**ABSTRACT**

A stroke, sometimes called a brain attack, occurs when something blocks blood supply to part of the brain or when a blood vessel in the brain bursts. In either case, parts of the brain become damaged or die. A stroke can cause lasting brain damage, long-term disability, or even death. The brain controls our movements, stores our memories, and is the source of our thoughts, emotions, and language. The brain also controls many functions of the body, like breathing and digestion. To work properly, your brain needs oxygen. Your arteries deliver oxygen-rich blood to all parts of your brain. If something happens to block the flow of blood, brain cells start to die **within minutes,** because they can’t get oxygen. This causes a stroke.

This project focuses on developing a web application for Stroke Prediction using Machine Learning. A stroke, a severe medical condition, occurs when blood supply to the brain is disrupted, leading to potentially life-threatening consequences. Various factors contribute to the risk of stroke, such as hypertension, diabetes, and lifestyle choices. To enhance stroke prediction, we employ machine learning algorithms, including Logistic Regression, K-Nearest Neighbors (KNN), Extratree Classifier, Gaussian Naive Bayes, Decision Tree and Random Forest Classifier. These algorithms analyze input data from users, considering diverse risk factors.

Machine learning plays a pivotal role by leveraging historical data to identify patterns and correlations, aiding in accurate stroke prediction. The web application utilizes this predictive capability, allowing users to input relevant health information for a personalized risk assessment. The expected outcome of the project is a user-friendly application with a high accuracy rate in predicting stroke probabilities. By integrating multiple algorithms, we aim to achieve a robust model that provides valuable insights, potentially enabling early intervention and preventive measures for individuals at risk of stroke.

**CHAPTER 1**

**INTRODUCTION**

**1.1 Overview**

The different body parts and how they function are the foundation of human life. A hazardous condition that ends human lives is stroke. After the age of 65, this condition is frequently discovered. Heart attacks influence the working of the heart, and strokes affect the brain similarly. One of these two conditions—a blood supply restriction to the brain or the rupture and bleeding of brain blood vessels—is what causes strokes. If there is a rupture or a blockage, blood and oxygen cannot reach the brain's tissues. In both industrialized and developing nations, it is currently the fifth greatest cause of death. A stroke victim's chances of making a full recovery are improved the earlier they receive medical care. Any stroke victim needs to see a doctor right away. Otherwise, it will result in death, permanent disability, and brain damage. Patients can develop stoke for a variety of reasons. Diet, inactivity, alcohol, tobacco, personal history, medical history, and complications are the main causes of stroke, according to National Heart, Lung, and Blood Institute. Predicting brain strokes is crucial for early detection and prevention of potentially life-threatening events. Brain strokes, also known as cerebrovascular accidents, occur when blood flow to the brain is interrupted, either by a blockage (ischemic stroke) or bleeding (hemorrhagic stroke). Predictive models for brain stroke aim to identify individuals at high risk based on various factors, including medical history, lifestyle choices, and physiological indicators. These models leverage advanced techniques such as machine learning algorithms and statistical analysis to analyze large datasets and identify patterns that precede stroke occurrence. By utilizing these predictive models, healthcare professionals can intervene proactively, implementing preventative measures and lifestyle modifications to mitigate stroke risk factors in at-risk individuals. Effective stroke prediction not only saves lives but also reduces the long-term impact and burden on healthcare systems by preventing debilitating strokes and associated complications.

* 1. **Symptoms of Brain Stroke**
* **Trouble speaking and understanding what others are saying**. A person having a stroke may be confused, slur words or may not be able to understand speech.
* **Numbness, weakness or paralysis in the face, arm or leg**. This often affects just one side of the body. The person can try to raise both arms over the head. If one arm begins to fall, it may be a sign of a stroke. Also, one side of the mouth may droop when trying to smile.
* **Problems seeing in one or both eyes.** The person may suddenly have blurred or blackened vision in one or both eyes. Or the person may see double.
* **Headache.** A sudden, severe headache may be a symptom of a stroke. Vomiting, dizziness and a change in consciousness may occur with the headache.
* **Trouble walking.** Someone having a stroke may stumble or lose balance or coordination.

**1.2.1 Controllable Risk Factors**

* **High Blood Pressure:** A blood pressure reading of 140/90 or above increases stroke risk 4-6 times.
* **High Cholesterol:** High levels of cholesterol in the bloodstream can clog arteries and cause a stroke or heart attack.
* **Diabetes:** Diabetes increases stroke risk 2-4 times.
* **Atrial Fibrillation (AF):** irregular heartbeats. AF increases stroke risk up to 6 times.
* **Tobacco Use/Smoking:** Smoking doubles the risk of stroke.
* **Alcohol Use:** Drinking more than 2 drinks per day may increase stroke risk by 50%.
* **Physical Inactivity or Obesity:** Excess weight puts a strain on the entire circulatory system. It also makes people more likely to have high cholesterol, high blood pressure and diabetes -- all of which can increase your risk for stroke.

**1.2.2 Uncontrollable Risk Factors**

* **Age.** Your risk of stroke increases with age. After the age of 55, your stroke risk doubles for every decade.
* **Gender.** Stroke is more common in men than women. But more women than men die from stroke.
* **Race.** If you are African American, your risk is twice the rate for whites. If you are Hispanic or Asian/Pacific Islander, your stroke risk is also higher than Caucasians.
* **Family History.** If someone in your family has had a stroke, you have a higher risk of stroke yourself.
* **Previous Stroke or Transient Ischemic Attack.** If you have already had a stroke or a transient ischemic attack (ministroke), you have a 25-40% chance of having another stroke in the next 5 year.

**1.3 Brain Stroke: The Two-Faced Culprit**

Brain Strokes, often dubbed “brain attacks,” come in two distinct types: ischemic and hemorrhagic.

1. **Ischemic Stroke:** Imagine a tiny traffic jam inside your brain’s highways. Ischemic strokes occur when a blood clot or plaque build-up blocks the flow of oxygen-rich blood within an artery. This blockade can be triggered by conditions like sclerosis, clotting disorders, heart defects, or microvascular diseases.
2. **Hemorrhagic Stroke:** Now, picture a catastrophic explosion in a narrow alley. Hemorrhagic strokes happen when a blood vessel bursts open leaking blood inside the brain.

High blood pressure, brain aneurysms, and brain tumours are common instigators of this explosive event.

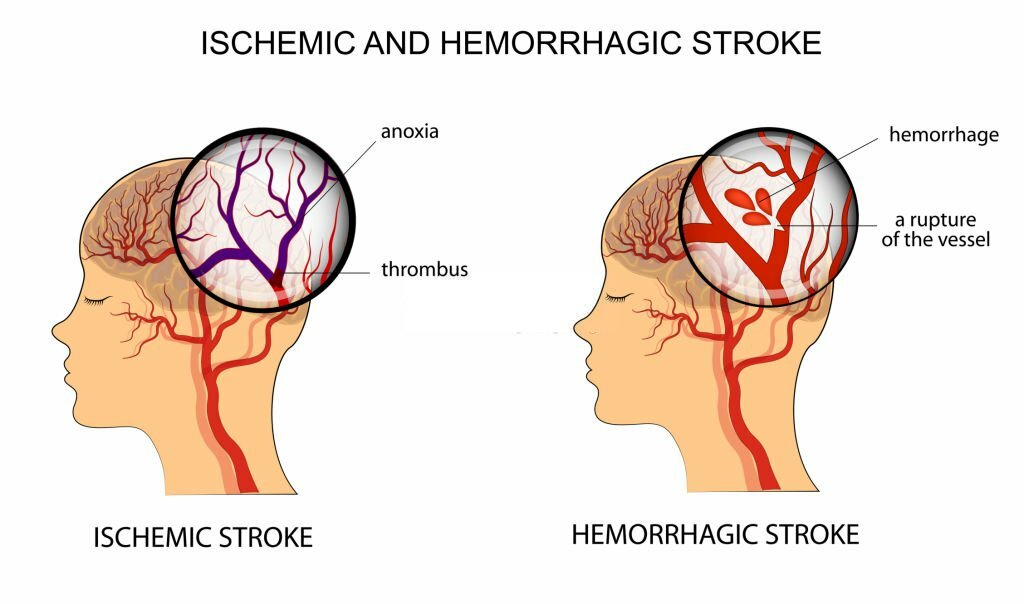


Figure 1.1: Ischemic And Hemorrhagic Stroke

Our brain is a demanding organ, and it needs a constant supply of oxygen. To cater to the brain’s demand, arteries supply oxygen-rich blood. If something blocks this essential flow for even a minute, our brain pays the price, potentially leading to lasting brain damage, long-term disability, or even death.

**1.4 Brain Stroke Treatment**

Brain Stroke treatment hinges on whether it’s an ischemic or hemorrhagic stroke. Sum Ultimate Neurology Department is equipped with advanced technologies and tools for treatment, therapies and surgeries. Depending on the type of stroke, the course of action will be recommended:

1. **Ischemic Stroke:**

* **Emergency IV Medication:**This is a therapy with drugs administered within 4.5 hours from symptom onset to dissolve the clot causing the stroke. Sooner the medication is administrated, better the chances of survival and reduced complications.
* **Emergency Endovascular Procedures:** This is done in two ways and could significantly improve the outcome and reduce long-term disability. One, a catheter is inserted to deliver medication directly to the brain where the stroke is happening. Two, direct clot removal with a stent retriever, particularly for large clots that can’t be completely dissolved with emergency medication.
* **Carotid Endarterectomy:**Surgery to remove plaque blocking a carotid artery.
* Angioplasty and Stents: Surgery involves temporarily inserting and inflating a tiny balloon into the clogged artery to widen the area and inserting stents to avoid the possibility of narrowing again.

1. **Hemorrhagic Stroke:**

* **Emergency Measures:** If the patient has taken blood-thinning medications to prevent blood clots, then drugs or transfusion of blood products are given to counteract its effects, lower blood pressure, and prevent seizures.
* **Surgery:** Removing the blood and repairing damaged blood vessels.
* Surgical Clipping: Clamping the aneurysm at its base to keep it from rapturing or stopping blood flow.
* **Coiling:** Placing detachable coils into an aneurysm to block blood flow that can cause further blood clots.
* **Surgical AVM Removal:** Removing tangled blood vessels to eliminate or lower the risk of rapture.
* **Stereotactic Radiosurgery:** Using focused radiation to repair blood vessel malformations.

**1.5 Aim and Objectives**

The aim of our project is to apply principles of machine learning over large existing data sets to effectively predict individual stroke risk with accuracy.

**Objectives**

* To collect datasets of brain stroke from hospitals, dataset repositories
* Design and develop an algorithm for early detection of brain stroke to take precautions.
* Developing a web application for early stroke detection, prediction and providing remedies to the above solution.

**1.6 Scope of Project**

The scope of a brain stroke prediction project can include various aspects such as data collection, feature selection, model development (like machine learning algorithms), validation, and deployment. It involves analyzing factors like medical history, lifestyle habits, and possibly genetic predispositions to predict stroke risk. Additionally, integrating real-time monitoring systems or wearable devices could enhance prediction accuracy.

**CHAPTER 2**

**LITERATURE SURVEY**

**Title:** Prediction of Brain Stroke using Machine Learning Algorithms and Deep Neural Network Techniques

**Year:**  2023

**Authors:** Senjuti Rahman, Mehedi Hasan, and Ajay Krishno Sarkar

**Description:** In this paper Early detection of the numerous stroke warning symptoms can lessen the stroke's severity. The main objective of this study is to forecast the possibility of a brain stroke occurring at an early stage using deep learning and machine learning techniques. To gauge the effectiveness of the algorithm, a reliable dataset for stroke prediction was taken from the Kaggle website. Several classification models, including Extreme Gradient Boosting (XGBoost), Ada Boost, Light Gradient Boosting Machine, Random Forest, Decision Tree, Logistic Regression, K Neighbors, SVM - Linear Kernel, Naive Bayes, and deep neural networks (3-layer and 4-layer ANN) were successfully used in this study for classification tasks. The Random Forest classifier has 99% classification accuracy, which was the highest (among the machine learning classifiers). The three layer deep neural network (4-Layer ANN) has produced a higher accuracy of 92.39% than the three-layer ANN method utilizing the selected features as input. The research's findings showed that machine learning techniques outperformed deep neural networks.

**Title:** Stroke prediction using SVM

**Year:** 2016

**Authors:** Jeena, R S, Kumar, Sukesh

**Description:** In this paper Early diagnosis of stroke is essential for timely prevention and treatment. Investigation shows that measures extracted from various risk parameters carry valuable information for the prediction of stroke. This research work investigates the various physiological parameters that are used as risk factors for the prediction of stroke. Data was collected from International Stroke Trial database and was successfully trained and tested using Support Vector Machine (SVM). In this work, we have implemented SVM with different kernel functions and found that linear kernel gave an accuracy of 90 %.

**Title:** Prediction of Stroke Using Deep Learning Model

**Year:** 2017

**Authors:** Pattanapong Chantamit-o-pas, Madhu latha

**Description:** This research reports predictive analytical techniques for stroke using deep learning model applied on heart disease dataset. The atrial fibrillation symptoms in heart patients are a major risk factor of stroke and share common variables to predict stroke. The outcomes of this research are more accurate than medical scoring systems currently in use for warning heart patients if they are likely to develop stroke.

