T.C.

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Department

BRAIN TUMOR DETECTION AND CLASSIFICATION USING IMAGE PROCESSING TECHNIQUES

Sultan Bahr Fayyadh

Master of Science

Supervisor

Asst. Prof. Abdullahi Abdu IBRAHIM

Istanbul, 2021

BRAIN TUMOR DETECTION AND CLASSIFICATION USING IMAGE PROCESSING TECHNIQUES

by

Sultan Bahr Fayyadh

Information Technologies

Submitted to the Institute of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2021

The thesis titled “BRAIN TUMOR DETECTION AND CLASSIFICATION USING IMAGE PROCESSING TECHNIQUES” prepared and presented by “Sultan Bahr Fayyadh” was accepted as a Master of Science Thesis in Information Technologies.

____

Asst. Prof. Abdullahi Abdu IBRAHIM

Supervisor

Thesis Defense Jury Members:

Asst. Prof. Abdullahi Abdu IBRAHIM	School of Engineering and Natural Sciences, Altinbas University	__________________
Asst. Prof. Hakan KOYUNCU	School of Engineering and Natural Sciences, Altinbas University	__________________
Assoc. Prof. Mohammad ABUBAKAR	School of Business and Social Sciences, Antalya Bilim University	__________________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Approval Date of Institute of Graduate Studies: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Sultan Bahr Fayyadh

Signature

ACKNOWLEDGEMENTS

First of all, praise is to Allah, the lord of the entire creation, on all the blessing was the aid in achieving this work to its end. I would like to express my appreciation to my supervisor, Dr. Abdullahi Abdu Ibrahim for his faithful guidance, valuable instructions, and constructive comments which have made the completion of this work possible.

Also, I would like to express my gratitude and thanks to all the teaching staff who have taught me. Special thanks are extended to all my friends for their help. Special thanks to the members of the evaluation committee for discussing my thesis. My final words go to my family. I want to thank my parents, my brothers, and my sisters for their love and guidance which helped me in achieving my goals.

Sultan Bahr Fayyadh

ABSTRACT

BRAIN TUMOR DETECTION AND CLASSIFICATION USING IMAGE PROCESSING TECHNIQUES

FAYYADH, Sultan Bahr Fayyadh,

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Asst. Prof. Dr. Abdullahi Abdu Ibrahim

Co-Supervisor:

Date: 04/2021

Pages: 90

[Brain Tumor is the abnormal and uncontrolled growth of tissues or cells in the brain, the brain tumor is dangerous and life-threatening disease, Detection of the diseases through image processing is done by using an integrated approach working methods of processing MRI images of brain tumor entering it and distinguishes this approach if the brain is normal or abnormal, computer systems have been used in this area to analyze medical information, analyzing and extracting the most important features of the brain tumor and focusing on image analysis and processing techniques to distinguish between different diseases based on the symptoms of each disease.

This work adopts two proposed approaches for detecting brain tumor using image processing and deep learning techniques with makes a comparison between these two approaches. This work was planned to some an important and common group of brain tumor, including Glioma, Meningioma, Pituitary Adenoma, and Nerve Sheath. These kinds of brain tumors are the most popular in the world. The dataset contains 3000 images related to malignant (normal) and benign (abnormal) each one has 1500 image. In the first proposed approach, where several steps are used in the form of stages, which are include, the image acquisition stage, image pre-processing, image segmentation, image post-processing, extraction the features, and the classification stage.

Class support vector machine (SVM) algorithm was used to perform the classification process in the second proposed approach, the convolution neural network (CNN) was used through which the brain tumor are classified according to a special structure of this algorithm consisting of several layers. In these two proposed approaches, the tumor were classified to deferent classes was detected.

The obtained results from the comparison between the two proposed approaches in terms of performance and accuracy showed the preference of the second approach which adopted the deep learning and using the CNN algorithm, over the first approach, because the overall accuracy rate that obtained from the second proposed approach was (98,29%). While the overall accuracy rate that obtained from the first proposed approach was (68.9%). So, the second proposed approach is more accurate and powerful in the process of detecting and classifying brain tumor.]

Keywords: Image Processing, Segmentation, Classification, Medical Image, Deep learning, CNN Algorithm

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

AI	:	Artificial Intelligence
WHO	:	World Health Organization
MRI	:	Magnetic Resonance Image
CNN	:	Convolutional Neural Network
SVM	:	Support Vector Machine
ANN	:	Artificial Neural Network
GLCM	:	Gray Level Co-occurrence Matrix
CAD	:	Computer Aided System
ML	:	Machine Learning

1. INTRODUCTION

1.1 OVERVIEW

Cancer is one of the main causes of death in Iraq, as it is the second cause of death after vascular and heart disease, at an average rate (11%) of all deaths in Iraq, and this proportion is expected to increase in Iraq, and it is worth noting that a large percentage of deaths With this disease, it occurs early and young age stage of life [1]

Early diagnosis and treatment of brain cancer have been found to be the most active ways of minimizing death-rate [2].

Rapid advances in digital recognition and computational intelligence techniques have significantly improved the understanding of medical images and helped to early detection.

Over the last three decades, the use of computers in medical applications has increased. Computers were generally used in this field to read, scan, store, process, retrieve, and analyze medical information. They were also used as an instrument in patient diagnosis and care.

More recently, modern electronic imaging techniques were used to enhance image quality and images can be analyzed to demonstrate points of contention or to extracting useful clinical characteristics which can provide reliable information to validate the human judgment process.

An expert system is a computer program that has the knowledge of an expert system built into a series of rules. The expert system then finds the solution to a problem, such as the diagnosis of a medical disease.

For this reason, the fields of analysis, detection and treatment of cancer diseases have become an important field of research, the use of computers and their techniques to analyze and detect this disease, and medical imaging techniques are one of the most important techniques that help in this field as shown in Figure (1.1), It would already have life-saving effects on our ability to recognize cancer at an early stage and to diagnose the disease more accurately. However, in order to help enhance the efficiency and accuracy of diagnosis and treatment, image processing technology is widely used to analyze and recognize tumors, evaluate diagnostic accuracy and predict cancer development.

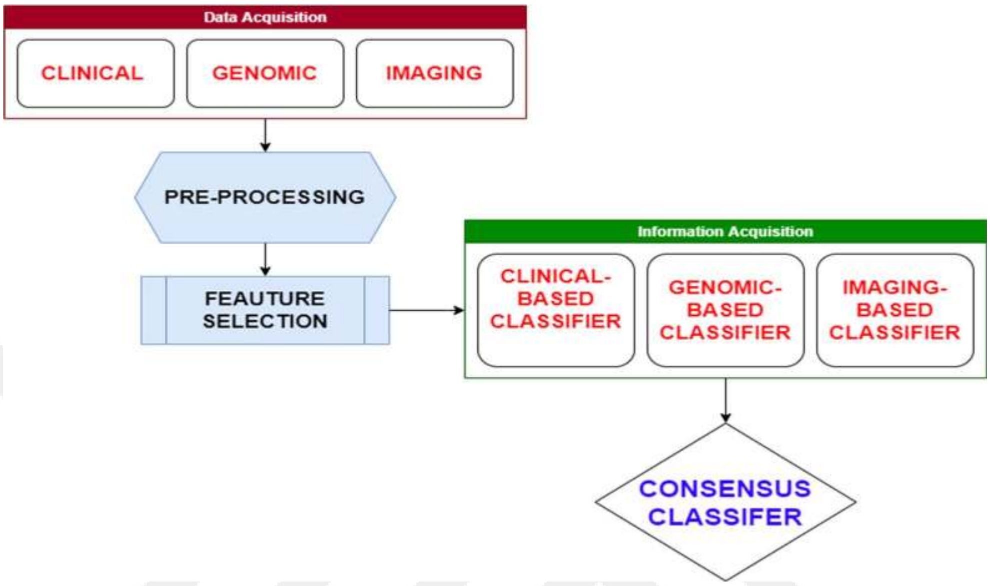

Figure 1.1: Types of Diseases Analyze Classification.

Brain Tumor is the abnormal and uncontrolled growth of tissues or cells in the brain, the brain tumor is dangerous and life-threatening disease, World Health Organization (WHO) classified the brain tumor based on malignancy. There are more than 130 kinds of brain and nervous system tumors. These may be malignant (cancerous) or benign (noncancerous), in any case, brain tumors maybe harmful or life-threatening. Tumors are categorized on the basis of cell origin and cell activity. Many clinicians use the classification system of the (WHO) to determine the type of brain tumor. The most popular forms of primary brain tumors include Glioma, Meningioma, Pituitary Adenoma, and Nerve Sheath [1].

1.2 INTRODUCTION TO IMAGE PROCESSING:

The Image processing expresses a form of signal processing, in which the images are entered for processing, and in the output it is in the form of either features or parameters specific to that image. It is used in many sections of the most important preprocessing in the classification and detection of objects in images. Moreover, It is a digital image where the application includes an individual being in a visual frame. In other words, the pictures are to be studied and operated upon by the people. The categories for image processing are many of including improving images, compressing, segmentation, analyzing them, and other different categories [3].

Detection of the diseases through image processing is done by using an integrated approach working methods of processing MRI images of brain tumor entering it and distinguishes this approach if the brain is normal or abnormal. It was also mentioned previously that manually detecting the brain tumor is an inaccurate and ineffective work, but using modern techniques for image processing. Analyzing the image and extracting the most important characteristics of these diseases and relying on image analysis and processing techniques to distinguish between different diseases based on the symptoms of the disease [4].

Thus, designing a fully automated and effective medical diagnosis framework is a major challenge. Generally, such a medical diagnostic system is capable of performing three types of tasks are the extraction features, the selection features and the classification.

1. **Feature extraction:** this stage is responsible for the extraction of all possible features of the MRI medical image that are intended to be useful in the diagnosis of brain tumors, without affecting the drawbacks of excessive dimensionality.
2. **Feature selection**: this stage is responsible for reducing dimensionality by dropping unnecessary data and looking for the best significant features to prevent the computational complexity and minimize computing complexity.
3. **Classification:** the issue arises when an image has to be allocated to a predefined category or class depend on a set of features observed. Figure 1.2 displays the standard architecture of the framework.

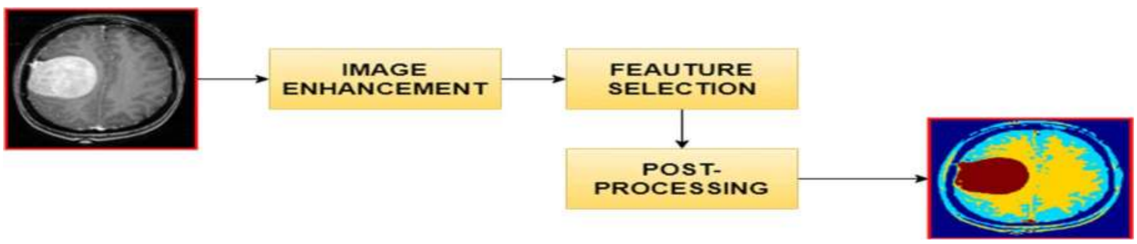

Figure 1.2: Standard architecture of the framework.

1.2.1 Problem Definition:

The present existing technology does not give any information about the tumor detection and classification, The classification of tumor is still performed manually, time-consuming, and dependent on human decisions, and May it’s leading to human errors, Problem is Brain tumors vary in size, shape, appearances, colour, location and orientation, which is the reason why tumor segmentation is challenging, Our system aims to perform classification techniques in order to classify the tumor to Normal (non-tumorous) or Abnormal (tumorous)

1.2.2 Research Objectives:

1. To build an efficient system for automating detecting brain tumors which is less time consuming and more accurate than the manual system.
2. To extract the significant features for brain tumors in medical images.
3. To perform segmentation techniques in order to detect and Extract the tumor.
4. To perform classification techniques in order to classify the tumor to Normal (non-tumorous) or Abnormal (tumorous)
5. To evaluate the performance of the proposed model.

1.2.3 Research Questions:

The following research questions have been formulated to set the direction for this research:

1. How to extract significant features of brain tumor in medical images that would be effective in identifying the normal tissues from abnormal tissues regions?
2. What are the good techniques for brain tumor in digital images?
3. How to evaluate the performance of the proposed model?
4. How would the proposed model perform compared with other techniques?

1.2.4 Scope:

Our aim is to develop an automated system for detection and classification brain tumors. The system can be used by neurosurgeons and healthcare specialists. The system incorporates image processing, pattern analysis, and computer vision techniques, Sensitivity, precision and efficacy of brain tumor testing are expected to increase. The clear understanding and experimental measurements of the above phases allows the production of supplement instruments which can aid in the early detection or tracking of treatment applications.

1.2.5 Thesis Organization:

The proposal is organized in six chapters as follows:

Chapter 1 gives an overview on information of the research. It describes the problem background and the problem statement, followed by research significance, objectives and questions. The chapter also presents the research scope and thesis organization.

Chapter 2 introduces the literature review related to medical images techniques, as well as the feature extraction methods and brain tumors classification aspects.

Chapter 3 provides a detailed description of the research methodology that has been followed to achieve the research objectives. The chapter introduces the research framework with its phases. It explains each phase and the techniques used to achieve its objective. It also demonstrates the research design, the research data and the performance evaluation and evaluation metrics for the detection model

Chapter 4 introduces the research techniques design and implementation. The research uses different algorithms in the research phases to attain its objective. It explains the procedure of each technique.

Chapter 5 displays the results discussion and evaluation. The results of all techniques are discussed, analyzed, evaluated against different works and compared with developed techniques and other research.

Chapter 6 demonstrates the conclusion and future works.

2. LITERATURE REVIEW

Brain Tumor is the abnormal and uncontrolled growth of tissues or cells in the brain, the brain tumor is dangerous and life-threatening disease, World Health Organization (WHO) classified the brain tumor based on malignancy. There are more than 130 kinds of brain and nervous system tumors. These may be malignant (cancerous) or benign (noncancerous), in any case, brain tumors maybe harmful or life-threatening. Tumors are categorized on the basis of cell origin and cell activity. Many clinicians use the classification system of the (WHO) to determine the type of brain tumor. The most popular forms of primary brain tumors include Glioma, Meningioma, Pituitary Adenoma, and Nerve Sheath [1].

The use of computers in medical applications has increased significantly over the last years. Typically, computers have been used in this area to read, scan, store, process, retrieve and analyze medical information. They have also been used as a method for the diagnosis and treatment of patients.

More recently, computerized image processing techniques have been used to improve the picture quality and images can be analyzed to highlighted areas of interest or to extract meaningful diagnostic features that can provide objective evidence to aid the human decision-making process. The integration of the two types of systems – databases and computers– aided feature extraction and allowed diagnostic information to be attached to patient records. This information can, then, be easily searched and cross referenced.

For this reason, the fields of analysis, detection and treatment of cancer diseases have become an important field of research, the use of computers and their techniques to analyze and detect this disease, and medical imaging techniques are one of the most important techniques that help in this field, It has already would have life-saving Effects on our ability to detect brain tumor and to diagnose the disease more accurately. However, in order to enhance the efficacy and accuracy of diagnosis and treatment, image processing concept has been successfully used to analyze and recognize the disease, evaluate treatment efficacy, and predict brain tumor development.

In [5] The authors has proposed an automated method for brain tumor segmentations from MRI images, they combined Chang-Vese model with the number of iteration for applying the region of interest (ROI) segmentation, the proposed method give promising result that could be helpful for clinicians and can be used in other various medical imaging modalities for obtaining precise clinical information rapidly.

In [6] they constricted a method for detection and classifying the brain tumor type, the MRI images used in the project since it gives a clear structures of brain, Machine learning predictions plays a key role in the proposed framework, the kernel support vector machine (SVM) was used the implemented model shows the good accuracy in classifying.

According to [7] they analyses a various medical image techniques and there properties in the context of brain tumor, they employed thresholding segmentation technique to detection brain tumor regions, the identifications is done efficiently.

Also the authors in [8] used Fourier transform, wavelet transform, ridgelet transform and curvelet transform techniques to represent the images. They shows that Fourier transform can only represent frequency and information domain, wavelet transform can represent both time and frequency domain signal, ridgelet transform can deals with images that without curves, curvelet transform can represent image with different scale and angles, thus, they enhanced curvelet transform to support dynamic texture classification, the proposed method shows significant performance and have better diagnosis of brain tumor.

Ahuja et al. [9] describe how pixel neighborhood elements can be used for images segmentation. For each, its neighbors are first identified in a window that contains tumor with its background that was grouped from MRI or CT scan images. A vector of these neighbors as individual gray values.

In [75] the authors proposed a framework for classify the brain tumors by combines statistical features and neural network algorithm, they use region of interest (ROI) to segment the tumors, and back propagation neural network to classification, and they obtained a total accuracy 91.9%.

The implementation of machine learning methods to medical image processing and pattern recognition problems has been generally accepted. Among the techniques developed for classification, Bayes Classifier, Decision Trees, K-Nearest Neighbors (KNN) Classifier, Artificial Neural Networks (ANN) Support Vector Machines (SVMs) and Expectation Maximization (EM) Classification Scheme [10].

In [11] a new approach for automated diagnosis and classification of Magnetic Resonance (MR) human brain images is proposed. The proposed method uses discrete wavelet for feature extraction, employs PCA for feature selection and exploits the capability of Back propagation neural network (BPN) and radial basis function neural network (RBFN) to classify brain MRI images automatically.

In [76] they used a convolution neural network algorithm (CNN) to classify the brain tumor from MRI images, they applied a numerus block to design the network such as convolution layer, pooling layer, and fully connected layer, the experiment achieved high average accuracy rate 94.39% using CNN model.

Classification was applied for that in [12] Probabilistic Neural Network (PNN). It is implemented because the training speed is several times faster than the Back Propagation Neural Network (BPN), the efficiency of the PNN Classifier has been evaluated in terms of training performance and classification accuracy. The Probabilistic Neural Network gives a fast and accurate classification and is a promising tool for tumor classification. The findings observed are substantially better than the results published Study using Wavelet Transform and other classification techniques.

In [13] The authors use a popular classification technique, Decision Trees due to their ability to model non-linear dependencies in the features, and their intuitive graphical representation of the learned model (as opposed to, for example, ANNs). Decision Trees perform classification by making a set of decisions based on the features, beginning from a root „node" and following decision made to other nodes in the tree where new decisions are made, leading finally where a classification is made.

In [14] Support vector machines (SVM), the technique of pattern recognition has developed from the theory of statistical learning. The basic idea of applying SVMs to solve classification problems is to turn an input space into a higher dimensional feature space through a non-linear mapping function and then create a distance matrix with the maximum distance from the closest points of the training set.

In the situation of linear separable data, the SVM is trying to find the one that distinguishes the training data with the maximum distance from their closest points between all support vectors that reduce the training error.

In [15], Sachdeva et al. established a Computer - aided diagnosis method to support radiologists in the multi-classification of brain tumors. A new hybrid machine learning method based on the Genetic Algorithm (GA) and the Support Vector Machine (SVM) for the classification of brain tumors is proposed. The texture and strength characteristics of the tumors are taken as input. Genetic algorithm was used to choose the set of most insightful input features.

