STROKE CLASSIFICATION AND DETECTION USING NEUROIMAGE-BASED ML MODELS

A PROJECT REPORT

Submitted by

P NAVEEN [RA2111026010085]
M POORNA PRADEEP REDDY [RA2111026010094]

Under the Guidance of

Dr. G. SENTHIL KUMAR

Associate Professor Department of Computational Intelligence

in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE ENGINEERING
with specialization in Artificial intelligence
and Machine Learning

DEPARTMENT OF COMPUTATIONAL INTELLIGENCE
COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE ANDTECHNOLOGY
KATTANKULATHUR- 603 203

May 2025

Department of Computational Intelligence
SRM Institute of Science & Technology
Own Work* Declaration Form

This sheet must be filled in (each box ticked to show that the condition has been met). It must be signed and dated along with your student registration number and included with all assignments you submit – work will not be marked unless this is done.

To be completed by the student for all assessments

Degree/ Course : B.Tech / Computer Science and Engineering with AI &ML

Student Name : P.NAVEEN , M. POORNA PRADEEP REDDY

Registration Number : RA2111026010085, RA2111026010094

Title of Work : Stroke Classification and Detection Using Neuroimage-Based ML Models

I / We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism, as listed in the University Website, Regulations, and the Education Committee guidelines.

I / We confirm that all the work contained in this assessment is my / our own except where indicated, and that I / We have met the following conditions:

- Clearly referenced / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook /University website

I understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:	
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above.	
P. NAVEEN RA2111026010085	M. POORNA PRADEEP REDDY RA2111026010094

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that 18CSP107L - Minor Project [18CSP108L- Internship] report titled "**Stroke Classification and Detection Using Neuroimage-Based ML Models**" is the Bonafide work of "**P.NAVEEN [RA2111026010085], M. POORNA PRADEEP REDDY [RA2111026010094]**" who carried out the project work[internship] under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr. G. Senthil Kumar	**DR. R. ANNIE UTHRA**
SUPERVISOR & ASSOCIATE PROFESSOR	**PROFESSOR & HEAD**
DEPARTMENT OF COMPUTATIONAL INTELLIGENCE	**DEPARTMENT OF COMPUTATIONAL INTELLIGENCE**

Internal Examiner **External Examiner**

ACKNOWLEDGEMENTS

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Scienceand Technology, for the facilities extended for the project work and his continued support. We extend our sincere thanks to **Dr. Leenus Jesu Martin M ,** Dean-CET, SRM Institute of Science and Technology, forhis invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Com putting, SRMInstitute of Science and Technology, for her support throughout the project work.

We encompass our sincere thanks to , **Dr. M. Pushpalatha**, Professor and Associate Chairperson, School of Computing and **Dr. C. Lakshmi,** Professor and Associate Chairperson, School of Computing, SRM Institute of Science and Technology, for their invaluable support.

We are incredibly grateful to our Head of the Department, **Dr. R. Annie Uthra**, Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, for her suggestions and encouragement atall the stages of the project work.

We want to convey our thanks to our Project Coordinators, Panel Head, and Panel Members, Department of Computational Intelligence, SRM Institute of Science and Technology, for their inputs during theproject reviews and support.

We register our immeasurable thanks to our Faculty Advisor **Dr. S. Amudha** , Department of Computational Intelligence, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. G. Senthil Kumar**, Department of Computational Intelligence, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under his / her mentorship. He / She provided us with the freedom and support to explore the research topics of our interest. His / Her passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff and students of Computational Intelligence, School of Computing, S.R.M Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents,family members, and friends for their unconditional love, constant support and encouragement.

P NAVEEN

M POORNA PRADEEP REDDY

ABSTRACT

This project applies machine learning (ML) approaches to automatic stroke detection and automatic stroke classification based on publicly accessible neuroimaging data, principally computed tomography (CT) scans. Stroke is a large public health challenge worldwide, and it is one of the leading causes of mortality and long-term disability. Appropriate treatment hinges on correct and timely diagnosis, yet typical diagnostics such as neuroimage interpretation through manual interpretation of images are dependent on equipment and specialist availability and not always conveniently available at resource-poor and low-resource locations in remote sites. The objectives of this study are to produce a low-budget and high-scale solution based on ML models which facilitate early stroke diagnosis and classification of stroke subtype with the overall goal of improving patient outcome and treatment. The system must have the capability of distinguishing between hemorrhagic and ischemic stroke as the two classes of stroke must be treated differently. The system combines standard machine learning classifiers such as Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) with more enhanced deep models such as Convolutional Neural Networks (CNN). These models acquire the ability of sweeping through the CT scans and therefore reduce their dependence on human bias or the extraction of manual features. The image preprocessing part of this task is very integral and involves image size standardization, intensity normalization, and removing noise.

All these operations enable the data to be consistent as well as to improve the efficiency of the model. Due to reasons of the class imbalance within stroke data and particularly the sparse hemorrhages, data augmentation measures such as data synthesis, flipping, and rotating are employed. Cost-sensitive training and weighted loss functions can as well be enforced to further address class imbalance. For enhancing model robustness, CNN models will use transfer learning during the fine-tuning of pre-trained models for stroke classification. Through this, accurate training on small-scale datasets is ensured and also taking advantage of the learned features of large-scale image databases. The model's performance will be evaluated with advanced metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) to ensure fair comprehension of sensitivity and specificity. The research will establish a robust and efficient ML-based diagnostic system that can be used to enable early stroke detection and subtype classification. With a simple and fast stroke diagnosis system, the proposed system can empower clinicians, reduce decision-making complexities, and ultimately lower the burden of stroke-related morbidity and mortality.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

KNN	K Nearest Neighbours
CNN	Convolutional Neural Networks
MRI-scan	Magnetic Resonance Imaging Scan
GPU	Graphics Processing Unit
CT-scan	Computed Tomography
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 Introduction to Stroke Classification and Detection Using Neuroimage-Based ML Models

Stroke was recognized as a critical neurological condition that resulted from an abrupt interruption in cerebral blood flow. This disruption often led to severe consequences, including long-term disability or death, making stroke a major global health concern[1]. The sudden onset of stroke symptoms demanded urgent medical intervention, as early treatment significantly influenced patient outcomes. Therefore, the timely identification and classification of stroke were considered essential for reducing morbidity and mortality rates associated with the condition [2].

Traditionally, medical imaging techniques such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) had been widely employed for stroke diagnosis. These modalities provided detailed insights into brain structures and were capable of distinguishing between different types of strokes, namely ischemic and hemorrhagic [3]. Despite their diagnostic efficacy, these imaging tools were often limited by various factors including high costs, equipment accessibility, and the requirement for expert radiological interpretation. Consequently, patients in resource-limited settings were frequently unable to benefit from early and accurate stroke diagnoses [4].

To address these limitations, the development of a machine learning (ML)-based platform for stroke detection and classification was proposed. This approach was intended to leverage open-access neuroimaging datasets, primarily using CT scans, to train and validate ML models capable of identifying acute stroke cases [5]. By focusing on a model that could operate with minimal infrastructure and without the need for highly specialized knowledge, the initiative aimed to democratize access to stroke screening and facilitate more rapid clinical decision-making in underserved regions [6].

During this research, a range of ML algorithms was explored, from conventional classification techniques to advanced deep learning frameworks. These models were trained to not only detect the presence of a stroke but also to classify its type accurately [7]. Particular emphasis was placed on differentiating between ischemic and hemorrhagic strokes, as each required distinct clinical intervention. The performance of these models was evaluated through rigorous validation methods to ensure reliability and clinical relevance [8].

The outcome of this research was envisioned to contribute to the creation of a practical, scalable tool that could be implemented within various healthcare systems [9]. By offering an accessible and efficient means of stroke detection, this platform had the potential to enhance the speed and accuracy of stroke diagnosis, particularly in regions lacking specialized medical personnel and equipment. Ultimately, the proposed solution was expected to support timely treatment initiation and improve health outcomes for stroke patients globally [10].

1.2 Motivation

The motivation for this project was derived from the pressing need to improve the speed, accuracy, and accessibility of stroke diagnosis, particularly in settings where medical resources were scarce. Stroke had been identified as one of the leading causes of death and long-term disability across the globe. The window for effective intervention following stroke onset was narrow, making rapid and accurate diagnosis essential for minimizing brain damage and improving patient outcomes. However, delays in diagnosis were frequently observed, especially in under-resourced regions, where infrastructure and skilled personnel were often lacking.

In routine clinical practice, the diagnosis of stroke continued to rely heavily on the manual interpretation of CT and MRI scans by experienced radiologists. While this method had proven to be effective in many settings, it was also inherently time-consuming and subject to human error. Diagnostic accuracy was influenced by the expertise and experience of the interpreting radiologist, which introduced variability and potential for misdiagnosis. In emergency situations, this dependence on human interpretation posed a significant challenge, potentially delaying critical interventions.

Moreover, in rural and economically disadvantaged regions, access to advanced neuroimaging equipment and trained radiologists remained limited or nonexistent. As a result, many stroke cases either went undetected or were identified too late for effective treatment. The disparity in diagnostic capabilities between urban and rural health systems highlighted a serious gap in equitable stroke care. Existing automated systems designed for stroke detection were found to lack the necessary levels of sensitivity, reliability, and the ability to accurately classify stroke subtypes, further compounding this issue.

These challenges underscored the need for a more reliable, scalable, and accessible diagnostic solution. It was believed that machine learning (ML) could be harnessed to address these shortcomings, by offering a data-driven approach to stroke classification and detection. ML algorithms, when trained on large, high-quality neuroimaging datasets, had the potential to replicate and even surpass human-level accuracy in image interpretation. Additionally, ML models could be deployed in low-resource environments, providing immediate diagnostic assistance without requiring constant expert oversight.

Thus, a project was proposed that aimed to develop a robust, cost-effective, and efficient ML-based platform for stroke detection and classification. Publicly available CT scan datasets were to be used to train and validate these models, ensuring both accessibility and reproducibility. By leveraging advanced ML techniques, this initiative sought to improve diagnostic speed and accuracy, reduce dependence on specialist expertise, and ultimately enable better-informed and timelier clinical decisions. This approach was envisioned as a vital step toward bridging diagnostic gaps and improving stroke care outcomes worldwide.

1.3 Sustainable Development Goal of the Project

This research project was strategically aligned with **Sustainable Development Goal (SDG) 3: Good Health and Well-being**, which aimed to ensure healthy lives and promote well-being for all individuals at all ages. Given that stroke remained a leading cause of death and long-term disability worldwide, urgent innovation in diagnostic and treatment pathways had been necessitated. By proposing a machine learning-based diagnostic platform utilizing neuroimaging data, the project was positioned to support the global effort to enhance healthcare systems. Early detection through automated imaging analysis was expected to enable faster clinical response, thereby promoting better health outcomes and contributing to universal well-being.

A direct contribution was made to Target 3.4, which aimed to reduce premature mortality from non-communicable diseases (NCDs) through prevention, treatment, and mental health promotion. As stroke represented a major NCD, the project addressed the need for timely and accurate diagnosis using machine learning algorithms to overcome limitations associated with manual interpretation. This approach was intended to minimize diagnostic delays, reduce disability, and increase survival rates, thereby reinforcing the objective of reducing the global burden of NCDs.

Additionally, alignment with Target 3.8, which emphasized universal health coverage and access to essential services, was achieved through the project's use of open-source neuroimaging datasets and the development of a low-cost, infrastructure-light implementation. The proposed model was designed to function with minimal reliance on expert radiologists, making it especially suitable for underserved and low-resource settings. By integrating artificial intelligence into clinical workflows, this initiative aimed to reduce healthcare disparities and deliver equitable, scalable, and sustainable stroke diagnostic solutions worldwide.

CHAPTER 2

LITERATURE SURVEY BRAIN STROKE DETECTION

S.NO	TITLE	METHODOLOGY	Identification of gaps and limitations
1.	"J. Chaki and M. Woźniak, Deep Learning and Artificial Intelligence in Action : A Review on Brain Stroke Detection, Diagnosis, and Intelligent Post-Stroke Rehabilitation Management"	• The authors reviewed research from 2019 to 2023 that used artificial intelligence (AI) and deep learning for detecting and diagnosing strokes from brain images. • They compared different methods and explained how well these AI systems work in stroke treatment and recovery.	• Most AI models work well on certain datasets but don't perform as well when tested on different types of brain scans or patient groups. • It is still difficult to use these AI systems in real hospitals because of challenges like high costs, complex models, and privacy concerns.
2.	"M. A. Saleem et al., Innovations in Stroke Identification: A Machine Learning-Based Diagnostic Model Using Neuroimages".	• The researchers developed a machine learning model that analyzes neuroimaging data, such as CT and MRI scans, to detect and classify strokes. • They used image preprocessing, feature extraction, and classification algorithms (like CNNs) to build a system that can automatically identify stroke types.	• The system's accuracy depends on the quality and diversity of the data; limited or imbalanced datasets can affect performance. • The model still needs testing in real clinical environments to ensure it works well outside of research settings.

3.	"F. Akbarifar, S. P. Dukelow, A. Jin, P. Mousavi and S. H. Scott, Optimizing Stroke Detection Using Evidential Networks and Uncertainty-Based Refinement".	• The authors used evidential networks, a type of machine learning model, to detect strokes by incorporating uncertainty into the decision-making process, improving model reliability. • They focused on refining the model by using uncertainty-based techniques that allow the system to adjust and improve its predictions as new information is processed.	• The system may struggle with accurately detecting strokes in cases where the data is noisy or lacks sufficient detail. • More work is needed to improve how the model handles real-world variability, such as different types of brain scans or patient conditions.
4.	"T. Janyalikit and C. A. Ratanamahatana, Time Series Shape let-Based Movement Intention Detection Toward Asynchronous BCI for Stroke Rehabilitation".	• The authors developed a time series shapelet-based approach to detect movement intentions in stroke patients, focusing on detecting brain activity related to voluntary movements. • They integrated this method into a brain-computer interface (BCI) system for asynchronous stroke rehabilitation, aiming to help patients control devices or exoskeletons based on their movement intentions.	• The method relies on the accurate detection of movement intentions, which can be challenging when the brain signals are weak or noisy. • The system still needs further validation to ensure it works effectively across diverse patient populations and different levels of stroke severity.

5.	“Y. Zhou et al., In Vivo Transcranial Acoustoelectric Brain Imaging of Different Deep Brain Stimulation Currents".	• Studied live imaging of sound and electric signals in the rat brain with different DBS currents. • This research tested four types of DC brain stim to find which best used the brain stim, reaching a spatial detail of 2 mm and a time detail of 10 ms.	• Limited to animal models (rat brains), which may not fully reflect human brain responses. The research's spatial and temporal resolution, while impressive, is based on current limitations in imaging technology.
6.	“C. Chen et al.,Tracking the Immediate and Short-Term Effects of Continuous Theta Burst Stimulation on Dynamic Brain States".	• They used cTBS on 22 healthy people. We checked brain activity with EEG at three times: before treatment, right after cTBS, and 80 minutes later. • They found changing brain states using a Hidden Markov Model (HMM) and saw that cTBS led to changes in brain states and better motor skills.	• The research only looks at people who are well, so the results may not fit sick groups. More studies are needed to check these results with different sick people.
7.	"A Siamese Convolutional Neural Network for Identifying Mild Traumatic Brain Injury and Predicting Recovery".	• This research builds a framework using Siamese Convolutional Neural Network (SCNN) for finding mild traumatic brain injury (mTBI) and watching how one gets better over time. • It uses pictures of the brain from mice and offers high accuracy in classifying (96.5%) and provides insights into recovery tracking.	• The framework was tested only on animal models, limiting its applicability to human cases. • The model's generalizability across various mTBI severity levels and recovery timelines requires further investigation.

8.	"L. Hu et al., Subject-Independent Wearable P300 Brain–Computer Interface Based on Convolutional Neural Network and Metric Learning".	• Created a wearable P300 brain-computer tool that works for anyone using CNNs and metric learning techniques. • Achieved high accuracy without calibration (73.23%) and even better with fine-tuning (78.75%).	• While the method demonstrates feasibility, the generalization across diverse populations with varying EEG characteristics is still uncertain. • The system's performance might degrade in real-world, highly variable environments.
9.	"S. Mamoon, Z. Xia, A. Alfakih and J. Lu, UCLN: Toward the Causal Understanding of Brain Disorders With Temporal Lag Dynamics".	• Introduced a Unified Causal and Temporal Lag Network (UCLN) for estimating causal relationships and temporal lags in brain networks using resting-state fMRI. • The approach outperforms traditional methods in Alzheimer's classification and identifies key biomarkers.	• The research focuses on Alzheimer's and may not fully apply to other neurodegenerative diseases. • The method's dependency on accurate temporal lag estimation may introduce challenges in interpreting complex brain dynamics.
10.	"Distinct Time-Resolved Brain-Wide Coactivations in Oxygenated and Deoxygenated Hemoglobin".	• The research compares resting-state brain activity in terms of oxygenated (HbO) and deoxygenated hemoglobin (HbR) using diffuse optical tomography. • It highlights the differences in time-resolved coactivation patterns (CAPs) and reveals that HbR better preserves brain-wide neuronal activations.	• The method requires advanced equipment like functional near-infrared spectroscopy (fNIRS), which may limit its accessibility in clinical settings. • The research does not account for the potential influence of other variables like age and neurodegenerative diseases.

