A project report on

Brain Tumor Detection Using Machine Learning

Submitted By

Mr. Purva Prasad Ghade

M.Sc. (Computer Science) Semester-IV

Submitted to

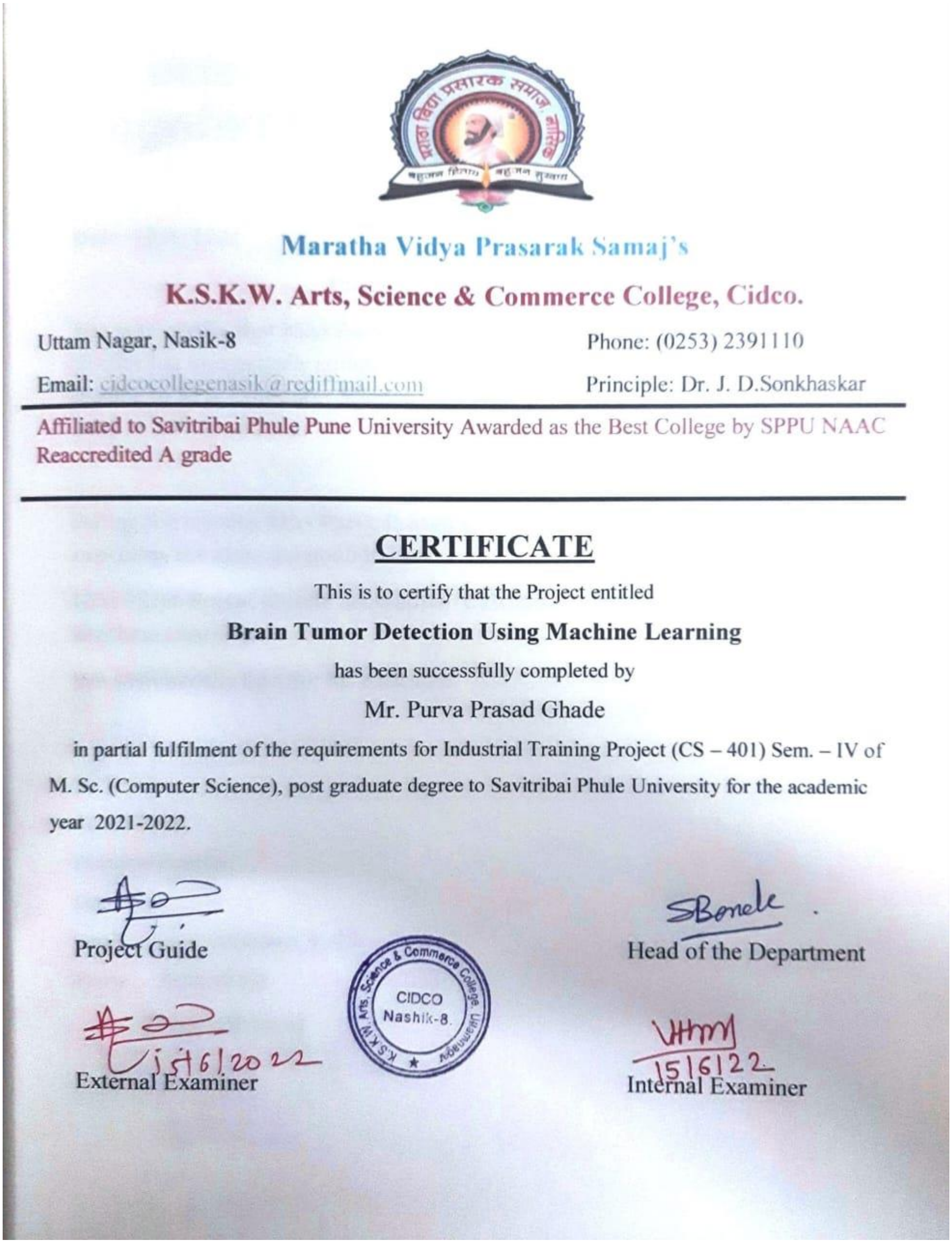
Savitribai Phule Pune University

for partial fulfillment of the requirement of M.Sc(Computer Science)

K.S.K.W. Art's, Commerce & Science College Cidco, Nashik

Guided by :

Smt.A.S.Bachhav



Maratha Vidya Prasarak Samaj's

K.S.K.W. Arts, Science & Commerce College, Cidco.

Uttam Nagar, Nasik-8 Phone: (0253) 2391110

Email: cidcocollegenasik@rediffmail.com Principle: Dr. J. D.Sonkhaskar

Affiliated to Savitribai Phule Pune University Awarded as the Best College by SPPU NAAC Reaccredited A grade

CERTIFICATE

This is to certify that the Project entitled

Brain Tumor Detection Using Machine Learning

has been successfully completed by

Mr. Purva Prasad Ghade

in partial fulfilment of the requirements for Industrial Training Project (CS – 401) Sem. – IV of M. Sc. (Computer Science), post graduate degree to Savitribai Phule University for the academic year 2021-2022.

Project Guide

SBonde
Head of the Department

K.S.K.W. Arts, Science & Commerce College, Uttamnagar
CIDCO
Nashik-8.

15/6/2022
External Examiner

15/6/22
Internal Examiner

Acknowledgement

We are happy to present the project “**Brain Tumor Detection Using Machine Learning”**in Python ML. A project titled “**Brain Tumor Detection Using Machine Learning**” would not have been completed without the valuable guidance and encouragement of **Smt A S Bachhav** and all staff members. We acknowledge them for their moral support.

This project is substantially upgrading our skill of software development which we intend to good use in developing better system in future. In conclusion, we would like to express our thanks to management for providing us all facilities for completion of our project. Finally we extend our thanks our all M.sc. (Computer Science) staff, Classmates and also thank to our parents to help us all time.

And special thanks to my friends who helped us for successful completion of this project. They pointed out errors and suggested changes, helped in many ways, give us the idea of this project.

INDEX

Abstract

A brain tumor is a disease caused due to the growth of abnormal cells in the brain.

There are two main categories of brain tumor; they are non-cancerous (**benign**) brain tumor and cancerous (**malignant**) brain tumor.

Manual method of visual inspection of MRI images is tedious and inaccurate.

Histological grading, based on a stereotactic biopsy test, is the gold standard and the convention for detecting the grade of a brain tumor.

The biopsy procedure requires the neurosurgeon to drill a small hole into the skull from which the tissue is collected. There are many risk factors involving the biopsy test, including bleeding from the tumor and brain causing infection, seizures, severe migraine, stroke, coma and even death.

But the main concern with the stereotactic biopsy is that it is not 100% accurate which may result in a serious diagnostic error followed by a wrong clinical management of the disease. Tumor biopsy being challenging for brain tumor patients, non-invasive imaging techniques like Magnetic Resonance Imaging (MRI) have been extensively employed in diagnosing brain tumors.

Therefore, automatic classification methods are essential to prevent the death rate of human. It saves radiologist time and provides better accuracy.

Automated detection of tumor in MRI is very important as it provides information about abnormal tissues which is necessary for planning treatment.

With the growth of Artificial Intelligence, Deep learning models are used to diagnose the brain tumor by taking the images of magnetic resonance imaging(MRI). MRI is a type of scanning method that uses strong magnetic fields and radio waves to produce detailed images of the inner body.

Machine Learning algorithm based on Neural Network(CNN) is used to detect brain tumor.

Introduction

Brain tumor is one of the most rigorous diseases in the medical science. An effective and efficient analysis is always a key concern for the radiologist in the premature phase of tumor growth. Histological grading, based on a stereotactic biopsy test, is the gold standard and the convention for detecting the grade of a brain tumor. The biopsy procedure requires the neurosurgeon to drill a small hole into the skull from which the tissue is collected. There are many risk factors involving the biopsy test, including bleeding from the tumor and brain causing infection, seizures, severe migraine, stroke, coma and even death. But the main concern with the stereotactic biopsy is that it is not 100% accurate which may result in a serious diagnostic error followed by a wrong clinical management of the disease. Tumor biopsy being challenging for brain tumor patients, non-invasive imaging techniques like Magnetic Resonance Imaging (MRI) have been extensively employed in diagnosing brain tumors. Therefore, development of systems for the detection and prediction of the grade of tumors based on MRI data has become necessary.

