A

Major

Project On

RECOGNIZING HUMAN BRAIN TUMORS

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

BY

Gayathri Jadhav

(17601A0534)

Arupula Umarani

(176H1A0501)

Kongala Praveen

(177R5A0503)

Under the Guidance of

K.RANJITH REDDY

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)
Recognized Under Section 2(f) & 12(B) of the UGC Act.1956,
Kandlakoya (V), Medchal Road, Hyderabad-501401.

2017-2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "**RECOGNIZING HUMAN BRAIN TUMORS**" being submitted by **Gayathri Jadhav 17601A0534, Arupula Umarani 176H1A0501, Kongala Praveen 177R5A0503,** bearing the**, ,** roll number in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by him/her under our guidance and supervision during the year 2020-21.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Mr.K.Ranjith Reddy
Assistant Professor
INTERNAL GUIDE

Dr. A. Raji Reddy
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voice Examination held on_____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We take this opportunity to express our profound gratitude and deep regard to our guide

Mr.K.Ranjith Reddy, Assistant Professor for her exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by her shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) Coordinators: **Ms. Sandhya, Mr. BP Deepak Kumar** and **Mr. J. Narasimha Rao** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to the Head of the Department **Dr. K. Srujan Raju** for providing excellent infrastructure and a nice atmosphere for completing this project successfully. We are obliged to our Director **Dr. A. Raji Reddy** for being cooperative throughout the course of this project. We would like to express our sincere gratitude to our Chairman Sri. **Ch. Gopal Reddy** for his encouragement throughout the course of this project the guidance and support received from all the members of **CMR TECHNICAL CAMPUS** who contributed and who are contributing to this project, was vital for the success of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement without which this assignment would not be possible. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project.

Gayathri Jadhav (17601A0534)
Arupula Umarani (176H1A0501)
Kongala Praveen (177R5A0503)

ABSTRACT

Deep Learning is the newest and the current trend of the machine learning field that paid a lot of the researchers' attention in the recent few years. As a proven powerful machine learning tool, deep learning was widely used in several applications for solving various complex problems that require extremely high accuracy and sensitivity, particularly in the medical field. In general, the brain tumor is one of the most common and aggressive malignant tumor diseases which is leading to a very short expected life if it is diagnosed at a higher grade. Based on that, brain tumor grading is a very critical step after detecting the tumor in order to achieve an effective treating plan. In this paper, we used Convolutional Neural Network (CNN) which is one of the most widely used deep learning architectures for classifying a dataset of 3064 T1 weighted contrast-enhanced brain MR images for grading (classifying) the brain tumors into three classes (Glioma, Meningioma, and Pituitary Tumor). The proposed CNN classifier is a powerful tool and its overall performance with an accuracy of 98.93%.

LIST OF FIGURES

LIST OF SCREENSHOTS

.

TABLE OF CONTENT

1. INTRODUCTION

1. INTRODUCTION

Brain tumors can be classified into two types: benign (noncancerous) and malignant (cancerous). The malignant tumors can quickly spread to other tissues in the brain and lead to worsening the patient's condition [1]. When most of the cells are old or damaged, they are destroyed and replaced by new cells. If damaged and old cells are not eliminated with generating the new cells, it can cause problems. The production of additional cells often results in the formation of a mass of tissue, which refers to the growth or tumor. Brain tumor detection is very complicated and difficult due to the size, shape, location and type of tumor in the brain. Diagnosis of brain tumors in the early stages of the tumor's start is difficult because it cannot accurately measure the size and resolution of the tumor [2]. However, if the tumor is diagnosed and treated early in the tumor formation process, the chance of patient's treatment is very high. Therefore, the treatment of tumor depends on the timely diagnosis of the tumor [3]. The diagnosis is usually done by a medical examination, with computer tomography or magnetic imaging. MRI imaging is a method that provides accurate images of the brain and is one of the most common and important methods for diagnosing and evaluating the patient's brain. In the field of Medical Detection Systems (MDS), MRI images provide better results than other imaging techniques such as Computed Tomography (CT), due to their higher contrast in soft tissue in humans [4]. The proposed technique has used CNN to identify and categorize the tumor from brain images of the brain. The main difference between the main channels of the neural network with the normal neural network is that it is able to automatically and locally extract the feature from each image [5]. These types of networks consist of neurons with weights and biases that can be learned [6]. Due to the results of CNN on the dataset, in order to improve the proposed method. Machine learning algorithm is used to feature extraction. The algorithm used was the clustering algorithm applied on data set, and then the images are applied to the CNN. The results showed that the proposed method has been successful. The purpose of extracting the property before applying to the CNN is that in some images fatty masses are considered as tumors, or in some images the tumor is mistakenly considered to be fat and should have increased medical error. Extracting the attribute initially and before applying the CNN leads to improved network accuracy and increased accuracy. The advancement in medical technologies helps the clinical experts to facilitate more efficient e-health care systems to the patients. There is a number of medical domains where e-health care systems are beneficial [1]. Computer vision-based applications of biomedical imaging are more importance as they provide recognition information to the radiologist for batter treatment-related problems. Different medical imaging techniques and methods that include X-ray, Magnetic Resonance

Imaging (MRIs), Ultrasound, and Computed Tomography (CT), have a great influence on the diagnosis and treatment process of patients [2, 3].The formation of abnormal groups of cells inside the brain or near it leads to the initialization of a brain tumor. The abnormal cells abrupt the processing of the brain and affect the health of a patient. Brain imaging analysis, diagnosis, and treatment with adopted medical imaging techniques are the main focus of research for the researcher, radiologist and clinical experts [5]. The analysis of brain images is considered imperative because diseases of the brain called brain tumors are fatal and responsible for a large number of deaths in developed countries; for instance, according to the National Brain Tumor Foundation (NBTF), 29,000 people are diagnosed with brain tumor in the United States (US) with brain tumor and 13,000 of those patients die per annum [6]. A number of advanced Magnetic Resonance Imaging (MRI) techniques that include Diffusion Tensor Imaging (DTI), MR Spectroscopy (MRS) and Perfusion MR are used for the analysis of brain tumor through MRI [7–9]. Brain tumor is broadly classified into two types: cancerous tumors, known as malignant tumors, and noncancerous tumors, known as benign tumors. Malignant tumors are further classified into grades I to IV by World Health Organization (WHO) [10]. A Grade-I tumor is called Pilocytic Astrocytoma, Grade-II tumor is Low-Grade Astrocytoma, Grade-III tumor is Anaplastic Astrocytoma and Grade-IV tumor is Glioblastoma. Grade-I tumors and Grade-II tumors are semi-malignant tumors with less aggressiveness. Grade-III and Grade-IV are malignant tumors and highly affect the health of the patient and may lead to the death of tumor patients [11]. A variety of image-processing techniques and methods have been used for the diagnosis and treatment of a brain tumor. Segmentation is the fundamental step in image processing techniques and is used to extract the infected region of brain tissue from MRIs [12]. Segmentation of the tumor region is an important task for cancer diagnosis, treatment, and the evaluation of treatment outcomes. A vast number of semi-automatic and automatic segmentation methods and techniques are used for tumor segmentation [13]. MRI contains methods with multiple sequence that include T1-weighted (TI) and T1-weighted contrast-enhanced (T1c), T2-weighted and T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) techniques, which are employed for the segmentation of brain tumor. MRI shave various features that are adopted in brain tumor segmentation studies that include image textures [14], local histograms [15], and structure tensor eigenvalues [16]. Machine learning methods such as Support Vector Machines (SVMs) [17–19] and Random Forest (RF) [14–16, 20] are commonly used for pattern classification in tumor segmentation studies. Deep-learning-based techniques and methods are becoming popular in brain tumor segmentation studies, as their performance is superior in image analysis fields, such as object detection [21], image classification [22] and semantic segmentation [23–25]. Deep learning techniques have achieved state-of-the-art performance for automatic segmentation of brain tumors through multi-model MRIs [1]. The Convolutional Neural Network (CNN) is a powerful method for image recognition and

Prediction. However, CNN is mostly used for brain tumor segmentation, classification, and prediction of survival time for patients [26–28]. More deep-learning-based methods that are utilized for tumor segmentation, classification, and prediction include Stacked De-Noising Auto encoders [29] and Convolutional Restricted Boltzmann Machine [30]. Among all the deep learning methods and techniques, CNNs perform batter for image segmentation, classification, and prediction. Two-Dimensional CNNs (2D-CNNs) [31–35] and 3D-CNNs [16, 36, 37], were both adopted to build brain tumor segmentation, classification, and prediction methods. Segmentation methods classify the image patch into different classes, such as necrosis, healthy tissues, edema, enhancing core and non-enhancing core. Different tumor cells show distinct phenotypic and morphological information for segmentation, classification, and prediction, including gene expression, motility, cellular morphology, metabolism metastatic potential, and proliferation. This paper presents a review of various methods, techniques, frameworks, architectures, algorithms and critical studies using deep learning for segmentation, classification, and survival time prediction. Survey taxonomy describes the methods, techniques, systems, algorithms, frameworks, and architectures that are based on tumor segmentation, evaluation, and features exploration for tumor prediction and its classification. The review performs an analysis of the features extraction techniques, dataset utilized, tools, languages, and libraries that are used for implementation, recognition and evaluation measures. The issues and research gaps in various existing research problems include the key issues in tumor recognition for monitoring, recognition procedures and treatment plans for cancer patients. The application of deep learning to brain tumor analysis first appears in conferences and workshops, and then in journals. The number of research papers grew rapidly from 2015 to onward. This topic has now became dominant at different conferences and journals. Figure1 illustrates the development of deep learning applications to brain tumor analysis. Figure2 presents a literature-based taxonomy for brain tumor analysis The development of deep learning application to brain tumor analysis motivated us to present a comprehensive review in all fields of brain tumor that includes segmentation, prediction, classification, both from methodology-driven and applications perspective. The review also includes a never view of all there search publications in tabular form that helps readers to quickly assess the field. Consequently, this review presents a dedicated discussion section to the readers that covers the state-of-the-art successful development, open research challenges and an overview of future research directions. Their view includes a large number of research papers, most of them recent, presenting an extensive variety of deep learning applications in brain tumor analysis to identify the most relevant contribution (“deep learning”AND“BrainTumor”) in the title and abstract query performed. Additionally, MICCAI workshop papers related to brain tumors have also been included in this review. In summary, the aim of this review is (a) to show the deep learning development in the entire field of brain tumor, (b) the identification of open research challenges for

successful deep learning methods for brain tumor tasks, (c) to highlight the successful deep learning contribution to brain tumor analysis.

1.1 BRAIN TUMOR SEGMENTATION:

Brain tumor segmentation is performed to extract the tumor region from the images for the further classification and prediction of brain tumors. Different Machine ML/DL methods are proposed for the segmentation of tumorized cells. Some of these ML methods use manually segmented images for the training, which is costly, time-consuming and needs medical expertise. Two types of data, fully-annotated and weakly annotated data, train the deep learning methods for segmentation. A method that uses these two types of data, presented by V. Rao, adds an additional branch to the segmentation network for image-level classification. The method also studies the weakly annotated images to learn to avoid features that are irrelevant for the segmentation task. Deep Neural Network (DNN) is applied on the Pixel wise multimodal image representation that includes T1, T1c, T2, and Flair for the segmentation. DNN learns from each pixel of the image and segments the brain region more accurately. Table 5describes the overview of recent development for brain tumor segmentation. State-of-the-art neuroimaging techniques are available for the detection of visible and invisible tumor cells. The variability in the shape and size of the tumor increases difficulties for automatic image segmentation. A hybrid Random Forest and Support Vector Machine (RF-SVM)-based method learns from the complex characteristics of the tumor lesion. RF-SVM consists of two-stage cascade in the first stage, random forest learns from the tumor label space and, at the second stage, the predicted features are fed into the SVM for classification. RF-SVM performs well as it is used solely for the segmentation. Fully Convolutional Network (FCN) is used for segmentation of the tumor region and modifies the network with bounce structural chart to facilitate the semantic requirements for segmentation. Three-dimensional CNN is used for segmentation of the brain tumor. S. Kumar uses UNET and crops the image when fed into the network for better results. The interactive deep-learning-based frame work consists of the integration of CNNs into the bounding box and the scribble-based image segmentation

1.2 MACHINE LEARNING ALGORITHM:

Three machine learning classification model Decision Tree, Random forest and Support vector machine has been selected to detect phishing websites.

Decision Tree Algorithm [5] One of the most widely used algorithm in machine learning technology. Decision tree algorithm is easy to understand and also easy to implement. Decision tree begins its

work by choosing best splitter from the available attributes for classification which is considered as a root of the tree. Algorithm continues to build tree until it finds the leaf node. Decision tree creates training model which is used to predict target value or class in tree representation each internal node of the tree belongs to attribute and each leaf node of the tree belongs to class label. In decision tree algorithm, gini index and information gain methods are used to calculate these nodes.

Random Forest Algorithm [6] Random forest algorithm is one of the most powerful algorithms in machine learning technology and it is based on concept of decision tree algorithm. Random forest algorithm creates the forest with number of decision trees. High number of tree gives high detection accuracy. Creation of trees are based on bootstrap method. In bootstrap method features and samples of dataset are randomly selected with replacement to construct single tree. Among randomly selected features, random forest algorithm will choose best splitter for the classification and like decision tree algorithm; Random forest algorithm also uses gini index and information gain methods to find the best splitter. This process will get continue until random forest creates n number of trees. Each tree in forest predicts the target value and then algorithm will calculate the votes for each predicted target. Finally random forest algorithm considers high voted predicted target as a final prediction.

Support Vector Machine Algorithm [7] Support vector machine is another powerful algorithm in machine learning technology. In support vector machine algorithm each data item is plotted as a point in n-dimensional space and support vector machine algorithm constructs separating line for classification of two classes, this separating line is well known as hyperplane. Support vector machine seeks for the closest points called as support vectors and once it finds the closest point it draws a line connecting to them. Support vector machine then construct separating line which bisects and perpendicular to the connecting line. In order to classify data perfectly the margin should be maximum. Here the margin is a distance between hyperplane and support vectors. In real scenario it is not possible to separate complex and nonlinear data, to solve this problem support vector machine uses kernel trick which transforms lower dimensional space to higher dimensional space.

Phishing is one of the major problems of the information security. It can occur in two ways, either by receiving suspicious emails that lead to the fraudulent site or by users accessing links that go directly to a phishing website. However, the two methods are common in one thing, which is the attacker targets human vulnerabilities rather than software vulnerabilities. Phishing can be described as fraudsters that try to manipulate the user into giving them their personal information such as username, password, and a credit card number. These scams are leading to economic and financial crises for users [4]. In the early 90s, phishers created a false account with a fake identity and fake credit card on the America Online (AOL) company that provided a web portal and was an online service provider. In this way, the phishers could be exploiting its services without any cost to them. Since then, in the mid-90s, AOL strengthened

its system to prevent phishers. Unfortunately, the phishers used another method, Stealing valid accounts by acting as an AOL employee and requesting users provide their password for security purposes. This occurred either by email or via instant message services.
Recently, there have been several studies trying to solve the phishing problem. They can be categorized into four categories: blacklist, heuristic, content analysis, and machine learning techniques. The blacklist

Compares the URL with an existing database that contains a list of phishing website URLs. Because of the rapid increase of phishing websites, the blacklist approach has become inefficient in deciding whether each URL is a phishing website or not, and this kind of delay can lead to zero-day attacks from new phishing sites [4]. The heuristics approach uses the signature databases of any known attacks, to match it with the signature of a heuristic pattern. The trade-off of using heuristics is failing to detect novel attacks, as it is easy to bypass the signatures through obfuscation. Also, updating the signature database is slow considering the growth of novel attacks, especially zero-day attacks [7]. Content analysis is a content-based approach in detecting phishing websites, using well-known algorithms such as term frequency/inverse document frequency (TF-IDF). It analyses the text-based content of a page itself to decide whether the website is phishing or not. Additionally, measuring website traffic using Alexa is another method that has been implemented by researchers to detect phishing websites.

