
CHAPTER 1

INTRODUCTION

A stroke happens when the blood flow to different parts of the brain is disrupted or diminished, which causes the cells in one of these parts of the brain to stop getting the nutrients, vitamins, and oxygen they require and die. A stroke is a medical emergency that requires prompt attention from medical professionals. To avoid damage to the affected area of the brain as well as headaches in other parts of the body, early detection and appropriate care are essential. With the development of science, it may be feasible to predict the likelihood of having a stroke by using machine learning. Machine learning algorithms are confident in their ability to provide accurate assessments and make accurate predictions. During a stroke, the brain is deprived of oxygen and nutrients, which causes brain cells to stop functioning and eventually die. It can now result in death and is extremely expensive in terms of medical care and permanent disability. Data mining generally plays a significant role in the medical industry's ability to forecast ailments. The use of AI in medicine is heavily incorporated in this research. A device learning model may use patient data facts to suggest a variety of realistic expectations.

1.1 Problem Statement

Stroke is among the most serious life-threatening diseases for people over 65 and the second leading cause of death globally and the fourth leading cause of death in India. Similar to how a heart attack damages the heart, it damages the brain. When a stroke sickness manifests, it may cause mortality in addition to enormous medical expenses and long-term damage. Every four minutes, a stroke claims the life of a person. However, if we could identify or predict the incidence of stroke in its early stages, up to 80% of strokes might be prevented.

A stroke is a blood clot or brain haemorrhage that could permanently affect speech, mobility, cognition, or vision. A stroke may cause permanent brain damage, problems, and, in rare cases, death, making it a medical emergency. Strokes are often categorized as ischemic embolic or haemorrhagic. A blood clot that forms outside of the patient's brain, usually in the patient's heart, and travels through the circulatory system to lodge in smaller brain arteries is known as an ischemic embolic stroke.

Another type of brain stroke known as haemorrhagic stroke happens when an artery in the brain spills or bursts. Strokes happen suddenly, but many of the illnesses they cause take time to manifest. Since of this, age is the most evident risk factor for stroke. Although stroke can happen at any age, it is more likely as we become older because the possibility of a blockage or fracture grows with each passing year. The profile includes information on age, systolic blood pressure, BMI, cholesterol, diabetes, smoking status and intensity, physical activity, alcohol use, history (hypertension, coronary heart disease), and family history (stroke, coronary heart ailment).

The goal of developing "Early Stroke Identification" is to enhance people's knowledge of strokes. People are often ignorant of the many types of strokes, How Stroke Happens. Nowadays, many with hectic schedules neglect their health, resulting in serious health consequences. People who live in cities take care of their health by eating good foods, exercising, and working daily, making them less susceptible to illness than those who do not.

The programme focused on the current system and attempted to fix Stroke-related concerns. There are several causes of stroke, including drinking, smoking, lifestyle, work pressure, and a low nutrient count in the body. Stroke may also be caused by a family history of stroke. An ischemic stroke, also known as a cerebral infarction, is the most frequent form of stroke. This is caused by a decrease in the flow of oxygen and nutrients to the brain. In this application, users can check their chances of having a stroke in a matter of minutes by providing basic information about their daily routine, such as whether they smoke, drink, and their BMI (body mass index) in the system, and the results will be displayed in terms of percentage based on the Machine Learning Algorithms (Neurologist). The application includes information on doctors and their contact information so that users may call them in an emergency.

The proposed system is designed so that anybody may utilise the interface. It is a simple interface that everyone may use without difficulty. Keeping in mind that the system is used by both urban and rural users. When the user submits his information, the analysis is performed using the Machine Learning algorithm, which forecasts the likelihood of Stroke based on the proportion of Barriers. Were users now the odds of having a stroke, in case of an emergency, the details of a Specialised Doctor have been supplied in the system where they may set an appointment and consult as soon as possible in case of an emergency.

Stroke is the second highest cause of mortality globally, and it continues to be a significant health burden for individuals and national healthcare systems. Hypertension, heart illness, diabetes, dysregulation of glucose metabolism, atrial fibrillation, and lifestyle variables are all potentially modifiable risk factors for stroke. As a result, the research intends to apply machine learning ideas to massive current data sets in order to accurately predict stroke based on possibly modifiable risk factors. The application would then be developed to provide a tailored warning based on each user's degree of stroke risk as well as a lifestyle corrective message regarding the stroke risk factors.

The programme may be used commercially by hospitals and doctors on their websites and Android apps to deliver services connected to stroke illnesses, allowing doctors to diagnose and treat patients. The information gathered from users may be utilised for research purposes, such as analysing the data gathered from users and determining the causes of the Stroke. The research conducted may assist forecast the early detection of stroke, which can save millions of lives. Doctors or hospitals may warn consumers before it's too late and take the necessary procedures to treat the sickness.

The programme is designed for adults in their forties and fifties, since strokes are more common in these age groups. This programme is made user-friendly by considering the types of users who would use it. The programme provides additional advantages to users since it predicts the early detection of stroke, which may be beneficial to users because they can take action quickly and save their lives. Even if the individual is experiencing typical symptoms, the stroke usually happens unexpectedly. As a result, it is preferable to see a doctor at regular intervals and apply this sort of application.

The "Early Identification of Stroke" application is new to the business since no work has been done in this sector in comparison to other risk analysis applications such as Heart, Cancer, and so on. The only application available is Stroke Risk meter, an Android application that requires installation and requires the user to log in to obtain the results. There are few applications available on the internet due to a lack of research in the field of stroke and a lack of awareness about the disease. As of today, there are no reliable online applications that make accurate predictions. The "Early Detection of Stroke" application distinguishes out from the competition due to its simple user interface and the contact information for physicians in case of an emergency.

The software is written in Python 3.7. Many more tools, including as NumPy, Pandas, Matplotlib, and Seaborn, are required to construct an application in Python. The project's goal is to forecast using Machine Learning Algorithms such as Logistic Regression, Decision Tree Classification, Random Forest Classification, and Nave Bayes. The prediction may be done for the best output using this method, which Nave byes has demonstrated to be the best algorithm for Stroke Prediction.

The Flask Framework is used to create the Front-End Application, which is a python code bridge between the web page and the learned machine learning. HTML code is used to create the web page. This application includes an input form that uses the user's input data to forecast the occurrence of strokes. CSS and Bootstrap are used to improve the user experience. The project is hosted on the Heroku platform. The back-end database is connected via SQL light.

1.2 Objectives

The primary goal of this research is to build a prediction model for stroke using machine learning methods. To get a better understanding of the data, categorical features, numerical features, and multicollinearity analysis will be performed. SVM, Decision Tree, Random Forest, K-nearest neighbour, and Logistic regression are among the models investigated. We can anticipate a patient's stroke using this approach by examining their symptoms such as age, gender, blood pressure, and so on. Finally, a higher performing algorithm for stroke prediction is selected, and a basic Graphical User Interface is constructed using Flask.

1.3 Scope of Early Identification of Stroke

In order to estimate the risk of stroke in senior people and people who are addicted to the risk factors mentioned in the project, this study applies a prediction model. The same project might be developed in the future to deliver the stroke percent based on the outcomes of the current study. By obtaining pertinent risk problem information and speaking with experts, this research may also be used to identify stroke risks in young and underage people. We can predict a patient's stroke using this technology and start the right treatment before it's too late. Numerous people are impacted by this condition. By anticipating strokes, we can avoid brain damage. By employing neural networks to train the model, this project may be further developed. Using more precise metrics will allow for more accurate performance comparisons. This study only uses textual data, which may or may not be reliable for

predicting strokes. In the future, it would be more effective to predict the likelihood of stroke by gathering a collection of images, such as brain CT scans. This experiment shows how well several machine learning algorithms predict stroke based on a variety of physiological factors.

- Identify customers' demands and necessities and develop technologies to meet those needs.
- Believe in discovering new opportunities that optimize corporate value, consistent growth, and sustainability.
- Allow clients to adapt to changing market dynamics more quickly and intuitively.
- Allow to become nimbler and more competitive by using new technology.
- Assist clients in bringing exceptional goods to market in much less time and at a much lower cost. Collaborate with our customers on their ongoing success.

1.4 Company Profile IndataAi

Software Solutions profile IndataAi Software Solutions was Established in the year 2019. The works in the domain of expert website expansion and mobile application company carrying practice web development cellular software, e-trade software, internet design, HTML, iPhone/iOS/android apps, and search engine marketing services to our worldwide clients. The company as is clear from the lots of enterprise sites on the net, internet sites are available in a change of outlines and scopes. An internet site may be custom-constructed by an internet

development organization to outfit the unique requirements and options that a corporate has. Distinct from the websites established over the past span, these times websites are created by taking a mobile-first style. This means evolving your internet site for slighter devices after which adjusting it to contact on a computer. We develop custom Android apps tailored to the unique needs and targets of our customers. We have a group of talented Android developers who construct strong Android apps with the use of flexible programming languages like Kotlin and Java. As a main Android app development organization in Bangalore, we provide exceptional Android app improvement services throughout all-important enterprise verticals. We deployed excessive-quality and smoothly applied Android apps for many reputed customers all around the world.

The billing software automates the billing and invoicing process. It guarantees well-timed collection of payments, tracks a couple of invoices, flags any unpaid invoices, and

decreases manual records entry errors. The billing and invoice software program is utilized in numerous industries including utilities, telecommunications, and professional offerings companies which include legal and medical billing services. Due to the variety of industries served, maximum companies provide customizable systems that could support numerous business models, pricing structures, features also can range significantly based on the corporation the software program is meant to serve.

1.4.1 Vision and Mission of InDataAi

Vision

The vision is the most important international internet expansion organization. Focused on consistent innovation is the key to completing the remaining aim of fulfilment and rising as a worldwide identified organization through presenting superior quality services and solutions. They keep on with the subsequent values in turning in vision.

- DataAi exploring new chances that maximize the business value, reliable development, and sustainability using essential abilities
- Conference expectancies of our clients, employees, and associates
- The business enterprise in which openness, trust, cooperation, simplicity, and invention are valued and helped
- Worldwide identified for offering excellent facilities to its clients globally while following essential ideals.
- Practical IT Solutions to the valued international clients. Persist with the following ideas in handing over our mission:
- Recognize the needs and necessities of clients and make technology to solve the needs.
- Believe to discover new possibilities that maximize the business value, steady growth, and sustainability.
- Empower customers to respond quicker and more intuitively to converting marketplace dynamics
- Enable them to become extra agile and competitive through leveraging new technologies
- Help customers to bring great products to the marketplace in much less time and at much less cost Collaborate with our customers on their continued success.

Mission

Being an expert web, cellular s application, and virtual advertising company, task is to offer client-centric, outcome-focused, s cost-viable advanced & practical IT Solutions to valued clients.

- Recognize the requirements and necessities of clients and make technology to solve the requirements.
- Believe to discover new possibilities that make the most of the professional value, steady development, and sustainability.
- Authorize customers to respond quicker and more spontaneously to converting marketplace subtleties
- Enable them to become extra agile and competitive finished leveraging new technologies

1.4.2 Projects/Product completed/under development by InDataAi

Domains of Projects handled by the company

- i. Cinewinks ERP
- ii. Insur Hubli (LIC)
- iii. Dhanushri Oil Mills (CRM)
- iv. Computation Book Stall (Dharwad)
- v. Doddamani Valmiki Matrimony (Hubli)
- vi. APMC Byadgi Cinewinks

ERP Enterprise Resource Planning (ERP)

Description: Device consists of various sets of software programs that can be used to integrate the business features in a company or organization. Given that those solutions provide an included way to the needs of an organization, those systems are in excessive demand through all organizations.

Insure Hubli (LIC)

Description: The occasion is s about the numerous changes that occurred in the Indian Life Insurance sector d after its privatization. Till privatization, the Life Insurance Corporation of India (LIC) f was the only corporation provided that life insurance services d in India. LIC sold

its policies as tax tools and not as a product offering protection against risk. Most of the purchasers were under-insured with flexibility or transparency within the amenities provided.

Dhanush Oil Mills (CRM)

Description: Customer Relationship Management tools can assist you in handling all advertising and marketing strategies and improve your capacity to acquire new clients and keep the existing ones.

Summary

In the above, have discussed the introduction of the application Stroke Prediction where the application briefly discussed the depth of the work. Description of application explains the work in detail on each module and the above section focuses on details internship company profile and the project which is handled by the company details explained and working domain these all points covers in the above section.

Literature Survey provides an overview of related work and also the related tools used in the previous work and existing system. The tools in this technological era are upgraded and users need a more interactive user interface and experience.

System Requirement Specification will provide an overview of every specification needed and it is used for the use. This System Requirement Specification will have all the technical aspects and perspectives of stakeholders.

The design will be describing how the application is designed. And how the interaction is happening between the different modules of the application. This design chapter is very helpful for the development phase of the application. The major concept of the implementation is to provide a good overview and easy usage of the application being developed. The application developed is very easy to use and it should feel good while using this application.

The Software testing will provide the behavior of the application. In this sense, the application must work according to the user requirement, and it should meet the user input without any error or any other bugs. The application must work according to the behavior of both the student and staff requirements.

Future Enhancement: involves the additional features and the up-gradation of the system that can be done in the future. That enhances the user experience.

References: Lists the websites and references referred to during the project work

CHAPTER 2

LITERATURE SURVEY

Many research has recently been conducted to identify strokes. Two distinct research topics are the classification of stroke based on a medical diagnosis and the classification of stroke based on facial movements. This section's goal is to evaluate the earlier studies on the subject. Two common methods of stroke detection are the use of tools like wristbands to monitor sleeping habits or image processing:

Accuracy has been determined for each of these signals individually with 400-800 strokes per 100,000 individuals, 15 million new acute strokes yearly, 28,500,000 disability-adjusted life years, and 28-30-day case mortality rates ranging from 17 to 35 percent, stroke is the second greatest cause of death and the primary cause of adult disability globally. With the number of fatalities from heart disease and stroke predicted to climb to five million in 2020 from three million in 1998, the burden of stroke is likely to become worse. This is owing to the continuous demographic and health changes that will lead to an increase in the risk factors for vascular disease and the senior population. 85 percent of stroke fatalities globally occur in underdeveloped countries. Stroke has severe social and economic implications. In the United States of America (USA), the cost of stroke was anticipated to be as high as \$49.4 billion in 2002, whilst expenditures upon discharge were expected to equal 2.9 billion Euros in France.

Researchers have identified stroke using image processing skills. To analyse 69 4 distinct photos of facial motions, researchers compared the facial movements of stroke patients with those of healthy people. The Naive Bayes, Support Vector Machine, and Random Forest machine learning models used in this technique all had accuracy levels that were close to 95%.

In contrast, researchers used the Support Vector Classifier to conduct a comparable inquiry on

32 patient MRI scans and properly (88 percent) identify a stroke. [1]

Instead of utilizing images to predict the incidence of strokes, researchers used hardware devices to collect characteristics that are primarily focused on a person's sleeping patterns. To predict, researchers used wristbands to collect information on 44 people—14 stroke patients and 30 healthy people—including indicators like sleep frequency and intensity.

SVM and K-Nearest Neighbors (KNN) models were used in this strategy, and SVM outperformed KNN in terms of relative prediction accuracy. According to a related experiment conducted utilizing the sleep data of 225 people, the Support Vector Machine (SVM) outperforms models like the

Neural Network (NN), Random Forest, and Naive Bayes on the assessment metrics of False Negative Value and True Positive Value. This research sought to divide user stroke propensities into five groups and calculate the 10-year stroke prediction probability. [2]

Researchers were able to predict the chance of someone suffering a stroke by analysing 357 unique traits of 1800 persons. These participants were randomly selected from the general population (of which 200 were stroke patients). After completing feature selection with the help of Decision Tree and dimensionality reduction with the help of Principle Component Analysis, the Artificial Neural Network 5 model was finally put into action. The accuracy of this model was 95 percent, making it effective in identifying patients who were at risk of having a stroke. According to the findings of this research, a head CT scan without contrast is the gold standard for performing the first imaging test on patients who have symptoms of stroke or head trauma. During this research, a variety of deep learning algorithms for automated detection were conceived of, created, and tested. The disadvantage of using this method is that it takes a significant amount of time to compile all 357 attributes. This research project and article aimed to construct a model equation to develop a pre-diagnosis algorithm for strokes by making use of risk factors that may be within the patient's control. [3]

To predict stroke, researchers looked at several psychological variables as well as the Gene Diagnostic Disease database. Neural Network outperformed models such as Decision Tree and Bayesian Classifier when applied to the 1000 entries of the Gene Diagnostic Disease database with an accuracy of 97 percent. Support Vector Machine (SVM) with a linear kernel was able to predict stroke using psychological parameters from about 350 samples with an accuracy of 90 percent; however, Neural Network outperformed SVM with an accuracy of 97 percent. In this research, the method of principal component analysis was used to minimize the dimensions, define the aspects that are most essential for predicting stroke disease, and determine whether the patient has the sickness or not [4].

