkers.

MODEL

Hyperparameter Optimization with Keras Tuner

Objective: The main goal is to enhance the performance of the machine learning model by systematically searching for the best set of hyperparameters, such as learning rate, number of layers, and batch size. This process is critical in improving model accuracy and generalization.

Bayesian Optimization:

- A probabilistic model is used to guide the search for the optimal hyperparameters. Bayesian optimization intelligently chooses the next hyperparameter configurations based on previous trials, significantly reducing the search space and improving optimization efficiency.

Model Hyperparameter Recording:

- During the search for the best epoch, it is crucial to record and store the model's hyperparameters for reproducibility and tracking. This ensures that the best configurations can be reused or analyzed further in future training.

Model Saving and Training:

- After finding the optimal hyperparameters, the model is saved using save_model for later use or deployment. The training phase (fit) involves training the model with the best hyperparameters to achieve superior performance.

Results Analysis:

- The average values are key in understanding the relationship between MRI scan features and MGMT promoter methylation status.
- The results are examined to ensure the model's predictions are consistent across different instances and that it generalizes well to unseen data.

Testing on Test Set:

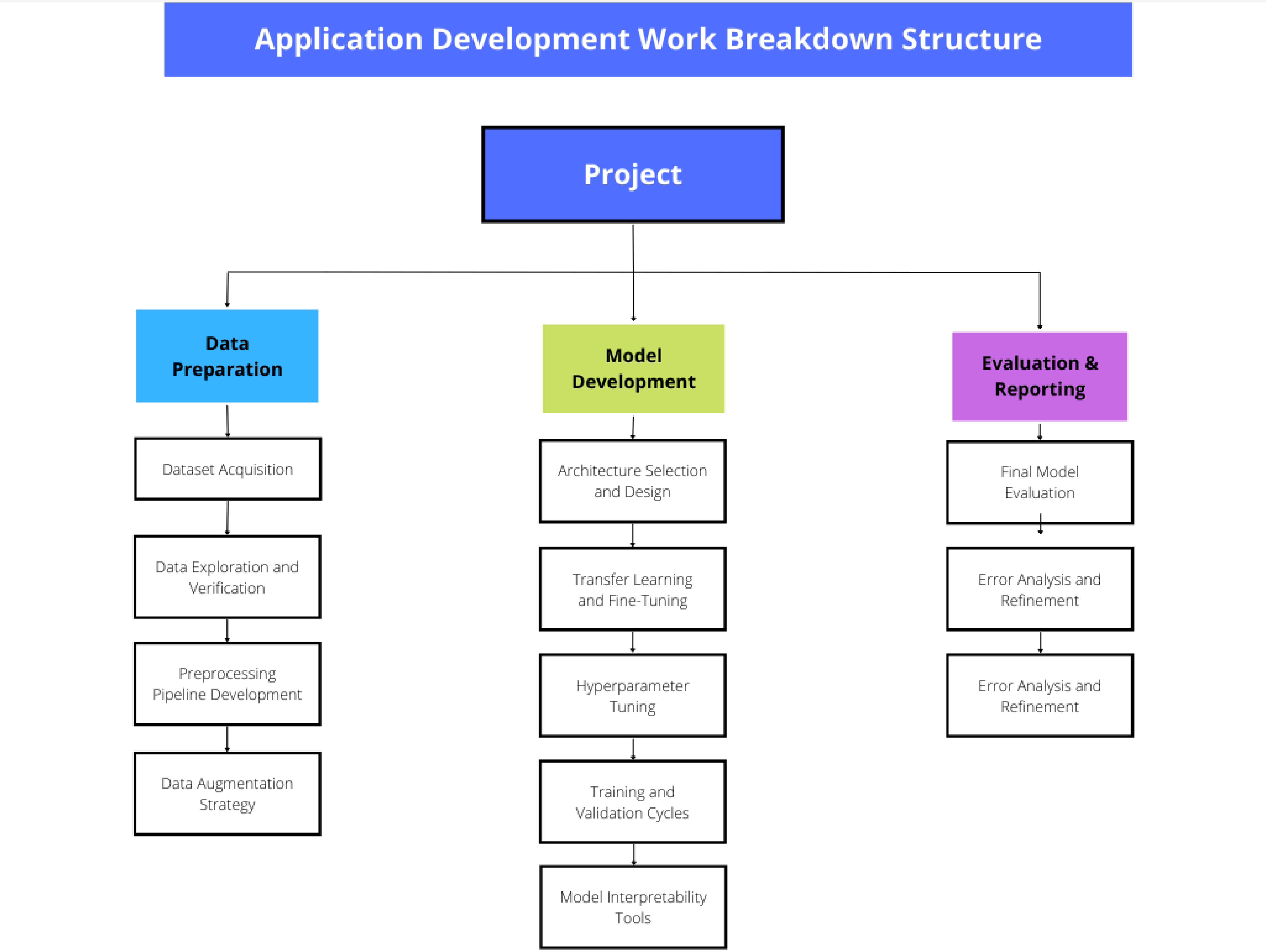
- The final model is tested on a separate test set to assess its ability to predict the MGMT methylation status in new, unseen MRI data. The performance on the test set is critical for evaluating the model's real-world applicability.