Machine learning takes advantage of its predictive power. It learns the characteristics of the phishing website and then predicts new phishing characteristics. There are several techniques, such as nave Bayes (NB), decision tree (DT), support vector machines (SVM), RF, artificial neural network (ANN), and Bayesian net (BN). The accuracy of phishing detection varies from one algorithm to another.

1.3 LITERATURE STUDY:

Recently, Machine learning (ML) and Deep Learning (DL) methods are widely been used for detection and grading brain tumors using different imaging modalities, especially those acquired using MRI. In this section, the most recent and related research works on the paper topic are presented. Mohsen, Heba, et all. [8] Propose a system that combines discrete wavelet transform (DWT) features and deep learning (DL) techniques. They have used fuzzy c-mean method for segmenting the brain tumor, and for each detected lesion the DWT was applied to extract the features, where these features are fed into the principal component analysis (PCA) for feature dimension reduction and finally the selected features are then fed to deep neural networks (DNN). The results show that they achieve an accuracy rate of

96.97% and a sensitivity of 97.0 %. Widhiarso, Wijang, Yohannes Yohannes, and Cendy Prakarsah [10] presented a brain tumor classification system using a convolutional neural network (CNN) and Gray Level Co-occurrence Matrix (GLCM) based features. They extracted four features (Energy, Correlation, Contrast, and Homogeneity) from four angles (0°, 45°, 90°, and 135°) for each image and then these features are fed into CNN, they tested their methodology on four different datasets (Mg-Gl, Mg-Pt, Gl-Pt, and Mg-Gl-Pt) and the best accuracy achieved was82.27% for Gl-Pt dataset using two sets of features; contrast with homogeneity and contrast with correlation. Seetha, J., and S. S. Raja [12] proposed a deep CNN based system for automated brain tumor detection and grading. The system is based on Fuzzy C-Means (FCM) for brain segmentation and based on these segmented regions a texture and shape features were extracted then these features were fed into SVM and DNN classifiers. The results showed that the system achieved a rate of 97.5% accuracy. On the other hand, Cheng, Jun, et al. [13] enhanced the performance of the brain tumor classification process using region of interest (ROI) augmentation and fine ring-form partition. They applied these enhancements to different feature extractions methods which are intensity histogram, GLCM, and the bag-of-words (BoW) where these features vectors are fed into a classifier. The experimental results showed that the accuracy enhanced from 71.39% to 78.18%, and 83.54% to 87.54%, and 89.72% to 91.28% for intensity histogram, GLCM, and BoW respectively. Sasikala, M., and N. Kumaravel. [17] Proposed a genetic algorithm feature selection for feature dimension reduction of wavelet features set. The method is based on selecting optimal features vector that can be fed into the selected classifier such as an artificial neural network (ANN). The results show that the genetic algorithm selected only 4 of 29 features and achieved an accuracy of 98% using only the selected features. Khawaldeh, Saed, et al. [23] proposed a system for non-invasive grading of glioma brain tumors using a modified version of AlexNet CNN. The classification process was done using whole-brain MRI images and the labels of the images were at the image level, not the pixel level. The experimental results showed that the method achieved a reasonable performance with an accuracy of 91.16%. Sajjad, Muhammad et al. [24] proposed an extensive data augmentation method fused with CNN for brain tumor classification. The method used for multi-grade classification of brain tumors using segmented brain tumor MRI images. They used pre trained VGG-19 CNN architecture for classification using transferee learning and achieved an overall accuracy of values 87.38% and 90.67% for data before and after augmentation respectively. While Özyurt, Fatih et al. [25] combine the CNN with neutrosophic expert maximum fuzzy (NS-CNN) sure entropy for brain tumor classification. They used the neutrosophic set – expert maximum fuzzy-sure method for brain tumor segmentation then these images are fed to CNN to extract features and fed them to SVM classifiers to be classified as benign or malignant. They achieved an average success of 95.62%.

In [7], an automated method is used to identify and categorize MRI images. This method is based on

the Super Pixel Technique and the classification of each Super Pixel. Extremely randomized trees (ERT) classifier is compared with SVM to classify each super pixel into tumor and normal. This method has two datasets, which are 19 MRI FLAIR images and BRATS 2012 dataset. The results demonstrate the good performance of this method using ERT classifier. In [8], an automatic classification method is used to identify a tumor using a CNN with 3×3 small kernels. The method obtained simultaneously the first position for the complete, core, and enhancing regions in dice similarity, coefficient metric (0.88, 0.83, 0.77), at the BRATS Challenge 2013. In [9], Alexnet model CNN is used to simultaneously diagnose MS and normal tumors. The CNN was able to accurately classify 98.67% images correctly into three classes. In [10], a multi-stage Fuzzy C-Means (FCM) framework was proposed to segment brain tumors

from MRI images. In [11], an efficient and effective method which uses CNNs used for classification and segmentation. The proposed method, used Image-Net for extract features. The results obtained 97.5% accuracy for classification and 84% accuracy for segmentation. In [12], multiphase MRI images in tumor grading have been studied and a comparison has been made between the results of deep learning structures and base neural networks. The results show that the network performance based on the sensitivity and specificity of CNN improved by 18% compared to the neural networks. In [13], a deep learning-based supervised method is introduced to detect synthetic aperture radar (SAR) image changes. This method provided a dataset with an appropriate data volume and diversity for training the DBN using input images and the images obtained from applying the morphological operators on them. The detection performance of this method indicates the appropriability of deep learning based algorithms for solving the change detection problems. In paper [14], a completely automated brain tumor classification method is proposed based on DNN. The proposed networks have been designed to be used in low-grade and high glioblastoma disease images. In this paper, a new architecture of CNN is presented. The proposed a cascading architecture is proposed in which the output of a core CNN is used as an additional source of information for the next CNN.

1.4 PYTHON

Python is a programming language, which means it ’a language both people and computers can understand. Python was developed by a Dutch software engineer named Guido van Rossum, who created the language to solve some problems he saw in computer languages of the time.

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, and a syntax that allows programmers to express concepts in fewer lines of code, notably

using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

Python interpreters are available for many operating systems. C Python, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations. C Python is managed by the non-profit Python Software Foundation.

You Can Use Python for Pretty Much Anything

One significant advantage of learning Python is that it's a general-purpose language that can be applied in a large variety of projects. Below are just some of the most common fields where Python has found its use:

- Data science
- Scientific and mathematical computing
- Web development
- Computer graphics
- Basic game development
- Mapping and geography (GIS software)

Python Is Widely Used in Data Science

Python's ecosystem is growing over the years and it's more and more capable of the statistical analysis. It's the best compromise between scale and sophistication (in terms od data processing).

Python emphasizes productivity and readability.

Python is used by programmers that want to delve into data analysis or apply statistical techniques (and by devs that turn to data science)

There are plenty of Python scientific packages for data visualization, machine learning, natural language processing, complex data analysis and more. All of these factors make Python a great tool for scientific computing and a solid alternative for commercial packages such as MatLab. The most popular libraries and tools for data science are:

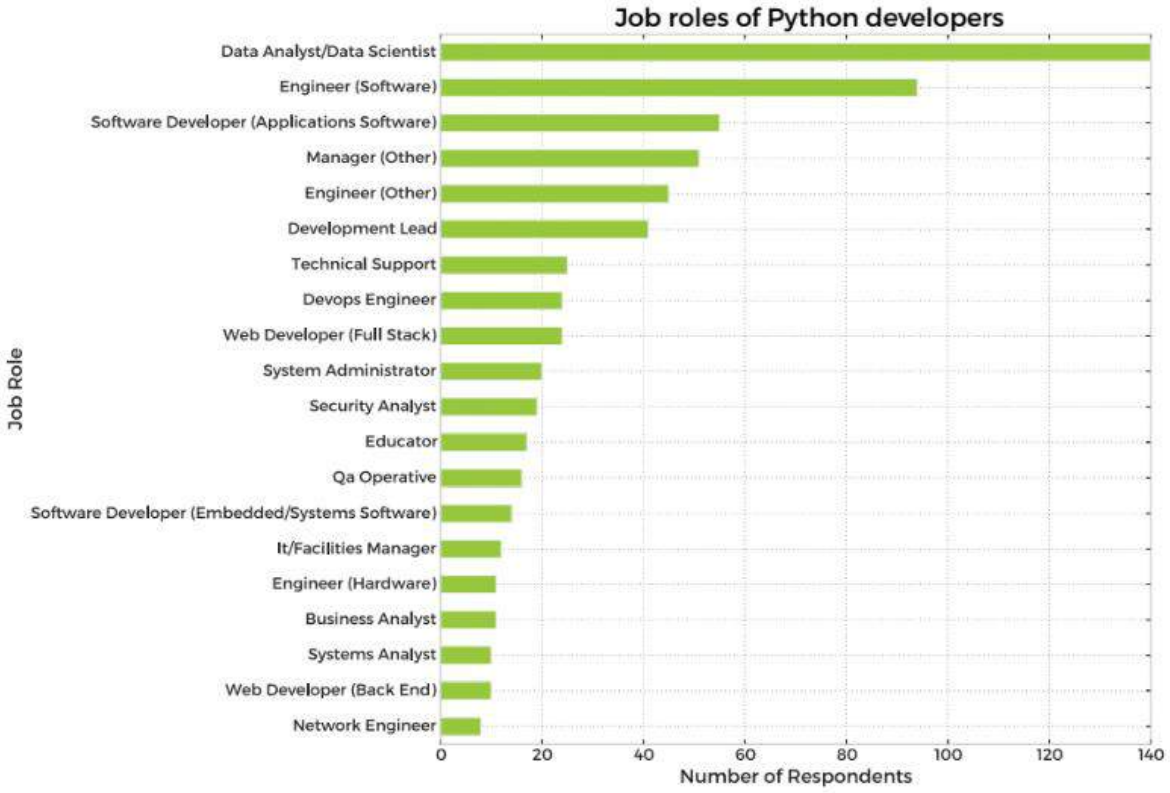
Pandas: a library for data manipulation and analysis. The library provides data structures and operations for manipulating numerical tables and time series.

NumPy: the fundamental package for scientific computing with Python, adding support for large, multi-dimensional arrays and matrices, along with a large library of high-level mathematical functions to operate on these arrays.

SciPy: a library used by scientists, analysts, and engineers doing scientific computing and technical computing.

Being a free, cross-platform, general-purpose and high-level programming language, Python has been widely adopted by the scientific community. Scientists value Python for its precise and efficient syntax, relatively flat learning curve and the fact that it integrates well with other languages (e.g. C/C++).

As a result of this popularity there are plenty of Python scientific packages for data visualization, machine learning, natural language processing, complex data analysis and more. All of these factors make Python a great tool for scientific computing and a solid alternative for commercial packages such as MatLab.



Here's our list of the most popular Python scientific libraries and tools

Astropy

The Astropy Project is a collection of packages designed for use in astronomy. The core astropy package contains functionality aimed at professional astronomers and astrophysicists, but may be useful to anyone developing astronomy software.

Biopython

Biopython is a collection of non-commercial Python tools for computational biology and bioinformatics. It contains classes to represent biological sequences and sequence annotations, and it is able to read and write to a variety of file formats.

Cubes

Cubes is a light-weight Python framework and set of tools for the development of reporting and analytical applications, Online Analytical Processing (OLAP), multidimensional analysis and browsing of aggregated data.

Deap

DEAP is an evolutionary computation framework for rapid prototyping and testing of ideas. It incorporates the data structures and tools required to implement most common evolutionary computation techniques such as genetic algorithm, genetic programming, evolution strategies, particle swarm optimization, differential evolution and estimation of distribution algorithm.

Scoop

SCOOP is a Python module for distributing concurrent parallel tasks on various environments, from heterogeneous grids of workstations to supercomputers.

PsychoPy

PsychoPy is a package for the generation of experiments for neuroscience and experimental psychology. PsychoPy is designed to allow the presentation of stimuli and collection of data for a wide range of neuroscience, psychology and psychophysics experiments.

Pandas

Pandas is a library for data manipulation and analysis. The library provides data structures and operations for manipulating numerical tables and time series.

Mlpy

Mlpy is a machine learning library built on top of NumPy/SciPy, the GNU Scientific Libraries. Mlpy provides a wide range of machine learning methods for supervised and unsupervised problems and it is aimed at finding a reasonable compromise between modularity, maintainability, reproducibility, usability and efficiency.

matplotlib

Matplotlib is a python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib allows you to generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, and more.

NumPy

NumPy is the fundamental package for scientific computing with Python, adding support for large,

multi-dimensional arrays and matrices, along with a large library of high-level mathematical functions to operate on these arrays.

NetworkX

NetworkX is a library for studying graphs which helps you create, manipulate, and study the structure, dynamics, and functions of complex networks.

TomoPy

TomoPy is an open-sourced Python toolbox to perform tomographic data processing and image reconstruction tasks. TomoPy provides a collaborative framework for the analysis of synchrotron tomographic data with the goal to unify the effort of different facilities and beamlines performing similar tasks.

Theano

Theano is a numerical computation Python library. Theano allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently.

SymPy

SymPy is a library for symbolic computation and includes features ranging from basic symbolic

arithmetic to calculus, algebra, discrete mathematics and quantum physics. It provides computer algebra capabilities either as a standalone application, as a library to other applications, or live on the web.

SciPy

SciPy is a library used by scientists, analysts, and engineers doing scientific computing and technical computing. SciPy contains modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other tasks common in science and engineering.

Scikit-learn

Scikit-learn is a machine learning library. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Scikit-image

Scikit-image is a image processing library. It includes algorithms for segmentation, geometric transformations, color space manipulation, analysis, filtering, morphology, feature detection, and more.

ScientificPython

ScientificPython is a collection of modules for scientific computing. It contains support for geometry, mathematical functions, statistics, physical units, IO, visualization, and parallelization.

SageMath

SageMath is mathematical software with features covering many aspects of mathematics, including algebra, combinatorics, numerical mathematics, number theory, and calculus. SageMath uses the Python, supporting procedural, functional and object-oriented constructs.

Veusz

Veusz is a scientific plotting and graphing package designed to produce publication-quality plots in popular vector formats, including PDF, PostScript and SVG.

Graph-tool

Graph-tool is a module for the manipulation and statistical analysis of graphs.

SunPy

SunPy is a data-analysis environment specializing in providing the software necessary to analyze solar and heliospheric data in Python.

Bokeh

Bokeh is a Python interactive visualization library that targets modern web browsers for presentation. Bokeh can help anyone who would like to quickly and easily create interactive plots, dashboards, and data applications. Its goal is to provide elegant, concise construction of novel graphics in the style of D3.js, but also deliver this capability with high-performance interactivity over very large or streaming datasets.

TensorFlow

TensorFlow is an open source software library for machine learning across a range of tasks, developed by Google to meet their needs for systems capable of building and training neural networks to detect and decipher patterns and correlations, analogous to the learning and reasoning which humans use. It is currently used for both research and production at Google products, often replacing the role of its closed-source predecessor, DistBelief.

