VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Project Report on

Assistive tool : Feature Extraction for Stroke outcome and prognosis prediction using MRI Images & Clinical Dataset

In partial fulfillment of the Fourth Year (Semester–VII), Bachelor of Engineering

(B.E.) Degree in Computer Engineering at the University of Mumbai
Academic Year 2023-2024

Project Mentor
Dr. (Mrs.) Gresha Bhatia

Submitted by
Manasi Shah (D17A 61)
Kaushik Sahasranaman (D17A 31)
Riya Nadagire (D17A 48)
Chaitanya Sondur (D17A 66)

(2023-24)

VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

CERTIFICATE OF APPROVAL

This is to certify that ***Manasi Shah, Kaushik Sahasranaman, Riya Nadagire, Chaitanya Sondur*** of Fourth Year Computer Engineering, studying under the University of Mumbai have satisfactorily presented the project on “ ” as a part of the coursework of PROJECT-I for Semester-VII under the guidance of ***Dr. (Mrs.) Gresha Bhatia*** in the year 2023-2024.

14/10/2023

Date

Internal Examiner

External Examiner

Project Mentor

Dr. (Mrs.) Gresha Bhatia

Head of the Department

Dr. (Mrs.) Nupur Giri

Principal

Dr. J. M. Nair

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Deputy HoD **Dr. (Mrs) Gresha Bhatia** for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:-

Course Outcome	Description of the Course Outcome
CO 1	Do literature survey/industrial visit and identify the problem of the selected project topic.
CO2	Apply basic engineering fundamental in the domain of practical applications for problem identification, formulation and solution
CO 3	Attempt & Design a problem solution in a right approach to complex problems
CO 4	Cultivate the habit of working in a team
CO 5	Correlate the theoretical and experimental/simulations results and draw the proper inferences
CO 6	Demonstrate the knowledge, skills and attitudes of a professional engineer & Prepare report as per the standard guidelines.

ABSTRACT

Stroke is a debilitating neurological condition that affects millions of individuals worldwide, leading to a range of physical, cognitive, and emotional impairments. Early and accurate prediction of stroke outcomes and prognosis is crucial for optimizing patient care and rehabilitation planning. This project aims to develop an assistive tool that leverages the power of advanced technology to enhance stroke prognosis prediction.Our tool integrates cutting-edge feature extraction techniques with data from MRI images and clinical datasets to provide clinicians and healthcare professionals with a robust predictive framework. By extracting and analyzing pertinent features from both imaging and clinical data, this tool will enable better-informed decisions regarding treatment, post-stroke management, and long-term care.The key components of our project include:MRI Image Processing, Clinical Dataset Analysis, Machine Learning Models, User-Friendly Interface.By combining MRI image analysis and clinical dataset features, our assistive tool will provide a comprehensive view of a stroke patient's condition. This, in turn, will empower healthcare providers to make more personalized and effective decisions, thereby improving patient care and enhancing the rehabilitation and recovery process.

INDEX

Chapter 1 : Introduction

This section gives a brief overview of the project, focusing on performing prognosis and extracting features from MRI images and clinical data.

1.1 Introduction

Stroke is a major global health concern, with debilitating consequences for those affected and a significant burden on healthcare systems. According to the World Health Organization, stroke is the second leading cause of death and a leading cause of disability worldwide. The impact of a stroke extends far beyond the individual; it affects families, communities, and healthcare providers, making the accurate prediction of stroke outcomes and prognosis an imperative challenge. This project addresses the pressing need for more accurate and data-driven methods to predict stroke outcomes and prognosis. Stroke outcomes can vary significantly, ranging from minimal impairment to severe disabilities, and predicting the course of recovery is notoriously complex. Moreover, the effectiveness of rehabilitation and post-stroke care significantly depends on the ability to anticipate an individual's unique needs and potential for recovery.

The significance of this project lies in its potential to revolutionize stroke care and management by leveraging advanced technology, specifically MRI images and clinical data, to create an assistive tool that enhances stroke prognosis prediction. Accurate prognosis prediction offers several benefits, including with a more precise understanding of an individual's prognosis, healthcare providers can develop personalized treatment plans that address the specific needs and challenges a patient may face, hospitals and rehabilitation centers can optimize resource allocation by providing intensive care and rehabilitation services to patients who need them the most, predictive insights empower healthcare professionals to have informed discussions with patients and their families, alleviating anxiety and helping them prepare for the recovery journey, the data collected and analyzed through this tool can contribute to a broader understanding of stroke outcomes and potentially lead to innovations in stroke treatment and management.

1.2 Objectives

The primary objective of this project is to develop an assistive tool that combines advanced feature extraction techniques for MRI images and clinical data analysis to improve the accuracy of stroke prognosis prediction. Key project goals include:

- Developing robust MRI image processing algorithms for feature extraction.
- Analyzing clinical datasets to identify relevant features and indicators of stroke prognosis.
- Building machine learning models that integrate data from both sources to predict stroke outcomes.
- Designing a user-friendly interface for healthcare professionals to access and utilize predictive insights.

By achieving these objectives, this project aspires to empower healthcare providers with a valuable tool that enhances their ability to make well-informed decisions, ultimately improving patient care and contributing to the ongoing effort to alleviate the burden of stroke on individuals and society.

1.3 Drawback of the existing system

The existing systems for stroke outcome prediction solely rely on MRI images for analysis.However, this approach overlooks a crucial dimension of patient health: structured clinical data. By not incorporating information such as NIH scores, BMI measurements,age, gender, and demographic details, the existing systems miss out on a comprehensive understanding of the patient's condition. This limitation may lead to less accurate predictions and hinder the ability to provide personalized and precise medical recommendations.

1.3 Problem Definition

To identify the major causes of stroke and the related features associated with it by training and testing comprehensive datasets. The major features will be extracted from MRI images and other clinical information, which helps in the better diagnosis of the project.

1.4 Relevance Of The Project

The relevance of the project "Assistive Tool for Stroke Outcome and Prognosis Prediction using MRI Images and Clinical Dataset" lies in its potential to address critical challenges in stroke care, advance medical science, and significantly improve patient outcomes.

1. Improving Patient Care and Rehabilitation: Stroke is a life-altering event that affects not only the individuals who experience it but also their families and communities. The ability to accurately predict stroke outcomes and prognosis is crucial for tailoring treatment plans, optimizing resource allocation in healthcare settings, and providing emotional support to patients and their loved ones. By developing an assistive tool that harnesses the power of advanced technology to predict stroke outcomes, this project aims to empower healthcare providers with the knowledge they need to make more informed decisions. This tool will facilitate the design of personalized rehabilitation programs, enhancing the quality of care provided to stroke survivors and helping them regain their independence and quality of life.

2. Advancing Medical Research and Knowledge: Beyond its immediate benefits for patient care, this project also holds the potential to contribute to the broader field of medical research. The data collected and analyzed through the assistive tool can be a valuable resource for researchers studying stroke outcomes and related factors. By expanding our understanding of the variables that influence stroke recovery, this project can drive advancements in stroke treatment and management. Furthermore, the insights gained from this project could lead to the development of innovative therapies and interventions to enhance stroke recovery and, ultimately, reduce the long-term societal and economic burden of

stroke. In this way, the project is not only relevant on a clinical level but also has the potential to influence healthcare practices and policies.

In conclusion, the project's relevance is underscored by its potential to enhance patient care and rehabilitation while simultaneously contributing to the advancement of medical research and the development of more effective treatments and strategies for stroke management. It addresses a pressing healthcare challenge with far-reaching implications for stroke survivors and the healthcare community as a whole.

1.5 Methodology Used

To provide a holistic view of patient health, the system merges two independent datasets, one structured and one unstructured. The structured dataset contains a wide range of clinical factors, such as NIH scores, BMI measurements, smoking habits, demographic data, level of impairment, growth type and location. Input images, such as MRI scans and physician-provided scans, supplement this data by providing a detailed visual portrayal of the patient's condition.