Validation and Model Performance:

- In addition to testing on a standard test set, cross-validation techniques could be used to further ensure the robustness and generalizability of the model. Metrics such as accuracy, precision, recall, and AUC are calculated to evaluate its predictive capabilities.

Additional Considerations:

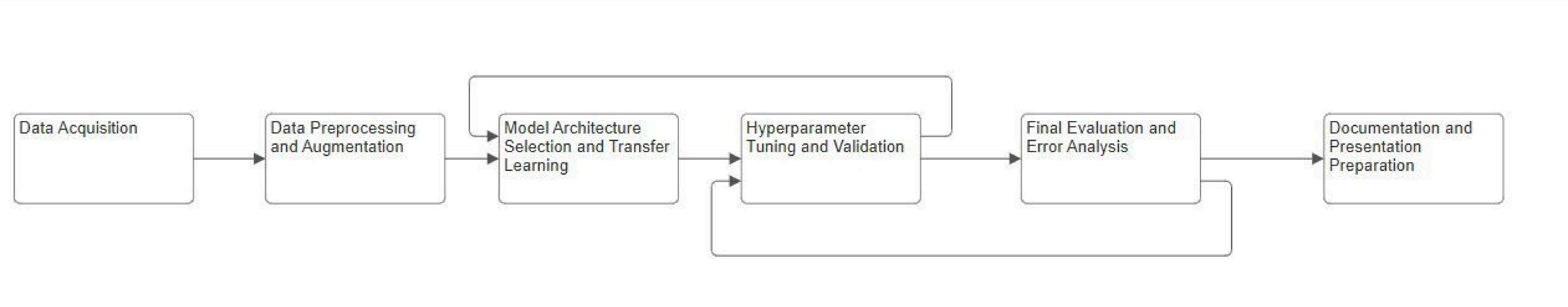
- **Overfitting:** To prevent overfitting, the model employs techniques like data augmentation, dropout, and early stopping. These techniques help to ensure that the model generalizes well to new data.
- The optimization process also ensures that the model can handle variations in MRI scans, ensuring its adaptability to real-world clinical environments.


Application Development Work Breakdown Structure
Project
Data Preparation
Dataset Acquisition
Data Exploration and Verification
Preprocessing Pipeline Development
Data Augmentation Strategy
Model Development
Architecture Selection and Design
Transfer Learning and Fine-Tuning
Hyperparameter Tuning
Training and Validation Cycles
Model Interpretability Tools
Evaluation & Reporting
Final Model Evaluation
Error Analysis and Refinement
Error Analysis and Refinement