Researchers used data from a total of 250 and 120 people, respectively, to forecast the acute phase of a stroke as well as the risk that follows it (such as ultimate mortality). According to the study reported in Popukaylo, Random Forest and XGBoost outperformed other models including Support Vector Machine (SVM), Decision Tree, and Logistic Regression, achieving an accuracy of almost 90%. (2019). Contrarily, Random Forest was shown to be superior to all other models in the study by Garcia-Terriza et al. (2019), which examined 6 distinct performance metrics. These other models included Support Vector Machine, Naive Bayes, Logistic Regression, KNN, Decision Tree, and Neural Network. In this case, the dimension is reduced using the principal component analysis method, the features are chosen using the decision tree approach, and the classification model is built using the back propagation neural network methodology. [5]

The previously described algorithms have had their scaling issues resolved; nevertheless, because of the limited size of the dataset, there is still a risk of overfitting occurring. In addition, these methodologies do not take into account the societal elements that can play a role in determining the incidence of the stroke while they are making their predictions. As a result of the fact that datasets reflecting the actual world are not always balanced, minority classes may be regarded as oddities in severely imbalanced datasets. In addition, the problem of stroke detection is generally tackled with the use of models of machine learning. Machine learning techniques are utilized in situations involving the identification of anomalies that occur across several domain Researchers used models like Support Vector Machine (SVM) and Artificial Neural Network to interpret mammography pictures (ANN). On a relative scale of F1 score, ANN demonstrated a 10% improvement over SVM in recognizing microcalcification clusters, which are crucial in the early detection of stroke. The researchers that participated in the study examined a highly unbalanced dataset while taking faults as anomalies that were injected offline using an experiment known as fault injection to pinpoint problems with the cloud architecture. The results showed that a variety of models, including Support Vector Machine (SVM), Naive Bayes, Decision Tree, Logistic Regression, and Random Forest, could detect anomalies with an average F1 - score of 91%. Similar to this, several categorization models to predict the monsoon based on measures of recall and positive predictive value, including Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors, Decision Tree, and Artificial Neural Network. and found that the score generated by Random Forest was just right—neither too high nor too low. Over two million NBA players' shots from the 2014–2015 season were

examined by researchers. They were successful in predicting the shots with an accuracy rate of 68 and 57 percent, respectively, using models like XGBoost and Random Forest. [6]

Study to evaluate the efficacy of automated early ischemic stroke detection systems. Their study's main goal was to develop a technique for automating the detection of primary ischemic stroke using convolutional neural networks (CNN). To train and evaluate the CNN model, they collected 256 images. They were able to raise the caliber of the final picture generated by their system by using the data lengthening approach. Their CNN approach has an accuracy rate of 90%, as measured by CNN. Through their research, Sung and colleagues created a stroke severity measure. There was a total of 3577 people who had an acute ischemic stroke, and data was gathered on them. They used several different data mining methods, one of which was linear regression, to create the models that they used for their predictions. They were more accurate with their predictions than the k-nearest neighbour method was (95 percent confidence interval). Using machine learning, were able to provide an accurate prediction about the functional outcome of an ischemic stroke. They used this strategy with a patient who sadly lost away three months after being admitted to the hospital. They determined that the AUC value was more than 90, researched to determine the risk of having a stroke. The researchers evaluated the data and made their predictions using several statistical methods, including Naive Bayes, decision trees, and neural networks. During their research, they examined the pointer's accuracy as well as its area under the curve (AUC). All of these algorithms were put together in the form of decision trees, and the naive Bayes method was shown to provide the best accurate results. Adam and his co-workers researched to classify what constitutes an ischemic stroke. They identified ischemic strokes by using the decision tree methodology and the k-nearest neighbor method. Medical specialists determined, throughout their study on the classification of strokes, that the decision tree strategy was the most successful method. [7]

It was speculated that the majority of research had an accuracy rate of about 90 percent, which is a very respectable figure. Our investigation stands out from others because, to get the best possible results, we used a variety of well-established methods of machine learning. Random forest (RF), decision tree (DT), the voting classifier (VC), and logistic regression each have relative F1 ratings of 96, 94, 91, and 87 percent, making them the most successful algorithms (LR). The models that were utilized in this investigation had an accuracy % that was much greater than the accuracy percentage of the models that were

used in prior studies; this suggests that these models are more trustworthy. They perform admirably in a variety of model comparisons, and the conclusions of the study might be used to design the plan. As was said earlier, the primary contribution made by this work is the application of multiple different machine learning models to a dataset that is available to the public without charge. The majority of the researchers in the earlier investigations used a considerable model to forecast the stroke disease, used support vector machines equipped with the appropriate kernel functions to do stroke analysis. Following the removal of conflicting and duplicate data during the pre-processing stage, the prediction was made using a total of 350 inputs. MATLAB was used to do the analysis, and the results showed an accuracy of 91 percent. According to Chutima Jalayondeja, the three models that were taken into account for the prediction utilizing demographic data were Decision Tree, Nave Bayes, and Neural Network. It was discovered that the Decision Tree had the best accuracy and the lowest false positive (FP) rate. A low false-positive rate indicates a high accuracy in predicting if the patient had a stroke but it was not a stroke. On the other hand, a high false-negative rate predicts that patients did not have a stroke when they did have a stroke. FN is dangerous since it may end in mortality for the patient because the patient has a stroke, even though the test findings reflect the opposite. In terms of precision, the Decision Tree was taken into account; nonetheless, the Neural Network was selected as the best option [8]. This was owing to the Neural Network's high FP value and low FN value.

Studies came up with a novel automated feature selection method that selects robust features based on a recommended heuristic called the conservative mean. In addition to this, they hypothesized a predictive analysis based on the Bayesian model and referred to as Bayesian Rule Lists (BRL). They merged it with SVM support vector machines to enhance the area that was contained under the ROC Curve (AUC). For stroke prediction, they used the dataset from the Cardiovascular Health Study (CHS) and compared the cox proportional hazards model to a machine learning approach. In addition to this, they devised a technique known as margin based censored regression, which combines censored regression with margin-based classifiers to provide a concordance index that is more accurate than that produced by the cox model. Both the concordance index and the area under the curve (AUC) were areas in which our method demonstrated superiority over the current state of the art. When there are many gaps in the available data and just a few risk factors are fully understood, this technique may be used to clinically anticipate a variety of diseases. However, they discovered that because their approach for selecting features

evaluated the performance of each feature independently, it may not perform well in other datasets with characteristics that are tightly connected. To solve this issue, they used a technique called conservative mean features selection, followed by an algorithm called L1 regularised feature selection, to prune the features. This allowed them to generate a distribution across permutations based on a massive collection of data that had been processed. The pre-processed data reduce the model space for the various groups of fragments, which enables the approach to scale with a smaller quantity of data while maintaining a sizeable feature set. With the assistance of the BRL technique, one can achieve a high level of accuracy, precision, and tractability and colleagues developed a multimodal analytic approach and used it to physiological data to predict strokes. Photoplethysmography, arterial blood pressure, and electrocardiography (ECG) were some of the measures. They incorporated all three signals and asserted that using a multi-model analysis delivers a better accuracy for stroke prediction [9].

In the past, a large number of academics have made stroke predictions using machine learning. Text mining and a machine learning classifier were the two methods that. used to classify the stroke conditions of 507 individuals. For training, they investigated a variety of machine learning methods, one of which being Artificial Neural Networks (ANN). They concluded that the SGD algorithm produced the greatest result, which was 95 percent. The research was carried out by Amini and colleagues to determine how often strokes occur. They classified 50 risk factors for stroke, diabetes, cardiovascular disease, smoking, hyperlipidaemia, and alcohol consumption based on the results of their study of 807 persons who were either healthy or sick. They used the K-nearest neighbor algorithm in conjunction with the c4.5 decision tree method. Both of these approaches had an accuracy rate of 95%. (94 percent accuracy). Cheng et al. presented their findings from a study that attempted to forecast the results of an ischemic stroke. In their investigation, they made use of two ANN models, 82 patient data sets, with accuracy values ranging from 79 to 95 percent researched to see whether or not the death rate of a stroke patient could be predicted. The incidence of stroke was determined based on the data from 15,099 participants who participated in the research. To identify strokes, they used a method based on deep neural networks. The authors used PCA so that they could extract data from the medical records and make predictions about strokes. The area under the curve that they have is 83 percent of the total (AUC). In the research that Singh and colleagues [10] conducted, they predicted strokes using artificial intelligence. In the research that they did

use the information from the cardiovascular health study (CHS), they made use of a brand-new technique to predict strokes. In addition to that, they used the method of a decision tree to carry out a feature extraction, which was then followed by a principal component analysis. In this particular scenario, a neural network classification technique was employed to construct the model, and the results achieved an accuracy rate of 97 percent. [10]

2.1 Existing Systems

Before the patient may undergo the procedure known as decompressive hemicraniectomy, it is necessary to ascertain the extent of the stroke-affected area, which is also often referred to as the infarction volume. A system was able to accurately predict the infarction volume growth rate using just a single CT scan and a few clinical data points. This was the only thing that was missing. The estimate of the infarction volume at a certain point in time is the primary problem that is brought up by this. As a result of the integration of Principal Component Analysis (PCA) and an Adaptive Neuro-Fuzzy 0 percent Plagiarized 100 percent Unique Inference System, it has been shown that the approach that is advocated is superior in terms of its ability to accurately forecast the infarction volume. This is because the strategy that has been advised is an altogether new one (ANFIS). The Fuzzy Inference System and the Artificial Neural Network (ANN) are both components of a method that is collectively referred to as the Adaptive Neuro Fuzzy Inference System (ANFIS) (FIS). This approach is used to analyse infarction development patterns in stroke patients who had a substantial arterial obstruction in their anterior circulation. With the information gained from this examination, an estimate of the infarction growth rate and the infarction volume at a certain time may then be made. Principal Component Analysis is a method that may be used to reduce the dimensionality of a dataset. This can be accomplished in several ways (PCA), which does so with the least amount of information that is lost in the process.

The approach that was recommended leverages feature reduction to anticipate the growth rate of the second infarction from the smaller data set, which ultimately produces better results. It is erroneous and inefficient at the same time. The objective was to make a decision on medical treatment, which is a highly specialized work that may be quite challenging due to the number of factors that need to be taken into account, especially in the case of unusual illnesses or diseases that have similar symptoms. On the dataset of the Cardiovascular Health Study (CHS), multiple strategies for predicting stroke are compared with the methodology. The use of artificial intelligence (AI) in the medical field is an

important topic. An artificial intelligence system would analyse the patient's data to generate a set of appropriate predictions. In this case, the strategy of principal component analysis is used to minimize the dimension, and the decision tree algorithm is used to determine the characteristics that are most relevant for the prediction of stroke sickness. Both of these methods are described in more detail below. additional usages backward transmission by using a neural network classification strategy, a classification model may be constructed. The neural network found in the brain served as inspiration for the development of a primitive kind of electronic neural network known as a neural network for classification. Neural networks give a higher level of accuracy in classification than other methods, such as decision trees and naive Bayes. The CHS dataset has a large number of attributes, and there are also a large number of missing values; as a result, it may be extremely challenging to properly use this information.

2.2 Proposed System

After being processed, the data are now in a state where they may be used in the construction of models. Methods of machine learning and a dataset that has already been pre-processed are necessities for the building of the model. LR, classification, classification, and Natives byes are some of the methods that are used in this process. The accuracy measures of accuracy score, precision score, recall score, and F1 score are used to assess and contrast the four different models that were created. It is necessary to reprocess the data in order to eliminate undesirable noise and outliers from the dataset before creating a model. These elements could lead to the model deviating from the instruction it was expected to receive. All the problems that impede the model from operating more effectively have been fixed at this point. Before the modelling process can start, the data must first be cleansed and categorised after the relevant information has been gathered. The dataset that was used had a total of twelve distinct features, as was previously mentioned. To start, the column id is left out because including it would have no effect on how the model is built. The dataset is then scrutinised for any occurrences of null values, and any found are then filled in. In this instance, the average of the numbers in the data column is used to fill in the null values in the BMI column.

The textual literals in the dataset are converted into easily digestible integer values via the process of label encoding. Since the computer is often trained on numbers, the strings need to be converted to integers first. The data obtained is organized using five columns that are of the string data type. During the label encoding process, each text is encoded, which transforms the whole dataset into a list of numbers. The dataset that is

being utilized to predict strokes is quite unevenly distributed. The dataset contains 5110 rows, 249 of which suggest the possibility of a stroke and 4861 of which provide evidence that suggests there was not a stroke. It is possible to obtain accuracy by using such data to train a machine-level model; however, other accuracy measures like precision and recall are not sufficient to achieve accuracy. If such unequal data is not handled in the right manner, the findings will be wrong, and the forecast will be of no use. As a result, addressing this unequal data is the first step that has to be taken to develop an efficient model.

The most common sickness that science is aware of is a stroke, and its incidence is growing at an unsustainable rate. Using the dataset for stroke prediction that is accessible to the public, the study evaluated four machine learning methods that are often used for predicting the likelihood of a subsequent brain stroke. These approaches are as follows In the context of a random forest, a decision tree Regression modelling,

Random Forest. The RF classification technique was selected as the optimal classification strategy. RFs are composed of a large number of autonomous decision trees, each of which was trained independently using a different subset of the provided data. During the training process, the outputs of the decision trees are collected as the trees themselves are produced. The outcome of a process that is known as voting is what ultimately determines the forecast that is made by this algorithm. Within the context of this operation, each DT is needed to cast a vote for one of the two output classes (in this case, stroke or no stroke). The RF approach, which chooses the category that received the most votes overall, is responsible for making the result prediction.

The Decision Tree Classification by use of DT is utilized to solve problems relating to both regression and classification. Additionally, since each of the variables used as input already has a matching variable serving as output, this method may be classified as a supervised learning model. It looks like a tree. During this process, the data are continually segmented into different categories according to a set of parameters. The decision node and the leaf node make up the two different elements that make up a decision tree. At the first node, the data are separated into two groups, and at the second node, the result is generated.

Logistic Regression The supervised learning approach makes extensive use of LR, making it one of the most used machine learning algorithms. It is a method of forecasting that makes use of several independent variables to make projections about a categorical dependent variable. Logistic regression is a statistical method that may predict the outcome of an experiment using a categorical dependent variable. Therefore, either discrete or

categorical output is required. The offered probability values range from 0 to 1, although they may be true or false, yes or no, 0 or 1, etc. Both logistic and linear regression has a wide range of potential applications. On the other hand, linear regression is used to address problems with regression, while linear reclassification is used to handle classification issues.

2.3 Feasibility Study

Subsequently, after carrying out the project and thoroughly investigating all essential or pre-existing structural components, the next responsibility is to conduct a probability study for the creation of the product. Even if time is limited and the necessary materials are missing, it is still possible to do tasks. In the part devoted to feasibility, every feasible approach to resolving the problem at hand is taken into consideration. The proposed solution must not only fulfil all of the operator requirements, but it also has to be flexible enough to easily accommodate any modifications that may be necessary for the future.

2.3.1 Economical Feasibility

This is an essential aspect that must be taken into consideration while developing a project. The technology was chosen because it could be produced for the least amount of money. All charges for hardware and software must get approval from the organization. In general, anticipate that the association would generate more money from the proposed organization than it will spend initially and, in the future, to administer the system. This is in contrast to our previous prediction, which said that the association would lose money.

2.3.2 Technical Feasibility

This involved teaching individuals about the capabilities of the system, its purpose, and any possible limits that it may have. During this proof of concept, we examined the entire functionality that the system was supposed to provide, as outlined in the System Requirements Specification, and we discovered that everything is likely to work with various front-end, middle-backend, and back-end stages. Specifically, we found that the system would be able to provide the functionality that was required.

2.3.3 Operational Feasibility

The system being proposed is made operationally feasible as a result of the following considerations. The user is eligible for extra advantages, and they will be able to go on with

their life even if they relocate to a new place. By making use of the pre-disaster information offered by the prediction, they may be able to be better prepared for the future. The expense of implementing the proposed method is almost inconsequential when measured against the benefits that are received. The most significant problem for the consumer is that just going to the website gives them access to numerous new updates connected to several different disasters. Therefore, it is feasible from a technological standpoint. Because the system includes a graphical user interface, even the most unskilled user will find it easy to use and understand how to do anything on their own. Additionally, studies have shown that workers who maintain a particular degree of fitness are better able to learn the foundations of the business, which increases the likelihood that they will be happy working there.

2.4 Technologies and Tools

2.4.1 Technologies

Flask

Python is the programming language of choice for developing the Flask web application framework's toolkit and the Jinja2 template engine serve as the basis for the Flask web framework. Python is the programming language of choice for developing the Flask web application framework. Flask is built on top of the WSGI toolkit and the Jinja2 template engine. Both of these components serve as the framework\'. The creation of Python web applications may now be done by the Web Server Gateway Interface (WSGI) standard. The WSGI protocol outlines a standard method for communication between web servers and the applications that run on such servers. It is a WSGI toolkit that does utility activities such as requesting and responding to objects. As a direct consequence of this, a web framework may be constructed on top of it. One of the pillars that support the Flask framework is the Werkzeug framework. Jinja2 is a Python templating engine that has gained a lot of popularity. A web templating system combines a pre-existing template with an individual data source to generate dynamic web pages. The term "micro-framework" is often used when referring to Flask. Its goal is to simplify while retaining the flexibility of an application's basic functionality. Flask does not provide either capability for the validation of forms or an in-built abstraction layer for the management of database operations. Instead, Flask allows developers to make use of extensions to provide their applications with comparable capabilities.

Python

Python is a well-known computer programming language that may be used for a variety of purposes, including the analysis of data, the automation of processes, the development of websites and apps, and more. Even though Python wasn't developed with a specific problem in mind when it was first developed, the fact that it is a general-purpose programming language means that it may be used to construct a wide variety of applications. Because of its adaptability and user-friendliness, it has quickly ascended to the very top of the list of programming languages that are employed in the modern world. According to the results of a survey in the year 2021, it was the second-most popular programming language among developers. Python is often put to use for a wide variety of purposes, including the development of apps and websites, the automation of routine processes, and the analysis and presentation of data. Because it is so easy to learn, a lot of people who are not programmers have discovered that Python may be useful for a variety of day-to-day tasks, including the management of money. This includes accountants and scientists. Python has established itself as a standard in the field of data science, making it possible for data analysts and other specialists to use it to do sophisticated statistical calculations, create machine learning algorithms, manage and analyse data, and a variety of other operations. Python can generate a wide range of data visualizations, such as line and bar graphs, pie charts, histograms, and plots in three dimensions. In addition, Python offers a range of libraries, such as TensorFlow and Keras, that assist programmers in the creation of data analysis and machine learning applications in a way that is both more efficient and quicker.

Sqlite3

By using the sqlite3 module, it is possible to connect it with Python. Gerhard Haring is the artist who created this. It offers a SQL interface that is compliant with the definition of the DB-API 2.0 standards that may be found in PEP 249. Since this module's migration was completed in Python versions 2.5 and above, there is no longer a need to redownload it since it is already installed. If you want to make use of the sqlite3 module, you will first need to construct a pointer object that is responsible for processing all of the SQL statements. Additionally, you will need to construct a connected object that is the database's representation. Database offer a variety of services that enable users to manage large volumes of data from the internet and information with a large volume of input and output in a regular file, such as a regular text file. This is made possible by the fact that databases store this information in a format that is similar to a regular

file. MySQL is used by a large number of websites. A language that employs SQL is called a query language. SQLite is a "light" version of the database. It works using syntax that is somewhat similar to that of SQL. SQLite is a database engine that is characterized by its high reliability, integration, autonomy, and stress. This is widely considered to be the most popular database engine on the face of the whole world. The SQLite databases are retrieved with the help of Sqlite3, which is a library for the Python programming language. It is compatible with the database that has been included with the Python package from version 2.5 and is intended to function with that database.

Jupyter Notebook

A user, scientific researcher, scholar, or analyst can produce and share a document known as the Notebook, which may include live programs, documentation, graphs, plots, and visualizations. Jupyter Notebook is an open-source web application, and its usage is free to anybody. The users of Jupyter Notebook have access to support for more than 40 different programming languages. These supported languages include some of the most popular programming languages, such as Python, R, and Julia. The user is provided with the option to save the notebook in a variety of file formats, including PDF, HTML, Python, Markdown, or as a job file. Anyone who is a data scientist, data engineer, data analyst, machine learning scientist, research scholar, scientific researcher, or a general user who wants to do any sort of scientific computation, data processing, or visualization-related work can use the Jupyter Notebook.

VS Code

The source code editor included with Visual Studio Code is blazingly quick, making it ideal for use in day-to-day operations. Syntax highlighting, bracket matching, auto indentation, box selection, snippets, and many other features help you become quickly productive with VS Code, which supports hundreds of different languages and offers many more features as well. You may explore your code more easily thanks to intuitive keyboard shortcuts that are simple to customize and keyboard shortcut mappings that have been given by the community. The simplicity of a source code editor is combined with the power of Visual Studio Code.

Django

Django is a web framework written in the open-source programming language Python that enables quick website creation that is also pragmatic, maintainable, and has a clean design. A web application framework is a toolkit that contains all of the components necessary for the building of an application. The primary objective of the Django framework is to provide developers the ability to concentrate on new aspects of the program rather than devoting their time to improving aspects of the application that are already complete. It offers a greater number of features than the majority of the other frameworks that are now available on the market. It takes care of a significant amount of the difficulty that is connected with web development, which enables customers to focus on developing the components that are necessary for their application.