Data Preprocessing:

- Image Normalization: This is applied to standardize pixel values, enhancing comparability across MRI images. This step ensures consistent data input for the subsequent analysis.
- Resizing: Resizing involves adjusting MRI images to a uniform resolution, facilitating efficient model training and improving convergence.
- Intensity Adjustment: Intensity adjustment techniques are used to enhance image contrast, making subtle features more discernible and aiding in accurate stroke detection.
- Removal of Metadata and Identifiable Information: By carefully eliminating metadata and identifiable patient information, data privacy is maintained, adhering to ethical standards in medical record handling.

Data Cleaning Techniques:

- Data Anonymization: Data anonymization protects patient privacy by removing personally identifiable information from medical records, ensuring compliance with confidentiality regulations.Normalization and Scaling: Numerical data is normalized and scaled to a common range, preventing features with larger scales from dominating the analysis.
- Duplicates Removal:Removing duplicated records ensures the dataset's integrity, preventing biases that might arise from redundant information.
- Outlier Detection:Outliers in data, potentially indicating measurement errors or anomalies, are identified and addressed to prevent skewed analysis.
- Missing Data Handling: Missing data is managed through imputation or removal, maintaining dataset completeness and improving model robustness.

Model Development

- Model Selection: Model selection involves choosing an appropriate deep learning architecture based on factors like complexity and compatibility with stroke detection tasks.
- Training and Evaluation: In model development, the selected model is trained using preprocessed data and evaluated on unseen test data to measure its performance accurately.

Data Visualization

- The data visualization component of our web application for stroke analysis serves as a critical interface for understanding the complex interaction between extracted features from MRI scan images and other relevant clinical information.
- The application transforms complex data sets into visually appealing representations by leveraging the tremendous capabilities of Tableau / Power BI.
- Users can traverse interactive dashboards that reveal relationships between MRI-derived parameters like lesion location and size and clinical data like NIH scores, BMI, smoking habits, and demographic information.

Chapter 2 : Literature Survey

This literature survey performed aims to comprehensively review the existing knowledge related to prognosis of stroke with the help of MRI Images. By this, we aim to provide comprehensive understanding of the current state of research in this field, identify challenges and highlight potential directions for further studies.

2.1 Research papers referred

1. Prognosis of ischemic stroke predicted by machine learning based on multi-modal MRI radiomics (2023)
   - Preprocessing techniques: skull stripping, lesion segmentation, normalization
   - Algorithms: Random forest, XGBoost, Logistic Regression
   - Stroke Outcomes: the model was able to correctly predict stroke prognosis in 83.1% of the cases.
   - Prognosis prediction: age, NIHSS score, and lesion location were independent predictors of stroke prognosis. This means that these factors were associated with stroke prognosis even after accounting for other factors.
   - Performance/Accuracy: LightGBM - 83.1%, random forest - 81.2%, XGBoost - 80.5%, SVM 79.8%, Logistic Regression - 78.1%

2. CNN-Res: deep learning framework for segmentation of acute ischemic stroke lesions on multimodal MRI images (2023)
   - Preprocessing techniques: image resizing, image normalization
   - Algorithms: CNN, Random forest classifier
   - Stroke Outcomes: the CNN-random forest classifier was able to predict stroke with an accuracy of 85.7% on a test set of 100 patients.

- Prognosis prediction: did not evaluate the ability of the CNN-random forest classifier to predict stroke prognosis.
- Performance/Accuracy: able to correctly predict stroke in 85.7% of the cases.

3. Mobile AI Stroke Health App: A Novel Mobile Intelligent Edge Computing Engine based on Deep Learning models for Stroke Prediction – Research and Industry Perspective (2021)
   - Algorithms: Sparse Auto-Encoders Deep Learning (DL) technique and Group Handling method (GMDH) neural networks.
   - Performance/Accuracy: Sparse Auto-Encoders reaches almost 98% for Stroke Diagnosis and GMDH Neural Networks proves to be a good technique for monitoring the EMG signal of the same patient case with average accuracies 98.60% to average 96.68% of the signal prediction.

4. Stroke Lesion Detection and Analysis in MRI Images Based on Deep Learning (2021)
   - Dataset: two local hospital datasets of 300 cases and over 1000 images. NLP for extracting lesion are from MRI reports and classifying them to predict.
   - Preprocessing techniques: Mark bounding box for lesions. Data analysis and visualization is performed
   - Algorithms: Faster R-CNN, YOLOv3, and SSD
   - Performance/Accuracy: Deep learning methods including Faster R-CNN, SSD, and YOLOv3 networks are conducted for automatic lesion detection with a precision of 89.77%.

5. Deep Learning-Based Acute Ischemic Stroke Lesion Segmentation Method on Multimodal MR Images Using a Few Fully Labeled Subjects (2021)
   - Preprocessing techniques: resampling, normalization, and data augmentation.
   - Algorithms: convolutional neural network (CNN) to segment stroke lesions on multimodal MRI images

- Stroke Outcomes: The CNN was able to achieve a mean dice coefficient of 0.84 on a test set of 179 subjects.
- Performance/Accuracy: 84% for CNN

6. Predicting outcome and recovery after stroke with lesions extracted from MRI images (2013)
   - Preprocessing techniques: skull stripping, lesion segmentation, normalization
   - Algorithms: Gaussian process model
   - Stroke Outcomes: correlation coefficient of 0.85 between the predicted and actual speech production scores.
   - Performance/Accuracy: high accuracy

7. MRI Radiomics and Predictive Models in Assessing Ischemic Stroke Outcome—A Systematic Review (2023)
   - Algorithms: Radionomics Analysis
   - Stroke Outcomes and Prognosis prediction: PROBAST tool was used to assess the risk of bias potential high risk of bias in participants selection was identified
   - Performance/Accuracy: Radiomics quality score (RQS) was also applied to evaluate the methodological quality of radiomics studies the results varying from an area under the ROC curve (AUC) of 0.80 (95% CI, 0.75–0.86) to an AUC of 0.92 (95% CI, 0.87–0.97 - reflecting a moderate methodological quality.

8. Machine Learning for Brain Stroke: A Review (2020)
   - Dataset : CT images are a frequently used dataset in stroke.
   - Algorithms: to classify state-of-arts on ML techniques for brain stroke into 4 categories based on

     their functionalities or similarity, and then review studies of each category systematically. A total of 39 studies were identified from the results of ScienceDirect web scientific database on ML for brain stroke from the year 2007 to 2019. Support Vector Machine (SVM) is obtained as optimal models in 10 studies for stroke problems

- Stroke Outcomes: SVM and Random Forests are efficient techniques used under each category.

9. Predicting the Severity of Neurological Impairment Caused by Ischemic Stroke Using Deep Learning Based on Diffusion-Weighted Images (2022)
   - Dataset : This retrospective study included 851 ischemic stroke patients (711 patients in the training

set and 140 patients in the test set).

   - Preprocessing techniques: The patients' NIHSS scores, which reflect the severity of neurological impairment, were reviewed upon admission and on Day 7 of hospitalization and were classified intotwo stages (stage 1 for NIHSS < 5 and stage 2 for NIHSS ≥ 5)
   - Algorithms: A 3D-CNN was trained to predict the stage of NIHSS based on different preprocessed DWI images
   - Performance/Accuracy: proposed model obtained better performance in predicting the NIHSS stage on Day 7 of hospitalization than that at admission (best AUC 0.895 vs. 0.846).
   - Proposed 3D-CNN model can effectively predict the neurological severity of IS using DWI images and performs better in predicting the NIHSS stage on Day 7 of hospitalization. The model also obtained promising performance in subgroup analysis, which can potentially help clinical decision making.

10. MRI radiomic features-based machine learning approach to classify ischemic stroke onset time (2021)
   - Aim: to investigate the ability of MRI radiomics features-based machine learning (ML) models to classify the time since stroke onset (TSS), which could aid in stroke assessment and treatment options.

- Dataset: This study involved 84 patients with acute ischemic stroke due to anterior circulation artery occlusion (51 in the training cohort and 33 in the independent test cohort).
- Preprocessing techniques: Region of infarct segmentation was manually outlined by 3D-slicer software. Image processing including registration, normalization and radiomics features calculation were done in R (version 3.6.1).
- Algorithms: A total of 4312 radiomic features from each image sequence were captured and used in six ML models to estimate stroke onset time for binary classification (≤ 4.5 h)
- Performance/Accuracy: The deep learning model-based DWI/ADC radiomic features performed the best for binary TSS classification in the independent test cohort, with an AUC of 0.754, accuracy of 0.788, sensitivity of 0.952, specificity of 0.500, positive predictive value of 0.769, and negative predictive value of 0.857, respectively. Furthermore, adding clinical information did not improve the performance of the DWI/ADC-based deep learning model.

11. Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: an automated machine learning approach (2023)
    - Dataset: The dataset comprised 4147 patients from a multicenter stroke registry, with 1268 (30.6%) experiencing unfavorable outcomes
    - Preprocessing techniques: Age, initial NIHSS, and early neurologic deterioration were identified as the most important clinical features, Image preprocessing
    - Algorithms: logistic regression (LR), random forest (RF), light gradient boosting machine (LGBM), and multi-layer perceptron (MLP), deep three-dimensional DenseNet (CNN model) on imaging data with three channels (DWI, ADC, and ground truth lesion mask),
    - Performance/Accuracy: We used the intersection over union (IoU) and Dice similarity coefficient

      (DSC) as performance evaluation metrics for lesion area detection in the segmentation model (Model S)

12. Development and clinical application of a deep learning model to identify acute infarct on magnetic resonance imaging (2022)
   - The ability to quickly identify the presence of acute infarct and quantify the volume on magnetic
   - resonance imaging (MRI) has important treatment implications
   - Dataset: It was trained on 6,657 MRI studies from Massachusetts General Hospital (MGH; Boston, USA)
   - Preprocessing techniques: All studies were labelled positive or negative for infarct (classification annotation) with 377 having the region of interest outlined (segmentation annotation).
   - Performance/Accuracy: The model performed better when trained on classification and segmentation annotations (area under the receiver operating curve [AUROC] 0.995 [95% CI 0.992–0.998] and median Dice coefficient for segmentation overlap of 0.797 [IQR 0.642–0.861]) compared to segmentation annotations alone (AUROC 0.982 [95% CI 0.972–0.990] and Dice coefficient 0.776 [IQR 0.584–0.857])

2.2 Interaction with domain experts

- Interaction with Dr. Bindu Menon, HOD and Prof. Neurology, Apollo Speciality Hospitals
- Discussion on the literature study presented during the meet by the student team and their mentors.
- Dr. Bindu Menon expressed her concerns on the dataset requirements.
- The discussion further led to the determination of prominent features that help in the detection of stroke.
- The challenging areas were explored in terms of the deterministic and nondeterministic locations.
- The features extracted from the MRI images cannot be the only factor that can be considered for the functional independence of accurate stroke diagnosis, was also discussed.

- Other prominent features need to be considered such as patient clinical data including comorbidities, demographic information, level of disability, the type of growth, location, severity using NIH score and different types of MRI images,were also discussed.
- Dr. Bindu Menon mentioned the stroke help app for the better understanding of stroke recovery steps.

2.3 Patent search

1. Patent Title : System and method for predicting stroke outcome using machine learning

- Patent Number: US11161972B2
- Publication Date: October 25, 2022
- Summary: This patent describes a system and method for predicting stroke outcome using machine learning. The system uses a variety of data sources, including MRI images, clinical data, and demographic data, to train a machine learning model to predict stroke outcome. The model can then be used to predict the outcome of future strokes, which can help clinicians to make informed decisions about patient care.

2. Patent Title: System and method for detecting and diagnosing stroke using artificial intelligence

- Patent Number: US11082276B2
- Publication Date: August 2, 2022
- Summary: This patent describes a system and method for detecting and diagnosing stroke using artificial intelligence (AI). The system uses a variety of data sources, including MRI images, clinical data, and demographic data, to train an AI model to detect and diagnose stroke. The model can then be used to detect and diagnose strokes in future patients, which can help clinicians to provide timely and effective care.

3. Patent Title: System and method for predicting stroke risk using machine learning

- Patent Number: US10948231B2
- Publication Date: March 15, 2022
- Summary: This patent describes a system and method for predicting stroke risk using machine learning. The system uses a variety of data sources, including clinical data, demographic data, and lifestyle data, to train a machine learning model to predict stroke risk. The model can then be used to predict the stroke risk of future patients, which can help clinicians to identify individuals who are at high risk for stroke and to implement preventive measures.

4. Patent Title: System and method for predicting stroke recurrence using machine learning

- Patent Number: US10905036B2
- Publication Date: February 1, 2022
- Summary: This patent describes a system and method for predicting stroke recurrence using machine learning. The system uses a variety of data sources, including MRI images, clinical data, and demographic data, to train a machine learning model to predict stroke recurrence. The model can then be used to predict the risk of stroke recurrence in future patients, which can help clinicians to develop personalized treatment plans and to monitor patients for signs of recurrence.

Chapter 3: Requirement Of Proposed System

The section gives a very detailed explanation of the working of our system and explains the intricacies of the technologies used.

3.1 Functional Requirements

- Collect and integrate a diverse dataset of MRI images and corresponding clinical data.
- Perform image preprocessing tasks like normalization, registration, and artifact removal for MRI images.
- Implement algorithms for extracting relevant features from MRI images, including texture, shape, and intensity characteristics.
- Enable the model to predict stroke outcomes and prognosis based on the integrated feature set.

3.2 Non-Functional Requirements

- The predictive model should exhibit a high level of accuracy in stroke outcome and prognosis predictions.
- Patient data must be stored and transmitted securely, adhering to privacy regulations and healthcare industry standards.
- The model's predictions should be interpretable and explainable, allowing clinicians to understand the factors influencing the outcomes.
- The system should integrate seamlessly with existing healthcare infrastructure, including electronic health record (EHR) systems.

3.3 Constraints

- The availability and capabilities of hardware resources, including processing power, memory, and storage, may impose constraints on the system's performance and scalability.
- Ethical guidelines and considerations may impose constraints on data collection, sharing, and research practices, particularly when dealing with sensitive patient information.
- Compatibility and integration with existing hospital or healthcare information systems may present technical challenges.
- The computational complexity of feature extraction and model training may impose constraints on the time required to process and analyze data.

3.4 Hardware and Software Requirements

Hardware Requirements:

- **Intel Core i5 Processor or Higher:**

  A powerful processor is crucial for handling the computational demands of feature extraction from MRI images, especially when using techniques like Convolutional Neural Networks (CNNs) or other deep learning methods. The processor's speed and efficiency directly impact the speed of data processing.
- **NVIDIA GPU:**

  A dedicated GPU from NVIDIA is specified, which indicates that the project may involve deep learning tasks. This is particularly important for accelerating the training and inference of deep learning models, which can be highly parallelizable and benefit greatly from GPU acceleration. It's especially relevant when working with CNNs for image analysis.
- **RAM >= 8GB:**

  Adequate RAM (Random Access Memory) is crucial for efficiently processing large datasets. In this project, with the potential need to load and manipulate high-resolution MRI images and clinical data simultaneously, having at least 8GB of RAM helps in ensuring smooth operations and prevents memory-related bottlenecks.

- **Laptop / Android / iOS based Smart Phones:**

  This indicates that the project is designed to be accessible on various platforms, including laptops, Android devices, and iOS devices. It suggests a user-friendly interface that is compatible with different operating systems. This ensures broader accessibility and ease of use for healthcare professionals who may have different preferences or work with different devices.

Software Requirements:

- **Operating System:**

  Windows, Linux, or macOS: The project should be compatible with these common operating systems to ensure widespread accessibility.

- **Python:**

  Python is a versatile and widely used programming language in data science and machine learning. It provides a rich ecosystem of libraries and frameworks for image processing, machine learning, and data analysis.

- **Integrated Development Environment (IDE):**

  Google Colab: Google Colab: Free cloud-based Jupyter notebooks with GPU/TPU support for coding and ML.

  Visual Studio Code: A lightweight and extensible code editor that supports Python development.

- **Machine Learning Libraries:**

  Key libraries for model development include:

  Scikit-learn: Provides a wide range of machine learning algorithms for classification, regression, and more.

  TensorFlow or PyTorch: For building and training deep learning models.

  Keras: A high-level neural networks API, often used with TensorFlow or Theano.