In [77] the authors combined structural to detect the brain tumor from MRI images, they used extreme learning machines (ELM) to classify the images to their related extracted features and convolution neural network (CNN) to classify the tumors of the brain, and they gained result 93.68% accuracy rate of using KE-CNN.

In [16], a new improved approach is introduced by combining LDA & PCA for the reduction of features and SVM is used for the classification of MRI images. High precision for the selection and extraction of features is achieved compared to the previous work indicated in the literature discussed above.

In [17] Joshi et al. introduced a rapid identification and classification of brain cancer using the Neuro Fuzzy logic classifier. Texture characteristics are used for the preparation of the artificial neural network (ANN). Co-occurrence matrixes in various directions are determined and the Gray Level Co-occurrence Matrix (GLCM) features are extracted from the matrix.

In [18] the developed scheme explains how to extract the brain tumor from the patient's body using K-means as a clustering algorithm. In the second stage, they performs the classification of MRI brain image using decision tree and SVM classification algorithms and predict which better classification technique is used.

A variety of methods have been used to segment and predict the grade and volume of brain tumors. In [19] EI papageevgious et.al they suggested a fuzzy cognitive map (FCM) to determine the grade value of brain tumors.

In [20] a new approach to automatic brain tumor recognition based on k-means and possibilistic c-means clustering with region detection, which distinguishes brain tumors from healthy tissues in magnetic resonance images. Magnetic resonance images (MRI) features images used for tumor detection consist of T1-weighted and T2-weighted images for each axial region of the brain. The presented scheme involves three stages, including pre-processing, segmentation and extraction of features.

The implementation of the neuro-fuzzy segmentation method of the MRI images is applied in this report to detect different tissues such as white matter, gray matter, csf and tumor. The benefit of the linear self-organizing map and the fuzzy c-means algorithms are used to classify the image the performance of the ANFIS classifier was evaluated in terms of training performance and accuracy of the classification. Here the result verified that the proposed ANFIS classifier was more accurate than was 90 % has the ability to detect tumors.

In [78] the authors used MRI images to diagnose the tumors in the brain, they proposed automated brain tumor detection using convolution neural network (CNN) algorithm, by using a small kernel, rectified linear unit (ReLU) layer, pooling layer and fully connected layer, they achieved a high rate accuracy 97.5 % overall.

In [22] the authors suggest an automated and precise method for the classification of normal and abnormal magnetic resonance (MRI) images of the human brain. Rippled transform Type-I (RT), an effective multi-scale geometric analysis (MGA) method for digital images, is used to represent the main features of brain MR images. The search space of the image representative function vector is reduced by the principal component analysis (PCA). Algorithms less intensive support vector machine (SVM) called least square-SVM (LS-SVM) is used to classify the brain MR images.

3.THEORETICAL BACKGROUND

3.1 DIAGNOSIS OF BRAIN TUMORS:

Brain tumors are a group of tumors that arise primarily from the brain itself or its coverings or as a secondary seeding from other organs. They are considered among the ten most common tumors seen in the clinical practice. The diagnosis of brain tumors is based on the clinical presentation, radiological characters, histology and grading (obtained from the specimen of the tumor taken during surgery). The histologic type is based primarily on a histological assessment of cell types and tissue patterns recognized by conventional light microscopy. Immunohistochemistry and electron microscopy are important additional tools to define the tumor cell types. The grading of brain tumors intends to provide a general estimate of a lesion's anticipated biologic behavior.

3.2 MODERN NEUROIMAGING TECHNIQUES:

Modern neuroimaging techniques, including Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), have greatly enhanced the ability to a non-invasively detect and diagnose brain tumors and to visualize them in their very early stage. Each tumor type possesses special patterns in both CT and MRI which help reaching diagnosis with high accuracy. However, a completely accurate prediction of tumor grade by imaging appearance is not possible. Thus, the differentiation and characterization of these tumors still often depend on basic principles of neuroanatomy and neuropathology. A complete understanding of the basics as well as of the principles underlying the techniques will greatly improve the radiologist's diagnostic ability and accuracy [24].

Some brain tumors are treated by surgery alone; some need both surgery and radiotherapy. Other tumors requiring surgery, radiotherapy and chemotherapy are infrequently curable. The tremendous advances in imaging procedures have revolutionized our ability to detect and diagnose brain abnormalities and have influenced the development of therapeutic options. The results of a high-resolution imaging not only direct medical and surgical treatments but also favorably affect the outcome when the modality is used appropriately [24].

3.2.1 Magnetic Resonance Imaging (MRI):

The Magnetic resonance imaging (MRI) is an imaging technique used mainly in medical situations to achieve high quality images from the inside of the human body. Since its establishment in the early 1980s, MRI has been the primary mode of imaging for brain tumor detection and evaluation [25].

3.2.1.1 Processing and Advantages of MR Images:

The image represents a display of the MR signal. The signal intensity depends on both tissue and machine parameters. The latter can dramatically alter the grey scale on an MR image and they include proton density, T_1 and T_2 relaxation times and blood flow. Two main methods are used in the reconstruction of an MR image: 1) Back-projection: whereby a number of projections using different angles of rotation are made through a selected section. 2) Two-dimensional Fourier transform: using the mathematical process called Fourier transformation, the average MR signal from each picture element (pixel) is calculated. Typically, a pulse sequence is repeated 128 times, each with two averages, to obtain an image with an array of 128 x 256 pixels. Data can, then, be interpolated to a 256 x 256 or a 512 x 512 matrix for the final image display, in figure (3.1) shows a sagittal plan of human brain.

The main advantages of an MR imaging include:

1. Excellent soft tissue contrast.

2. Enables images to be made, at any orientation, such as sagittal, coronal, oblique or transverse axial without repositioning of the patient (multiplane imaging capacity).

3. Free from bone and air artifacts.

4. Does not employ ionizing radiation.

5. Has potential for tissue characterization.

6. Has potential for flow measurements.

7. Provides both anatomical and physiological information.

8. Yields images with a significant pathological contrast.

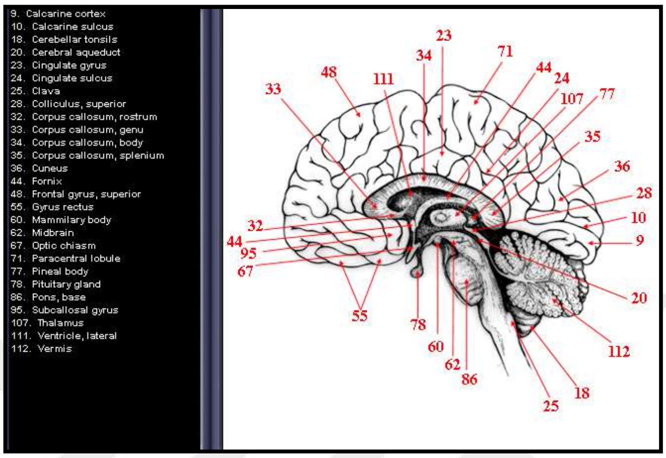

Figure 3.1: Sagittal plan of human brain [25].

3.2.2 Computed Tomography (CT):

This method of forming images from X-rays was developed and introduced into a clinical use by the British physicist Godfrey Hounsfield in 1972. In the conventional imaging with x-rays (planar imaging), an X-ray tube is placed on one side of the body, a suitable detector (usually film) is placed on the other, and an image is formed of the distribution of the photons that are transmitted through the body. In contrary, the X-ray output, in case of CT, is collimated to a very narrow beam. While this beam passing through the patient, it is partially absorbed, and the remaining photons of the X-ray beam fall on radiation detectors instead of X-ray film. The detector response is directly related to the number of photons impinging on it and so to tissue density. When they strike the detector, the X-ray photons are converted to scintillations which, then, quantified and recorded digitally. The information is fed into a computer which produces different readings as the X-ray beam is traversed around the subject. The first machines had only two detectors and used sharply collimated beams of X-rays. The modern CT generations use a fan beam of X-rays and multiple detectors. The patient to be imaged is placed in-between the X-ray source and the detectors. As the source-detector combination is moved linearly across the patient, the detector registers the intensity of the X-rays exiting the patient at each position. The X-ray intensity, versus position for one pass of the source-detector combination across the patient, is called a projection or view. After this projection has been acquired, the source-detector combination is rotated slightly to a new angle and a second projection or view is acquired. If sufficient views are taken, the distribution of attenuation coefficients within the cross section (slice) can be determined and presented as a tomographic image. This "image reconstruction from projections" is a mathematically complex process and is carried out by using a computer; hence the name computer-assisted tomography, another name applied to the CT. Thus, one-dimensional sets of data gathered at many different angles around the patient (the projections) can be combined mathematically (image reconstruction) to form the final cross-sectional CT image. The mathematical method used in an image reconstruction process is called filtered back projection which yields a potentially artifact free final gray-scale image. The mathematic theory of image reconstruction from projections is credited to J. Radon (1917); however, it was G. Hounsfield (1972) who first demonstrated the practical application of the methodology for a clinical use. A.M. Cormack applied the mathematic theory of Radon to the specific problem of CT and built a laboratory model. As a result, Nobel Prize for Medicine has been awarded jointly to Dr. Hounsfield and Professor Cormack in 1979 [25].

Similar to an MRI, two sets of the CT images are taken; before (CT without contrast) and after injection of contrast agent (CT with contrast) and three descriptive terms are utilized for the description of tumor's appearance on the CT image. If the abnormal tissue is similar in its gray value to the normal surrounding brain, it is called isodense; otherwise it could be either hyperdense or hypodense in appearance [25].

3.3 DIGITAL IMAGE PROCESSING

The image might be characterized as a function of two-dimensional f (x, y), where y and x are locative (plane) organizes, the plentifulness (f) at each couple of directions (x, y) is known as the grey level or intensity image at that point. When (x, y) the intensity portions estimations for the most part limited discrete amounts, it considers the image a digital image. The field of digital images processing alludes to processing digital images by methods for a digital computer. Note that an advanced image is made out of a limited number of contents, every one of which has a specific area f (x, y) one image is worth more than ten thousand information unknown and value. These components are called image elements, panels, and pixels. Pixel is the term utilized most broadly to mean the elements of a digital image [26].

3.3.1 Typical Structure of a Brain Tumor Detection Approach

The general structure of brain tumor detection consists of the following main stages:

1.**Image acquisition**: In this stage, images were taken from dataset for brain tumor that collected [26].

2.**Image pre-processing**: The primary goal of image preprocessing is an improvement of the image data that contains undesired distortions or enhance some image feature important for additional processing.

3.**Image segmentation**: This technique is used for the conversion of the digital image into several segments having some similarity. Image segmentation assists in the detection of objects and the boundary line of the image [26].

4.**Feature extraction**: The goal of the feature extraction process is to extract the most important features in the image after the image post-processing and these features are used to determine the specific sample used in the next stage [26].

5.**Classification:** It is considered from the important steps because, from this step, it can classify the tumor according to features that get it from brain image, for classification step used many algorithms like machine learning algorithms. Figure (3.2) views the general structure for detection of brain tumor and classification [26].

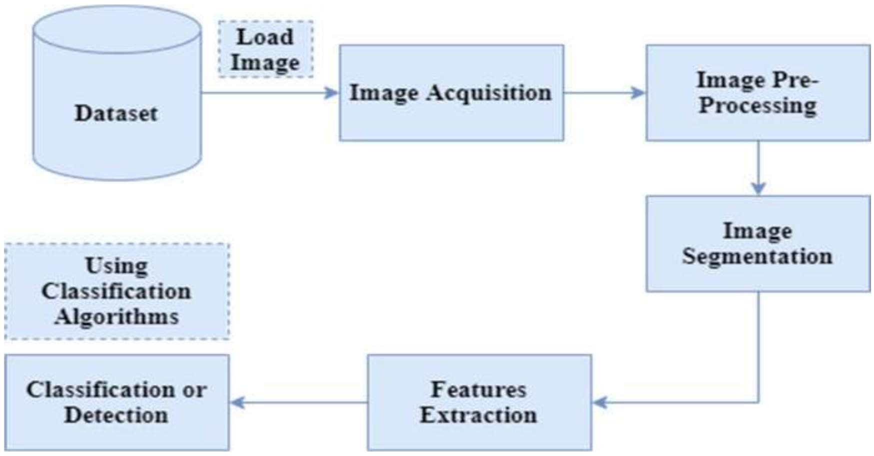

Figure 3.2: General structure of brain tumor detection.

3.4 IMAGE ACQUISITION

In image processing, it is characterized as the activity of retrieving an image from some source, generally a hardware-based for processing. It is the initial stage in the work process grouping, because, without an image, no processing is conceivable; the image that is gained is natural [26].

3.5 IMAGE PRE-PROCESSING

The image pre-processing is one of the basic stages in image processing. It improves image information by noise, the background removing and furthermore stifling of unwanted deformations. The improvement of image characteristic for investigation and processing also may image are resized from size to another [27]. An example of pre-processing is the process of converting from one color space to another.

3.6 IMAGE SEGMENTATION

According to the various processing image techniques, the segmentation image assumes an indispensable job in stage to do the given image will be analyzed. The segmentation image is the essential stage to images dissects and the information extract from it. This project bargains about essential standards of the techniques by the segment of the image. The segmentation has gotten an unmistakable goal in computer vision and image analysis. The segmentation methods of images like edge detection thresholding, clustering and region growing are account for this investigation. The segmentation algorithms depend on two characteristic discontinuity and similarity. Here must highlight on the technique that is broadly utilized to image segment [24].

3.6.1 Segmentation by Feature Based Clustering

Clustering is a step of division the gatherings according to their attributes. A cluster as a base includes a gathering of matching pixels that has a sit with a particular region and not the same as different areas. The term information clustering as equivalent words like cluster analysis, automatic classification, numerical taxonomy, and typological analysis. Images are gathered according to their substance. The techniques of clustering are generally sorted as partitional algorithms and hierarchical algorithm. In content-based hierarchical, the gathering is done dependent on the acquired features of the pixels example texture, shape etc. There are various clustering techniques use the most generally by the K-means algorithm [28]

3.6.2 K-Means Clustering Algorithm

K-Means algorithm is an unsupervised learning algorithm. It's one of the prevalent techniques. The k-means clustering, it parcels data an assortment into group's data of k number [28]. It arranges given information of arrangement into k number of cluster disjoint. K-means algorithm involves two separate parts.

In the first step, is account the k centroid and in the second step, it get every point to the cluster that has the closest centroid from the separate point data. That's have various techniques to define closest centroid is the separated and one of the most utilized techniques is Euclidean separation, when the grouping is finished it recalculates the new centroid of any cluster and up on that centroid. Another Euclidean space is determined between any middle and each data point and the points a signs in the cluster that have little Euclidean space. Every cluster in the parcel is defined use the objects part and its centroid are seen in figure (3.3) [29].

The point of centroid for any cluster that the aggregate distance from every object that make the minimized of cluster. So K-means is an algorithm iterative, which minimizes the sum of the distance from any item to its centroid cluster, total clusters. Give us a chance to think an image with a resolution of x * y and the image should be cluster into cluster of k number. Let p(x, y) be Ck be the cluster centers and an information pixel to be cluster. The k-means clustering algorithm is summarized in the following steps [29]:

1. Begin a number of cluster center and k.

2. The image of all pixel, account the Euclidean distance (ED), between the center and every the image of pixel utilizing the equation given below. Center and every pixel of an image utilizing the equation (3.1) [29] given below.

$$ED = \sqrt{(X_1 - C_1)^2 + (X_2 - C_2)^2} + \cdots + (X_n - C_k)^2 \qquad \text{………………(3.1)}$$

3. Allocate all the pixels to the nearest center depended on ED.

4. After all, pixels have been assigned; recalculate the new position of the center by calculate the mean utilizing the equation (3.5) [29] given below.

$$C_k = \frac{1}{K} \sum_{Y \epsilon Ck} \sum_{Y \epsilon Ck} P(x, y) \qquad \text{………………(3.2)}$$

Where Ck represented the cluster centers

5. Repeat the process unto it satisfies the tolerance or error value.

6. Reshape the cluster pixels into the image.

To facilitate the K-Means algorithm here will show it in the following figure (3.3) [30].

Although k-means has the advantage of being easy to implement, it has a few downsides. The quality of the last clustering outcomes relies upon the subjective choice of the initial centroid. So if the initial centroid is arbitrarily picked, it will get a diverse outcome for various initial centers. The initial centers will be intentionally picked with the goal that gets need segmentation. Furthermore, computational complexity nature is another term which has to reflect while structuring the K-means clustering. It depends on the number of information components, amount of cluster and number of iteration [30]. Anyway segmentation image is field advance for realizing and image analyzes that is a usually way to give semantic high level [32].

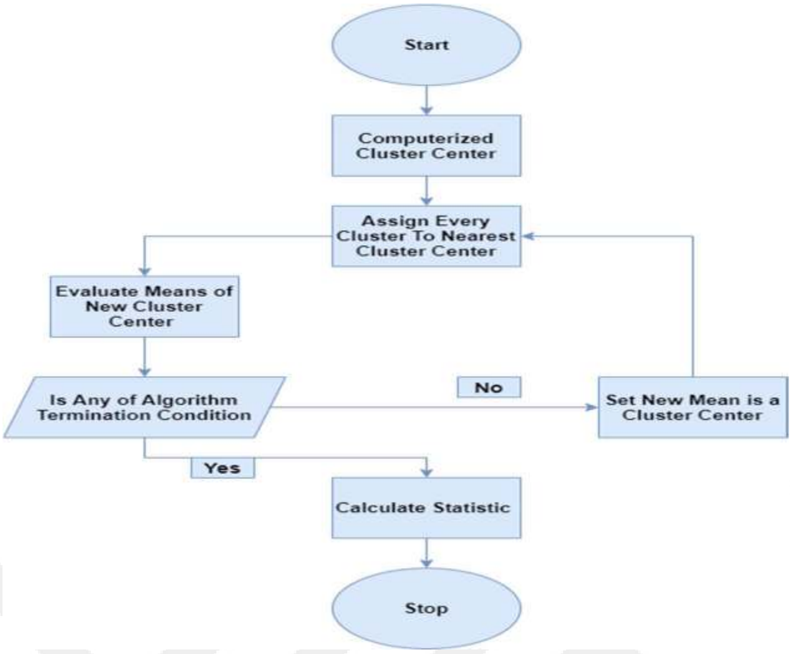

Figure 3.3: Scheme of K means clustering algorithm.

3.7 FEATURE EXTRACTION

Features extraction usually is done after the pre-processing stage and segmentation stage. The essential assignment of model recognition is to take an information design and effectively dole out it as one of the conceivable output classes. This procedure can be isolated into two general stages: Feature selection and classification. Feature selection is basic to the entire procedure since the classifier won't have the option to perceive from inadequately selected features. Extraction feature is a significant progress in the evaluation of each model classification of the viable data that portrays of every class. The procedure indicates that important features are released from objects to vectors of feature. These vectors of feature are utilized to see the classifiers information unit with the objective yield unit. It gets a classifier simpler to series between various classes by pick the features gander. Feature extraction is a procedure retrieves the ultimate significant information from the original data. The finding arrangement of the element when characterizes state of an element furthermore, remarkably [35].