Nilearn

Nilearn is a Python module for fast and easy statistical learning on NeuroImaging data. Nilearn makes it easy to use many advanced machine learning, pattern recognition and multivariate statistical techniques on neuroimaging data for applications such as MVPA (Mutli-Voxel Pattern Analysis), decoding, predictive modelling, functional connectivity, brain parcellations, connectomes.

Dmelt

DataMelt, or DMelt, is a software for numeric computation, statistics, analysis of large data volumes ("big data") and scientific visualization. The program can be used in many areas, such as natural

sciences, engineering, modeling and analysis of financial markets. DMelt can be used with several scripting languages including Python/Jython, BeanShell, Groovy, Ruby, as well as with Java.

Python-weka-wrapper

Weka is a suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. It contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions. The python-weka-wrapper package makes it easy to run Weka algorithms and filters from within Python.

Dask

Dask is a flexible parallel computing library for analytic computing composed of two components: 1) dynamic task scheduling optimized for computation, optimized for interactive computational workloads, and 2) Big Data collections like parallel arrays, dataframes, and lists that extend common interfaces like NumPy, Pandas, or Python iterators to larger-than-memory or distributed environments.

Python Saves Time

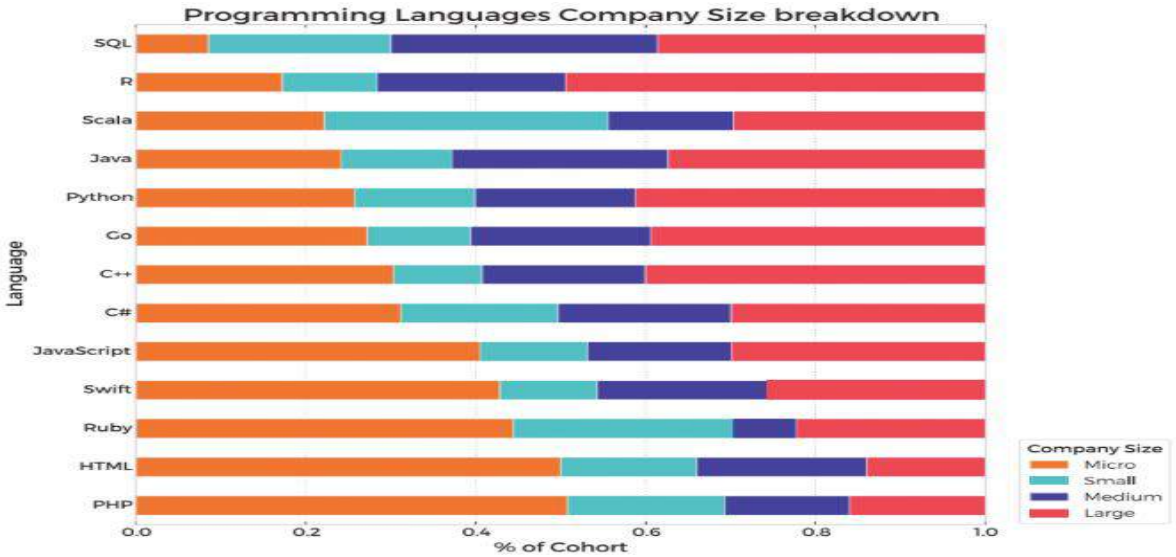
Even the classic “Hello, world” program illustrates this point:

```
print("Hello, world")
```

For comparison, this is what the same program looks like in Java:

```
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello, world");
    }
}
```

All the Big Names Use Python



Python Keywords and Identifier

Keywords are the reserved words in Python.

We cannot use a keyword as variable name, function name or any other identifier. They are used to define the syntax and structure of the Python language.

In Python, keywords are case sensitive.

There are 33 keywords in Python 3.3. This number can vary slightly in course of time.

All the keywords except True, False and None are in lowercase and they must be written as it is. The list of all the keywords is given below.

Keywords in Python programming language

False	class	finally	is	return
None	continue	for	lambda	try
True	def	from	nonlocal	while
and	del	global	not	with
as	elif	if	or	yield
assert	else	import	pass	
break	except	in	raise	

Identifier is the name given to entities like class, functions, variables etc. in Python. It helps differentiating one entity from another.

Rules for writing identifiers

Identifiers can be a combination of letters in lowercase (a to z) or uppercase (A to Z) or digits (0 to 9) or an underscore (_). Names like myClass, var_1 and print_this_to_screen, all are valid example.

An identifier cannot start with a digit. 1variable is invalid, but variable1 is perfectly fine.

Keywords cannot be used as identifiers.

```
>>> global = 1
  File "<interactive input>", line 1
    global = 1
           ^
```

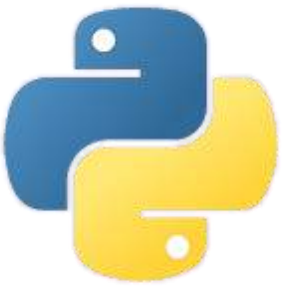
SyntaxError: invalid syntax

We cannot use special symbols like !, @, #, $, % etc. in our identifier.

```
>>> a@ = 0
  File "<interactive input>", line 1
    a@ = 0
     ^
SyntaxError: invalid syntax
```

Identifier can be of any length.

Python is a broadly utilized abnormal state programming dialect for universally useful programming Python highlights a dynamic sort framework and programmed memory administration and backings various programming ideal models, including object-arranged, basic, utilitarian programming, and procedural styles. It has an expansive and complete standard library. Python mediators are accessible for some working frameworks, permitting Python code to keep running on a wide assortment of frameworks.



Python Logo

2. SYSTEM ANALYSIS

2 SYSTEM ANALYSIS

2.1 PROBLEM DEFINITION

The deep learning concept to perform an automated brain tumors classification using brain MRI images and measure its performance. The proposed methodology aims to differentiate between normal brain and some types of brain tumors such as glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumors using brain MRI images

2.2 EXISTING SYSTEM

Many attempts have been by researchers in the past for analyzing human barin tumor with the available methods but the methods has not produced enough results for the tumor identification. When the methods were applied on the image scans of the brain, the results are not even satisfying.

2.2.1 DISADVANTAGES

The system was not able to measure the tumor availability
Working with image is a big bottle neck an finding the resources from the dataset is more difficult.

2.3 PROPOSED SYSTEM

Our proposed methodology based on the DNN learning architecture for classification where the classifier is identifying the brain tumors in brain MRIs.

The proposed methodology uses a set of features extracted by the discrete wavelet transform (DWT) feature extraction technique from the segmented brain MRI images, to train the DNN classifier for brain tumors classification.

2.3.1 ADVANTAGES OF THE PROPOSED SYSTEM

Makes the prediction of the disease in earlier way this helps us to save the patient’s life.
Image analysis will help us to understand more about the brain and the diseases causes.

3. SOFTWARE REQUIREMENTS SPECIFICATION

3. SOFTWARE REQUIREMENTS SPECIFICATION

The reason for this SRS record is to distinguish the necessities and functionalities for Intelligent Network Backup Tool. The SRS will characterize how our group and the customer consider the last item and the attributes or usefulness it must have. This record additionally makes a note of the discretionary prerequisites which we intend to execute yet are not required for the working of the venture.

This stage assesses the required necessities for the Images Processing for an orderly method for assessing the prerequisites a few procedures are included. The initial step associated with dissecting the prerequisites of the framework is perceiving the idea of framework for a solid examination and all the case are defined to better comprehend the investigation of the dataset.

INTENDED AUDIENCE AND READING SUGGESTIONS

This record is proposed for extend engineers, directors, clients, analyzers and documentation journalists. This report goes for examining plan and execution imperatives, conditions, framework highlights, outside interface prerequisites and other non-utilitarian necessities.

IDENTIFICATION OF NEEDS

The first and imperative need for a business firm or an association is to know how they are performing in the market and parallelly they have to know how to conquer their rivals in the market. To do as such we have to investigation our information in view of all the accessible variables

3.1 FEASIBILITY STUDY

A credibility contemplate expects to fair-mindedly and soundly uncover the qualities and inadequacies of a present business or proposed meander, openings and threats present in nature, the benefits required to bring through, and in the long run the prospects for advance. In its most clear terms, the two criteria to judge believability are incurred significant injury required and motivator to the fulfilled.

An inside and out arranged feasibility ponder should give a recorded establishment of the business or wander, a delineation of the thing or organization, accounting explanations, purposes of

enthusiasm of the operations and organization, publicizing examination and game plans, budgetary data, authentic necessities and cost duties. All things considered, plausibility looks at go before specific change and wander utilization. There are three sorts of attainability

- Economical Feasibility
- Technical Feasibility
- Operational Feasibility

3.1.1 ECONOMICAL FEASIBILITY

The electronic structure manages the present existing system's data stream and technique absolutely and should make each one of the reports of the manual structure other than a substantial gathering of the other organization reports. It should be filled in as an electronic application with specific web server and database server. Advance a segment of the associated trades happen in different ranges. Open source programming like TOMCAT, JAVA, MySQL and Linux is used to restrict the cost for the Customer. No extraordinary wander need to manage the instrument.

3.1.2 TECHNICAL FEASIBILITY

Surveying the particular probability is the trickiest bit of a believability consider. This is in light of the fact that, starting at the present moment, not a lot of point by point layout of the system, making it difficult to get to issues like execution, costs on (by excellence of the kind of development to be passed on) et cetera.

Different issues must be considered while doing a particular examination. Grasp the differing progressions required in the proposed system. Before starting the wander, we should be clear about what are the advances that are to be required for the change of the new system. Check whether the affiliation by and by has the required advancements. Is the required development open with the affiliation?
In case so is the utmost sufficient?
For instance – "Will the present printer have the ability to manage the new reports and structures required for the new system?"

3.1.3 OPERATIONAL FEASIBILITY

Proposed wanders are profitable just if they can be changed into information systems that will meet the affiliations working necessities. Simply communicated, this trial of probability asks with reference to whether the structure will work when it is made and presented. Are there genuine obstacles to Implementation? Here are questions that will help test the operational achievability of a wander.

- Is there sufficient help for the wander from organization from customers? In case the present structure is particularly cherished and used to the extent that individuals won't have the ability to see purposes behind change, there may be resistance.
- Are the present business methodologies qualified to the customer? If they are not, Users may welcome a change that will accomplish a more operational and supportive systems.
- Have the customer been locked in with the orchestrating and change of the wander? Early commitment decreases the chances of impenetrability to the structure.

Different issues must be considered while doing a particular examination. Grasp the differing progressions required in the proposed system. Before starting the wander, we should be clear about what are the advances that are to be required for the change of the new system. Check whether the affiliation by and by has the required advancements. Is the required development open with the affiliation?
In case so is the utmost sufficient?

- For instance – "Will the present printer have the ability to manage the new reports and structures required for the new system?".

3.2 HARDWARE & SOFTWARE REQUIREMENTS

3.2.1 SOFTWARE REQUIREMENTS

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements,

Operating system	:	Windows
Languages	:	Python
Frame work	:	Jupyter
IDE	:	Anacond

3.2.2 HARDWARE REQUIREMENTS

Hardware interfaces specify the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

- Processor : Pentium 4.
- Hard disk : 500 GB and Above.
- RAM : 4GB and Above.
- Monitor : 5 inches or above.

4. SYSTEM DESIGN

4. SYSTEM DESIGN

The System Design Document describes the system requirements, operating environment, system and subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and external interfaces.

This section describes the system in narrative form using non-technical terms. It should provide a high-level system architecture diagram showing a subsystem breakout of the system, if applicable. The high-level system architecture or subsystem diagrams should, if applicable, show interfaces to external systems. Supply a high-level context diagram for the system and subsystems, if applicable. Refer to the requirements trace ability matrix (RTM) in the Functional Requirements Document (FRD), to identify the allocation of the functional requirements into this design document.

This section describes any constraints in the system design (reference any trade-off analyses conducted such, as resource use versus productivity, or conflicts with other systems) and includes any assumptions made by the project team in developing the system design.

The organization code and title of the key points of contact (and alternates if appropriate) for the information system development effort. These points of contact should include the Project Manager, System Proponent, User Organization, Quality Assurance (QA) Manager, Security Manager, and Configuration Manager, as appropriate.

4.1 SYSTEM ARCHITECTURE

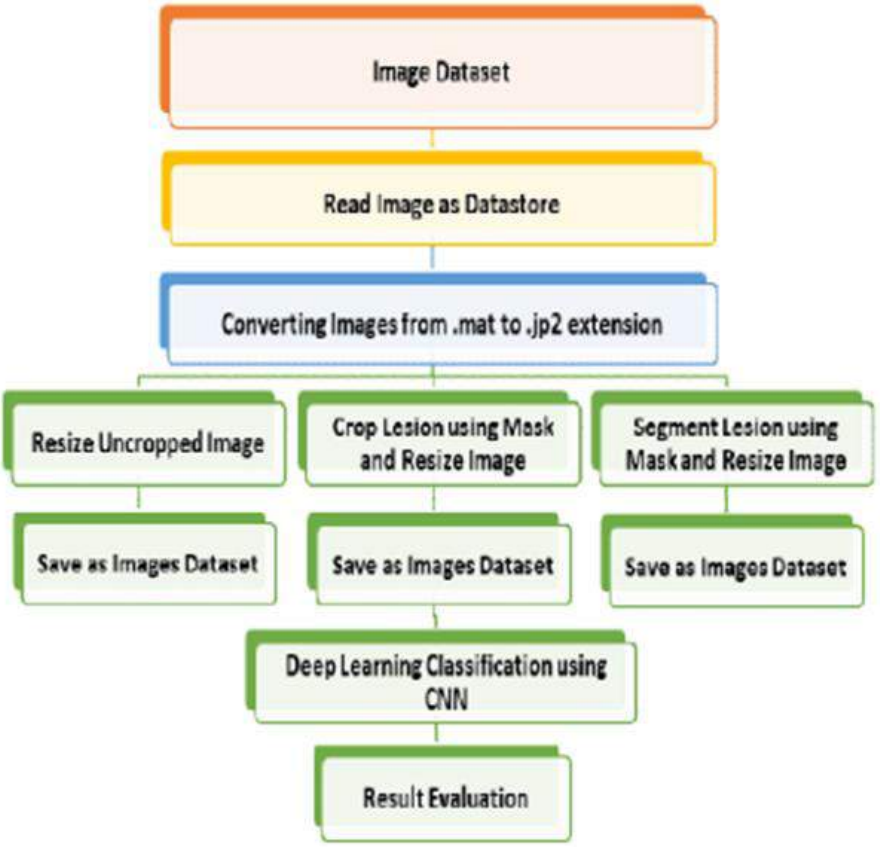


Figure 4.1: System Architecture

4.2 MODULES

Dataset The data set images used in this paper include brain MRI images of 153 patients, including normal and brain tumors patients who referred to imaging centres because of headaches After examination and diagnosis of the doctor, the collected images included brain images of 80 healthy patients Include 1321 images which has 56 images for testing data and 515 images for the train data.73 patient tumors Include 571 images which has 170 images for test data and 1151 images for the train data. Of the total number of patients with brain tumor.

FEATURE EXTRACTION In machine learning and image processing, feature are created from the initial dataset. Which facilitates the learning process. When the input data of an algorithm is too large, it can be converted to a smaller set of features. The process of extracting a subset from the primary features set is called feature extraction [15]. The selected features include information about the input data, so that the reduced representation of the agent can be done instead of the full initial data. One of the important application of feature extraction is in the image processing, which are used to distinguish the desired segments or the shape (features) of a digital image or video stream. Disease, 86 were women

and 68 were men, whose age range from 8 to 66 years old. Of a total of 153 patients, 1892 images were collected, 1666 images for train data and 226 images for test images. The collected images originally had an initial size of 224×224.