**CHAPTER 3**

**CAPSTONE PROJECT**

**3.1 Capstone Project Planning**

The capstone project is designed to consolidate final-year students’ learning with valuable hands-on experience to help develop them into well-prepared and well-rounded graduates. Students work together in small groups to come up with innovative solutions for real-life problems, all while gaining valuable insights into the demands and responsibilities of the working world. This gives students a chance to bring their leadership and management skills alive and understand the consequences of their decisions in a ‘safe space’. By the increase demand and the effectiveness of the capstone project which is done by the students around the globe, it has become an important component of many universities' degree programs. It can take many different shapes, but the goal is the same. The capstone project provides a once-in-a-lifetime opportunity to conduct autonomous group study in order to develop a novel solution to a real-world problem. A project of this size and scope might be difficult, but it can also be extremely gratifying.

In many universities the capstone project is given as the final assignment and, because it follows so much of practical approaches & applications and due to its very high capacity to help students in improving their professional knowledge and skills, it serves an important role in preparing students for the professional industry. By doing the capstone project, students can easily consolidate their learning on the real-world problem and it provides the students a very valuable hands-on experience which allows them to become well-prepared and well-rounded graduates.

Students collaborate in small groups to develop creative solutions to real-world challenges, all while learning about the demands and responsibilities of the workplace. This allows students to practice their leadership and management abilities in a secure environment while also understanding the implications of their decisions.

Planning a capstone project is an important step towards successful completion. Here are some steps you can follow to plan your capstone project effectively:

**1. Understand the Requirements:** Start by thoroughly understanding the requirements and guidelines set by your educational institution or program. This includes understanding the scope, deliverables, timeline, and any specific formatting or documentation requirements.

**2. Choose a Topic:** Select a topic that aligns with your academic interests, professional goals, and the requirements of your program. Consider a topic that challenges you, allows for meaningful research, and offers practical application.

**3. Define Aim and Objectives:** Clearly define the objectives of your capstone project. What do you want to achieve through your research or implementation? Develop specific research questions or problem statements that will guide your project.

**4. Conduct Literature Survey:** Before diving into your project, conduct thorough background research on your topic. Read scholarly articles, books, and other relevant resources to gain a solid understanding of existing knowledge, theories, and practices related to your chosen area.

**5. Develop a Project Proposal:** Write a project proposal that outlines the purpose, objectives, research questions, methodology, and expected outcomes of your capstone project. This proposal will serve as a roadmap for your work and will require approval from your advisor or instructor.

**6. Create a Timeline:** Break down your project into smaller tasks and create a timeline or project schedule. Identify key milestones, such as literature review completion, data collection, analysis, and report writing. Assign deadlines to each task to ensure timely progress.

**7. Gather Resources:** Identify the resources and tools you will need to complete your project. This may include access to libraries, databases, software, hardware, or collaboration with individuals or organizations. Ensure you have everything you need before starting the project.

**8. Dataset Collection and Analysis:** If your project involves data collection, plan how you will collect, organize, and analyze the data. Determine the appropriate research methods or techniques and ensure you have the necessary permissions or ethical approvals, if applicable.

**9. Implementation of the Project:** If your capstone project involves implementing a solution or creating a prototype, outline the steps required for implementation. Consider potential challenges, resource requirements, and test or validation procedures.

**10. Documentation and Reporting:** Plan how you will document your project progress and outcomes. Determine the format for your final report or presentation, including the sections, visuals, and supporting materials required. Allocate time for revisions and proofreading.

**11. Review and Feedback:** Seek regular feedback from your Project Guide, Coordinator, or peers throughout the project. Share your progress, discuss any challenges you encounter, and incorporate their suggestions or recommendations to improve the quality of your work.

**12. Finalize and Present:** Complete your capstone project according to your plan and make any necessary revisions or improvements. Prepare a presentation or defense, if required, to showcase your project's findings, conclusions, and recommendations.

**3.1.1 Work breakdown structure (WBS)**

A work breakdown structure (WBS) is a hierarchical decomposition of the project deliverables and tasks that need to be accomplished. Here's a phase of a work breakdown structure for a capstone project:

**Phase 1: Project Initiation**

* Define project goals and objectives
* Identify project stakeholders
* Develop project charter

**Phase 2:**  **Literature Survey and Planning**

* Conduct literature review
* Define research questions and objectives
* Develop research methodology
* Create project schedule and timeline
* Identify required resources and tools

**Phase 3: Dataset Collection**

* Identify data sources
* Gather relevant data and information
* Perform data preprocessing and cleaning
* Validate data quality

**Phase 4: Model Selection and Development**

* Identify suitable machine learning algorithms
* Implement and train the selected model
* Optimize model parameters and hyperparameters
* Validate the model's performance

**Phase 5: Evaluation and Analysis**

* Evaluate model accuracy and performance metrics
* Perform statistical analysis of results
* Identify strengths and limitations of the model
* Interpret findings and draw conclusions

**Phase 6: Documentation and Reporting**

* Prepare project documentation, including the report, thesis, or presentation
* Summarize research methodology and findings
* Present results and recommendations

**Phase 7: Project Review and Reflection**

* Review the project outcomes against the initial goals and objectives
* Reflect on the project process and lessons learned
* Identify areas for improvement or further research
* Document project review and reflection

**Phase 8: Project Presentation and Defense**

* Create a compelling presentation of the project
* Practice the presentation and improve delivery skills
* Prepare for the project defense or presentation to stakeholders
* Address questions and feedback during the defense

**Phase 9: Project Closure**

* Complete any necessary project documentation or final reports
* Archive project materials and data
* Conduct a final project review and sign-off
* Celebrate project completion and acknowledge contributors

It's important to note that the specific tasks and structure of the work breakdown structure may vary depending on the nature of the capstone project and the requirements set by the educational institution or program. The WBS should be customized and tailored to the specific project and its objectives.

**3.1.2 Timeline Development – Schedule**

A project timeline is a visual list of tasks or activities placed in chronological order, which lets project managers view the entirety of the project plan in one place. A project timeline typically takes the form of a horizontal bar chart, where each task is given a name and a corresponding start and end date.

A project timeline provides an in-depth overview of the entire project from start to finish. You can see when a task starts and when it’s due—and importantly, whether or not it’s dependent on another task.

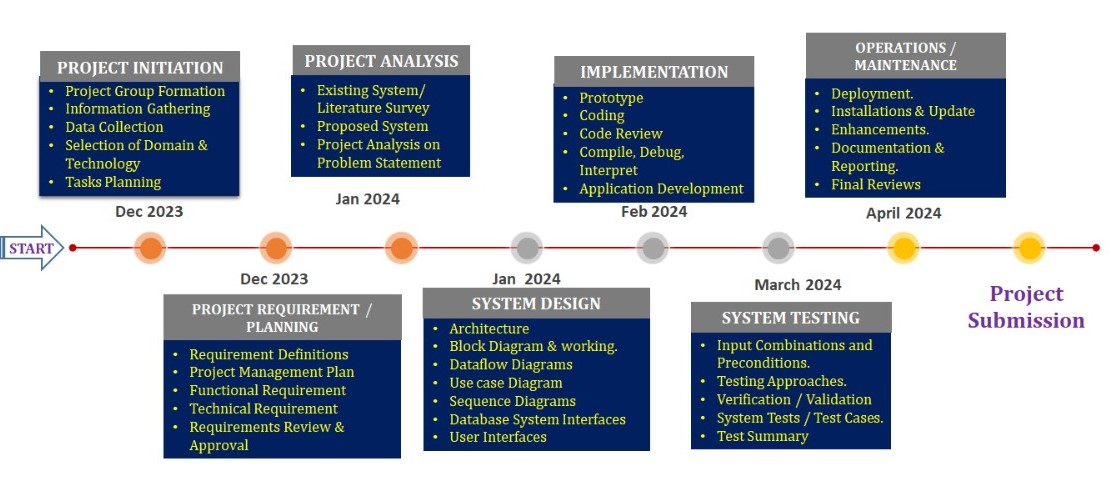


Figure 3.1 : Timeline for the Capstone Project

Using project timeline software is generally regarded as a necessity for keeping a modern project on schedule. It allows managers and teams to see the status of their tasks; that is, whether they’re in progress, overdue or done.

At its most basic, project timeline software creates a graphic representation of the project on a timeline with tasks and milestones plotted across. Tasks can be assigned start and end dates, and display a duration bar that indicates the planned amount of time to complete the tasks. More powerful software offerings can also include further options to manage tasks.

**3.1.3 Capstone Project Risks Assessment**

Risk assessment is a crucial component of any capstone project, as it helps identify potential risks and develop strategies to mitigate or manage them. Here's an example paragraph outlining the risk assessment for a capstone project:

"The capstone project entails several risks that need to be carefully assessed and managed to ensure successful project completion. One potential risk is the availability and quality of data required for the project. There is a possibility of encountering challenges in sourcing the necessary data or obtaining data of sufficient quality, which could impact the accuracy and reliability of the project outcomes. To mitigate this risk, a thorough data collection plan will be developed, including identifying multiple data sources and ensuring data validation and cleaning processes are in place. Another risk to consider is the complexity of the chosen machine learning algorithms and their implementation. If the selected algorithms are too complex or lack sufficient documentation and support, it could pose challenges in developing and optimizing the models. To address this risk, an extensive literature review will be conducted to identify well-established algorithms, and consultation with experts or mentors will be sought to ensure proper implementation. Additionally, time management and scheduling risks should be considered, as delays in task completion can lead to a cascading effect on the overall project timeline. Regular monitoring, setting realistic deadlines, and incorporating buffer time for unexpected delays will be essential strategies to mitigate this risk. By proactively assessing and addressing these risks, the capstone project can proceed with a greater likelihood of success and minimize potential obstacles along the way."

Risk assessment is a crucial step in any capstone project to identify potential risks and develop strategies to mitigate or manage them. Here is an example of a risk assessment for a capstone project:

* **Data Availability and Quality**

One of the key risks in a capstone project is the availability and quality of data. There is a possibility of facing challenges in accessing the required data or obtaining data of sufficient quality. To mitigate this risk, it is important to thoroughly research and identify potential data sources, establish data collection protocols, and perform data validation and cleaning processes to ensure the reliability and accuracy of the data.

* **Technical Challenges**

Capstone projects often involve implementing complex algorithms or utilizing advanced technologies. Technical challenges may arise during the development and implementation phase, such as software compatibility issues, limitations of selected tools or frameworks, or difficulties in integrating different components. To address this risk, it is important to conduct a thorough feasibility analysis, ensure compatibility of tools and technologies, and seek guidance from mentors or experts in the field.

* **Time Management and Project Scope**

Managing time effectively and adhering to project scope can be challenging in a capstone project, particularly when faced with unexpected obstacles or scope creep. Failure to manage time and scope effectively can lead to delays, compromised project outcomes, or incomplete deliverables. To mitigate this risk, it is crucial to develop a detailed project plan with realistic timelines, regularly monitor progress, and adjust the plan as needed to accommodate any unforeseen challenges or changes in scope.

* **Communication and Collaboration**

Capstone projects often involve teamwork and collaboration among team members or stakeholders. Communication gaps, conflicts, or a lack of coordination can pose risks to project success. To address this risk, establishing effective communication channels, setting clear expectations, and fostering a collaborative environment are important. Regular team meetings, progress updates, and conflict resolution mechanisms should be implemented to ensure effective communication and collaboration throughout the project.

* **Ethical and Legal Considerations**

Depending on the nature of the capstone project, there may be ethical and legal considerations to address, such as data privacy, intellectual property rights, or compliance with research ethics guidelines. It is crucial to thoroughly understand and adhere to relevant ethical and legal frameworks, seek necessary approvals if required, and ensure the project is conducted in an ethical and responsible manner.

By proactively identifying and assessing these risks, a capstone project can be better prepared to mitigate or manage them effectively. Regular monitoring, open communication, and adaptability are key to addressing risks and ensuring the successful completion of the project.

**3.2 System Requirements Specification**

Prerequisites exam is simple for undertaking improvement. Prerequisites need to be archived, vast, quantifiable, and testable and characterised to some extent of detail adequate for framework plan. Necessities can be engineering, underlying, social, realistic, and beneficial. A Software Requirements Specification (SRS), product requirements specific in attaining the depiction planned reason and the weather for programming being worked on.

**3.2.1 Functional Requirements**

The tools to execute the Python programs can be many, among that we can go with Visual Studio, Anaconda Navigator (Jupyter Notebook) or any IDLE based on Python. The online tool from Google can be an effective solution towards the execution of Python coding.

Functional requirements describe the system functionality, while the non-functional requirements describe system properties and constraints.

Functional requirements capture the intended behaviour of the system. This behaviour may be expressed as services, tasks, or the functions the system is required to perform. This lays out important concepts and discusses capturing functional requirements in such a way they can drive architectural decisions and be used to validate the architecture. Features may be additional functionality, or differ from basic functionality along some quality attribute. In the proposed system, concert assesses the compliance of a workflow by analysing the five established elements required to check for the rule adherence in workflows: activities, data, location, resources, and time limits. A rule describes which activities may, must or must not be performed on what object which roles. In addition, a rule can further prescribe the order of activities i.e. which activities have to happen before or after other activities.

* **Approach 1: Jupyter Notebook (Anaconda Navigator)**

This tool is also known as IPython Notebook, and it is Open-Source Distribution Software and provides the platform for development of web applications, computational interactive and specific environment for the users to create notebook documentations. It support for individual code execution, browser based interoperability, can plot various graphs using python libraries and also support for many open source libraries like Bootstrap, JQuery, Tornado, Matplotlib, Seaborn and others.

The features of Jupyter Notebook can be listed as:

* Flexible Notebook Interface
* Useful tool in Machine learning, Deep learning and Ai based Application and model Design.
* Creating and sharing the computational Documents.