The objective significant of the extraction feature is to evolve a lot of features, which expands the rate of recognition with the least sum components and create a similar feature set for a variety of instance for an assortment of an example of a similar image. The generally utilized feature extraction strategies are Gradient feature, Template matching, Graph description, Deformable templates, Zoning, Histograms, Fourier descriptors, Contour profiles, Unitary Image transforms, Geometric minute invariants, and Gabor features etc [35]. This introduction summarizes that feature extraction is a data minimization procedure to discover a subset of useful image dependent variables that are used and relied on in the data classification process. That will mention here some types of features and some methods used to extract them. Images are normally represented as shape, colour, and texture features. The most famous are colour and texture features.

3.7.1 Texture Feature

The texture is an important depiction of an immense arrangement of images. It is a helpful representation of the image's wide scale. It has been acknowledged, for the most part, that systems of human visual by texture for interpretation. In most ways, a pixel property is colored whereas texture must assume from a pixels group. Countless methods to extract the texture features were proposed in light of the domain when the texture feature is extracted, it’s extensively grouped into spatial extraction strategies of texture feature and extraction techniques of spectral texture feature. Texture feature is taken away using registering the statics pixel or searching the nearby pixel construction in genuine image space, though last transform image into the domain frequency and afterwards computes feature from the image is transformed. The texture is represented by two kinds of essential methodologies that are utilized in image processing which it spectral methodology and statistical methodology. A statistical methodology for texture portrayal texture represented in this methodology depends on the distribution of pixels and the connection between image pixels [26].

3.7.2 Statistical Based Features

The statistical in this project are classified into the first and second-order statistic [35].

3.8 FIRST-ORDER FEATURES

These features usually depended on the color, some of these features are:

1. Mean

Describes this feature the medium value, providing information for the total image pixels, it expresses in the equation (3.3) below [36]:

$$\mu = \frac{1}{N} \sum_{j=1}^{N} Pij \qquad \text{……..………………..……… (3.3)}$$

While N represents the number of pixels in the image and (P_{ij}) is the value of the j-th image pixel in the i-th color channel.

2. Standard deviation

The standard of deviation is additionally named the contrast square root, viewing the image contrast. It displays the spread of information the picture with a rising standard deviation and has an elevation contrast ratio. The standard deviation (SD) is described in equation (3.4) [36].

$$(\sigma)\ or\ (SD) = \sqrt{\left(\frac{1}{N} \sum_{j=1}^{N} (P_{ij} - \mu)^2\right)} \qquad \text{…………..…. (3.4)}$$

3. Entropy

The entropy is a measure that discloses to what number of bits it needs to code the image information, it is represented by the following equation (3.5) [36].

$$Entropy = -\sum_{i=0}^{N-1} P_i(x) log_2[P_i(x)] \qquad \text{…………..…. (3.5)}$$

Where Pi is the probability associated with image pixels.

4. Root Mean Square (RMS)

The Root Mean Square (RMS) calculates the amount of every column or row of the insert overall vectors of a predetermined dimension of the insert or of all insert. The Root Mean Square amount to j column of an M- by -N insert matrix u is obtained using the equation (3.6) below [37]:

$$RMS = \sqrt{\frac{\sum_{i=1}^{M}|u_{ij}|^2}{M}} \qquad \text{(3.6)}$$

Where M is the number of samples

5. Variance

Variance is the desire of the squared deviation of a random variable from its mean. Casually, it measures how far a lot of (random) numbers are spread, out from their average value. Variance is the square of standard deviation. The equation for obtaining theVariance is (3.7) [36]:

$$Variance = (SD)^2 \qquad \text{(3.7)}$$

(SD) represented the Standard Deviation

6. Kurtosis

Kurtosis depicts the state of probability dissemination, and similar to skewness, there are various methods for measuring it for a hypothetical distribution and comparing methods for evaluating it from a sample from a population. It is represented in equation (3.8) [37].

$$\text{Kurtosis} = \sqrt[4]{\left(\frac{1}{N}\sum_{j=1}^{N}(P_{ij} - \mu)^4\right)} \qquad \text{(3.8)}$$

7. Skewness

It aims to measure how asymmetrical the color distribution is and thus gives details on the shape of the color distribution. Its equation (3.9) represented below [36].

$$\text{Skewness} = \sqrt[3]{\left(\frac{1}{N}\sum_{j=1}^{N}(P_{ij} - \mu)^3\right)} \qquad \text{(3.9)}$$

3.9 SECOND-ORDER FEATURES

The first-order statistics of features produced obtain data introduced with the gray level appropriation of image. It's may that don't get any data concerning the overall positions of the diverse gray levels into the picture. These features shan't have the choice to calculate whether total minimum-value gray levels are placed together, or it is exchanged with the minimum-value gray levels. Some event of gray level design must be depicted using a matrix of frequencies relative Pθ, d (I_1, I_2). It how depicts frequently two pixels with gray levels I_1, I_2 appear in the window isolated by a distance d in direction θ [38]. There are more, methods for computing second-order features such as Gray-Level-Co-Occurrence-Matrix (GLCM).

3.10GRAY-LEVEL-CO-OCCURRENCE-MATRIX (GLCM)

GLCM focuses on co-occurrence matrix gray level. It maps the pixel brightness of a picture which takes place. It was offered by Harlick in 1970. It is additionally called Gray tone Spatial Reliance Matrix (GSDM). GLCM texture gets the connection between two pixels one after another, called the reference and the neighbor pixel. A matrix is developed at distance d=1 and at angles in degrees (0, 45, 90,135) as shown in figure (3.4). GLCM clarifies the distance and angular spatial relationship over a picture sub-region of a particular size. It is prepared from grayscale values are shown in figure (3.5) explained the GLCM generating process for angle 0 with degree 1. It is considered how regularly a pixel with gray level (gray scale intensity or gray tone) values come either vertically, horizontally and diagonally to leveled the pixels with the value j Harlick additionally offered various measures for example entropy, vitality, differentiates relationship and so on. These measurements at various angles and the figure (3.18) which explained an example the GLCM generating process for angle 0 with degree 1. GLCM direction are: Horizontal (0) Vertical (90) Diagonal a) bottom left to upper right (- 45) b) top left to base right (- 135) .They are declared as P0, P45, P90 and P135 respectively [54].

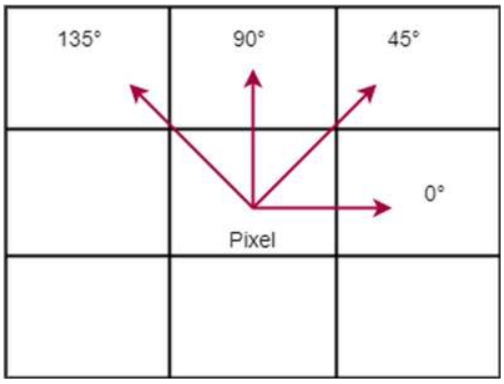

Figure 3.4: Angle direction in GLCM.

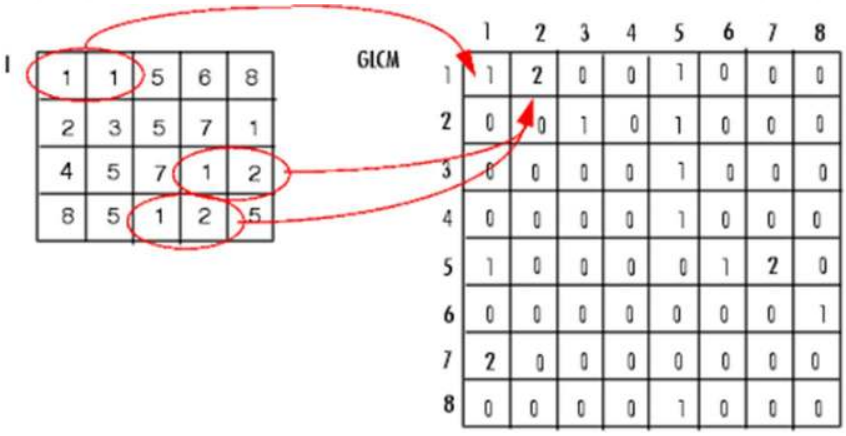

Figure 3.5: Process used to generating GLCM.

Harlick proposed many features textures are determined the probability matrix up on the matrix co-occurrence [40], some of these feature:

1. Contrast

The using this feature to estimate the contrast picture by the differences of the local measure, if the image has a high contrast of, the rate of contrast will be high as the equation viewed (3.10) [41].

$$Contrast = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} (i-j)^2 P(i,j) \qquad \text{………… ……….. (3.10)}$$

Here(i , j) the-spatial assimilate co-ordinates of the function P (i, j) the GLCM matrics, where (N_g) represent the gray tone.

2. Correlation

Correlation is an important feature since they can show a predictive relationship that can be employed in practice. This indicates statistic the proportion of gray linear tone displays in the image, its equation (3.11) [42].

$$Correaltion = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} \frac{(i-\mu i)(j-\mu j)}{\sigma i \sigma j} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(3.11)$$

3. Energy

It has different names, such as uniformity or Angular Second Moment (ASM), which can be characterized as the total of the squares of entries in the

GLCM. It estimated homogeneity of the image. When the value is high, the image has high homogeneity. It represented by equation (3.12) [41]:

$$Energy = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} P(i,j) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.12)$$

4. Homogeneity

Homogeneity or Inverse Difference Moment (IDM) is a feature measure the picture's homogeneity, its value being high at the point when the picture elements are similar. There is an inverse association with the contrast, implying that when the contrast value is high, its value is low, as in equation(3.13) [41]:

$$Homogeneity = \frac{\sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} P(i,j)}{1+(i-j)^2} \qquad \ldots\ldots\ldots\ldots\ldots (3.13)$$

3.11 CLASSIFICATION OF BRAIN TUMOR

Classification of image alludes to this images labelling into one of the numbers of previously classes. The type of image is most significant as a basic stage, for the processing high-level. The classifier is utilized for classifying images depending on their features. Classification is a significant module in brain tumor detection structures. Brain tumor detection approach that recognize brain tumor utilizing an image, thus classification here is characterized as a procedure of categorizing brain tumor images dependent on distinguished diseases [44]. In the proposed approach, the machine learning algorithms used to classify and detect brain tumor. Here will talk about machine learning, its types, and some algorithms that will be used.

3.12 MACHINE LEARNING(ML)

Machine learning is a subset of artificial intelligence that by re-iterative procedures allows a computer system to derive from the environment and improve itself. Machine learning methods sort out the information gain from it gather insights and make predictions reliant on the data analyzed without the need for additional programming that is definite. ML is utilized to teach machines how to deal with the information all the more efficiently. Some of the time after viewing the information, it can't interpret the pattern or concentrate data from the information. All things considered, it applies ML

[43] With the wealth of datasets accessible, the interest in machine learning is in the ascent. Numerous ventures from medication to the military to agriculture apply ML to separate significant data and classification for that data [44]. ML is divided into several common types as follows:

1. Supervised Learning:

In this type, the algorithm produces a function that plans the inputs to wanted outputs. The first standard definition of the learning task supervised is divided issue the learner is wanted to learn a function that vector maps change to one of the various types use looking at the function of manyinput_output cases [45].

2. Unsupervised Learning

It is used to draw conclusions from datasets comprising of input information without labelled reactions. The most well-known unsupervised learning strategy is cluster analysis, which is utilized for exploratory information investigation to discover hidden patterns or grouping in the information [45].

3. Semi-Supervised Learning

Its merges both labelled and unlabeled patterns to produce a suitable function or classifier.

4. Reinforcement Learning

In this type, the algorithm learns a pattern of acceptable behavior given a perception of the environment. Each action has some effect on nature, and the environment gives input that aides the learning algorithm.

5. Transduction

Like supervised learning, however, doesn't unequivocally generate a function rather attempts to predict new outputs dependent on training inputs, training outputs and the new inputs.

6. Learning to Learn

In this type, by relying on the previous experience the algorithm learns its own inductive [45].

The proposed approach used the first type of machine learning (supervised learning) that will clarify and some of its algorithms that have been used in detail.

3.13 SUPERVISED LEARNING

The machine learning assignment of a function learning a maps an input to an output dependent on model output-input.[45] the function derives from training labelled information comprising of many examples of training. Every model is a couple comprising of an insert object (vector commonly) and the ideal output amount (additionally called the "supervisory signal"). An algorithm analyzes of supervised learning down the training produces and information a function deduced that utilized for modern mapping models. Ideal situation will be consider algorithm to effectively decide the labels type for concealed instances. This wanted the algorithm learning of sum up from the training information concealed circumstances at "sensible" way [46].

Among the most important algorithms in the field of supervised learning is a support vector machine (SVM) algorithm and also within the deep learning algorithm that will be explained with its algorithms in the most important of which are the Artificial neural network(ANN) and the convolutional neural network(CNN), but before that the SVM algorithm will be detailed.

3.14 SUPPORT VECTOR MACHINE (SVM) ALGORITHM

The algorithm was invention in (1995), introduced this algorithm to resolve the kind of specific regression and classification problems. This method is based on the learning statistical method [47]. Its work to determine the optimal hyperplane that maximizes the classes' margin. SVM has become one of the classification's most popular methods. It is one of the supervised classification methods in which the inputs series with their labels is given. The structures defined by the vector features of these inputs. This algorithm produces a hyperplane distinguishing the two groups in order to achieve total separation between these classes. In figure (3.6) there are two vectors parallel to the classifier that passes via several stages. The distance between these two parallel vectors is called the (margin). The points on boundary vectors are called support vectors [48].

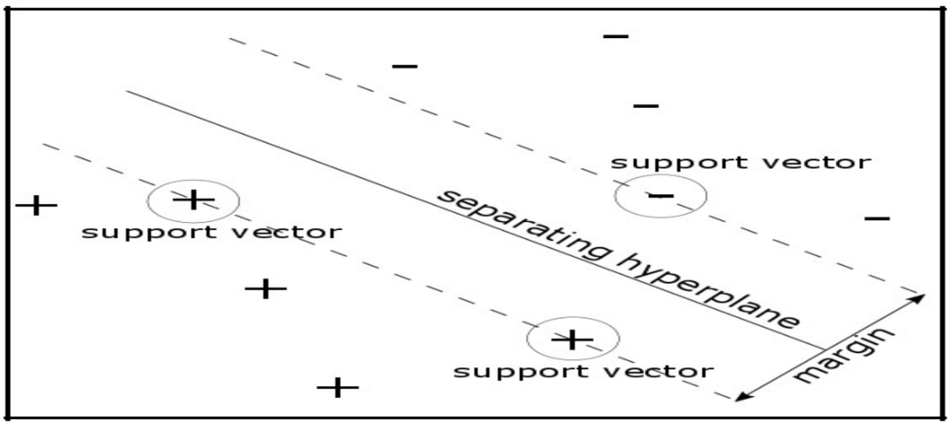

Figure 3.6: The SVM hyperplane between two classes [48].

The hyperplane is illustrated in figure (3.6) the separating data points from two groups as + and -. The vectors and margin of support are labeled.

SVM algorithm is divided into two kinds linear and non-linear SVM, the non-linear SVM illustrate in the following section:

3.14.1 Non-Linear SVM

In most cases, the linear classification does not consider the appropriate classification approach for the non-linear classification used in such situations, where a non-linear kernel function will be used. Linear SVM are fast to train and implement, but with many training examples and not too many features they appear to underperform on complicated datasets [50]. In many applications, non-linear SVM can be more consistent for quality across different problems and the preferred choice, although they lack critical power [51].

A- Kernel Function

To move testing samples and training to feature space high-dimensional feature space, functions kernel are implemented. This section, functions kernel are described to replace functions of mapping because the kernel computation is most effective than the function of mapping, computation period usually saved to exchange functions of mapping with the used of kernels [52]. SVM implements the kernel function, K ($x_{n,}$ x_i), which transforms the raw data space into a higher-dimensional new space. It process involves the dot product transformation function ϕ(x) (Equation (3.14). The target is the information that can be feasibly extracted, that has been converted into a higher dimension. It is possible to write the hyperplane function in formula (3.15) [52].

$$K(Xn, Xi) = \emptyset(Xn)\emptyset(Xi) \qquad \text{………………………….(3.14)}$$

$$f(Xi) = \sum_{n=1}^{N} a_n y_n k(Xn, Xi) + b \qquad \text{…………………….. (3.15)}$$

Where is a lag range multiple, is support vector information, and is a membership class label (+1, −1) with n = 1, 2, 3,……., N,[51].In this analysis, the three commonly used functions of the kernel at the linear, polynomial, and (RBF) Radial Basis Function, it's calculating in the following equations [52]:

1- **Linear** $K(X_n, X_i) = (X_n \,.\, X_i)$ …………..(3.16)

2- **Polynomial-of degree d** $K(X_n, X_i) = (X_n \,.\, X_i)^d$ …………..(3.17)

3- **Radial Basis Function (RBF)** $K(X_n, X_i) = \exp\frac{-\|Xn - Xi\|^2}{2\sigma^2}$ …………..(3.18)

B- Multi-Class SVM

The main requirement of the support vector machine algorithm was prepared to distinguish two classes, but there are conditions in life that need to more than two classes to be classified. Multiclass classification issues (k > 2) are usually broken down into a set of binary issues so that generic support vector machine can be applied directly. One-versus-rest (1VR) [54] and one-versus-one (1V1) [53] approach are two representative organization schemes for multi-class SVM. Both (1VR) and (1V1) are individual cases of the Error-Correcting Output Codes (ECOC) which disintegrates the multi-class problem into a predefined set of binary issues.

1. One-Versus-Rest (1VR)

The (1VR) strategy for k-classification constructs separates binary classifiers. The binary classifier m-th is learned to utilize the m-th type information as examples confirm and negative examples are k − 1 the remaining types. The binary classifier, which gives the greatest output value, specifies the class label during the test. The unbalanced learning set is a main issue of the (1VR) strategy. Assuming that total types have an equivalent number of examples training, in every individual classifier, the rate of plus to minus examples is 1k−1. This scenario, the initial problem's balance is missed [54].

2. One-Versus-one (1V1)

Another classical strategy to multi-type of classification is to decompose (1V1) or pair wise [53]. It tests all potential pair wise classifiers and therefore causes individual binary classifiers k (k − 1)/2. Introducing every classifier to an example of a test would give the having won class one vote, the best test is labeled with greatest votes for class. Each size of the classifiers generated by the (1V1) almost is greater than the against-one of rest strategy. The measure of

QP is smaller in each classifier, however, making it possible to train quickly. However, the (1V1) strategy is more symmetrical compared to the one-versus-rest method [55].

3.14.2 Advantages VS Disadvantages of SVM Algorithm

A. Advantage

1. It is marked with its high accuracy especially when used to classify two classes.
2. In this algorithm, over fitting is less important and is considered strong to noise.

B. Disadvantages of SVM Algorithm

1. A time that consumes
2. It is designed to solve two classes problems if have more than two classes, the accuracy decreases sometimes, and in this case, that must use 1VR solutions or other methods
3. It is computationally hard and expensive.