11.	"Khan, A. F., et al. (2024). Distinct Time-Resolved Brain-Wide Coactivations in Oxygenated and Deoxygenated Hemoglobin."	• The researchers used a brain scan method called fNIRS to measure blood flow changes (oxygenated and deoxygenated hemoglobin) in 13 healthy people. • They looked at how brain areas activated together over time using a method called Coactivation Pattern (CAP) analysis.	• Only 13 people were tested, so the results might not apply to everyone. • The study only included healthy people, so we don't know how this would work for people with brain disorders.

2.1 Limitations Identified from Literature Survey

The limitations identified in the studies on brain stroke detection reveal several challenges faced by current methodologies and technologies. There are many challenges in using neuroimaging data and machine learning to detect strokes. One major threat is that there are usually more non-stroke images than stroke cases, which can make models biased. The images can also have noise or errors, like movement during a scan, which affects accuracy. Getting brain scans is expensive and time-consuming, so it’s hard to collect large amounts of data [3].

Since strokes can affect different parts of the brain in different ways, it’s tough to build one model that works for everyone. Data privacy rules and the need for patient consent also make it harder to access good datasets. It's important for models to tell the difference between ischemic and hemorrhagic strokes, which isn't always easy. Brain scans are complex and often in 3D, which means strong computers are needed to analyze them. Some systems only use single images and ignore useful time-based information from dynamic scans[4].

Real-time processing is difficult, which can delay help during emergencies. Strokes can happen alongside other problems like tumors, making it harder to spot stroke-specific signs. Different hospitals use different scanning methods, which affects how well a model works everywhere. It’s also tricky to include extra patient information like age or symptoms, even though it’s helpful. Many models don’t explain how they make decisions, so doctors might not trust them. Some models only work well with the data they were trained on and need updates to be used elsewhere. Stroke severity and recovery vary a lot, so predicting outcomes is tough. [5]

Most models are made for adults, so they don’t work well for children. In poorer areas, the equipment needed to use these models might not be available. Adding these systems to real hospital settings is also a challenge. Without long-term data, it's hard to track how a patient recovers. Lastly, the ways we test these models in research don’t always match real-life situations in clinics.[1]

2.2 Research Objectives

- **To Develop a CNN-based diagnostic model:** To design and implement a Convolutional Neural Network (CNN) model for the automated identification and classification of stroke regions in MRI neuroimages, ensuring the system effectively distinguishes between ischemic and hemorrhagic strokes.
- **To Improve stroke detection accuracy:** To enhance the accuracy and precision of stroke detection through the application of deep learning techniques, minimizing misclassification errors and improving the reliability of the model for identifying stroke-affected areas in early-stage MRI scans.
- **To Reduce diagnostic time:** To create an efficient diagnostic tool that significantly reduces the time required for stroke identification, providing healthcare professionals with faster

results, enabling prompt treatment decisions, and improving patient care outcomes.

- **To Evaluate model performance:** To assess the performance of the CNN model by conducting rigorous validation and testing on a diverse MRI dataset, ensuring that the model achieves high sensitivity, specificity, and overall accuracy in stroke detection under various conditions.

Finally, the system

2.3 Product Backlog

S.No	USER STORIES OF BRAIN STROKE IDENTIFICATION
US 1.	As a user, I want to gather publicly available neuroimaging datasets so I can use real-world data to train and test machine learning models.
US 2.	As user, I want to collect and organize relevant datasets from different sources so I can prepare the data for analysis and training.
US 3.	As a user, I want to remove non-brain areas from CT scans (skull stripping) so the model focuses only on important brain structures.
US 4.	As a user, I want to normalize image brightness and intensity so all scans have a consistent format for better model training.
US 5.	As a user, I want to reduce noise in CT images using filters so the scans are clearer and easier for models to analyze.
US 6.	As a user, I want to label brain regions as ischemic, hemorrhagic, or normal so the model can learn from specific stroke-affected areas.
US 7.	As a user, I want to use pre-labeled (segmented) brain images to train CNN models so they can focus on key features in specific regions.
US 8.	As a user, I want to connect the segmentation results with the classification step so that the model uses stroke-affected areas for better predictions.
US 9.	As a user, I want to build and compare SVM, k-NN, and CNN models so I can choose the most accurate one for stroke prediction.
US 10.	As a user, I want to classify brain scans into ischemic, hemorrhagic, and normal types so the system can help doctors make quick decisions.
US 11.	As a user, I want to apply data augmentation and class balancing so the models don’t get biased due to uneven class sizes.
US 12.	As a user, I want to create a clear and scalable stroke prediction system so it can be used easily in real hospitals and clinics.

Table 2.1 Product Backlog

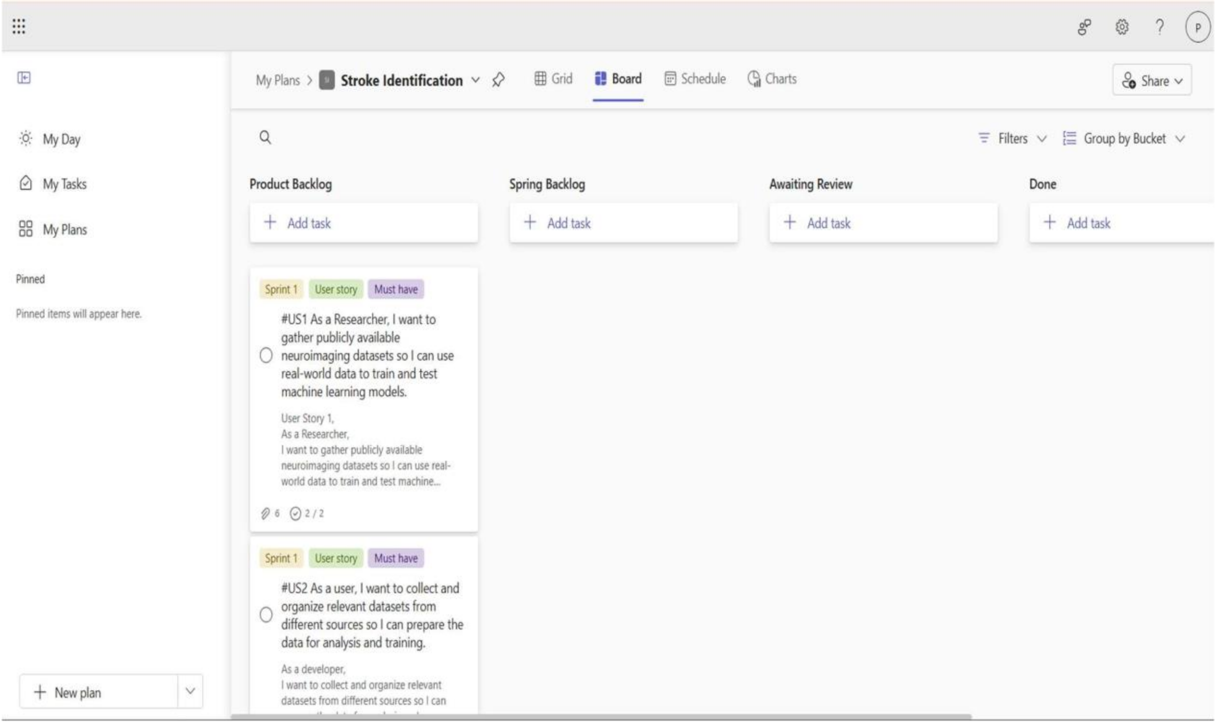


Figure 2.1 MS Planner Board of Brain Stroke Identification

2.4 Plan of Action

The following figure 2.1 depicts the plan of action of the project

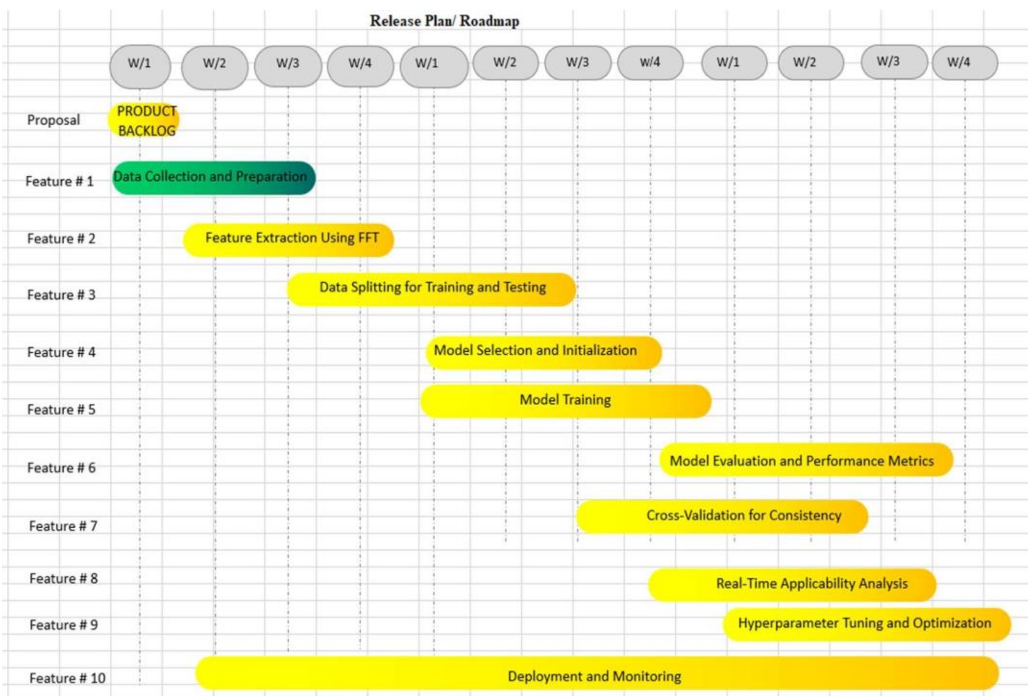


Figure 2.2 Plan of Action

CHAPTER 3
SPRINT PLANNING AND EXECUTION

3.1 SPRINT 1

3.1.1 OBJECTIVES WITH USER STORIES OF SPRINT 1:

The goal of the first sprint was to prepare a refined dataset of brain scans for stroke analysis. This dataset was intended to be used to train models to detect and classify strokes. Free CT and MRI brain scan datasets were sourced from reliable and trusted platforms. These datasets included both stroke and non-stroke cases, covering a variety of stroke types and imaging methods. Once collected, the scans were organized based on scan type, stroke presence, and the quality of accompanying annotations. The data was then prepared for detailed examination. An essential part of this sprint involved preprocessing the scans to make them suitable for model training. Non-brain elements such as the skull and scalp were removed, allowing the model to focus solely on brain tissue. Additionally, the brightness and contrast of the scans were normalized to reduce variability caused by differing imaging techniques. These preprocessing steps were crucial for improving the robustness and generalizability of the model. By the end of the sprint, a well-structured and clean dataset was prepared for the next phases, including feature extraction, model training, and evaluation.

S.NO	Detailed User Stories
US 1	As a user, I want to gather publicly available neuroimaging datasets so I can use real-world data to train and test machine learning models.
US 2	As user, I want to collect and organize relevant datasets from different sources so I can prepare the data for analysis and training.
US 3	As a user, I want to remove non-brain areas from CT scans (skull stripping) so the model focuses only on important brain structures.
US 4	As a user, I want to normalize image brightness and intensity so all scans have a consistent format for better model training.