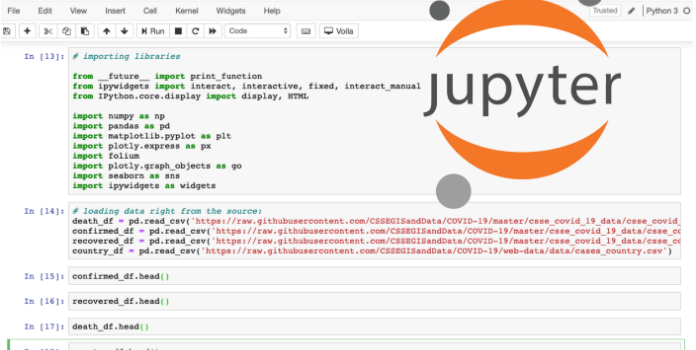


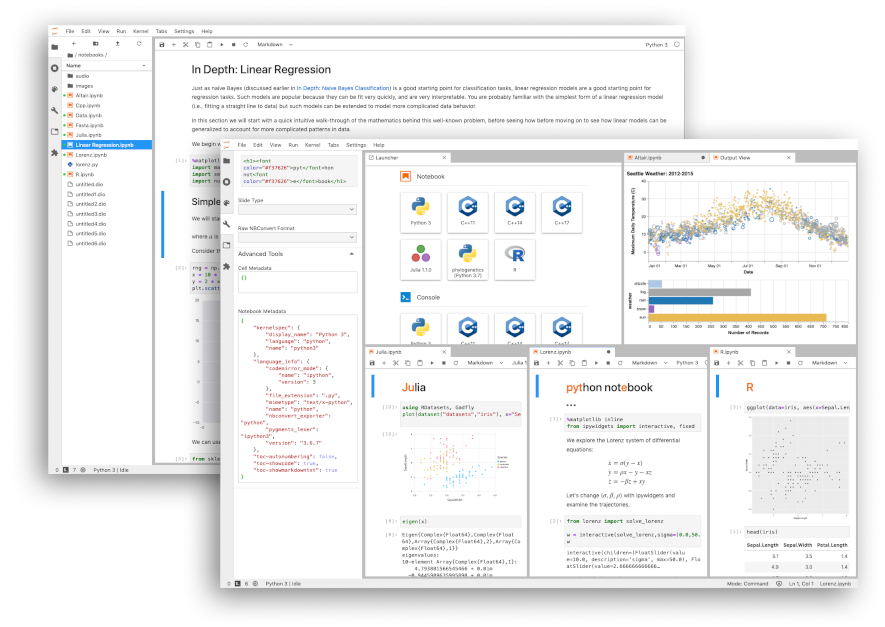
Figure 3.2: Jupyter Notebook Dashboards ****

Figure 3.3: Notebook support for plotting

* **Approach 2: Python IDLE**

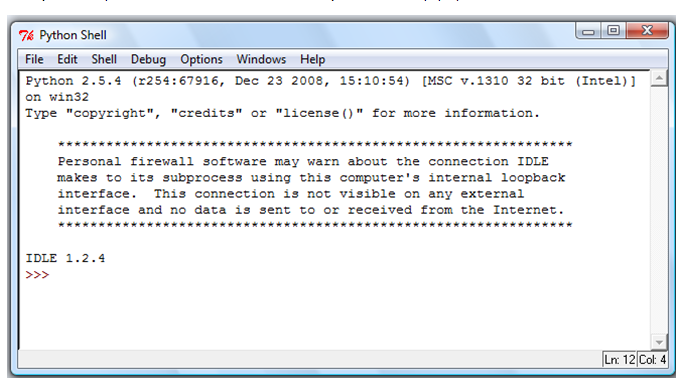
**Python IDLE (Python Integrated Development and Learning Environment) help is writing the code very effectively and efficiently and helpful tool to the Python learning who wants to start from the scratch and beginners can have an advantage to execute the code easily. This is a powerful interpreter and compiler to run the code.**

**It’s an Interactive Interpreter also known as shell, which executes the python written code, reads the input, evaluate the statements and print the output on the standard output screen provided.**

**File Editor Help to edit the code, save the program in text files and store as .py file.**



**Figure 3.4: Python IDLE Download Page**

****

**Figure 3.5: Python IDLE prompt to write and execute code**

* **Approach 3: Google Colab**

**Google Colab, also called as Colab in short is a powerful Machine Learning, Deep Learning and Data Analysis Tool that allows mixing the Python script along with text document. Rich support for Plotting the graphs, Diagram, Charts, Import Images, HTML Tags Support and LATEX format API conversions. Additional functional is it works on cloud model where document can be accessed and run on any platform independent of framework design and operating system. The runtime support for Virtual Hard Disk space and 12GB of RAM to execute the application is very excited feature of Colab. The uploading of files is very easy in this application so that it connects to the runtime.**

**Some of the important feature is:**

* **Remote Desktop Connection**
* **Runtime Environment**
* **Dataset Upload Features**
* **I/O operations and Operating System API Support**
* **General Processing Unit (GPU) availability**

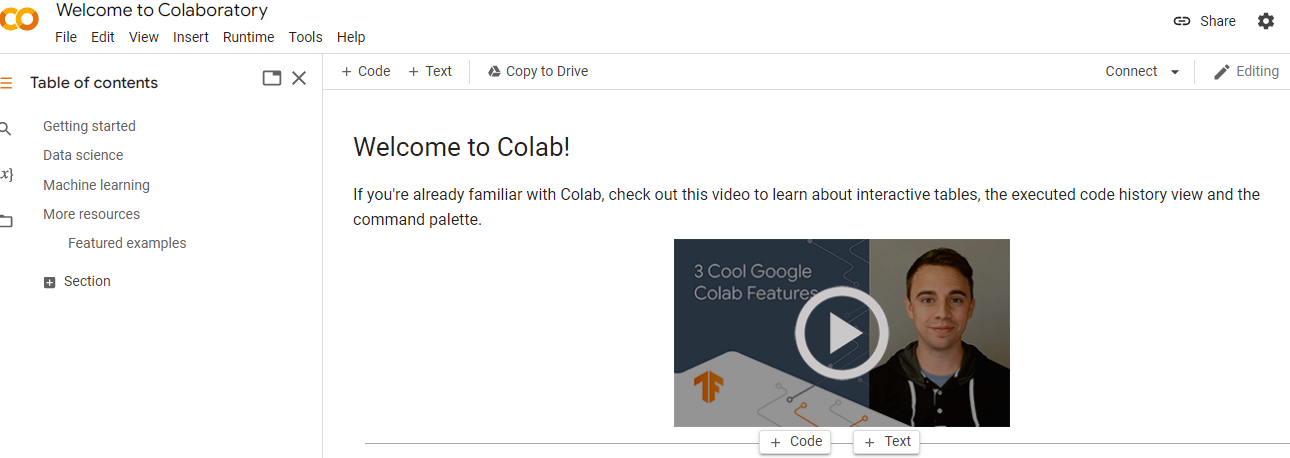


Figure 3.6: Welcome page of Google Colab



Figure 3.7: Upload the Notebook File

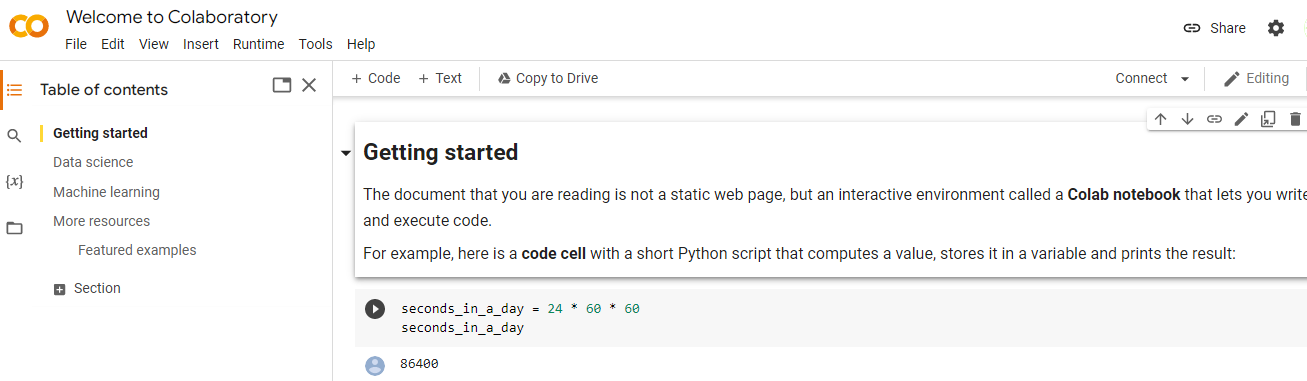


Figure 3.8: Start the Application Page

Requirements analysis is critical for project development. Requirements must be documented, actionable, measurable, testable and defined to a level of detail sufficient for system design. Requirements can be architectural, structural, behavioural, functional, and functional.

A software requirements specification (SRS) is a comprehensive description of the intended purpose and the environment for software under development.

**3.2.2 Non-functional requirements:**

**Hardware Requirements**

Processor : 3.0 GHz and Above

Output Devices : Monitor (LCD)

Input Devices : Keyboard

Hard Disk : 1 TB

**Software Requirements**

Scripting language : Python Programming

Scripting Tool : Anaconda Navigator (Jupyter Notebook) & Google Colab

Operating System : Microsoft Windows 7, 8 or 10

Dataset : Brain Stroke Dataset

Machine Learning Packages : Numpy, Pandas, Matplotlib, Seaborn Packages etc.

**3.2.3: User Input (Dataset)**

This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relavant information about the patient.

**Attribute Information**

* gender: "Male", "Female" or "Other"
* age: age of the patient
* hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
* heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
* ever\_married: "No" or "Yes"
* work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"
* avg\_glucose\_level: average glucose level in blood
* bmi: body mass index
* smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*
* stroke: 1 if the patient had a stroke or 0 if not.

**3.3 Design Specification**

A design specification is a written document that explains your product and specifies what you want it to perform as well as how the user should interact with it. While it may seem laborious to write everything down, it is the most crucial thing you can do in the early stages of product design and development. This is because the design specs direct the whole scope of your product development.

As the term suggests, a design specification is a document that outlines a project’s design needs. For the design team, this is a vital document since half of the success of a project depends on it. After extensive study, the document is created by the client and contains extremely comprehensive, specific design requirements that must be consistently applied until the project is completed. With such a list of requirements at hand, designers will be able to bring what the client sees to live.

**3.3.1 Existing System**

The existing systems for brain stroke prediction encompass a variety of approaches, ranging from traditional risk assessment methods to more advanced predictive models leveraging machine learning techniques.

* **Clinical Risk Scores:** Various clinical risk scores, such as the CHA2DS2-VASc score for atrial fibrillation-related stroke or the Framingham Stroke Risk Score, utilize demographic and clinical factors to estimate an individual's risk of stroke over a certain period. These scores are often based on statistical analyses of large cohort studies and provide a standardized approach to risk assessment in clinical practice.
* **Imaging-Based Predictive Models:** Advanced imaging techniques, such as magnetic resonance imaging (MRI) and computed tomography (CT), can detect structural and functional changes in the brain associated with stroke risk. Machine learning algorithms trained on imaging data can analyze features extracted from neuroimaging scans to predict stroke occurrence or identify individuals with preclinical signs of cerebrovascular disease.
* **Electronic Health Record (EHR) Data Analysis:** Healthcare institutions often maintain electronic health records containing comprehensive information about patients' medical history, diagnoses, medications, and laboratory results. Analyzing EHR data using machine learning algorithms allows for the development of predictive models that leverage a wide range of clinical and demographic variables to assess stroke risk and guide preventive interventions.
* **Wearable Devices and Remote Monitoring:** Wearable devices, such as smartwatches and fitness trackers, can continuously monitor physiological parameters such as heart rate, blood pressure, and physical activity levels

**3.3.2 Proposed System**

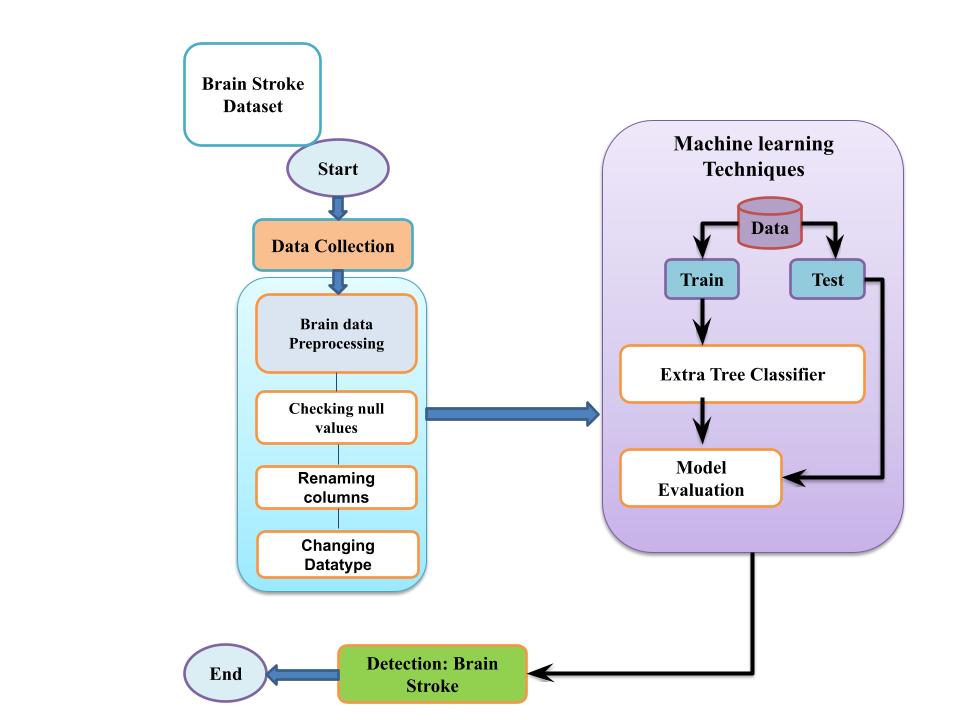


Figure 3.9: Proposed System

**Data Collection and Integration:**

* Gather diverse data sources, including electronic health records (EHRs), medical imaging (MRI, CT scans), laboratory results, genetic information, lifestyle factors, and wearable device data.
* Establish secure mechanisms for data storage, transmission, and integration, ensuring compliance with privacy regulations and data protection standards.