Motivation

The main motivation behind Brain tumor detection is to not only detect tumor but it can also classify types of tumor. So it can be useful in cases such as we have to sure the tumor is positive or negative, it can detect tumor from image and return the result tumor is positive or not. This project deals with such a system, which uses computer, based procedures to detect tumor blocks and classify the type of tumor using Convolution Neural Network Algorithm for MRI images of different patients.

Problem Statement

But at first sight of the imaging modality like in Magnetic Resonance Imaging (MRI), the proper visualisation of the tumor cells and its differentiation with its nearby soft tissues is somewhat difficult task which may be due to the presence of low illumination in imaging modalities or its large presence of data or several complexity and variance of tumors-like unstructured shape, viable size and unpredictable locations of the tumor. Automated defect detection in medical imaging using machine learning has become the emergent field in several medical diagnostic applications. Its application in the detection of brain tumor in MRI is very crucial as it provides information about abnormal tissues which is necessary for planning treatment.Studies in the recent literature have also reported that automatic computerized detection and diagnosis of the disease, based on medical image analysis, could be a good alternative as it would save radiologist time and also obtain a tested accuracy. Furthermore, if computer algorithms can provide robust and quantitative measurements of tumor depiction, these automated measurements will greatly aid in the clinical management of brain tumors by freeing physicians from the burden of the manual depiction of tumors. The machine learning based approaches like Deep ConvNets in radiology and other medical science fields plays an important role to diagnose the disease in much simpler way as never done before and hence providing a feasible alternative to surgical biopsy for brain tumors . In this project, we attempted at detecting and classifying the brain tumor and comparing the results of binary and multi class classification of brain tumor with and without Transfer Learning (use of pre-trained Keras models like VGG16, ResNet50 and Inception v3) using Convolutional Neural Network (CNN) architecture.

Purpose/Objectives and Goals

Automated defect detection in medical imaging using machine learning has become the emergent field in several medical diagnostic applications. Its application in the detection of brain tumor in MRI is very crucial as it provides information about abnormal tissues which is necessary for planning treatment.For easy use of the system, the project also provides Front End and Back end so that Doctor can conveniently interact with the patient and also store his history.

- The main aim of the applications is tumor identification.
- The main reason behind the development of this application is to provide proper treatment as soon as possible and protect the human life which is in danger.
- This application is helpful to doctors as well as patient.
- The manual identification is not so fast, more accurate and efficient for user. To overcome those problem this application is design.
- It is user friendly application

Literature Survey

Paper-1: Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM

• Publication Year: 6 March 2017

• Author: Nilesh Bhaskarrao Bahadure, Arun Kumar Ray, and Har Pal Thethi

• Journal Name: Hindawi International Journal of Biomedical Imaging

• Summary: In this paper using MR images of the brain, we segmented brain tissues into normal tissues such as white matter, gray matter, cerebrospinal fluid (background), and tumor-infected tissues. We used pre-processing to improve the signal-to-noise ratio and to eliminate the effect of unwanted noise. We can used the skull stripping algorithm its based on threshold technique for improve the skull stripping performance.

Paper-2: A Survey on Brain Tumor Detection Using Image Processing Techniques

• Publication Year: 2017

• Author: Luxit Kapoor, Sanjeev Thakur

• Journal Name: IEEE 7th International Conference on Cloud Computing, Data Science & Engineering

• Summary: This paper surveys the various techniques that are part of Medical Image Processing and are prominently used in discovering brain tumors from MRI Images. Based on that research this Paper was written listing the various techniques in use. A brief description of each technique is also provided. Also of All the various steps involved in the process of detecting Tumors, Segmentation is the most significant.

Paper-3: Identification of Brain Tumor using Image Processing Techniques

- Publication Year: 11 September 2017

- Author: Praveen Gamage

- Journal Name: Research gate

- Summary: This paper survey of Identifying brain tumors through MRI images can be categorized into four different sections; pre-processing, image segmentation, Feature extraction and image classification.

Paper-4: Review of Brain Tumor Detection from MRI Images

- Publication Year: 2016

- Author: Deepa, Akansha Singh

- Journal Name: IEEE International Conference on Computing for Sustainable Global Development

- Summary: In this paper, some of the recent research work done on the Brain tumor detection and segmentation is reviewed. Different Techniques used by various researchers to detect the brain Tumor from the MRI images are described. By this review we found that automation of brain tumor detection and Segmentation from the MRI images is one of the most active Research areas.

Paper-5: An efficient Brain Tumor Detection from MRI Images using Entropy Measures

- Publication Year: December 23-25, 2016

- Author: Devendra Somwanshi , Ashutosh Kumar, Pratima Sharma, Deepika Joshi

- Journal Name: IEEE International Conference on Recent Advances and Innovations in Engineering

- Summary: In this paper, we have investigated the different Entropy functions for tumor segmentation and its detection from various MRI images. The different threshold values are obtained depend on the particular definition of the entropy.

Project Scope and Limitations

Project Scope

1. It is considered as the best ml technique for image classification due to high accuracy.

2. Image pre-processing required is much less compared to other algorithms.

3. It is used over feed forward neural networks as it can be trained better in case of complex images to have higher accuracies.

4. It reduces images to a form which is easier to process without losing features which are critical for a good prediction by applying relevant filters and reusability of weights

5. It can automatically learn to perform any task just by going through the training data i.e. there no need for prior knowledge

6. There is no need for specialised hand-crafted image features like that in case of SVM, Random Forest etc.

Limitations:

1. It requires a large training data.

2. It requires appropriate model.

3. It is time consuming.

4. It is a tedious and exhaustive procedure.

5. While convolutional networks have already existed for a long time, their success was limited due to the size of the considered network.

Solution-Transfer Learning for inadequate data which will replace the last fully connected layer with pre-trained ConvNet with new fully connected layer.

Existing System and Limitations

There are several existing of techniques are available for brain tumor segmentation and classification to detect the brain tumor. There are many techniques available presents a study of existing techniques for brain tumor detection and their advantages and limitations. To overcome these limitations, propose a Convolution Neural Network (CNN) based classifier. CNN based classifier does the comparison between trained and test data, from this to get the simplest result.

Project Perspective

- To provide doctors good software to identify tumor and their causes.
- Save patient’s time.
- Provide a solution appropriately at early stages.
- Get timely consultation.

Features

- The website reduces the manual work to maintain record efficiency
- Save patient’s time.
- This web application is fully functional and flexible
- Easy to use because of the friendly Graphical User Interface.

Requirement Analysis

Software Requirements:

- **Windows:** Python 3.6.2 or above, PIP and NumPy 1.13.1
- **Python**
- **PIP:** It is the package management system used to install and manage software packages written in Python.
- **NumPy:** NumPy is a general-purpose array-processing package. It provides a high performance multidimensional array object, and tools for working with these arrays
- **Anaconda:** Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution includes data-science packages suitable for Windows, Linux, and macOS. Anaconda distribution comes with 1,500 packages selected from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator, as a graphical alternative to the command-line interface (CLI).
- **Jupyter Notebook:** Anaconda distribution comes with 1,500 packages selected from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator, as a graphical alternative to the command line interface (CLI). A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text mathematics, plots and rich media, usually ending with the ". Ipynb" extension.
- **Tensor Flow:** Tensor flow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google. 31
- **Keras:** Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R,

Theano, or Plaid ML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code.