Table 3.1 Detailed User Stories of sprint 1

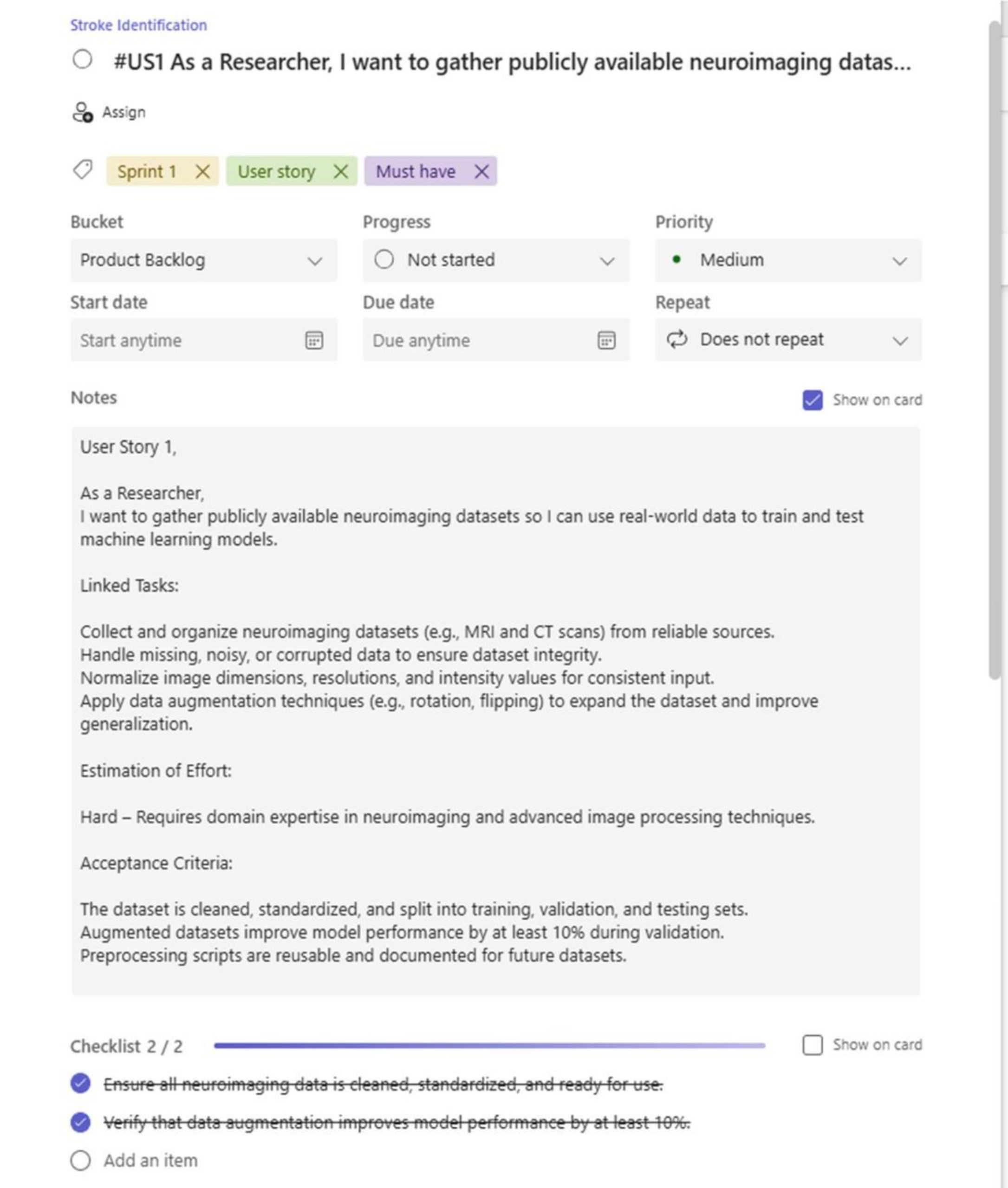


Figure 3.1 user story for collecting the datasets

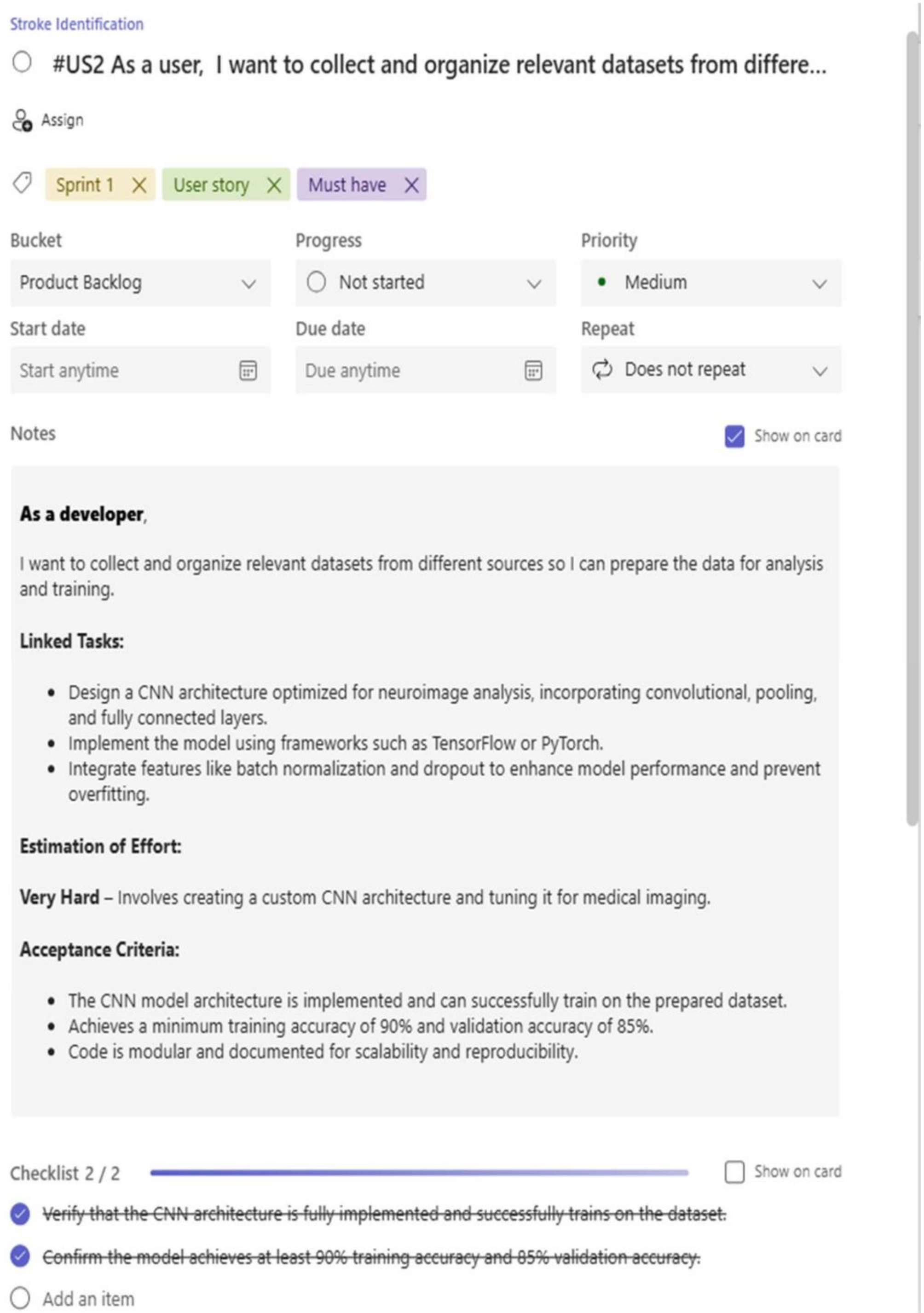


Figure 3.2 User story for preparing data for training

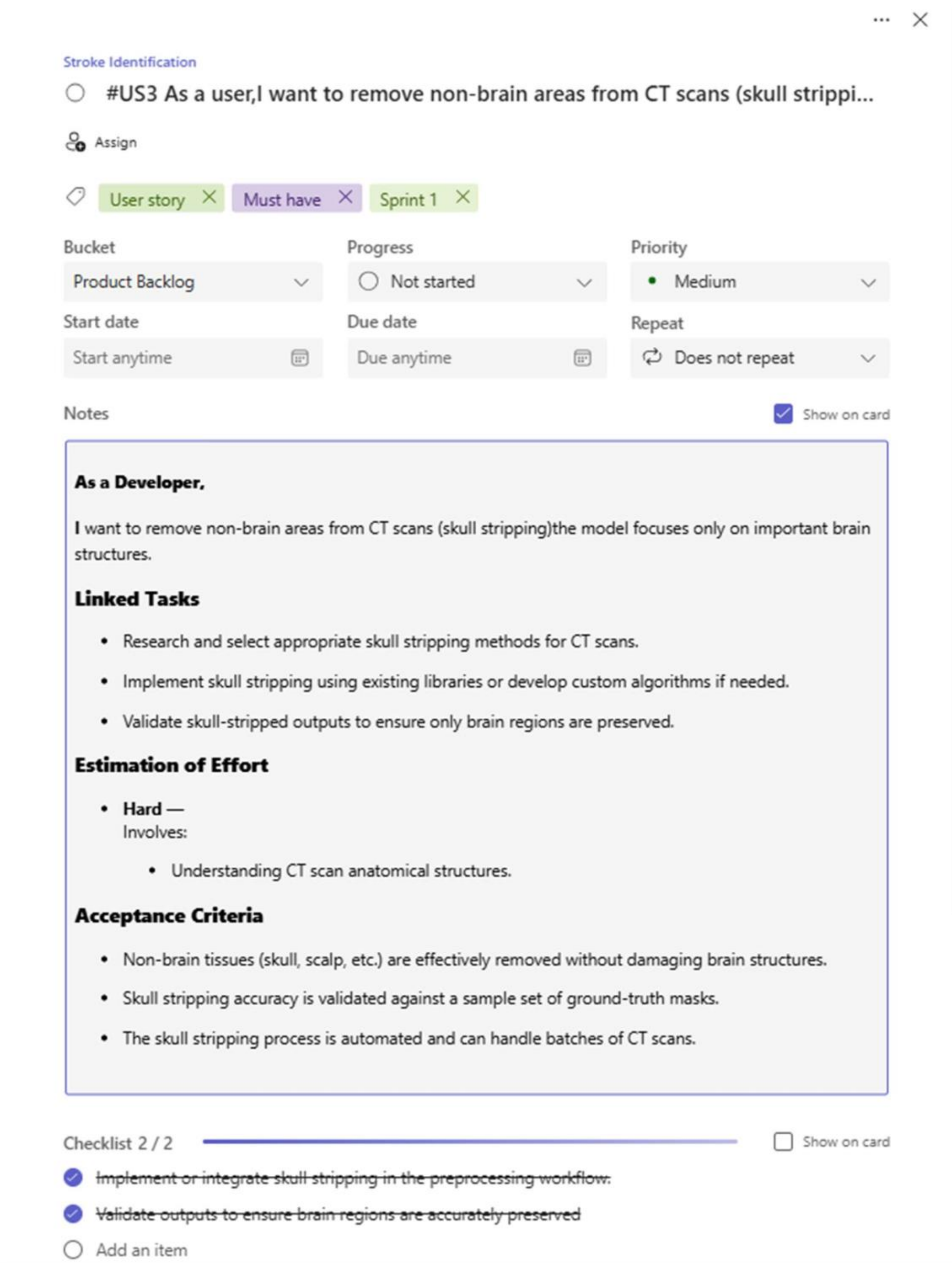


Figure 3.3 user story for CT scans model

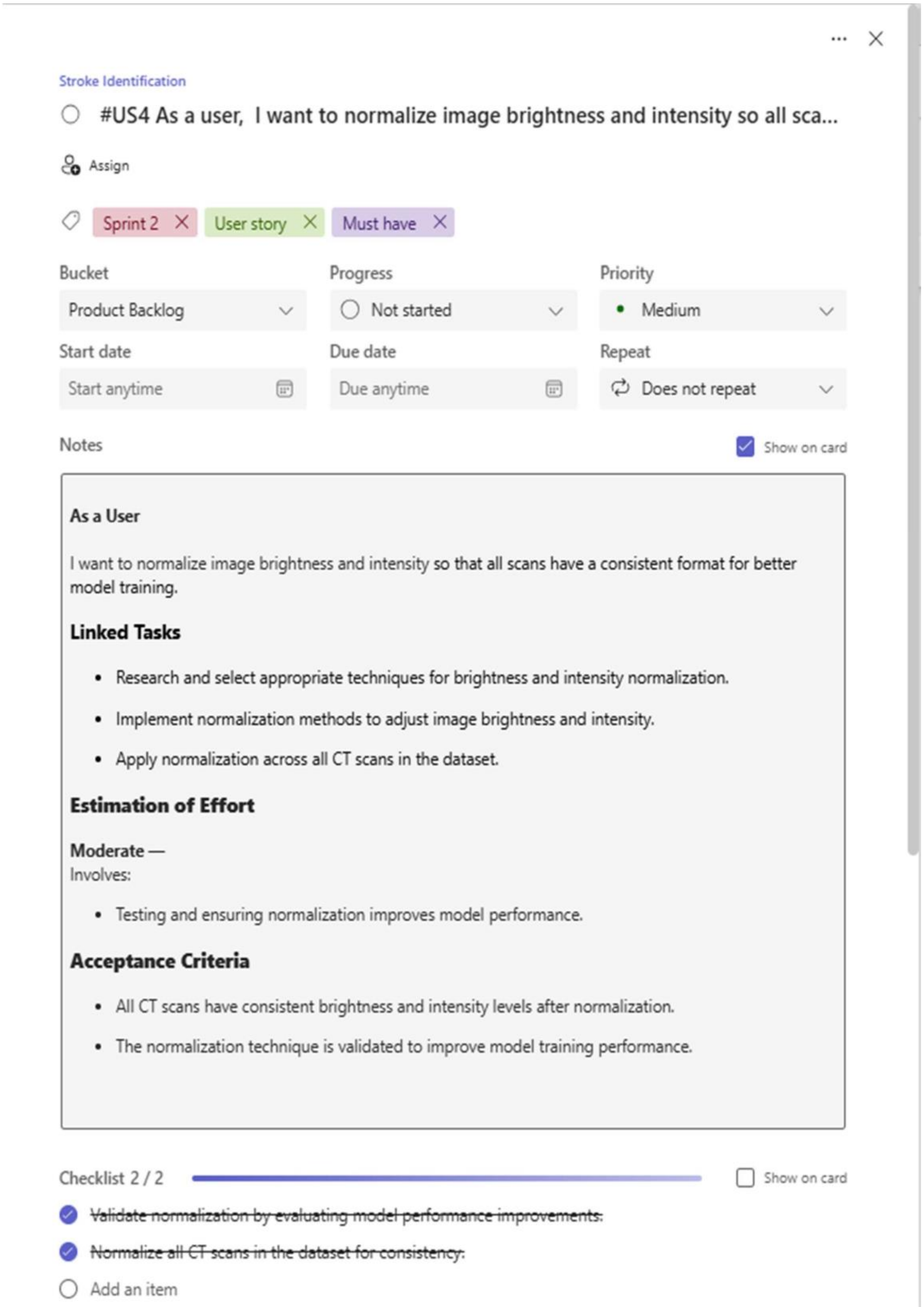


Figure 3.4 user story for normalizing image brightness

3.1.2 FUNCTIONAL DOCUMENT

3.1.2.1 Introduction:

Today, machine learning (ML) can help doctors quickly find brain problems by analyzing CT scans. To do this well, we need real CT images from hospitals and public sources. This project aims to collect these brain scan images, clean them up, and prepare them for ML models, making it easier to build tools that can detect strokes and other brain issues accurately.

3.1.2.2 Product Goal:

The goal is to create a clean and organized set of brain scan images that can be used to train and test machine learning models. The system will collect CT scan data, clean it (remove skulls, adjust brightness, reduce noise), and organize it so that models can learn from it better. This will help create ML tools that can assist doctors faster and more accurately.

3.1.2.3 Who Will Use It:

- **Users:** Data scientists, engineers, doctors, medical researchers, hospitals, and universities.
- **Where:** It will be useful worldwide, especially where medical research is growing, like North America, Europe, Asia, and Africa.

3.1.2.4 How It Will Work:

1. **Collect and Organize Data:**

   Gather brain scan images from trusted websites and organize them neatly.

2. **Preprocessing Images:**
   - **Remove Skull (Skull Stripping):** Keep only the brain part in the image.
   - **Normalize Images:** Make brightness and contrast similar for all images.
   - **Reduce Noise:** Clean up images using filters to remove random specks.

3. **Prepare for Machine Learning:**

   Label the images properly and split them into training and testing groups.

4. **Feedback and Improvement:**

   Allow users to suggest improvements to make the system better over time.

3.1.2.5 Features:

1. **Collect and Organize Datasets:**

   Users can easily find and use brain scan images.

   - **User Story***:* As a user, I want to collect and organize datasets for training.

2. **Skull Stripping:**

   Automatically remove parts outside the brain.

   - **User Story:** As a preprocessing engineer, I want to remove non-brain areas from CT scans.

3. **Normalize Brightness and Intensity:**

   Make all images look similar.

   - **User Story***:* As a machine learning engineer, I want to normalize image brightness and intensity.

4. **Reduce Noise:**

   Make images cleaner and clearer.

   - **User Story:** As a medical imaging specialist, I want to reduce noise in CT images.

3.1.2.6 Authorization Matrix :

Role	See Data	Edit Images	Upload Data	Give Feedback	Train Models
Data Scientist	Yes	No	No	Yes	Yes
Preprocessing Engineer	Yes	Yes	No	Yes	Yes
ML Engineer	Yes	No	No	Yes	Yes
ML Specialist	Yes	No	No	Yes	No
Administrator	Yes	Yes	Yes	Yes	Yes

Table 3.2 Authorization Matrix

3.1.2.7 Assumptions:

1. **Availability of Publicly Accessible Datasets:**

   Access to diverse and high-quality neuroimaging datasets is assumed.

2. **Data Privacy and Compliance:**

   All datasets used are de-identified and comply with medical data privacy standards (e.g., HIPAA, GDPR).

3. **Standardization across Sources:**

   While sources may differ, a common preprocessing standard will be applied to create uniformity across all datasets.

4. **User Engagement:**

   Researchers, engineers, and specialists are willing to engage by providing feedback and reporting errors in the datasets.

5. **Infrastructure Support:**

   There is access to sufficient computational resources for preprocessing and training ML models.

3.1.3 ARCHITECTURE DOCUMENT

3.1.3.1 Application

1. **Microservices:**

- **Architecture:**
  The system will be divided into small independent microservices. Each microservice will handle a specific task in the stroke classification process.
- **Components:**
  - **Data Ingestion Service:**
    Reads CT scan images from DICOM files and other medical imaging formats.
  - **Preprocessing Service:**
    Handles data augmentation, skull stripping, image normalization, and noise reduction to prepare the CT scans.
  - **Feature Detection Service:**
    Extracts important features from the preprocessed images using techniques like SIFT, SURF, or CNN-based feature extractors.
  - **Model Training Service:**
    Trains machine learning models like CNN, SVM, or Random Forest for classifying stroke types.
  - **Classification Service:**
    Applies trained models to classify new CT scans into different stroke categories (e.g., hemorrhagic, ischemic).
  - **Evaluation Service:**
    Calculates model metrics (accuracy, precision, recall) and generates evaluation reports.
  - **Explainability Service (XAI):**
    Visualizes model predictions using tools like Grad-CAM to make decisions understandable to users.
  - **User Interface Service:**
    Web application where users (doctors, researchers) can upload CT scans and view classification results.
  - **API Gateway:**
    Manages all requests from users and routes them to the correct microservice.

3.1.3.2 System Architecture :

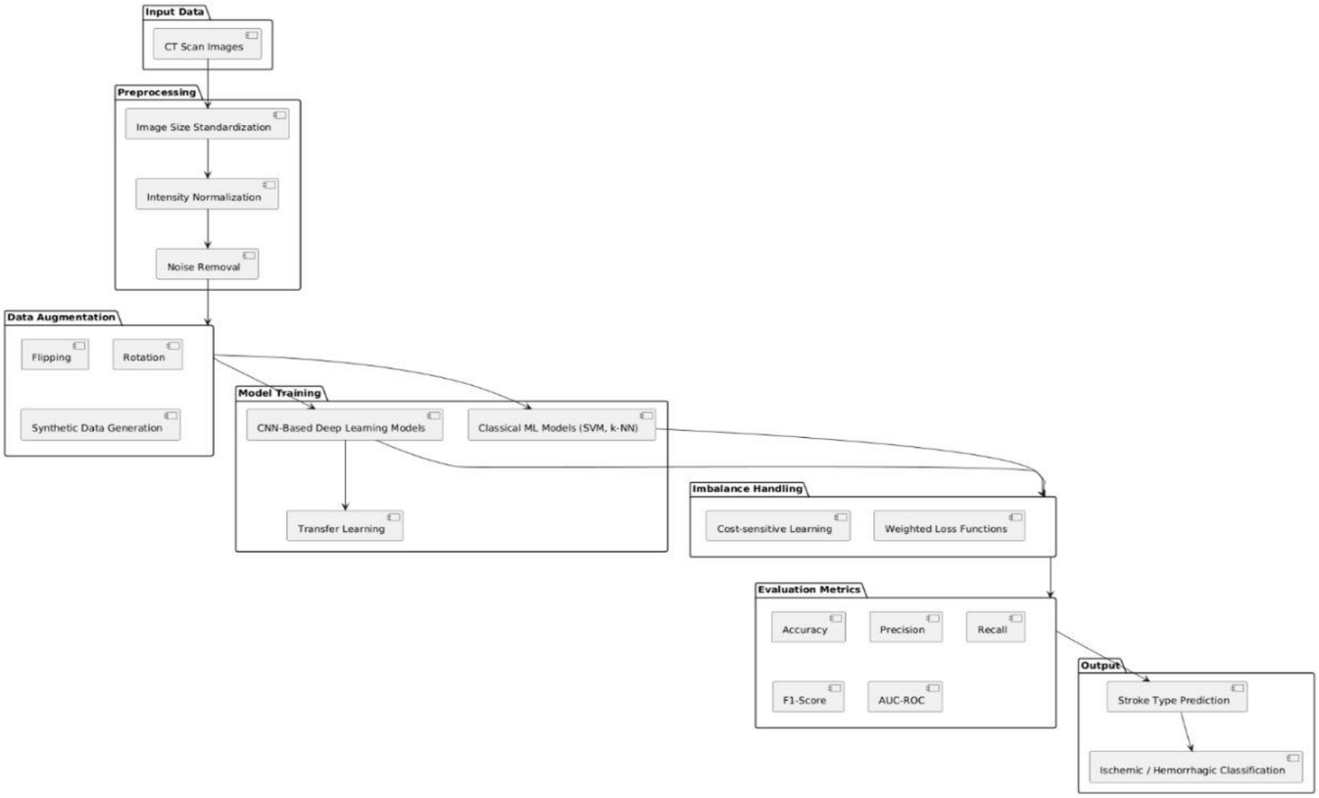


Figure 3.5 System Architecture

3.1.3.3 Data Exchange Contract:

1. Frequency of Data Exchanges:

- **Real-Time:** CT scan ingestion and immediate preprocessing and classification.
- **Batch:** Periodic updates for training datasets (new CT scans collected weekly or monthly).
- **On-Demand:** Manual retraining of models when the user requests or new labeled datasets are added.