**Feature Selection and Engineering:**

* Identify informative features relevant to stroke prediction from the collected data sources.
* Perform feature engineering to extract meaningful insights and create new features that capture relevant patterns and relationships.
* Use domain knowledge and data-driven techniques to select a subset of features that contribute most to the predictive performance of the model.

**Machine Learning Model Development:**

* Choose appropriate machine learning algorithms, such as logistic regression, random forests, support vector machines, or deep learning models, for stroke prediction.
* Train the models using labeled data, incorporating both traditional statistical approaches and advanced machine learning techniques.
* Optimize hyperparameters and model architecture to enhance predictive performance and generalization capabilities.

**Model Evaluation:**

* Validate the predictive model using independent datasets and external validation cohorts to assess its generalization ability and robustness.
* Evaluate the system's performance using appropriate metrics, such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), or precision-recall curve. After evaluating model it predicts the result whether the person having brain stroke or not

**3.3.3 Architecture Diagram**

In machine learning (ML), an architecture diagram typically represents the high-level structure and flow of a machine learning system, including its components, data flow, and interactions. It provides a visual overview of how data is processed, models are trained, and predictions are made within the system.

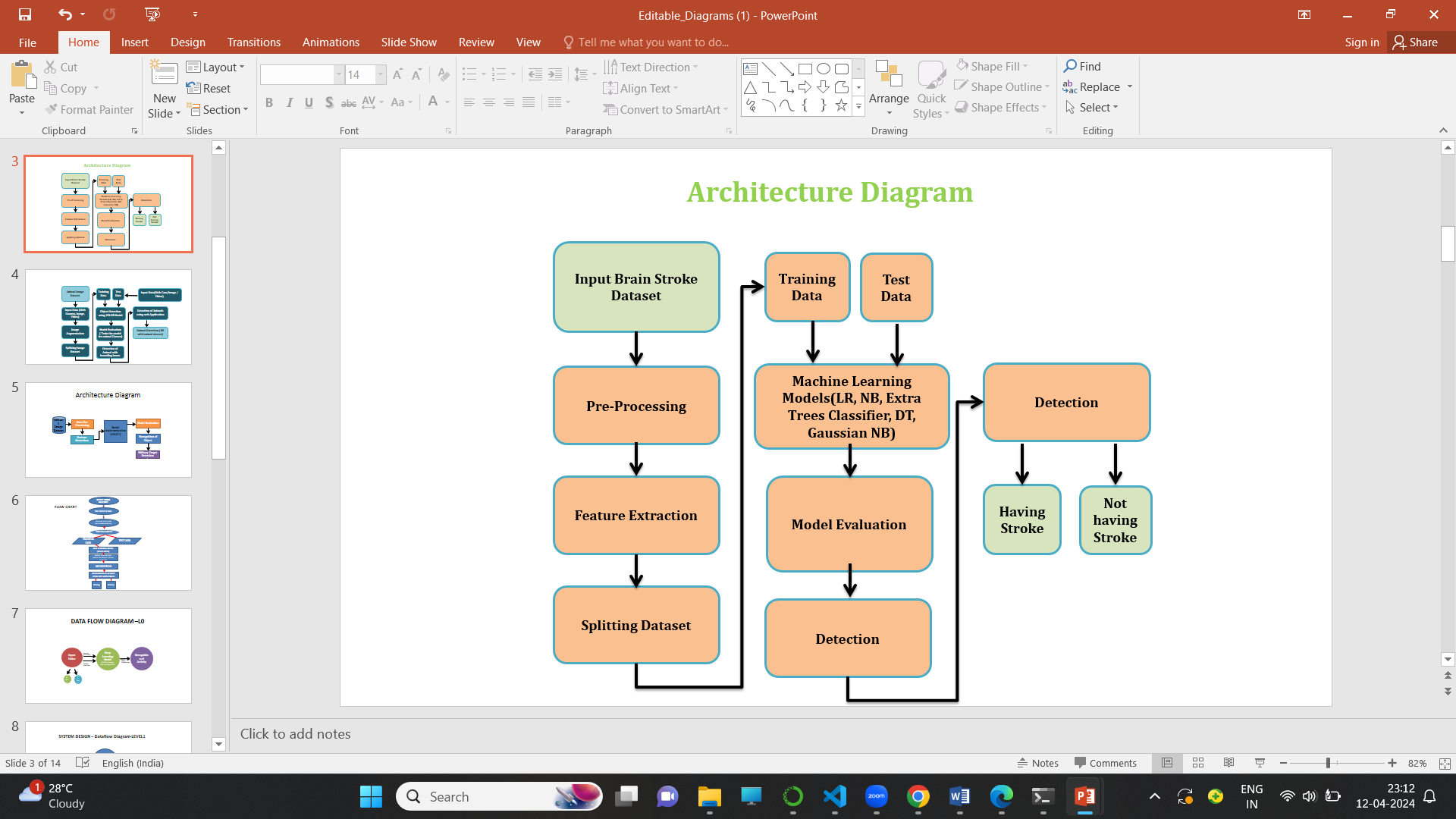


Figure 3.10: Architecture Diagram

This architecture outlines the key components and steps involved in building a brain stroke prediction system using machine learning. It starts with data collection and preprocessing, followed by model training and evaluation. Once the model is deployed, it provides real-time predictions and integrates with clinical workflows. Continuous monitoring and maintenance are essential to ensure the system's performance and compliance with ethical and regulatory standards.

**CHAPTER 4**

**APPROACH AND METHODOLOGY**

**4.1 Technologies to be used**

**4.1.1 Machine Learning**

Machine learning is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to learn and make predictions or decisions based on data without being explicitly programmed. It encompasses a broad range of techniques and approaches for analyzing and extracting patterns from data to perform tasks such as classification, regression, clustering, and reinforcement learning. Here's an overview of some commonly used models in machine learning

**Decision Tree**

A decision tree in machine learning is a versatile, interpretable algorithm used for predictive modelling. It structures decisions based on input data, making it suitable for both classification and regression tasks. This article delves into the components, terminologies, construction, and advantages of decision trees, exploring their applications and learning algorithms.

A decision tree is a type of supervised learning algorithm that is commonly used in machine learning to model and predict outcomes based on input data. It is a tree-like structure where each internal node tests on attribute, each branch corresponds to attribute value and each leaf node represents the final decision or prediction. The decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems.

**Logistic Regression**

* Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
* Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
* Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.
* In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
* The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
* Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.

**K-Nearest Neighbors (KNN)**

The K-Nearest Neighbors (KNN) algorithm is a supervised machine learning method employed to tackle classification and regression problems. Evelyn Fix and Joseph Hodges developed this algorithm in 1951, which was subsequently expanded by Thomas Cover. The article explores the fundamentals, workings, and implementation of the KNN algorithm. KNN is one of the most basic yet essential classification algorithms in machine learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining, and intrusion detection.

It is widely disposable in real-life scenarios since it is non-parametric, meaning it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data). We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.

**ExtraTrees Classifier**

ExtraTreesClassifier is an ensemble learning method fundamentally based on decision trees. ExtraTreesClassifier, like RandomForest, randomizes certain decisions and subsets of data to minimize over-learning from the data and overfitting. Extra Trees is like Random Forest, in that it builds multiple trees and splits nodes using random subsets of features, but with two key differences: it does not bootstrap observations (meaning it samples without replacement), and nodes are split on random splits, not best splits.

* 1. **Methodology**

1. **Data Collection and Preprocessing:**

* Gather diverse data sources relevant to stroke prediction, including electronic health records (EHRs), medical imaging (MRI, CT scans), laboratory results, lifestyle factors, genetic information, and wearable device data.
* Preprocess the data to handle missing values, outliers, and inconsistencies.
* Perform feature engineering to extract meaningful features from the raw data, such as demographics, medical history, physiological indicators, and imaging characteristics.

**2. Model Development:**

* Select appropriate machine learning algorithms for stroke prediction, considering factors such as the nature of the data, the complexity of the problem, and computational resources available.
* Split the data into training, validation, and test sets to evaluate model performance.
* Train machine learning models using the training data, tuning hyperparameters and optimizing model architecture as needed.
* Evaluate model performance using appropriate metrics such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), or precision-recall curve.

**3. Validation and Evaluation:**

* Validate the trained models using the validation dataset to assess their generalization ability and robustness.
* Conduct external validation using independent datasets or cross-validation techniques to further validate model performance.
* Evaluate the clinical utility of the predictive models, considering factors such as interpretability, ease of integration into clinical workflows, and potential impact on patient outcomes.

**4. Model Deployment and Integration:**

* Deploy the trained models into clinical workflows or decision support systems for real-time risk assessment.
* Integrate the predictive models with electronic health record systems or other healthcare IT infrastructure to enable seamless data exchange and interoperability.
* Provide user-friendly interfaces or dashboards for healthcare providers to interact with the predictive models and interpret the results effectively.

**4.3 System Implementation**

The purpose of System Implementation can be summarized as follows: making the new system available to a prepared set of users (the deployment), and positioning on-going support and maintenance of the system within the Performing Organization (the transition).

**Systems implementation** is the process of:

1. Defining how the information system should be built (i.e., physical system design),
2. Ensuring that the information system is operational and used,
3. Ensuring that the information system meets quality standard (i.e., quality assurance).

**The implementation can have classified into different modules of project and are listed as:**

1. Data Collection
2. Data Preprocessing
3. Feature Selection / Extraction
4. Model Training And Testing
5. Machine Learning Model Deployment
6. Performance Evaluation
7. Comparison Study of the Model
8. Exploratory Data Analysis (EDA)

4.3.1 **Code Implementation**

* **Importing the packages:**

When we implement the project, this is first step. Importing packages refers to the process of including external libraries or modules into your programming environment to gain access to their pre-written code and functionality. In most programming languages, including Python, packages provide a way to organize and distribute reusable code.

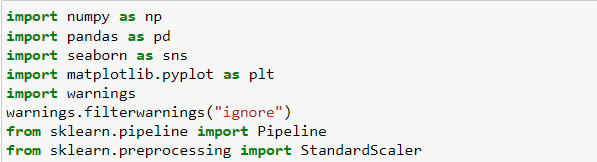


Figure 4.1: Importing Packages

* **Loading the dataset**

Once you've chosen your tools, you can load the dataset into your programming environment. For example, if you're using Python and your dataset is in CSV format, you can use Pandas to read the CSV file into a Data Frame.

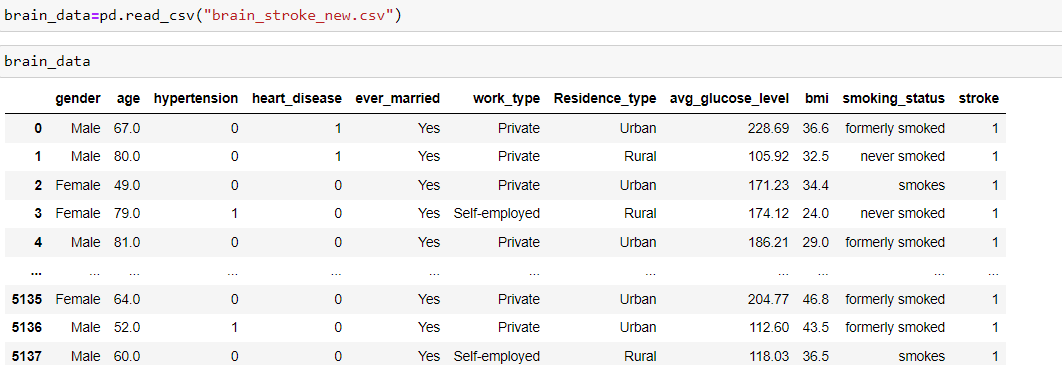


Figure 4.2: Loading Dataset

* **Data Pre-processing:**

Data preprocessing is a crucial step in the machine learning pipeline that involves cleaning and transforming raw data into a format suitable for training machine learning models. It aims to improve the quality of the data, mitigate issues such as missing values or outliers, and make the data compatible with the algorithms being used.