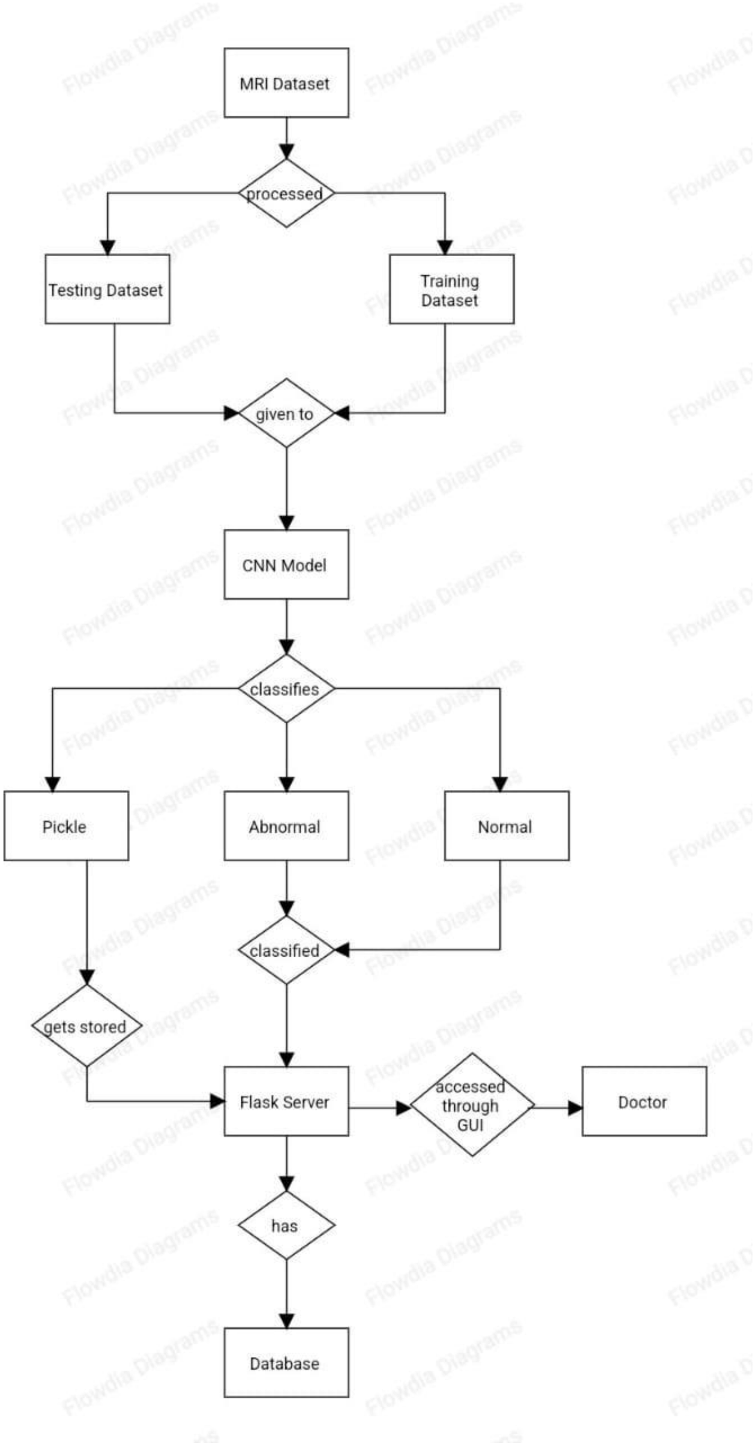
- **OpenCV:** OpenCV (Open source computer vision) is a library of programming functions mainly aimed at real-time computer vision. Originally developed by Intel, it was later supported by willow garage then Itseez (which was later acquired by Intel). The library is cross platform and free for use under the open source BSD license. OpenCV supports some models from deep learning frameworks like TensorFlow, Torch, PyTorch (after converting to an ONNX model) and Caffe according to a defined list of supported layers. It promotes Open Vision Capsules. Which is a portable format, compatible with all other formats.

Hardware Requirements:

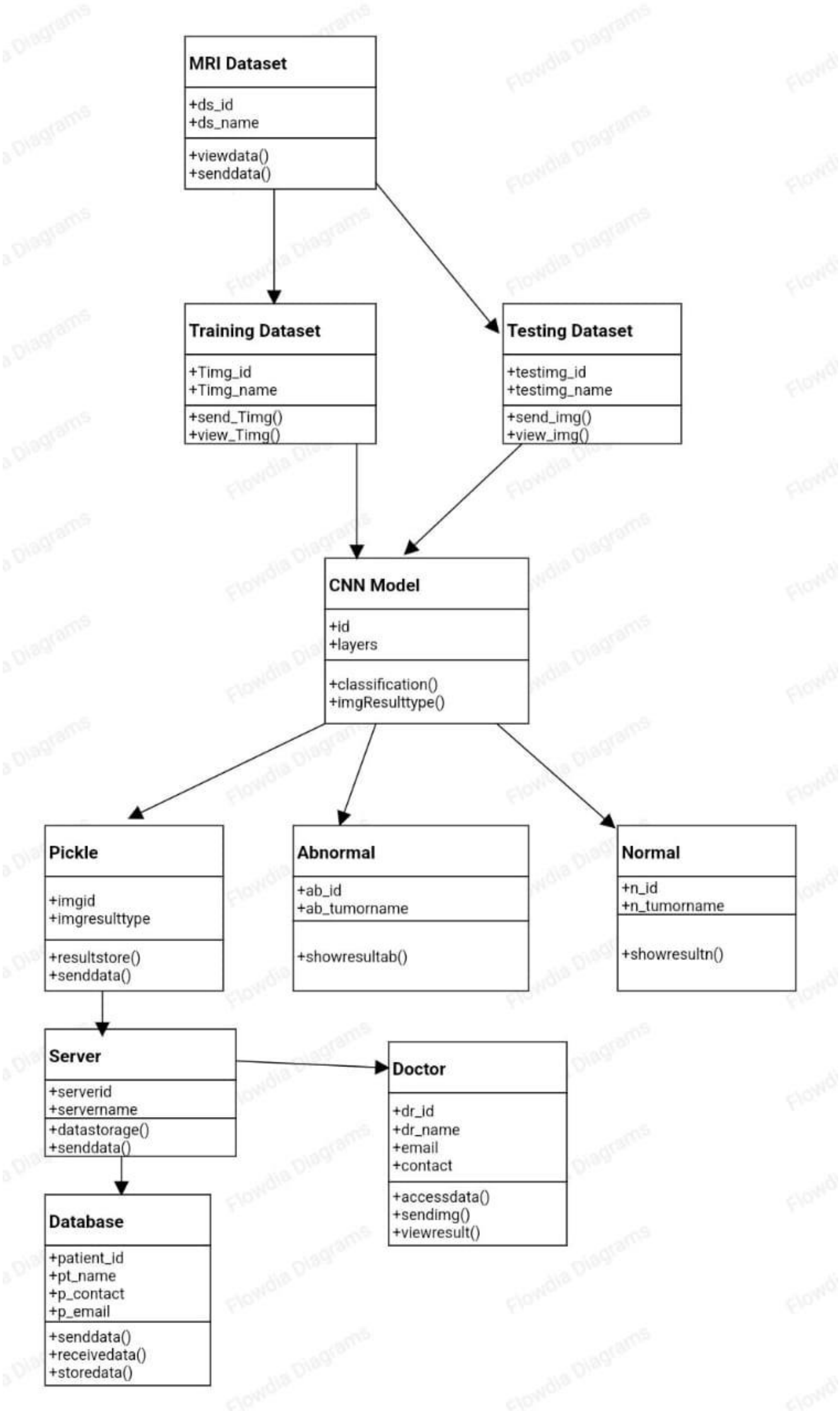
- Processor: Intel core i5 or above.
- 64-bit, quad-core, 2.5 GHz minimum per core
- Ram: 4 GB or more
- Hard disk: 10 GB of available space or more.
- Display: Dual XGA (1024 x 768) or higher resolution monitors
- Operating system: Windows