Summary

The literature study takes up the whole of the second chapter, which also includes a synopsis of the review based on the historical data from the system. It provides an analysis of the problems that are inherent in the current system. In this section, we will discuss how the proposed system fills in any gaps that were created by the currently used one. It discusses the advantages and disadvantages of the existing system as well as the proposed alternative. After that, the findings of the feasibility study will indicate whether or not we should proceed with the proposed system. After this is some information about the tools that were utilized to develop this system as well as the capabilities of those tools.

CHAPTER 3

SOFTWARE REQUIREMENTS SPECIFICATIONS

The foundation of the software development process is a specification, which is a fundamental document. It highlights a system's key features in addition to outlining the requirements for using it. An SRS is a company's documented understanding of the system requirements and dependencies of a client or potential client at a specific point in time (typically before any real design or development work). It is a two-way insurance policy that guarantees both the client and the business are aware of each other's expectations at all times. In order to save time when rebuilding a software system, the Software Requirement Specification (SRS) is used to outline how the client and suppliers behave. Along with details on the product's risk, it offers an estimate of the product's cost. Furthermore, the Software Requirement Specification contains both functional and non-functional requirements.

The SRS also acts as a strategy for completing a project with the least amount of expense growth possible. The SRS is frequently referred to as the "parent" document because it serves as the foundation for all other project management documents that come after it, including design specifications, statements of work, software architecture specifications, testing and validation plans, and documentation plans. It is important to note that an SRS only includes functional and non-functional requirements; it excludes design concepts, potential answers to operational or technical problems, and any other information that isn't considered part of the development team's assessment of the customer's system requirements.

3.1 Users

The Stakeholders are intervened and needs of system requirements are being collected and then categorized into different modules keeping users in view for their operational purpose mainly the following stakeholders are identified in the system

i. Admin

ii. User

iii. Visitor

i. Admin

The admin is also the stakeholder among the two. The admin role in visitor management is he can create a profile and add doctors to the profiles based on their designation and he can manage the profiles of the doctors and can view how many visited the profile he creates, remove, update, delete the profiles of doctors and the appointments of the visitors. The Database used in the Application is Sql lite 3, Only the administrator will have the right to modify the database accordingly

ii. Users

User are Stakeholder in the Stroke Prediction as the prediction model is done completely based on the user model. The user has many functionalities as of admin he can view the page can use the prediction model where he needs to enter the input and enters the queries which have been asked after entering the values, and he can view the results. The user can even book an appointment with the doctor in case of emergency These are the functions that are given to the users

iii. Visitors

Visitor The among the stakeholder in the Stroke Prediction Application the visitor has the functionality to view the doctors and book appointments with the doctors. The visitor needs to enter his name and phone number to book an appointment with the doctor. The visitor can also view the current availability of a doctor whether the doctor is busy or available based upon this he can confirm the appointments with the doctor. The visitor can view the Dashboard to view the appointment

3.2 Functional Requirements

The term "functional requirements" refers to a description of the services provided by the programme. These criteria are also known as functional requirements. Functional requirements assist us in determining if the application provides all of the features stated in this document. It even assists us in identifying the needs that were overlooked. Functional requirements describe the core system behaviours. They describe what the system performs in response to the inputs and outputs provided by the users. Many components are addressed in the section on functional requirements, such as data input, computations, and business operations, and they typically determine the system's behaviour.

3.2.1 Data Collection

Data play a crucial role in the development of predictive systems. Your entire project depends on it since your system is based on it. The first and most crucial step that must be correctly finished is data selection. For the purpose of the study, we collected data from the government's website. These documents were open to everybody. Several other websites offer this kind of information. We chose the dataset based on the various components and restrictions we intended to include in our prediction engine.

3.2.1 Data Cleaning

Immediately after choosing the record. Since the data set could be in a different format, the following step is to clean the data and convert it to the appropriate format. Several datasets from various sources in a variety of file formats may also be used. They must first be converted to the proper format or type that the type prediction system accepts in order to be used. This step is essential because introducing limits that the prediction system does not need into the dataset could make the system more complex and take longer to process. Data cleansing is also necessary because the dataset could contain null and garbage values. Changing the garbage values when the data is converted is the solution to this problem. There are various methods for doing this.

3.2.3 Data Pre-Processing

Data pre-processing is a method that does all of the transformation work required to convert imprecise data into intelligible data. We gather real-world data, which is often inconsistent or incomplete. You may also be deficient in some traits or habits. So we clean the data and prepare it for our algorithms to make predictions easier. Eliminate the incidence of mistakes as well.

3.2.4 Data Analysis

The process of examining data to extract information to help decision-making is known as data analysis. To extract usable and meaningful information from our data collection, we use a variety of methodologies and procedures. We worked on several prediction methods to get certain predicted values from the same. Model of Prediction The stage of developing the predictive model include working with various statistical approaches, testing these models, and then utilising anticipated values to get meaningful information that may aid the end-user in any manner, such as predicting post-disasters or detecting safe zones.

3.2.5 Data Interpretation

The execution of the procedure by which the end-user may notice some helpful ideas/information about catastrophes or trends and assist them conclude is known as data interpretation. It comprises of the following steps: digesting the results of the data analysis, determining what we can deduct from them, and lastly considering how we may apply them, i.e. finish them. Along with DATA VISUALIZATION and DATA INTERPRETATION, we will learn about the many crimes that occur in various states and localities in this section. Data visualisation is the process of transforming textual data into excellent visuals such as bar charts, scatter charts, and pie charts for better interpretation and modification by the end-user. We create charts for different calamities and damages.

3.2.6 Data Management

The system must manage geo-referenced data, such as points of interest (POIs), maps, and vector data (polygons, lines, and points), as well as other data, such as text documents or status information on other associated units. It must include report formats that aid in the evaluation of particular items such as critical infrastructure such as airports or power plants. These forms assist field assessment professionals in asking pertinent questions and communicating the data to others.

3.2.8 Synchronization

The system must synchronise data among its instances. Such cases might be found locally in the field or elsewhere in the globe. In particular, the system must deal with irregular network connection and utilise connectivity whenever it becomes available.

3.2.9 Algorithms

A functional requirement describes the function of a software system as well as how the system must work when given certain inputs or when faced with certain circumstances. It is possible that additional specialised functionality, such as computations, data manipulation and processing, and so on, will be included. The following is a list of the functional requirements that need to be met by this system: -

Methodology of Machine Learning, Visualization of Assets, Methodology of Machine Learning. Before agreeing on a model, the modeller might use this process to determine

the "performance ceiling" for the data set. In many circumstances, a variety of models will perform similarly, allowing the practitioner to consider the advantages of various techniques.

Among the approaches used in our projects are

- i. Decision Tree
- ii. Random Forest
- iii. Logistic Regression

3.2.9.1 Decision Tree

Decision tree is a tool for supporting decisions that uses a tree-like graph or model to represent options and their outcomes, including utility, resource costs, and outcomes of chance events. It's one method of demonstrating an algorithm with only conditional control statements. The use of a decision tree is crucial when working with multidimensional data. One of the best and most popular supervised learning techniques is the usage of tree-based learning algorithms. Prediction models using tree-based techniques provide good accuracy, stability, and simplicity of understanding. They are more effective than linear models at mapping nonlinear interactions. Any problem they encounter can be overcome.

3.2.9.2 Logistic Regression

One of the ML algorithms used most frequently in the supervised learning approach is LR. It is a method of forecasting where a group of independent factors are used to predict a categorical dependent variable. Using logistic regression, the result of a categorical dependent variable is predicted. As a result, the output must be discrete or categorical. True or false, yes or no, 0 or 1, and so on, with probability values between 0 and 1. The applications of logistic and linear regression are extremely similar. While linear regression is used to address problems with regression, LR is used to address classification-related issues.

3.2.9.3 Random Forest

The classification algorithm applied was RF classification. A large number of separate decision trees that were trained on a random sampling of data make up RFs. These trees are created during training, and decision tree outputs are compiled. Voting is the process used to determine the final prediction made by this algorithm. Each DT must choose one of the two output classes

in this process (in this case, stroke or no stroke). The final forecast is determined by the RF approach, which chooses the class with the most votes.

3.3 Non-Functional Requirements

Non-functional requirements describe the quality of software. They are in charge of aspects like maintainability, performance, security, portability, and dependability. They handle the critical concerns that reflect the software system's quality. Non-functional needs that are not adequately addressed may result in dissatisfied customers, software inconsistencies, and cost overruns. Performance restrictions, operational limits, interface restrictions, and many other categories are used to categorise non-functional needs. Physical constraints, the system's interactions with people and the environment, cost, the calibre of the design, and other factors are examples of these. These have a number of advantages since they help to ensure legal compliance as well as dependability, scalability, and performance. Additionally, they save money, maintain a simple operating system for the software, and offer a respectable user experience. Non-functional needs are those that have no direct connection to the specific function that the system offers. Instead of specifying specific behaviours, they establish the criteria that may be used to assess how well a system functions. They may be connected to characteristics of emergent systems like dependability, response speed, or store occupancy. User expectations and cost constraints lead to non-functional needs.

Because of company rules, the necessity for compatibility with other software and hardware systems, or external forces such as: -

- Scalability
- Performance
- User Requirement
- Response Time
- Maintainability
- Usability

3.3.1 Scalability

The analytics platform must be applicable to any size machine or facility. The solution must be able to add assets without requiring any more hardware, software, or dedicated worker hours.

3.3.2 Performance

The goal of an industrial analytics platform is to offer accurate and timely data to the manufacturing plant.

3.3.3 User Requirements

To setup the network, a user interface is required. There should be a choice for the user. Viewing the performance parameters is an option. The system should be user-friendly, with the client application visible in the system tray and the user just needing to click to pick any settings.

3.3.4 Response Time

The time between a system query and the response to that question is referred to as response time. Response time, a system performance metric, can be used to assess service requests generated by a variety of technologies. Perhaps quick reactions are necessary for effective computing.

3.3.5 Maintainability

Maintainability is an essential quality trait, but it is a tough concept to grasp since it requires numerous assessments. Estimating software maintainability is a kind of quality estimate. Maintainability is a collection of characteristics that influence the amount of work required to make the specified changes.

3.3.6 Usability

One issue that designers of interactive systems face is catering to the broad spectrum of consumers who will utilise a certain application. Understanding the user is essential for creating a useful interface. There are numerous approaches to tackling this issue, including better design processes that employ "intuitive" interface designs, adaptable interfaces, and better training and user support resources.

3.3.7 Data Requirements

Information Machine learning has the potential to allow new types of predictive analytics and to incorporate algorithm-driven knowledge into a wide range of software applications. However, none of this is feasible without the proper data being recorded and analysed.

Machine learning algorithms analyse massive amounts of data in order to discover complex patterns about people, health, transactions, and events, among other things. This knowledge is subsequently fed into a prediction model. Comparing an entity to the model may indicate if it is performing within permissible bounds or displaying an abnormality. Machine learning is used to tackle problems with well-defined boundaries, such as classification and grouping. During development, a machine learning algorithm learns on so-called training data. The first and most critical stage in scoping data needs is determining what data are required for a certain product or feature. Machine learning models are simply mathematical functions that take characteristics as inputs, generate predictions as outputs, and learn how to best match predictions to patterns found in training data. In most circumstances, more data is preferable than less. Transfer learning may be useful when there is little or no data available. In summary, transfer learning enables you to take data and/or machine

If gathering labelled data involves money (or time), set a target for where you want to go (in terms of model quality/performance) and a budget for how much money/time you are willing/able to invest. More data will not assist at some point. To illustrate these claims, consider the following simplified graph, which emphasises the different scenarios in which additional data may or may not be advantageous. The assumption is that most ML issues are on the steep side of this curve, implying that gathering more data will result in improved performance. However, if there is a large amount of labelled training data already available, there may be diminishing returns, i.e. training on additional data does not enhance model quality.

The dataset shown below is pre-defined in the UCI Machine Learning Repository. The dataset contains 73 characteristics, of which choose those that contribute the most to our prediction output and in which we are most interested. Irrelevant characteristics in our data may reduce model accuracy and cause our model to train based on irrelevant information.

Summary

The Software Requirement Specifications are described in this chapter. These criteria specify the product's performance, which is necessary to fulfil all of the users' demands. It defines the functions of Users, Administrators, and Producers. Following that, we have functional requirements, which outline the services provided by the programme. We offer numerous features for consumers, manufacturers, and so forth. Following the functional need, there are user interface requirements. Following that, we have hardware and software needs that demand on distinct specs. At the conclusion of the chapter, we have non-functional requirements that define the quality characteristics of the product.

CHAPTER 4

SYSTEM DESIGN

Depending on the needs, System Design is a way to define the various components of a tool, such as modules, architecture, components, and their interfaces, as well as records for a tool. It is a process for defining, developing, and designing systems that adhere to the particular objectives and standards of a business or organisation. A tool plan is necessary for a well-taken walk that is clear and concise. A bottom-up or top-down approach is desired to remember all the tool's interconnected variables. We use the data flow diagrams from the investigation phase and the design phase to create the system. For designing, we employ universal access techniques. We think about the top-down strategy. We choose the entities, their attributes, and the connections between the entities during the design process. We are in charge of both the system's conceptual and physical design.

The most creative and challenging phase of the life cycle is system design. System format is the way of designing the elements of a device consisting of the architecture, modules, and components. The critical cause of the System Design way is to give sufficient positive information and the facts about the device and its device elements to permit the implementation of consistent with the architectural entities as defined in the models and views of the device architecture. In this phase, a high-level format form is furnished, and its communication is furnished as a context diagram for the critical aspect interactions at the beneficial level with the architectural format and beneficial components.

The output's production process and format are the designer's objectives. It also includes examples of input and output. It is necessary to design additional input files and database files to comply with the specifications of the proposed output. Construction and testing of the programme is used to manage the processing phases. As a final step before implementation, management records and assesses information on the system's justification as well as an assessment of the prospective system's impact on the user and the organisation. Any project's design phase is the most crucial. Moving away from the area of concern is the first step towards the solution vicinity. The input to the format phase is the specifications of the device to be designed. The output of the top-level format is the architectural format or the device format for the software program application device to be built. A format should be very clear, verifiable, complete, traceable, inexperienced, and simple.

4.1 Architectural Design for Stroke Prediction

The architectural format is a visual representation of device pieces. The gadget's shape format is the first stage of the format. It specifies the gadget's structural, behavioural, and other aspects. The architecture format aids device designers and developers in visualizing data in the form of diagrams. The overall form of the gadget is to ensure that the gadget is thrilling and meets the demands of the patron. A somewhat form diagram is a blueprint that is utilized to explain the patterns that will be used for the duration of the device.

Every element that contributes to building up a part of the system is represented visually and this Visual representation is called an Architecture diagram. The main advantage of drawing an architecture diagram is that it helps each and every one who uses the system to know about the layout of the system and understand it in a better way. It can be considered as something similar to the blueprint of the system. The whole system with all its components is shown in the architecture diagram. We all know that a person understands the diagrams easily when it is compared to the information, so the diagrammatic representation is the best way to make users understand the system. Different types of architecture diagrams include Application architecture diagrams, Integration architecture diagrams, deployment architecture diagrams, Devops architecture diagrams, website architecture diagrams, and many more. These diagrams say about the overview of the systems information. They don't give a detailed explanation. An architecture diagram helps to visualize complex processes easily. It helps the users to learn about the system.

4.1.1 Architectural Design for Stroke Prediction

A system's structure, behaviour, and viewpoints are described by its system architecture, which is a conceptual model. An architecture description is a formal description and representation of a system that is set up to facilitate analysis of its components' structures and actions. A conceptual data structure is referred to as architectural design, and the project's general logical structure is broken up into processing modules.

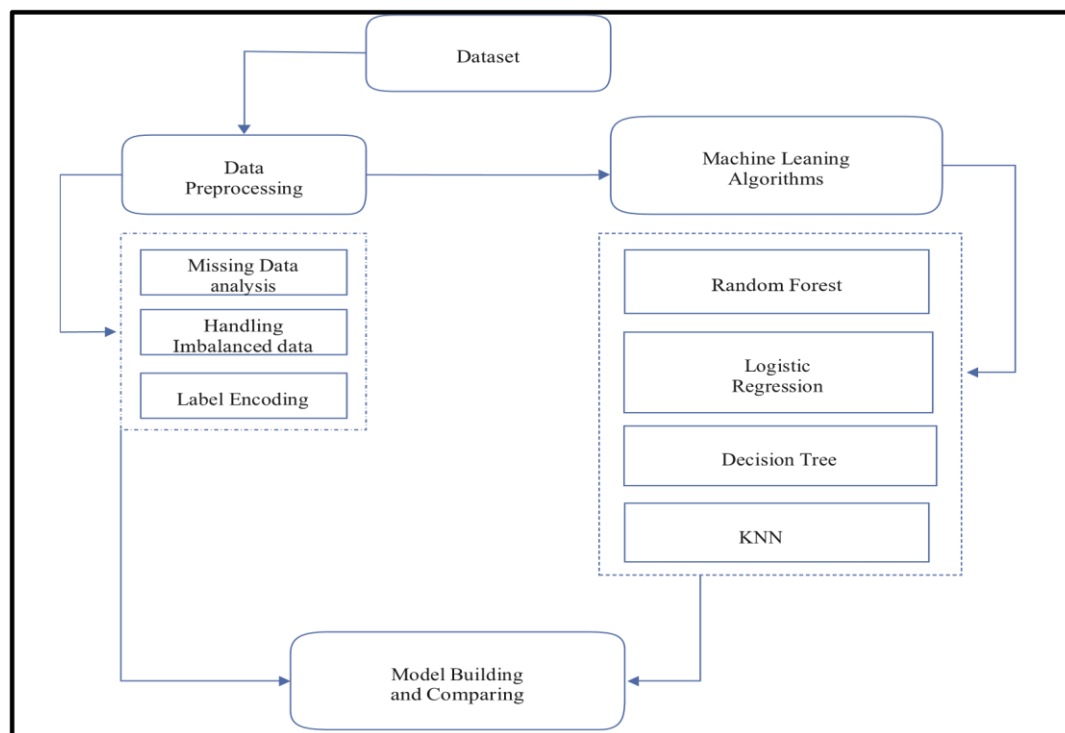


Fig 4.1 Architectural Design for Early Identification of Stroke

Fig 4.1 shows the overall logical structure of the project with the following modules Input data: Risk factors like age, gender, hypertension, heart disease, BMI, Smoking status, and Glucose level. Machine Learning Techniques: Logistic Regression, Decision Tree, Random Forest Analysis: Prediction and analysis of stroke whose performance is based on machine learning techniques. Management Suggestions and improvement of stroke victims.

4.1.2 Architectural Design for Visitor Management

The Architecture Diagram for visitor management follows when a visitor enters the application he should register after registering he can view the doctors and check the availability of the doctor if the doctor is free and can book the appointment with the doctor. The admin's role is that can manage the profile of the doctors and can manage meetings the admin can add or delete the doctors the has also had a login module where he needs to log in first and then he can do the changes. The admin can manage the database.