- **Image Processing Libraries:**

  Libraries like OpenCV and SimpleITK may be used for tasks like image preprocessing, registration, and feature extraction.

- **Version Control:**

  Git and platforms like GitHub or GitLab for version control and collaboration.

3.5 Techniques utilized till date for proposed system

The techniques utilized till date for the proposed system include data collection from the ATLAS 2.0 dataset, 3D image visualization using FSLeyes, and preprocessing tasks such as normalization, intensity adjustment, and image manipulation through slicing and cropping.

3.6 Tools utilized till date for the proposed system

- **FSLeyes:**

  FSLeyes is a specialized software tool designed for visualizing neuroimaging data, making it a valuable resource for viewing 3D medical images in this project.

- **Preprocessing Libraries and Tools:**

  Various Python libraries and tools are likely employed for preprocessing tasks such as normalization, intensity adjustment, slicing, and cropping. These could include:

  NumPy for numerical operations and array manipulation.

  OpenCV for image processing tasks like normalization and intensity adjustment.

  SimpleITK for more advanced medical image processing tasks.

- **Python Programming Language:**

  Python serves as the primary programming language for implementing these techniques, as it offers a wide range of libraries and tools suitable for handling medical image data and related Operations.

  These techniques and tools collectively form the initial stages of processing and analyzing the ATLAS 2.0 dataset for stroke outcome prediction in the proposed system.

3.7 Project Proposal

- **Objective:**

  Develop a predictive model using MRI images and clinical data for accurate stroke outcome and prognosis prediction.

- **Methodology:**

  Utilize advanced feature extraction techniques on MRI images and clinical data.

  Integrate features to build a machine learning model for outcome prediction.

  Evaluate and validate the model's performance using appropriate metrics.

- **Expected Deliverables:**

  Trained predictive model with documentation.

  Visualization and reporting of model performance.

  Comprehensive project documentation including code and methodologies.

- **Resources:**

  Hardware: Intel Core i5 Processor, NVIDIA GPU, 8GB RAM.

  Software: Python, TensorFlow, Keras, Scikit-learn, OpenCV, SimpleITK.

Chapter 4: Proposed Design

The section explains the architecture and gives a granular view of the working module along with the Gantt chart.

4.1 Block diagram of the proposed system

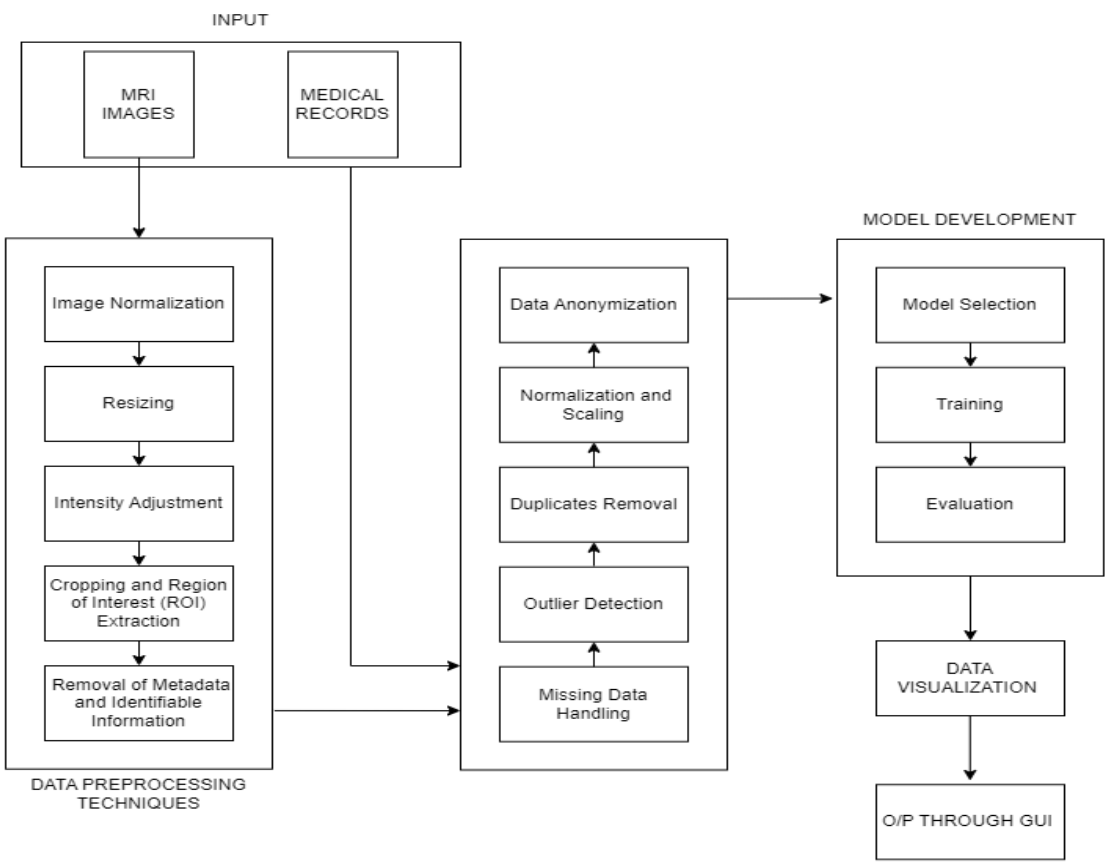


Figure 1: Block Diagram

- **Dataset Acquisition:** To provide a holistic view of patient health, the system merges two independent datasets, one structured and one unstructured. The structured dataset contains a wide range of clinical factors ,such as NIH scores, BMI measurements, smoking habits, demographic data, level of impairment, growth type and location. Input images, such as MRI scans and physician-provided scans, supplement this data by providing a detailed visual portrayal of the patient's condition.

- **Image Normalization:** Image normalization is applied to standardize pixel values, enhancing comparability across MRI Images. This step ensures consistent data input for the subsequent analysis.
- **Resizing:** Resizing involves adjusting MRI images to a uniform resolution, facilitating efficient model training and improving convergence.
- **Intensity Adjustment:**Intensity adjustment techniques are used to enhance image contrast, making subtle features more discernible and aiding in accurate stroke detection.
- **Removal of Metadata and Identifiable Information:** By carefully eliminating metadata and identifiable patient information, data privacy is maintained, adhering to ethical standards in medical record handling.
- **Data Anonymization:** Data anonymization protects patient privacy by removing personally identifiable information from Medical records, ensuring compliance with confidentiality regulations.
- **Normalization and Scaling:** Numerical data is normalized and scaled to a common range, preventing features with larger scales from dominating the analysis.
- **Duplicates Removal:** Removing duplicated records ensures the dataset's integrity, preventing biases that might arise from redundant information.
- **Outlier Detection:** Outliers in data, potentially indicating measurement errors or anomalies, are identified and addressed to prevent skewed analysis.
- **Missing Data Handling:**Missing data is managed through imputation or removal, maintaining dataset completeness and improving model robustness.
- **Model Selection:** Model selection involves choosing an appropriate deep learning architecture based on factors likecomplexity and compatibility with stroke detection tasks.

- **Training and Evaluation:** In model development, the selected model is trained using preprocessed data and evaluated on Unseen test data to measure its performance accurately.
- **Data Visualization-** The data visualization component of our web application for stroke analysis serves as a critical Interface for understanding the complex interaction between extracted features from MRI scan images and other relevant clinical information. The application transforms complex data sets into visually appealing representations by leveraging the tremendous capabilities of Tableau / Power BI. Users can traverse interactive dashboards that reveal relationships between MRI-derived parameters like lesion location and size and clinical data like NIH scores, BMI, smoking habits,and demographic information. Healthcare practitioners obtain a better knowledge of how these variables interact by using dynamic charts, graphs, and heatmaps, allowing them to spot patterns, trends, and potential risk factors related to stroke.