1. **Checking null values**

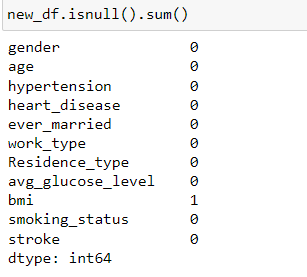


Figure 4.3: Checking Null Values

1. **Renaming the column**

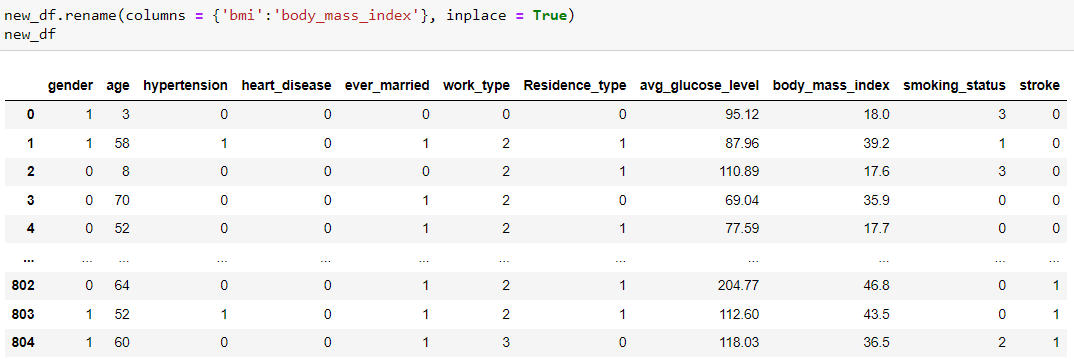


Figure 4.4: Renaming the Column

1. **Changing the datatype**

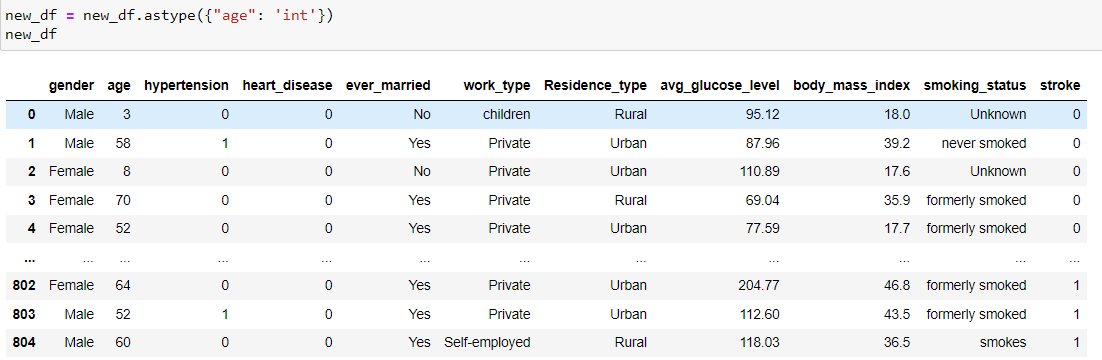


Figure 4.5: Changing the Datatype

* **Accuracy Score:**

The accuracy score is a common metric used to evaluate the performance of classification models in machine learning. It represents the proportion of correctly classified instances out of the total instances in the dataset. In other words, it measures how accurately the model predicts the class labels of the data.



Figure 4.6: Accuracy Score

* **Comparison of models**

Comparing models is a critical step in the machine learning workflow, helping you select the best-performing model for your particular task. In this project ExtraTree classifier gives more accuracy.

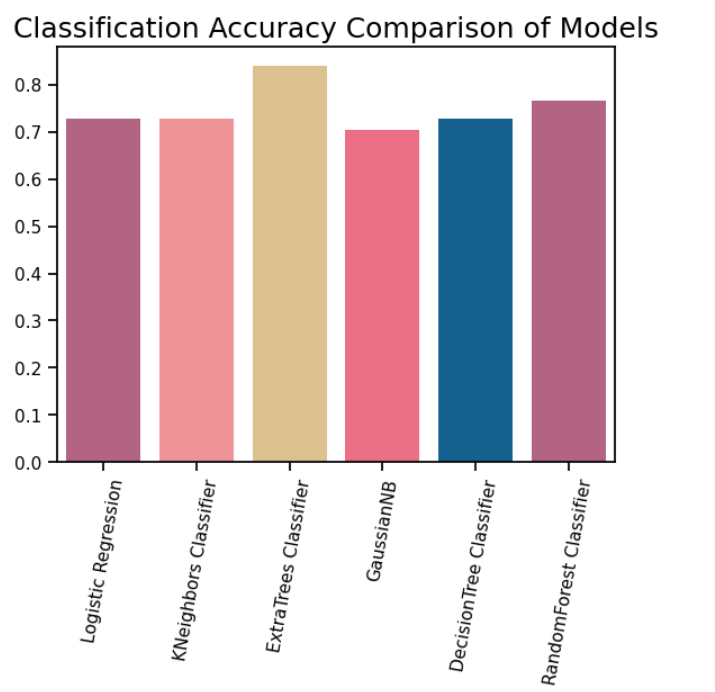


Figure 4.7: Comparison Models

* **Results**

1. **Brain Stroke - Positive**

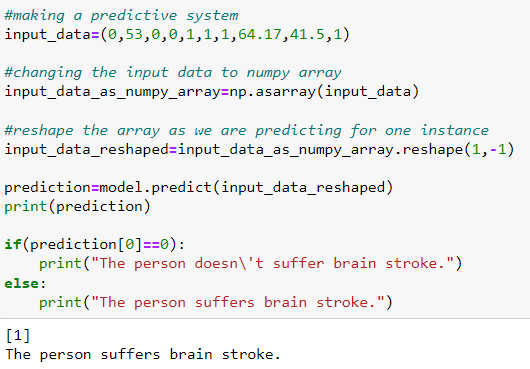


Figure 4.8: Positive Prediction

1. **Brain Stroke –Negative**

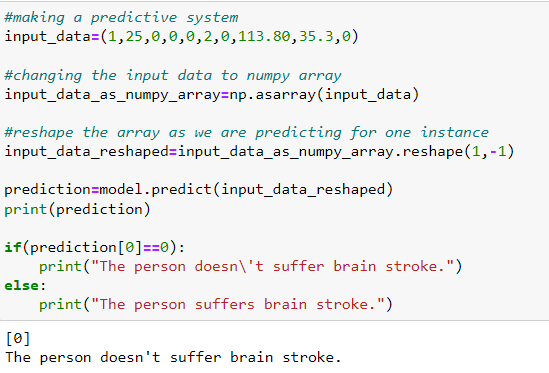


Figure 4.9: Negative Prediction

**CHAPTER 5**

**SYSTEM TESTING AND VALIDATION**

**5.1 Test Plan**

A test plan is a comprehensive document that outlines the objectives, approach, scope, and schedule of testing activities for a software project. It serves as a roadmap for the testing phase, providing guidance to the testing team on how to conduct testing activities and achieve the desired quality goals. The test plan is typically created during the early stages of the project and serves as a reference throughout the testing process.

A well-defined test plan typically includes the following components:

**Test Overview**

Provides an overview of the software project, its purpose, and the objectives of the testing effort. It outlines the scope of the testing, including the features or modules to be tested and any exclusion.

**Test Objectives**

Clearly defines the specific goals and objectives of the testing phase, such as ensuring functional correctness, validating system performance, or verifying compliance with specific standards or regulations.

**Test Approach**

Describes the overall strategy for testing, including the testing techniques, methodologies, and tools to be used. It defines the types of tests to be performed, such as functional testing, performance testing, security testing, or user acceptance testing.

**Test Environment**

Specifies the hardware, software, and network configurations required for testing. It includes information about the test environment setup, including test data, test tools, and any specific dependencies or prerequisites.

**Test Deliverables**

Identifies the documents or artifacts to be produced as part of the testing process, such as test cases, test scripts, test data, test reports, and any other relevant documentation.

**Test Schedule**

Outlines the timeline and sequence of testing activities, including milestones, deadlines, and dependencies on other project activities. It includes estimates for test preparation, execution, and reporting.

**Test Execution and Evaluation**

The testing and evaluation describes the procedures for executing tests, recording test results, and evaluating the outcome. It includes criteria for determining test completion, exit criteria for each test phase, and criteria for determining the success or failure of individual tests.

The test plan is a living document that may be updated and refined as the project progresses, based on changes in requirements, scope, or other project factors. It serves as a reference for all testing activities, ensuring a structured and systematic approach to achieve the desired quality standards.

**5.2 Test Approach**

A testing approach outlines the overall strategy and methodologies that will be employed to conduct testing activities for a software project. It defines the framework within which testing will be executed and helps ensure that the software meets the required quality standards. Here are some common testing approaches:

**Waterfall Testing Approach**

In the waterfall model, testing is typically performed in a sequential manner, following each phase of the development lifecycle. It starts with requirements verification, followed by system testing, integration testing, and finally, acceptance testing. Each phase must be completed before moving to the next one.

**Iterative Testing Approach**

In an iterative development model, testing is conducted in multiple iterations or sprints. Testing starts early in the development cycle and is performed continuously throughout each iteration. This approach allows for regular feedback and flexibility in adapting to changes, making it suitable for Agile methodologies like Scrum.

**Incremental Testing Approach**

Similar to the iterative approach, the incremental approach involves breaking the software into small functional increments. Each increment undergoes testing before being integrated with the existing system. This approach allows for early detection of defects and ensures that each increment is tested thoroughly.

**Parallel Testing Approach**

Parallel testing involves creating multiple test environments that mimic the production environment. Different versions or configurations of the software are tested simultaneously, allowing for a comparison of results and identification of discrepancies. This approach is useful for testing compatibility, performance, or migration scenarios.

**Exploratory Testing Approach**

Exploratory testing is a more informal and ad hoc approach where testers explore the software dynamically, without predefined test scripts or plans. Testers use their domain knowledge, intuition, and experience to identify defects and provide feedback. It is particularly effective for uncovering usability issues and validating user experience.

**Risk-based Testing Approach**

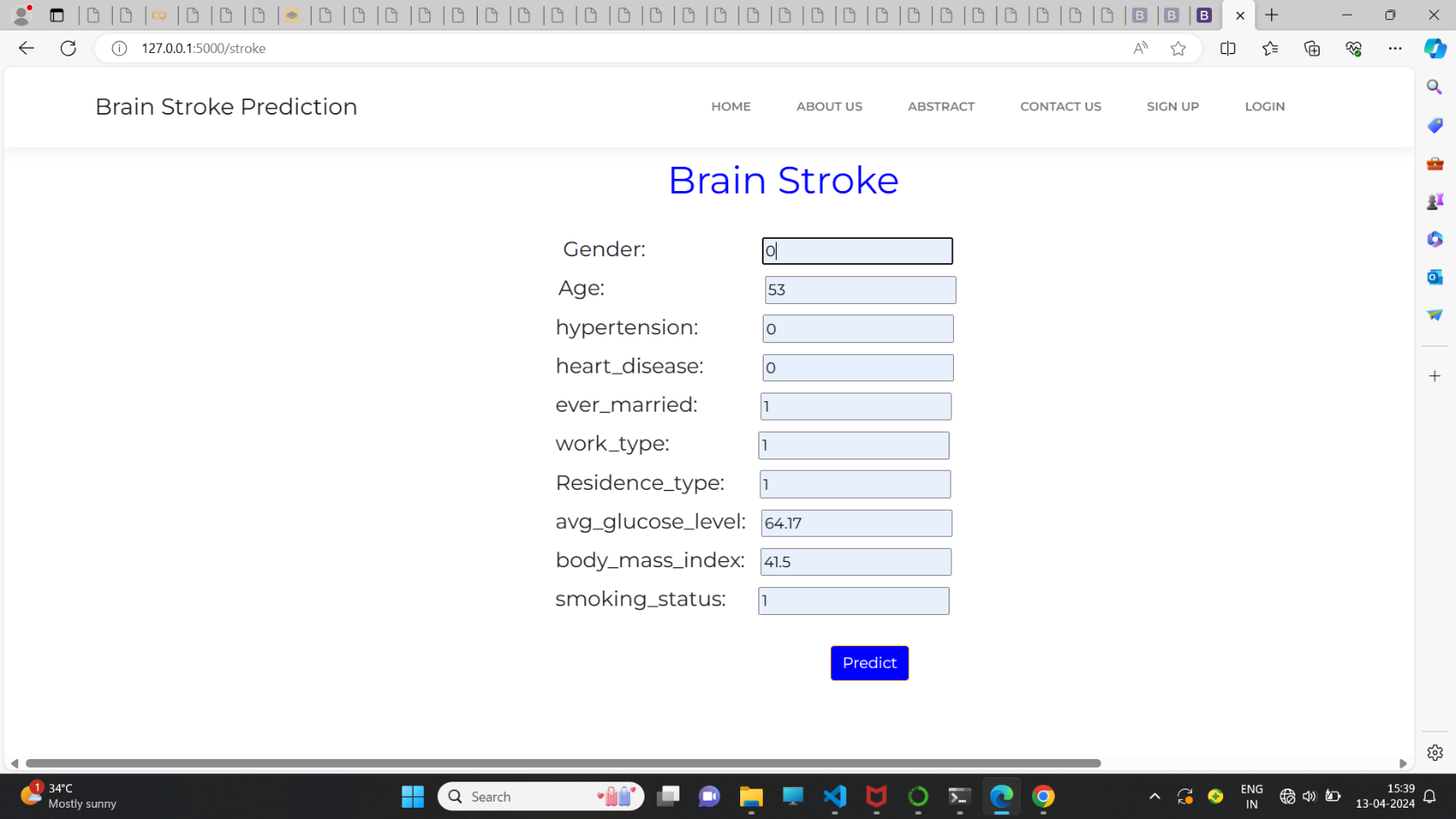
Risk-based testing focuses on identifying and prioritizing tests based on the risks associated with the software. Risks are assessed by considering factors such as the impact of a failure, the likelihood of occurrence, and the level of test coverage required. This approach ensures that testing efforts are focused on critical areas of the software.