2. Data Sets:

- **Training Data:** Pre-labeled CT scans showing types of strokes or healthy brains.
- **Preprocessing Outputs:** Skull-stripped, normalized, and noise-reduced versions of the scans.
- **Feature Sets:** Extracted features ready for training or classification.
- **Evaluation Reports:** Performance metrics and error analysis after model training.

3. Mode of Exchanges:

- **API:** REST APIs to upload scans, fetch classification results, and request reports.
- **File Uploads:** DICOM or NIfTI file uploads for adding new CT scans.
- **Queue:** Message queues (e.g., RabbitMQ, Kafka) to manage scan uploads, preprocessing tasks, and classification events.
- **Database:** Storage of preprocessed data, extracted features, model versions, and user interaction logs.

3.14 OUTCOME OF OBJECTIVE

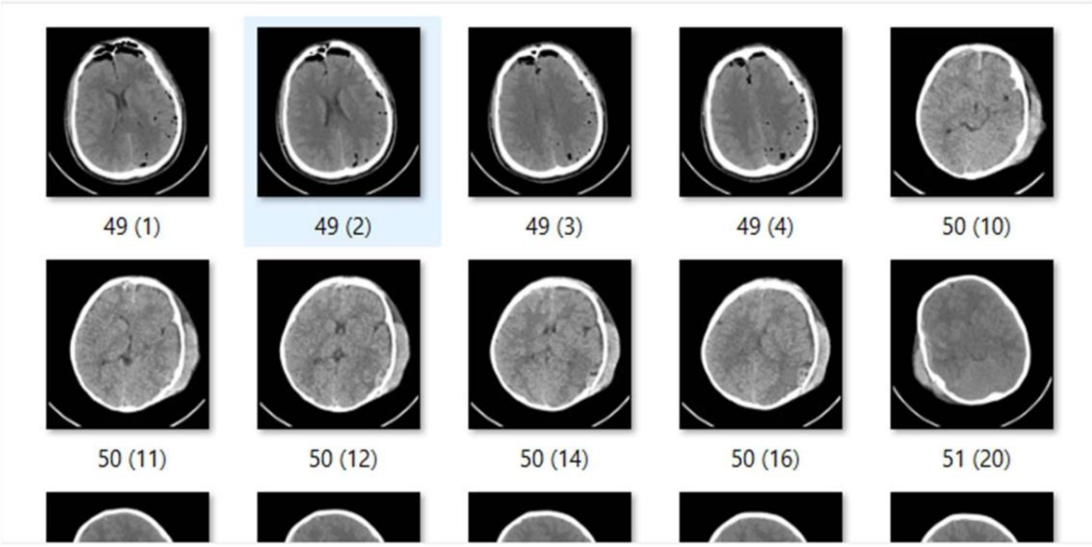


Fig 3.6 Brain MRI Scan Dataset

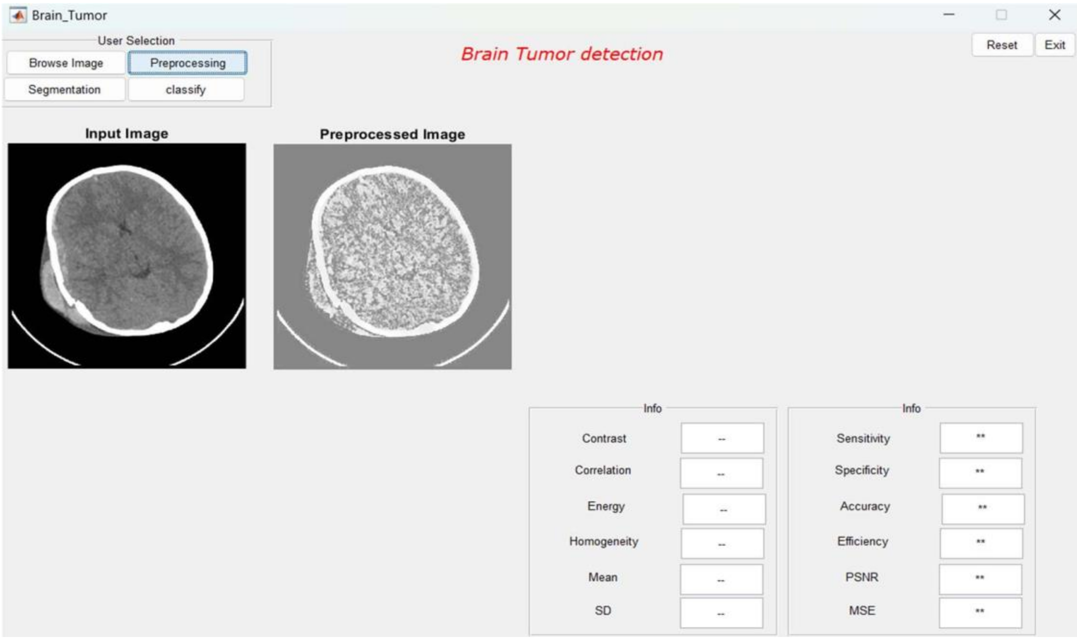


Fig 3.7 Brain Scan Input and Pre-processed Image

The result of these user stories was the creation of an organized, high-standard, and systematic neuroimaging dataset that could be used for machine learning model development. By compiling CT and MRI scans already available in the public domain from a variety of reputable sources, the project ensured that the model was developed on a dataset providing real-world, clinically relevant data for stroke detection and classification. The datasets were organized by modality, diagnosis, and metadata, serving as the first step toward effective data management and analysis. The process of skull stripping was used to remove all non-brain structures from the scans, which allowed the models to focus on learning from the most useful and relevant brain regions, reduced noise, and enabled more efficient learning. Normalization of image brightness and intensity was performed to standardize the scans and minimize variations caused by different imaging sources or acquisition settings. All of these preprocessing steps together resulted in a clean, standardized dataset that significantly improved the training and testing quality of the machine learning models, contributing to the development of a robust and accurate system for stroke diagnosis.

3.14.1 SPRINT RETROSPECTIVE

	Sprint Retrospective			
Liked	**Learned**	**Lacked**	**Longed For**	
Share aspects of the sprint that you enjoyed or found particularly effective.	*Discuss lessons learned, whether they are related to processes, technical aspects, or teamwork.*	*Identify areas where the team felt a lack of resources, support, or information.*	*Discuss any desires or expectations that the team had but were not met during the sprint.*	*Guidelines*
Discovery of reliable neuroimaging data repositories helped accelerate dataset gathering.	Validating metadata and licensing is crucial for dataset usability.	Some datasets lacked consistent metadata and standard structures	Easier access to high-quality, standardized neuroimaging datasets.	Validate licensing, completeness, and metadata standards before dataset acceptance.
Successful skull stripping methods significantly improved focus on brain regions in CT scans.	Importance of validating skull-stripped outputs against ground-truth masks.	Initial skull stripping methods occasionally removed brain tissue.	A more robust, automated skull stripping pipeline.	Implement batch validation checks after skull stripping.

Figure 3.8 sprint retrospective

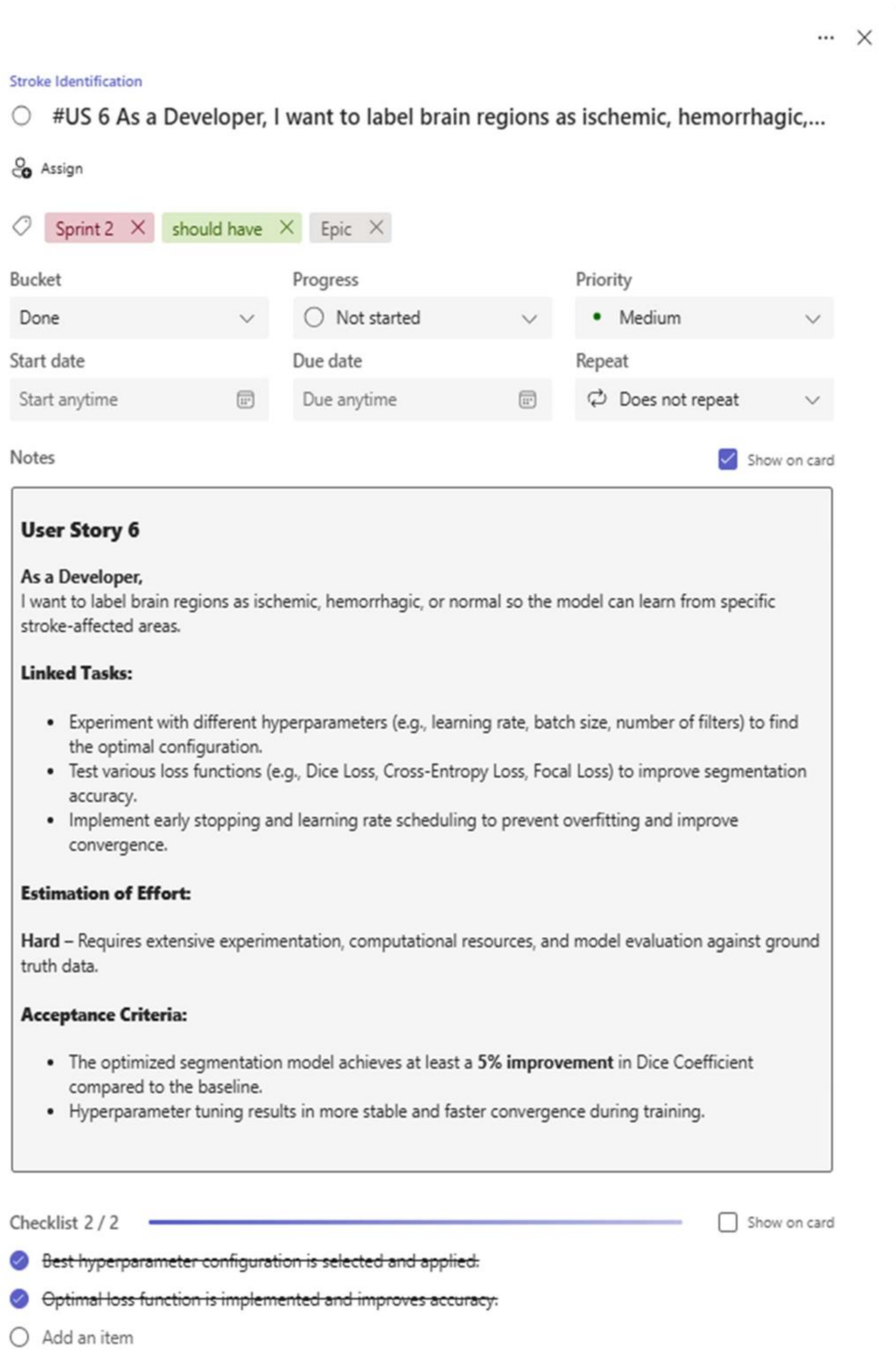
3.2 SPRINT 2

3.2.1 OBJECTIVES WITH USER STORIES OF SPRINT 2

The goal of the second sprint was to improve the accuracy and clinical value of stroke classification models by labeling and segmenting specific brain regions. Brain scans were labeled as ischemic, hemorrhagic, or normal so that the model could learn from the areas actually affected by a stroke, rather than using the entire brain image. This helped the model tell the difference between stroke types based on the location and structure of the damaged tissue. Pre-labeled or semi-labeled brain images made it easier for the model to focus on important areas and ignore noise, which sped up the learning process. The segmentation results were then passed directly into the classification stage, allowing the model to make better predictions based on the severity and location of damage. This method led to more accurate and understandable results and helped create a smarter model that works more like a doctor in real-life stroke diagnosis.

S.No	**Detailed User Stories**
US 6	As a user, I want to label brain regions as ischemic, hemorrhagic, or normal so the model can learn from specific stroke-affected areas.
US 7	As a user, I want to use pre-labeled (segmented) brain images to train CNN models so they can focus on key features in specific regions.
US 8	As a user, I want to connect the segmentation results with the classification step so that the model uses stroke-affected areas for better predictions.

Table 3.3 Detailed User Stories of sprint 2


Stroke Identification

#US 6 As a Developer, I want to label brain regions as ischemic, hemorrhagic,...

Assign

Sprint 2 | should have | Epic

Bucket	Progress	Priority
Done	Not started	Medium
Start date	**Due date**	**Repeat**
Start anytime	Due anytime	Does not repeat

Notes

Show on card

User Story 6

As a Developer,
I want to label brain regions as ischemic, hemorrhagic, or normal so the model can learn from specific stroke-affected areas.

Linked Tasks:

- Experiment with different hyperparameters (e.g., learning rate, batch size, number of filters) to find the optimal configuration.
- Test various loss functions (e.g., Dice Loss, Cross-Entropy Loss, Focal Loss) to improve segmentation accuracy.
- Implement early stopping and learning rate scheduling to prevent overfitting and improve convergence.

Estimation of Effort:

Hard – Requires extensive experimentation, computational resources, and model evaluation against ground truth data.

Acceptance Criteria:

- The optimized segmentation model achieves at least a **5% improvement** in Dice Coefficient compared to the baseline.
- Hyperparameter tuning results in more stable and faster convergence during training.