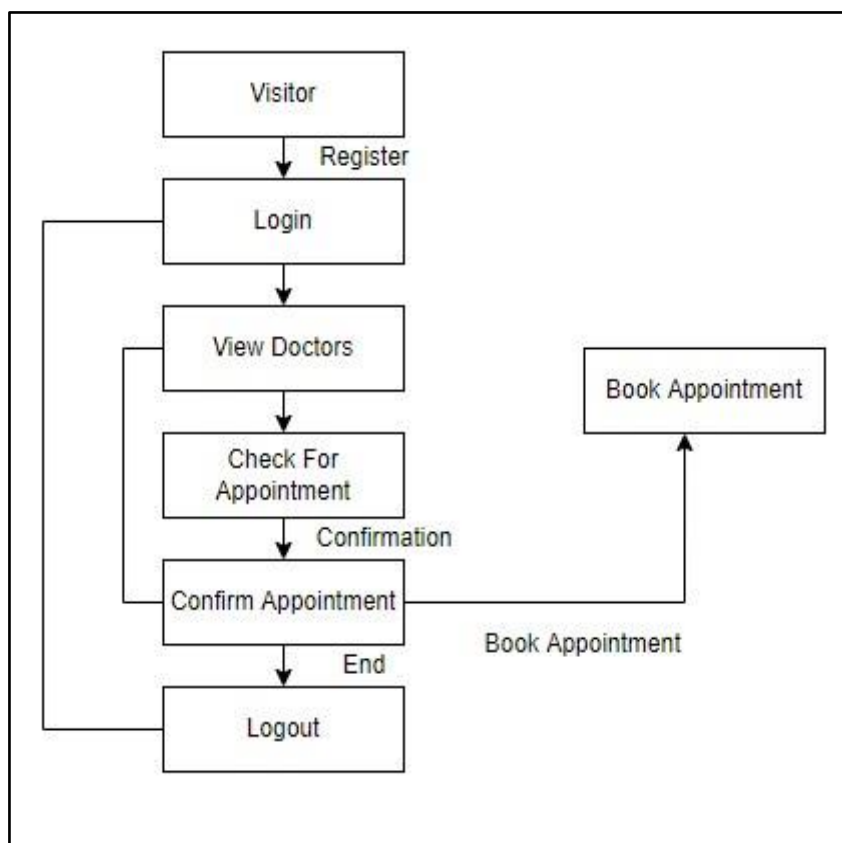


Fig 4.2 Architectural Design for Visitor Management System

4.2 Context Diagram

The DFD can be used to carry out a device or software program application software at any diploma of abstraction. DFDs can be partitioned into levels that constitute growing records of waft and realistic detail. Levels in DFD are numbered 0, 1, 2, or beyond. Here, we're capable of regularly seeing 3 levels withinside the facts waft diagram, which are: 0-diploma DFD, 1diploma DFD, and 2-diploma DFD.

DFD is the abbreviation for Data Flow Diagram. The flow of data of a system or a process is represented by DFD. It also gives insight into the inputs and outputs of each entity and the process itself. DFD does not have a control flow and no loops or decision rules are present. Specific operations depending on the type of data can be explained by a flowchart data Flow diagrams can be represented in several ways. The DFD belongs to structured-analysis modelling tools. Data Flow diagrams are very popular because they help us to visualize the major steps and data involved in software-system processes.

DFD Level 0 for Stroke (Data Flow Diagram)

It gives an abstract view of the system. It shows the relationship between the internal and external entities in a single process. In this diagram, the whole system is represented as one bubble and the input and output interactions are represented using arrows. Gradually, when we go deeper the single bubble is then split into many bubbles and their functionality is studied in detail. This process continues and the bubbles are divided and documented in a more detailed format.

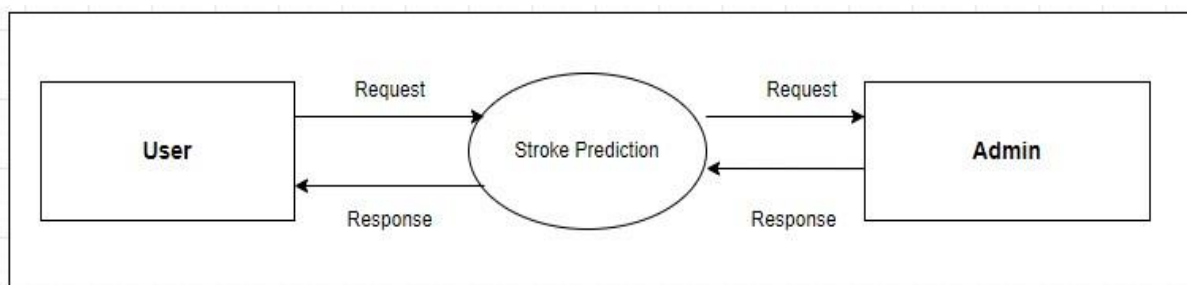


Fig 4.3 DFD Level-0 For Stroke

As illustrated in figure 4.3 the data flow diagram, briefly describes the system. It tells about the users who input the information about the users into the Stroke Prediction System which in turn will send the request to the admin. In turn to this request, the admin sends the response to the system which will be gradually sent to the user.

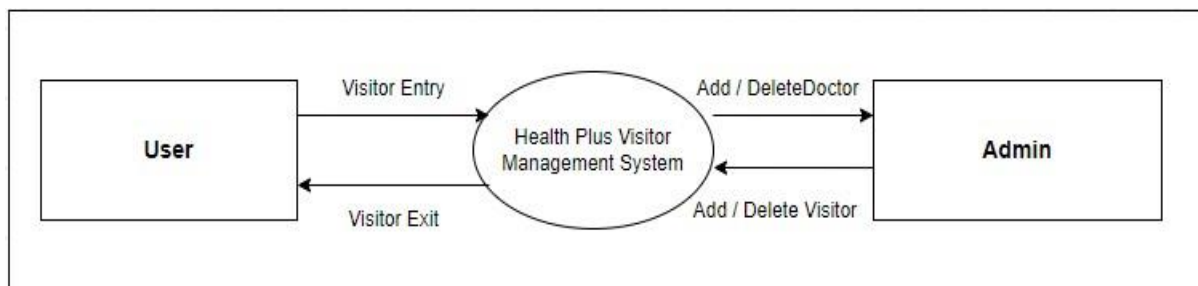


Fig 4.4 Level-Zero DFD for Health Plus Visitor Management

As illustrated in figure 4.4 the data flow diagram, briefly describes the system. It tells about the users who input the information about the users into the Health Plus Visitor Management which in turn will send the request to the admin. In turn to this request, the admin sends the response to the system which will be gradually sent to the user.

DFD 1-Level for Stroke Prediction

In level one of the DFD, the DFD single bubble of the system is now divided into multiple bubbles. This level includes the subprocesses. This helps to give the users more information

about the breakdown of processes. It will highlight the main functions of the system at this level.

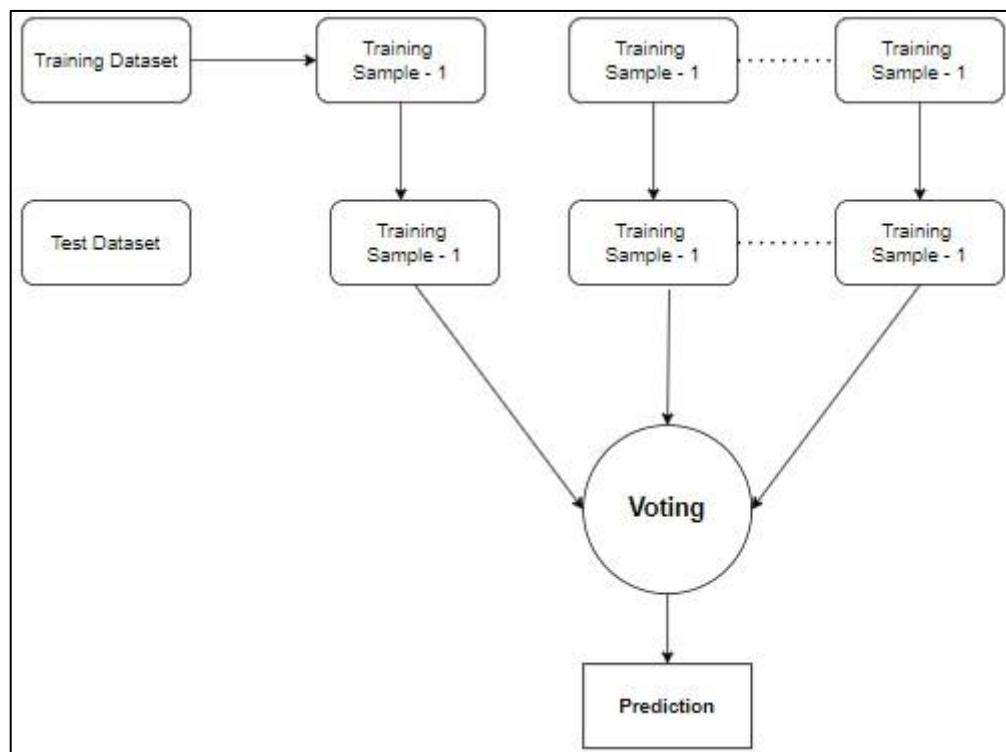


Fig 4.5 DFD Level-1

As illustrated in Fig 4.5 the data flow diagram, there are two entities, the customer and the User. As concerned with the internal entities, there is manage customer information and manage Once the User enters the Inputs the model splits into two parts test and Train. The inputs by the users and sent into Training samples which are then merged and then the prediction is based on the voting classifier which combines the user inputs and then predicts the results using the algorithms

DFD 2-Level for Prediction

When it comes to DFD level 2, it goes even deeper than that of level 1. It helps to understand the functioning of the system better. It makes the process of studying the system in detail easier. When there is a necessity to record or plan the specific detail of some component of the system, we can opt for this level.

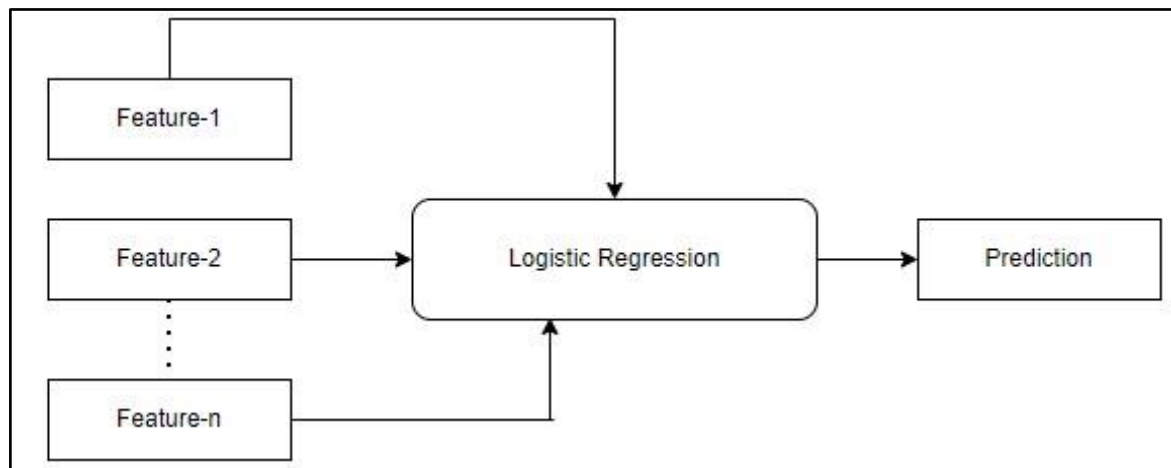


Fig 4.6 DFD Level-2

As illustrated in Fig 4.6 there is a Dataflow of the Algorithm on how the Prediction works there are ways in which the prediction works basically when the user enters the input the prediction is done based on Machine Learning There are many algorithms that have different functions but the prediction accuracy is same so in the level 2 Data flow Diagram it is shown how the Algorithm works based on the user input

4.4 Module Level Description for Stroke Prediction

4.4.1 User

User The Among the Stakeholder in the Stroke Prediction as the prediction model is done completely based on the user model. The user has many functionalities as of admin he can view the page can use the prediction model where he needs to enter the input and enters the queries which have been asked after entering the values, and he can view the results. The user can even book an appointment with the doctor in case of emergency These are the functions that are given to the users

4.4.1 Module Level Description for Visitor Management System

4.4.1.1 Admin

The admin is also the stakeholder among the two. The admin role in visitor management is he can create a profile and add doctors to the profiles based on their designation and he can manage the profiles of the doctors and can view how many visited the profile he creates, remove, update, delete the profiles of doctors and the appointments of the visitors. The Database used in the Application is Sql lite 3, Only the administrator will have the right to modify the database accordingly

4.4.1.1 Visitor

Visitor The among the stakeholder in the Stroke Prediction Application the visitor has the functionality to view the doctors and book appointments with the doctors. The visitor needs to enter his name and phone number to book an appointment with the doctor. The visitor can also view the current availability of a doctor whether the doctor is busy or available based upon this he can confirm the appointments with the doctor. The visitor can view the Dashboard to view the appointment

4.5 Collaboration Diagram

The collaboration diagram takes advantage of the connection between two or more distinct system elements. It provides information on the object-oriented programming-compliant object architecture that is present in the system. There are numerous characteristics of an object. The communication diagram is an alternative name for the cooperation diagram. The many system components' objects are interconnected. The collaboration diagram is created when it is necessary to explain the relationship between various things. The sequence diagram and collaboration both provide the same information, but in distinct ways. In fact, the collaboration diagram works well for use case analysis in the system. Here, the System's structural aspect is given emphasis. Communication diagrams and collaboration diagrams are synonyms. They can show how objects interact in order to carry out a certain use case activity or a particular use case aspect. Designers can explain and identify the responsibilities of objects executing a certain flow of actions in a use case using collaboration diagrams. They serve as the primary information source for determining the functions and interfaces of classes. The process of creating collaboration diagrams begins with identifying the design components needed to incorporate the functionality of interface features. Then, a model is created using the interactions between these components. A collaboration diagram demonstrates how different lifelines connect within a structure's systemic dimensions. Collaboration is a collection of explicitly stated processes and the mechanisms that connect them. To complete any task, they cooperate.

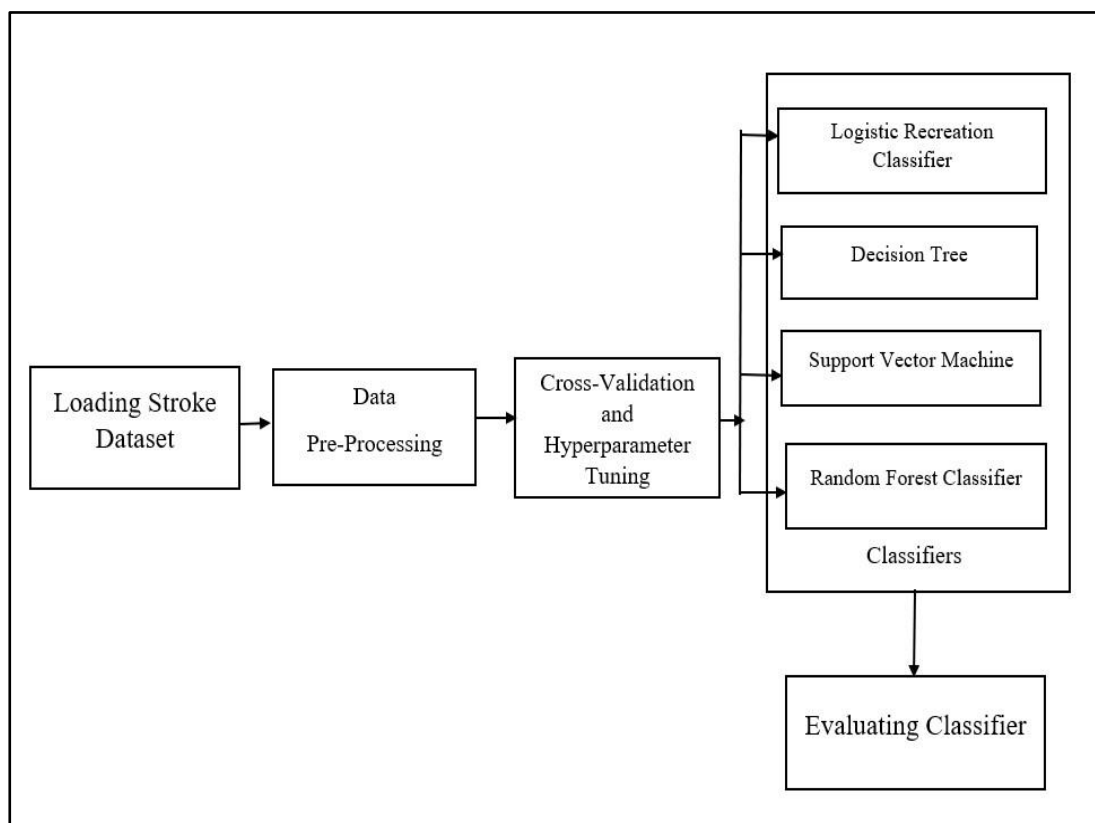


Fig 4.7 Collaboration Diagram

As illustrated in Fig 4.7, the collaboration of the different objects in the Stroke Prediction system, the users of the application namely customers who use the application to get to know about whether they are affected by stroke or not, the various modules of the system have collaborated in such a way that it should perform the tasks when the user gives the inputs it should show whether the user has a stroke or not

Summary

This chapter describes the system design, which tells about the process of designing various components, interfaces, and architecture that meets the requirements of the user. Next to this can see how the external entities and internal entities react with each other with the help of a context diagram. which contains various DFD levels. Proceeding this we have the Entity relationship diagram that shows us how different entities are related in the system. After this, we have the module-level description that includes information about the admin, producer, and customer. Then we have the collaboration diagram to learn about the relationship between two or more objects.

Chapter 5

DETAILED DESIGN

The chapter discusses detailed diagrams related to the “Early Identification of Stroke” application and the diagram like the use case diagram, sequence diagram, and database design. Use case mapping is a process of briefing the particulars of an organization and the operators inside that classification. It is frequently accessible as a graphical depiction of the communications among the numerous basics of a scheme. Using case figures requires the arrangement of the proceedings and how individuals measure flow. Though, though the use of instance figure prepares not to define how these proceedings are applied. A custom instance is a procedure used in classifications scrutiny to classify, simplify, and unify organization necessities. In this framework, the stretch "structure" mentions rather what is established or functioned, such as the pack and the postal instruction auction of products. Usage circumstance figures are cast-off in UML, a typical representation for demonstrating physical-creation matters and organizations. There are several compensations of a usage situation illustration over comparable figures such as flowcharts.

5.1 Use Case Diagram

Use case mapping is a process of briefing the particulars of an organization and the operators inside that classification. It is frequently accessible as a graphical depiction of the communications among the numerous basics of a scheme. Using case figures require proceedings in an arrangement and how individuals measure flow; Though, though the uses instance figure prepares not to define how these proceedings are applied. The requirements that are listed here are mainly related to the design aspect. It gives the overall structural overview of the system. It helps to recognize the factors that are influencing internally and externally. Using a use case diagram helps to identify the functionalities by analysing the requirements. To describe in simple words, use case diagrams are the functionalities of the system which are written in a well-understood manner. In the use case diagram, we come across actors. Anything which interacts with the system is considered to be an actor. These actors can be humans or internal or external applications. In the use case diagram, the functionalities are depicted as the use cases and the users as the actors. The relationship between these two is marked in the use case diagram.