**Regression Testing Approach**

Regression testing involves retesting the software after modifications or changes have been made to ensure that existing functionalities have not been affected. It aims to identify any unintended side effects or regressions caused by changes made during development. Automated testing tools are often used to perform repetitive regression tests efficiently.

The choice of a testing approach depends on various factors, including project requirements, development methodology, time constraints, and available resources. It is common to combine multiple approaches or tailor them to suit the specific needs of the project. The testing approach should be documented in the test plan and regularly reviewed and adjusted as the project progresses.

**Finding Results**



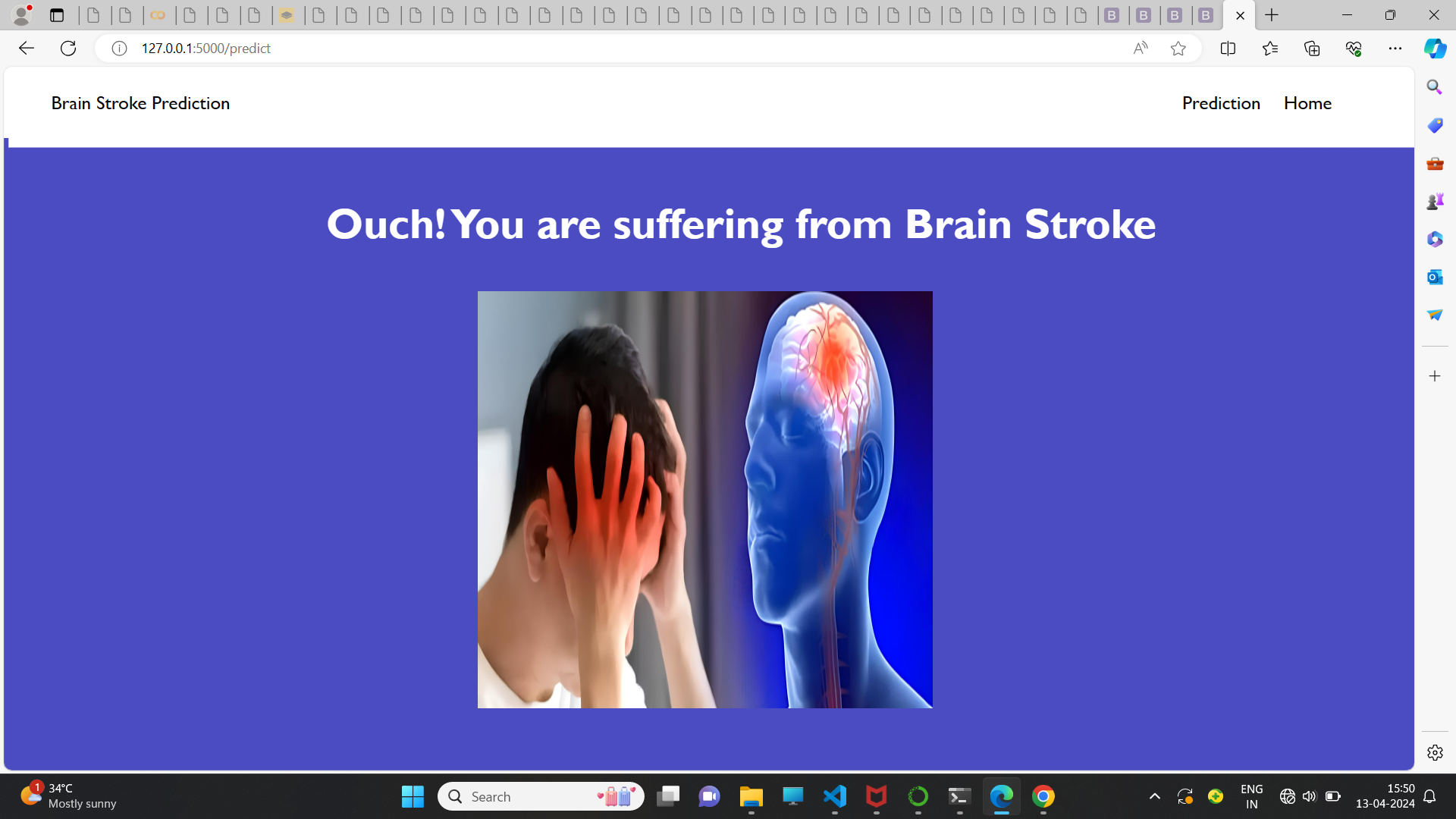


Figure 5.1: Positive Prediction

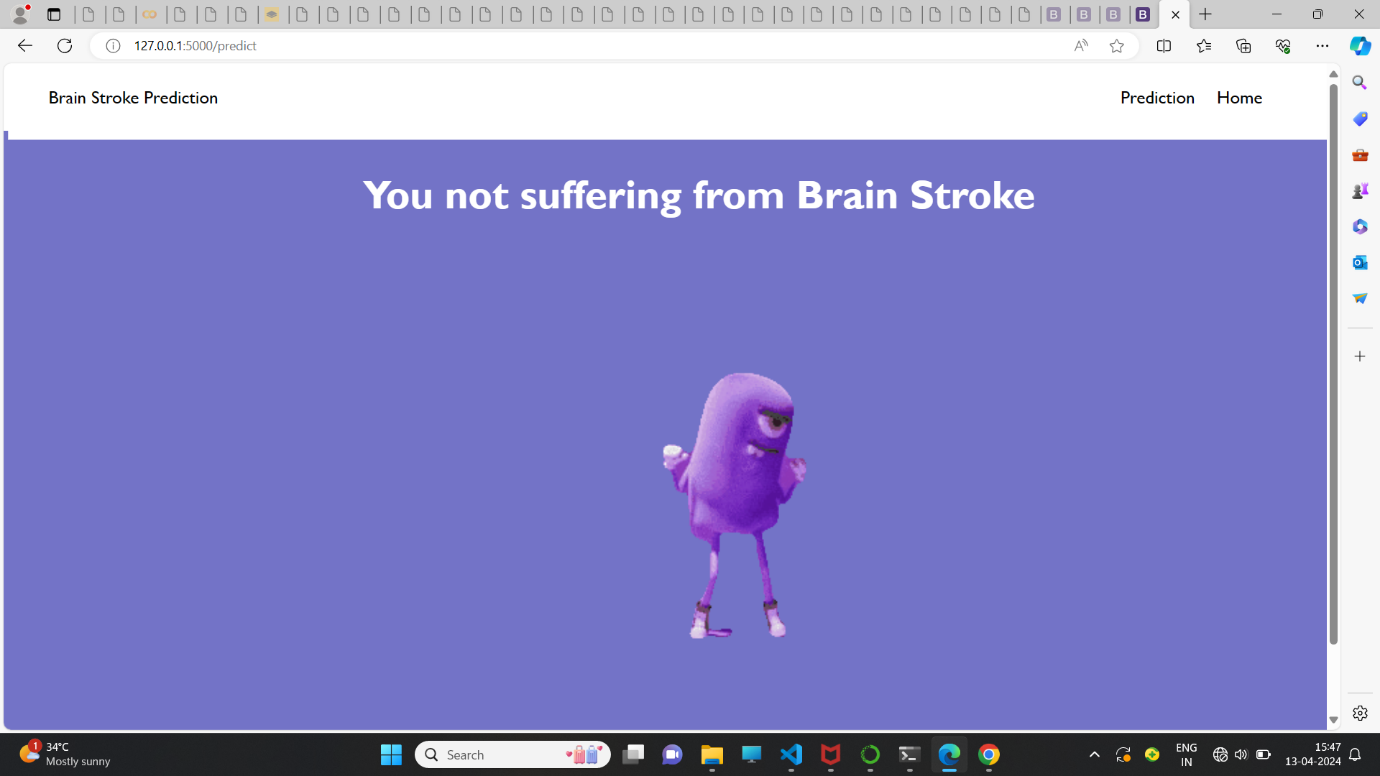
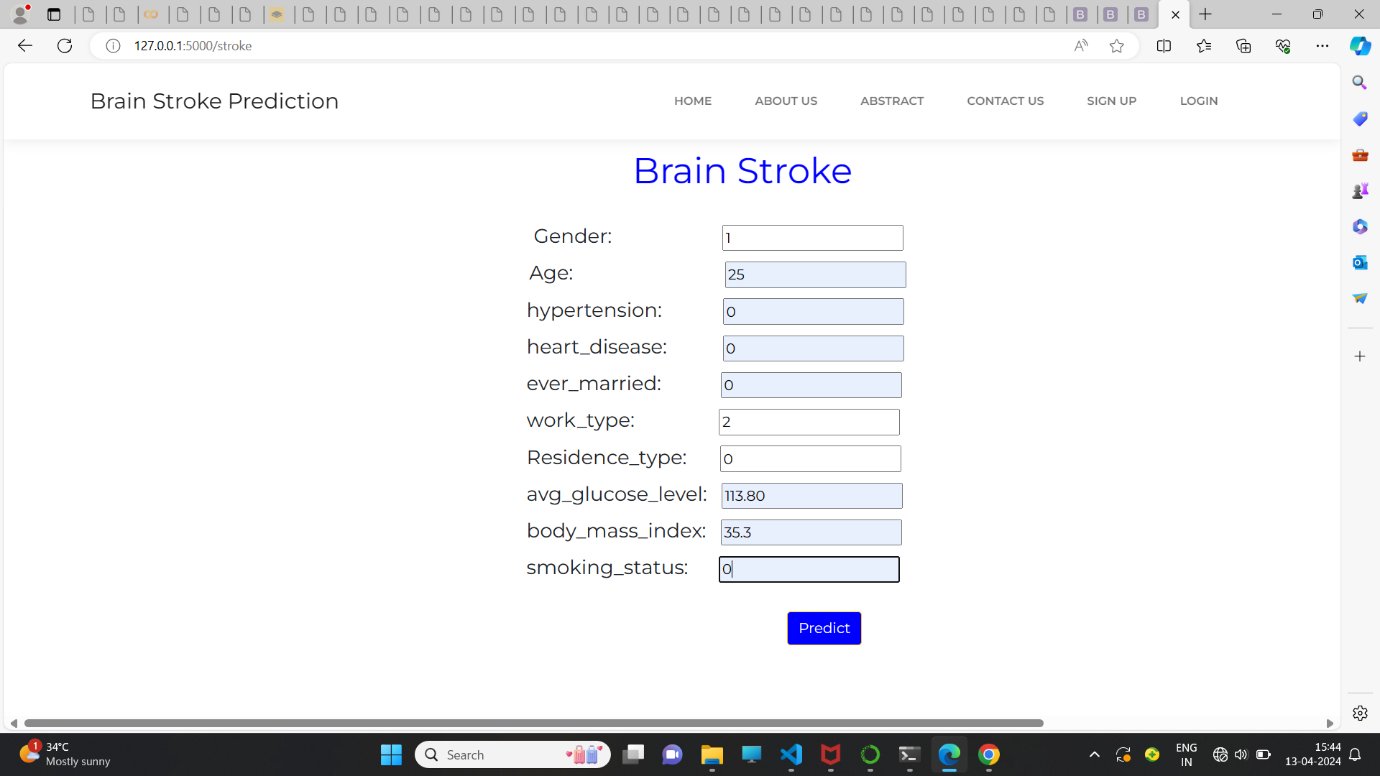


Figure 5.2: Negative Prediction

**CHAPTER 6**

**BUSINESS ASPECTS**

**6.1 Applications of the Project**

* **Early Intervention and Prevention:** Identifying individuals at high risk of stroke allows for early intervention and preventive measures. Healthcare providers can offer targeted interventions such as medication management, lifestyle modifications, and patient education to reduce modifiable risk factors and prevent stroke occurrence.
* **Improved Patient Outcomes:** Early detection of stroke risk enables healthcare providers to implement timely interventions, potentially reducing the severity of strokes and improving patient outcomes. By identifying high-risk individuals and providing appropriate care, the project can contribute to better long-term health outcomes for patients.
* **Resource Allocation Optimization:** Predictive models for stroke risk assessment help optimize resource allocation within healthcare systems. By identifying individuals at highest risk of stroke, healthcare providers can prioritize interventions and allocate resources more efficiently, reducing the overall burden on healthcare facilities and improving cost-effectiveness.
* **Personalized Healthcare:** The project facilitates personalized healthcare by tailoring preventive strategies to the individual characteristics and risk profiles of each patient. By considering a wide range of factors such as medical history, lifestyle choices, and physiological indicators, healthcare providers can offer targeted interventions that are more effective and relevant to each patient's needs.
* **Public Health Initiatives:** Stroke prediction efforts contribute to broader public health initiatives aimed at reducing the overall burden of stroke on society. By identifying population-level trends and risk factors associated with stroke occurrence, public health agencies can develop targeted interventions and health promotion campaigns to raise awareness, prevent strokes, and improve community health outcomes.
* **Clinical Decision Support:** Machine learning-based predictive models can serve as decision support tools for healthcare providers, assisting in risk assessment, treatment planning, and patient management
  1. **Features of the Project**
* **Data Collection:** Gather relevant data sources such as electronic health records (EHRs), medical imaging (MRI, CT scans), laboratory results, lifestyle factors, genetic information, and wearable device data.
* **Data Preprocessing:** Cleanse, preprocess, and standardize the collected data to handle missing values, outliers, and inconsistencies. This may involve data normalization, feature scaling, and handling of categorical variables.
* **Feature Engineering:** Extract informative features from the data that are relevant for predicting stroke risk. This could include demographic information, medical history, physiological indicators, imaging characteristics, and lifestyle factors.
* **Machine Learning Models**: Develop machine learning models to predict the likelihood of stroke occurrence based on the extracted features. Common algorithms include logistic regression, random forests, support vector machines, and neural networks.
* **Model Evaluation:** Assess the performance of the predictive models using appropriate evaluation metrics such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), or precision-recall curve.
* **Real-time Prediction:** Implement real-time prediction capabilities to assess stroke risk based on incoming data streams. Continuously monitor patients' risk profiles and provide timely alerts and recommendations for preventive interventions.
  1. **CONCLUSION**

brain stroke prediction holds significant promise in improving healthcare outcomes by enabling early detection, intervention, and prevention of stroke-related events. Through the utilization of diverse data sources, advanced machine learning techniques, and integration with clinical workflows, predictive models can accurately assess an individual's risk of experiencing a stroke.