System Model : UML Diagrams

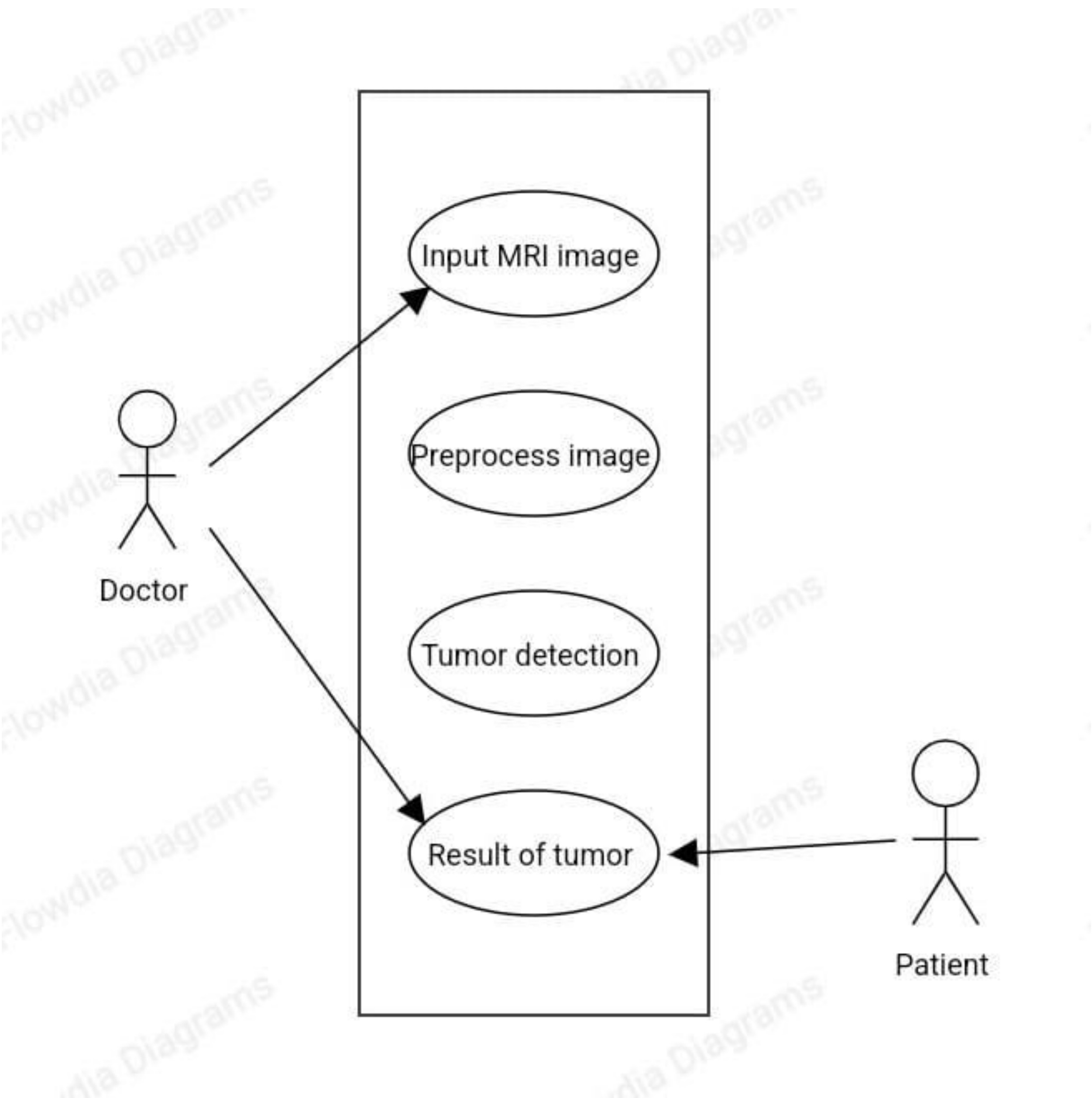
- **Entity Relationship Diagram**



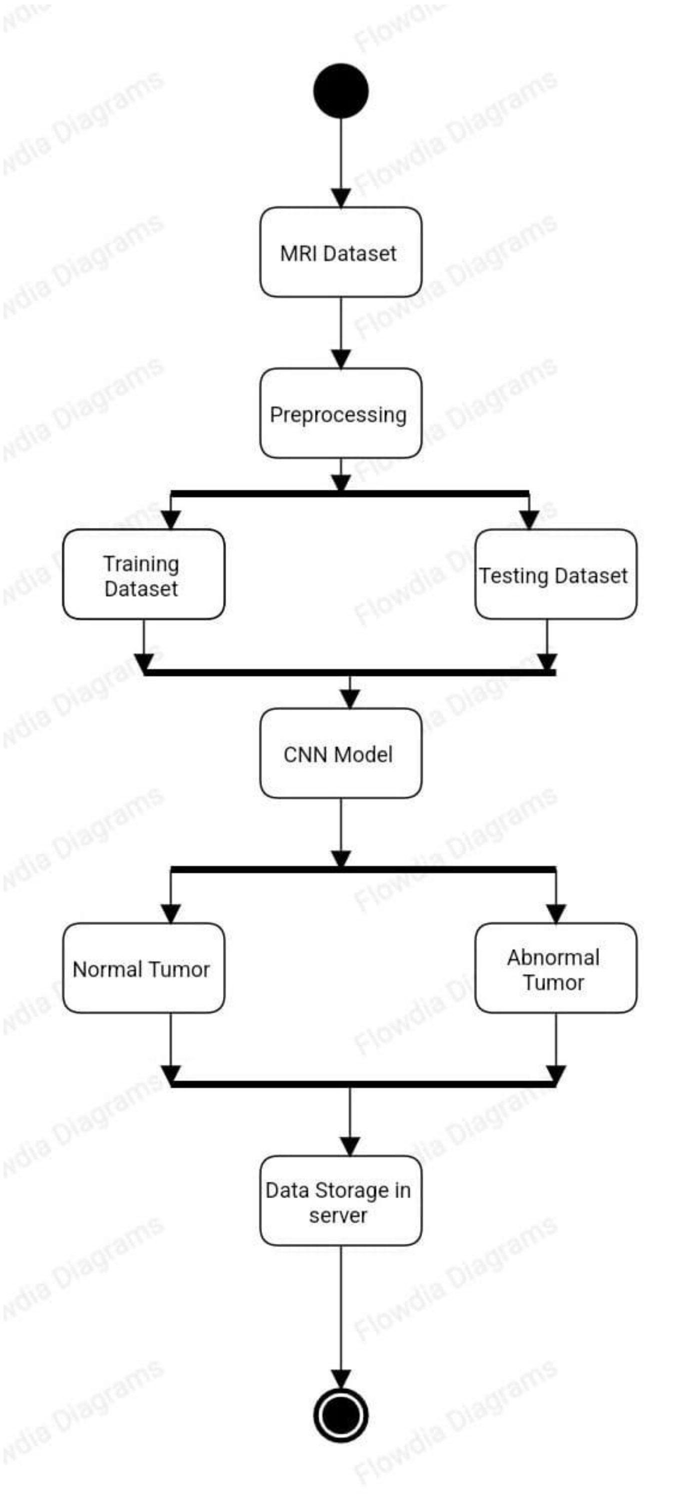
- ## **Class Diagram**



- **Use Case Diagram**

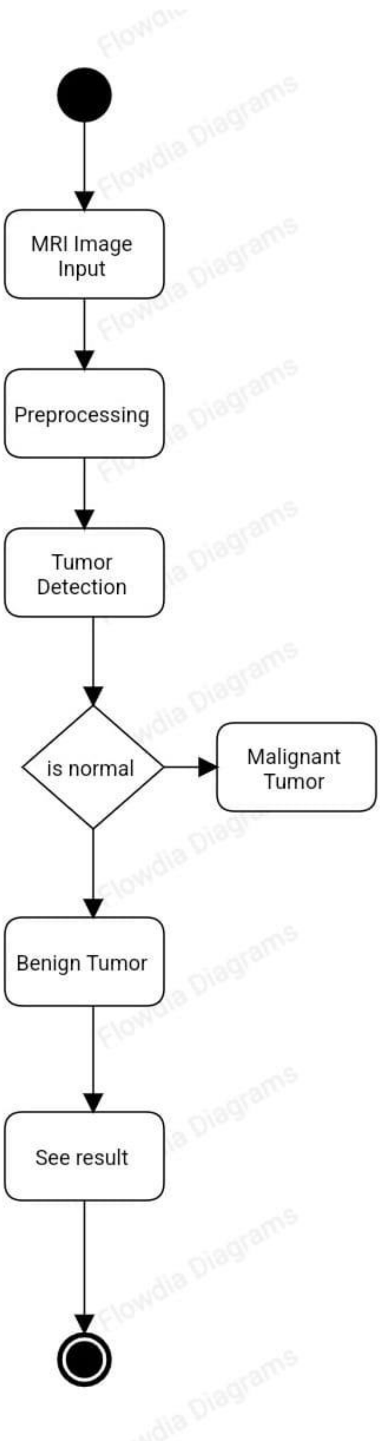


- **Activity Diagram :**



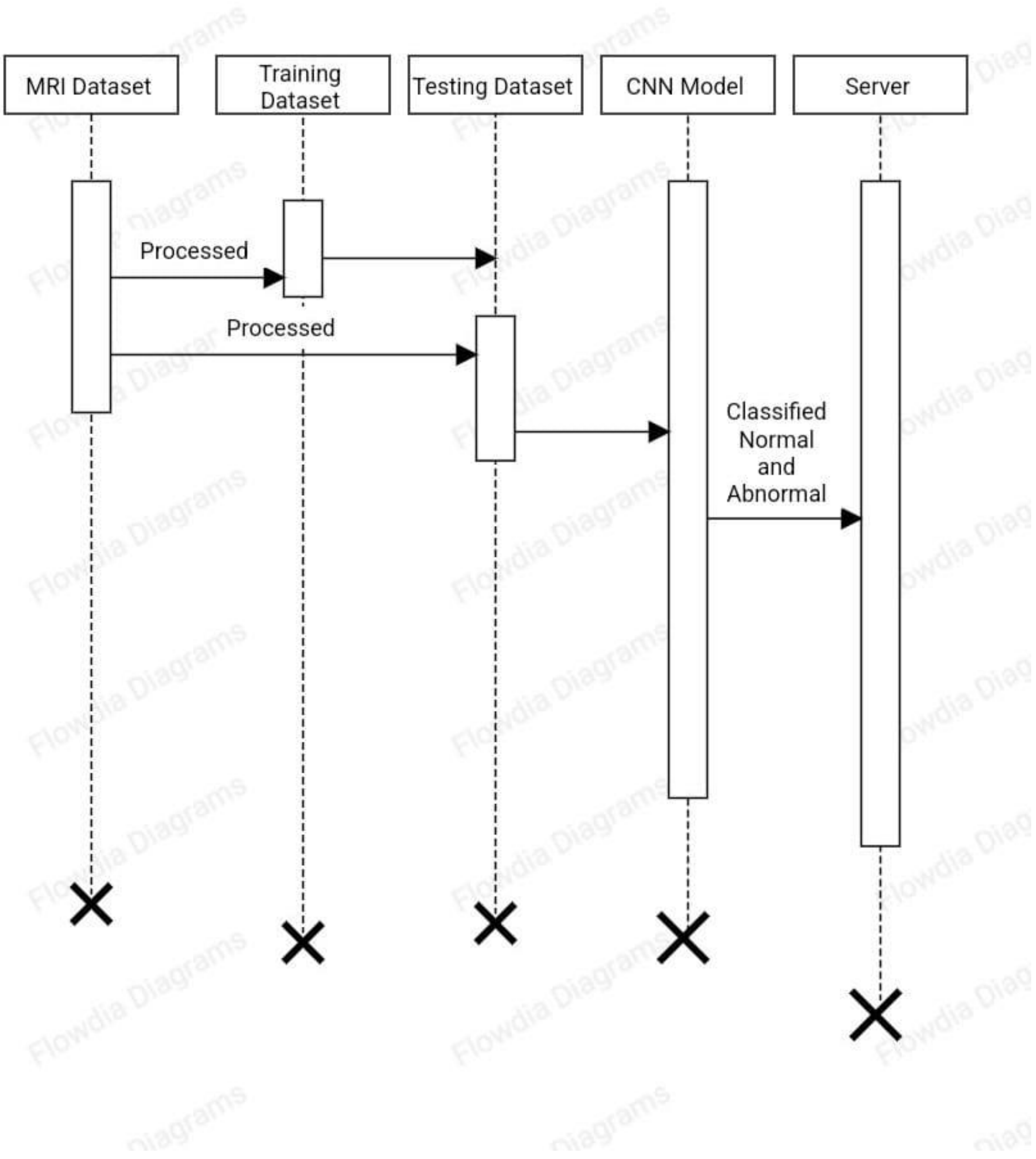
Activity Diagram of System