5.1.1 Use case of Admin for Prediction

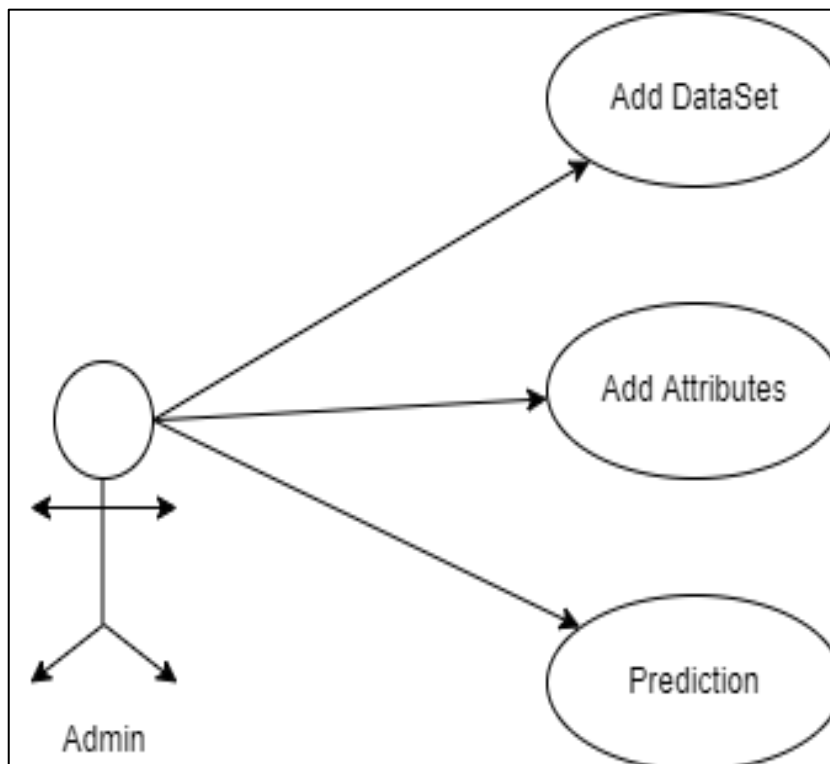


Fig 5.1 Use case of Admin for Prediction

The above Fig 5.1 of the admin use-case diagram explains about accessibilities of the admin where the admin application of the Early Identification of Stroke Has Several Such has been He can Add Datasets Or Dataset. The Prediction Model is done based upon the Previously available Datasets Were the Admin test and train the dataset using the algorithms. There are many attributes used to Predict the Stroke Such as age, gender, BMI, etc Using these the Prediction is done and the output is predicted whether the Patient has a Stroke or not.

5.1.2 Use case of User for Prediction

Fig 5.1.2 The use of User side prediction, the user is the only one who uses the software here the user can Visit the application, the user can view the features of the application like services provided by the health care centre he can view the contact details of the health care centre, later on, he can visit the prediction page the user can enter the required values and check for the prediction if he has any symptoms, he can book an appointment with the doctor.

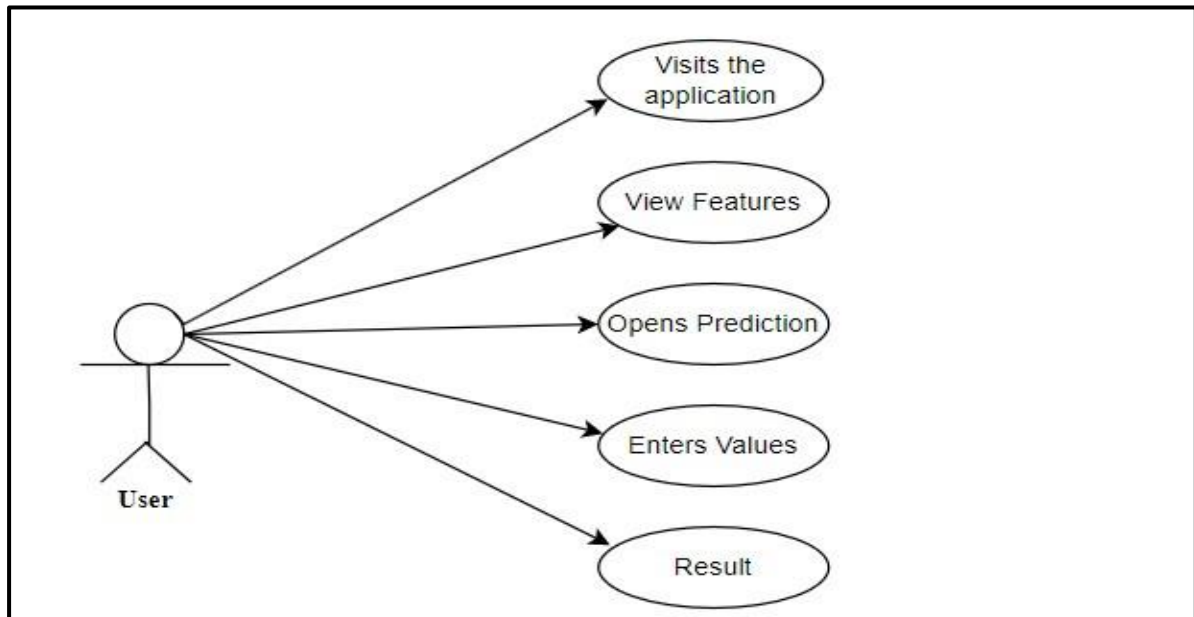


Fig 5.2 Use case of User prediction

5.1.3 Use case of Admin for Visitor Management

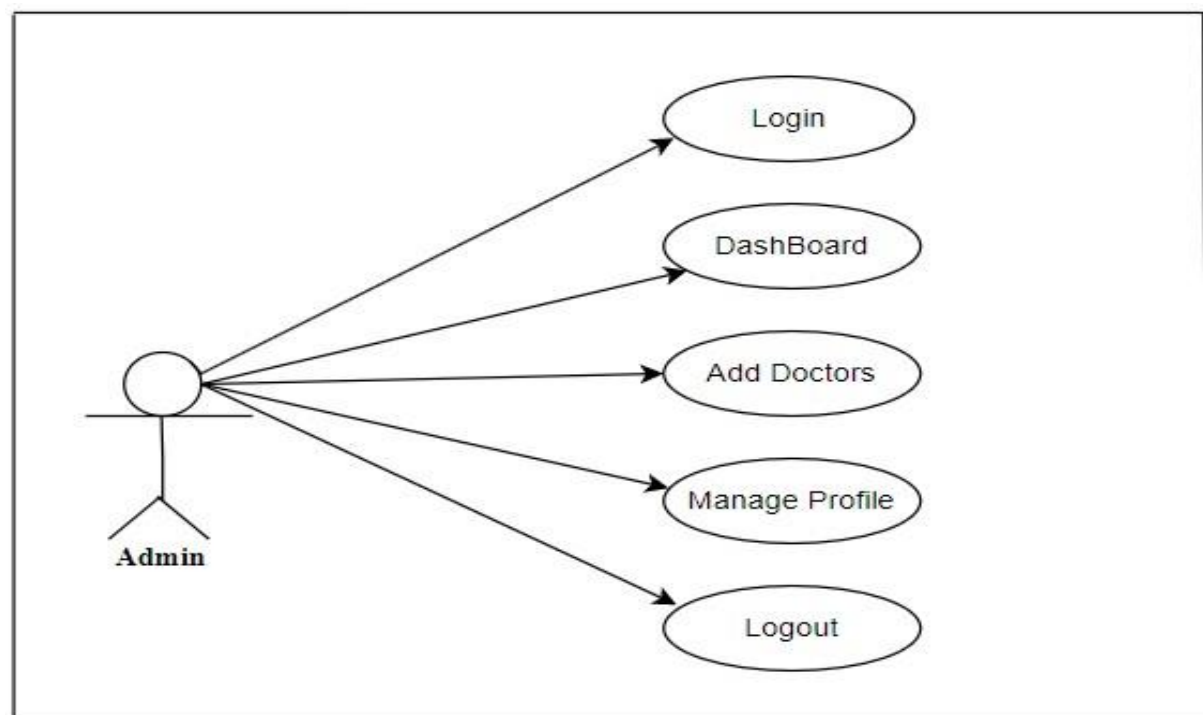


Fig 5.3 Use case of Admin for Visitor Management

The checkout option is openly available on the dashboard, someone else may misuse the option and checkout for others. Assume we use a verification system for checkout on the dashboard, the visitor may forget to checkout after leaving and as it requires verification, the admin will be shown as BUSY on the dashboard for an unknown time. If we make the

Checkout option manual, means the visitor is required to go to reception and ask for checkout, then also the visitor may forget to visit reception after leaving. Also, our system is made to be automatic and this will make it manual. The admin is the only person involved in the meeting after the visitor, he also knows that exactly when the visitor left, he can be given the option to check out the visitor.

5.1.4 Use case of User for Visitor Management

The organization's receptionist is primarily responsible for running the visitors' record management system. The numerous user interfaces that the system's user interacts with are depicted in the use case model diagram in (Fig 5.4) below. A thorough security strategy must include the capability of automatically screening and tracking visitors. The following features should be included in a computer-based visitor record management system

- Register visitors (visitors' check-in)
- Check-out visitors

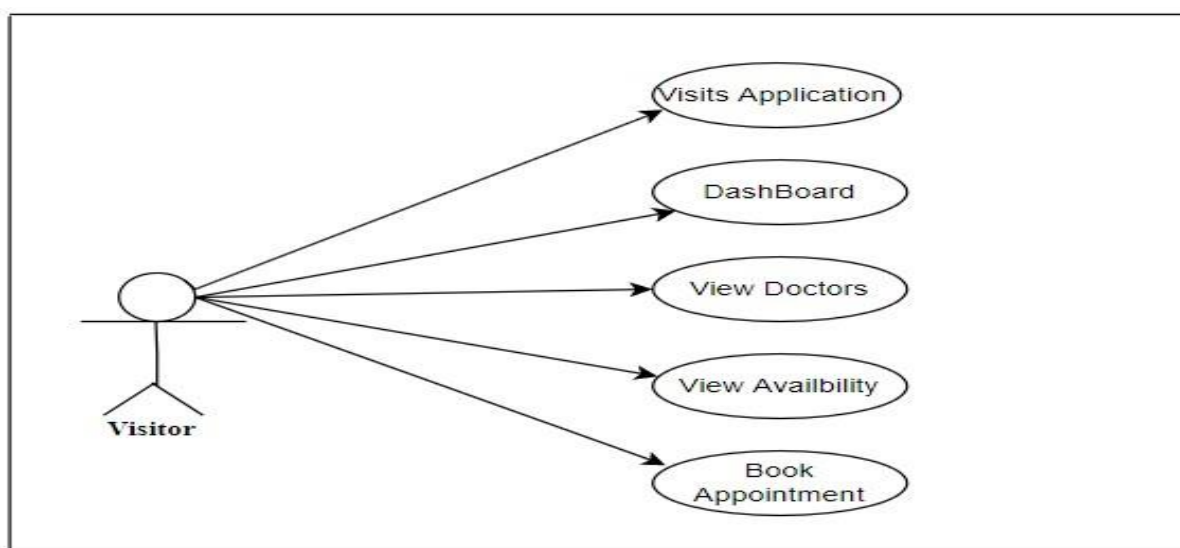


Fig 5.4 Use case of User for Visitor Management

5.2 Sequence Diagram

A sequence diagram, which explains how and in what order a group of devices interact, is a type of interplay diagram. Software developers and business professionals use these diagrams to understand the specifications for a new device or to document an existing process. Sometimes occasion diagrams or event scenarios are used to describe sequence diagrams. The sequence diagram displays the communication between the entities and is arranged in chronological order. It is used to carry out the function, displays the system's objects and modules, alters the variety of messages sent between the objects. Equivalent

used logistic regression, random forest, and decision tree. The best and most accurate algorithm has to be selected and use a random forest for prediction. If an error occurs then adjust the weight and check if the error is very small then get the prediction.

Sequence Diagram for Visitor Management

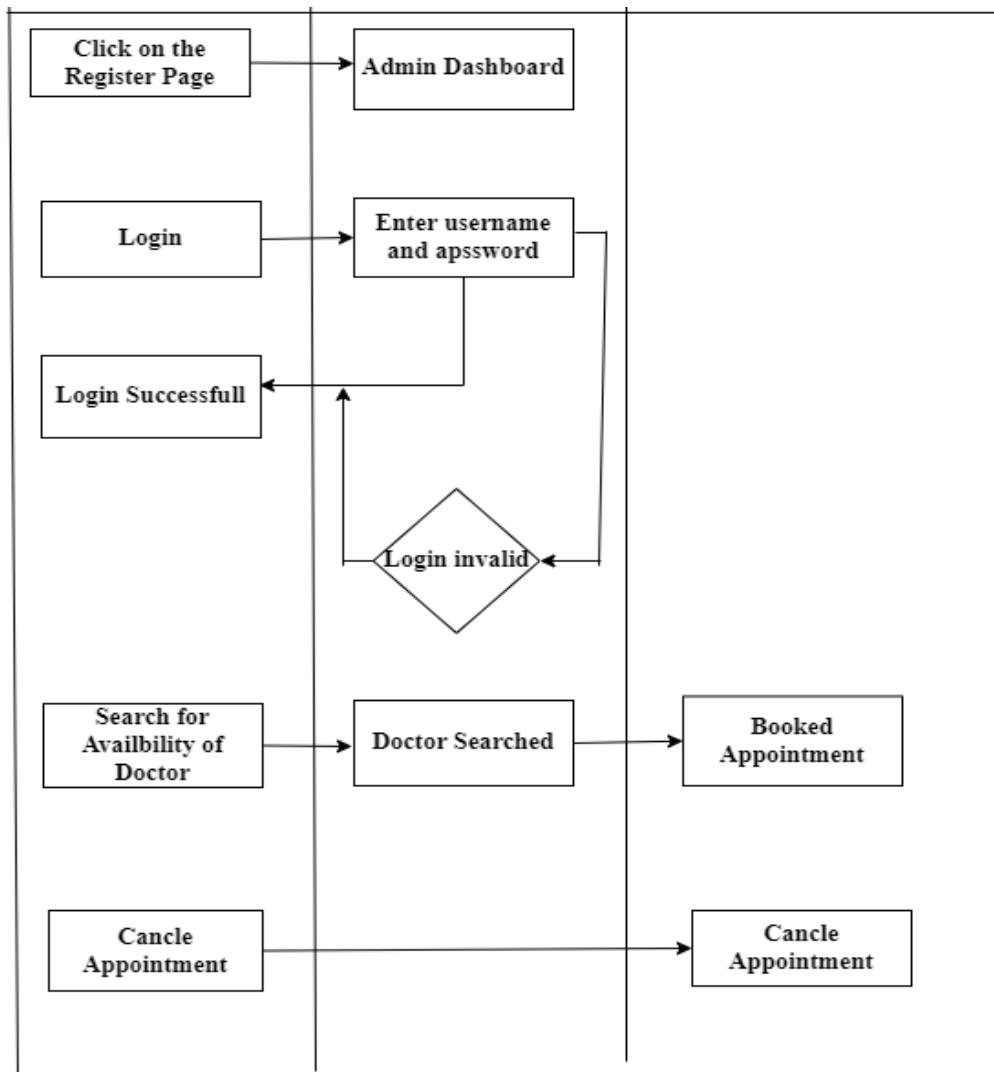


Fig 5.6 Sequence Diagram for Visitor Management

The layout of the database tables, the types of fields and their lengths, and the manner in which the data should be structured based on user requirements are all specified by the database design. The overarching goal is to provide all users with an information access system that is user-friendly, speedy, economical, and adaptable. During the process of designing the database, the following factors were taken into account-controlled redundancy, data independence, accuracy and integrity, large amounts of information at a

low cost, backup and recovery from failure, privacy and security, system performance, efficiency, and effectiveness.

5.3 Activity Diagram

To describe the dynamic aspects of the system, an activity diagram is important. It can be considered a flowchart fundamentally. It indicates the flow from one operation to another operation of the system. The activity diagram deals with all kinds of flow activities. There is a necessity to understand the components such as associations, conditions, activities, constraints, etc. After recognizing all the aspects, we can draft the whole flow. In simple words, it tells about operations. The reason for designing the activity diagram is to show how the flow of activities takes place in the system, explain the branched, collateral, and simultaneous glide of the system, to explain the one to another sequence. Activity diagrams help to build the executable system. They do not show any message flow, and this is the only missing aspect in the activity diagram.

5.3.1 Activity diagram for User

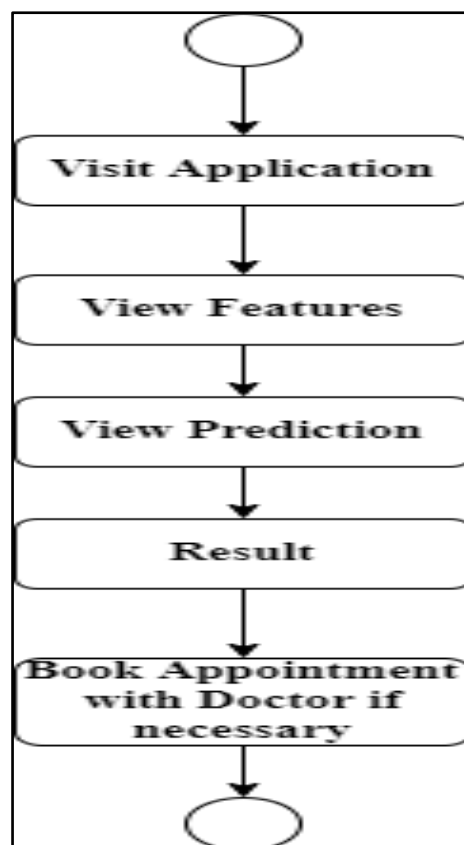


Fig 5.7 Activity diagram for User

Fig 5.7, User module for Stroke Prediction, the User will fill in the details given on the prediction page like age, gender, hypertension, heart disease, Residency, work type, etc. The prediction will get according to the user-entered data.

5.8 Activity diagram for Admin

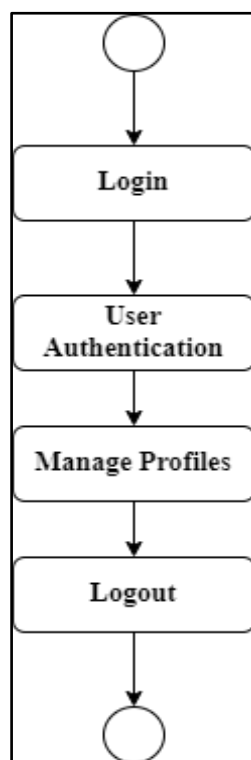


Fig 5.8 Activity diagram for Admin

The above-shown Activity Diagram for the Admin describes all sets of activities done by the Admin. He needs to login into the portal, using a valid set of credentials. Once he enters the portal he can deal with the users and manage them. He has the privilege to add new users and Doctors to the portal. Admin also has the authority to delete the user when necessary.