3.15 DEEP LEARNING

Deep learning (hierarchical learning or deep machine learning) is a department of machine learning based on specifics of algorithms that aim to model high-level data abstractions using multiple layers of complex processing with systems or otherwise consisting of multiple non-linear transformations [57]. An analysis (e.g. an image) can be depicted in many ways as a vector of intensity values for every pixel, or as a set of edges, areas of a particular shape, etc in a more abstract manner. Many characterizations make learning tasks easier (e.g. recognition of brain tumor, recognition of the face, or recognition of facial expression) [58]. So, it is a term that includes a specific approach to neural network construction and learning. Since the 1950s, neural networks have existed and, such as nuclear fusion, they have been an extremely ambitious laboratory concept whose practical application has been constantly delayed.

They take a number matrix (which can represent pixels, word, or shapes of audio). On that set encompass a sequence of functions and output one or even more numbers as outputs. The outputs are typically a prediction of certain properties that you are trying to determine from the data, such as whether or not an image is a cat picture [59]. Deep learning techniques are dependent on distributed representations. The fundamental theory that underlies distributed representations is that experimental information is generated by layer-structured factor correlations. Deep learning incorporates the assumption that the levels of abstraction or structure apply to these component layers. Various layer numbers and sizes of layers can be used to provide various abstraction products [60].

Deep learning uses this concept of explanatory hierarchical variables in which higher-level, more abstract principles are learned from lower-level ones such architectures are often designed using a greedy model of layer-by-layer. Deep learning is attempting to isolate these concepts and to determine what features are useful for learning. Its techniques eliminate function engineering in supervised learning activities by transforming the data into compact intermediate formats close to main components and obtaining layered architectures that remove consistency in representation. Deep learning algorithms are many and the most famous of these algorithms are Artificial Neural Networks (ANNs) and Convolution Neural Network (CNN) [60].

3.16 CONVOLUTION NEURAL NETWORK (CNN)

The Convolutional Neural Networks are the most common machine learning techniques in which several layers are automatically trained [61], have been shown to be highly successful and are additionally the generality by various applications of vision computer. The general structure of the convolutional neural networks system is viewed in figure (3.7) [62].

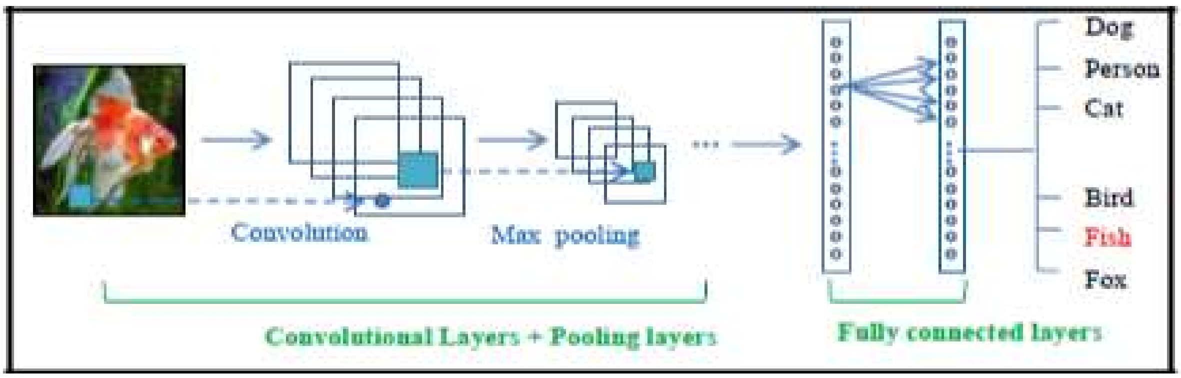

Figure 3.7: The general structure of the CNN system [62].

Commonly, CNN consisting of three basic neural layers, convolutional layers, pooling layers, and layers that are fully connected. Various types of slices take various roles. The figure 3.8 is showed a general convolutional neural networks structure for classification of image is presented layer by layer [63]. For train the network, there are two stages: forward and backward. The forward stage's main objective is to specify the input image in every layer with the proposed parameters (weights and bias). Later the output of the prediction is used with the ground of the truth labels to calculate the estimate of loss. Second, relative to the cost of loss, the backward stage determines with chain rules the gradients of per parameter. All parameters are modified and prepared for the next forward calculation dependent on the gradients. After enough variations of the stages forward and backward, the learning network can be ended [62].

CNN and traditional Neural Networks (NNs) have major differences; the best is CNN, due to these differences, which are among the advantages of CNN:

1_ CNN input can handle either 3dimensions or 2D and 1D sizes focused on configurations and parameters, while NN input is just a 1D array

2_ CNN lists more layers than regular Neural Network.

3.16.1 Basic Structure of CNN

Convolution neural network is identical to artificial neural network, as both are composed of self-optimized neurons, introduced by inputs and executing non-linear transformation. Convolution neural network is commonly utilized in pattern recognition on objects relative to artificial neural networks, because it encodes image-specific features in system architecture. In the convolution neural network, there are five fundamental elements: the input layer, convolution layer, non-linear layer, pooling layer and fully connected layer [63].

1. Input Layer

The stage includes the inputted image as well as their pixel values. These images are the images of brain tumors that entered in matrix form.

2. Convolution Layers

The convolution neural network (CNN) uses separate kernels in the convolutional layer to transform the whole image to the appropriate features map, producing different feature maps [62].

In another meaning a flip and gathering over the kernel. This is extended to two dimensions [82]:

$$S(i,j) = \sum_m \sum_n I(m,n)F(i-m, j-n) \qquad \text{.......... (3.19)}$$

In another meaning flip the filter (F) up-down and left-right and sum over all products are shown in an example about convolution layer in figure (3.10). The convolution method has three key advantages [65]:

1) The weight distribution functions in a similar map of feature to decrease the number of parameters.

2) Spatial learns communication associations between nearby pixels.

3) Invariance to the object's position.

The primary idea is that the conventional replacement of convolution layer with a tiny multilayer softmax made up of completely many connected layers with functions nonlinear, substituting linear filtrate with neural networks nonlinear. This approach produces best results when classifying images.

Figure (3.8) shown a, b, c convolution example process with a kernel measurement of [3×3], with no padding and one stride, which the stride means the step scale to vertically and horizontally traverse. The kernel is extended around the insert tensor, and the component-wise function between each object of the tensor input and kernel is measured at every point and averaged up to produce the output amount in the related direction of the tensor output called the map feature [66].

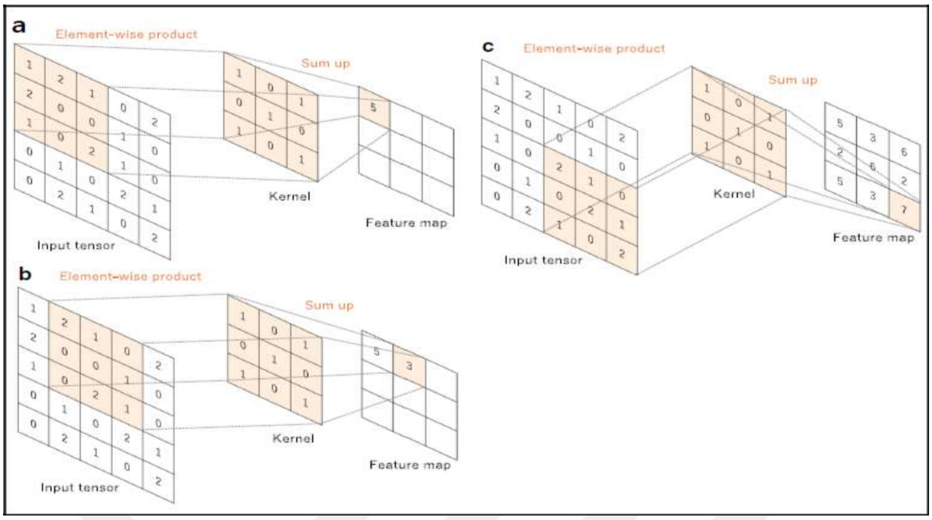

Figure 3.8: An example of convolution operation [66].

3. Activation Function

The activation function also is applied in this work, the non-linear translation function in the artificial neural network is sigmoid or hyperbolic tangent, this layer uses many types of activation functions the most popular is Rectified Linear Units (ReLU). Though, if the sparsity of the data is greater, the outcome will be better for image processing. This layer uses many types of activation functions the most popular is Rectified Linear Units (ReLU).

A. Rectified Linear Units (ReLU)

Rectified linear units (ReLU) are generally used as nonlinear transition, the following equation is used for rectified linear units [84].

$$y = Max\ (x, 0) \qquad \text{………………………} \ (3.20)$$

So that the output is in the same size as the input. Rectified linear unit raises the decision function's non-linear properties and has no negative impact on the convolution layer's sensitive fields. The training rate of the rectified linear model is much higher compared to other non-linear units. The figure (3.9) an example of the Rectified Linear Units (ReLU) [67].

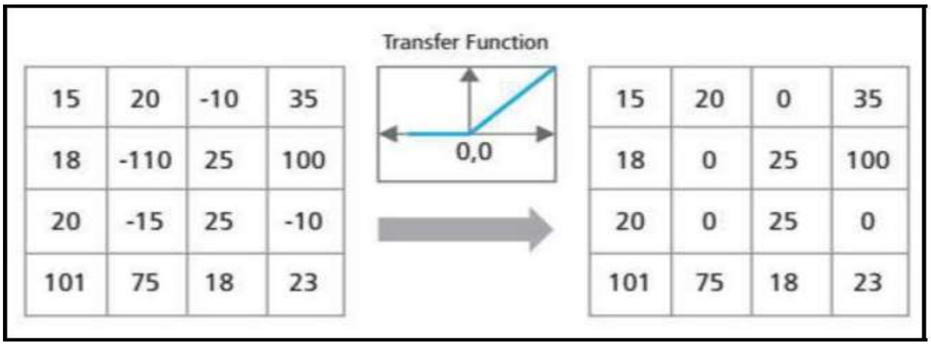

Figure 3.9: An Example of ReLU Transformation [67].

4. Pooling layers

It's usually follows the layer of convolutional and used to minimize parameters of the network and the dimensions of feature maps. Pooling layers also are symmetric in translation, as are convolutionary layers, so their measurements take into account adjacent pixels. Average pooling

And max pooling are the most commonly used techniques. The following figure (3.10) explains the processes of both kinds of pooling.

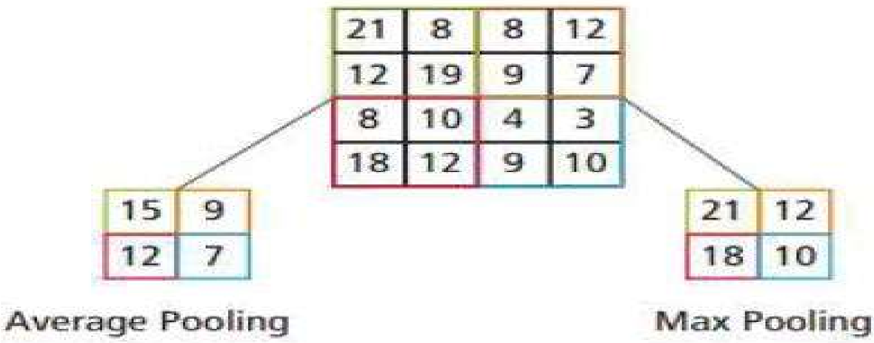

Figure 3.10: Two Classic Pooling Methods [67].

The pooling layer applies the max-pooling function with a 2 x 2 kernel and a stage of 2 along the spatial dimensions of the input, according to the reference cases. It is based on the care about the pooling layer's destructive functionality. Through this process, the feature map is reduced to

25 % of the previous size while keeping the depth capacity at its standard size. It deducts the layer to use the intervening region to move through the whole spatial dimensionality of the data. If the stage is adjusted to 3 with a kernel size set to 3, the model's output will be effectively reduced [68].

5. Fully Connected Layer

The data is accessed on the final layer of the neural convolution network (CNN), Often the last layers of a CNN are fully connected layers, so mainly the two neighboring layers inside the network will directly connect by a fully connected layer. The objective from this layer is, to sum up, the weights of the features from the previous layers and show the likelihood of per class. For example, if there is a neural convolution network for gender classification and the result matrix is a probability of [0.8, 0.2], this means that there is 80 % probability of male gender and 20 % probability of female gender [69]. Also, there are several other concepts related to the CNN structure such as MLP, normalization layer and softmax function.

6. Multi-Layer Perception (MLP)

Finishing a CNN with a multi-layer perception is popular. In practice, this is just one pixel for every filter for convolutional layers. Also used in the term Deep Neural Networks which is the product of adding layers that expanding the network size. The decision boundary is no large limited to a hyperplane when incorporating hidden layers but can lead to non-linear solutions as shown in figure (3.11) [70].

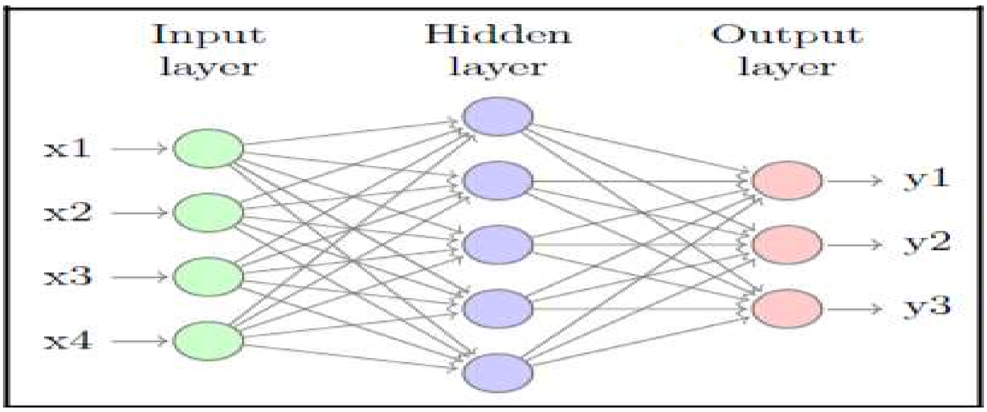

Figure 3.11: Schematic representation of an MLP [70].

7. Normalization Layer

The input layer is normalized by modifying and scaling the activations, for e.g., if it has features from 0 to 1 and also from 1 to 1000, also must normalize those features to accelerate learning. If it supports the input layer, why not do the alike for the values in the hidden layers that shift all the time and increase the training pace by 10 times or more. There are various types of normalization, the most famous of which is the batch normalization.

Batch normalization: It is a process of improving the speed, execution and stability of neural networks. In a neural network, batch normalization is performed by a normalization phase that adjusts the means and variances of the inputs of each layer, computing the mini batch mean () by the equation (3.21), m is the values of this activation in a mini batch [71]:

$$\mu B = \frac{1}{m} \sum_{i=1}^{m} xi \qquad \text{……………. (3.21)}$$

While computing the mini batch variance by the equation (3.22):

$$\mu B = \frac{1}{m} \sum_{i=1}^{m} xi \qquad \text{……………. (3.22)}$$

At the last, normalize the layer inputs by using the prior calculated batch statistics as in the equation (3.23) [88]:

$$\overline{xi} = \frac{Xi - \mu B}{\sqrt{\sigma^2 B + \epsilon}} \qquad \text{..................................... (3.23)}$$

8. Softmax

The output of the neural network can be difficult to understand. It's typical to end CNN with a softmax function on classification purposes. The following equation (3.24) expresses the Softmax [70]:

$$yi = \frac{e^{Zi}}{\sum_{j=1}^{N} e^{Zi}} \qquad \text{……………………...(3.24)}$$

Where the N is the number of the input Z_i is the input layer value, this normalization of the total amount of outputs to one can be defined as the likelihood of the input belonging to class i and therefore the softmax output y_i [70].

3.16.2 Training a Network

It is the method of obtaining kernels in convolution weights and layers in fully connected layers, which reduces the gap between defined ground truth tables in the training dataset and output predictions. The back propagation technique is the approach widely used to train neural networks where the concept of loss and the optimization of gradient descent algorithms play important tasks. A model efficiency under different weights and kernels is determined by absence function by forwarding propagation on learnable parameters and a training dataset, namely kernels and weights, are modified by the loss value by, among other items, an favor technique called stochastic gradient descent with momentum (SGDM) [66].

3.16.3 Stochastic Gradient Descent with Momentum (SGDM)

Stochastic gradient descent (frequently abbreviated SGD) is an iterative approach to optimize an actual function with appropriate smoothness (e.g. differentiable or sub differentiable) attributes. It can be considered as a stochastic estimate of gradient descent optimization as it substitutes the true gradient (obtained from the whole dataset) with an average of it (computed from a randomly picked data sub-set). Stochastic gradient descent in machine learning has become a significant method of optimization. The trouble of reducing a function objective in the form of average is known by both statistical analysis and machine learning [73]:

$$F(x) = \frac{1}{n} \sum_{i=1}^{n} F_i(x) \qquad \text{.........................(3.25)}$$

When the x parameter that decreases F (x) to be determined. Any function summand F (i) is generally joined in the i-th check in the set of data (used for training). A "batch" or standard tendency descent technique would carry the following equation [72].

$$x = x - \eta \nabla f(x) = x - \eta \sum_{i=1}^{n} \quad \nabla fi(x)/n \qquad \text{..................... (3.26)}$$

Represent a stage measure (sometimes named the rate of learning in machine learning) [72].

3.16.4 Loss Function

This function is known as an estimate function, calculates consistency through the network's output estimates by assigned ground truth labels and forward propagation. Cross entropy is the widely utilized loss function for multiclass grouping, while mean squared error is usually extended to continuous values for regression. One of the hyper parameters is a form of loss function and it requires to be decided according to the tasks provided [66].

3.16.5 Back-propagation

The back-propagation network is considered to be mostly used in the domain of neural network models. The core architecture of a multi-layer neural feed forward network based on a back-propagation training algorithm (build). The output sends out, and then the errors of the whole network are measured and re-propagated back to a certain layer. Throughout this case, the network receives the signal by the neurons in the input layer, and the production of the entire neurons is generated by the specific learning NN paradigms in the output layer of the network. The output can be repeated by one or more intermediary hidden layers. The idea of a back-propagation learning algorithm is generally utilized to fine-tune training parameters in which the errors should be decreased [73].

3.17 EVALUATION CRITERIA

Performance evaluation of classification algorithms conducted in various methods from the confusion matrix of these methods involves information on predicted and observed classification produced by the classification method. The confusion matrix includes data from which output can be determined. The table below displays the confusion matrix of two classes [74]

Table 3.1: Confusion matrix Table

	TP	TN	FP	FN
Actual				
Tested				

The following equation can be used to calculate accuracy [74]:

$$Accurcy = \frac{a+b}{a+b+c+d} \qquad (3.27)$$

Or use the equation below

$$Accurcy = \frac{number\ of\ correctly\ classified\ images}{total\ number\ of\ images} \times 100\% \qquad (3.28)$$

4.THE PROPOSED FRAMEWORKS

4.1 THE PROPOSED APPROACH DESIGN

Our proposed frameworks aims to identify and segment the brain tumors in the medical image; the work includes two proposed approaches. They are separated based; each one is different from the other.