4.2. Modular diagram representation of the proposed system

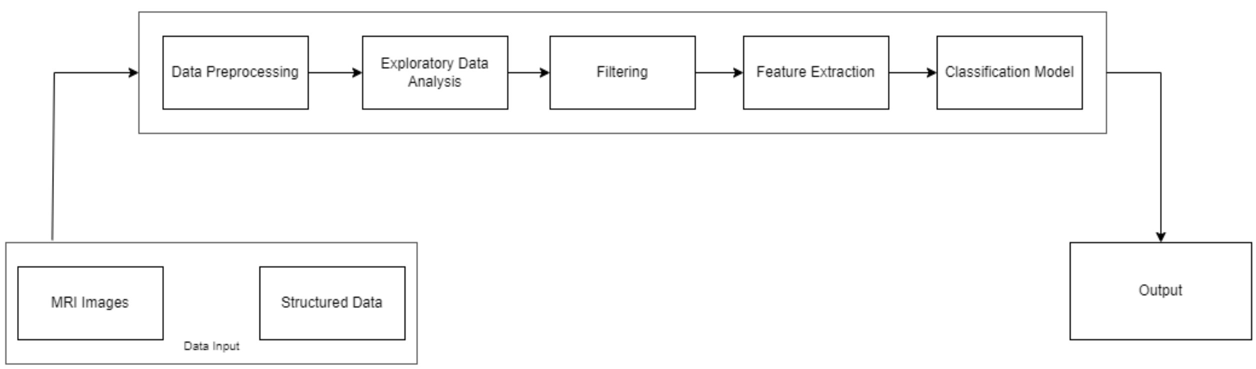


Figure 2: Modular diagram

- Data Input:
  MRI Images: These will be the primary source of image data for your analysis.
  Structured Data: Clinical information such as patient demographics, medical history, and other relevant information.

- Data Preprocessing:

  Data Cleaning: Removing any noisy or irrelevant data.

  Image Preprocessing: Enhancing and standardizing MRI images if necessary.

  Data Integration: Combining MRI images with structured data for analysis.

- Exploratory Data Analysis (EDA):

  This step involves data visualization and statistical analysis to gain insights into your data. It can help you understand the distribution of data, relationships between variables, and identify potential patterns or outliers.

- Filtering: Filtering can involve the removal of irrelevant data or noise from your dataset to improve mode performance.

- Feature Extraction: In the context of MRI images, feature extraction could involve techniques like extracting texture features, shape features, or specific regions of interest from the images. For structured data, you can engineer new features or select relevant features that might be useful in the classification task.

- Classification Model: Build a machine learning or deep learning model for the classification task. In your case, you're trying to identify the major causes of stroke.

  You can use various algorithms such as Random Forest, Support Vector Machines, or Convolutional Neural Networks (CNNs) for image data.

- Training and Testing: Split your dataset into training and testing sets to evaluate the model's performance.You may use techniques like cross-validation to ensure the model's robustness.

- Web App: Develop a web application to provide a user-friendly interface for running the trained model. Users can input MRI images and clinical data, and the app will provide predictions or diagnostic results.

- Output: The web app will display the results to the users, such as the identified causes of stroke and any relevant features associated with them.

- Project Diagnosis: Use the model's predictions and insights to aid in the diagnosis of stroke causes and support medical professionals in their decision-making.

4.3 Design of the proposed system with proper explanation of each

Data Flow Diagram

Level 0

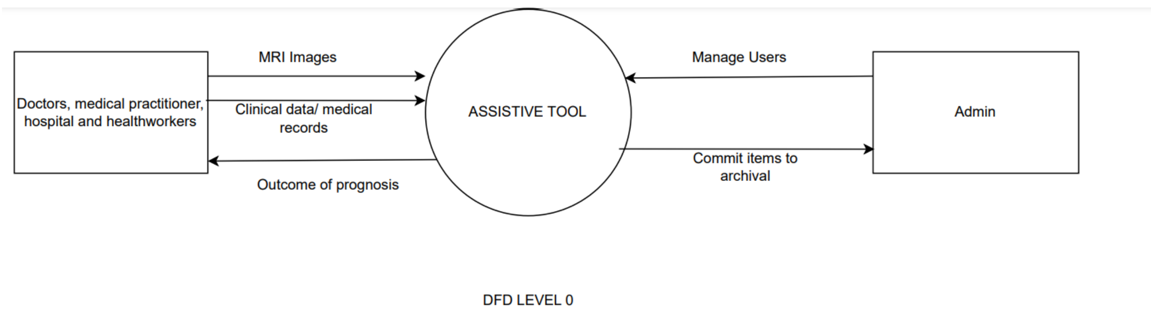


Figure 3: DFD level 0

The end users of this tool will be the doctors, medical practitioners, hospitals and health workers. They will be giving MRI Images of the patient and medical records as the input to the Assistive tool and will receive the prognosis outcome as the output. The administrator will be incharge of managing users and storing, maintaining the results in an archival, from where these previously attained outcomes can be easily retrieved.

Level 1

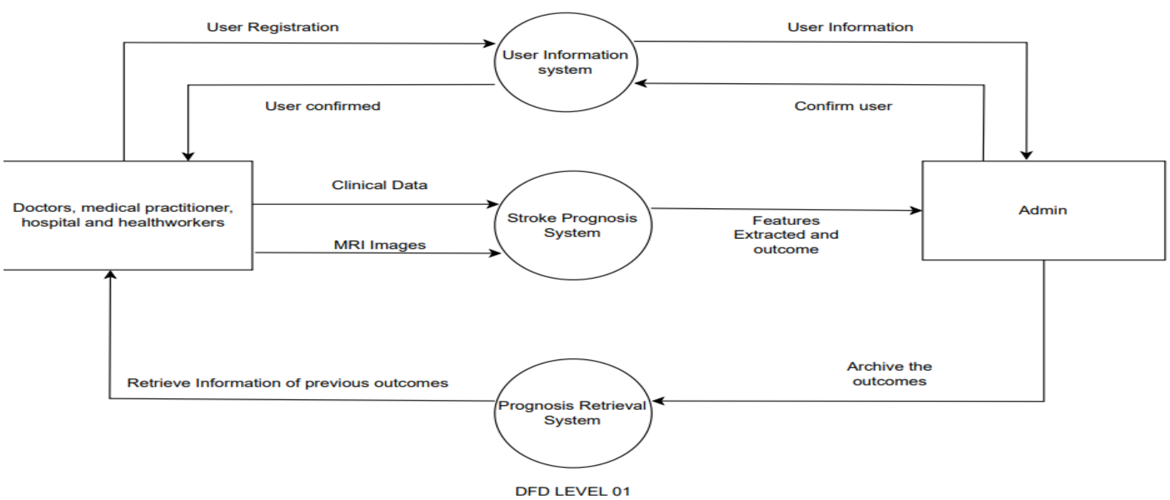


Figure 4: DFD- level 1

The project consists of three systems:

User Information System:- The user registers him or herself and this information is sent to the administrator via the user information system. The admin then confirms the user and this confirmation is sent back to the registered user who can now give inputs to the Stroke prognosis system.

Stroke Prognosis System:- This takes the MRI Images, clinical data as input and attains the features extracted as the outcome which is also sent to the administrator to commit to the archival database.

Prognosis Retrieval System:- This system consists of the previous outcomes i.e the prognosis and features extracted from the data and MRI Images. It is written by the admin and can be accessed by the user if they are authenticated.