4.2.1 SIMULATION

In few cases, some areas of fat in the pictures are mistakenly detected as tumor, or the tumors may not be seen by the physician; the most exact diagnosis is completely depended on the physician's skill. In this paper, the CNN has been used for tumor detection through brain images. There were additional margins of the images gathered from the imaging centers. These margins were cropped to prevent the noise of the images. One of the main reasons for using the feature extraction technique and combining it with the CNN is to retrieve the feature extraction of the images in order to increase the accuracy of the network. According to the results of the CNN on the initial images, in order to improve the network accuract, in this study, a new method which is a combination of Clustering algorithm for feature extraction and CNN is proposed.

4.2.2 FEATURE EXTRACTION METHOD

Convolutional neural method initially, the images were applied to the CNN without any feature extraction methods. The size of the input images is initially 227×227. The Alexnet architect was used to identify and classify the images, which consisted of 5 Convolutional layers and 3 layers of Sub-sampling layers, Normalization layers, Normalization layers, Fully Connected layers and lastly layer the classification layer

4.3 UML DIAGRAMS

UML (Unified Modeling Language) is a standard vernacular for choosing, envisioning, making, and specifying the collectibles of programming structures. UML is a pictorial vernacular used to make programming blue prints. It is in like way used to exhibit non programming structures similarly like process stream in a gathering unit and so forth.

UML is not a programming vernacular yet rather instruments can be utilized to make code in different tongues utilizing UML graphs. UML has an incite relationship with question composed examination and outline. UML expect a fundamental part in portraying trade viewpoints of a structure.

4.3.1 USE CASE DIAGRAM

The use case graph is for demonstrating the direct of the structure. This chart contains the course of action of use cases, performing pros and their relationship. This chart might be utilized to address the static perspective of the structure.

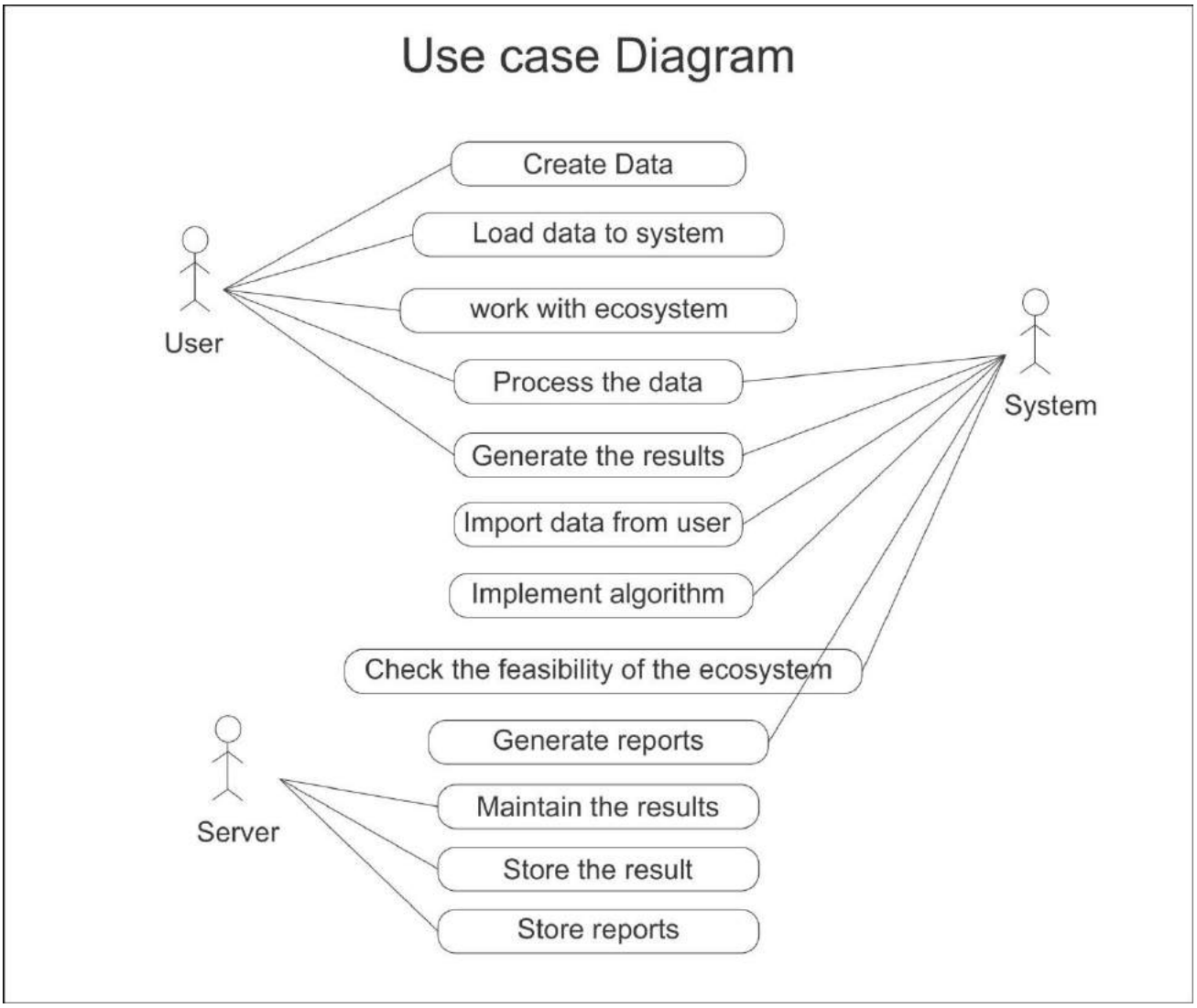


Fig: 4.3.1: Use Case Diagram

In the above diagram, the performing specialists are customer, structure, client, server, Hadoop and data cleaning. The client exchanges the data to the system which disengages the data into squares and gives the data to Hadoop. By then Hadoop does the data cleaning which is just performing data connection and data repairing, by then the results will be secured. These results can be seen using Hadoop and can be secured in server for future reason. The gained results can be created as reports by the customer.

Here we first take the data set from the bank which has various details of the accountant such as applicant personal details, income, co-applicant income, credit score etc.

4.3.2 CLASS DIAGRAM

The class graph is the most normally pulled in layout UML. It addresses the static course of action perspective of the structure. It solidifies the strategy of classes, interfaces, joint attempts and their affiliations.

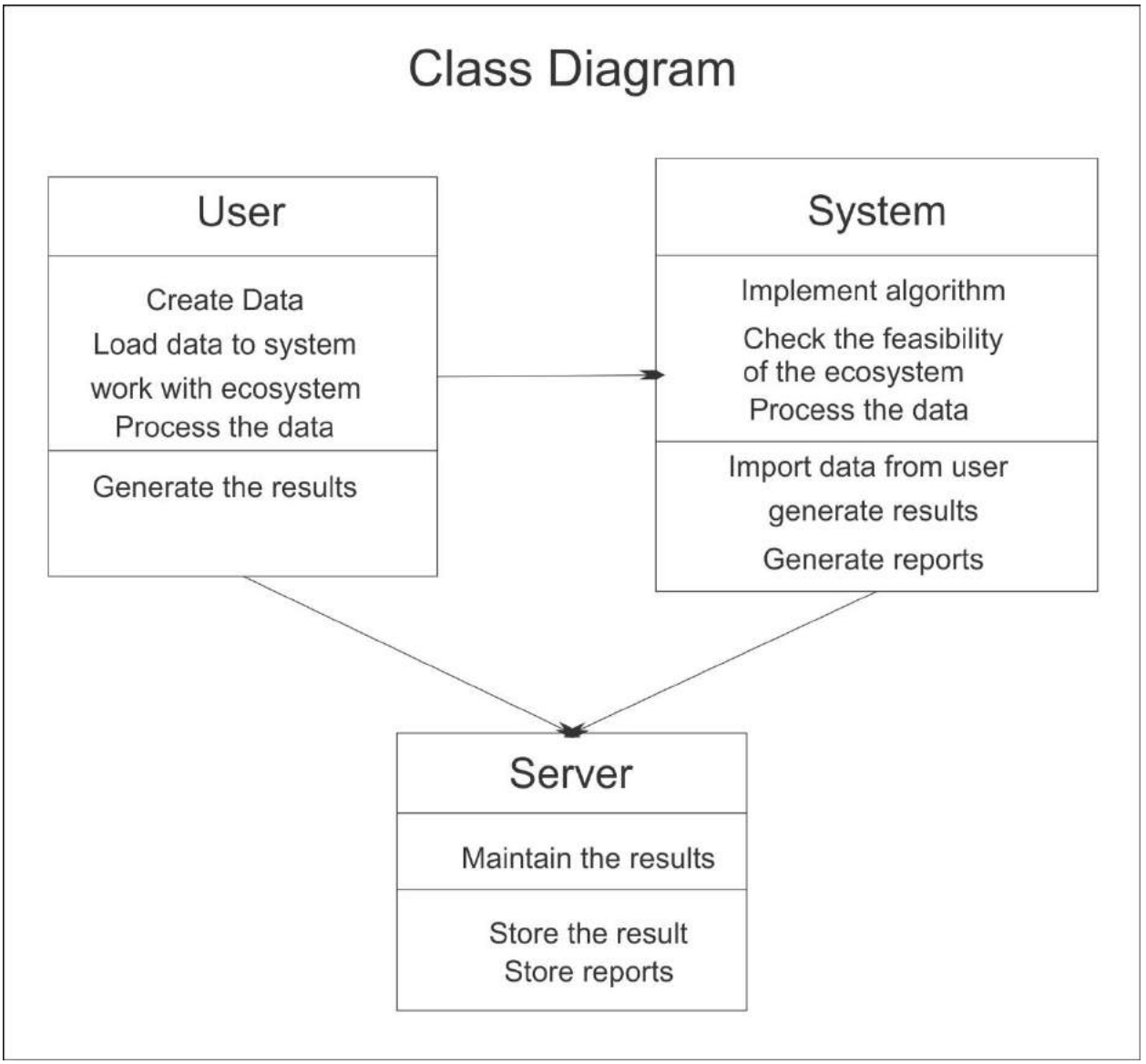


Fig: 4.3.2: Class Diagram

In the above class diagram, the relationship that is the dependence between each one of the classes is sketched out. Additionally, even the operations performed in each and every class is similarly appeared.

4.3.3 SEQUENCE DIAGRAM

This is a cooperation design which tends to the time requesting of messages. It includes set of parts and the messages sent and gotten by the instance of parts. This chart is utilized to address the dynamic perspective of the structure.

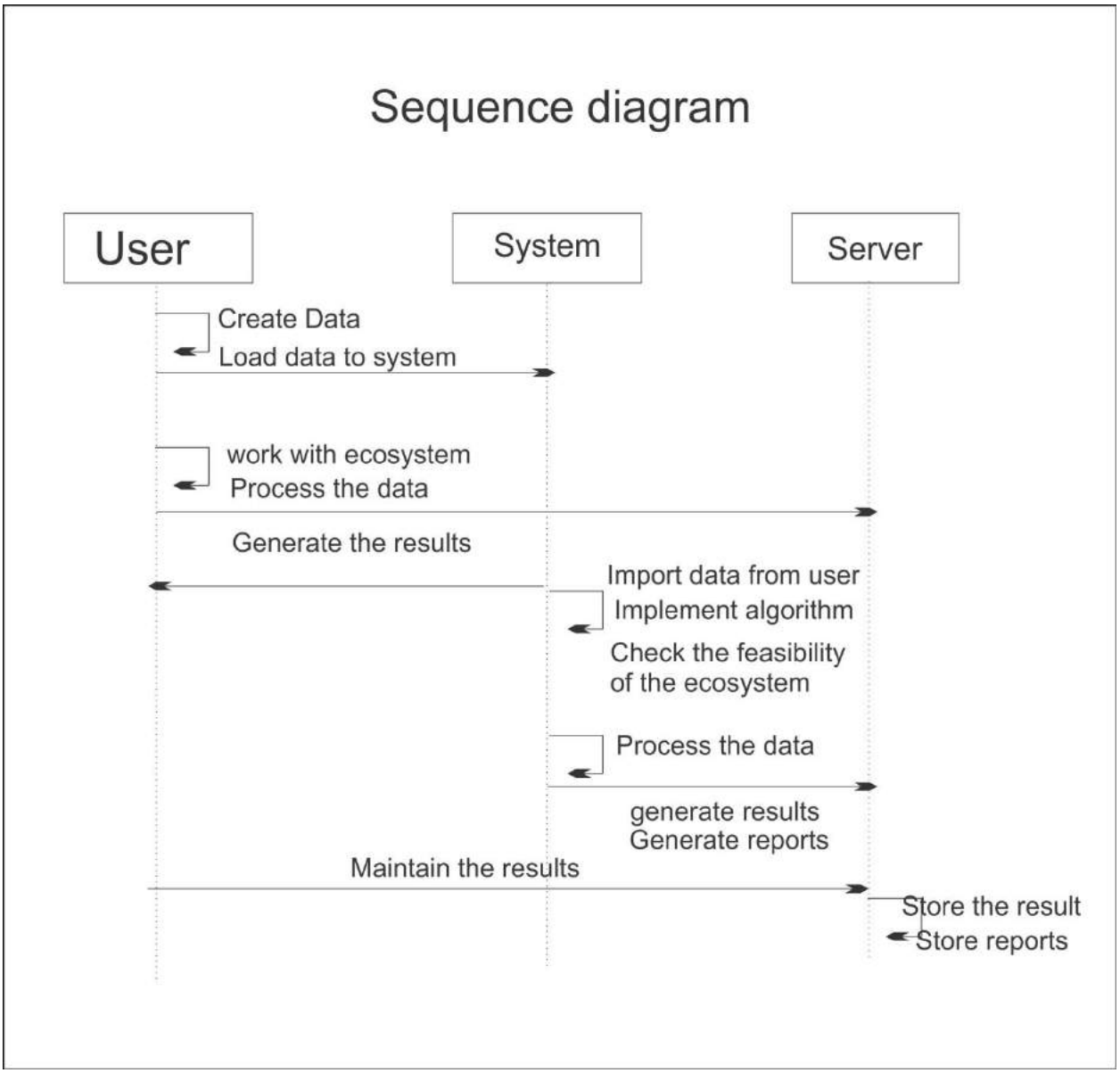


Fig: 4.3.3: Sequence Diagram

A succession outline indicates question communications masterminded in time arrangement. In the above graph, there are five articles cooperating with each other. Each protest has a vertical dashed line which speaks to the presence of a question over some undefined time frame. This graph has additionally a tall, thin rectangle which is called center of control that demonstrates the timeframe amid which a protest is playing out an activity, either specifically or through a subordinate system.

4.3.4 COLLABRATION DIAGRAM

This is a support format, which tends to the principal relationship of articles that send and get messages. It incorporates set of parts, connectors that interface the parts and the messages sent and get by those parts. This graph is utilized to address the dynamic perspective of the framework.