5.3.2 Activity diagram for Visitors

When A Visitor Checks in the Visitor Management Portal he first lands on the Dashboard where he can see the profiles of the doctors to visit the doctor's page the visitor must enter the

Profile Manager Page which lands into the doctor's profile Where the Visitor can view the Doctors profile and can check the Availability of if the doctors are busy, it prompts busy if they are available free The Visitors can Book Appointment, he has to add his name phone

number etc once the Appointment is confirmed The Visitor can check the confirmation in Dashboard.

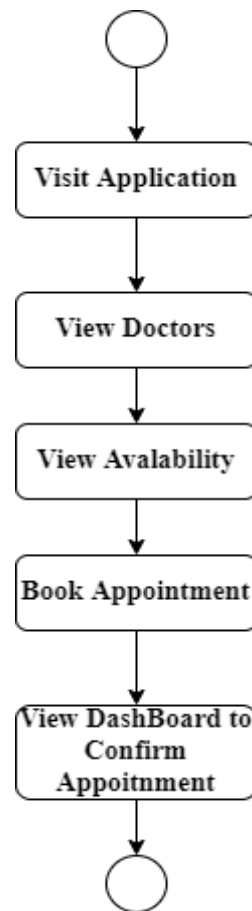


Fig 5.9 Activity diagram for Visitors

Summary

In this chapter different detailed designs are explained. In the Use case diagram users' and admin's relationships are explained in the form of actors and how the individual modules perform the tasks. A sequence diagram will give information about what are the steps will involve during the usage of the application. The activity diagram gives an idea of the different activities involved in this application.

CHAPTER 6

IMPLEMENTATION

Software implementation refers to the process of accommodating a software application to check its workflow and adaptability capacity. It can also be regarded as the process of executing the plans or the process of bringing strategies into action. Implementations apply to all the software, may be simple or complex. The process of implementation also applies to the significant updates which are required. The implementation makes the work more productive and efficient. The most regular kinds of implementations are Direct, parallel run phased, pilot, etc.

Implementation is important because it brings a plan or a concept into reality. To make a proper implementation clear communication between everyone in the team is necessary. If Implementation is not done properly, this may lead to server downtime and cause downtime loss to the organization or the company. There are some steps to follow to make a successful implementation. Firstly, outline all the tasks you need to complete as a part of the project. Communicate each step with all the necessary employees and assign the responsibility to each one. In the testing, the environment tries the newly released software which is a virtual space. In the middle of this make sure to train the employees about the process of implementation. Let everyone who needs to use the software get trained well. As the time of implementation is concerned, make sure to shut down all the existing software to check with ease. Be ready with all the necessary tools and instructions before the process of implementation. Lastly, make sure to take the feedback from the team so that it helps you to make the process still easier.

The implementation stage requires the following tasks.

- Careful planning.
- Research the system and constraints.
- Design of techniques to obtain the changeover.
- Evaluation of the changeover technique.
- Accurate decisions concerning the selection of the platform

- Suitable selection of the language for application development

6.1 Code Snippet

Algorithms that are Used For prediction

The prediction is the one that is used to predict the Stroke. The prediction is for the user to predict the Early identification of stroke The steps for the prediction are:

- i. Collect Historical Data
- ii. Cleaning & Preparing Data
- iii. Analysing & Building Models
- iv. Merging Models & Accuracy Calculation

i. Collect Historical Data

Data from the past is required for solving any problem involving prediction or classification. In the course of this project, the historical dataset pertaining to each catastrophe must be compiled once per year. It would be inefficient to collect data on a daily basis by hand; therefore, a Python script is executed on a distant server.

ii. Cleaning & Preparing Data

After the data is collected, need to clean & prepare the data according to the model's requirements. In any machine learning problem, this is the most important step and the most time-consuming. this used various statistical techniques & logic and implemented them.

iii. Analysing & Building Models

Data preparation is followed by analyzing the data, uncovering hidden trends, and then applying various predictive & classification models to the training set. These included Random Forest, Logistic Regression, Decision tree, KNN, SVM, and a combination of these models to increase the accuracy. Further statistical models and trend analyzer models have been built to increase the accuracy of the ML algorithms for this task.

iv. Merging Models & Accuracy Calculation

Having built various models, then have to test the models on testing set and calculate the savings or loss done on each query put by the user.

Decision Tree Classifier

```
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()

dt.fit(x_train_std, y_train)
x_train.columns
Index(['gender', 'age', 'hypertension', 'heart_disease', 'work_type',
      'Residence_type', 'avg_glucose_level', 'bmi', 'smoking_status'],
      dtype='object')
y_pred_dt = dt.predict(x_test_std)

from sklearn.metrics import accuracy_score

ac_dt = accuracy_score(y_test, y_pred_dt)
print("Accuracy using decision tree classification algorithm = " + str(ac_dt*100) + " %")
Accuracy using decision tree classification algorithm = 92.08173690932313 %
```

DT is used to find a solution to each regression and classification issue by using a class. In addition, since each of the variables serving as input already has a matching variable serving as output, this method constitutes a supervised learning version. It looks like a tree. In the context of this tactic, information is continuously segmented according to a set of criteria. There are two components that make up a decision tree; they are the decision node and the leaf node. At the first node, the data are separated into their component parts, and at the second, the final product is generated. It has been determined that the accuracy of the decision tree classification approach is 92.08 percent.

Logistics Regression

```
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()

lr.fit(x_train_std, y_train)

y_pred_lr = lr.predict(x_test_std)

ac_lr = accuracy_score(y_test, y_pred_lr)
|Accuracy using logistic regression algorithm = 94.89144316730524 %
```

Logistic Regression One of the machine learning algorithms known as LR is the one that gets the most use in the context of the supervised learning technique. This way of learning is also known as "classical" learning. It is a method of forecasting that makes use of a variety of elements that are not connected to one another in order to develop a prediction about a categorical variable. The categorical variable in question is the subject of the prediction. This approach has improved quite a bit. Through the use of logistic regression, one is able to generate hypotheses on the results of an investigation. It's possible that a categorical dependent variable was responsible for these results. As a result of this, the output may either be discrete or categorical depending on the conditions that are present at the time. The possible values may be anything like yes or no, zero or one, true or false, etc.; nonetheless, the range of values is from zero to one. The fact that the applications of logistic regression and linear regression are substantially similar to one another shouldn't come as much of a surprise to anybody. The method of linear regression (LR) is used in order to address the problems associated with classification, and the LR method is also utilised in order to address the issues associated with regression. In this case, the statistical technique known as linear regression is applied.

When using the Logistic Regression classification technique, the accuracy rate is 94.89 percent.

KNN K Nearest Neighbor

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()

knn.fit(x_train_std, y_train)

y_pred_knn = knn.predict(x_test_std)

ac_knn = accuracy_score(y_test, y_pred_knn)
print("Accuracy using k nearest neighbours algorithm = " + str(ac_knn*100) + " %")
Accuracy using k nearest neighbours algorithm = 94.6360153256705 %
```

The K-Nearest Neighbor computation is one of the machine learning algorithms that is considered to be among the most fundamental due to the focus that is placed on proximity in the supervised learning strategy. The NN computation takes into account the degree to which the new case or information is similar to cases that have already been examined and then places the new case in the classification that is most similar to the classifications that have already been examined. The NN computation stores each individual piece of information that can be retrieved and assigns a ranking to the subsequent information point based on how close it is to the most recent one. This suggests that if new information arrives, it has a tendency to be easily described into a good suite categorization by making use of the K-NN approach. This is because of how the K-NN method works. The K-Nearest Neighbor (K-NN) technique can be used for Regression in relation to Classification; however, it is more typically employed for Classification problems. [Case in point:] [Case in point:] [Case in point:] [Case in point:]

Accuracy using the KNN classification algorithm is 94.63%

Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()

rf.fit(x_train_std, y_train)

y_pred_rf = rf.predict(x_test_std)

ac_rf = accuracy_score(y_test, y_pred_rf)
print("Accuracy using random forest classification algorithm = " + str(ac_rf*100) + "%")
Accuracy using RandomForestClassifier algorithm = 94.6360153256705 %
```

The Random Forest RF classifications were decided upon as the most suitable approach to the categorising process. RFs are built from a large number of autonomous decision trees, each of which was trained separately by employing a distinct portion of the available data. These trees are formed during the training process, and the outputs of the decision trees are compiled when training has been completed. Voting is the mechanism that is employed in order to select the ultimate prediction that is created by this algorithm. [Case in point:] [Case in point:] [Case in point:] During this process, each DT is obliged to vote for one of the two output classes. The class that receives the most votes wins (in this case, stroke or no stroke). The approach known as RF, which identifies the category that was awarded the greatest number of votes, is the one that, in the end, determines what the conclusive forecast will be.

Accuracy using the Random Forest classification algorithm is 94.63%

SVM Support Vector Machine

The Support Vector Machine, often known as SVM, is a well-known kind of computation used in Supervised Learning. This type of calculation may be used to problems involving Classification as well as Regression. On the other hand, its primary use is in the field of machine learning, namely for classification problems. Calculating the optimal line or choice limit that can partition an n-layered space into classes is the purpose of the SVM computation. This is done so that we may easily place newly acquired data in the appropriate classification in the future without too much of a struggle. The term "hyperplane" refers to this constraint on the greatest option.

Accuracy using the SVM classification algorithm is 92.63%

```
from sklearn.svm import SVC
svc = SVC()

svc.fit(x_train_std, y_train)

y_pred_svc = svc.predict(x_test_std)

ac_svc = accuracy_score(y_test, y_pred_svc)
print("Accuracy using support vector classification algorithm = " + str(ac_svc*100) + " %")
Accuracy using RandomForestClassifier algorithm = 92.6360153256705 %
```

6.1.1 Code for Stroke Prediction

Source codes are instructions that are readable and understandable by humans. It can be considered as the instructions for the computer to perform some tasks. Source codes are generally written by programmers using various tools and specific coding languages. To make the computers understand the code written by the humans, they are compiled using the compiler tools. Once the source code turns into the code which is understood by the machines, it's said to be the object code. Object code generally consists of 1s and 0s and this code is not understandable by humans. Source code can be in various programming languages like C, C++, java, python, and others.

Install Flask

```
from flask import Flask, render_template, request
import os, joblib, pickle
import numpy as np
```

Flask is a web application system written in Python. Flask depends on the WSGI tool compartment and the Jinja2 format engine. Flask is a web structure, a Python module that allows you to foster web applications without any problem. It's having a little and simple to-broaden centre it's a microframework that does exclude an ORM (Object Relational Manager) or such highlights.

Steps to install flask

Install pip flask in terminal

Run the application in terminal

Launch the application in the chrome browser

Attributes for Prediction

```
gender = int(request.form['gender'])
age = int(request.form['age'])
hypertension = int(request.form['hypertension'])
heart_disease = int(request.form['heart_disease'])
work_type = int(request.form['work_type'])
Residence_type = int(request.form['Residence_type'])
avg_glucose_level = float(request.form['avg_glucose_level'])
bmi = float(request.form['bmi'])
smoking_status = int(request.form['smoking_status'])
```

The Above code represents what are the attributes used in the stroke prediction application using these attributes we can predict whether a person has the symptoms of a stroke. These are the attributes that are commonly used to determine the Stroke in the early stage. The attributes are stored in the dataset which is trained and tested using the Five types of algorithms used are explained above

Pickle file

```
lr = pickle.load(open('finalized_model.pkl', 'rb'))
y_pred_lr = lr.predict(x)
```

A Python object structure may be serialised and deserialized with the help of the pickle module, which is part of Python. Any Python article has the potential to be salted with the intention that it might potentially save the user money on the circle. First, it "serialises" the article, and then it composes it such that it may be recorded. This is what pickle does. A python object (list, dict, and so on) may be transformed into a person stream by the use of a procedure known as pickling. This character stream, it is presumed, provides all of the data necessary to recreate the object in another Python script.

Dataset

Table 6.1 Dataset

1	gender	age	hypertensi	heart_dise	ever_marr	work_type	Residence	avg_glucose	bmi	smoking_s	stroke
2	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly sm	1
3	Female	61	0	0	Yes	Self-emplo	Rural	202.21	N/A	never smo	1
4	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smo	1
5	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
6	Female	79	1	0	Yes	Self-emplo	Rural	174.12	24	never smo	1
7	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly sm	1
8	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smo	1
9	Female	69	0	0	No	Private	Urban	94.39	22.8	never smo	1
10	Female	59	0	0	Yes	Private	Rural	76.15	N/A	Unknown	1
11	Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
12	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smo	1
13	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1

The Dataset used to predict the Early identification of stroke the dataset contains more than 5000 values which include age, gender, hypertension, married status, work type, and residence. Ang glucose level, heart disease, smoking status. All these Dataset are trained and tested using the Five algorithms to predict the accurate accuracy amongst them Random Forest classifier is proven to be the best which gives high accuracy rates compared to other algorithms

Index Page

```

</div>
<div class="collapse navbar-collapse navbar-right" id="myNavbar">
  <ul class="nav navbar-nav">
    <li class="active"><a href="#banner">Home</a></li>
    <li class=""><a href="#service">Services</a></li>
    <li class=""><a href="#about">About</a></li>
    <li class=""><a href="#testimonial">Health Topics</a></li>
    <li class=""><a href="#contact">Contact</a></li>
    <li class=""><a href="/home">Prediction</a></li>
  </ul>
</div>
</div>
</div>
</nav>

```

The above code explains the index page when a user visits the page for the first time the lands in the Index where the user gets to know about the application like what are the features in the application like the user has options to explore About us, Contact, Services, Prediction and Book appointment Pages where he can visit each page and check out what are the information provided. The user interface is designed simply for users to use and navigate from one page to another page

Services

```
<section id="service" class="section-padding">
  <div class="container">
    <div class="row">
      <div class="col-md-20 col-sm-6">
        <h2 class="ser-title">Our Service</h2>
        <hr class="botm-line">
        <p></p>
      <!-- <div class="col-md-5 col-sm-4">
        <div class="service-info">
          <div class="icon">
            <i class="fa fa-stethoscope"></i>
          </div>
          <div class="icon-info">
            <h4>24 Hour Support</h4>
            <p>Operates rapid diagnostic and laboratory procedures 24x7x365.</p>
          </div>
        </div>
        <div class="service-info">
          <div class="icon">
            <i class="fa fa-ambulance"></i>
          </div>
          <div class="icon-info">
            <h4>Emergency Services</h4>
            <p>The situation is further complicated by people's lack of awareness on stroke symptoms. People suffering from stroke ;
          </p>
          </div>
        </div>
      </div>
    </div>
  </div>
```

The Services page provides the information about what are the services provided by the health Care before landing into the services page the user logs into the index button where the services page button is provided when the user clicks the page he lands on this page. The user knows about the services of health the care system such as 24x7 ambulance services, Emergency Operation Ambulance service services provided by the health care centre.

Stroke

```
<div class="jumbotron hero-nature">
  <h1 class="hero-title">You have been diagnosed with Stroke Risk</h1>
  <p>The model has diagnosed you with Stroke Risk based on your Information. Please consult a Physician Immediately.</p>
  <p>Take Care :) </p>
</div>
```

This Page Displays when The User is Having the chances of Stroke when the user logs into the index page there the prediction option is available when the user clicks into the but that

the loan into the prediction page there the user needs to enter the values which have been asked after entering the value. The values are analyzed by Machine Learning Algorithms using this if the user has symptoms of stroke this page is displayed

```
<div class="jumbotron hero-nature">
  <h1 class="hero-title">You have been diagnosed with no Stroke Risk...!</h1>
  <p>The model has diagnosed you with No Stroke Risk based on your information. However it might be better to talk to a Physician regardless during proper interval.</p>
  <p>Take Care Stay Healthy :)</p>
```

This Page Displays when The User is Having the chances of Stroke when the user logs into the index page there the prediction option is available when the user clicks the button that lands on the prediction page there the user needs to enter the values which have been asked after entering the value. The values are analysed using Machine Learning Algorithms using this if the user has no symptoms of stroke this page is displayed

6.1.2 Source Code for Visitor Management

```
<!DOCTYPE html>
<html>
  <head>
    <title>Admin Login</title>
    <link
      rel="stylesheet"
      href="https://stackpath.bootstrapcdn.com/bootstrap/4.1.3/css/bootstrap.min.css"
      integrity="sha384-MCw98/SFnGE8fJT3GXwEOngsV7Zt27NXFoaoApmYm81iuXoPkFOJwJ8ERdknLPMO"
      crossorigin="anonymous"
    />
    <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.3.1/jquery.min.js"></script>
    <link
      rel="stylesheet"
      href="https://use.fontawesome.com/releases/v5.6.1/css/all.css"
      integrity="sha384-gfdkjb5BdAXd+lj+gudLWI+BXq4IuLW5IT+brZEZsLFm++aCMIF1V92rMkPaX4PP"
      crossorigin="anonymous"
    />
  </head>

  <body>
    <div class="container h-100">
      <div class="d-flex justify-content-center h-100">
        <div class="user_card mb-5">
          <div class="d-flex justify-content-center">
            <div class="brand_logo_container">
              
            </div>
          </div>
          <div class="d-flex justify-content-center form_container">
            <form action="/admin_login/" method="post">
```

The above code represents the login Page for the admin where the admin needs the enter the credentials provided if he enters the wrong id and password it prompts, please enter a

valid id and the password. After entering the user and password the admin has the access to add doctors and visitors and manage appointments. The admin has the option the add delete and modifies both the Visitors and the Doctors.

6.2. Screenshots

6.2.1 Screenshots for Prediction of Stroke

The screenshot shows what is displayed on the computer screen when we open certain functionality. In this section, the screenshot represents the capture of all the important software aspects that we have used to build. Mainly the screenshots contain the website layout. Each image in this section explains about different functionality of the website. Using the images to explain the website makes the user completely understand the system in a detailed and easy way. All the images are displayed here to show all the functionalities such as the login page for producers, users, admins, etc. It shows the product page, order page, etc. Similarly, it shows all the pages available on the website from the user's concern.



Fig 6.1 Home Page

The above-mentioned Screenshot is from the home page of Stroke Prediction. When the user visits the website, he lands on the home page. The User gets to know about the Page and the functionality provided on the page. At the right top corner of the website we have the buttons like Home, About, Health Topics, Contact, Prediction, and Book Appointment which have their functionalities. When a user clicks on the home page, he again lands on the same page itself. . When the button Contact is clicked, this navigates the user to the

contact page which has information about the Medi lab Plus contact details. And lastly, the Book Appointment button navigates to the Visitor Management page where the user can Search For Doctors And Book The Appointment According to Availability.