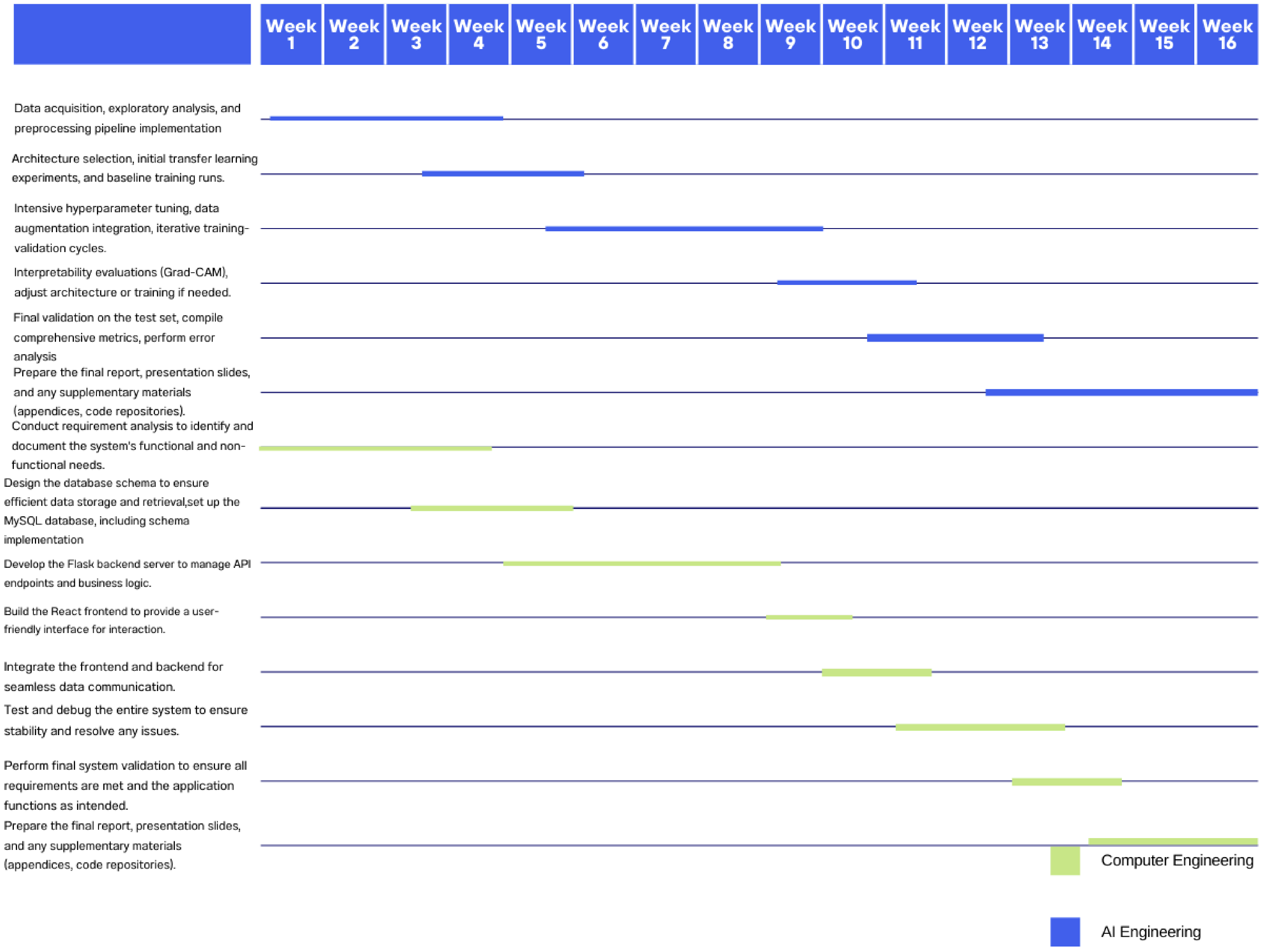
Responsibility Matrix

	Barış	Cengiz	Ömer	Benan	Deniz	Mert
Data Preparation		X		X	X	
Model Development				X	X	X
Integrating to App	X	X	X			X
Evaluation and Reporting	X		X			

PROJECT NETWORK (PN)



Gantt Chart


Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10
Week 11
Week 12
Week 13
Week 14
Week 15
Week 16
Data acquisition, exploratory analysis, and preprocessing pipeline implementation
Architecture selection, initial transfer learning experiments, and baseline training runs.
Intensive hyperparameter tuning, data augmentation integration, iterative training-validation cycles.
Interpretability evaluations (Grad-CAM), adjust architecture or training if needed.
Final validation on the test set, compile comprehensive metrics, perform error analysis
Prepare the final report, presentation slides, and any supplementary materials (appendices, code repositories).
Conduct requirement analysis to identify and document the system's functional and non-functional needs.
Design the database schema to ensure efficient data storage and retrieval,set up the MySQL database, including schema implementation
Develop the Flask backend server to manage API endpoints and business logic.
Build the React frontend to provide a user-friendly interface for interaction.
Integrate the frontend and backend for seamless data communication.
Test and debug the entire system to ensure stability and resolve any issues.
Perform final system validation to ensure all requirements are met and the application functions as intended.
Prepare the final report, presentation slides, and any supplementary materials (appendices, code repositories).
Computer Engineering
AI Engineering