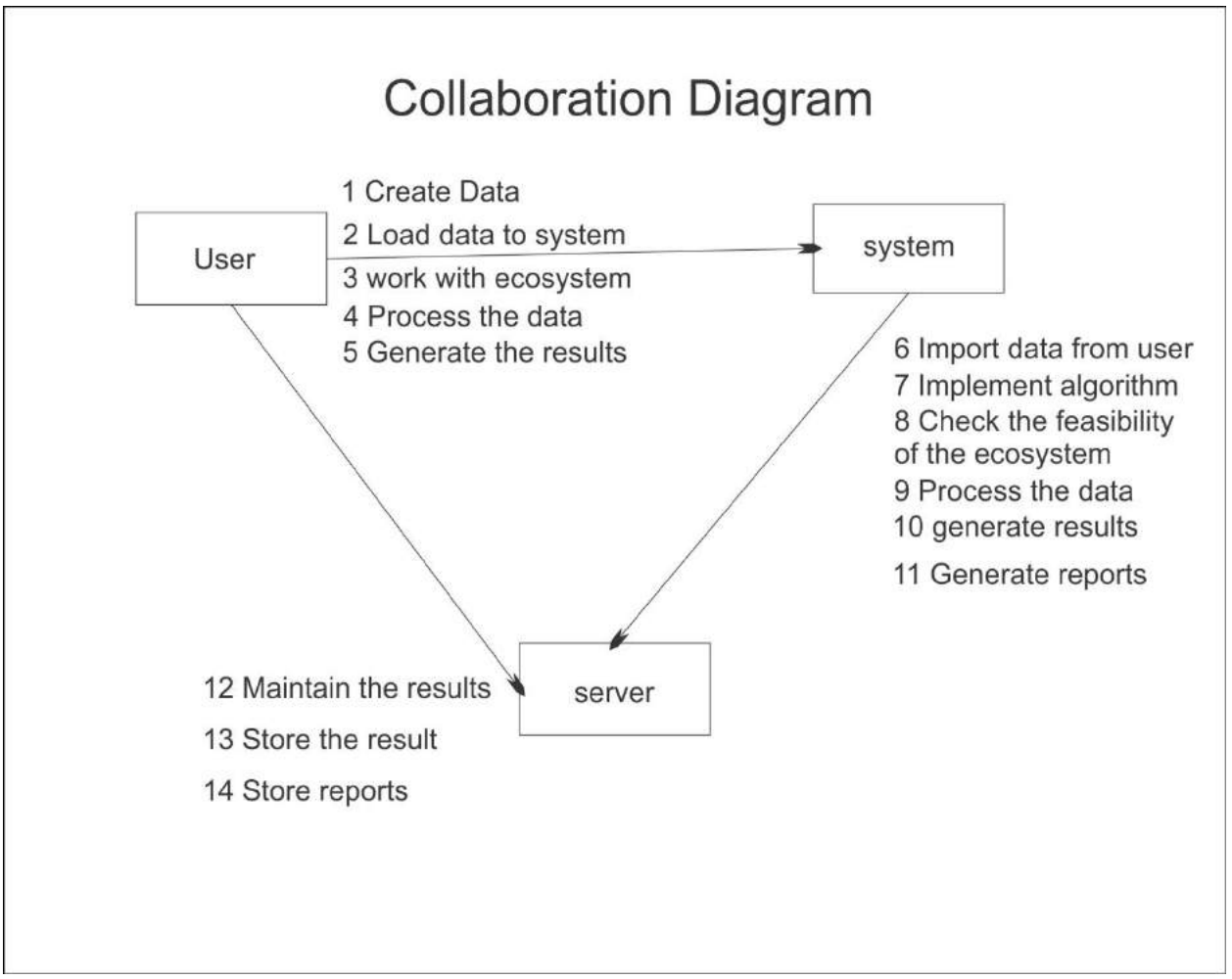


Fig: 4.3.4: Collaboration Diagram

The joint effort outline contains articles, way and arrangement number. In the above graph, there are five questions specifically customer, client, framework, Hadoop and server. These items are connected to each other utilizing a way. A succession number show the time request of a message.

4.3.5 STATE CHART DIAGRAM

The state graph contains the game-plan of states, occasions and exercises. This graph is noteworthy for tending to the lead of the interface, class and made effort. The key centralization of state outline is to show the occasion sort out lead of the request. The state follows diagram the dynamic perspective of the framework.

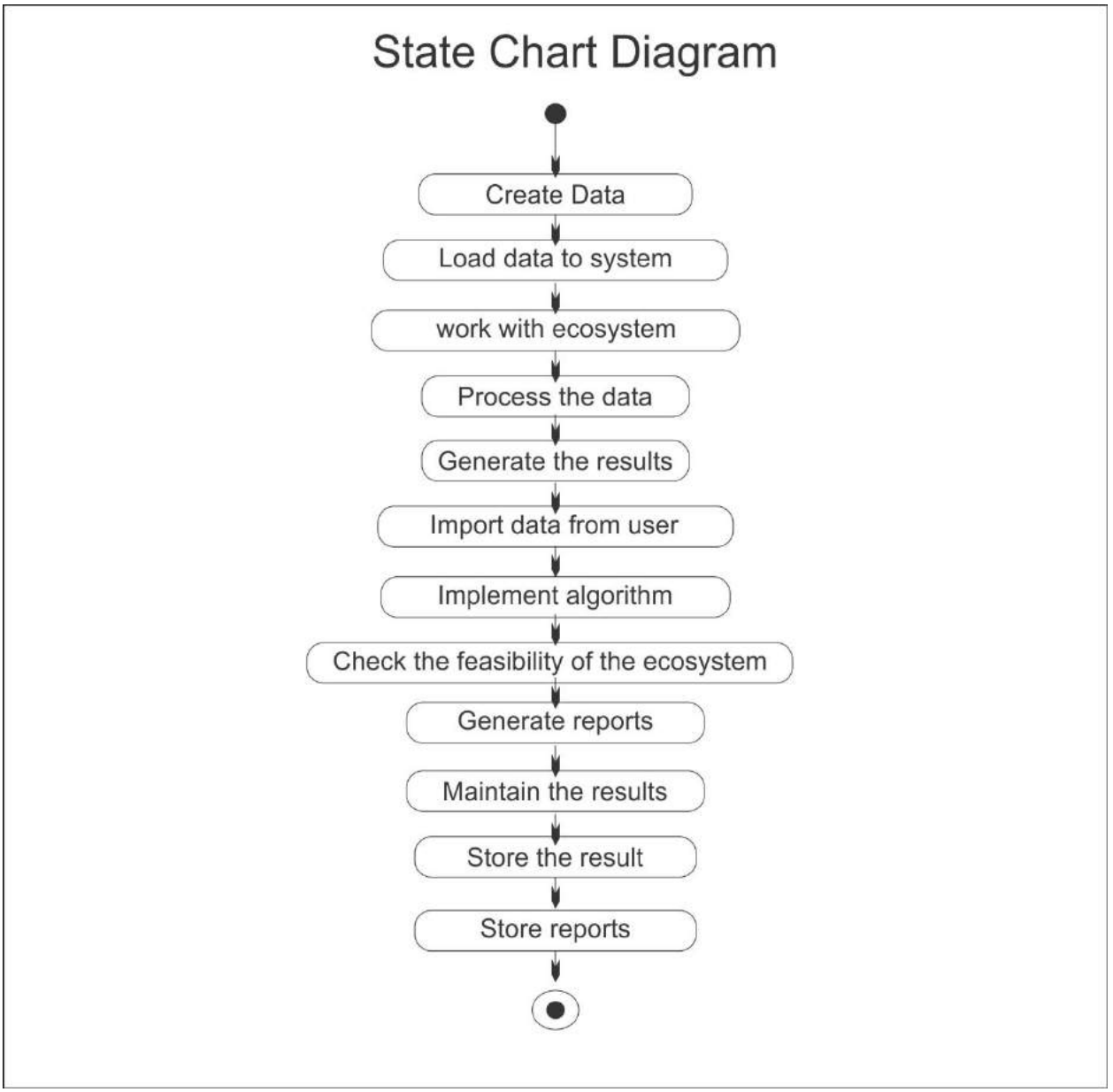


Fig: 4.3.5: State Chart Diagram

A state outline graph contains two components called states and progress. States speak to circumstances amid the life of a question. We can without much of a stretch outline a state in Smart Draw by utilizing a rectangle with adjusted corners. Change is a strong bolt speaks to the way between various conditions of a question. Name the change with the occasion that activated it and the activity those outcomes from it.

4.3.6 COMPONENT DIAGRAM

The imperative portion of part format is segment. This diagram demonstrates within parts, connectors and ports that understand the piece. Precisely when section is instantiated, duplicates of inside parts are besides instantiated.

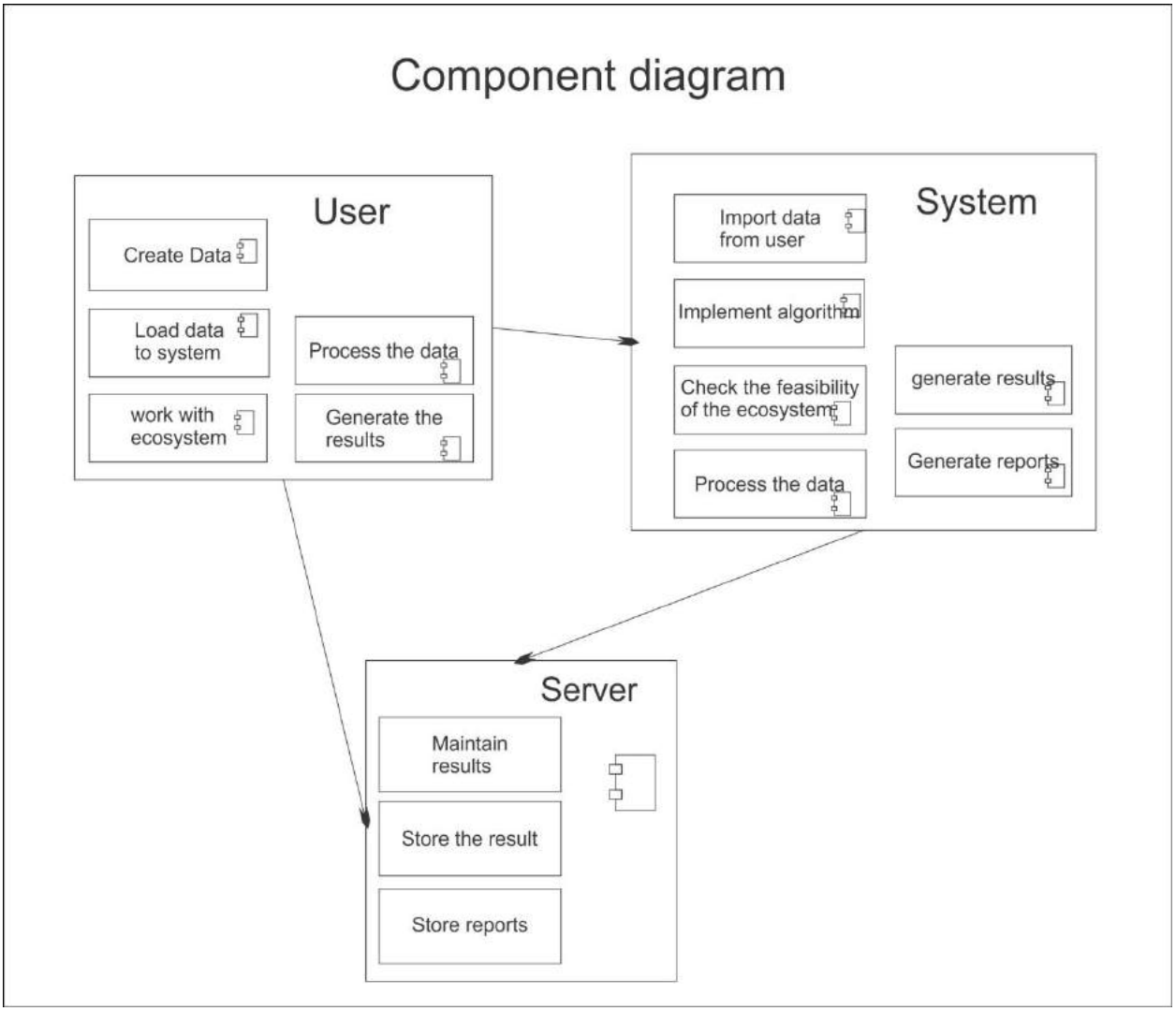


Fig: 4.3.6: Component Diagram

A part outline is spoken to utilizing segment. A part is a physical building piece of the framework. It is spoken to as a rectangle with tab. Part outline portrays the inward handling of the venture. The information is sent to the Hadoop where sqoop is utilized for information cleaning and the reports are produced utilizing hive.

4.3.7 DEPLOYMENT DIAGRAM

The fundamental fragment in game-plan layout is a middle point. The strategy of focus focuses and their relationship with other is tended to utilizing sending plot. The sending outline is identified with the area diagram, that is one focus purpose obviously of activity format frequently includes no short of what one sections. This outline is in like way critical for tending to the static perspective of the framework.

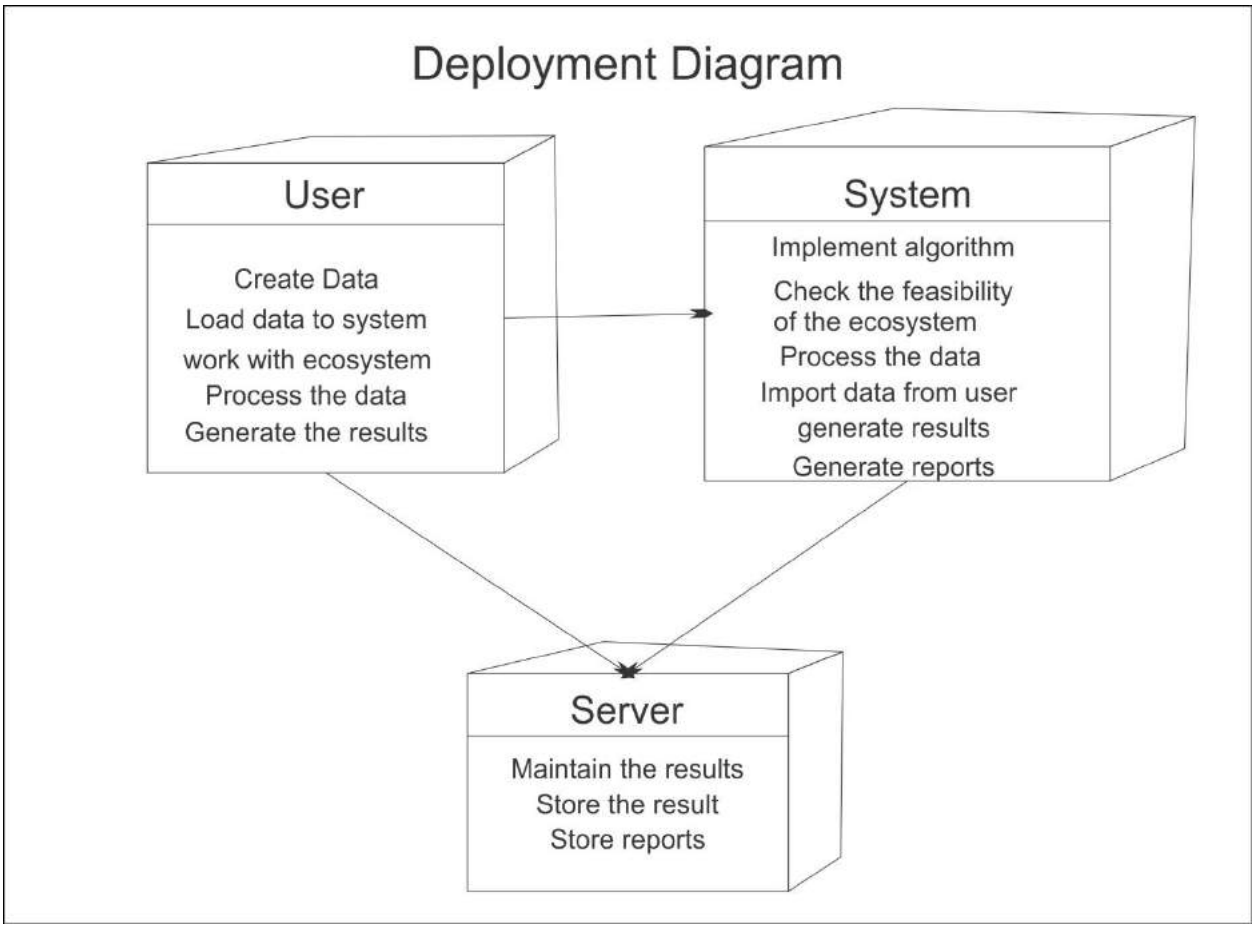


Fig: 4.3.7: Deployment Diagram

An arrangement graph is spoken to utilizing hub. A hub is a physical asset that executes code parts. They are likewise used to portray run time handling of hubs. The information is sent to the Hadoop where sqoop is utilized for information cleaning and the reports are produced utilizing hive.