Level 2

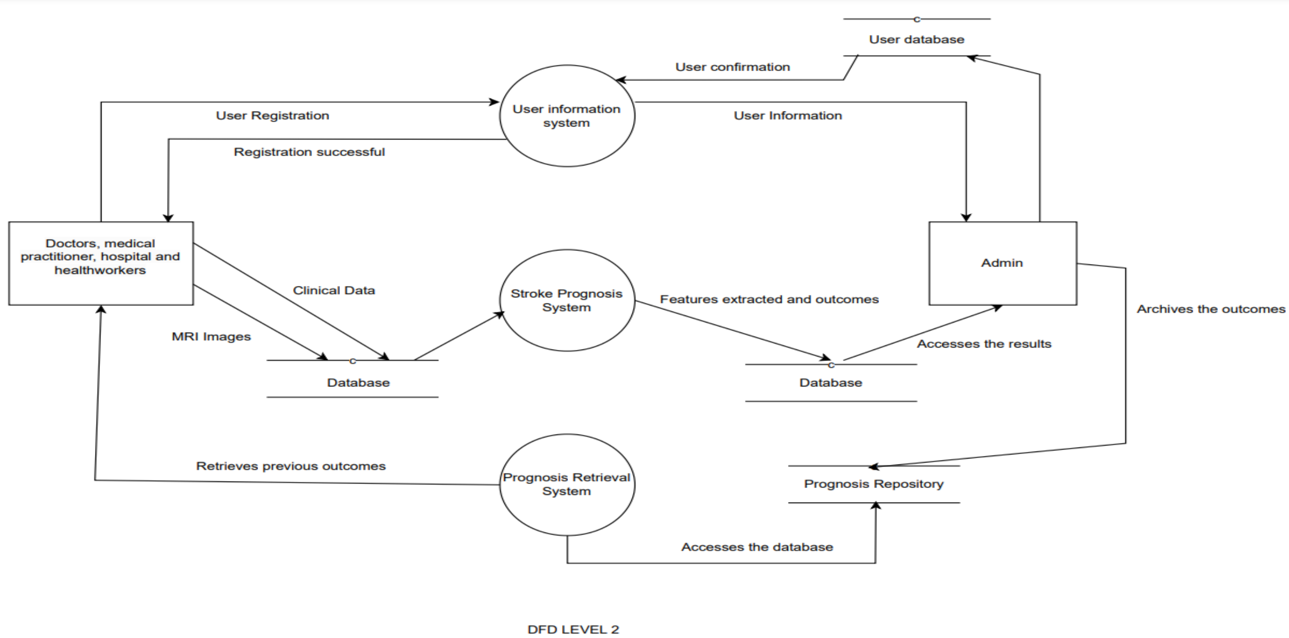


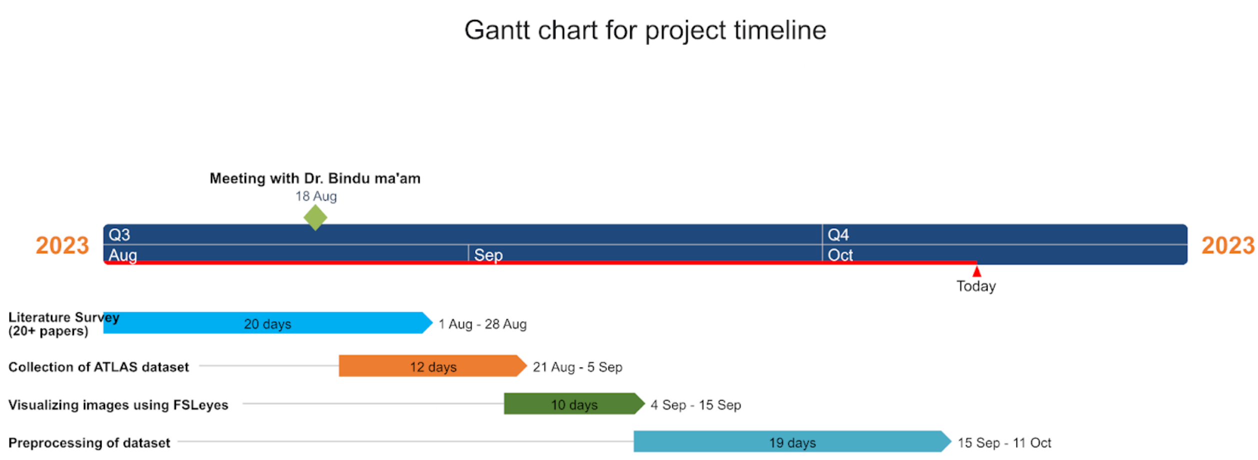
Figure 5 : DFD- level 2

The user information is stored in the user database where it can be accessed by the administrator. On storing the information of a new user in the database, he or she receives the confirmation of an authenticated user and can access the Stroke Prognosis System.

The MRI Images and Clinical data are stored in the database from where, the input to the model is provided. The outcomes and features extracted are again stored in a database which can be accessed by the administrator.

The admin can archive these results in the prognosis repository and the user can access these outcomes from the database using the Prognosis Retrieval System.

4.4. Project Scheduling & Tracking using Timeline / Gnatt Chart



Chapter 5: Proposed Results and Discussions

So far, we have analyzed some MRI images obtained from the ATLAS dataset. The dataset contains the 3D MRI images of stroke patients. Along with the images, we also have a file which mentions the image name and the lesion location in medical terms. By assessing this, we are able to manually identify the location of the lesion.

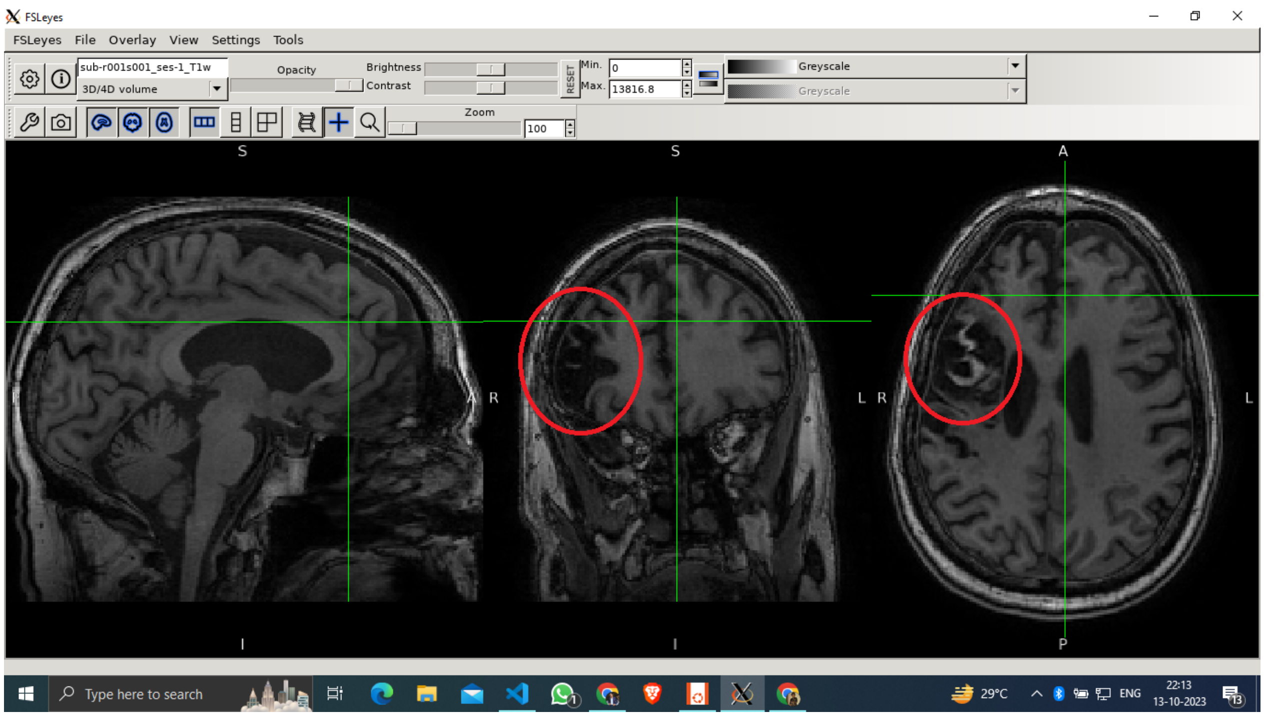


Figure 6:The location of the lesion is highlighted manually

We have used various image pre-processing techniques so far like normalization, resizing and intensity adjustment. All these techniques allow us to reduce the overall cost when we run the algorithm.