Cost Estimates Using NVIDIA L4 GPU on Google Cloud

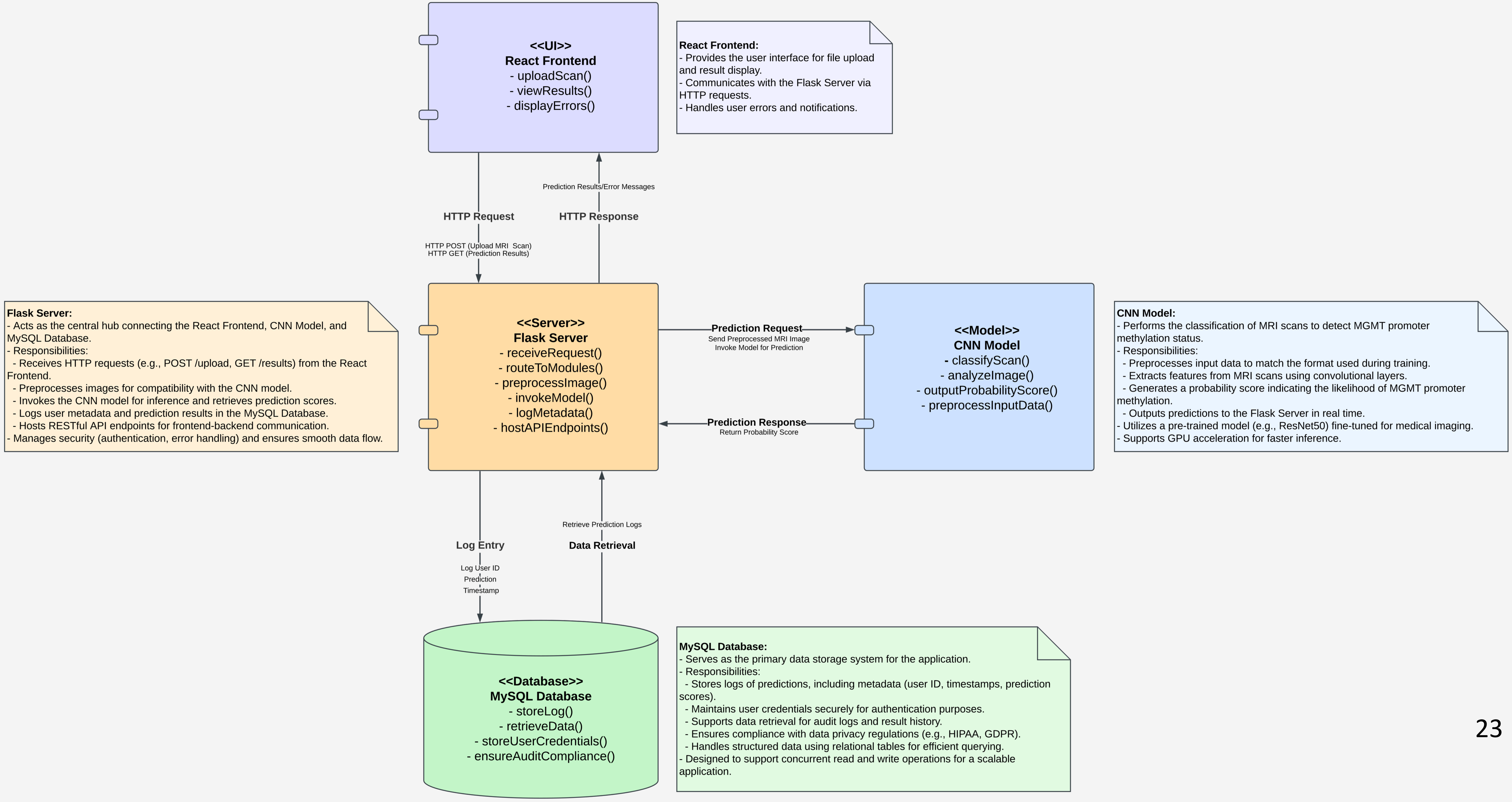
GPU Instance (1 month): ~$516.99

Storage & Misc. Expenses: ~$50–$100

RISK ASSESSMENT

Risk	Likelihood	Impact	Risk Level	Mitigation Strategies
Limited Data Quality or Availability	Medium	High	High	Implement robust preprocessing, consider synthetic data augmentation, and exclude unreliable cases if necessary.
Overfitting to Training Data	High	Medium	High	Use early stopping, dropout layers, regularization techniques, and augmented validation protocols to ensure generalization.
Cloud Cost Overruns	Medium	High	High	Track usage closely, use efficient training techniques, and stop experiments when diminishing returns are observed.
Computational Bottlenecks or Delays	Medium	Medium	Medium	Profile code, use mixed-precision training, and schedule experiments during off-peak hours for cost savings.

COMPONENT DIAGRAM



FLASK - INTEGRATION FRAMEWORK

Purpose: Acts as the central hub for integrating multiple subsystems.

- Coordinates communication between the database, React frontend, and AI model.

Technology:

- Python-based lightweight web framework.
- Integrates with Gunicorn for handling concurrent requests.

Functionality:

- Routes:
  - `/predict`: Processes MRI scans and returns prediction probabilities.
  - `/log`: Logs user activities and prediction metadata into the database.
- Facilitates seamless interaction between subsystems, ensuring efficient data flow.

Example Libraries:

- **Flask:** API routing and request handling.
- **Flask-CORS:** Enables cross-origin requests.
- **Flask-RESTful:** Simplifies RESTful API development.
- **Flask-SQLAlchemy:** ORM for managing database interactions.

DATABASE - MySQL

Purpose: Logs predictions and maintains metadata for traceability.

Technology:

- Relational database with ACID compliance.
- Secure user authentication and role-based access.

Integration:

- Stores logs for:
  - User ID
  - Timestamp
  - Prediction probability
- Ensures GDPR/HIPAA compliance.

Technical Highlights:

- Uses indexed queries for efficient log retrieval.
- Backup and encryption mechanisms for sensitive medical data.

DEPLOYMENT - Gunicorn, Nginx, and Certbot