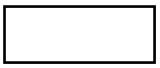
4.4 DATA FLOW DIAGRAMS

An information stream design (DFD) is a graphical portrayal of the "stream" of information through a data framework, demonstrating its strategy edges. A DFD is a significant part of the time utilized as a preparatory stroll to make an overview of the framework, which can later be cleared up. DFDs can in like way be utilized for the depiction of information prepare. A DFD indicates what sort of data will be sense of duty regarding and yield from the structure, where the information will begin from and go to, and where the information will be secured. It doesn't demonstrate data about the organizing of process or data about whether strategy will work in game-plan or in parallel.

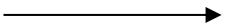
DFD Symbols:

In the DFD, there are four symbols

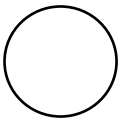
- A square defines a source or destination of system data.



- An arrow identifies data flow. It is the pipeline through which the information flows.



- A circle represents a process that transforms incoming data flow into outgoing data flow.



- An open rectangle is a data store, data at rest or a temporary repository of data.

4.4.1 LEVEL 0: SYSTEM INPUT/OUTPUT LEVEL

A level 0 DFD describes the system wide boundaries, dealing input to and output flow from the system and major processes.

Fig 4.4.1 Level 0 DFD

DFD Level 0 is in like way called a Context Diagram. It's a urgent review of the entire structure or process being bankrupt down or appeared. It's required to be an at first watch, demonstrating the framework as a particular surprising state handle, with its relationship to outside substances.

4.4.2 LEVEL 1: SUB SYSTEM LEVEL DATA FLOW

Level 1 DFD delineates the accompanying level of purposes of enthusiasm with the data stream between subsystems. The Level 1 DFD exhibits how the system is secluded into sub-structures (shapes), each of which oversees no less than one of the data streams to or from an outside pro, and which together give most of the helpfulness of the system as a rule.

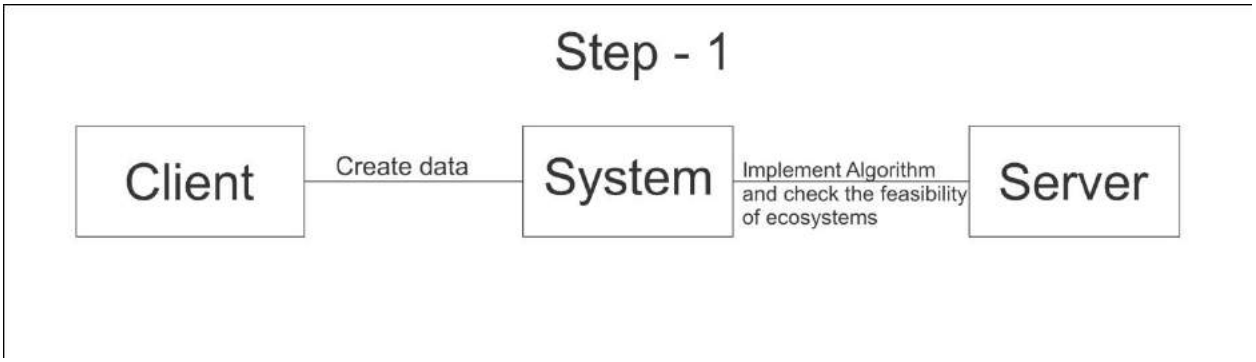


Fig 4.4.2 Level 1 DFD

4.4.3 LEVEL 2: FILE LEVEL DETAIL DATA FLOW

Plausibility and danger examination are connected here from various perspectives. The level 2 DFD elucidates the fundamental level of understanding about the system's working.

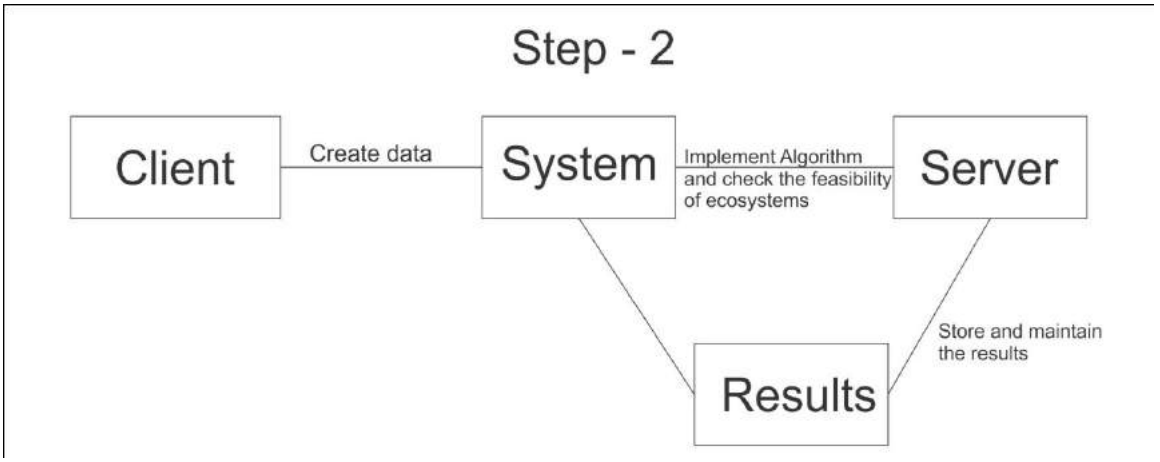


Fig 4.4.3 Level 2 DFD

5. IMPLEMENTATION

5. IMPLEMENTATION

5.1 SAMPLE CODE

```
import matplotlib.pyplot as plt
import seaborn as sns
import keras
from keras.models import Sequential
from keras.layers import Dense, Conv2D , MaxPool2D , Flatten , Dropout
from keras.preprocessing.image import ImageDataGenerator
from keras.optimizers import Adam
from sklearn.metrics import classification_report,confusion_matrix
import tensorflow as tf
import cv2
import os
import numpy as np
import pandas as pd

labels = ['nodules', 'none']
img_size = 224
def get_data(data_dir):
 data = []
 for label in labels:
  path = os.path.join(data_dir, label)
  class_num = labels.index(label)
  for img in os.listdir(path):
   try:
    img_arr = cv2.imread(os.path.join(path, img))[...,::-1] #convert BGR to RGB format
    resized_arr = cv2.resize(img_arr, (img_size, img_size)) # Reshaping images to preferred size
    data.append([resized_arr, class_num])
   except Exception as e:
    print(e)
 return np.array(data)
```

```
train = get_data('C:/Users/sravan/Desktop/2021/satya/input/train')
val = get_data('C:/Users/sravan/Desktop/2021/satya/input/test')

l = []
for i in train:
 if(i[1] == 0):
  l.append("nodules")
 else:
  l.append("none")

sns.set_style('darkgrid')
sns.countplot(l)
plt.show()

plt.figure(figsize = (5,5))
plt.imshow(train[1][0])
plt.title(labels[train[0][1]])
plt.show()

plt.figure(figsize = (5,5))
plt.imshow(train[-1][0])
plt.title(labels[train[-1][1]])
plt.show()



x_train = []
y_train = []
x_val = []
y_val = []

for feature, label in train:
 x_train.append(feature)
 y_train.append(label)

for feature, label in val:
```

```
 x_val.append(feature)
 y_val.append(label)

x_train = np.array(x_train) / 255
x_val = np.array(x_val) / 255

x_train.reshape(-1, img_size, img_size, 1)
y_train = np.array(y_train)

x_val.reshape(-1, img_size, img_size, 1)
y_val = np.array(y_val)


datagen = ImageDataGenerator(
        featurewise_center=False,  # set input mean to 0 over the dataset
        samplewise_center=False,  # set each sample mean to 0
        featurewise_std_normalization=False,  # divide inputs by std of the dataset
        samplewise_std_normalization=False,  # divide each input by its std
        zca_whitening=False,  # apply ZCA whitening
        rotation_range = 30,  # randomly rotate images in the range (degrees, 0 to 180)
        zoom_range = 0.2, # Randomly zoom image
        width_shift_range=0.1,  # randomly shift images horizontally (fraction of total width)
        height_shift_range=0.1,  # randomly shift images vertically (fraction of total height)
        horizontal_flip = True,  # randomly flip images
        vertical_flip=False)  # randomly flip images


datagen.fit(x_train)

model = Sequential()
model.add(Conv2D(32,3,padding="same", activation="relu", input_shape=(224,224,3)))
model.add(MaxPool2D())

model.add(Conv2D(32, 3, padding="same", activation="relu"))
model.add(MaxPool2D())
```

```
model.add(Conv2D(64, 3, padding="same", activation="relu"))
model.add(MaxPool2D())
model.add(Dropout(0.4))

model.add(Flatten())
model.add(Dense(128,activation="relu"))
model.add(Dense(2, activation="softmax"))

model.summary()



opt = Adam(lr=0.000001)
model.compile(optimizer = opt , loss =
tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True) , metrics = ['accuracy'])



history = model.fit(x_train,y_train,epochs = 500 , validation_data = (x_val, y_val))
#history = model.fit(x_train,y_train,epochs = 50 , validation_data = (x_val, y_val))
#model.fit(x_train,y_train,epochs = 30)




epochs_range = range(500)

plt.figure(figsize=(15, 15))
plt.subplot(2, 2, 1)
plt.plot(epochs_range, acc, label='Training Accuracy')
plt.plot(epochs_range, val_acc, label='Validation Accuracy')
plt.legend(loc='lower right')
plt.title('Training and Validation Accuracy')

plt.subplot(2, 2, 2)
plt.plot(epochs_range, loss, label='Training Loss')
plt.plot(epochs_range, val_loss, label='Validation Loss')
plt.legend(loc='upper right')
```

```
plt.title('Training and Validation Loss')
plt.show()

predictions = model.predict_classes(x_val)
predictions = predictions.reshape(1,-1)[0]
print(classification_report(y_val, predictions, target_names = ['Nodules (Class 0)','None (Class 1)']))

predictions = model.predict([x_val]) # Make prediction
predictions = model.predict_classes(x_val)

data = {'actual':y_val,'pred':predictions}
result = pd.DataFrame(data)
```

6. SCREENSHOTS

6.1 TUMOR IMAGE

```
plt.figure(figsize = (5,5))
plt.imshow(train[-1][0])
plt.title(labels[train[-1][1]])
plt.show()
```

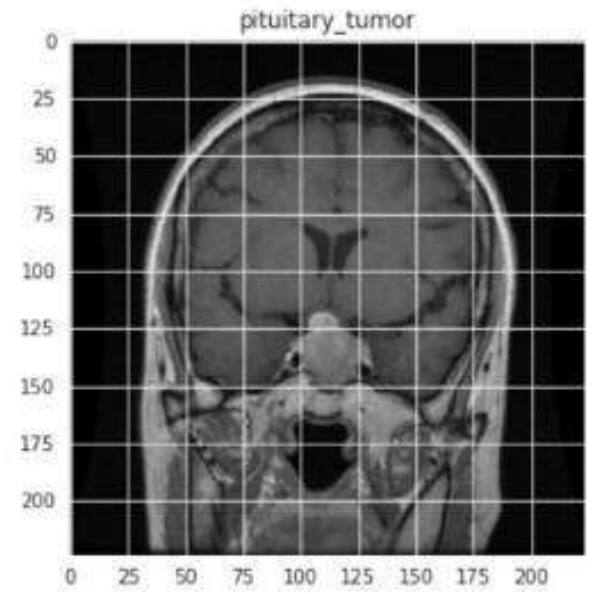


Figure 6.1 Tumor image

6.2 NO TUMOR IMAGE

```
[ ]  plt.figure(figsize = (5,5))
     plt.imshow(train[1][0])
     plt.title(labels[train[0][1]])
     plt.show()
```

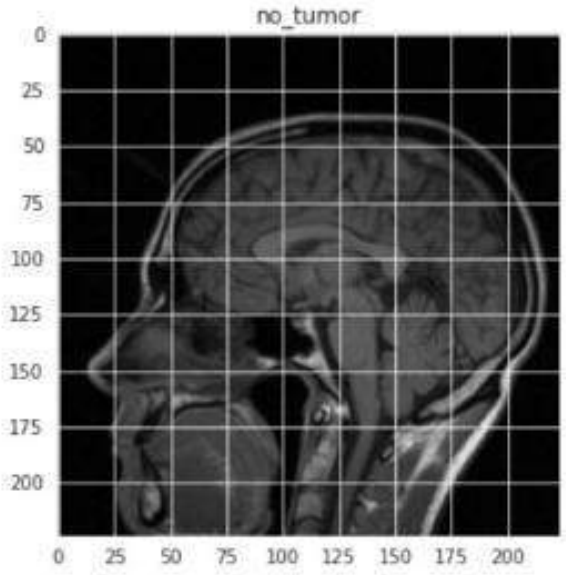


Figure 6.2 No Tumor Image

6.3 FINAL RESULT

```
[ ] result
```

	actual	pred
0	0	1
1	0	1
2	0	1
3	0	1
4	0	1
...	...	...
174	1	1
175	1	1
176	1	1
177	1	1
178	1	1

179 rows × 2 columns

Figure 6.3 Final Result

7. TESTING

7. TESTING

7.1 INTRODUCTION TO TESTING

Testing is a procedure, which uncovers blunders in the program. Programming testing is a basic component of programming quality affirmation and speaks to a definitive audit of determination, outline and coding. The expanding perceivability of programming as a framework component and chaperon costs related with a product disappointment are propelling variables for we arranged, through testing. Testing is the way toward executing a program with the plan of finding a mistake. The plan of tests for programming and other built items can be as trying as the underlying outline of the item itself It is the significant quality measure utilized amid programming improvement. Amid testing, the program is executed with an arrangement of experiments and the yield of the program for the experiments is assessed to decide whether the program is executing as it is relied upon to perform.