- **Activity Diagram**

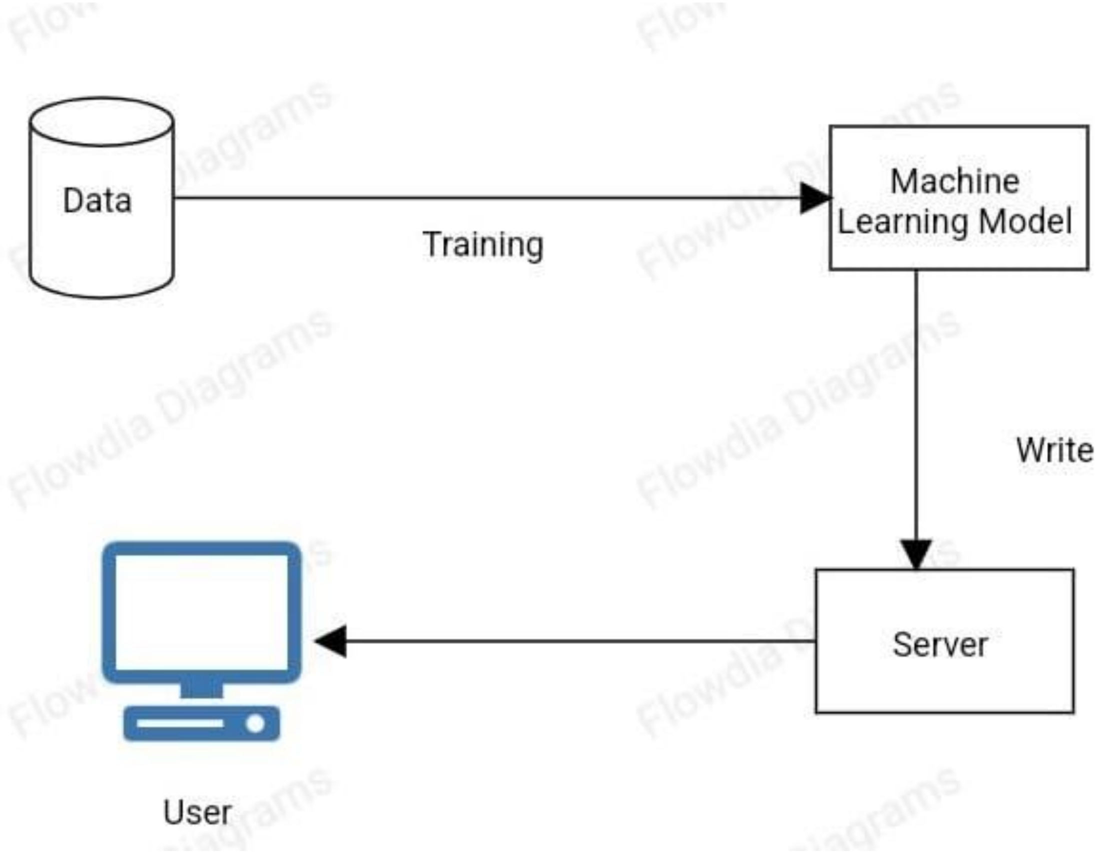


Activity Diagram of Doctor

- Sequence Diagram



- Deployment Diagram



Data Model

Table Name : MRI Dataset

Sr.no	Attribute	Datatype	Constraints	Description
1	id	int	Primary key	Image Identifier
2	name	Varchar(20)	Not null	MRI name
3	type	Varchar(20)	Not null	Tumor type