Checklist 2 / 2

Show on card

~~Best hyperparameter configuration is selected and applied.~~

~~Optimal loss function is implemented and improves accuracy.~~

Add an item

Figure 3.9 User story for Identifying Brain regions

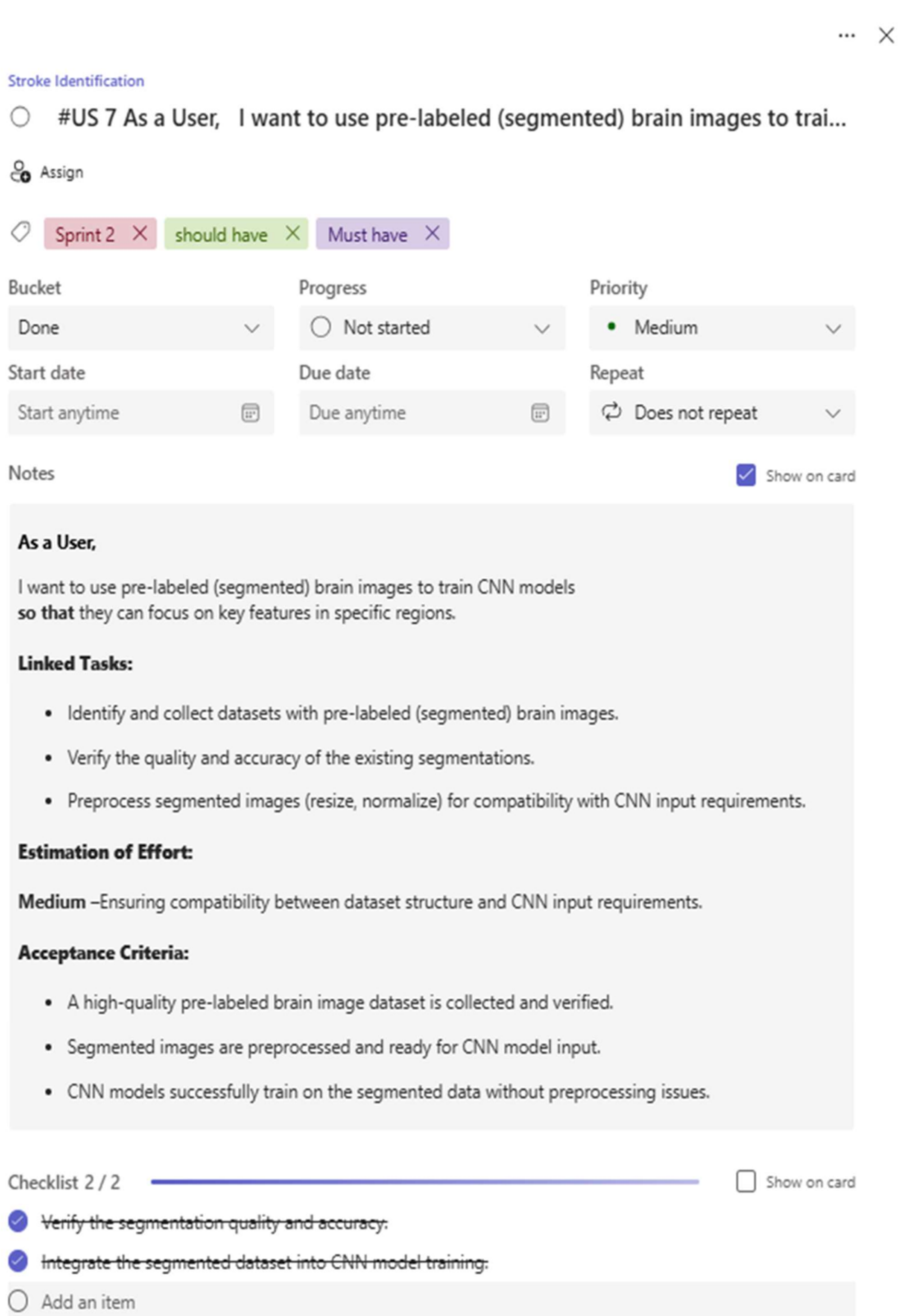


Figure 3.10 user story for pre-labeled images to train CNN models

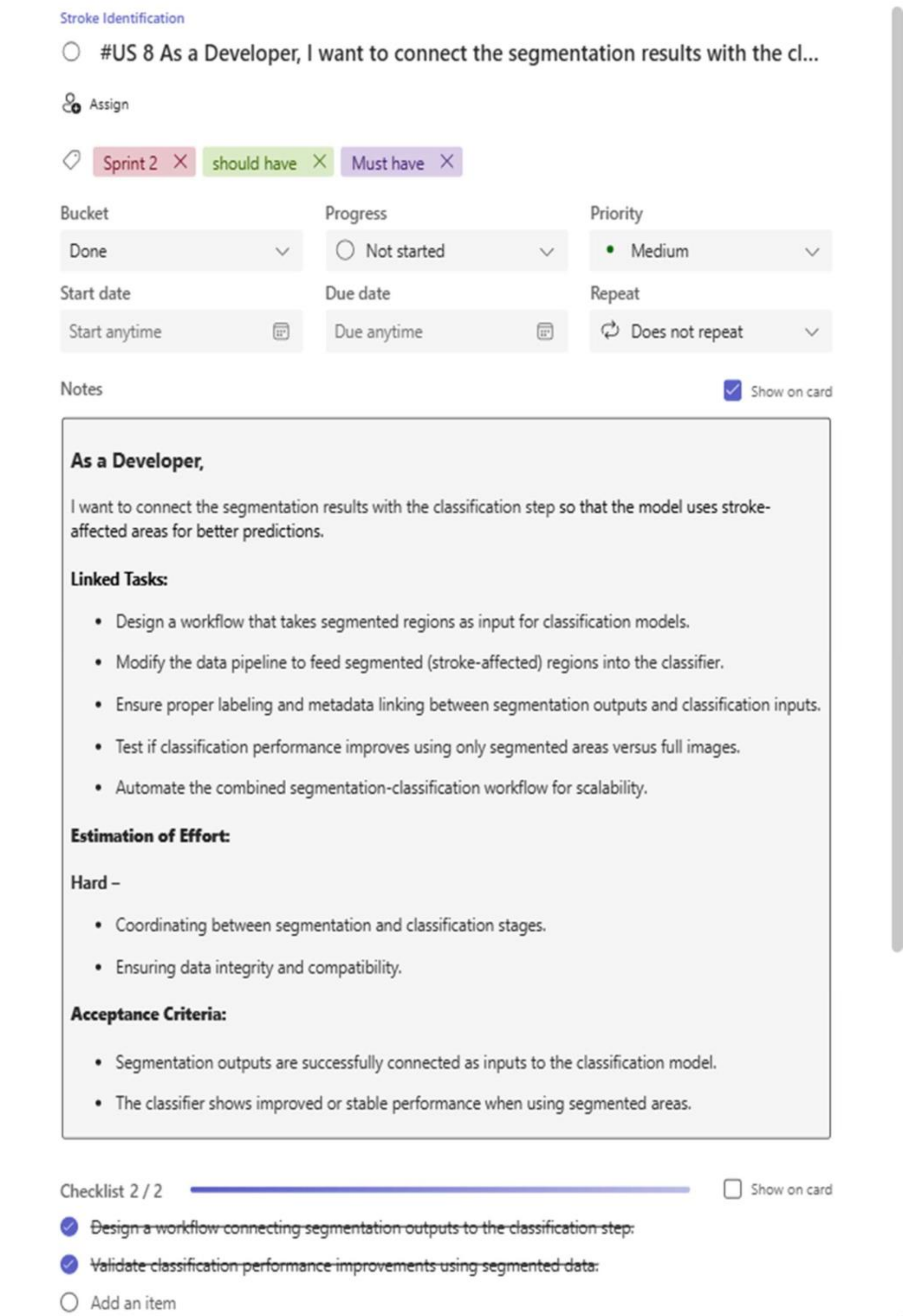


Figure 3.11 user story for segmentation results

3.2.2 FUNCTIONAL DOCUMENT

3.2.2.1 Introduction:

Stroke remains a major cause of disability and death worldwide. Early detection and precise classification of strokes, whether ischemic or hemorrhagic, are critical for timely treatment. This project focuses on developing a machine learning-based system that utilizes neuroimaging data to detect and classify strokes. By integrating segmentation and classification models, the system aims to improve prediction accuracy and assist medical professionals in diagnosis.

3.2.2.2 Product Goal:

The goal is to create a smart system that can first find stroke-affected areas in brain images by using segmentation. Then, it will classify the type of stroke, either ischemic or hemorrhagic, based on the labeled images. This system will help doctors make faster and more accurate decisions by giving them reliable, automatic stroke detection support.

3.2.2.3 Demography

Users:

- Medical Researchers.
- Radiologists and Neurologists.
- Data Scientists in Healthcare.
- AI/ML Engineers specializing in Medical Imaging.

Location:

- Hospitals and Research Institutes.
- Stroke Centers.
- Universities focusing on biomedical research.
- AI research laboratories across North America, Europe, and Asia.

3.2.2.4 Business Processes:

1. Algorithm Evaluation and Comparison:

- Neuroimaging data (MRI, CT scans) are collected and stored in a secure database.

2. Performance Report Generation:

- A Researchers manually label regions of the brain as ischemic, hemorrhagic, or normal.
- Reports can be customized based on specific metrics or datasets.

3. Research Guidance and Insights:

- Segmented datasets are used by data scientists to train deep learning models (e.g., CNNs).

4. Collaboration and Feedback:

- The classification model predicts stroke type based on segmented regions.

5. Data Management and Security:

- Medical professionals validate model outputs before integrating them into clinical workflows.

3.2.2.5 Features

1. Description

The system tags brain areas on images in three parts: stroke with blocked blood flow, stroke with bleeding, and healthy tissue. These tagged images train a CNN to pick up important details for clinics. The test model uses data to guess stroke type better.

User Stories:

- I want to label brain regions as ischemic, hemorrhagic, or normal so the model can learn from specific stroke-affected areas.
- I want to use pre-labeled (segmented) brain images to train CNN models so they can focus on key features in specific regions.
- I want to connect the segmentation results with the classification step so that the model uses stroke-affected areas for better predictions.

3.2.2.6 Authorization Matrix:

Role	View Images	Segment Images	Train Models	Access Predictions	Admin Rights
Segmentation Researcher	Yes	Yes	No	No	No
Data Scientist	Yes	No	Yes	Yes	No
System Architect	Yes	Yes	Yes	Yes	Yes
Medical Reviewer	Yes	No	No	Yes	No

Table 3.4 Authorization Matrix

3.2.2.7 Assumptions

- High-quality, annotated neuroimages are available for segmentation and classification.
- Collaboration among segmentation experts, data scientists, and system architects is active throughout development.
- Hardware infrastructure (GPUs, servers) supports high-volume training and inference tasks.
- Regulatory compliance (HIPAA, GDPR) is maintained for handling sensitive patient data.
- Users have basic training in using the system interfaces and understanding output results.

3.2.2 ARCHITECTURE DOCUMENT

3.2.3.1 Application :

In this application, patients are scanned with MRI and the images saved with the medical data of patients. The medical professionals who perform and report on the MRI scans, for example radiologists, provide the scan reports and analysis. The segmentation researchers will take the MRI scans and use those images to label/refer specific parts of the brain as ischaemic, haemorrhagic, or normal. The images that have labelled areas will be provided to data scientists so they can train a convolutional neural network (CNN) to classify the brain regions that were impacted during the stroke, hoping that they will learn to focus on features relevant to the stroke impacted area. Finally the system engineers will link the segmentation outputs to the classification models assuming the predictions will be made on rightly segmented brain parts. The modelling whichever it looks like should provide a primary betterment to the overall stroke detection and classification.

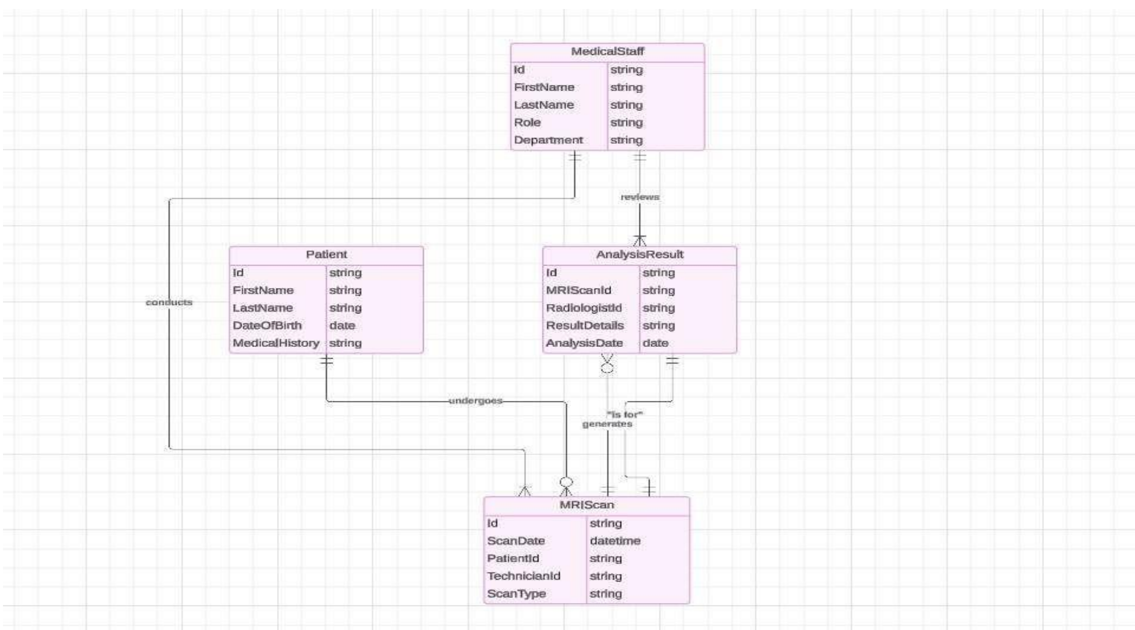


Figure 3.12 Entity Relationship Diagram for patient to medical staff

In this application, after a patient has an MRI scan, the images are saved in the system. Segmentation researchers open these images and label the brain areas as ischemic, hemorrhagic, or normal. They use tools to mark the parts of the brain that show stroke signs. This labeling helps in creating a dataset that shows exactly where the problem areas are. Later, data scientists use these labeled images to train CNN models, which learn to detect strokes by focusing on these important parts. The system architect then links the segmentation output with the classification model, making sure the predictions are based on the labeled regions, helping doctors get faster and more accurate results.

3.2.3.3. Training the Model :

To create the dataset for stroke classification and detection, first, a variety of MRI brain scans are collected from patients. Segmentation researchers label the medical terminology to brain regions on each scan as ischemic, hemorrhagic, or normal. Next, the label heads are pre-processed through resizing, normalizing, and further augmenting to increase the dataset size. Now, the pre-processed images of the data are ready to be built through a Convolutional Neural Network (CNN) by simply taking the input data and allowing it to learn the patterns/features surrounding a stroke. The classification model receives the data as a training set, then the images are further divided into three sections - training set/validation dataset/testing dataset, which will help with learning, parameter tuning, and eventually model evaluation for accuracy. Finally, our model will classify an new, empty unseen MRI image to the correct categorized stroke help.

OUTCOME OF OBJECTIVE :

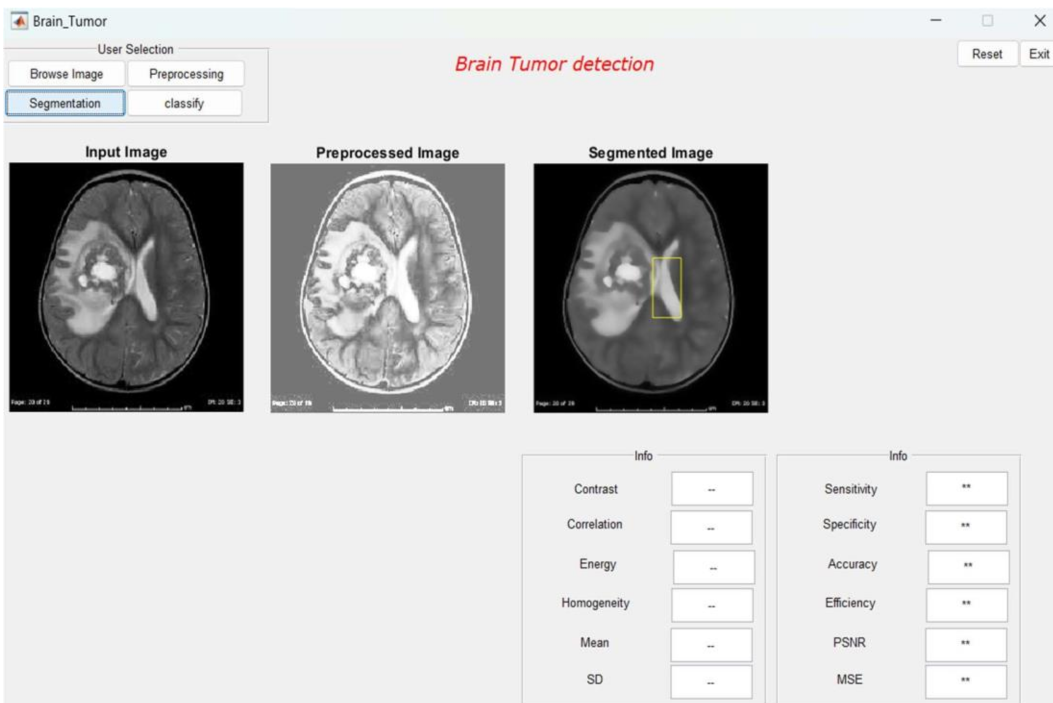


Figure 3.13 Outcome of Segmented Image

The anticipated benefit of these user stories was seen in the development of a more precise and efficient stroke detection and classification system, achieved through the use of region-specific information from neuroimages. Brain areas were labeled as ischemic, hemorrhagic, or normal, allowing stroke manifestations to be presented in a more fine-grained manner, thereby increasing the model's sensitivity to subtle differences. The CNN models were trained on brain images that had been pre-segmented into the defined categories, ensuring that only the most relevant features were used for classification. As a result, the computational load was reduced, and predictive outcomes were made more reliable. Additionally, the segmentation output was integrated into the classification pipeline, so that classification was carried out using clinically relevant brain regions, while irrelevant areas were excluded. In the end, a more accurate and generalizable stroke classification model was produced, designed to function as a dependable decision support system and automated diagnostic tool for clinical use.

3.2.4 SPRINT RETROSPECTIVE

	Sprint Retrospective			
Liked	**Learned**	**Lacked**	**Longed For**	
Share aspects of the sprint that you enjoyed or found particularly effective.	*Discuss lessons learned, whether they are related to processes, technical aspects, or teamwork.*	*Identify areas where the team felt a lack of resources, support, or information.*	*Discuss any desires or expectations that the team had but were not met during the sprint.*	*Guidelines*
Labeling brain regions (ischemic, hemorrhagic, normal) improved model learning on stroke-affected areas.	Hyperparameter tuning and selecting the right loss function significantly impact segmentation performance.	Early experiments lacked stable convergence and required extensive tuning.	More automated hyperparameter optimization frameworks.	Combine early stopping, learning rate scheduling, and customized loss functions for best segmentation
Pre-labeled segmented brain datasets made CNN training more focused and efficient.	Preprocessing (resizing, normalization) ensures smoother CNN model integration.	Quality inconsistencies across pre-labeled datasets complicated preprocessing.	More consistent, high-quality, standardized segmented datasets.	Always verify label accuracy and preprocess images before CNN training.