7.2 TESTING STRATEGIES

A technique for programming testing coordinates the outline of programming experiments into an all-around arranged arrangement of steps that outcome in fruitful improvement of the product. The procedure gives a guide that portrays the means to be taken, when, and how much exertion, time, and assets will be required. The procedure joins test arranging, experiment configuration, test execution, and test outcome gathering and assessment. The procedure gives direction to the specialist and an arrangement of points of reference for the chief. Due to time weights, advance must be quantifiable and issues must surface as ahead of schedule as would be prudent

7.2.1 UNIT TESTING

Unit testing focuses on the building blocks of the software system, that is, objects and sub system. There are Unit Testing is done on singular modules as they are finished and turned out to be executable. It is restricted just to the planner's prerequisites. It centers testing around the capacity or programming module. It concentrates on the interior preparing rationale and information structures. It is rearranged when a module is composed with high union

- Reduces the quantity of experiments
- Allows mistakes to be all the more effectively anticipated and revealed **Black Box 6.2.2**

7.2.2 TESTING

It is otherwise called Functional testing. A product testing strategy whereby the inward workings of the thing being tried are not known by the analyzer. For instance, in a discovery test on a product outline the analyzer just knows the information sources and what the normal results ought to be and not how the program touches base at those yields. The analyzer does not ever inspect the programming code and does not require any further learning of the program other than its determinations. In this system some experiments are produced as information conditions that completely execute every single practical prerequisite for the program. This testing has been utilizations to discover mistakes in the accompanying classifications:

- Incorrect or missing capacities
- Interface blunders
- Errors in information structure or outside database get to
- Performance blunders
- Initialization and end blunders.

In this testing just the yield is checked for rightness.

7.2.3 WHITE BOX TESTING

It is otherwise called Glass box, Structural, Clear box and Open box testing . A product testing procedure whereby express learning of the inner workings of the thing being tried are utilized to choose the test information. Not at all like discovery testing, white box testing utilizes particular learning of programming code to inspect yields. The test is precise just if the analyzer comprehends what the program should do. He or she would then be able to check whether the program veers from its expected objective. White box testing does not represent blunders caused by oversight, and all obvious code should likewise be discernable. For an entire programming examination, both white box and discovery tests are required.

In this the experiments are produced on the rationale of every module by drawing stream diagrams of that module and sensible choices are tried on every one of the cases. It has been utilizations to produce the experiments in the accompanying cases:

- Guarantee that every single freeway have been Executed.
- Execute every single intelligent choice on their actual and false Sides.

7.2.4 INTEGRATION TESTING

Coordination testing guarantees that product and subsystems cooperate an entirety. It tests the interface of the considerable number of modules to ensure that the modules carry on legitimately when coordinated together. It is characterized as a deliberate procedure for developing the product engineering. In the meantime reconciliation is happening, lead tests to reveal blunders related with interfaces. Its Objective is to take unit tried modules and assemble a program structure in view of the recommended outline

Two Approaches of Integration Testing

- Non-incremental Integration Testing
- Incremental Integration Testing

7.2.5 SYSTEM TESTING

Framework testing includes in-house testing of the whole framework before conveyance to the client. Its point is to fulfill the client the framework meets all necessities of the customer's determinations. This testing assesses working of framework from client perspective, with the assistance of particular report. It doesn't require any inward learning of framework like plan or structure of code.

It contains utilitarian and non-useful zones of utilization/item. Framework Testing is known as a super arrangement of a wide range of testing as all the significant sorts of testing are shrouded in it. In spite of the fact that attention on sorts of testing may differ on the premise of item, association procedures, course of events and necessities. Framework Testing is the start of genuine testing where you test an item all in all and not a module/highlight.

7.2.6 ACCEPTANCE TESTING

Acknowledgment testing, a testing method performed to decide if the product framework has met the prerequisite particulars. The principle motivation behind this test is to assess the framework's consistence with the business necessities and check in the event that it is has met the required criteria for conveyance to end clients. It is a pre-conveyance testing in which whole framework is tried at customer's site on genuine information to discover blunders. The acknowledgment test bodies of

evidence are executed against the test information or utilizing an acknowledgment test content and afterward the outcomes are contrasted and the normal ones.

The acknowledgment test exercises are completed in stages. Right off the bat, the essential tests are executed, and if the test outcomes are palatable then the execution of more intricate situations are done.

7.2.7 BLACK BOX TESTING

The method is used when knowledge of the specified function that a product has been designed to perform is known. The concept of black box is used to represent a system whose inside workings are not available to inspection. In a black box the test item is a "Black" , since its logic is unknown , all that is known is what goes in and what comes out or the input and output.

7.3 TEST APPROACH

A Test approach is the test system usage of a venture, characterizes how testing would be done. The decision of test methodologies or test technique is a standout amongst the most intense factor in the achievement of the test exertion and the precision of the test designs and gauges.
Testing should be possible in two ways

- Bottom up approach
- Top down approach

7.3.1 BOTTOM UP APPROACH

Testing can be performed beginning from littlest and most reduced level modules and continuing each one in turn. In this approach testing is directed from sub module to primary module, if the fundamental module is not built up a transitory program called DRIVERS is utilized to recreate the principle module. At the point when base level modules are tried consideration swings to those on the following level that utilization the lower level ones they are tried exclusively and afterward connected with the already inspected bring down level modules

7.3.2 TOP DOWN APPROACH

In this approach testing is directed from fundamental module to sub module. in the event that the sub module is not built up an impermanent program called STUB is utilized for mimic the sub module. This sort of testing begins from upper level modules. Since the nitty gritty exercises more often than not performed in the lower level schedules are not given stubs are composed. A stub is a module shell called by upper level module and that when achieved legitimately will restore a message to the calling module demonstrating that appropriate association happened.

7.4 VALIDATION

The way toward assessing programming amid the improvement procedure or toward the finish of the advancement procedure to decide if it fulfills determined business prerequisites. Approval Testing guarantees that the item really addresses the customer's issues. It can likewise be characterized as to exhibit that the item satisfies it’s proposed utilize when sent on proper condition.

The framework has been tried and actualized effectively and along these lines guaranteed that every one of the Prerequisites as recorded in the product necessities determination are totally satisfied.

7.5 TEST CASES

Experiments include an arrangement of steps, conditions and sources of info that can be utilized while performing testing undertakings. The principle expectation of this action is to guarantee whether a product passes or bombs as far as usefulness and different perspectives. The way toward creating experiments can likewise help discover issues in the prerequisites or plan of an application. Experiment goes about as the beginning stage for the test execution, and in the wake of applying an arrangement of information esteems, the application has a conclusive result and leaves the framework at some end point or otherwise called execution post condition.

8. CONCLUSION

8. CONCLUSION

8.1 PROJECT CONCLUSION

In this paper, a new method based on the combination of feature extraction algorithm and the CNN for tumor detection from brain images is presented. The CNN is capable of detecting a tumor. The CNN is very useful for selecting an auto-feature in medical images. Images collected at the centers were labeled by clinicians, then, tumor screenings were categorized into two normal and patient classes. A total of 1666 images were selected as train data and 226 images were taken as a test data. The proportion of image categorization in two classes was proportional from the ratio of patients to healthy subjects. Images were applied to the CNN after preprocessing. In order to evaluate the performance of the CNN, has been used by other classifiers such as the RBF classifier and the decision tree classifier in the CNN architecture.

9. BIBLIOGRAPHY

9. BIBLIOGRAPHY

9.1 REFERENCES

[1] M. Karuna and A. Joshi, Automatic detection and severity analysis of brain tumors using gui in matlab, International Journal of Research in Engineering and Technology, 10, pp. 586-594, 2013.

[2] KS. Aboody, A. Brown, et al, Neural stem cells display extensive tropism for pathology in adult brain Evidence from intracranial gliomas, Proceedings of the National Academy of Sciences, 97 (23), pp. 1284612851, 2000.

[3] A. joshi, D. H. Shah, et al, Survey of brain tumor detection techniques through MRI images, International Journal of Research in Engineering and Technology, 10, pp. 586-594, 2013.

[4] JP. Poonam, Review of image processing techniques for automatic detection of tumor in human brain, International Journal of Computer Science and Mobile Computing, 2(11), pp. 117-122, 2013.

[5] H. Cecotti and A. Graeser, Convolutional neural network with embedded Fourier transform for EEG classification, Pattern Recognition,19th International Conference on. IEEE, ICPR, pp. 1-14, 2008.

[6] R. Bayot and T. Gonalves, A survey on object classification using convolutional neural networks, 2015.

[7] M. Soltaninejad, et al, Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI, International journal of computer assisted radiology and surgery, 12(2), pp. 183-203, 2017.

[8] S. Pereira, et al, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE transactions on medical imaging, 35(5), pp. 1240-1251, 2016.

[9] Halimeh Siar, Mohammad Teshnehlab, Diagnosing and Classification Tumors and MS Simultaneous of Magnetic Resonance Images Using Convolution Neural Network, 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), 2019.

[10] https://github.com/CMRTCTeam12/Recognizing-human-brain-tumors/tree/main

The International Journal of Analytical and Experimental Modal analysis

An UGC-CARE Approved Group - II Journal

An ISO : 7021 - 2008 Certified Journal

ISSN NO: 0886-9367 / web : http://ijaema.com / e-mail: submitijaema@gmail.com

Certificate of Publication

This is to certify that the paper entitled

"RECOGNIZING HUMAN BRAIN TUMORS"

Authored by :

Kongala Praveen

From

CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

Has been published in

IJAEMA JOURNAL, VOLUME XIII, ISSUE VI, JUNE- 2021

Michal A. Olszewski Editor-In-Chief
IJAEMA JOURNAL

6.3 IMPACT FACTOR

ISO International Organization for Standardization 7021-2008

http://ijaema.com/

The International Journal of Analytical and Experimental Modal analysis

An UGC-CARE Approved Group - II Journal

An ISO : 7021 - 2008 Certified Journal

ISSN NO: 0886-9367 / web : http://ijaema.com / e-mail: submitijaema@gmail.com

Certificate of Publication

This is to certify that the paper entitled

"RECOGNIZING HUMAN BRAIN TUMORS"

Authored by :

Gayathri Jadhav

From

CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

Has been published in

IJAEMA JOURNAL, VOLUME XIII, ISSUE VI, JUNE- 2021

Michal A. Olszewski Editor-In-Chief
IJAEMA JOURNAL

6.3 IMPACT FACTOR

ISO International Organization for Standardization 7021-2008

http://ijaema.com/

The International Journal of Analytical and Experimental Modal analysis

An UGC-CARE Approved Group - II Journal

An ISO : 7021 - 2008 Certified Journal

ISSN NO: 0886-9367 / web : http://ijaema.com / e-mail: submitijaema@gmail.com

Certificate of Publication

This is to certify that the paper entitled

"RECOGNIZING HUMAN BRAIN TUMORS"

Authored by :

Arupula Umarani

From

CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

Has been published in

IJAEMA JOURNAL, VOLUME XIII, ISSUE VI, JUNE- 2021

Michal A. Olszewski Editor-In-Chief

IJAEMA JOURNAL

6.3 IMPACT FACTOR

ISO International Organization for Standardization 7021-2008

http://ijaema.com/

The International Journal of Analytical and Experimental Modal analysis

An UGC-CARE Approved Group - II Journal

An ISO : 7021 - 2008 Certified Journal

ISSN NO: 0886-9367 / web : http://ijaema.com / e-mail: submitijaema@gmail.com

Certificate of Publication

This is to certify that the paper entitled

"RECOGNIZING HUMAN BRAIN TUMORS"

Authored by :

K Ranjith Reddy, Associate Professor

From

CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

Has been published in

IJAEMA JOURNAL, VOLUME XIII, ISSUE VI, JUNE- 2021

Michal A. Olszewski Editor-In-Chief

IJAEMA JOURNAL

6.3 IMPACT FACTOR

ISO International Organization for Standardization 7021-2008

http://ijaema.com/

RECOGNIZING HUMAN BRAIN TUMORS

Gayathri Jadhav
UG Scholar, Computer Engineering Department, CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

Kongala Praveen
UG Scholar, Computer Engineering Department, CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

Arupula Umarani
UG Scholar, Computer Engineering Department, CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

K Ranjith Reddy
Associate Professor, CMR Technical Campus, UGC Autonomous, Kandlakoya (V), Medchal Road, Hyderabad-501401, INDIA

Abstract:

Deep Learning is the newest and the current trend of the machine learning field that paid a lot of the researchers' attention in the recent few years. As a proven powerful machine learning tool, deep learning was widely used in several applications for solving various complex problems that require extremely high accuracy and sensitivity, particularly in the medical field. In general, the brain tumor is one of the most common and aggressive malignant tumor diseases which is leading to a very short expected life if it is diagnosed at a higher grade. Based on that, brain tumor grading is a very critical step after detecting the tumor in order to achieve an effective treating plan. In this paper, we used Convolutional Neural Network (CNN) which is one of the most widely used deep learning architectures for classifying a dataset of 3064 T1 weighted contrast-enhanced brain MR images for grading (classifying) the brain tumors into three classes (Glioma, Meningioma, and Pituitary Tumor). The proposed CNN classifier is a powerful tool and its overall performance with an accuracy of 98.93%.

Keywords*—Deep Learning, Convolutional Neural Network (CNN), Mchine Learning, Brain Tumors.*

I Introduction

Brain tumors can be classified into two types: benign (noncancerous) and malignant (cancerous). The malignant tumors can quickly spread to other tissues in the brain and lead to worsening the patient's condition [1]. When most of the cells are old or damaged, they are destroyed and replaced by new cells. If damaged and old cells are not eliminated with generating the new cells, it can cause problems. The production of additional cells often results in the formation of a mass of tissue, which refers to the growth or tumor. Brain tumor detection is very complicated and difficult due to the size, shape, location and type of tumor in the brain. Diagnosis of brain tumors in the early stages of the tumor's start is difficult because it cannot

accurately measure the size and resolution of the tumor [2]. However, if the tumor is diagnosed and treated early in the tumor formation process, the chance of patient's treatment is very high. Therefore, the treatment of tumor depends on the timely diagnosis of the tumor [3]. The diagnosis is usually done by a medical examination, with computer tomography or magnetic imaging. MRI imaging is a method that provides accurate images of the brain and is one of the most common and important methods for diagnosing and evaluating the patient's brain. In the field of Medical Detection Systems (MDS), MRI images provide better results than other imaging techniques such as Computed Tomography (CT), due to their higher contrast in soft tissue in humans [4]. The proposed technique has used CNN to identify and categorize the tumor from brain images of the brain. The main difference between the main channels of the neural network with the normal neural network is that it is able to automatically and locally extract the feature from each image [5]. These types of networks consist of neurons with weights and biases that can be learned [6]. Due to the results of CNN on the dataset, in order to improve the proposed method. Machine learning algorithm is used to feature extraction. The algorithm used was the clustering algorithm applied on data set, and then the images are applied to the CNN. The results showed that the proposed method has been successful. The purpose of extracting the property before applying to the CNN is that in some images fatty masses are considered as tumors, or in some images the tumor is mistakenly considered to be fat and should have increased medical error. Extracting the attribute initially and before applying the CNN leads to improved network accuracy and increased accuracy. The advancement in medical technologies helps the clinical experts to facilitate more efficient e-health care systems to the patients. There is a number of medical domains where ehealth care systems are beneficial [1]. Computer vision-based applications of biomedical imaging are more importance as they provide recognition information to the radiologist for batter treatment-related problems. Different medical imaging techniques and methods that include X-ray, Magnetic Resonance. tumor segmentation, classification, and prediction methods. Segmentation methods classify the image patch into different classes, such as necrosis, healthy tissues, edema, enhancing core and non-enhancing core. Different tumor cells show distinct phenotypic and morphological information for

segmentation, classification, and prediction, including gene expression, motility, cellular morphology, metabolism metastatic potential, and proliferation. This paper presents a review of various methods, techniques, frameworks, architectures, algorithms and critical studies using deep learning for segmentation, classification, and survival time prediction. Survey taxonomy describes the methods, techniques, systems, algorithms, frameworks, and architectures that are based on tumor segmentation, evaluation, and features exploration for tumor prediction and its classification. The review performs an analysis of the features extraction techniques, dataset utilized, tools, languages, and libraries that are used for implementation, recognition and evaluation measures. The issues and research gaps in various existing research problems include the key issues in tumor recognition for monitoring, recognition procedures and treatment plans for cancer patients. The application of deep learning to brain tumor analysis first appears in conferences and workshops, and then in journals. The number of research papers grew rapidly from 2015 to onward. This topic has now became dominant at different conferences and journals. Figure1 illustrates the development of deep learning applications to brain tumor analysis. Figure2 presents a literature-based taxonomy for brain tumor analysis The development of deep learning application to brain tumor analysis motivated us to present a comprehensive review in all fields of brain tumor that includes segmentation, prediction, classification, both from methodology-driven and applications perspective. The review also includes a never view of all there search publications in tabular form that helps readers to quickly assess the field. Consequently, this review presents a dedicated discussion section to the readers that covers the state-of-the-art successful development, open research challenges and an overview of future research directions.