All details of the proposed frameworks are described in the following sections. According to figure (4.1) below shows, the general block diagram of the approach for detecting brain tumor and its stages will be explained later, the approach will be divided into two parts depending on the detection method.

The main steps of this approach are as below:

1.In the first proposed approach, the support vector machine (SVM) algorithm is applied and their stages will be done up to the SVM detection algorithm for brain tumors.

2.In the second proposed approach based on deep learning where convolutional neural network (CNN) algorithm is applied. It also has various stages depends on building this work.

3.This approach will detect tumors and take into account the most common tumor.

4.Through this approach and in the end, the brain tumors will be detected by applying the stages of the proposed methodology of the approach.

Figure (4.1) below illustrates the major stages of a general diagram which will be applied for the two used proposed approaches. In this general diagram, the stage (image acquisition) is a common stage between the two proposed approaches and will be discussed in detail later. In the first proposed approach, there are several stages that will have been carried out after image acquisition, these include, image pre-processing, image segmentation, image post-processing, feature extraction, and the detection stage.

In the second proposed approach, after image acquisition there is an image pre-processing stage, the CNN structure has been built, and after that training and testing operations will use the dataset to complete the detection process. Before explaining the two proposed approaches, the joint stage between these two proposed approaches, which is the (image acquisition) stage, will be explained.

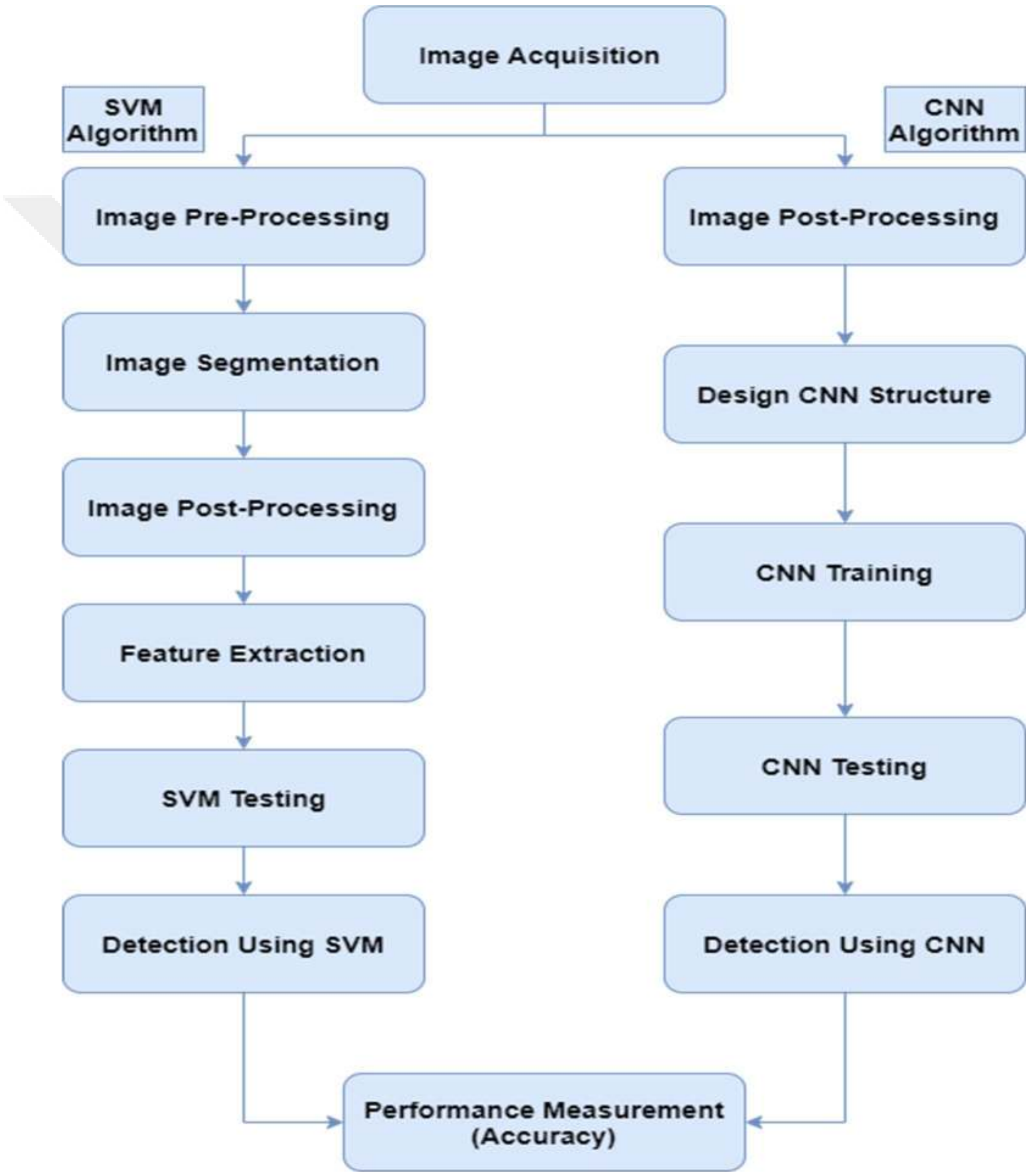

Figure 4.1: General block diagram for Brain Tumor detection approach.

4.2 IMAGE ACQUISITION STAGE

The difficulty of this work is to create a dataset, the proposed approach seeks to detect and classify brain tumors, and this is the main reasons why it is too difficult to create a dataset consisting of thousands of different images. For these reasons, we have recourse to the brain tumor treasury which is a highly popular and certified global site that contains thousands of images of brain tumors. In the proposed approach, three very important tumor types were chosen. These images and the specific types for each tumor will be classified according to the symptoms that have been mentioned in chapter 3. Every sample in the dataset was gray color space with size 256 * 256 pixels and the images in the JPG file format.

The approach is divided into two main sub approaches, the first proposed approach was using the Support Vector Machine (SVM) algorithm, while the second proposed approach using the Convolutional Neural Network (CNN) algorithm for detection the brain tumor, and they will be explained in details in the following sections.

4.3 THE FIRST PROPOSED APPROACH: (SUPPORT VECTOR MACHINE ALGORITHM (SVM))

SVM is an algorithm considered to be amongst the popular algorithms of machine learning techniques. Its definition, details, and equations are presented in chapter 3, section (3.17). This section uses SVM for classification after completion of all stages of the approach, these stages are shown in the figure (4.1) and a brief explanation for each stage is presented in the next section.

4.4 IMAGE PRE-PROCESSING

The pre-processing stages are carried to make the images more suitable for further processing. This process is always modifying or smoothing images to make them better and prepare them for the next stage of it is also used to perform another specific process that needs in the next stage. In this proposed approach, this process is needed to convert the images to the gray-scale level. We present a steps that aims to improving and enhancing the quality of the images, as well as removing undesirable distortions such as noise and staining saturation. For this context, we present a set of optimization techniques described in the following paragraphs.

4.4.1 Converting to the Gray Scale

In this stage, we're converting the MRI images of the brain tumor to the gray color space. Then we collect just the 1 channel space, which is the gray scale, which will be used in all of the following process.

4.4.1 Image Normalization

The goal of the image normalization is to resolve the color inconsistency that results in the overall color variance of the images. In digital mammograms of brain images, color variability is attributed to several causes, such as illumination, scanner type, and staining differences from one laboratory to another.

4.4.2 Noise Redaction and Removal

Noise in image processing can be characterized as a difference in brightness which does not fit its actual variance. Noise can occur during the collection of the image, like sensor noise, due to poor lighting. To check the noise, we randomly picked one picture from the dataset and calculated its histogram as seen below in Figure 4.2.

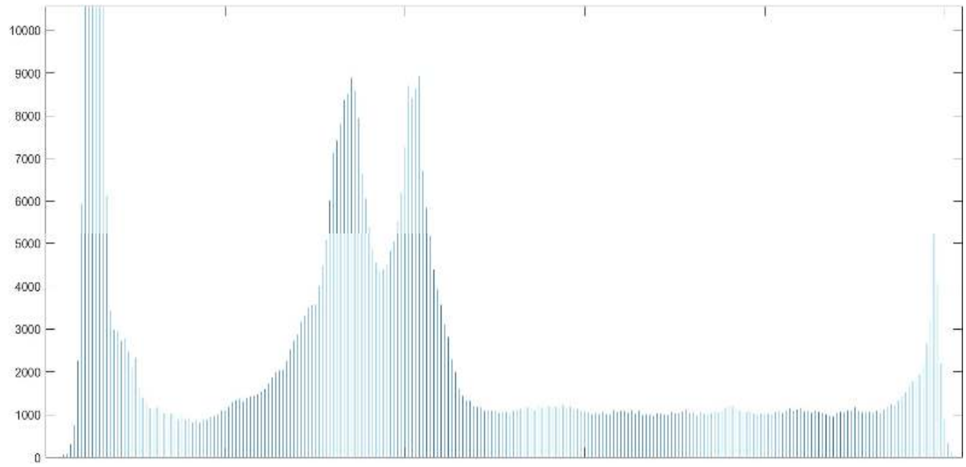

Figure 4.2: The histogram of random MRI image selected from the dataset.

As seen in Figure 4.2, the histogram provides a regular distribution that can be compared with a number of techniques. In this sense, we define the Mean and Median filters for noise reduction and image softening.

4.4.2.1 Mean Filter

Mean filter also called Average filter is most commonly used as a simple method for reducing noise and smoothing image. The process of mean filtering is simply to replace each pixel value

In an image with the mean 'average' value of its neighbors, including itself. This has the effect of eliminating pixel values which are unrepresentative of their surroundings. We Applied 3 × 3 averaging filter with coefficients of 1/9 is used to a random brain image to evaluate the Median Filter. A typical example of filtering is shown below in Figure (4.3).

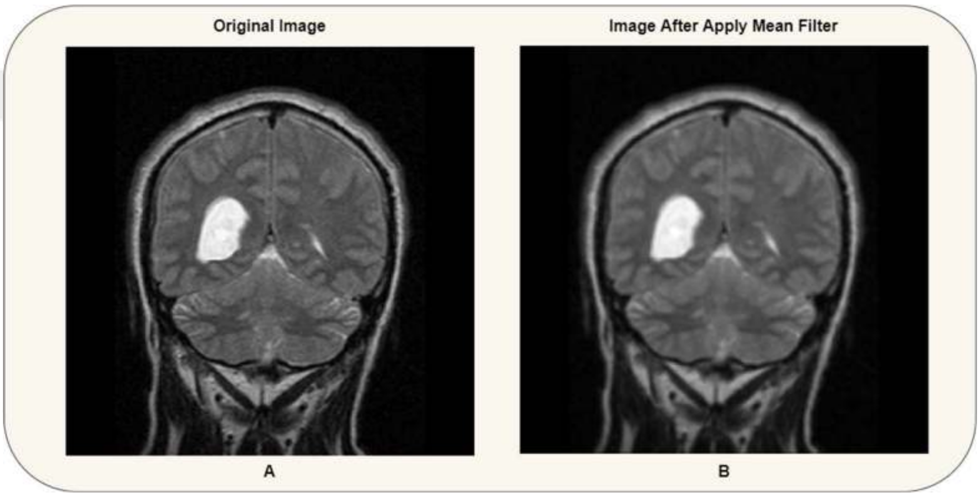

Figure 4.3: (A) original MRI image of brain – (B) Outputted image after applied mean filter for smoothing.

4.4.2.2 Median Filter

Median Filter is non-linear method use to remove noise from image and it's widely used as its very effective and good performance to remove noise, the process of median filter work by moving through the image pixel by pixel replaced each value with the median value of neighborhoods pixels. We Applied Salt & Pepper noise filter to a random brain image to evaluate the Median Filter. A typical example of filtering is shown in Figure (4.4).

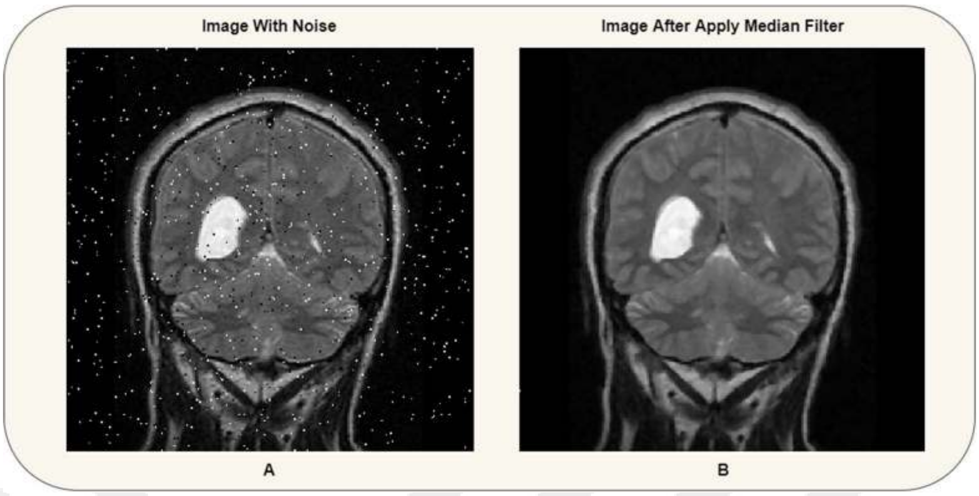

Figure 4.4: (a) an image affected with salt & pepper noise filter – (b) Outputted image after applied median.

4.5 IMAGE SEGMENTATION

In order to distinguish the affected area of the brain tumor, the image segmentation process is carried out. This stage is very important in order to determine the part that will be working on later for the process of extracting the selected features. The image segmentation was previously explained with their types in chapter 3, section (3.8). The proposed approach will give more trustworthy results using cluster subtraction by using the K-mean clustering algorithm.

4.5.1 Images Segmentation Using K-Mean Clustering Algorithm

In this stage, the segmentation of the image using the K-mean clustering algorithm. This algorithm was explained in detail in chapter 3, section (3.8.2). In this method, the image of the brain tumor is divided into k clusters where each remark has a proper place to the region with the nearest mean, pointing to the cluster as a concept. In this work, the Euclidean Distance measure was utilized; the clustering of k-means differentiates the sample image into three clusters with relation to the color pixel of each component. So, it needed to convert to gray space because the process of segmentation depends on the illumination of the color. All the nearly related colors will be in the cluster and will choose the cluster that segments the disease based on its color as shown in figure (4.5). The algorithm (4.1) clarifies the K-mean clustering algorithm for the segmentation of brain tumor images.

Algorithm(4.1) k-mean clustering for image segmentation	
Input:	Medical Image size 256 *256
K //	Number of cluster
ED //	Euclidian distance
Ci //	Cluster centroid
Output:	Segmented image
Step(1):	Begin
Step(2):	Select number of clusters k
Step(3):	Select the initial cluster centroid *ci*, *i* = 1... k randomly.
Step(4):	Compute the Euclidian distance using equation (3.4)
Step(5):	Go to the following points and measure the new centroid
Step(6):	Update new centroid by mean of every clusters point
Step(7):	Compute ED for the next check
Step(8):	Continue until no change in centroid
Step(9):	Return segmented image
Step(10):	End algorithm

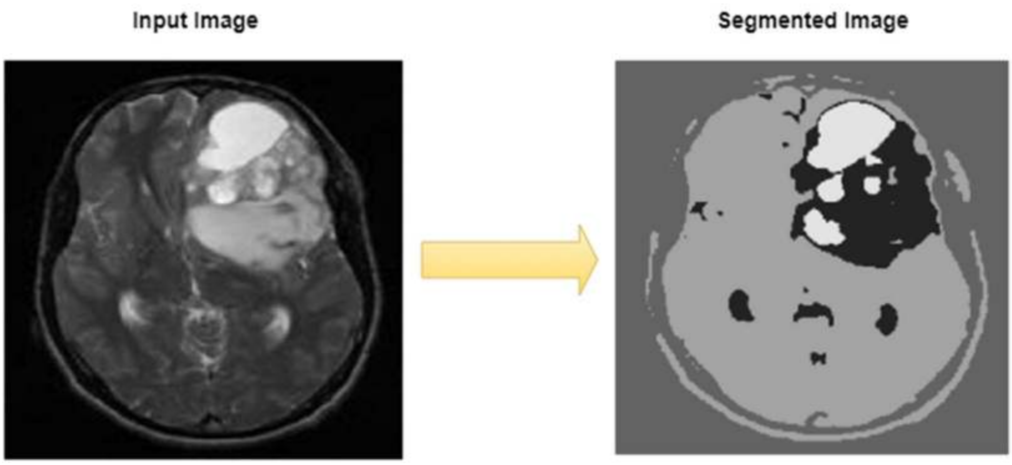

Figure 4.5: Image segmentation after using the k-means clustering algorithm.

4.6 FEATURE EXTRACTION STAGE

The different features are extracted after segmentation to characterize the infected area. Important features of the images are extracted for later use in the classification process, based on these features. There are several types of features that were explained previously in the third chapter, section (3.10). Extracted additional features are very important by utilizing the Gray Level Co-occurrence Matrix (GLCM) method, from which extracted four important features that were used in the classification process, upon features arrangement, at the beginning will be explained first-order features.

4.6.1 Statistical Features Extraction (First Order Features)

The statistical features are extracted from the images after performing the operations that were demonstrated previously in section (3.11). These features depended on the color moments of the image is used in the calculation, as illustrated in chapter two. The features are taken from the color image; which means the segmented images in the previous stage don't need to convert to gray-scale for certain features. These features are "mean", "standard deviation", "entropy", "root mean square"(RMS),"variance", "kurtosis" and "skewness", are shown in table (4.1). The table (4.1) explains each feature with its description and the number of the equation that was used for calculating it.

Table 4.1: First-order features

No.	Features	Description of features	Equation numbers in chapter two
1.	Mean	It's a medium value, (Average) or mean of matrix elements	(3.7)
2.	Standard Deviation	It is also named the square root of the contrast, showing the contrast of the image	(3.8)
3.	Entropy	It defines corresponding intensity level states which can be adapted by individual pixels	(3.9)
4.	RMS	It is described as the square root of the mean square	(3.10)
5.	Variance	It is the assumption that a random variable will squarely deviate from its mean	(3.11)

6.	Kurtosis	It represents the state of probability dissemination	(3.12)
7.	Skewness	It is a metric of the imbalance of the distribution of the likelihood of a real-valued arbitrary variable over its mean	(3.13)

4.6.2 Gray Level Co-occurrence Matrix (GLCM)

It is a matrix that described as the distribution of co-occurring pixel values (gray-scale values) across the image. GLCM can measure textural features to understand the nature of the information of the image. This method was explained in detail with its features in chapter 3 in section (3.13). GLCM of an image is calculated by using a displacement (d) value of (2) is chosen in order to extract the detailed textural features, with an angle, θ (0) as shown in the algorithm (4.2). Every feature is determined from GLCM for each block four main features are determined for each GLCM according to the equations described in chapter 3 for every feature. The important texture features applied in this stage are explained later. The vector of all features is saved in a dedicated database file, called “tumor”, to be used in the next stage are shown in the algorithm (4.2) which summarized the GLCM steps method.