Figure 3.14 sprint retrospective

3.3 SPRINT 3

3.3.1 OBJECTIVES WITH USER STORIES OF SPRINT 3

The goal of the third sprint was to develop a stroke prediction application that would be valid, reliable, and trustworthy for physicians. A variety of machine learning (ML) models were planned to be employed and tested. Deep learning models, such as Convolutional Neural Networks (CNNs), and traditional models like Support Vector Machines (SVMs), were intended to be used for classifying brain scans as ischemic, hemorrhagic, or normal. The information provided by these ML models was expected to support smart and safe decision-making in clinical care. As with any ML application, the validity and robustness of the trained models were prioritized for thorough testing. Data augmentation and oversampling techniques were to be applied to address potential biases in prediction, especially those caused by class imbalances. Additionally, the system was designed with the intention of being understandable, scalable, and easy to implement, making it suitable for real-world clinical environments where usability, trust, and flexibility are essential. Ultimately, a consistent and high-performing cognitive aid system was expected to be produced—one capable of reliably classifying strokes in a timely manner to support rapid diagnosis in hospitals and clinics.

S.No	**Detailed User Stories**
US 9	As a user, I want to build and compare SVM, and CNN models so I can choose the most accurate one for stroke prediction.
US 10	As a user, I want to classify brain scans into ischemic, hemorrhagic, and normal types so the system can help doctors make quick decisions.
US 11	As a user, I want to apply data augmentation and class balancing so the models don't get biased due to uneven class sizes.
US 12	As a user, I want to create a clear and scalable stroke prediction system so it can be used easily in real hospitals and clinics.

Table 3.5 Detailed User Stories of sprint 3

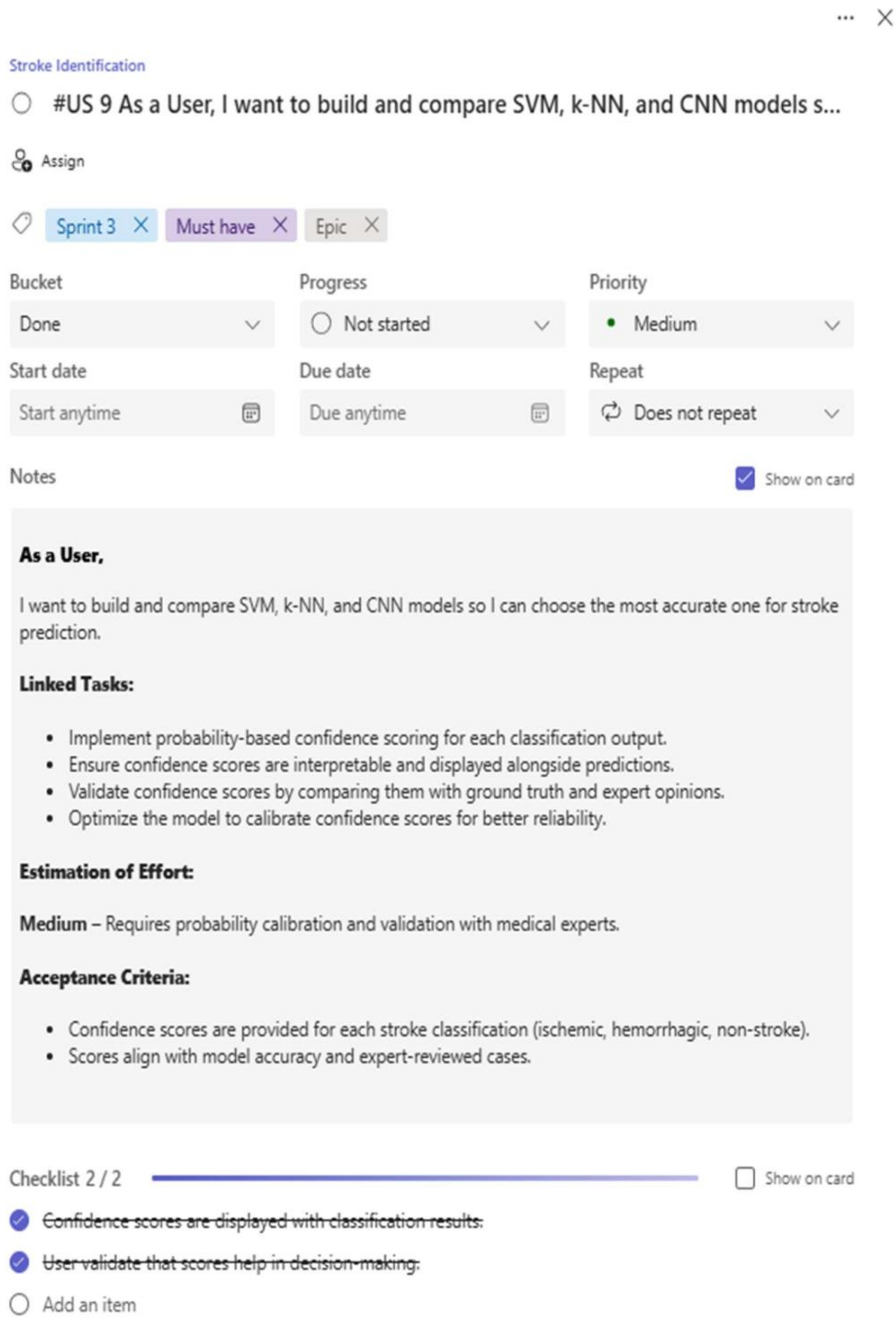


FIGURE 3.15 User story for build and compare of ML models

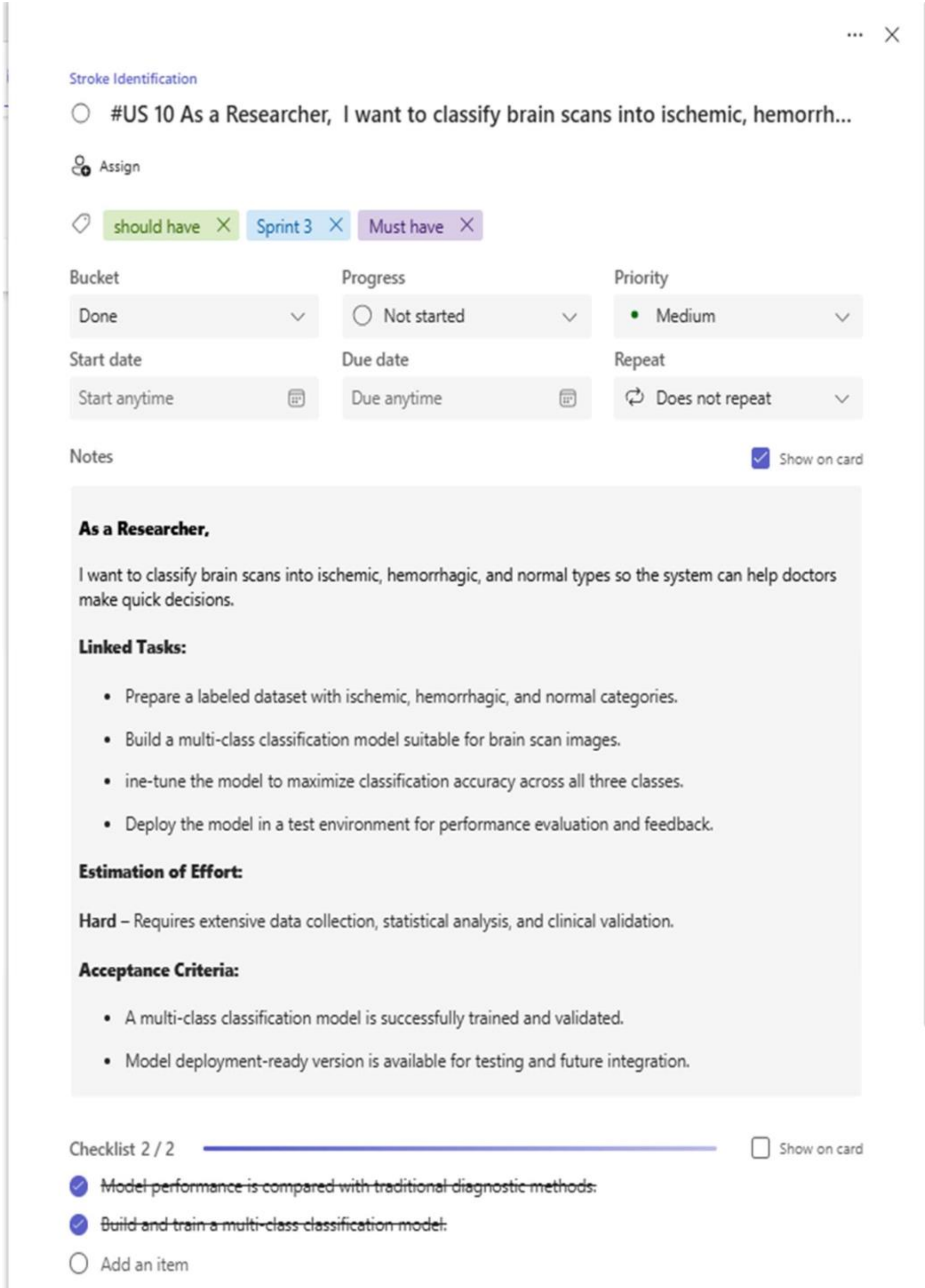


FIGURE 3.16 User story for classify brain scans

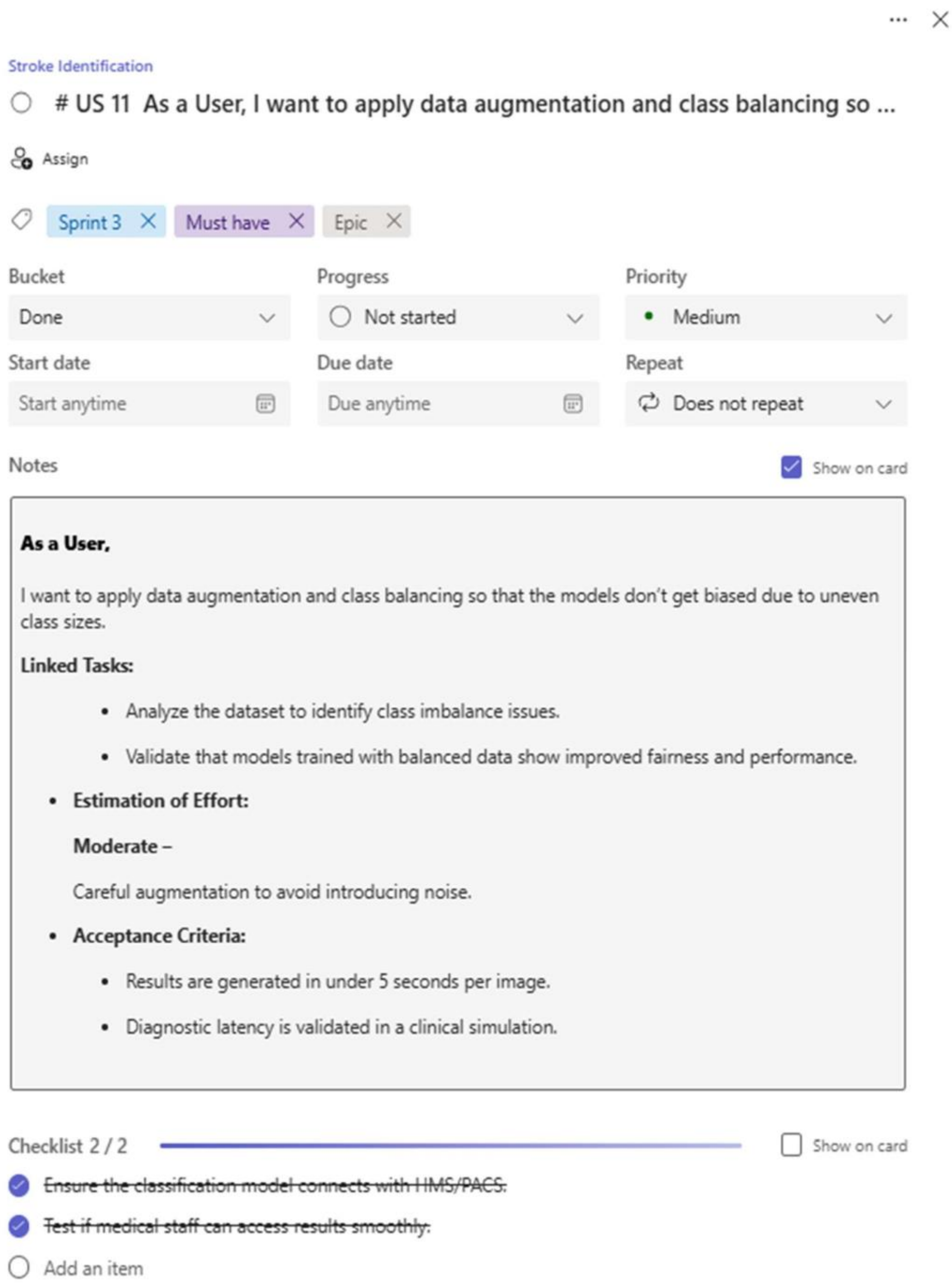


Fig 3.17 User Story for data augmentation and balancing

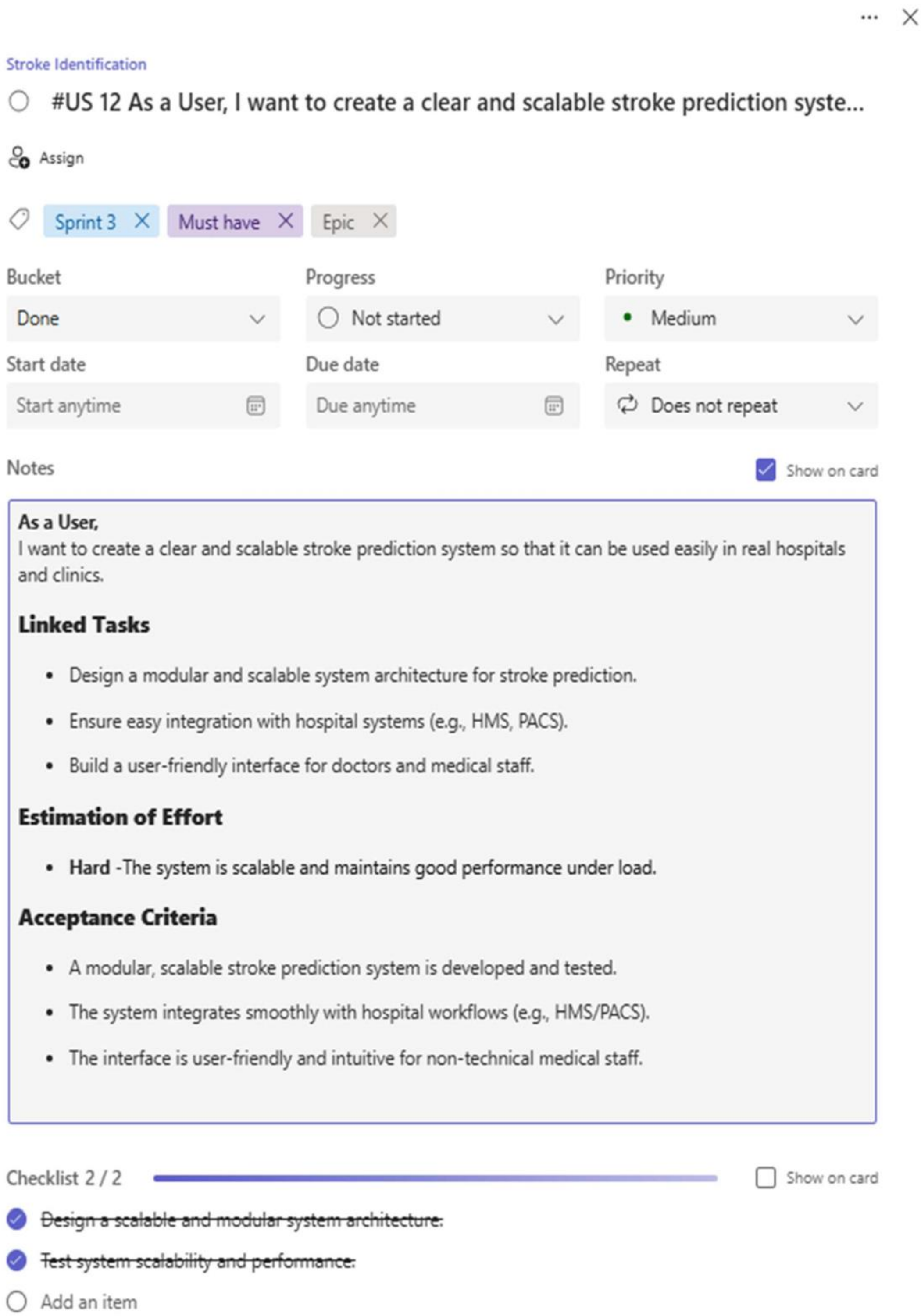


Fig 3.18 User Story for clear and stroke prediction system

3.3.2 FUNCTIONAL DOCUMENT

3.3.2.1. Introduction

This document outlines the functional requirements for the stroke prediction system, which leverages machine learning to classify brain scans and predict stroke types. The system aims to support healthcare professionals in making quick, accurate decisions.