Purpose: Ensures secure, efficient, and scalable deployment.

Technology:

- **Gunicorn:**
  - WSGI server to manage multiple Flask workers.
  - Handles concurrent requests efficiently.
- **Nginx:**
  - Acts as a reverse proxy.
  - Distributes incoming requests and serves static files (e.g., React frontend).
- **Certbot:**
  - Enables HTTPS with SSL/TLS encryption.
  - Automates certificate issuance and renewal.

Workflow:

1. User requests handled by Nginx.
2. Proxies API calls to Flask via Gunicorn.
3. Returns prediction results through secure HTTPS.

REACT - FRONTEND FRAMEWORK

Purpose: Provides a modern, responsive, and interactive user interface.

Technology:

- **React:** JavaScript library for building UI components.
- **React Router:** Enables seamless navigation without full-page reloads.

Features:

- **File Upload Page:** Allows users to upload MRI scans.
- **Result Display:** Shows prediction probabilities in a clean format.
- **Error Handling:** Alerts users for invalid inputs or system errors.

Workflow:

1. Users interact with React components for data entry.
2. Frontend validates and submits data to Flask endpoints.
3. Receives and displays JSON responses (e.g., predictions).

Deployment:

- Bundled using Webpack.
- Served via Nginx alongside the backend services.

CONCLUSION

Impact: This AI-driven approach to predicting MGMT promoter methylation status offers a significant step toward non-invasive, personalized treatment for brain tumor patients.

Model Evaluation:

- Precision and recall indicate the model's effectiveness in detecting methylated and unmethylated MGMT tumors.
- ROC and AUC curves show the discriminative power of the model.

Error Analysis: Minor errors were observed in cases where MRI sequences showed ambiguous tumor boundaries or artifacts.

Key Contributions:

- A robust 3D CNN that processes multi-sequence MRI data.
- A non-invasive method to predict molecular biomarkers like MGMT methylation.
- Enhanced clinical decision-making and faster treatment planning.

Next Steps:

- Further validation and integration into clinical practice.
- Expanding the model's capabilities to predict additional molecular markers and improve diagnostic accuracy.