Algorithm(4.2) Features Extraction based on GLCM method	
Input:	Gray Image (Brain tumor) \\\\ ImG
	(tumor) \\\file to save features.
	GT \\\ gray tone, d=2 \\\ distance between two pixels,
	w \\\ width, h \\\ hight,
Output:	Features vector (feat-disease)
Steps:	Begin
Step (1):	Set the angle direction to zero ,θ=0)
Step (2):	Compute the width value of gray tone, (w=GT.width)
Step (3):	Compute the height value of gray tone, (h=GT.height)
Step (4):	Loop start from zero to gray tone width value, (For i=0 to w)
Step (5):	Loop start from zero to gray tone height value (For j=0 to h)
Step (6):	Set the values of GM to zero's and c equel to zero,(GM(1 to h, 1 to w) = 0, c=0)
Step (7):	Loop start from zero to gray tone height value,(For i=1 to h)
Step (8):	Loop start from zero to gray tone width value -1,(For j=1 to w-1)
Step (9):	Compute the first gray image and put it in Temprory1, (Temp1=ImG(i,j))
Step (10):	Compute the second gray image and put it in Temprory2, (Temp2=ImG (i,j+1))

Step (11):	GM(Temp1,Temp2)= GM(Temp1,Temp2) +1
Step (12):	GM(Temp2,Temp1)= GM(Temp2,Temp1) +1
Step (13):	Increase the c value by 1 , (c = c +1)
Step (14):	End of first loop i
Step (15):	End of second loop j
Step (16):	Compute the probability matrix using Gm divided by c value, (probability_matrix[i][j] = GM/c)
Step (17):	Loop start from zero to gray tone hight value, (For i=0 to h)
Step (18):	Loop start from zero to gray tone width value, (For j=0 to w)
Step (19):	Compute sum of the summation of the probability matrix power 2 (sm=sum (sum(probability_matrix[i][j]) ^2)
Step (20):	End of first loop i
Step (21):	End of second loop j
Step (23):	Extract co-occurrence features from each matrix
Step (24):	Save features vector on “tumor” file
Step (25):	Return Features Vector (“tumor”)
Step (26):	End algorithm

Generating the GLCMs by using the gray co-occurrence properties function, it can obtain some statistics from them, such statistics give detail on an image's texture. Figure (4.6) shows the process of extracting features using the GLCM algorithm with different angles that used in this algorithm until the four features are extracted and saved in the features vector.

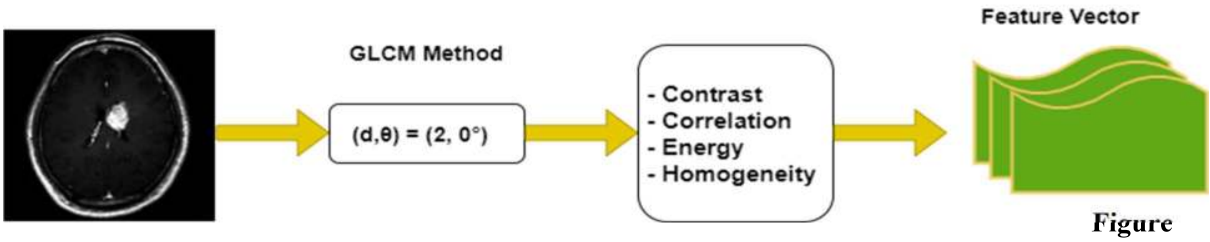

Figure 4.6: Features extraction from GLCM.

The table (4.2) below explains the (second-order features) that were extracted from the GLCM algorithm with the description and range for each of the four features, as well as the equations through which the features were calculated that are mentioned in chapter two.

Table 4.2: Second-order features.

No.	Features	Features describe and its range	Equation numbers in chapter 2
1.	Contrast	Its intensity contrast over the entire image between a pixel and its neighbor. Range = [0 (size(GLCM,1)-1)^2]	(3.14)
2.	Correlation	It estimates the linear reliance of neighboring pixel gray levels. Range = [-1 ... 1]	(3.15)
3.	Energy	It is the summing up of the GLCM squared components. Range = [0 ... 1].	(3.16)
4.	Homogeneity	It's the closeness of the GLCM component distribution to the diagonal GLCM. Range= [0 ... 1].	(3.17)

4.7 DETECTION STAGE USING SVM ALGORITHM

This stage is the last stage of the first proposed approach. The Detection process is the summary of the work through which the decision is made. The SVM algorithm is chosen, as one of the well-known traditional algorithms of supervised learning of machine learning algorithms. This algorithm was explained in detail with its equations in chapter three, section (3.17). The extracted features from the previous stages will be adopted in this stage. In the previous stage, 11 important features were extracted from the brain tumor image. They were saved in a function that is called in the classification process, which is a function (tumor) as mentioned in the previous algorithm (4.3) in this function; the seven features that were extracted before the GLCM method are also stored. The SVM algorithm as explained previously depends on the establishment of the hyperplane in order for the data to be separated, because this work needs more than two classes in the detection process, for this reason, must use the SVM method it was explained in the previous chapter in section (3.17) it is also called binary SVM. The following algorithm (4.3) shows the steps of the SVM Classifier.

Algorithm (4.3): Support vector machine classifier	
Input:	Training _set (Features function (tumor))
Output:	Class name
Step(1):	Begin

Step(2):	Establish the training label for all training sets and identify 2 classes
Step(3):	Specify the kernel function ((polynomial function), as calculated in the previous chapter in the equation (3.21)) that applied to map the training set to kernel space by hyperplane computation, Its calculation was also defined in the previous chapter in equation (3.19)
Step(4):	Segregated data to 1 against all method
Step(5):	Passes testing set on SVM depended on the hyperplane to identify the class name
Step(6):	Return class name
Step(7):	End algorithm

In the above algorithm (4.3), the steps of the algorithm was shown to obtain the training label with the features that have been extracted and saved in the features function on the basis of classification. Also, the kernel function was used, more than one kernel function was tried and the best results obtained using the polynomial function that was used to map the training datasets that are explained and calculated in chapter 3 in section (3.17) in equation (3.21). The data used in this work dividing into two parts for training and testing.

4.7.1 SVM Training

At this stage, the proposed approach has different types of brain tumor to be detected. In the training phase, the images obtained on the SVM algorithm are entered, in order to be classified in any class for any disease. Each SVM binary classifier is trained to utilize a vector of training data, every row associates to features extracted as an investigation from a class. Following the training stage, the SVM model is able to decide the right class for the input features vector.

4.7.2 SVM Testing

The trained SVM model was used as the classifier; the classifier produced independent test dataset was categorized to the classifier's test accuracy. In order to estimate and evaluate the quality of the model training that has been proposed as a specific group is used randomly for this stage which is the testing stage. In this stage, each row of features vector that was previously extracted was categorized and unlabeled here, which is in the training stage are labeled rows. The classified approach was created using the information provided in the training stage and in the features vector. Here, the sample is where each feature is related to a specific column, and a matrix of samples is formed. The sample size must be divided by the same size of the training data because the number of columns represents the number of features.

4.7.3 Detection for Brain Tumor in the First Proposed Approach (Performance Measurement)

In detection, there are several forms online or offline to conduct the classification process. In this work, the Offline form was performed to conduct the detection process, based on the previous steps and on the training dataset, and a user interface was created for the detection process as it will be presented in the next chapter.

4.8 THE SECOND PROPOSED APPROACH USING CNN ALGORITHM

Convolution Neural Network (CNN) is one of the very famous and powerful algorithms in the field of deep learning. It is used in several fields, the most important of which is the classification between several classes. CNN is a class of deep neural networks, most widely used for visual imaging analysis. First, the diagram for the first proposed approach to classify and detect the brain tumor will be displayed in figure (4.1). Then the algorithm (4.4) for CNN training will be showed. In this algorithm, the work of the second proposed approach and the processes that occurred to detect and classify brain tumor were presented starting from the first stage, which is pre-processing, as well as splitting dataset to training and testing, that have been chosen for the ratio It is 70 % for training and 30 % for testing, after testing all ratios, and this was the best ratio as it will be shown in the next chapter, and the network design stage that consists of several layers, also obtaining the best weights for the network after performing several operations in order to obtain the trained network to detect and classify brain tumor.

Algorithm (4.4): CNN training algorithm to classify brain tumor	
Input:	Brain tumor Image with size 256 * 256 (JPG format)
Output:	CNN with optimum weigh
Step(1):	Begin
Step(2):	Pre-processing stage in it the image is resizing from 256 * 256 to 128 * 128
Step(3):	Splitting the data into two parts, 70% for training and 30% for the testing.

Step(4):	CNN design which consisting of several layers: a) Input layer : RGB image with 128 * 128 size b) Convolution layer : Multiple filters were used with size 3*3 c) Nonlinear layer (Activation layer) : Using Rectified linear units (ReLU) d) Pooling layer :Using Max-pooling layer e) Normalize layer : using batch normalization f) Fully connected layer g) Softmax layer
Step(5):	For each pattern in the training dataset: a) Input current pattern (input Image with Label) b) Calculate the real output of the CNN through Softmax layer c) Calculate the error rate by comparing the real output with the desired output. d) Compare the performance goal with the error rate: 1. If the performance goal was not meet, change the connection weights by using the back-propagation learning algorithm. 2. Else, go to next image pattern. e) Stop condition: 1. If the performance goal was meeting with validation data or the maximum iteration was achieved, go to step (6). 2. Else, repeat step (5).
Step(6):	Return the CNN with the optimum weight.
Step(7):	End algorithm

CNN algorithm is explained in detail with all its parts in chapter 3 section (3.19). To facilitate and clarify the explanation more, the second proposed approach has been presented as a chart, as in figure (4.1) present the structure of the second proposed approach stages using the CNN algorithm for brain tumor detection.

The first stage, which is the images acquisition stage, it has been explained previously in the section (4.3), so the second proposed approach will start with the next stage, which is the pre-processing stage.

4.9 IMAGE PRE-PROCESSING

At this stage, the image size is resized before it enters the CNN. The purpose of reducing the size of the images to increase the speed of the training method and calculate the model test realistically. The MRI images are resized from 256 * 256 to 128 * 128 image size. The process of enhancing both input and variables assist improve the speed of the training process. Maintaining data integrity without prejudicing the original MRI image dat. The figure (4.7) is an example of changing the size of images from the proposed approach. The images were converted by using the resizing function of each width and height of the image from 256 * 256 to 128 * 128.

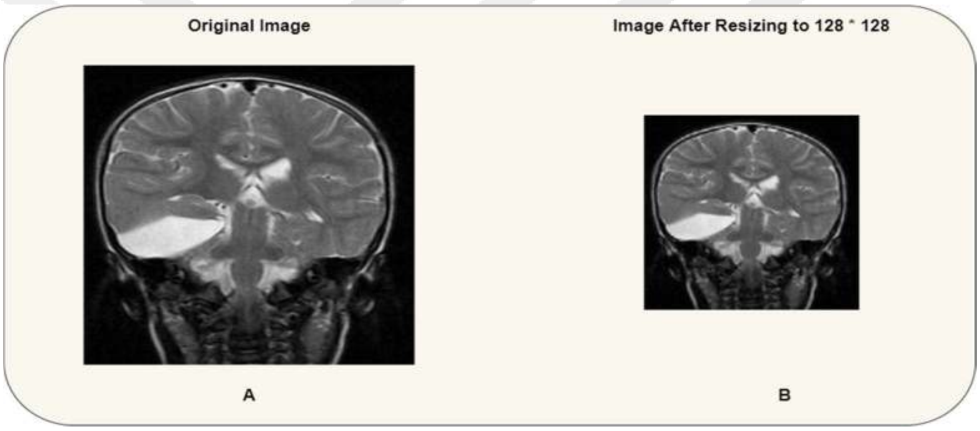

Figure 4.7: Resizing Brain Image.

In Figure (4.7), (a) indicates the original image with a size of 256 * 256, while (b) indicates the image of the tumor of the brain after changing its size to 128 * 128.

4.10DESIGN CONVOLUTION NEURAL NETWORK (CNN) STRUCTURE

The CNN is one of the well-known networks in the field of deep learning and contains many characteristics that distinguish it from the traditional neural network, such as the use of fewer parameters and the number of neurons is less in the CNN, so the training time is less, etc. In designing this network, there are several famous types of designs of the convolutional neural network, such as Alex Net, Google Net, LeNet, etc. So the structure is designing in the figure

(4.8) to suit the proposed approach and after changing many of the parameters and testing it, by choosing this design for the network for obtaining the best results.

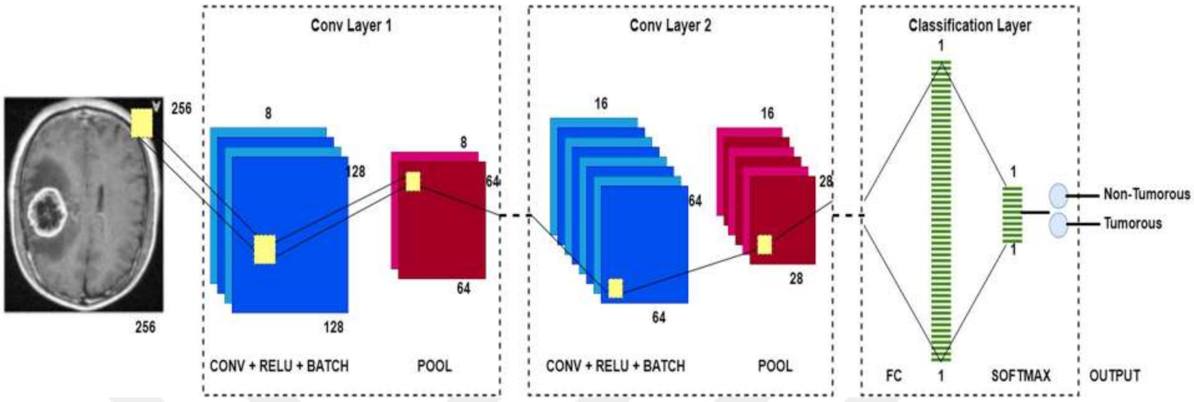

Figure 4.8: Structure of the CNN algorithm.

As shown in the figure (4.8) above, the structure of the CNN algorithm consists of several layers for each layer a specific work and a different structure, the structure is designed as follows:

4.10.1 Input Layer

The stage includes the inputted image as well as their pixel values. These images are the images of brain tumors that entered in matrix form, or in other words layers, whose size has been reduced in the previous stage to size 128 * 128. The input image is an image of the affected brain tumor and the process of converting it into matrixes to be dealt with it in the other stages.

4.10.2 Convolution Layer

Convolution Neural Network (CNN) use various kernels to convert the whole object into convolution layers, also the optimized feature maps to create deferent various of feature maps. Not only does CNN utilize kernel convolution, but it's also a staple of many smart vision algorithms as usual. It's a mechanism that takes a short array of numbers that named a kernel or filter; the filters are the shared weights and bias. These filters work on every section of the image it searching for a similar feature throughout the image.

The values are calculated according to the equation that was calculated in the previous chapter in part (3.19.1.2), which is the equation (3.23). Also as shown in the previous chapter also, that clarifies the work of the convolution layer. In the second proposed approach, four convolution layers are used. In the first convolution layer, 10 filters were used with dimension 3*3 with “the same padding”. The padding implemented to the inputs across the edges means the 'same' padding was established, so the output will be same as the input size. In second convolution, 20 filters were used with dimension 3*3 with "same" padding. In the third convolution layer, 64 filters were used with 3*3 dimensions with padding 'same', and in the fourth convolution layers, and 30 filters were used with 3*3 sizes with the same padding also, as shown in figure (4.8). Choosing the number of filters in each of the convolution layers, based on several experiments that prove the best result obtained of these numbers that were used in each level.

4.10.3 Activation Function

The activation function is implemented to the input of the CNN, and the node activation function describes the node output provided by the input or input set. There are different types of activation functions, in the second proposed approach using the best type after the experiment, which is the most famous type, which is the Rectified Linear Unit (ReLU), which was explained previously in chapter 3, section (3.19.1.3.A). In this function, negative values in the matrix resulting from the previous step are converted to 0 and positive values remain the same. It also showed with its calculation in the previous chapter in the equation (3.24), and its work also has been clarified more in the previous. They are used in the proposed approach after each level of the convolution layers were shown in figure (4.8).

4.10.4 Pooling Layer

Another building block to the CNN is a pooling layer; the purpose of this layer is to gradually decrease the spatial scale of the representation in order to reduce the number of parameters and the network calculation. A pooling layer sometimes follows a convolutionary layer which can be used to depreciate the size of the feature maps. Pooling layers also are invariant through interpretation, such as convolutionary layers, because their computations take that into account neighboring pixels were explained in detail in chapter 3 in section (3.19). Its most commonly used methods are the average pooling and max-pooling layers, in the proposed approach the Max-pooling layer is applied. As explained in the previous chapter in the example was shown previously, after performing the experiments, Max pooling was chosen because it gives better results. It takes the highest value in the matrix specified for pooling and passes the process to all the values of the matrix and therefore has another matrix with fewer dimensions. In the proposed approach 2*2 Max-pooling is used with two strides, the "stride" which means "the step scale to vertically and horizontally traverse the input" is shown in figure (4.9) to explain the Max-pooling layer process.

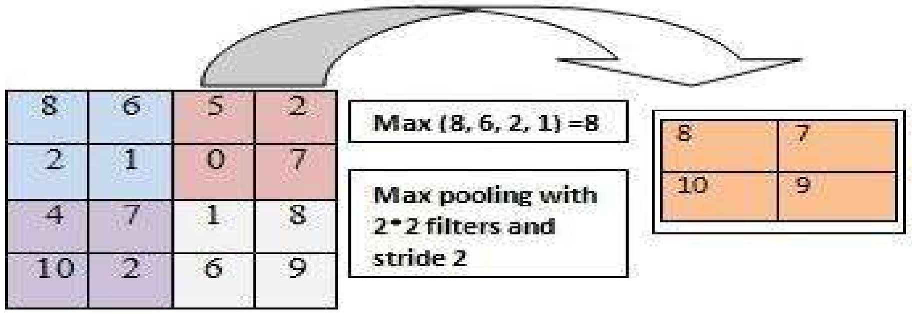

Figure 4.9: Max pooling layer.

4.10.5 Normalization Layer

It is a very important layer that was explained in chapter 3, section (3.18). The batch normalization layer is using in the second proposed approach; it forms norms any channel by means of a mini-batch. This can help diminish sensitivity to variations in the results. It eliminates the fitting problem, as it has a slight effect on regularization. Batch normalization optimizes the yield of the prior activation layer, to calculate the batch normalization, at first calculate the mean and the variance for the layer that was inserted, and then do scaling and shifting to calculate the batch normalization, as shown in the algorithm (4.5).