3.3.2.2. Product Goal

To create an accurate, efficient, and scalable stroke prediction system using machine learning models, enabling doctors to classify brain scans and make informed decisions quickly.

3.3.2.3. Demography

- **Users:** Model developers, medical AI researchers, data scientists, healthcare software integrators, and doctors.
- **Location:** Primarily used in hospitals and healthcare clinics for stroke diagnosis.

3.3.2.4. Business Processes

- The system will take neuroimaging data (brain scans) as input.
- It will apply machine learning models to classify the scans into ischemic, hemorrhagic, or normal categories.
- Data augmentation and class balancing will ensure accurate model performance even with uneven class sizes.
- The results will be displayed in an easy-to-interpret format for doctors, aiding quick decision-making.

3.3.2.5. Features

3.3.2.5.1 Feature 1: Stroke Prediction Model

Description: The system will use various machine learning models like SVM, k-NN, and CNN to classify brain scans into ischemic, hemorrhagic, or normal categories. The model will be trained using augmented data and balanced class distributions.

User Story:

- As a model developer, I want to build and compare SVM, k-NN, and CNN models to identify the most accurate stroke prediction model.
- As a medical AI researcher, I want the system to classify brain scans into ischemic, hemorrhagic, and normal types to help doctors make accurate and fast decisions.
- As a data scientist, I want to apply data augmentation and class balancing techniques to prevent model bias due to uneven class sizes.
- As a healthcare software integrator, I want the stroke prediction system to be clear, scalable, and easy to use in real hospitals and clinics.

3.3.2.6 Authorization Matrix :

Role	View Images	Segment Images	Train Classification Models	Access Classification Predictions	Admin Rights
Model Developer	Yes	No	Yes	No	No
Medical AI Researcher	Yes	No	No	Yes	No
Data Scientist	Yes	No	Yes	Yes	No
Healthcare Software Integrator	Yes	No	No	Yes	Yes

Table 3.6 Authorization Matrix

3.3.2.7 Assumptions:

- The system will have access to labeled medical datasets of brain scans.
- Sufficient computational resources will be available for training the machine learning models.
- Data privacy and security protocols will be implemented according to healthcare regulations.
- The system will be integrated with existing hospital software and hardware infrastructure.

3.3.3 ARCHITECTURE DOCUMENT

This flowchart outlines the process of diagnosing strokes using machine learning and neuroimaging data. This system provides a structured approach to developing a diagnostic model for stroke detection by emphasizing data quality enhancement, iterative refinement, and accuracy validation.

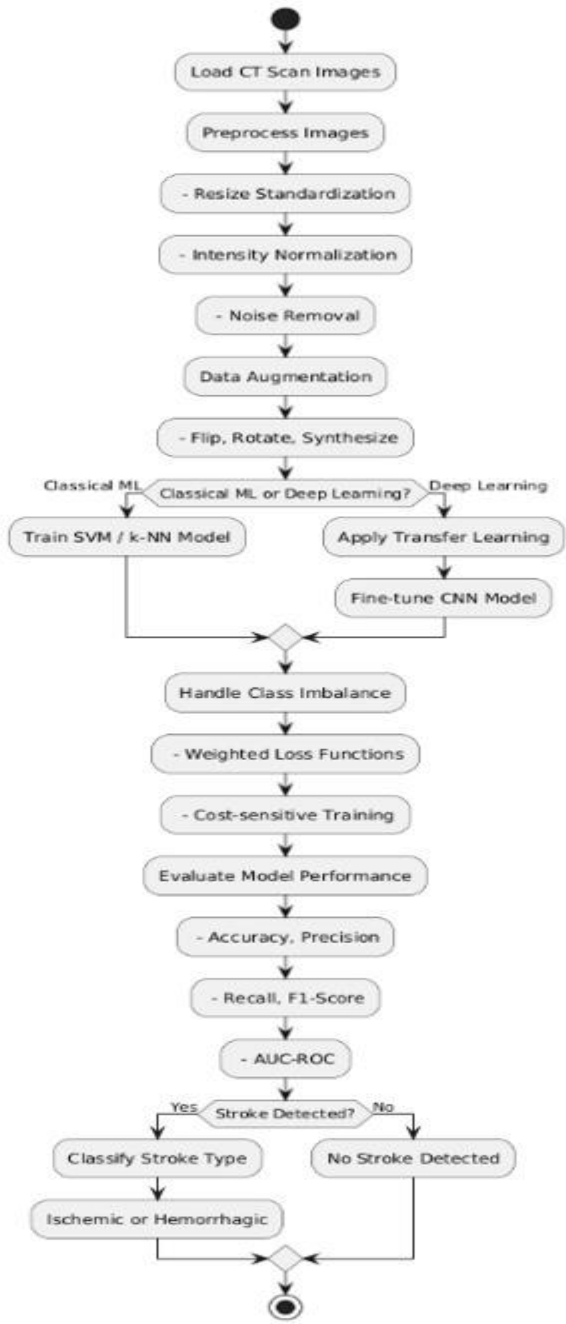


Figure 3.19 Activity Diagram

3.3.3.3 Testing the Model :

Testing models to find strokes in scans has many steps. First, collect many images of brains from CT or MRI scans. Some images show strokes, others do not. Clean up the images to make them better and make features more similar. Pull out key parts from the images using simple or advanced methods like deep learning. Pick a model to use, like CNN, SVM, or a mix. Train your model with some of the images. Check how it works with the rest of the images. Look at accuracy, sensitivity, and specificity to see how well it finds strokes and tells them apart from other cases. Doing this helps check if the model works well for real hospital work.

3.3.4 OUTCOME OF OBJECTIVE

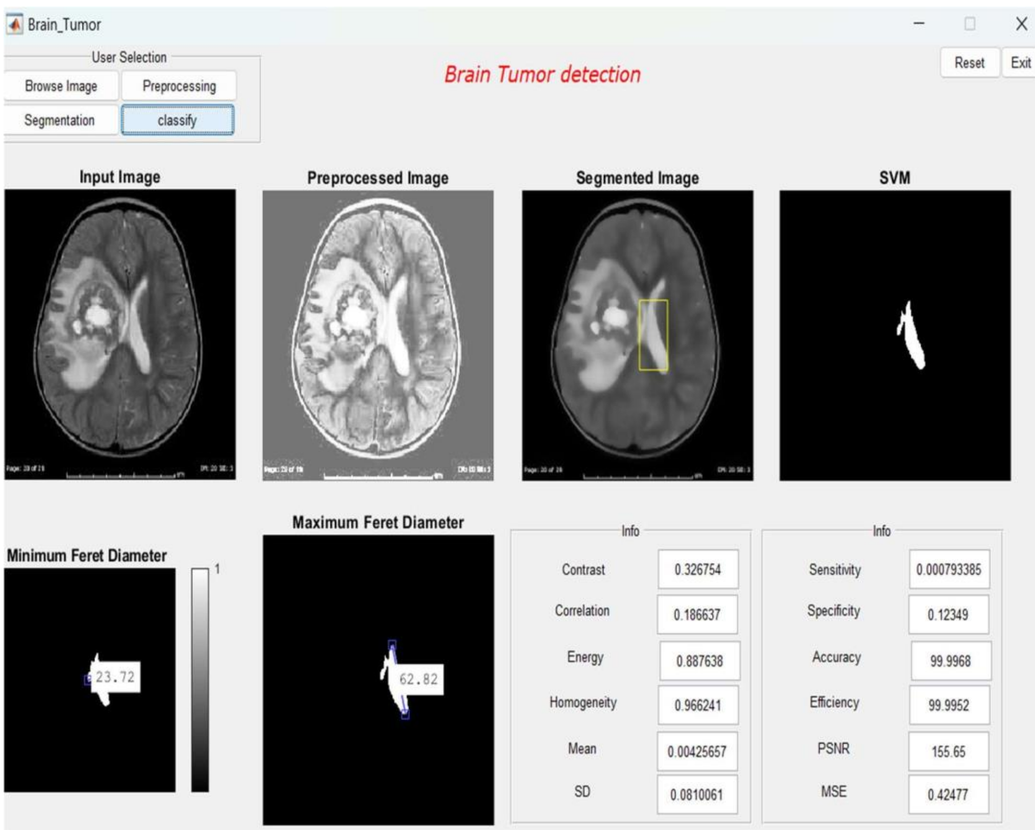


Fig 3.20 Outcome of Classified Image

This work was aimed at building a robust stroke prediction system in which brain scans were classified into three categories: ischemic, hemorrhagic, and normal. This system was designed to support doctors in making quick and informed decisions. Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models were used and evaluated to determine the most effective method for stroke classification. Additional data were incorporated, and class imbalances were addressed through balancing techniques to ensure that the models were trained on an improved dataset, allowing better generalization and reduced error rates. High accuracy, sensitivity, and specificity were expected to be achieved by the final system. It was designed to be scalable and user-friendly, making it suitable for integration into hospital environments. Through this approach, an AI-powered tool was established to assist in stroke detection, accelerate treatment, and enhance the quality of patient care.

3.3.5 SPRINT RETROSPECTIVE

	Sprint Retrospective			
Liked	**Learned**	**Lacked**	**Longed For**	
Share aspects of the sprint that you enjoyed or found particularly effective.	*Discuss lessons learned, whether they are related to processes, technical aspects, or teamwork.*	*Identify areas where the team felt a lack of resources, support, or information.*	*Discuss any desires or expectations that the team had but were not met during the sprint.*	*Guidelines*
Comparing SVM, k-NN, and CNN models provided valuable insights into model accuracy and confidence.	Probability calibration and confidence scoring improve the interpretability of classification results.	Initial confidence scores were poorly aligned with actual model accuracy.	Better automated tools for calibrating and validating confidence scores.	Always validate models statistically and clinically before deployment.
Multi-class classification into ischemic, hemorrhagic, and normal scans improved decision support potential.	Fine-tuning and careful validation are critical for reliable multi-class models.	Sourcing a balanced and clinically validated labeled dataset was challenging.	A richer, diverse dataset covering all stroke types and patient demographics.	Always validate models statistically and clinically before deployment.

Fig 3.21 Sprint retrospective

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Project Outcome

1. Objective:

The project aims to develop a reliable automated system for stroke classification and detection using neuroimaging data (CT and MRI scans). This system will analyze image features and patterns to identify stroke presence, type, and potentially affected areas, assisting healthcare professionals in faster diagnosis and treatment planning.

2. Data Processing and Feature Extraction:

- **Image Pre-processing:** To make brain pictures better and the same for models, we clean them up by cutting noise, taking out the bad parts, and making all pictures look alike.
- **Feature Extraction:** Images are processed to find important parts. You can pick features by hand using medical know-how or let deep learning models, like Convolutional Neural Networks (CNNs), do it. CNNs learn straight from image data on their own.
- **Feature Selection/Dimensionality Reduction:** Techniques like mRMR or PCA help pick key features and cut down data size, making models run better.

3. Dataset Preparation:

- A dataset of neuroimages (CT and MRI scans) is collected, including cases labeled as stroke and non-stroke, and potentially further classified by stroke subtypes (ischemic, hemorrhagic).
- The set is cut into train, check, and test sets. The train set helps teach the models. The check set tweaks model settings and helps pick the best one. The test set checks how the last model works with new data.

4. Model Comparison:

- **CNNs :**CNNs are highly effective for image analysis and are explored as a primary model for stroke detection, classifying images directly or extracting features for other classifiers. Different architectures are compared based on accuracy, speed, and computational requirements.
- **Traditional Machine Learning Models :**These models are explored, particularly in combination with feature extraction from CNNs or pre-processed images. Their performance is compared against CNNs and each other to identify the most suitable approach.
- **Hybrid Models :** Hybrid approaches combining different models, such as using genetic algorithms for feature selection networks for classification, are investigated to leverage the strengths of each model and potentially improve overall performance.

5. Model Validation:

- **Cross-validation :** K-fold cross-validation is applied to evaluate the models' robustness and generalization ability.
- **Performance Metrics:** We check models by looking at things like accuracy, how well they find stroke or miss non-stroke, how exact they are, recall, F1-score, and the area under the curve for ROC. These measures tell us how good they are at spotting cases of stroke and telling them apart from non-stroke ones.

6. Real-Time Applicability:

The potential for real-time integration of the best-performing model into clinical workflows is analyzed. This involves evaluating computational efficiency, processing speed, and resource requirements for live analysis of incoming neuroimages, aiming to expedite stroke diagnosis in real-time clinical settings.

7. Future Directions:

- **Hyperparameter Tuning:** Use best ways to tweak model settings for top performance like trying many options or smart tips.
- **Ensemble Methods:** Look at mixing many models with group methods. Use stacking, bagging, or boosting to take different strengths and maybe improve how well they predict overall.
- **Explainability and Interpretability:** Investigate methods to make the models more transparent and understandable, enhancing trust and facilitating clinical adoption.
- **Data Augmentation:** To boost the amount and range of training data, use data augmentation techniques. These methods can improve how well models perform and adapt.
- **Multi-modal Learning:** Explore integrating other data sources (e.g., patient demographics, clinical history) with neuroimaging data to improve prediction accuracy and provide more comprehensive insights.
- **Federated Learning:** Investigate using federated learning to train models on decentralized datasets across multiple institutions, enhancing data privacy and enabling collaborative model development.

CHAPTER 5

CONCLUSION AND FUTURE ENHANCEMENT

5.1 Conclusion:

In conclusion, the user stories reflect a well-structured and goal-oriented approach toward building an effective stroke detection and classification system using neuroimaging and machine learning. The process begins with the acquisition and organization of real-world CT and MRI brain scan datasets, which form the foundation for training reliable models. Critical preprocessing tasks such as skull stripping, noise reduction, and image normalization ensure that the data fed into the models is clean, consistent, and focused solely on the relevant brain regions. These steps help to minimize irrelevant variations and enhance the ability of machine learning algorithms to extract meaningful features for stroke classification.

Moreover, by incorporating region-specific labeling (ischemic, hemorrhagic, and normal) and integrating segmentation with classification, the system is designed to learn from clinically significant patterns within the brain. The evaluation and comparison of multiple models, including CNN, SVM, and k-NN, allow for the identification of the most accurate and efficient approach for stroke prediction. Through the use of data augmentation and class balancing techniques, model robustness and fairness are improved, preventing overfitting and bias. The end goal is the development of a clear, scalable, and trustworthy stroke prediction tool that supports healthcare professionals in making fast, informed decisions—ultimately leading to better patient outcomes in real-world hospital and clinical environments.

In the future, the system can be enhanced by incorporating longitudinal neuroimaging data to monitor stroke progression and recovery over time. This would allow the model not only to classify stroke types but also to predict outcomes and track patient improvement. Additionally, integration with electronic health records (EHR) could enable the use of clinical metadata such as age, gender, medical history, and symptoms to improve the contextual accuracy of predictions. Expanding the dataset to include pediatric and rare stroke cases could also increase the generalizability and applicability of the model across diverse populations.

Another potential enhancement is the development of a real-time diagnostic interface that can be embedded into clinical workflows. This interface could provide visual explanations (e.g., heatmaps or attention maps) showing which regions influenced the model’s decision, improving trust and transparency for clinicians. Furthermore, the system could be deployed on cloud-based platforms or edge devices to support remote or resource-constrained healthcare settings. Regular model updates through federated learning could also be introduced, allowing the system to learn from new data while preserving patient privacy. These improvements would make the stroke prediction system even more robust, interpretable, and accessible for global clinical adoption.