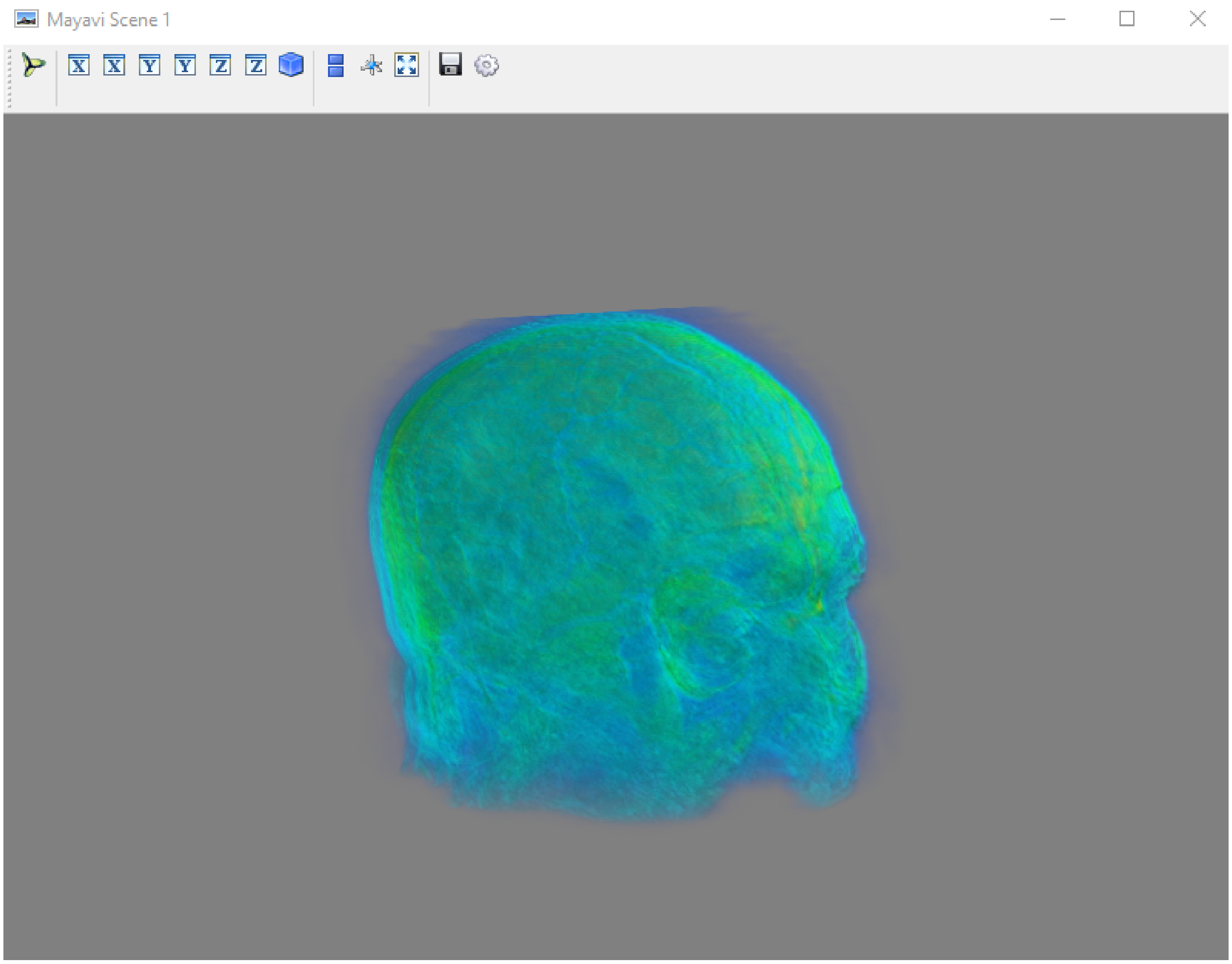


Figure 7: The volumetric intensity of the image

Given a 3D image, we can extract slices across all the 3 axes so that viewing the image becomes a lot easier. Since images have a lot of slices, we extract only a subset of the slices (like every 20th slice).

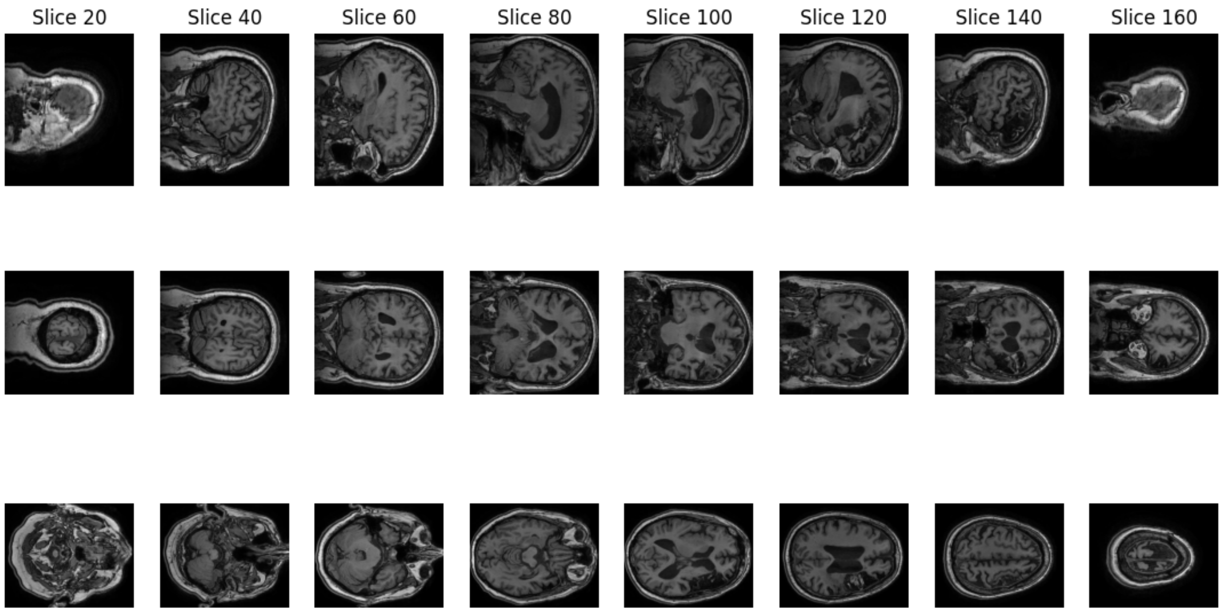


Figure 8: 3D images viewed in 2D slices before resizing

Resizing enables us to make sure the image slices are consistent across dimensions.

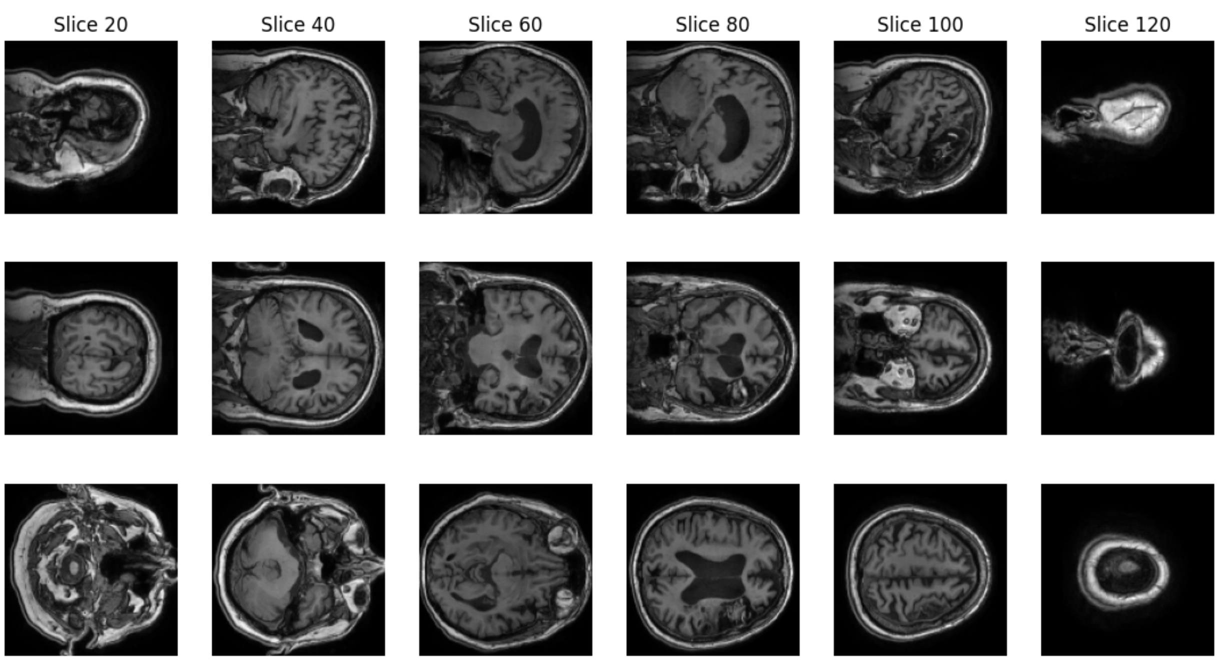


Figure 9: The resized image has less slices than the original image

Next, we apply intensity adjustment on the images. The primary goal of intensity adjustment is to enhance the visibility of specific features or structures within an image by mapping the original pixel intensity values to a new range of values.