Algorithm (4.5): Batch Normalization	
Input:	The matrix of the convolution layer, suppose C = (x_1 .... m)

Output:	Batch normalization (Y_i =BN), β(xi)
Step(1):	Begin
Step(2):	Calculate the mini-batch mean () of the layers input by using equation (3.25).
Step(3):	Compute the mini-batch variance () of the layers input by using equation (3.26).
Step(4):	Normalize the layer inputs utilizing the prior calculated batch statistics () by using equation (3.27).
Step(5):	Scale and shift to get output from the layer BN, (xi)
Step(6):	Return Batch Normalization
Step(7):	End algorithm

4.10.6 Fully Connected Layer

Often the last layers of a CNN are fully connected layers, in a traditional neural network; these would behave identically to the layers. The major difference is that the inputs would be in the form that CNN earlier stages would build. In the two neighboring layers, the cells were directly connected to the cells inside the fully connected network. So what this implies is that every one of neurons in the fully connected layers can collect input data components over time that would help it to predict the right class value in the softmax layer afterward are shown in figure (4.8). The objective from this layer is to summarize the weights of the features from the prior layers and show the value of per class, as was explained previously in the previous chapter in part (3.18). In the proposed approach 2 classes that will be produced from this process because according to the data used for training and determining the number of classes that have been extracted from this stage, these classes will be in the form of value for each class linked to the fully connected layer image and Softmax will be made for them in the last step as seen in figure (4.8).

4.10.7 Softmax Layer

The output of the neural network can be difficult to understand. It's typical to end convolution neural network (CNN) with the softmax layer for classification purposes, the result values are between the [0, 1] range which is good because that can prevent binary classification and fit as many classes or measurements in the CNN model that used. The Softmax function equation was mentioned in the previous chapter in part (3.18) in equation (3.28). After extracting the values in a fully connected phase, softmax layer will be assigned in all process that will be done according to the extracted features that have been via the previous layers which the MRI images of brain tumors have passed. Firstly extracting the exponential for the inputs, and then dividing the values by the sum of the exponential values in order to obtain probabilities through which to classify them as shown in the algorithm (3.8). So the greatest probability is the correct class that indicates the disease of the specific brain. According to these probabilities, the disease is classified among 2 classes of brain tumors, as shown in figure (4.8).

Algorithm (4.6): Softmax layer function	
Input:	Fully Connected Layer Values
Zi	// Fully Connected Layers Values
Output:	Softmax probabilities value for the classes of brain tumor
Step(1):	Begin
Step(2):	Calculate the exponential for every input in fully connected layer
Step(3):	Calculate the exponential summation for the class of input fully connected layer
Step(4):	Calculate the softmax function() by using equation (3.28) for all classes after calculating the exponential for each class in step(2), and divide each of them by the sum of the 2 classes after calculating the exponential for them in step(3), to predicate a true class that has the highest probability.
Step(5):	Return softmax probabilities value for the classes of brain tumor (yi)
Step(6):	End algorithm

4.11 CNN NETWORK TRAINING

Network training is a process for obtaining kernels in convolution layers and weights in fully connected layers, which reduces the differences in training data set between output predictions and defined area truth tables. As explained in the previous chapter in section (3.18.2), this stage is one of the most important stages in CNN because learning comes through this stage, the network that has been built will learn by extracting features from the MRI images of brain tumors in order to learn from these features for each image to be distinguished on its basis. In our proposed approach also used the pre-trained network method, the network is trained and the training is saved for use later, which saves time and effort and not to be re-trained again. In the training stage, the data are labeled with each type of brain tumor, and training is conducted on the foundation of these labeled, in respect of learning and training the network. The training stage, need three things for training, which is the training set that is obtained from the dataset, the layers in which the network was built, and the different training options that were made for training. Also, this structure has a very important function for evaluating the training process, which is the Loss function, which will also be used.

4.11.1 Training Set

The data is divided into two main sections, which are training data and testing data. The data is dividing using a function in the Matlab called "Split Each Label", this function divides the data according to the percentages determined by the user and randomly. It was randomly divided and not chained to obtain a better approach and data is randomly taken from the dataset to make detection later better and stronger. In the proposed approach the data was divided so that the training data would be 70% and the testing data is 30%, after testing all ratios, and this best ratio as it will be shown in the next chapter.

4.11.2 Layers (designing)

The layers are intended as the layers that were built in the construction of the CNN structure in the training stage, all of these layers that mentioned earlier will also be passed in order to extract the features and learn from them to be classified through these features. Training for all data for training will be passed through the layers that built.

4.11.3 Training Option (training algorithm)

In order for the training process to be completed, that need multiple options that were used in this process using Matlab program by parameters that are created in the training process these options are:

1. Stochastic Gradient Descent with Momentum (SGDM)

The method is utilized in the data training process, due to its good properties that fulfill the purpose of training. It is an optimization method utilized to train CNN and machine learning systems. SGDM is a tool that helps move vectors of gradients in true directions and thus contributes to faster convergence. It is one of the most famous optimization algorithms and it is used to train several states of the art systems. They were explained in the previous chapter in section (3.19) and illustrated in equations (3.29) and (3.30).

2. Max Epoch

The "Epoch" is a metric of how many times all training vectors are once utilized to update the weights. Concurrently in one epoch in the learning algorithm, before weights are upgraded. The maximum number of epochs that used for training is 2 epochs. In addition, the training dataset Divided into two epochs, for each epoch, 112 iteration and 224 iterations were divided into two epochs. As in the following figure (3.11), which illustrates the training process, this image was taken from the proposed approach during network training upon execution.

3. Shuffle

The data is shuffled to have various data per batch, this command is the training data is shuffled. There are many options are: "never" If no shuffling is implemented if the option is "once" the data will be shuffled only once in the training. If the option is "every-epoch" the dataset will be shuffled before each training epoch. In the second proposed approach, "every-epoch" is used for the shuffling option.

4. Plots

Plots to present when training, there are two options "training-progress" or "none" in (default); in the proposed approach was used "training progress" for plot option to see the progress of training.

5. Verbose

It is one of the training options and has two options, "True" or "False". If this is set to "true", the command window will be printed with the data on the training progress. If the option is "false" then command window will not be printed with the data on the training progress the default for this is "true", but in the proposed approach use the "false" option.

6. Validation Data

It's the information to use throughout the training for validation. This may be an Image data store with categorical labels, a Mini-Batch data store with specified responses, and a table where unless image paths or images are in the first column, a cell array{X, Y} where X is a numerical array with the input data and Y is a return array. In the proposed approach, this option was used (testing set, the dataset for the testing set. labels) for validation data. The testing set means is the group that was used for the later testing process, while the dataset is a set of datasets that have taken as a label form.

7. Validation Patience

The number of times the validation loss can be greater than or equal to the initially smallest loss before the network training is finished. There are two choices either an "Inf" choice or a "positive" choice. The definition of the "Inf" choice is the automatic ending of training the network, and this choice is the default choice for training the network. The other choice, which is positive, means specifying a "positive" integer for stop network training. In the proposed approach that uses the "Inf" choice.

4.11.4 Loss Function

A loss function, also known as a cost function, can be calculated the consistency between the network's output estimates by forwarding propagation and assigned area truth labels, as shown in figure (3.11). It helps in the optimization of CNN parameters. The important purpose here is to reduce a convolution neural network's loss by optimizing its parameters (weights). The loss is measured utilizing loss function, by a convolution neural network combining the target (actual) result and the predicted value with errors. As explained in the previous chapter in section (3.20.4), also the back-propagation that explained in the previous chapter in section (3.20.5), the loss function and accuracy function was also calculated and displayed by using the confusion matrix as it will be presented in next chapter.

4.12 CNN NETWORK TESTING

Testing is the process to test the used dataset to supply an equitable final evaluation of the trained dataset, which have been trained at CNN, and the feature that is extracted by learning neural network. The dataset also used that was allocated to the testing process when it was divided, based on these datasets and the training data for testing; the classification process for a brain tumor is done. Before the testing stage, there is the training stage; this means that the network is trained when inserting an image of a brain tumor. The disease will be determined because the network was previously learned and trained. So the main difference between training and testing is test data that are unlabeled, unlike training data that is labeled. In the proposed approach, uses this function in Matlab, CLASS = classify (Sample, Training Group), this categorizes every row of the dataset to one of the training groups. Sample and training must be arrays with the same column size. Training the Group is a group variable and the specific values determine groups, and each component determines the group to which the associated training row belongs. Here we have reached the last, which is the classification and detection of brain tumors.

4.13 BRAIN TUMOR DETECTION IN THE SECOND PROPOSED APPROACH (PERFORMANCE MEASUREMENT)

In this stage, the offline model was applied to conduct the detection process, based on the previous steps and on the trained network, and a user interface was created for the detection process for the brain tumor as it will be presented in the next chapter.

5.IMPLEMENTATION AND EXPERIMENTAL RESULTS

5.1 INTRODUCTION

The implementation and experimental results of the proposed approach were presented and described. This proposed approach "detection of brain tumor using image processing and deep learning techniques" is divided into two different sub approaches according to the used algorithms. The first proposed approach that was previously explained consists of several stages; the outcomes of these stages will be presented with the final results of classification using the SVM algorithm, in addition to the results of the second approach which is based on CNN algorithm. The results of these two approaches will be presented and a comparison will be made between the results of the two approaches. Moreover, a comparison with related works will be presented. The same dataset was used in the two sub approaches.

5.2 IMPLEMENTATION ENVIRONMENT

The proposed approach is implemented and evaluated by using MatlabR2018b and the proposed approach performed on the windows10 O.S, and the platform Intel Core i5, RAM 4 GB.

5.3 DATASET ACQUISITION (BRAIN TUMOR IMAGES)

The collected data consists of 3000 images related to different tumors which includes the number of images of these brain tumor. A small portion of these images that shown in figure (5.1), which includes a sample of images for the different types of brain images.

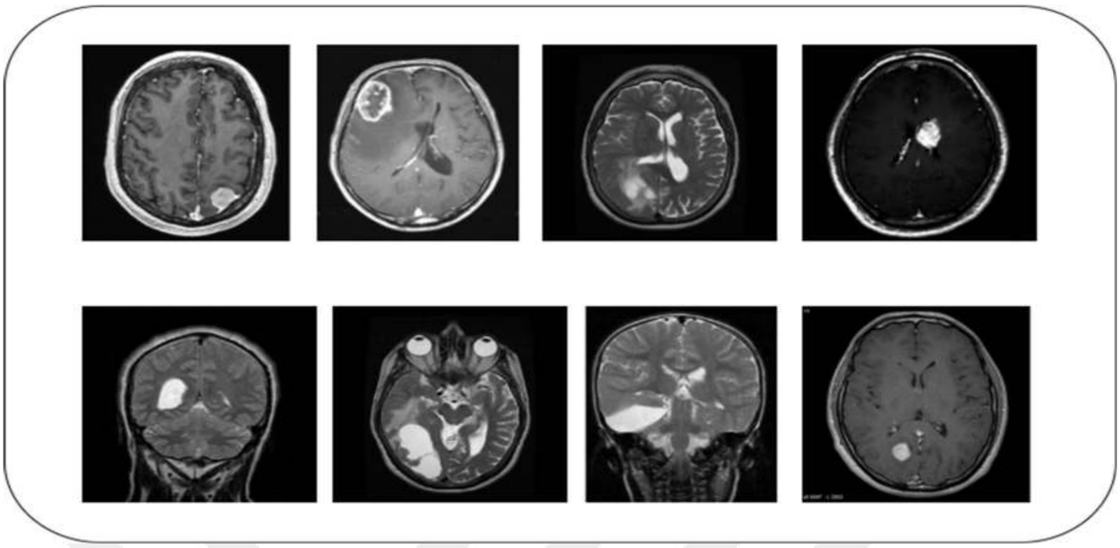

Figure 5.1: a sample of MRI images for the different types of brain images.

5.4 EVALUATION OF FIRST PROPOSED APPROACH

The first proposed approach that was explained in the fourth chapter, which consists of several stages, each stage and its results, up to the final results will be shown, it the classification stage with the SVM algorithm. The first stage is the pre-processing stage.

5.4.1 Result of Image Pre-Processing

At this stage, the images will be converted gray colour space and enhanced with different filter to reduce the noise and stain.

5.4.2 Result of Image Segmentation

At this stage, the K-means clustering algorithm was applied according to algorithm (3.2) in chapter three. Figure (5.2) shows a sample of some images from the proposed approach after applying the image segmentation process, where, (a) shows the original image of the affected brain , and (b) shows the image for which segmentation was made.

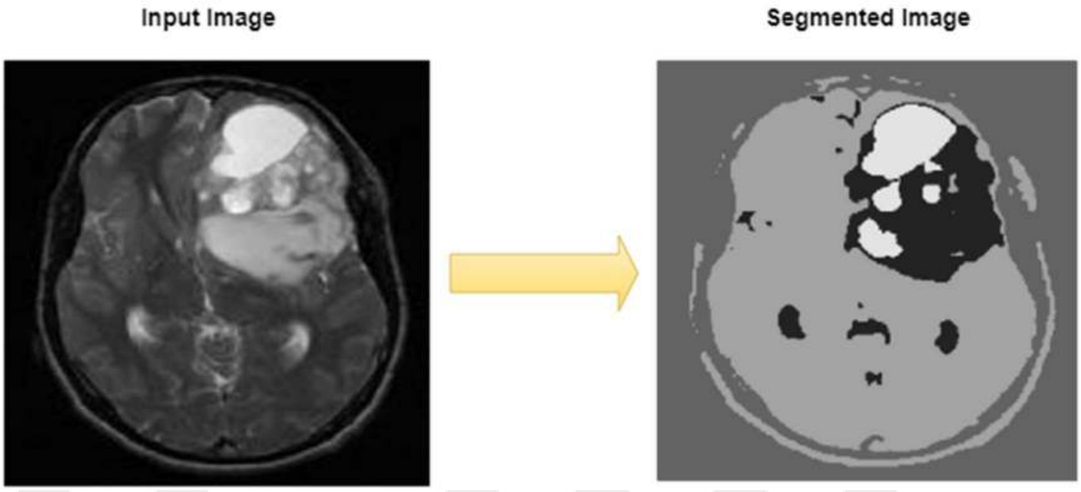

Figure 5.2: The images segmentation from the proposed approach.

5.4.3 Result of Feature Extraction

Processes continue after the image segmentation process, the process of extracting features that have been explained in the previous chapter in the first proposed approach comes with a feature that has been extracted. Seven features were extracted directly from the segmented images, while the remaining four features were extracted using the GLCM method after conversion to grayscale. Since having the number of images in the dataset is (3000), then each of these images has 11 different features, because of the data size is very large here will show a sample of these features, so in the table (5.1) which displays a samples from the proposed approach of these features were taken for 11 images randomly and each image 11 features as explained previously.

Table 5.1: The samples of features extracted from 11 random images.

NO:	SD	Entropy	RMS	Variance	Smoothness	Kurtosis	Skewness	IDM	Major Axis	Solidity	Circularity

0.904427571	0.932308843	0.973496408	0.938257817	0.927192281	0.923164245	0.832744917	0.971544129	1.05758666
0.917910448	0.960659898	0.964620187	0.950688073	0.944309927	0.934693878	0.932432432	0.961538462	0.977040816
24.14579	33.77568	37.04686	34.67206	24.76707	20.19143	21.08148	29.33037	23.05961
0.192732	0.487346	0.837045	0.641188	-0.23875	-0.48315	0.837354	-0.78879	0.622794
0.684572	0.750373	0.584274	0.796349	0.581913	0.742762	0.389772	0.815128	0.441837
4.093818	4.454436	4.69758	5.10555	3.483895	5.102073	3.888138	4.899505	3.479159
0.89201448	0.85146681	0.87671632	0.8971146	0.81927427	0.79720833	0.90039704	0.86560842	0.90189401
0.026812	0.024129	0.024429	0.024365	0.026596	0.029856	0.026626	0.02717	0.030026
0.166667	0.158114	0.158114	0.158114	0.166667	0.176777	0.166667	0.166667	0.176777
3.997665	3.56703	4.205701	4.005381	3.754407	3.661866	4.391248	3.85025	4.232195
0.164466	0.157342	0.157093	0.156465	0.166023	0.176297	0.163084	0.165447	0.17343
1	2	3	4	5	6	7	8	9

10	0.138141	3.88004	0.138675	0.018823	0.89289133	7.699749	0.969475	0.083646	59.13905	0.897219464	0.771670933
11	0.143455	3.987824	0.144338	0.02049	0.90744713	5.632477	0.86596	0.085668	59.72602	0.9214223	0.845336145

Figure (5.3) shows example from the proposed approach, where eleven features were extracted from a random image after conducting operations on them, according to the equations that mentioned in the previous chapters. These features will be used later in classification process.

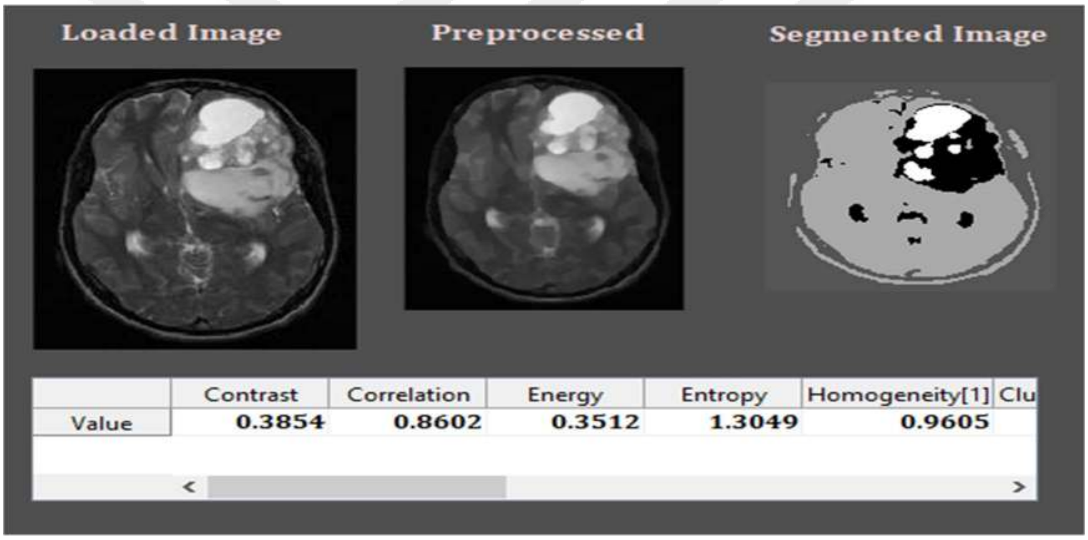

Figure 5.3: Example of features extraction from a random image.

5.4.4 Result of Training and Testing Using SVM Algorithm

After the process of extracting features, the approach reached the classification stage. In this stage, the brain tumor are classified according to the features that were extracted, and as the SVM algorithm was previously explained in the algorithm (3.5). In order to complete the classification process, the approach needs two stages that have been explained previously, which are the training stage and the testing stage. The results of these two stages will be displayed in order to then detect and classify the brain tumor. The confusion matrix is used to display the accuracy of the training and testing results in each class as well as for the accuracy rate of all classes. The training results will be displayed first and then the test results.