II. LITERATURE SURVEY

Recently, Machine learning (ML) and Deep Learning (DL) methods are widely been used for detection and grading brain tumors using different imaging modalities, especially those acquired using MRI. In this section, the most recent and related research works on the paper topic are presented. Mohsen, Heba, et all. [8] Propose a system that combines discrete wavelet transform (DWT) features and deep learning (DL) techniques. They have used fuzzy

c-mean method for segmenting the brain tumor, and for each detected lesion the DWT was applied to extract the features, where these features are fed into the principal component analysis (PCA) for feature dimension reduction and finally the selected features are then fed to deep neural networks (DNN). The results show that they achieve an accuracy rate of 96.97% and a sensitivity of 97.0 %. Widhiarso, Wijang, Yohannes Yohannes, and Cendy Prakarsah [10] presented a brain tumor classification system using a convolutional neural network (CNN) and Gray Level Co-occurrence Matrix (GLCM) based features. They extracted four features (Energy, Correlation, Contrast, and Homogeneity) from four angles (0°, 45°, 90°, and 135°) for each image and then these features are fed into CNN, they tested their methodology on four different datasets (Mg-Gl, Mg-Pt, GlPt, and Mg-Gl-Pt) and the best accuracy achieved was82.27% for Gl-Pt dataset using two sets of features; contrast with homogeneity and contrast with correlation. Seetha, J., and S. S. Raja [12] proposed a deep CNN based system for automated brain tumor detection and grading. The system is based on Fuzzy C-Means (FCM) for brain segmentation and based on these segmented regions a texture and shape features were extracted then these features were fed into SVM and DNN classifiers. The results showed that the system achieved a rate of 97.5% accuracy. On the other hand, Cheng, Jun, et al. [8] enhanced the performance of the brain tumor classification process using region of interest (ROI) augmentation and fine ring-form partition. They applied these enhancements to different feature extractions methods which are intensity histogram, GLCM, and the bag-of-words (BoW) where these features vectors are fed into a classifier. The experimental results showed that the accuracy enhanced from 71.39% to 78.18%, and 83.54% to 87.54%, and 89.72% to 91.28% for intensity histogram, GLCM, and BoW respectively. Sasikala, M., and N. Kumaravel. [7] Proposed a genetic algorithm feature selection for feature dimension reduction of wavelet features set. The method is based on selecting optimal features vector that can be fed into the selected classifier such as an artificial neural network (ANN). The results show that the genetic algorithm selected only 4 of 29 features and achieved an accuracy of 98% using only the selected features. Khawaldeh, Saed, et al. [3] proposed a system for non-invasive grading of glioma brain tumors using a modified version of AlexNet CNN. The classification process

was done using whole-brain MRI images and the labels of the images were at the image level, not the pixel level. The experimental results showed that the method achieved a reasonable performance with an accuracy of 91.16%. Sajjad, Muhammad et al. [4] proposed an extensive data augmentation method fused with CNN for brain tumor classification. The method used for multi-grade classification of brain tumors using segmented brain tumor MRI images. They used pre trained VGG19 CNN architecture for classification using transferee learning and achieved an overall accuracy of values 87.38% and 90.67% for data before and after augmentation respectively. While Özyurt, Fatih et al. [5] combine the CNN with neutrosophic expert maximum fuzzy (NS-CNN) sure entropy for brain tumor classification. They used the neutrosophic set – expert maximum fuzzy-sure method for brain tumor segmentation then these images are fed to CNN to extract features and fed them to SVM classifiers to be classified as benign or malignant. They achieved an average success of 95.62%. In [7], an automated method is used to identify and categorize MRI images. The CNN was able to accurately classify 98.67% images correctly into three classes. In [10], a multi-stage Fuzzy C-Means (FCM) framework was proposed to segment brain tumors from MRI images. In [11], an efficient and effective method which uses CNNs used for classification and segmentation. The proposed method, used Image-Net for extract features. The results obtained 97.5% accuracy for classification and 84% accuracy for segmentation. In [12], multiphase MRI images in tumor grading have been studied and a comparison has been made between the results of deep learning structures and base neural networks. The results show that the network performance based on the sensitivity and specificity of CNN improved by 18% compared to the neural networks. In [13], a deep learning-based supervised method is introduced to detect synthetic aperture radar (SAR) image changes. This method provided a dataset with an appropriate data volume and diversity for training the DBN using input images and the images obtained from applying the morphological operators on them. The detection performance of this method indicates the appropriability of deep learning based algorithms for solving the change detection problems. In paper [14], a completely automated brain tumor classification method is proposed based on DNN. The proposed networks have been designed to be used in low-grade and high glioblastoma disease images. In

this paper, a new architecture of CNN is presented. The proposed a cascading architecture is proposed in which the output of a core CNN is used as an additional source of information for the next CNN.

III. PROPOSED METHODOLOGY

Many attempts have been by researchers in the past for analyzing human barin tumor with the available methods but the methods has not produced enough results for the tumor identification. The system was not able to measure the tumor availability Working with image is a big bottle neck an finding the resources from the dataset is more difficult. When the methods were applied on the image scans of the brain, the results are not even satisfying. The deep learning concept to perform an automated brain tumors classification using brain MRI images and measure its performance. The proposed methodology aims to differentiate between normal brain and some types of brain tumors such as glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumors using brain MRI images. Our proposed methodology based on the DNN learning architecture for classification where the classifier is identifying the brain tumors in brain MRIs. The proposed methodology uses a set of features extracted by the discrete wavelet transform (DWT) feature extraction technique from the segmented brain MRI images, to train the DNN classifier for brain tumors classification. Makes the prediction of the disease in earlier way this helps us to save the patient's life. Image analysis will help us to understand more about the brain and the diseases causes.

Dataset The data set images used in this paper include brain MRI images of 153 patients, including normal and brain tumors patients who referred to imaging centres because of headaches After examination and diagnosis of the doctor, the collected images included brain images of 80 healthy patients Include 1321 images which has 56 images for testing data and 515 images for the train data.73 patient tumors Include 571 images which has 170 images for test data and 1151 images for the train data. Of the total number of patients with brain tumor. In machine learning and image processing, feature are created from the initial dataset. Which facilitates the learning process. When the input data of an algorithm is too large, it can be converted to a smaller set of features. The process of extracting a subset from the primary features set is called feature extraction [15]. The selected features

include information about the input data, so that the reduced representation of the agent can be done instead of the full initial data. One of the important application of feature extraction is in the image processing, which are used to distinguish the desired segments or the shape (features) of a digital image or video stream. Disease, 86 were women and 68 were men, whose age range from 8 to 66 years old. Of a total of 153 patients, 1892 images were collected, 1666 images for train data and 226 images for test images. The collected images originally had an initial size of 224×224. In few cases, some areas of fat in the pictures are mistakenly detected as tumor, or the tumors may not be seen by the physician; the most exact diagnosis is completely depended on the physician's skill. In this paper, the CNN has been used for tumor detection through brain images. There were additional margins of the images gathered from the imaging centers. These margins were cropped to prevent the noise of the images. One of the main reasons for using the feature extraction technique and combining it with the CNN is to retrieve the feature extraction of the images in order to increase the accuracy of the network. According to the results of the CNN on the initial images, in order to improve the network accuract, in this study, a new method which is a combination of Clustering algorithm for feature extraction and CNN is proposed. Convolutional neural method initially, the images were applied to the CNN without any feature extraction methods. The size of the input images is initially 227×227. The Alexnet architect was used to identify and classify the images, which consisted of 5 Convolutional layers and 3 layers of Subsampling layers, Normalization layers, Normalization layers, Fully Connected layers and lastly layer the classification layer.

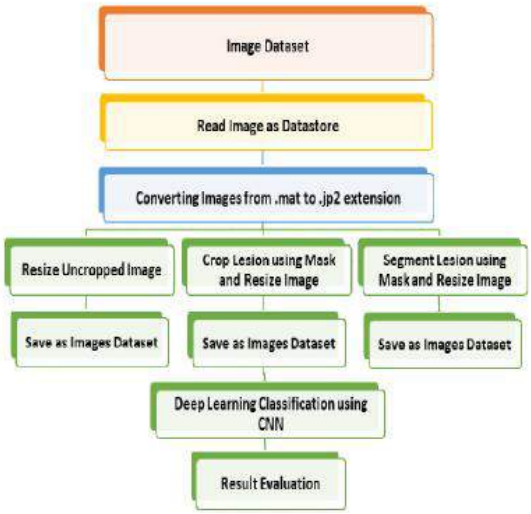


Fig 1: Project Architecture

IV RESULT ANALYSIS

After preprocessing and feature extraction of our dataset, 80% of the dataset was selected for training and 20% of the dataset was selected for testing.

For error calculation, we are using scikit-learn metrics.

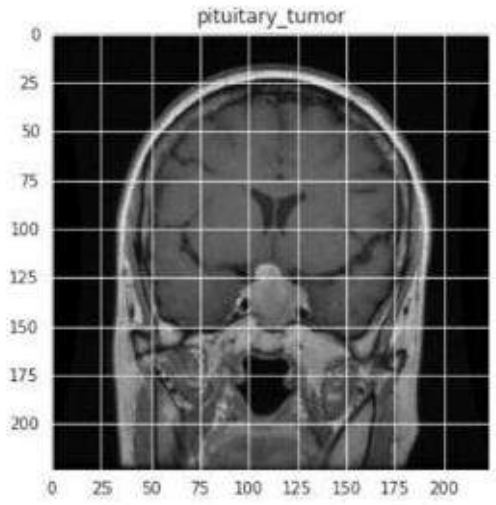


Fig 2: Tumor images

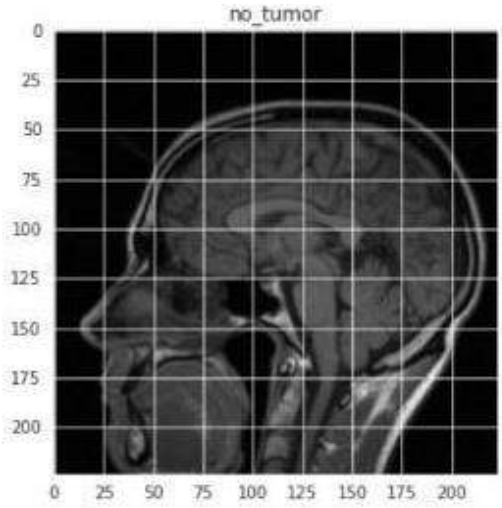


Fig 3: No Tumor Image

	actual	pred
0	0	1
1	0	1
2	0	1
3	0	1
4	0	1
...	...	...
174	1	1
175	1	1
176	1	1
177	1	1
178	1	1

179 rows × 2 columns

Fig 4 : Final Result

V CONCLUSION

In this paper, a new method based on the combination of feature extraction algorithm and the CNN for tumor detection from brain images is presented. The CNN is capable of detecting a tumor. The CNN is very useful for selecting an auto-feature in medical images. Images collected at the centers were labeled by clinicians, then, tumor screenings were categorized into two normal and patient classes. A total of 1666 images were selected as train data and 226 images were taken as a test data. The proportion of image categorization in two classes was proportional from the ratio of patients to healthy subjects. Images were applied to the CNN after preprocessing. In order to evaluate the performance of

the CNN, has been used by other classifiers such as the RBF classifier and the decision tree classifier in the CNN architecture..

VI REFERENCES

[1] M. Karuna and A. Joshi, Automatic detection and severity analysis of brain tumors using gui in matlab, International Journal of Research in Engineering and Technology, 10, pp. 586-594, 2013.

[2] KS. Aboody, A. Brown, et al, Neural stem cells display extensive tropism for pathology in adult brain Evidence from intracranial gliomas, Proceedings of the National Academy of Sciences, 97 (23), pp. 1284612851, 2000.

[3] A. joshi, D. H. Shah, et al, Survey of brain tumor detection techniques through MRI images, International Journal of Research in Engineering and Technology, 10, pp. 586-594, 2013.

[4] JP. Poonam, Review of image processing techniques for automatic detection of tumor in human brain, International Journal of Computer Science and Mobile Computing, 2(11), pp. 117-122, 2013.

[5] H. Cecotti and A. Graeser, Convolutional neural network with embedded Fourier transform for EEG classification, Pattern Recognition,19th International Conference on. IEEE, ICPR, pp. 1-14, 2008.

[6] R. Bayot and T. Gonalves, A survey on object classification using convolutional neural networks, 2015.

[7] M. Soltaninejad, et al, Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI, International journal of computer assisted radiology and surgery, 12(2), pp. 183-203, 2017.

[8] S. Pereira, et al, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE transactions on medical imaging, 35(5), pp. 1240-1251, 2016.

[9] Halimeh Siar, Mohammad Teshnehlab, Diagnosing and Classification Tumors and MS Simultaneous of Magnetic Resonance Images Using Convolution Neural Network, 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), 2019.

[10] L. Szilagyi, et al, Automatic brain tumor segmentation in multispectral MRI volumes using a fuzzy c-means cascade algorithm, In 2015 12th international conference on fuzzy systems and knowledge discovery (FSKD), IEEE, pp. 285-291, 2015.

[11] Y. Xu, et al, Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation, In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 947-951, 2015.

[12] Prasadu Peddi (2019), "AN EFFICIENT ANALYSIS OF STOCKS DATA USING MapReduce", ISSN: 1320-0682, Vol 6, issue 1, pp:22-34

[13] F. Samadi, G. Akbarizadeh, et al, Change Detection in SAR Images using Deep Belief Network: a New Training Approach based on Morphological Images, IET Image Processing, 2019.

[14] M. Havaei, et al, Brain tumor segmentation with deep neural networks, Medical image analysis, 35, 18-31, 2017.

[15] Prasadu Peddi (2017) "Design of Simulators for Job Group Resource Allocation Scheduling In Grid and Cloud Computing Environments", ISSN: 2319- 8753 volume 6 issue 8 pp: 17805-17811.

[16] W. Zhang, K. Itoh, et al, Parallel Distributed Processing Model with local Space-Invariant Interconnections and its Optical Architecture, Applied optics, 29(32), pp. 4790-4797, 1990.

[17] LY. LeCun, K. Kavukcuoglu, et al, Convolutional networks and applications in vision, In ISCAS, pp. 253-256, 2010.

[18] Prasadu Peddi (2020), "Public auditing mechanism to verify data integrity in cloud storage", International Journal of Emerging Trends in Engineering Research, vol 8, Issue 9, Pages 5220–5225.

[19] D. Ciresan, U. Meier, et al, Multi-column deep neural networks for image classification. Int.Conf. Computer Vision and Pattern Recognition, IEEE, pp. 1202.2745, 2012..