Table Name : Training Dataset

Sr.no	Attribute	Datatype	Constraints	Description
1	id	int	Primary key	Image identifier
2	name	Varchar(20)	Not null	MRI name
3	type	Varchar(20)	Not null	Tumor type

Table Name : Testing Dataset

Sr.no	Attribute	Datatype	Constraints	Description
1	id	int	Primary key	Image identifier
2	name	Varchar(20)	Not null	MRI name
3	type	Varchar(20)	Not null	Tumor type

Table Name : CNN Model

Sr.no	Attribute	Datatype	Constraints	Description
1	id	int	Primary key	Image identifier
2	trainingimgid	int	Foreign key	Training img id identifier
3	testingimgid	int	Foreign key	Testing img id identifier
4	type	Varchar(20)	Not null	Tumor type

Table Name : Database

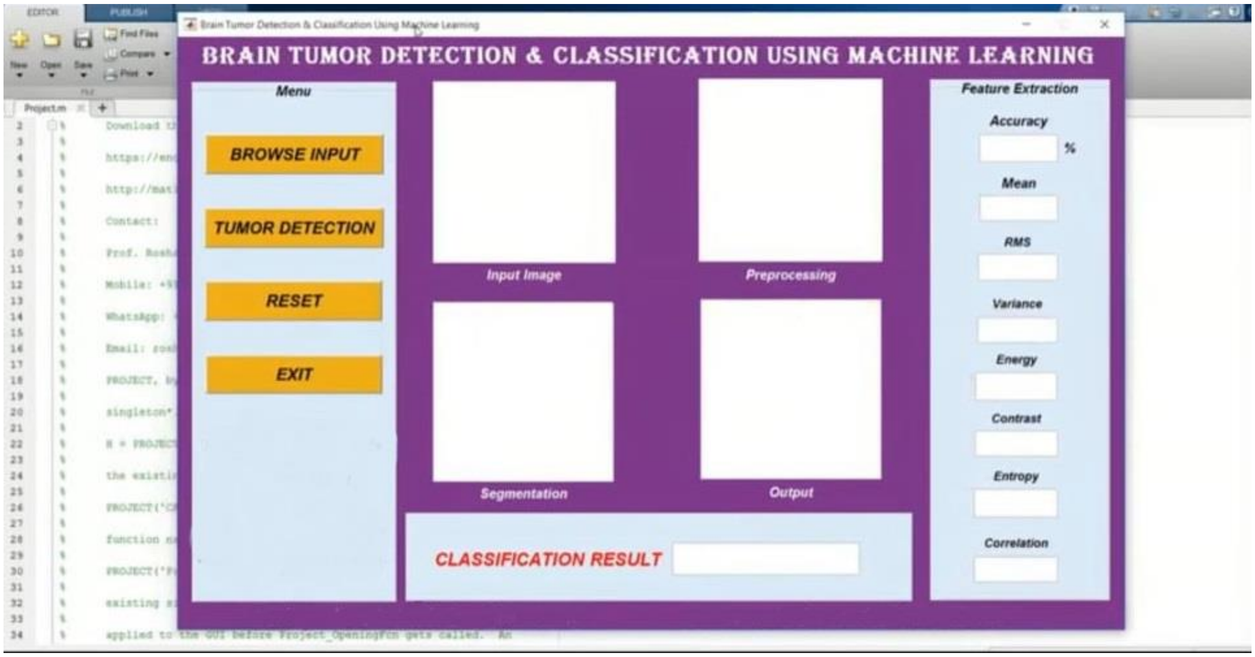
Sr.no	Attribute	Datatype	Constraints	Description
1	Patient_id	int	Primary key	Patient identifier
2	name	Varchar(20)	Not null	Patient name
3	email	Varchar(20)	Not null	Patient email
4	contact	Int(11)	Not null	Patient contact no
5	type	Varchar(20)	Not null	Tumor type

Table Name : Doctor

Sr.no	Attribute	Datatype	Constraints	Description
1	id	int	Primary key	Dr identifier
2	name	Varchar(20)	Not null	Dr name
3	email	Varchar(20)	Not null	Dr email id
4	contact	Int(11)	Not null	Dr contact no

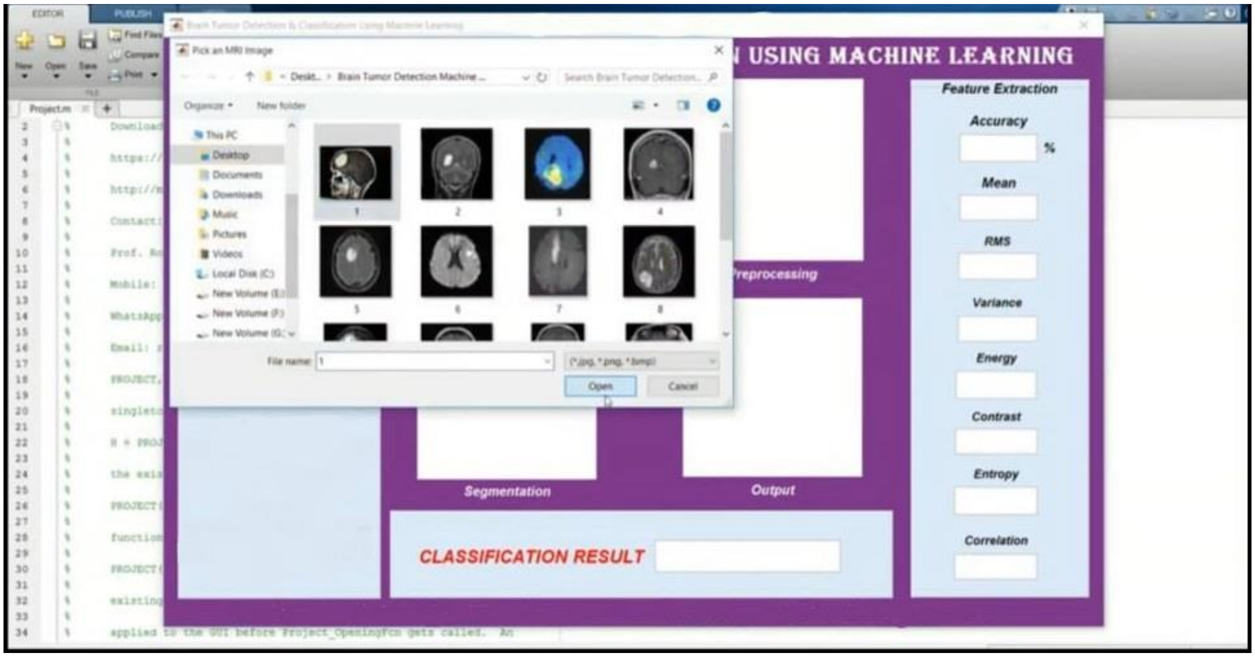
User Interface

Home page



```
%     Download t
%
%     https://en
%
%     http://mat
%
%     Contact:
%
%     Prof. Rosh
%
%     Mobile: +9
%
%     WhatsApp:
%
%     Email: ros
%
%     PROJECT, by
%
%     singleton*
%
%     H = PROJEC
%
%     the existi
%
%     PROJECT('C
%
%     function n
%
%     PROJECT('F
%
%     existing s
%
%     applied to the GUI before Project_OpeningFcn gets called.  An
```