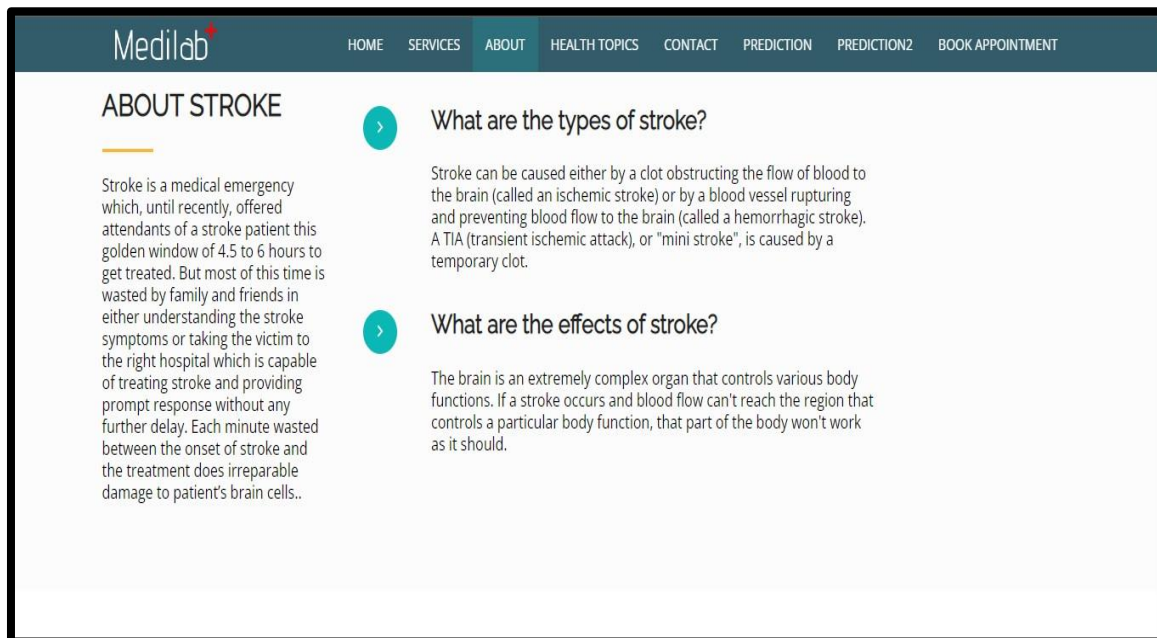


Fig 6.2 About Us

The above-mentioned Screenshot is from the About Us of Stroke Prediction. When The User enters the Home Page, he can view the About Us button Provided. The About Us page provides information about stroke-like how stroke occurs, what are the types of stroke-like ischemic stroke or hemorrhagic, and what are the effects of the Stroke. The User can view and can get the information about the Stroke

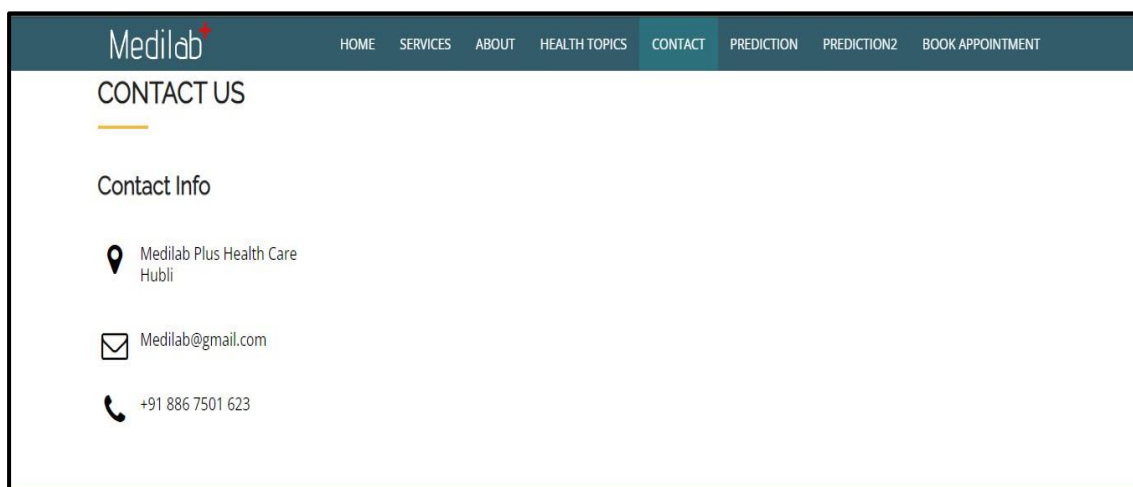
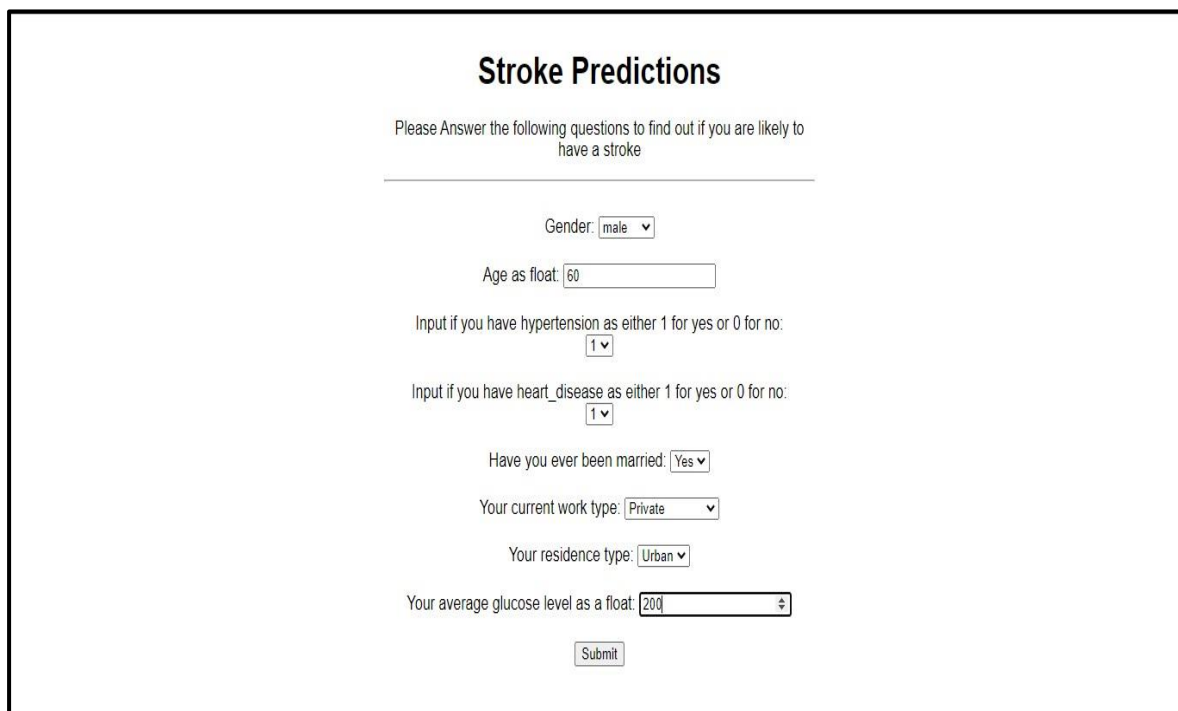


Fig 6.3: Contact

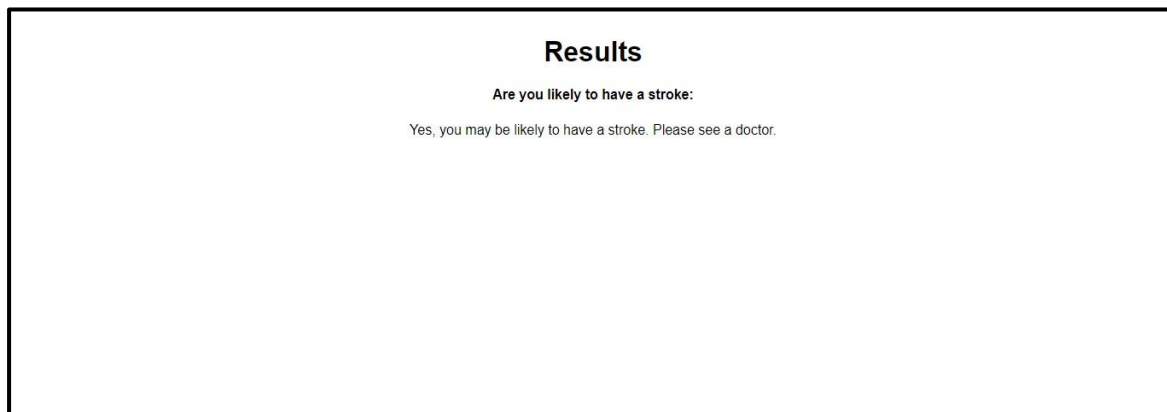
The Contact page provides the contact information of the health care centre such as the location of the centre, Phone number in case of emergency, and Emails address, and Useful social media links



The screenshot shows a web form titled "Stroke Predictions". Below the title is a subtitle: "Please Answer the following questions to find out if you are likely to have a stroke". The form contains several input fields and dropdown menus: "Gender" with a dropdown menu showing "male"; "Age as float" with a text input field containing "60"; "Input if you have hypertension as either 1 for yes or 0 for no:" with a dropdown menu showing "1"; "Input if you have heart_disease as either 1 for yes or 0 for no:" with a dropdown menu showing "1"; "Have you ever been married:" with a dropdown menu showing "Yes"; "Your current work type:" with a dropdown menu showing "Private"; "Your residence type:" with a dropdown menu showing "Urban"; and "Your average glucose level as a float:" with a text input field containing "200". At the bottom of the form is a "Submit" button.

Fig 6.4: Prediction Page

The user lands on the stroke prediction home page where he can view the Prediction button when the user clicks the button he lands on the prediction where he needs to enter the information about the gender of the User, Age, If the User has Hypertension(BP) he/she needs to click 1 if not 0 If the User has heart Disease he/she needs to click 1 if not 0, Next if the user is married he needs to click 1 if not 0 The user needs to specify the work type such as if he works in a private company, Government Employee, Self Employed, Never Worked, Next, the User information about where he lives if the user lives in an Urban area or Rural area according to his locality the user can choose the option The last user has to give the information about His Glucose Level(Blood Sugar) accurate value for better Prediction and Finally press the Submit button after pressing it he will be landed to Result in Page.



6.5 Result Page Having a Stroke

The Result page provides the information on whether The User is Having a stroke it displays “Yes you may be having a stroke Please see a doctor” the Page gives information user having a stroke. The stroke prediction is done using the machine learning algorithms such as Random Forest classification, SVM, Logistics Regression, KNN, and Decision Tree using these 5 algorithms the prediction is done

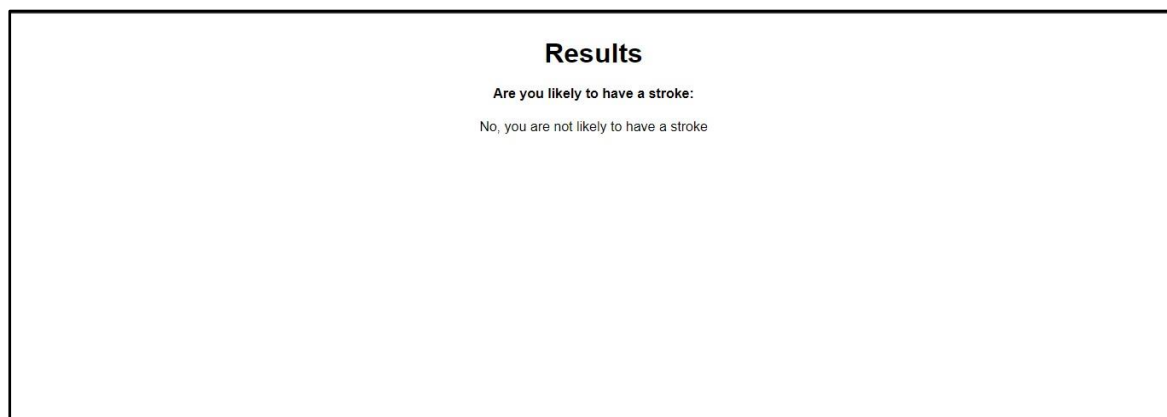


Fig 6.6Result Page Having a No Stroke

The Result page provides the information on whether The User is Having a stroke it displays

“No you are not likely to have a stroke Please see a doctor” the Page gives information user having a stroke. The stroke prediction is done using the machine learning algorithms such as Random Forest classification, SVM, Logistics Regression, KNN, and Decision Tree using these 5 algorithms the prediction is done

6.2.2 Screenshots for Visitor Management

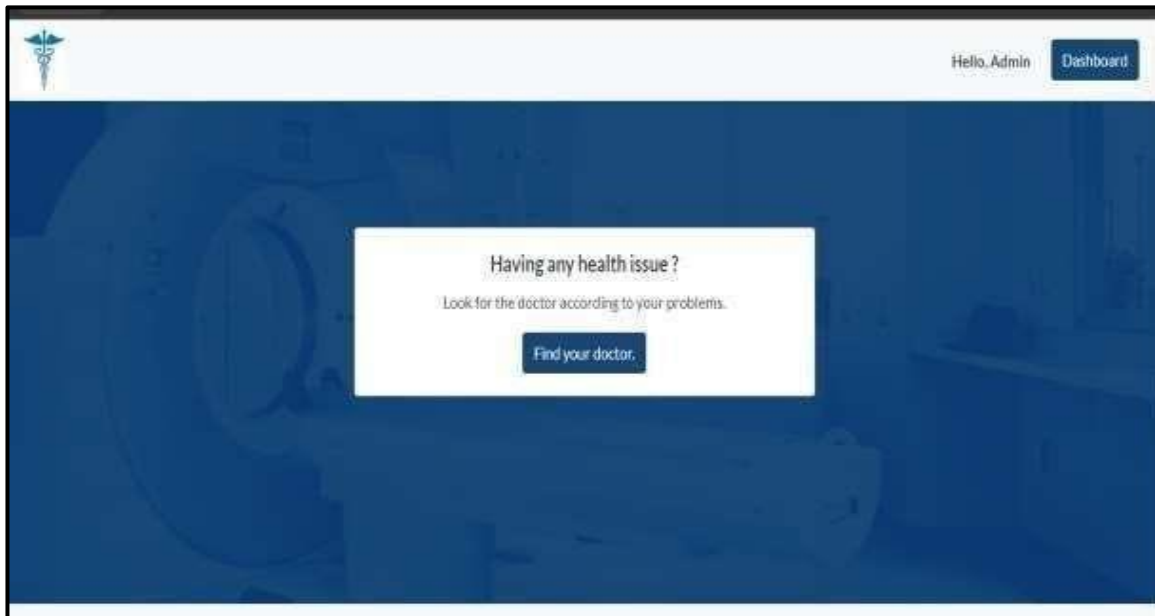


Fig 6.7.:Home Page

This is the Home Page illustrates the home page these many options are available like Dashboard Find Doctors Admin login. When the new Visitor can visit the Login through Dashboard and can log in through his login credentials and can book an appointment with the visitor on this page, only the admin can log in click on the admin option then login next they go to the administrative block



Fig 6.8 Login Page

This is a login page for the customers and the vendors who transact via bank accounts. So, the Login page has two options. One is for user login and another for vendor. This is an important page because the authentication of the vendor is done here. If anyhow, an attacker enters the login page, it would be dangerous. So, its security is the highest priority. This login page takes a strong password.

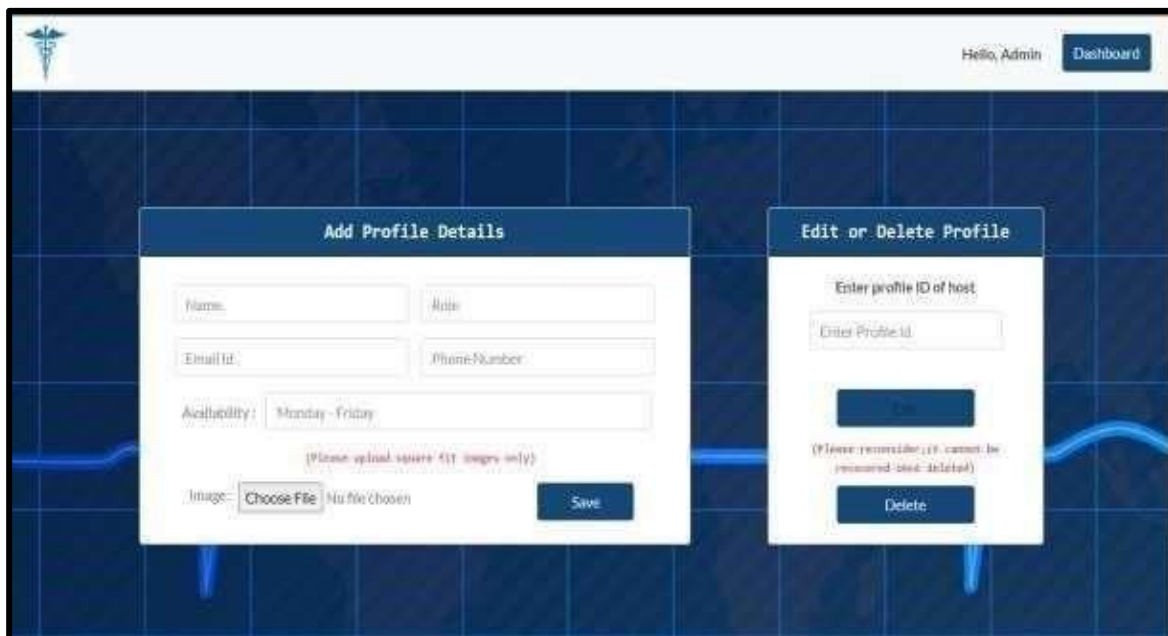
The screenshot shows a web application interface with a dark blue background and a grid pattern. At the top, there is a header bar with a logo on the left, the text 'Hello, Admin' in the center, and a 'Dashboard' button on the right. Below the header, there are two main white panels. The left panel is titled 'Add Profile Details' and contains several input fields: 'Name', 'Role', 'Email id', 'Phone Number', and 'Availability' (with a dropdown menu showing 'Monday - Friday'). There is also an 'Image' field with a 'Choose File' button and a 'No file chosen' status. A 'Save' button is at the bottom right of this panel. The right panel is titled 'Edit or Delete Profile' and contains an 'Enter profile ID of host' field, an 'Enter Profile Id' input field, a 'Save' button, and a 'Delete' button. Both panels have a red error message: '(Please upload square 4:3 images only)' and '(Please provide id; it cannot be removed once deleted)'.

Fig 6.9 Profile

This is the Profile page Where the Host can add their Profile Details Such as Name, Role, Email id, Phone Number, Availability, and Image and can save. The Profile is visible in the Dashboard which is can be accessed by both the Host and the Visitor.