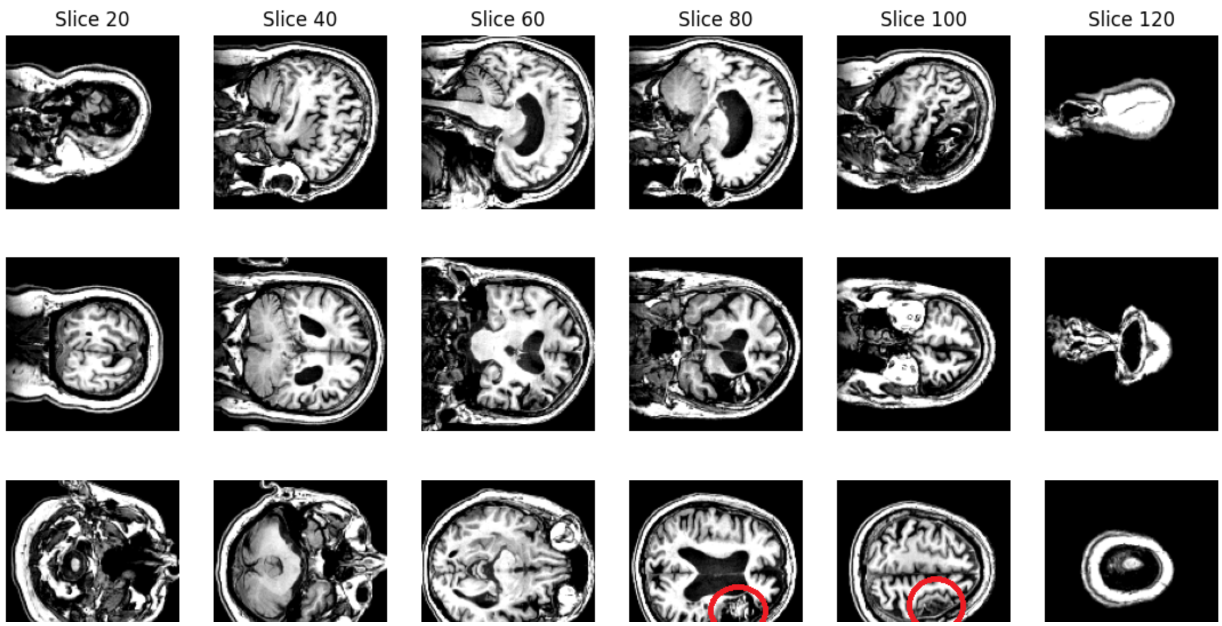


Figure 10: The intensity-adjusted images

We can see that the lesion is possibly present in the last rows of slice 80 and 100. It corresponds to the area which we saw earlier in Figure

Chapter 6 : Plan Of Action For the Next Semester

6.1 Work done till date

- Obtained MRI images dataset via ATLAS
- Performed data preprocessing : normalization, resizing and intensity adjustment. All these techniques allow us to reduce the overall cost when we run the algorithm.

6.2 Plan of action for project II

- Choose appropriate deep learning architecture based on complexity and compatibility for stroke feature detection/extraction tasks.

- Train selected model on preprocessed data and rigorously evaluated on unseen test data for accurate performance assessment.

- Use Tableau to perform Data Visualization and create interactive dashboards for visualizing complex interactions between MRI features and clinical data.

- Feature Detection:- Isolating relevant information to facilitate analysis. Techniques like Convolutional Neural Networks (CNNs) identify patterns, edges, and textures. This process aids in detecting structures, anomalies, or characteristics crucial for medical diagnosis and research.

Chapter 7 : Conclusion

In summary, our approach combines detailed clinical data with MRI images to predict stroke outcomes effectively. We carefully prepared the data for analysis, ensuring consistency, privacy, and completeness. Our chosen deep learning models, namely RCNN, 3D-CNN, and LSTM, excel in extracting important details from MRI images. Through clear visualization, we make it easy to understand how clinical data and MRI features relate. This method shows great potential in improving stroke assessment and enhancing medical diagnostics for better patient care.

Chapter 8 : References

[1] H. Yu et al., “Prognosis of ischemic stroke predicted by machine learning based on multi-modal MRI radiomics,” Frontiers in Psychiatry, vol. 13, Jan. 2023, doi: 10.3389/fpsyt.2022.1105496.

[2] Y. Gheibi, K. Shirini, S. N. Razavi, M. Farhoudi, and T. Samad-Soltani, “CNN-Res: deep learning framework for segmentation of acute ischemic stroke lesions on multimodal MRI images,” BMC Medical Informatics and Decision Making, vol. 23, no. 1, Sep. 2023, doi: 10.1186/s12911-023-02289-y.

[3] “Mobile AI Stroke Health App: A Novel Mobile Intelligent Edge Computing Engine based on Deep Learning models for Stroke Prediction – Research and Industry Perspective,” IEEE Conference Publication | IEEE Xplore, Oct. 29, 2021. https://ieeexplore.ieee.org/document/9811307

[4] S. Zhang, S. Xu, L. Tan, H. Wang, and J. Meng, “Stroke lesion detection and analysis in MRI images based on deep learning,” Journal of Healthcare Engineering, vol. 2021, pp. 1–9, Apr. 2021, doi: 10.1155/2021/5524769.

[5] B. Zhao et al., “Deep Learning-Based Acute Ischemic Stroke Lesion segmentation method on multimodal MR images using a few fully labeled subjects,” Computational and Mathematical Methods in Medicine, vol. 2021, pp. 1–13, Jan. 2021, doi: 10.1155/2021/3628179.

[6] T. M. H. Hope, M. L. Seghier, A. Leff, and C. J. Price, “Predicting outcome and recovery after stroke with lesions extracted from MRI images,” NeuroImage: Clinical, vol. 2, pp. 424–433, Jan. 2013, doi: 10.1016/j.nicl.2013.03.005.

[7] H. Dragos et al., “MRI Radiomics and Predictive Models in Assessing Ischemic Stroke Outcome—A Systematic Review,” Diagnostics, vol. 13, no. 5, p. 857, Feb. 2023, doi: 10.3390/diagnostics13050857.

[8] M. S. Sirsat, E. Fermé, and J. Câmara, “Machine Learning for Brain Stroke: A review,” Journal of Stroke and Cerebrovascular Diseases, vol. 29, no. 10, p. 105162, Oct. 2020, doi: 10.1016/j.jstrokecerebrovasdis.2020.105162.

[9] Y. Zhang et al., “MRI radiomic features-based machine learning approach to classify ischemic stroke onset time,” Journal of Neurology, vol. 269, no. 1, pp. 350–360, Jul. 2021, doi: 10.1007/s00415-021-10638-y.

[10] H. Jo et al., “Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: an automated machine learning approach,” Scientific Reports, vol. 13, no. 1, Oct. 2023, doi: 10.1038/s41598-023-44201-8.

[11] C. P. Bridge et al., “Development and clinical application of a deep learning model to identify acute infarct on magnetic resonance imaging,” Scientific Reports, vol. 12, no. 1, Feb. 2022, doi: 10.1038/s41598-022-06021-0.

[12] Y. Zeng, L. Chen, W. Zhao, and J. Liu, “Predicting the severity of neurological impairment caused by ischemic stroke using deep learning based on Diffusion-Weighted images,” Journal of Clinical Medicine, vol. 11, no. 14, p. 4008, Jul. 2022, doi: 10.3390/jcm11144008.

Chapter 9. Appendix

9.1 List of Figures