Image Input by doctor:

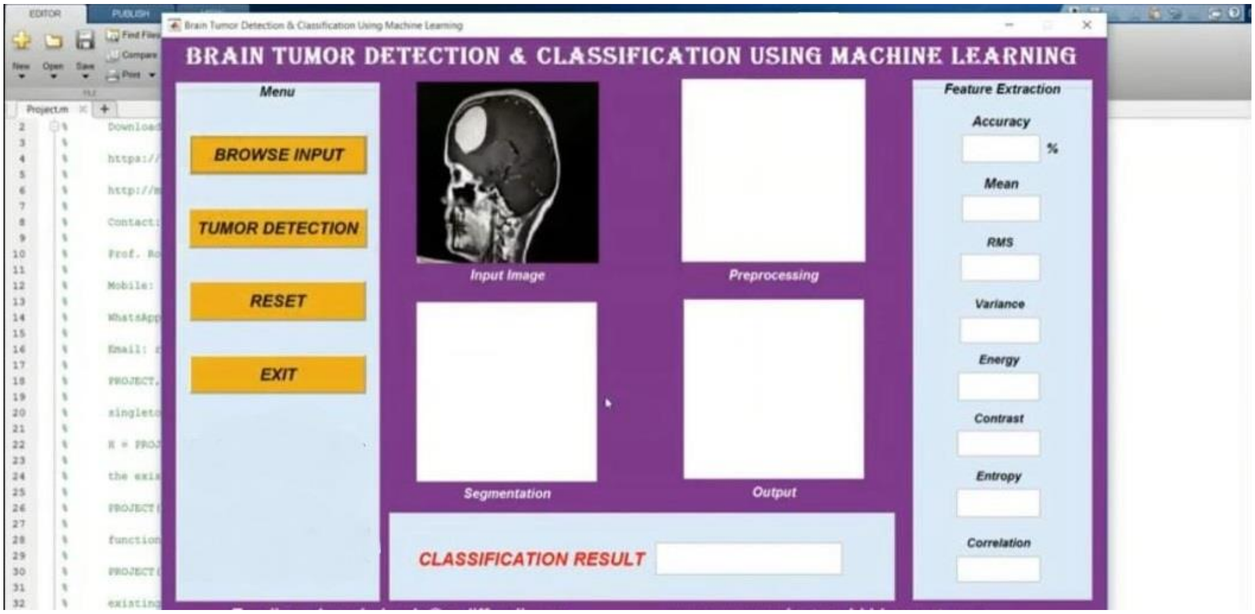


```
%       Download
%
%       https://
%
%       http://
%
%       Contact:
%
%       Prof. Ro
%
%       Mobile:
%
%       WhatsApp
%
%       Email: 
%
%       PROJECT,
%
%       singleto
%
%       H = PROJ
%
%       the exis
%
%       PROJECT(
%
%       function
%
%       PROJECT(
%
%       existing
%
%       applied to the GUI before Project_OpeningFcn gets called.  An
```

File name: 1

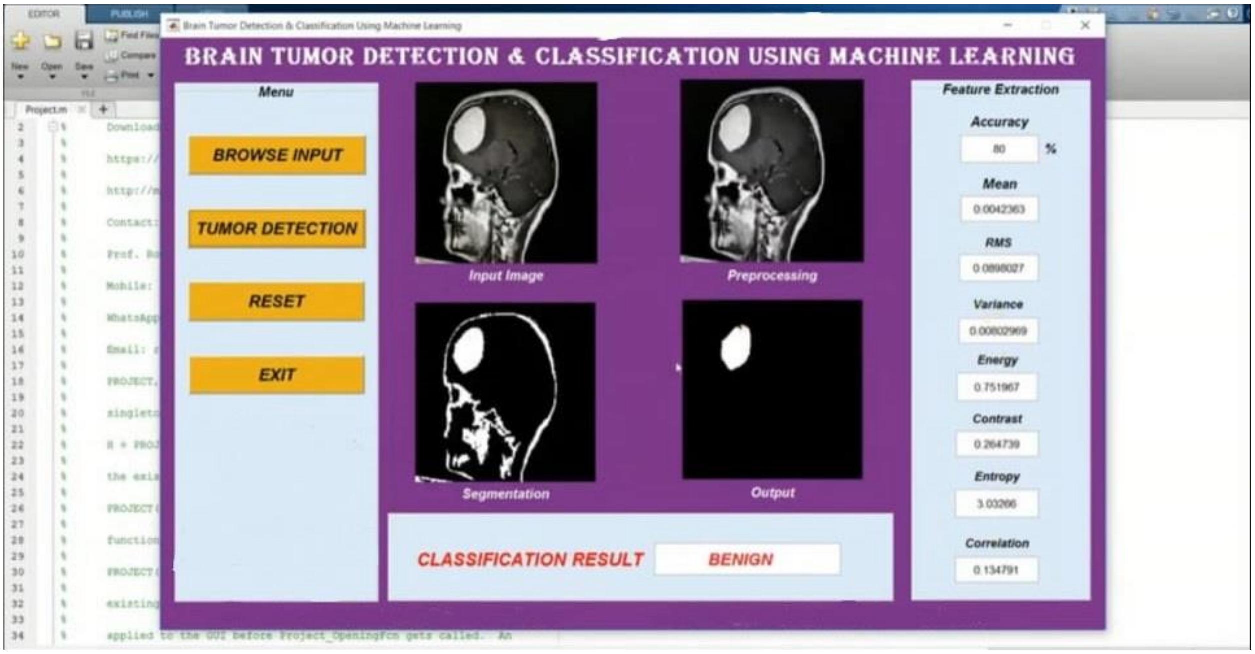
(*.jpg, *.png, *.bmp)

Input Image:



```
%       Download
%
%       https://
%
%       http://m
%
%       Contact:
%
%       Prof. Ro
%
%       Mobile:
%
%       WhatsApp
%
%       Email: r
%
%       PROJECT.
%
%       singleto
%
%       H = PROJ
%
%       the exis
%
%       PROJECT(
%
%       function
%
%       PROJECT(
%
%       existing
```

Image Processing and Classified Result :



```
%       Download
%       https://
%       http://m
%       Contact:
%       Prof. Bo
%       Mobile:
%       WhatsApp
%       Email: r
%       PROJECT,
%       singleto
%       H = PROJ
%       the exis
%       PROJECT(
%       function
%       PROJECT(
%       existing
%       applied to the GUI before Project_OpeningFcn gets called.  An
```