REFERENCES

1. J. Chaki and M. Woźniak, "Deep Learning and Artificial Intelligence in Action (2019–2023): A Review on Brain Stroke Detection, Diagnosis, and Intelligent Post-Stroke Rehabilitation Management," in IEEE Access, vol. 12, pp. 52161-52181, 2024, doi: 10.1109/ACCESS.2024.3383140.
2. M. A. Saleem et al., "Innovations in Stroke Identification: A Machine Learning-Based Diagnostic Model Using Neuroimages," in IEEE Access, vol. 12, pp. 35754-35764, 2024, doi: 10.1109/ACCESS.2024.3369673.
3. Y. Zhou et al., "In Vivo Transcranial Acoustoelectric Brain Imaging of Different Deep Brain Stimulation Currents," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 597-606, 2024, doi: 10.1109/TNSRE.2024.3356440.
4. C. Chen et al., "Tracking the Immediate and Short-Term Effects of Continuous Theta Burst Stimulation on Dynamic Brain States," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1386-1396, 2024, doi: 10.1109/TNSRE.2024.3378712.
5. X. Yuan et al., "Cross-Domain Identification of Multisite Major Depressive Disorder Using End-to-End Brain Dynamic Attention Network," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 33-42, 2024, doi: 10.1109/TNSRE.2023.3341923.
6. S. Mamoon, Z. Xia, A. Alfakih and J. Lu, "UCLN: Toward the Causal Understanding of Brain Disorders With Temporal Lag Dynamics," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 3729-3740, 2024, doi: 10.1109/TNSRE.2024.3471646.
7. L. Hu et al., "Subject-Independent Wearable P300 Brain–Computer Interface Based on Convolutional Neural Network and Metric Learning," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 3543-3553, 2024, doi: 10.1109/TNSRE.2024.3457502.
8. "Estimating Functional Brain Networks by Low-Rank Representation With Local Constraint" *IEEE Transactions on Neural Systems and Rehabilitation Engineering, Z. Li et al., 2024, vol. 32, pp. 684-695*
9. "A Siamese Convolutional Neural Network for Identifying Mild Traumatic Brain Injury and Predicting Recovery" *IEEE Transactions on Neural Systems and Rehabilitation Engineering, F. Koochaki et al., 2024, vol. 32, pp. 1779-1786*

10. "Distinct Time-Resolved Brain-Wide Coactivations in Oxygenated and Deoxygenated Hemoglobin" *IEEE Transactions on Biomedical Engineering, A. F. Khan et al., 2024, vol. 71, no. 8, pp. 2463-2472*
11. Fernandes, J. N. D., Cardoso, V. E. M., Comesaña-Campos, A., & Pinheira, A. (2024). Comprehensive Review: Machine and Deep Learning in Brain Stroke Diagnosis. *Sensors*, 24(13), 4355.
12. Zhou, Y., et al. (2024). In Vivo Transcranial Acoustoelectric Brain Imaging of Different Deep Brain Stimulation Currents. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 597-606.
13. Chen, C., et al. (2024). Tracking the Immediate and Short-Term Effects of Continuous Theta Burst Stimulation on Dynamic Brain States. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 1386-1396.
14. Yuan, X., et al. (2024). Cross-Domain Identification of Multisite Major Depressive Disorder Using End-to-End Brain Dynamic Attention Network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 33-42.
15. Mamoon, S., Xia, Z., Alfakih, A., & Lu, J. (2024). UCLN: Toward the Causal Understanding of Brain Disorders With Temporal Lag Dynamics. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 3729-3740.
16. Hu, L., et al. (2024). Subject-Independent Wearable P300 Brain–Computer Interface Based on Convolutional Neural Network and Metric Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 3543-3553.
17. Li, Z., et al. (2024). Estimating Functional Brain Networks by Low-Rank Representation With Local Constraint. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 684-695.
18. Koochaki, F., et al. (2024). A Siamese Convolutional Neural Network for Identifying Mild Traumatic Brain Injury and Predicting Recovery. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 1779-1786.
19. Khan, A. F., et al. (2024). Distinct Time-Resolved Brain-Wide Coactivations in Oxygenated and Deoxygenated Hemoglobin. *IEEE Transactions on Biomedical Engineering*, 71(8), 2463-2472.
20. Fernandes, J. N. D., et al. (2024). Comprehensive Review: Machine and Deep Learning in Brain Stroke Diagnosis. *Sensors*, 24(13), 4355.
21. Zhou, Y., et al. (2024). In Vivo Transcranial Acoustoelectric Brain Imaging of Different Deep Brain Stimulation Currents. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 597-606.

22. Chen, C., et al. (2024). Tracking the Immediate and Short-Term Effects of Continuous Theta Burst Stimulation on Dynamic Brain States. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 1386-1396.

23. Yuan, X., et al. (2024). Cross-Domain Identification of Multisite Major Depressive Disorder Using End-to-End Brain Dynamic Attention Network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 33-42.

24. Mamoon, S., Xia, Z., Alfakih, A., & Lu, J. (2024). UCLN: Toward the Causal Understanding of Brain Disorders With Temporal Lag Dynamics. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 3729-3740.

APPENDIX

A. SAMPLE CODE

```
function varargout = Brain_Tumor(varargin)
% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @Brain_Tumor_OpeningFcn, ...
                   'gui_OutputFcn',  @Brain_Tumor_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT

% --- Executes just before Brain_Tumor is made visible.
function Brain_Tumor_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to Brain_Tumor (see VARARGIN)
```

```
% Choose default command line output for Brain_Tumor
handles.output = hObject;
axes(handles.axes1); axis off
axes(handles.axes2); axis off
axes(handles.axes3); axis off
axes(handles.axes6); axis off
% axes(handles.axes7); axis off
axes(handles.axes10); axis off
axes(handles.axes9); axis off

set(handles.edit14,'String','--');
set(handles.edit15,'String','--');
set(handles.edit16,'String','--');
set(handles.edit17,'String','--');
set(handles.edit18,'String','--');
set(handles.edit19,'String','--');
```

Output

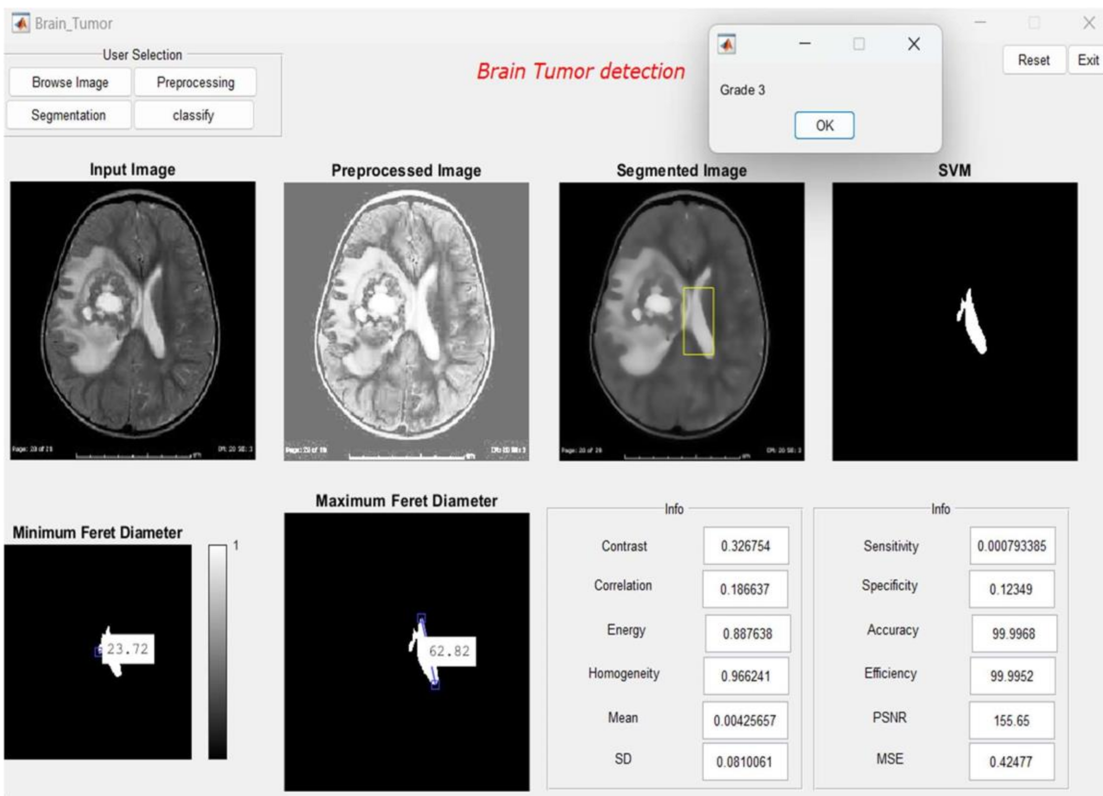


Fig 3.23 Output for Brain Stroke Detection

B. RESEARCH PAPER

Stroke Classification and Detection Using Neuroimage-Based ML Models

P.Naveen, M.Poorna Pradeep Reddy, *Dr. G.Senthil Kumar *Corresponding Author
Department of Computer Science and Engineering, SRM University, Kattankulathur-603203,India
mm8251@srmist.edu.in,pp5958@srmist.edu.in,Senthilg1@srmist.edu.in

Abstract—

Diseases have an immense impact on the health and lifestyle of the patient. Proper and early diagnosis is very much essential for treatment. The medical field currently employs a lot of manual interpretation of images. The process has errors and is inconsistent. The paper suggests an efficient system that identifies various diseases, including heart conditions and tumors, using human fundus images. The solution requires preprocessing the images to remove noise and irrelevant background using filtering and transformation techniques. A grey-level co-occurrence matrix is used for accurate segmentation, and texture and color features specific to different disease types are extracted. The system uses a Convolutional Neural Network (CNN) classification method for the identification of multiple diseases based on the processed features. This would imply that the new system improves diagnosis accuracy in respect to the diseases; cuts down human errors, provides reliable information and tools to practitioners, which might lead to improving patient outcome with earlier interventions.

Keywords— *Diagnosis, Diseases, MRI Images, Preprocessing, Segmentation, Features, CNN Classification.*

1.Introduction

Diseases have a profound impact on the health and lifestyle of patients, making early and accurate diagnosis crucial for effective treatment. The medical field currently relies heavily on manual interpretation of images, a process prone to errors and inconsistencies. This paper proposes an efficient system for identifying various diseases, including heart conditions and tumors, using human fundus images. The solution involves preprocessing the images to remove noise and irrelevant background [1] through filtering and transformation techniques. A grey-level co-occurrence matrix is employed for accurate segmentation, and texture and color features specific to different disease types are extracted. The system utilizes a convolutional neural network (CNN) classification method to identify multiple diseases based on the processed features. This approach aims to improve diagnosis accuracy, reduce human errors, and provide reliable information and tools to practitioners, potentially leading to better patient outcomes through earlier interventions.

Early diagnosis is often the course of treatment that determines the patient's prognosis and therefore cannot be overemphasized. Image interpretation, as in the medical field, has traditionally been time-consuming and prone to human error. These are among the many reasons that make the requirements for developing some automated systems that can assist healthcare professionals in making accurate and timely diagnoses important [2]. This system will employ the most current image processing techniques along with machine learning algorithms to meet the challenges posed by these issues. The system could accurately segment out various types of diseases and classify them by preprocessing the images to enhance the quality of the images and extracting relevant features. This automated approach not only streamlines the diagnostic process but also ensures consistency and reliability in the results.

The use of human fundus images is particularly advantageous in this context, as they provide a wealth of information about various diseases. Fundus images capture the interior surface of the eye, including the retina, optic disc, and macula, [3] which can reveal signs of conditions such as diabetic retinopathy, glaucoma, and age-related macular degeneration. By analyzing these images, the proposed system can detect abnormalities and provide valuable insights into the patient's health. The preprocessing step is crucial in this process, as it removes noise and enhances the image quality, making it easier to identify relevant features. The grey-level co-occurrence matrix is used for segmentation, which helps in isolating different regions of the image based on their texture and color properties.

Following image preprocessing and segmentation, the system will extract certain features indicative of various diseases. Such features will then be fed into a deep neural network for classification. The CNN is allowed to learn patterns and characteristics [4] associated with the type of disease through a large dataset of images of labeled classes. In fact, this training will enable the system to correctly identify multiple diseases based on the extracted features. Currently, CNNs have great potential in analyzing medical images since they have the ability to handle complex data with high levels of accuracy. Considering these benefits, the proposed system shall incorporate CNNs, aimed at enhancing the diagnostic capability and improving the outcomes of patients.

The proposed system offers a promising solution for improving the accuracy and efficiency of disease diagnosis using human fundus images [5]. By leveraging advanced image processing techniques and convolutional neural network classification, the system can reduce human errors and provide reliable information to healthcare professionals. This automated approach not only streamlines [6] the diagnostic process but also ensures consistency and reliability in the results. The potential benefits of this system include

Fig 3.24 Research Paper

C. CONFERENCE SUBMISSION

Paper ID : 2503836 Acceptance for conference & Publication IMSTEM 2025 - Submit the Paper as per the Paper Template & Submit the Abstract as per the format for Conference Proceedings. Reg External Inbox ×

RSP Research Hub <editorrspresearchhub@gmail.com>
to me

Mar 27, 2025, 10:31 AM

Dear Mr. P Naveen

Article IMSTEM Number: 2503836

Title: Fake News Detection Using Sentiment Analysis And Machine Learning Algorithms

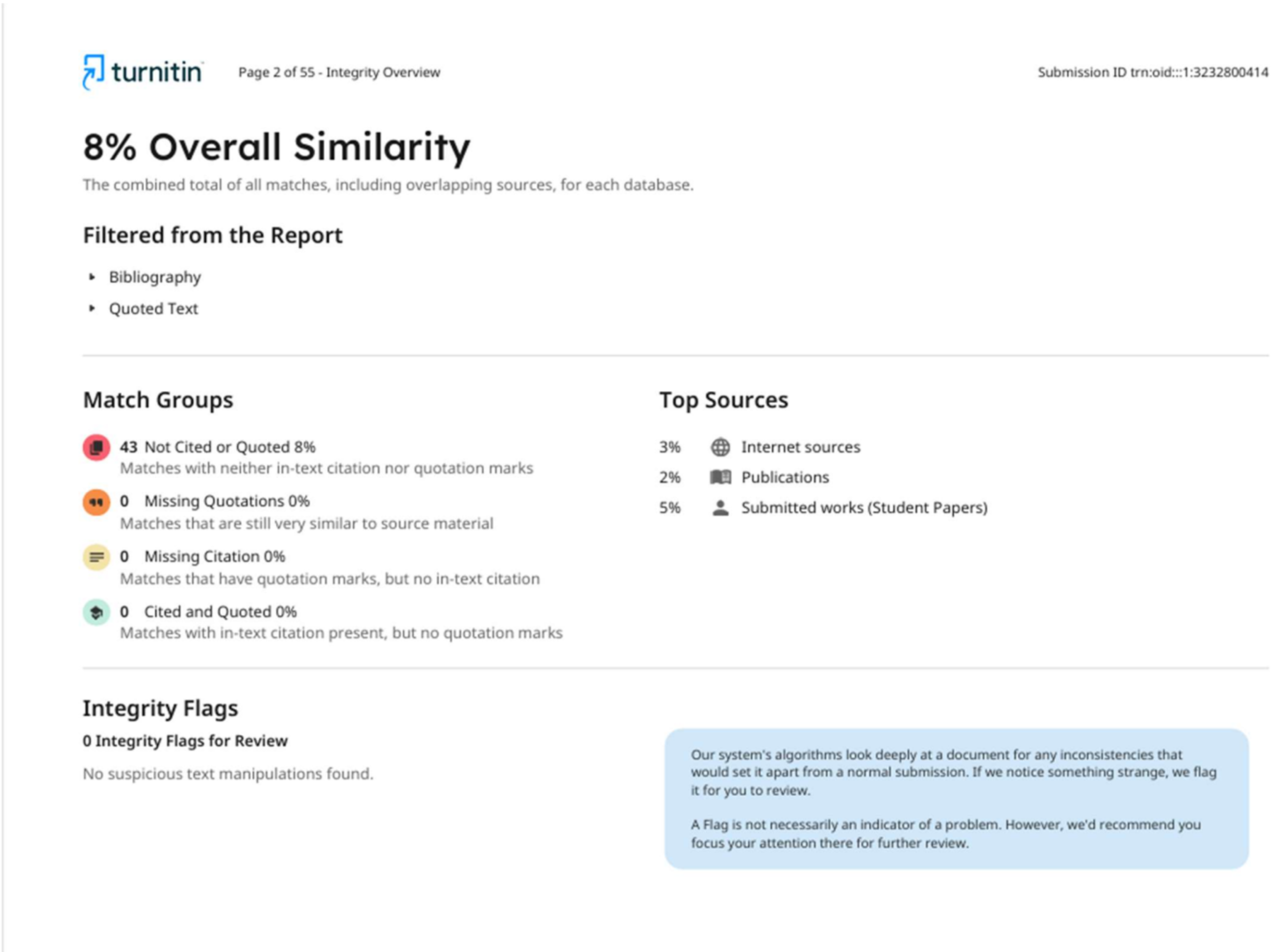
While Reply in the mail please mention your paper ID Without Fail

Warm greetings to all,

We are very much Thankful for your kind support and registration in our conference on **"International Conference on Innovations in Materials Science, Technology, Engineering, and Management for Sustainable Development" - (IMSTEM) 2025 Hybrid** mode on 30^{th} & 31^{st} March 2025 , Organized by the **St.Joseph College of Engineering ,OMR, Chennai, India & RSP Research Hub, Coimbatore, Tamil Nadu, India.**

I am pleased to inform you that your above article has been **acceptance for conference and publication in** International Journal with DOI Indexed, Google Scholar Indexed or Google Scholar Indexed Journal with ISSN Number Based on Reviewer Decision. Your article will be available online in its 'accepted' format in the month of April 2025 . If you have any questions, please contact the Publication Editor.

D. PLAGARISM REPORT



turnitin Page 2 of 55 - Integrity Overview Submission ID trn:oid:::1:3232800414

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

43 Not Cited or Quoted 8%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
2% Publications
5% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Fig 3.26 Plagiarism Report