Stroke is a potentially fatal medical condition that needs to be treated right away to prevent future consequences. The creation of a machine learning (ML)could help with stroke early diagnosis and subsequent reduction of its severe consequences. This study examines how well different machine learning (ML)predict stroke based on various biological factors. With a classification accuracy of 83%, ExtraTree classification exceeds the other investigated techniques. According to the study, the ExtraTree classifier performs better than other methods when forecasting brain strokes using cross-validation measures.

* 1. **Future Enhancements**
* **Integration of Multi-modal Data:** Incorporate diverse data types such as electronic health records (EHRs), medical imaging (MRI, CT scans), genetic information, lifestyle factors, and wearable device data. Utilize advanced techniques to handle and integrate heterogeneous data sources, capturing complementary information for more comprehensive risk assessment.
* **Advanced Feature Engineering:** Develop more sophisticated feature engineering techniques to extract informative features from the data. This may include the use of domain-specific knowledge, feature selection algorithms, dimensionality reduction techniques, or deep learning-based feature extraction methods to identify relevant patterns and relationships in the data.
* **Utilization of Deep Learning**: Leverage deep learning architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformer-based models to automatically learn hierarchical representations from raw data. Deep learning models can capture complex spatial and temporal dependencies in the data, leading to improved predictive performance.
* **Ensemble Learning:** Implement ensemble learning approaches such as model averaging, stacking, or boosting to combine predictions from multiple machine learning models. Ensemble methods can leverage the complementary strengths of different algorithms and mitigate the weaknesses of individual models, resulting in more robust and accurate predictions.

**REFERENCES**

* M. J. Ferdous and R. Shahriyar, "A Comparative Analysis for Stroke Risk Prediction Using Machine Learning Algorithms and Convolutional Neural Network Model," International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ECCE57851.2023.10101567, 2023.
* B. R. Gaidhani, R. R.Rajamenakshi, and S. Sonavane, "Brain Stroke Detection Using Convolutional Neural Network and Deep Learning Models,", 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 2019, pp. 242-249, doi: 10.1109/ICCT46177.2019.8969052, 2019.
* Gupta, Saumya & Raheja, Supriya, "Stroke Prediction using Machine Learning Methods". 553-558. 10.1109/Confluence52989.2022.9734197, 2022.
* T. I. Shoily, T. Islam, S. Jannat, S. A. Tanna, T. M. Alif and R. R. Ema, "Detection of Stroke Disease using Machine Learning Algorithms," 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944689, 2019.
* S. Kumar Satapathy, H. K. Kondaveeti and D. Parmar, "An Effective Framework for Predicting Stroke Prediction using Machine Learning Technique," Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2023, pp. 01-08, doi: 10.1109/ICECCT56650.2023.10179766, 2023.
* Jeena RS, Kumar S. Stroke prediction using SVM, International Conference on Control. Instrumentation, Communication and Computational Technologies (ICCICCT), 2016: 600–602
* Gary H, Gibbons L. National Heart, Lung and Blood Institute. 2022 [updated 2022 March 24]. Available from: https://www.nhlbi.nih.gov/health/stroke.