CLASSIFICATION RESULT: BENIGN

Accuracy: 80 %

Mean: 0.0042363

RMS: 0.0898027

Variance: 0.00802969

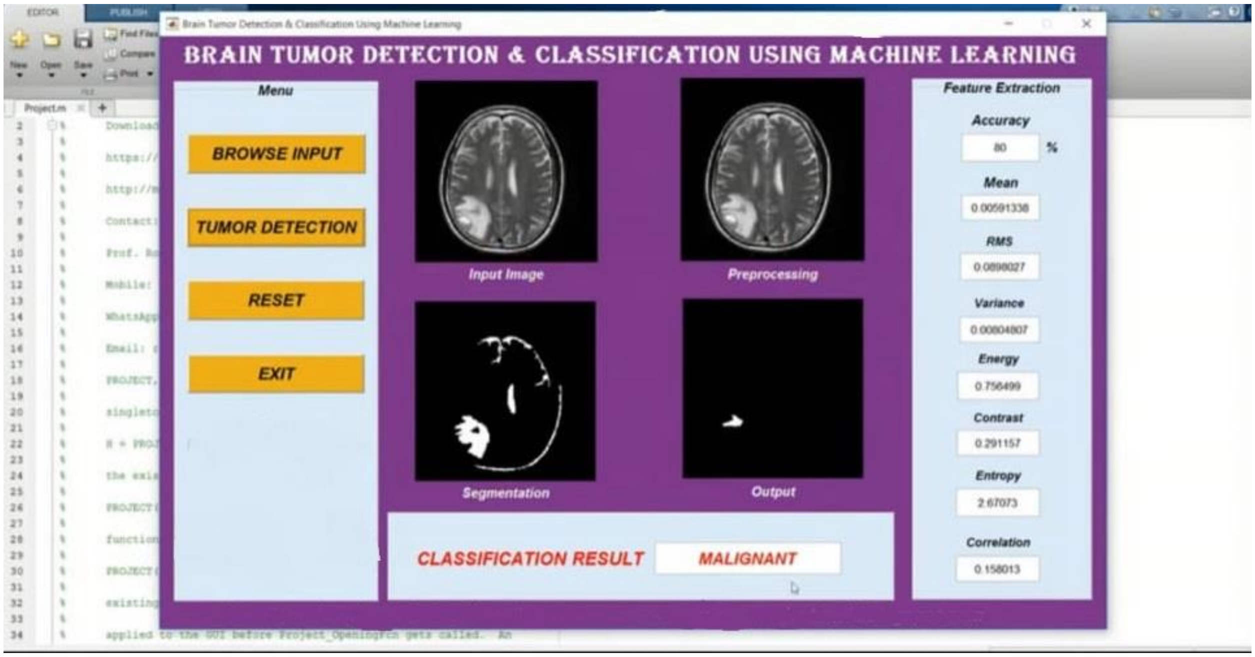
Energy: 0.751967

Contrast: 0.264739

Entropy: 3.03266

Correlation: 0.134791

Image Processing and Classified Result :



```
% Download
% https://
% http://
% Contact:
% Prof. Ro
% Mobile:
% WhatsApp
% Email:
% PROJECT,
% singleto
% H = PROJ
% the exis
% PROJECT(
% function
% PROJECT(
% existing
% applied to the GUI before Project_OpeningFcn gets called.  An
```

CLASSIFICATION RESULT: MALIGNANT

Accuracy: 80 %

Mean: 0.00591338

RMS: 0.0898027

Variance: 0.00804807

Energy: 0.756499

Contrast: 0.291157

Entropy: 2.67073

Correlation: 0.158013

Implementation Details

Hardware Specifications:

Hard Disk	:	500GB and Above
RAM	:	4GB and Above
Processor	:	I3 and Above
Output Devices	:	Monitor,Printer

Software Requirements:

Operating System	:	Windows 10 (64 bit)
Software	:	Python
Tools	:	Anaconda (Jupyter Note Book IDE)
Web Browser	:	Chrome or any

Outputs and Reports

The image data are added to the variable named data which is of ndarray datatype. The class labels of the images are also generated and stored in the variable data_target which is also an ndarray. The images are then added inside the dataframe. The image dataset is divided into training, validation and testing dataset. Figure 3 represents the accuracy and loss obtained when the CNN model is applied on the training and validation dataset When CNN model is applied on the training data for fifty epochs training accuracy obtained is 97.13% and a validation accuracy of 71.51 %. The same when applied on the testing data gives 80.77% accuracy.

Testing

SYSTEM TESTING:

Web application should sustain to heavy load Web performance testing should include web load testing

1. Web Application Testing Checklist
2. Functionality Testing
3. Usability Testing
4. Interface Testing
5. Compatibility Testing
6. Performance Testing
7. Security Testing

Web testing is software testing practice to test websites. By performing website testing an organization can make sure that web-based system is functioning properly and can be accepted by real-time users.

The UI design and functionality are the captions of website testing

1) Functional Testing:

Test for-all the links in web pages, database connection, forms used for submitting or getting information from the user through web pages, cookies testing etc.

a) Links
b) Internal Links
c) External Links
d) Mail Links
e) Forms
f) :Field validation
g) :Error message or wrong input
h) :Optional and mandatory fields

Database:

Testing will be done on database integrity.

2) Usability Testing:

It is process by which the human-computer interaction characteristics of a system are measure, and weakness are identified for correction.

It included the following:

a) The website should be easy to use.
b) The instructions provided should be very clear.

3) Interface Testing:

In web testing, the server-side interface should be tested. This is done by that communication is done properly. Compatibility of server with software, hardware, network and database should be tested.

a) Web server and application server interface.
b) Application server and database server interface.

4) Compatibility Testing:

The compatibility of website is very important testing aspect. See which compatibility test to be executed.

a) Browser compatibility.
b) Operating system compatibility.

5) Performance Testing:

The web application should sustain to heavy load. Web performance testing should include.

Web load testing.

Conclusion and Recommendation

Without pre-trained Keras model, the train accuracy is 97.5% and validation accuracy is 90.0%.The validation result had a best figure of 91.09% as accuracy.It is observed that without using pre-trained Keras model, although the training accuracy is >90%, the overall accuracy is low unlike where pre-trained model is used.

Also, when we trained our dataset without Transfer learning, the computation time was 40 min whereas when we used Transfer Learning, the computation time was 20min. Hence, training and computation time with pre-trained Keras model was 50% lesser than without.

Chances over over-fitting the dataset is higher when training the model from scratch rather than using pre-trained Keras.Keras also provides an easy interface for data augmentation.

Amongst the Keras models, it is seen that ResNet 50 has the best overall accuracy as well as F1 score.ResNet is a powerful backbone model that is used very frequently in many computer vision tasks.

Precision and Recall both cannot be improved as one comes at the cost of the other .So, we use F1 score too.

Transfer learning can only be applied if low-level features from Task 1(image recognition) can be helpful for Task 2(radiology diagnosis).
For a large dataset, Dice loss is preferred over Accuracy.

For small size of data, we should use simple models, pool data, clean up data, limit experimentation, use regularisation/model averaging ,confidence intervals and single number evaluation metric.

To avoid overfitting, we need to ensure we have plenty of testing and validation of data i.e. dataset is not generalised. This is solved by Data Augmentation. If the training accuracy too high, we can conclude that it the model might be over fitting the dataset. To avoid this, we can monitor testing accuracy, use outliers and noise, train longer, compare variance (=train performance-test performance).

Future Scope

Build an app-based user interface in hospitals which allows doctors to easily determine the impact of tumor and suggest treatment accordingly

Since performance and complexity of ConvNets depend on the input data representation we can try to predict the location as well as stage of the tumor from Volume based 3D images. By creating three dimensional (3D) anatomical models from individual patients, training, planning and computer guidance during surgery is improved.

Using VolumeNet with LOPO (Leave-One-Patient-Out) scheme has proved to give a high training as well as validation accuracy(>95%).In LOPO test scheme, in each iteration, one patient is used for testing and remaining patients are used for training the ConvNets, this iterates for each patient. Although LOPO test scheme is computationally expensive, using this we can have more training data which is required for ConvNets training. LOPO testing is robust and most applicable to our application, where we get test result for each individual patient. So, if classifier misclassifies a patient then we can further investigate it separately.

Improve testing accuracy and computation time by using classifier boosting techniques like using more number images with more data augmentation, fine-tuning hyper parameters, training for a longer time i.e. using more epochs, adding more appropriate layers etc.. Classifier boosting is done by building a model from the training data then creating a second model that attempts to correct the errors from the first model for faster prognosis. Such techniques can be used to raise the accuracy even higher and reach a level that will allow this tool to be a significant asset to any medical facility dealing with brain tumors.

For more complex datasets, we can use U-Net architecture rather than CNN where the max pooling layers are just replaced by upsampling ones.

Ultimately we would like to use very large and deep convolutional nets on video sequences where the temporal structure provides very helpful information that is missing or far less obvious in static images.

Unsupervised transfer learning may attract more and more attention in the future.

Bibliography/References

Reference Books:

1. The hundred page Machine learning Book - Andriy Burkov
2. Machine learning for hackers - Drew Conway and John Myles White

Youtube Videos

Websites:

1. www.w3schools.com
2. www.geeksforgeeks.org