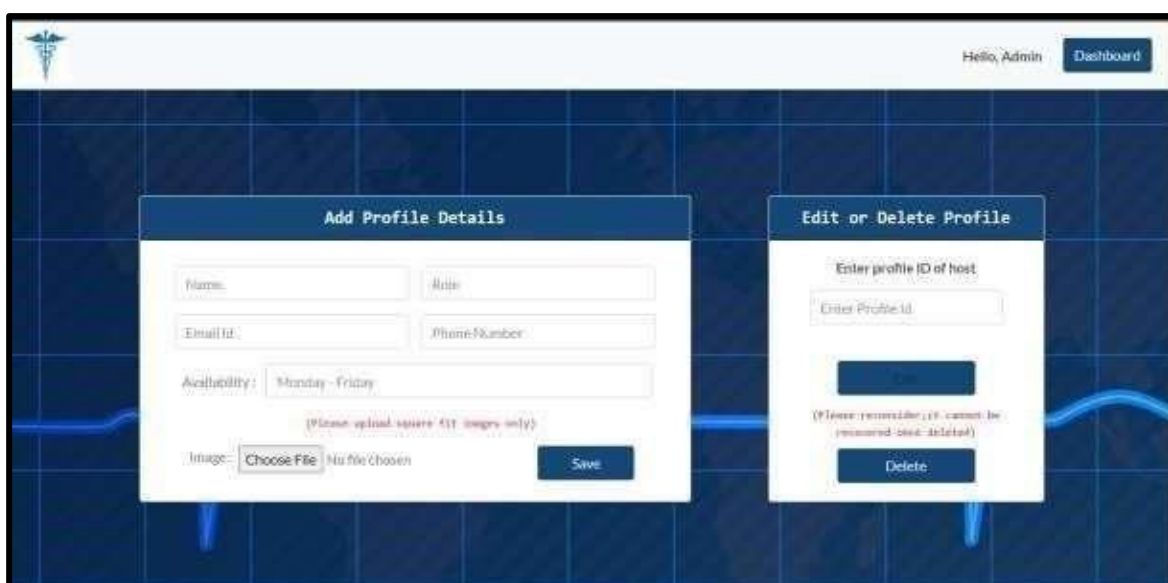
This screenshot is identical to the one in Fig 6.9, showing the same web application interface with the 'Add Profile Details' and 'Edit or Delete Profile' forms. The layout, including the header, grid background, and form elements, is the same.

Fig 6.10 Dashboard

This is the Dashboard Where the Host and Visitors Can access. The Visitor Can view the Doctor's Profile and can Arrange a meeting with the doctor With the Profile Id Available. The Host can check the visitor's profile and can manage the meeting history. The Dashboard has the profile Manager where he can add his details and his profile id will be generated which can be useful for users to select the visitors.

Summary

The software Implementation part has been described in the chapter. It refers to the process of accommodating software applications to check their workflow and adaptability capacity. It is all about executing plans to check the developed software. We have got the explanation about the source code, its meaning, and the explanation, proceeding with this we have the code which we have written to build the software. These codes have been written using the Python language. Next to this, we have the screenshots section which has got the screenshot of

the system we have built and explains the functionalities of the system. Each screenshot has its functionality to describe, seeing these pictures we can easily understand the functionality of each web page of the system. Various pictures are elaborating the pages such as the contact page, login page for the users and producers, admin dashboard, home pages, product tracking page, product analysis page, feedback page, etc. These will help the readers understand the system in a better manner.

CHAPTER 7

SOFTWARE TESTING

The testing step is very vital to the development of a successful system. In this step, before putting the new system into operation, a test run of the system is carried out with the goal of finding and removing any faults. After the coding for all of the system's applications has been finished, a test strategy has to be devised, and then one particular piece of test data needs to be executed.

7.1 Testing Approach

The testing approach's goal is to detect code faults in functions or units of work. In terms of flaw identification, testing is the most efficient, effective, and cost-effective step. Because there are several browser versions, browser compatibility must be maintained to ensure that the system is appropriately shown across all browsers. IE Tester may be used to compare the impact of different versions of Internet Explorer on the same machine. To test the system, browsers such as Firefox, Google, and Opera are installed. Finally, all of the exams have passed. Internal logic, error handling, field processing, computational calculations, data handling access, range of potential inputs and outputs, and cross-field editing are all validated using the unit test technical method. Using the application documentation as a reference, the developer walks through the code he or she has created, ensuring that all requirements and coding standards have been followed. Best practises dictate that the developer will have peers examine his or her code. The technical aspects of the software should be covered in unit test plans. Expected outcomes concentrate on internal code, as well as inputs and outputs from the code. In-depth, high-quality unit testing reduces the failure rate, which shortens subsequent testing stages. The goal is to ensure that each component of the programme works as intended. Individual pieces of source code are examined to see whether they are appropriate for usage using unit testing. A unit is the smallest testable component of a programme, which includes functions, procedures, classes, and interfaces. Software engineers generally write and run unit tests to confirm that code conforms to its design and works as expected. The purpose of unit testing is to isolate each component of the software and demonstrate that each component is accurate. Unit tests detect flaws early in the development process.

7.1.1 Unit Testing

Using unit testing, individual bits of source code are investigated to see whether or not they satisfy the criteria necessary for deployment. Functions, procedures, classes, and interfaces are all considered to be part of a program's units. A unit is the smallest tested component of a programme. Unit tests are routinely written and executed by software developers to ensure that code is working as intended and adheres to the design of the system. The aim of unit testing is to verify that each component of the programme is correct by isolating each component of the product and testing it individually. Early on in the development process, problems are uncovered via unit testing. When all of the modules have been created, we start the unit testing process. Through unit testing, we guarantee that each module operates appropriately when used alone. In the event that the unit testing turns up no issues, we will go on to the integration testing. During the stage of the software testing process known as "unit testing," the individual units or components of a programme or system are put through their paces. It is important to check and double-check that every aspect of the application is operating as it should. A unit is the smallest component of a piece of software that can be tested. In most cases, it will only have a single output and one or more inputs. In procedural programming, a unit may be anything from a single programme to a function to a method or perhaps something else entirely. Methods are the fundamental building blocks of object-oriented programming. They may be a part of a base/super class, an abstract class, or a derived/child class. Methods can also stand alone. The practise of examining a single, low-level programme on its own in isolation prior to combining it with further units is referred to as unit testing. Testing at the unit level guarantees that the code object satisfies the requirements of the technical design, that data is handled appropriately, and that all of the product's conceivable workflows are properly executed. This testing is typically carried out on the workstations used by the developers, and it is afterwards replicated in the environment used for server-based development. During this phase, there is likely to be some minimal interfacing of components performed by a developer.

7.1.2 Integration Testing

In the software testing paradigm, integration testing happens before system testing and after unit testing. Integration testing works by taking separate modules that have passed through the unit testing process and merging them into a group. The integration testing process will ensure that any issues, such as mistakes or defects, that arise as a result of module integration

are removed. Integration testing is concerned with the integration of processes within the system rather than the integration of the whole system. In the integration testing step, three things are built to guarantee that the module integration is successful and that it works smoothly: a test plan, test cases, and test data are created to properly test that the integration is successful. Test data is often utilised by test cases, but I've included each kind here.

Plan for Integration Testing: Integration testing examines component integration or interfaces, interactions with other portions of the system such as the operating system, file system, and hardware, and system interfaces. Integration testing is carried out by a dedicated integration tester or test team. Following unit testing, we incorporate several components. Then we ensure that the combined elements operate correctly together. If the project functions successfully after integration, we proceed to system testing.

7.1.3 Integration Test Cases

In order to ensure that the output of the combined modules is creating the expected output and is functioning exactly as it should, test cases are developed. These test cases are run to validate that the output is behaving as expected. This is an effective method for locating any bugs or faults that may have been introduced during the integration phase of the process. The tester will next go through this system in order to submit all of the data using the test case that was prepared. The test case will look at all inputs and outputs inside the integrated modules. The following is an example of a simple test case that I developed for a programme that I wrote when I was a student. In addition, I have included some test statistics inside the test case itself in order to demonstrate precisely how the programme operates. If you are working with a sizable application or programme, then it is possible that you will need to develop a great number of test cases in order to verify the functionality of the various components of the application. The many test cases are often grouped together and given the name test suites, which is short for test suite, which is a collection of test instances.

7.1.4 Test cases

A test case refers to the file or the document which has been designed for a specific test scenario. The document has got certain details such as the conditions which are required for pre-testing, the data used for the test, the expected results, etc. They are considered the initial point for the execution of the test. After checking and trying out various kinds of values, the

condition that is left out is said to be the post condition of the test cases. There are various kinds of configurations required for the test case such as the id of the test case, description of the test case, test data, steps, the result that is expected, etc. The test case is written for various reasons. To resist giving training to every new test engineer who joins the organization, the test cases are written so that he can easily understand them. It is also written when the test case is more dependent on some procedure instead of the people who perform that. Which means that there is a particular standard procedure to do this. Apart from these, there are still many advantages to writing test cases.

7.2 Test Case for Admin

Table 7.1 Test case for Admin

Test Case Number	Test Case	Expected Input	Expected Output	Result
1	Open The Project on Desktop	Click to open Project	Project Opens	Pass
2	Run the Project	Launch the project in terminal	Launches in Terminal	Pass
3	Run The HTTP address	Copy the address and run it in chrome	Runs in chrome	Pass
4	Launch The application	Opens in Chrome	Opens in chrome	Pass

Table 7.2 Test case for Admin

Test Case Number	Test Case	Expected Input	Expected Output	Result
1	Launch Application	Home page	Home page	Pass

2	Open Prediction page	Click to open the prediction page	Prediction Page Opens	Pass
3	Open Contact page	Click to open Contact page	Contact Page Opens	Pass
4	Open About Us page	Click to open the About page	About us page Opens	Pass
5	Open Services Page	Click to Open Services Page	Services Page Opens	Pass
6	Open Book Appointment	Click to Open Book Appointment	Book Appointment Opens	Pass

7.2.1 Test Case for Prediction

Table 7.3 Test Case for Prediction

Test Case Number	Test Case	Expected Input	Expected Output	Result
1	Test case to choose the gender	Select Gender	Gender Selected	Pass
2	Test case to choose Age	Select age	Age Selected	Pass
3	Test case to choose Hypertension	If yes enter 1 if no enters 0	Hypertension Selected	Pass
4	Test case to choose Hypertension	Enters 2	Prompts Message, please select valid input	Pass

5	Test case to choose Marital Status	If married yes if not select no	Input selected	Pass
6	Test case to choose Work type	Select work type private, govt job, self-employed	Input Selected	Pass
7	Test case to choose heart disease	If yes enter 1 if no enters 0	Input Selected	Pass
8	Test case to choose heart disease	Enters 2	Prompts Message, please select valid input	Pass
9	Test case to choose residence type	Select Urban or rural	Input Selected	Pass
10	Test case to choose Glucose level	Enter vale	Input Selected	Pass
11	Choose Submit button	Enters submit button redirects to Prediction Page	Input Selected	Pass
12	Prediction Page	Shows yes if the person has symptoms	Input Selected	Pass
13	Prediction Page	Shows no if the person no has symptoms	Input Selected	Pass

7.3 Test Case for Visitor Management Admin Test Cases

Table 7.4 Test Case for Visitor Management Admin Test Cases

Test Case Number	Test Case	Expected Input	Expected Output	Result
1	Launch Application	Dashboard	Dashboard displayed	Pass
2	Enter Credentials	Enters login id and password	Login in successful	Pass
3	Create Doctors Profile	Enter the doctor's name, id, etc	Entered successful	Pass
4	Delete Doctors Profile	Deletes doctor's profile through ID	Deleted successful	Pass
5	Visitors	View visitors list	Successful	Pass
6	View Dashboard	View Doctor's and visitors' appointment	Successful	Pass
7	Logout	Logout	Logout Successful	Pass

7.3.1 Test Case for Visitor Management Visitor Test Cases

Table 7.5 Test Case for Visitor Management Visitor Test Cases

Test Case Number	Test Case	Expected Input	Expected Output	Result
1	Visits application	Dashboard	Successful	Pass
2	View Doctors	Views Doctors in Profile sections	Clicks Profile button	Pass
3	Checks for availability	Checks for availability in profile sections	Input Selected	Pass
4	Book Appointment	Books Doctors appointment according to availability	Appointment Booked	Pass
5	Confirms Booking	Confirms booking in dashboard	Confirm successfully	Pass

Summary

This chapter is all about Software testing which is the last step of system development. This process makes sure that the system meets all the user requirements and has been developed without any errors or defects. The goal here is to recognize the errors. Here we have the test cases which refer to the file or the document that has been designed for a specific scenario. Test cases include admin test cases, product test cases, and customer test cases.

CHAPTER 8

CONCLUSION AND FUTURE ENHANCEMENTS

8.1 CONCLUSION

When it comes to identifying people who are at risk of having a stroke, a number of assessments and prediction models, including as Decision Tree, Random Forest, KNN, and SVM, performed well. Because of this, our research is able to forecast the likelihood of having a stroke by using a prediction model, and it also provides tailored warnings and lifestyle corrective messages through a web application. This motivates people who use medicinal products to enhance their motivation for health management and to improve the ways in which they behave in regards to their health.

Stroke is a serious medical illness that may result in death and has to be treated as quickly as possible in order to avoid further complications. The development of a model using machine learning might be of assistance in the early detection of strokes, which would then lead to a

decrease in the severity of the stroke's aftereffects. In this study, a number of different machine learning algorithms are compared against one another to see which is better at accurately predicting strokes based on a wide range of physiological parameters. The random forest classification method is superior to the other approaches that were investigated because to its superior accuracy of classification, which stands at 96%. According to the findings of the research, the random forest approach performs better than other algorithms in the prediction of brain strokes when cross-validation measures are used. In the future, the framework models may be enhanced by making use of a more extensive dataset in conjunction with other machine learning models, such as AdaBoost, SVM, and Bagging. Both the dependability and the aesthetics of the framework will be enhanced as a result of this change. In exchange for providing just some fundamental information, the general public may get assistance from the machine learning architecture in determining the likelihood of an adult patient experiencing a stroke in the future. In a perfect world, it would let patients get early treatment for strokes and start reconstructing their lives following the event.

In order to improve it, we may train models to forecast the sorts of stroke disorders and provide advice to users, as well as employ more advanced models.

8.2 FUTURE ENHANCEMENTS

The Future Enhancement for “Early Identification of Stroke” is to predict after stroke once stroke occurs the chances of recovery. Users often want to know how the recovery could progress and what might it look like in the future. Right now, there are not many tools available to help health professionals to predict the recovery of stroke.

There are some ways to prevent the stroke which can be Implemented in the Future

- Collect The Information about what type of stroke the people are suffering most
- Understanding what are the problem faced by the individuals after a stroke occurs
- Analysing The most Common Symptoms for the cause of Stroke To prevent it in the early Stage
- Understanding what are the group of age and the gender who are getting affected by Stroke
- To prepare a model and predict are effects after stroke by collecting information available from the previously affected stroke patients

A person who has had a stroke will have difficulty moving their hands and arms because of the effects of the stroke on their mobility. The recovery of stroke victims who only had modest problems may be predicted, but it is more difficult to determine how badly they were impacted by the stroke. In the future, we will hopefully have a better understanding of the processes of recovery, which will allow us to evaluate and forecast which therapies will be most effective for stroke survivors. Stroke survivors typically report experiencing changes in their mood as well as their way of thinking. They may find it harder to concentrate or worry more about their loss of self-assurance. As time passes, this problem may become better, but it also has the potential to develop worse. Because we don't have a lot of knowledge on this issue, the study that has been done so far has only followed individuals for a brief period of time following the event. It's possible that this issue may evolve over time, which would allow us to better predict how the survivors would react to therapy and how quickly they will recover. Stroke survivors and their families will be able to better prepare for the future if they have a better understanding of the impact that these impacts have on people's lives and recoveries, and scientists will be able to create more effective therapies for strokes at an earlier stage.

Appendix A

BIBLIOGRAPHY

- [1] Chang, C., Cheng, M. & Ma, M., 2018. Application of Machine Learning for Facial Stroke Detection. In 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), 23(1), pp. 1-5. Vijayalakshmi, V., Babu, M. & Lakshmi, R., 2018. KFCM Algorithm for Effective Brain Stroke Detection through SVM Classifier. IEEE International Conference on System, Computation, Automation, and Networking (ICSCA), pp. 1-6.
- [2] Jeon, S. et al., 2018. RISK-Sleep: Real-time Stroke Early Detection System During Sleep Using Wristbands. IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 4333-4339.
- [3] Singh, M. & Choudhary, P., 2017. Stroke Prediction using Artificial Intelligence. Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), 8(1), pp. 158-161.
- [4] Jeena, R. & Kumar, S., 2016. Stroke Prediction Using SVM. International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 600-602.
- [5] García-Terriza, L. et al., 2019. Comparison of different machine learning approaches to model stroke subtype classification and risk prediction. Institute of Electrical and Electronics Engineers (IEEE), pp. 1-10.
- [6] Gulenko, A. et al., 2016. Evaluating machine learning algorithms for anomaly detection in clouds. In 2016 IEEE International Conference on Big Data (Big Data), pp. 2716-2721. Guo, X. et al., 2008. On the Class Imbalance Problem. Fourth international conference on natural computation, 4(1), pp. 192-201.
- [7] M. S. Zulfikar, N. Kabir, A. A. Biswas, P. Chakraborty, and M. M. Rahman, "Predicting students' performance of the private universities of Bangladesh using machine learning approaches," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020
- [8] Jeena R.S and Dr.Sukesh Kumar "Stroke prediction using SVM", International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2016.
- [9] Benjamin Letham, Cynthia Rudin Tyler, H. McCormick and David Madigan "An interpretable model for stroke prediction using Bayesian analysis", *The Annals of Applied Statistics* 2015, Vol. 9, No. 3, 1350–1371, Institute of Mathematical Statistics, IEEE, 5 November 2015.

-
- [10] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Classification of stroke disease using machine learning algorithms," *Neural Computing & Applications*, vol. 32, no. 3, pp. 817–828, 2020.
- [14] <https://en.wikipedia.org/wiki/NumPy>
- [15] <https://docs.python.org/3/library/pickle.html>
- [16] <https://www.geeksforgeeks.org/python-introduction-matplotlib/>
- [17] <https://www.ibm.com/in-en/topics/logistic-regression>
- [18] <https://www.javatpoint.com/machine-learning-random-forest-algorithm>
- [19] <https://pypi.org/project/Flask/>
- [20] <https://djangopackages.org/>
- [21] https://www.tutorialspoint.com/sqlite/sqlite_python.htm#:~:text=SQLite3%20can%20be%20integrated%20with,x%20onwards.
- [22] <https://www.python.org/>
- [23] <https://www.geeksforgeeks.org/html-responsive-modal-login-form/>
- [24] <https://code.visualstudio.com/learn>
- [25] <https://jupyter.org/>

Acronyms used

1. KNN - K-Nearest Neighbors.
2. SVM - Support Vector Machine.
3. ANN - Artificial Neural Network.
4. CNN - Convolutional neural networks.
5. CHS - Cardiovascular Health Study.
6. ECG – Electrocardiography.
7. ANFIS - Adaptive Neuro Fuzzy Inference System.
8. PCA - Principal Component Analysis.
9. HTML - Hypertext Markup Language.
10. CSS - Cascading Style Sheets.
11. XML - Extensible Markup Language.
12. WSGI - Web Server Gateway Interface.
13. HTTP - Hypertext Transfer Protocol.
14. ML – Machine Learning.
15. AI - Artificial intelligence.
16. APIs - Application Programming Interface.
17. MATLAB - Matrix Laboratory
18. BMI – Body Mass Index
19. AUC - Area Under the curve
20. BRL - Bayesian Rule Lists
21. DFD - Data Flow Diagram.
22. UML - Unified Modelling Language.
23. SRS - software requirements specification.