**SNAP SHOTS**

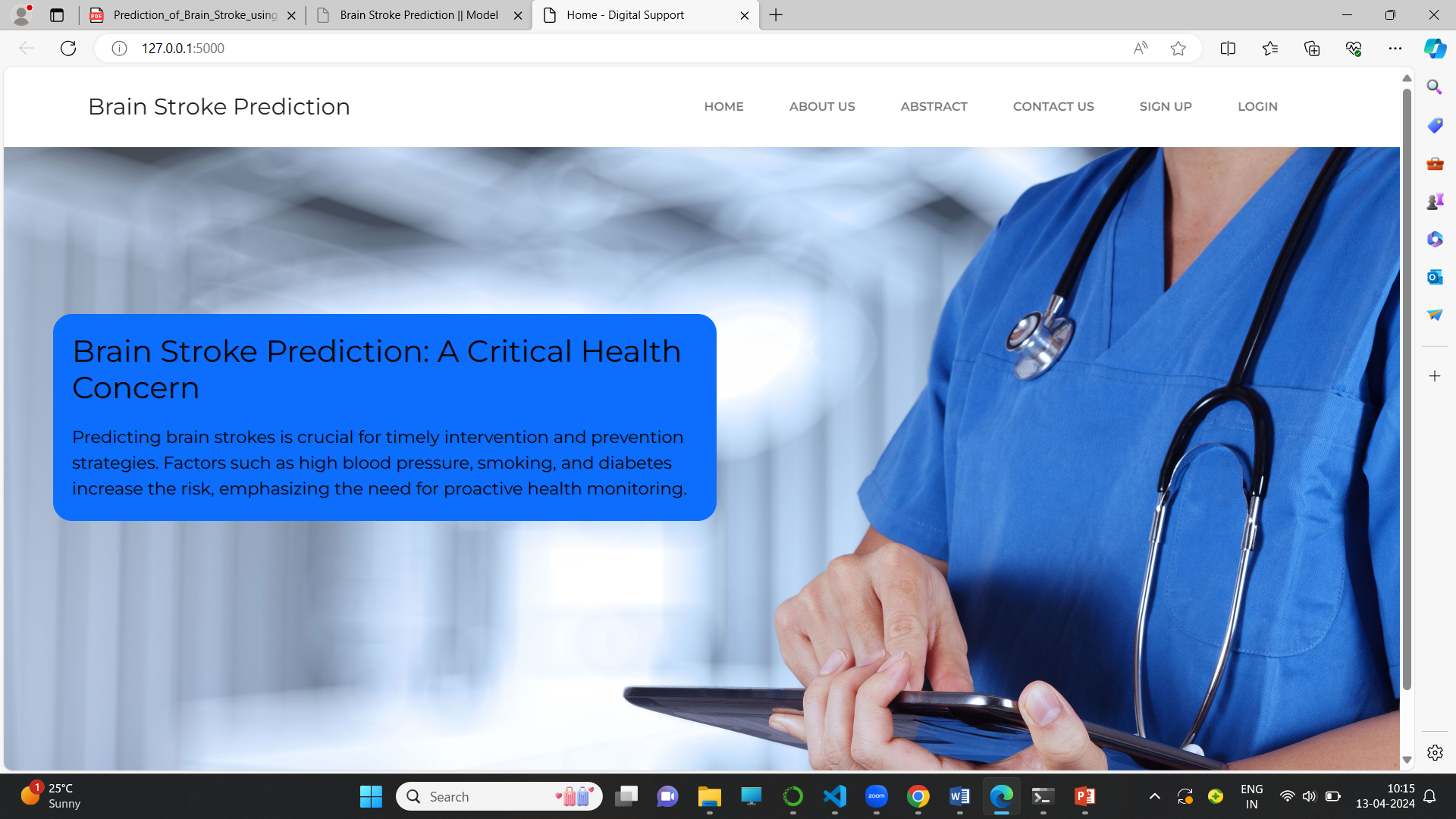


Figure: Home Page

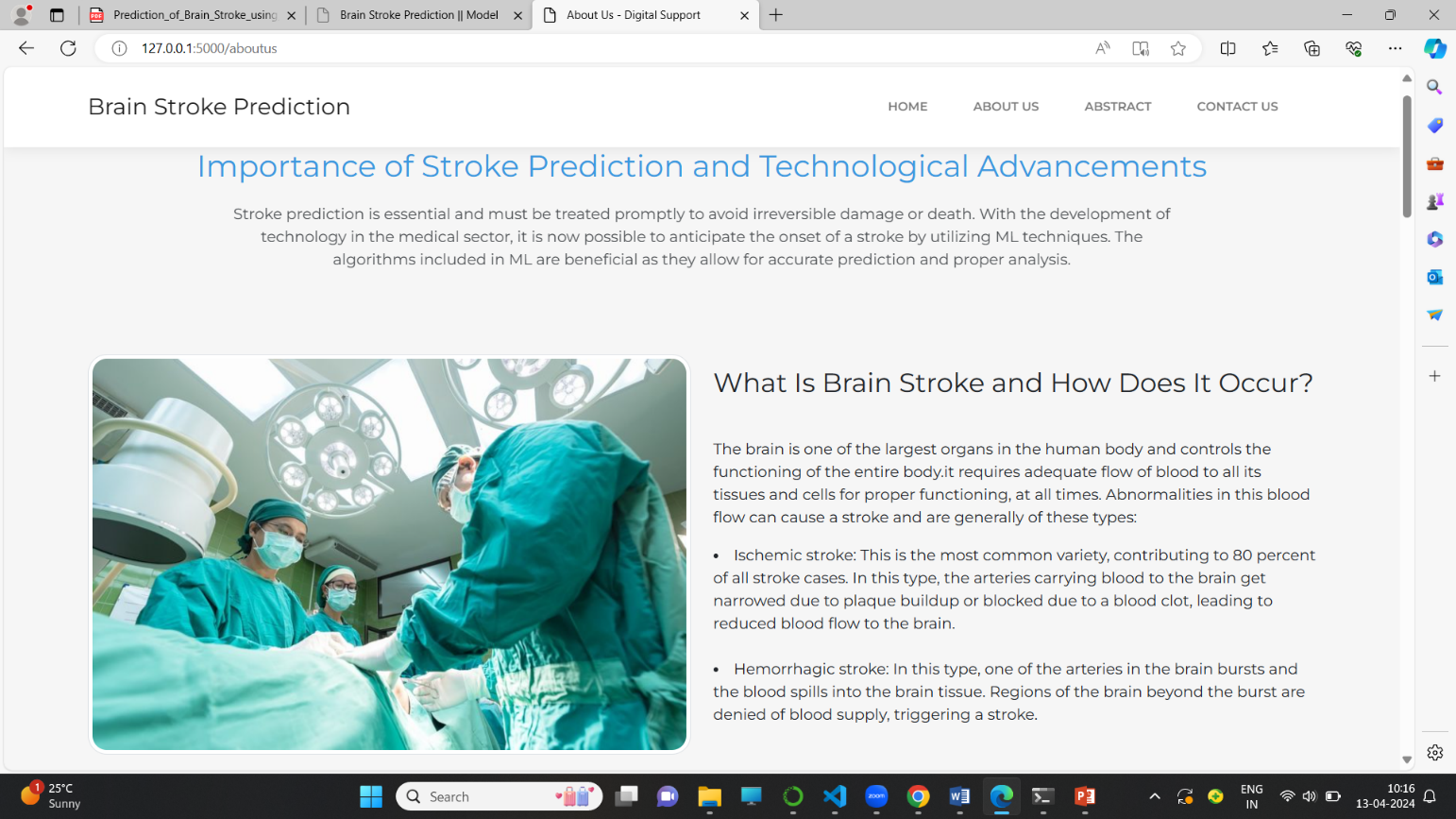


Figure: About Us Page

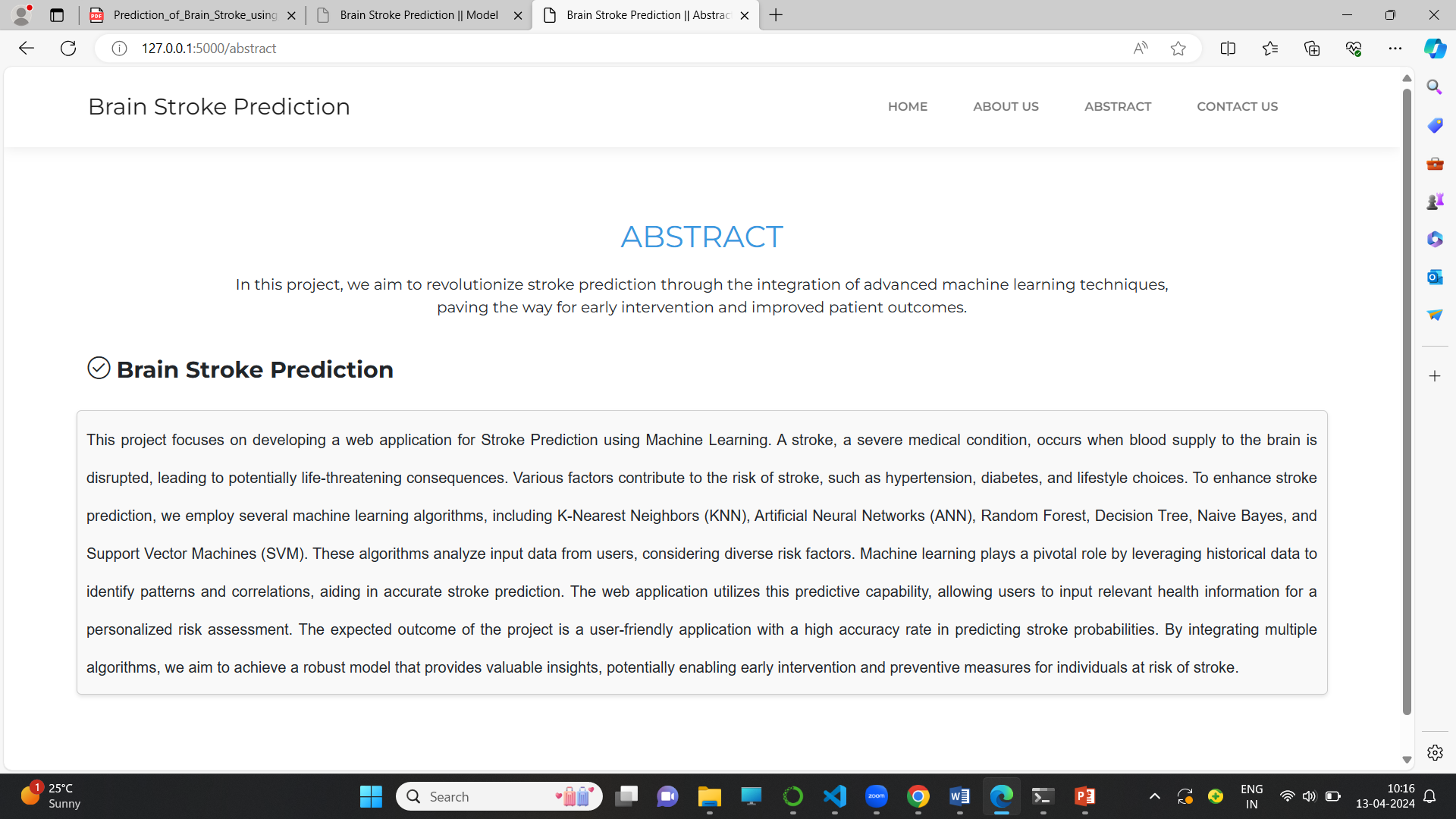


Figure: Abstract Page

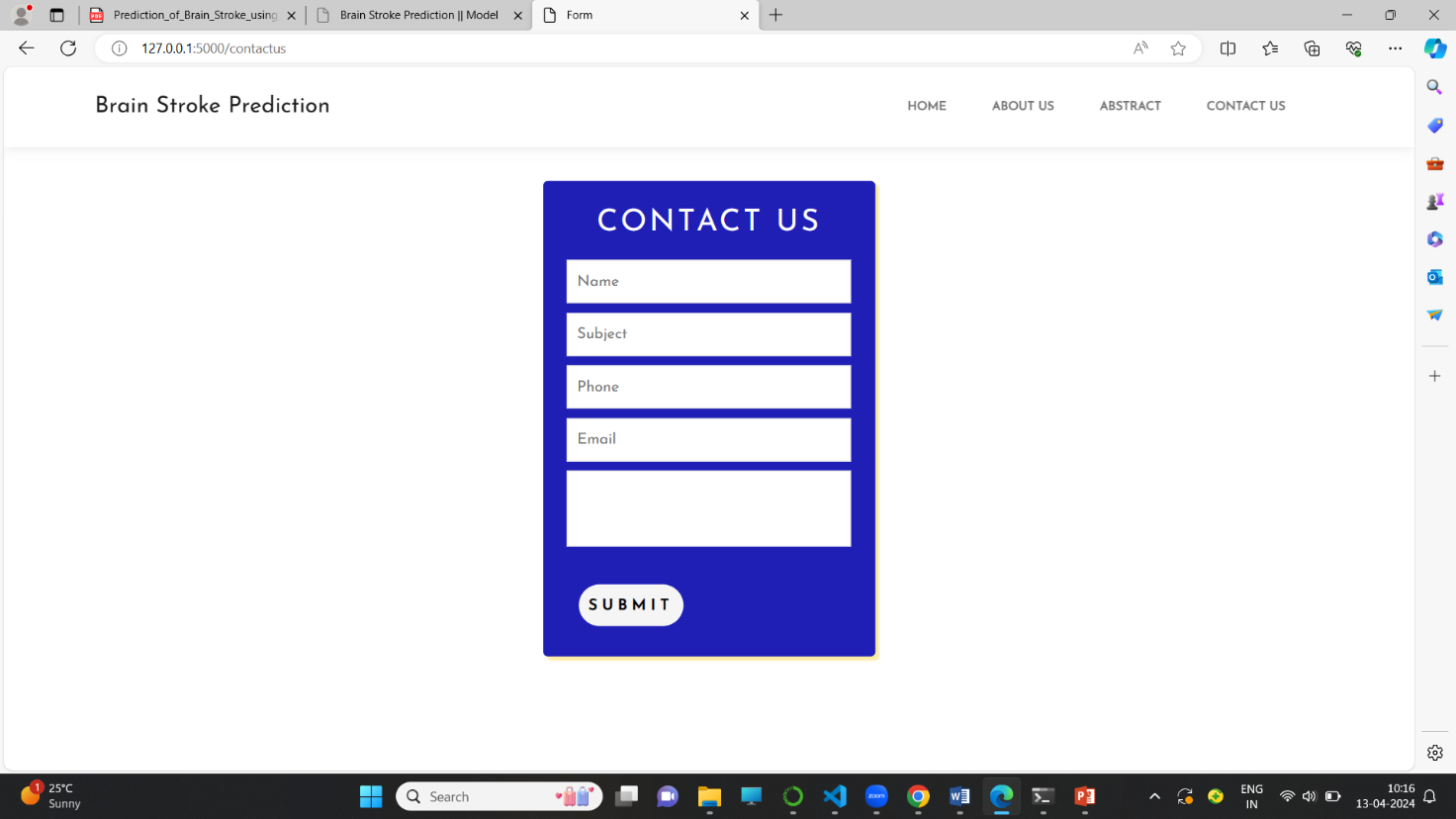


Figure: Contact Page

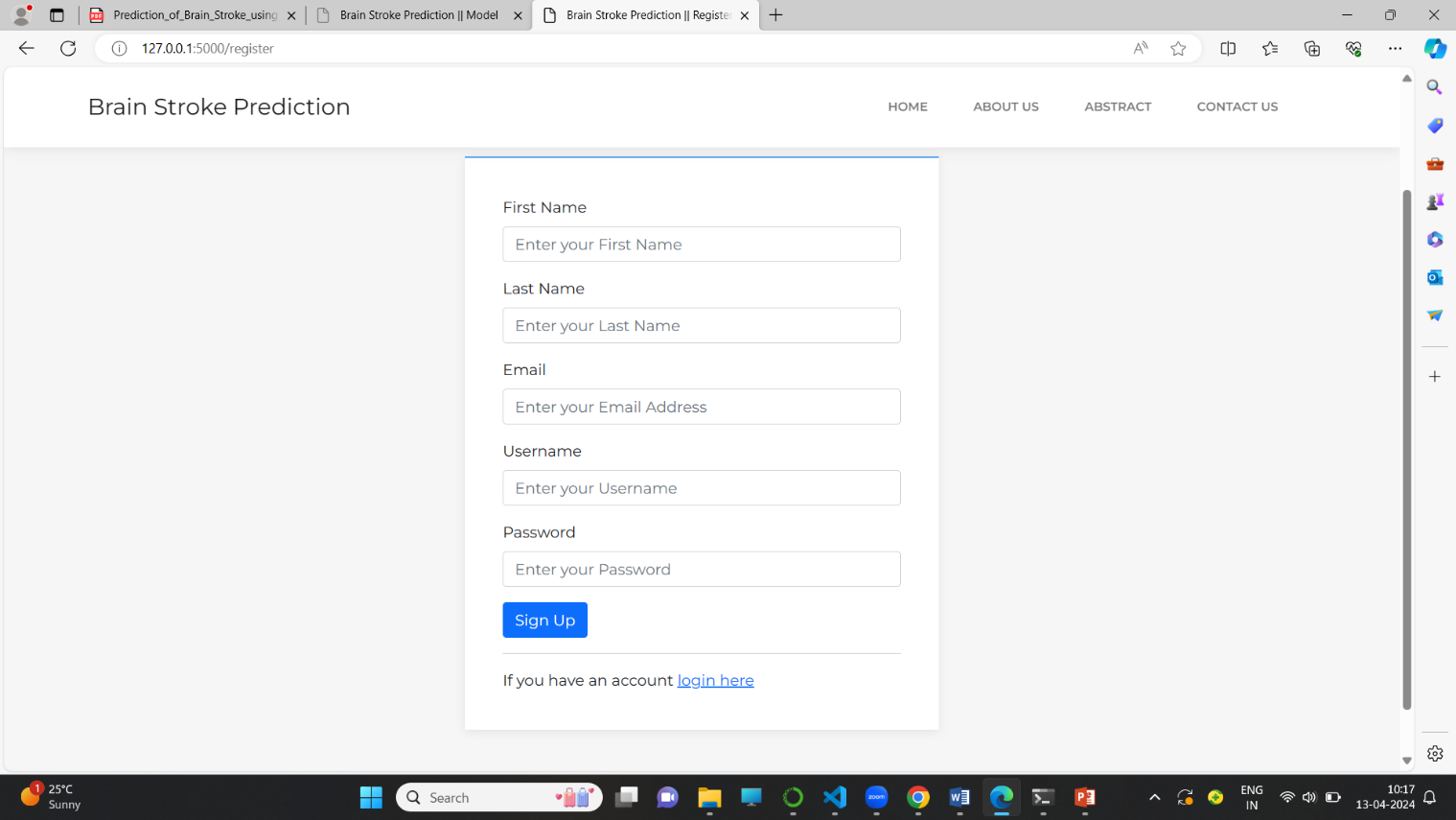


Figure: Register Page

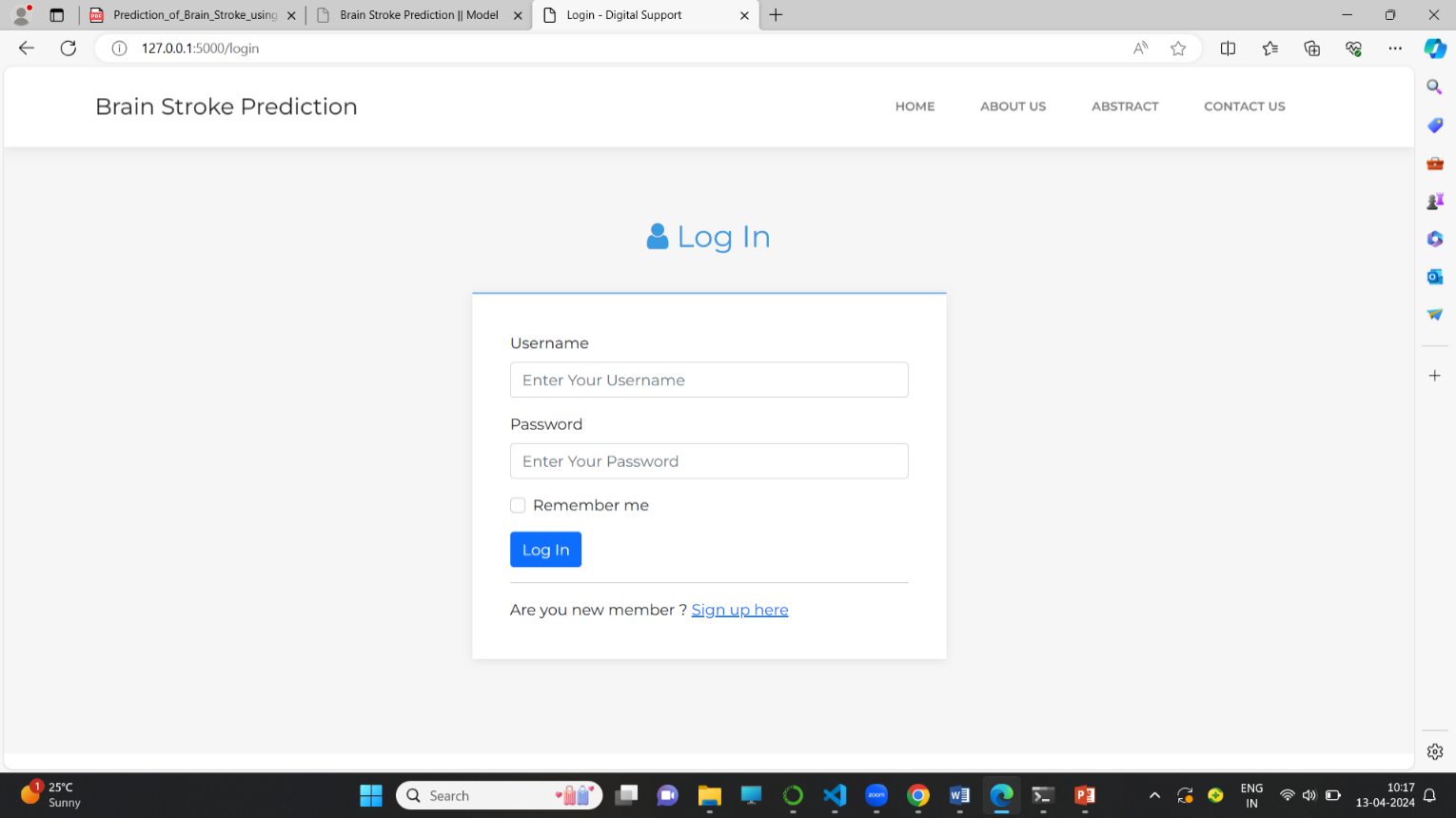


Figure: Login Page