A-Result of SVM Training

This operation was explained previously, where the training process for images is done in an SVM method after performing the operations and to detect the disease in the testing stage later. The display of the results using the confusion matrix as in figure (5.5), which shows the accuracy of the training results for each class, and also displays at the top of the matrix, the average cumulative accuracy of all classes at the training level is equal to (69.33). The overall accuracy and accuracy of each training class was determined according to that was mentioned in the second chapter in section (2.19) and according to the equations (2.27), (2.28) and (2.29).

Thus, the process of calculating the accuracy of all classes was done, as in figure (5.5), which shows the confusion matrix of the training process, to clarify the numbers that are in figure more, they will be shown in the comparison table (5.4), it will be presented later.

B- Result of SVM Testing

This stage is an examination of the approach by testing it with the remainder of the data that is not labeled in order to classify the brain tumor as explained in the previous chapter. At this stage also the results were presented in the form of a confusion matrix that shows the accuracy of each class in the testing stage also, the average overall accuracy of all classes is equal to (01.79), it is calculated in the same way as calculating the total accuracy of the training stage.

It should be note that the main reasons for low accuracy here are: the first reason is a large number of classes, there are deferent classes in the proposed approach, usually the SVM algorithm is ideal is when having two classes but when to use more than two classes, must use a SVM. After research on this topic, it was found that the accuracy is inversely proportional to the number of classes in the SVM algorithm, and this is one of the disadvantages of this algorithm. Also because using non-linear SVM, the number of classes is more than two because of this, it will be a great overlap in the data as have observed in the Confusion matrix, in the training, also in testing stages, in addition, the large data used in this approach also from these reasons, which led to an exacerbation of accuracy in the training and testing stages. This is the reasons that led to the search for another algorithm to detect brain tumor with greater accuracy. In order to prove these reasons that led to a lack of accuracy in the training and testing stages, the approach was trained and testing on two classes only for classification and high accuracy was obtained in the training and testing stages. These results were presented to show the reason for the lack of accuracy in relation to the first proposed approach, and how accuracy increased when only two classes are classified and smaller dataset, as in figure (5.6). In the figure (a) it displays the confusion Matrix, the accuracy rates was obtained in training (100%) and in the testing (100%), and in (b) the two confusion matrices were displayed and a high accuracy rate was obtained also in training (100%) and in the testing (90.32%).

Before performing the detection process, it should be noted here that an important note is that the kernel function that is relied upon is (polynomial function) due to the experience of the rest of the kernel functions and the best results were obtained with this function for classification the classes with the highest degree of accuracy in the preparation and testing phases as in the table (5.2), which shows a comparison between the types of kernel functions that were used in terms of the overall accuracy in training and testing.

Table 5.2: Comparison of different kernel functions for the overall accuracy rate.

No.	Kernel functions	The overall accuracy rate for all classes in training %	The overall accuracy rate for all classes in testing %
1.	Polynomial function	69.33%	61.79%
2.	Radial basis function (RBF)	63.22 %	54.53 %
3.	Linear function	43.68%	42.10 %

5.4.5 Brain tumor Detection in the First Proposed Approach (Using SVM Algorithm)

Following conducting the training and testing processes of the approach, the process of detection of brain tumor is carried out. The detection mechanism in the first proposed approach using the SVM algorithm for classification and after conducting all the previous operations that were mentioned, this disease is detected and classified, but the accuracy in detection is not high due to the lack of accuracy of the test and training as well, and the reasons have been explained, and for this reason, a second proposed approach has been built that is better than the first approach. Here, interfaces will be displayed from the first proposed approach that shows the detection process for brain tumor, in the form of stages, each of these stages has its own procedure, in order to facilitate the use of the approach for the user, as shown in the figure (5.4) which shows the approach for brain tumor detection using the SVM algorithm with all the stages shown to perform the classification process, from the image load button to the detection button.

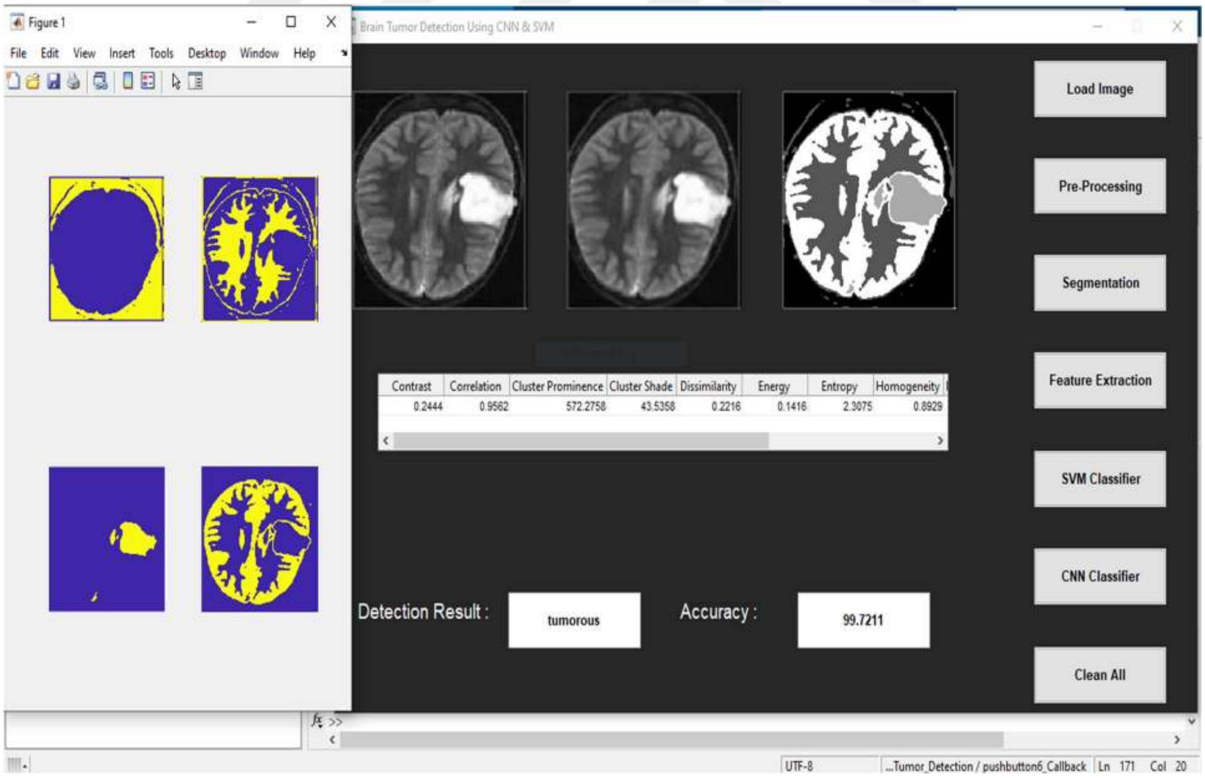

Figure 5.4: The interface of the proposed approach (brain tumor detection using SVM & CNN).

5.5 EVALUATION OF THE SECOND PROPOSED APPROACH

The lack of accuracy and the weak results that obtained in the first proposed approach made our seek for a second approach and be better than the first proposed approach, and this actually happened, here the results obtained from this approach will be presented through deep learning techniques, using the CNN algorithm. As this approach was explained in the previous chapter, which consists of multiple stages, the difference in this approach is that the convolutional neural network is that performs these stages, it means the process of extracting features until the classification process. That means just inserting MRI images of brain tumor on the network, and in the outputs, this is revealed and classified tumor that is why it was mentioned in this approach that there are only the end results obtained, there are no results for each stage as in the first proposed approach.

The first stage is the process of resizing images from 256 * 256 to 128 *128. Then these pictures enter the network in two stages, which are the training stage and the testing stage. As in figure (5.4) which shows an interface from the second proposed approach that clarifies the process of entering the dataset with its division into the testing and training stages, in order to complete the training and testing operations, after which the trained network is saved so that it can be used in the next stage, which is the stage of detection of brain tumor. The trained network is saved in order not to repeat the training process at each detection stage, thus saving time for each training process, the flowing figure (5.5) shows the CNN network design

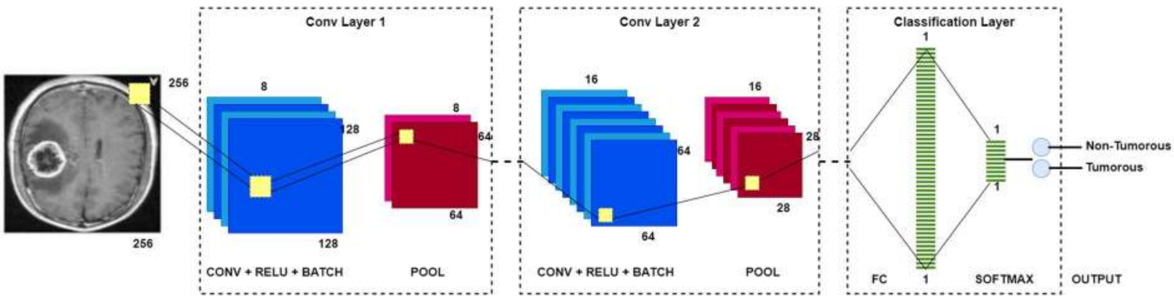

Figure 5.5: the CNN network design.

5.5.1 Result of the CNN Training

At this stage, the network training process was conducted to learn about the features of each disease and detect it at the detection stage. When the network is trained and all images pass on the CNN of all layers in order to teach this network, this is the main purpose of the training process. The main problem that occurred during the training is that it takes a long time, and also the speed of the computer and its properties have a big role in the time spent on the training the network. This stage was explained previously; here the results of the training process will be presented. As well as, the overall accuracy rate is calculated by Summarizing the accuracy of the class is seen in the confusion matrix and dividing it by its number and its result is (98.29%).

And assigned area truth labels by using the loss function that has been explained in the previous two chapters. In figure (5.6) a chart will be displayed the loss function and the accuracy. The loss function expressed in the red line, which is descending to the bottom, in contrast to the accuracy expressed in the blue line that goes from the bottom to the up. In addition, the number of epochs was used in the training process is 12 epochs, as for the number of iterations is 5040 For each epoch 420 iterations.

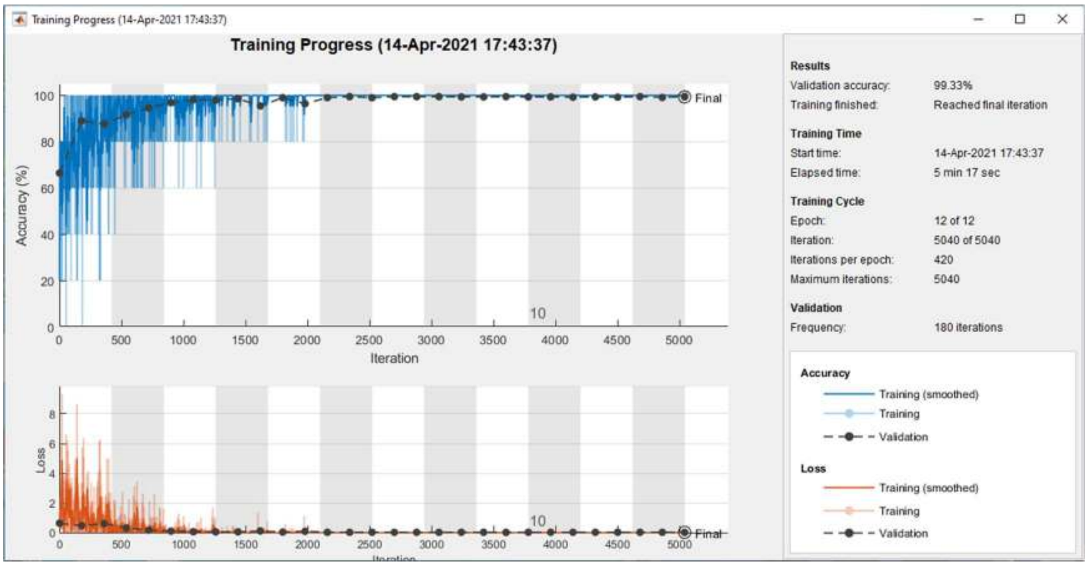

Figure 5.6: The Training, accuracy and Loss function progress.

5.5.2 Result of the CNN Testing

In this stage, the results of the testing were displayed. This process was explained in the previous chapter and how the convolutional neural network is tested in order to be used for the process of detection of brain tumor. In the testing stage, excellent results were obtained and presented as a confusion matrix, which shows the results of the accuracy of each class of the classes in the testing stage, to clarify the numbers that are in figure more, they will be shown in the comparison table (5.3), it will be presented later. An example from the proposed approach for calculating the accuracy of the first class in this matrix, as well as how the overall accuracy rate was calculated for all classes in the testing stage.

Thus the accuracy was calculated for the rest of the classes, Also the overall accuracy of the test process (98.29) which is at the top of the confusion matrix.

An important note that must be mentioned, the process of dividing the dataset, which was relied upon in our work, is 70% for the training dataset and 30% for test dataset. This division process is not arbitrary, but all data rates for training and testing have been tried and start with 50% only for training data collection and 50% for testing, until 90% of the training data versus 10% of the testing data. The highest percentage and the best result obtained were the 70% for training and 30% for the set of the testing dataset, especially for the testing process that depends on it in the classification and detection process of brain tumor, are shown in table (5.3) which illustrates these percentages versus total accuracy training data and test data.

Table 5.3: Comparison of the different ratio of the size of data in training and testing for the overall accuracy rate.

No.	Percentage of training dataset and testing dataset %	The overall accuracy rate for all classes in training %	The overall accuracy rate for all classes in testing %
1.	50% training _ 50% testing	99.72%	91.81%
2.	60% training _ 40% testing	99.71 %	94.32 %
3.	70% training _ 30% testing	98.29 %	98.2 %
4.	80% training _ 20% testing	99.91 %	94.25 %

5.	90% training _ 10% testing	99.95 %	95.51 %

5.5.3 Brain tumor Detection in the Second Proposed Approach (Using A CNN Algorithm)

The last stage in the second proposed approach is the detection stage of brain tumor, which is the most important stage of the approach. When the network has been trained and tested so that any image that is chosen will be classified according to the different classes of tumor , furthermore after obtaining high and excellent results in the training and testing, thus detection accuracy will be very high and the errors are almost non-existent or very rare.

In the figure (5.4) that showed interface of the second proposed approach used the same image that was used in the first proposed approach, which the brain tumor infected with disease that was detected and classified using the convolutional neural network algorithm, and selecting the classifier network that previously trained, tested and saved and when pressing the detection button, it will be classified the disease, and also the time taken for the detection process is shown, in our example it reached (0.062s).

5.6 COMPARISON, BETWEEN THE FIRST PROPOSED APPROACH WITH THE SECOND PROPOSED APPROACH

The comparison between the two proposed approaches is very important in order to show the strengths and weaknesses of each of them. Initially, when choosing this topic which is the detection of brain tumor, at first, the SVM algorithm was used in the first proposed approach, but after getting the results of accuracy is not high, then searched for an alternative approach is better than the first proposed approach in order to obtain a high detection accuracy and very few errors, unlike the first proposed approach. After experiments and research, the CNN algorithm was chosen in the second proposed approach for classification the brain tumor. It is worth noting that the same dataset were used, as well as the same data division for training and testing, which is 70% for training and 30% for testing for both algorithms. After showing the results of the two approaches at all stages and found that the best is the second proposed approach for the following reasons:

1- The high accuracy obtained from the second proposed method in the training and testing phases for each of the classes, as well as the average accuracy rate in these two stages as shown in Table (5.4) that shows the accuracy of all classes and overall accuracy in the two proposed approaches.

2- The time taken for the training process in the first proposed approach using the SVM algorithm is more than the time taken in the second proposed approach that used the CNN algorithm because the first proposed approach needs a long additional time to extract the features from the images. This is also an advantage of the CNN algorithm which extracts its features internally without the need for that extra time.

5.7 PROPOSED APPROACH VS. RELATED WORKS

After obtained high accuracy results on the large data of brain tumor images, and different classes that classified, a comparison was made between the second proposed approach and the work related to it, which showed the efficiency and high accuracy of the proposed approach. As the table (5.4) shows the comparison between the first and second proposed approach with some other work related to ours. It shows the high accuracy with the largest numbers of data for the proposed approach, it also illustrates the weaknesses of the related works.

Table 5.4: Comparison between the second proposed approach VS related work.

Author(s), Year	Ref. No.	Algorithm	Accuracy
Ismael et al, 2018	[75]	CNN	91.9%
Pei et al, 2020	[76]	CNN	94.39%
Pashaei et al, 2018	[77]	CNN	93.68%
Seetha & Raja, 2018	[78]	CNN	97.5%
The first proposed approach	-	SVM	68. 9%
The second proposed approach	-	**CNN**	**98. 29%**

6. CONCLUSIONS AND FUTURE WORKS

6.1 CONCLUSIONS

In this chapter, the proposed approach is summarized; the following conclusions were taken from collection of test results. Some of those conclusions are listed in the following:

1- The proposed approach is an automated method to detection and classification brain tumor with high accurate, fastest results based on computer modern technologies.

2- This work is conducted to obtain the results using two different proposed approaches each one used different algorithm for classification and make a comparison between them. The first proposed approach use support vector machine (SVM) algorithm and the second proposed approach use a Convolutional Neural Network (CNN) algorithm.

3- The same dataset is used in the two proposed approaches and the data is divided in the same percentage also, which was 30% for the testing and 70% for training; this was the best ratio after trying all percentages as in the table (5.2).

4- The results of the comparison showed the preference of the second proposed approach which is based on deep learning field using CNN algorithm, in terms of performance and accuracy. The obtained accuracy in testing phase for the second proposed approach was (98,29%), while the accuracy for the first proposed approach was (68. 9%).

In addition, the comparison explained the main reason for low accuracy in the first proposed approach, which is related to the number of classifiers that have been classified and also the large size of dataset. This has been proven by applying the SVM algorithm to classifying only two classes and on a smaller number of the dataset, which obtained a very high accuracy of (98%) in training and testing phases for tumor infected and uninfected as shown in figure (5.5). At the conclusion of the comparison, it is proven that the second proposed approach is the best, the fastest, most accurate and powerful approach for detecting and classifying brain tumor.

6.2 FUTURE WORKS

The proposed approach brain tumor detection is flexible approach and there are many suggestions that can be performed for future work as follows:

1- Improve the accuracy results, the system has to take into account the feedback from the user.

2- Design an expert system that can detect brain tumor in automated way depending on multiple techniques and presents ways to prevent these diseases and treatment.

3- Using various machine learning techniques such as (ResNet, AlexNet, GoogleNet, etc...) with comparison between them, also other optimizers for the proposed approach to obtain a better accuracy rate.

REFERENCES

[1]I. Country-specific, N. Method, and M. Country-specific, “Iraq source:Globocan 2018,” *Int. Agency resarch cancer/WHO.*, vol. 141, pp. 2018–2019, 2019.

[2]A. Yardimci, “Applications of soft computing to medical problems,” *ISDA 2009 - 9th Int. Conf. Intell. Syst. Des. Appl.*, pp. 614–619, 2009, doi: 10.1109/ISDA.2009.168.

[3]Deepika Choudhary, Anjali Lather and Sandeep Yadav, “Overview of Image Processing”, Computer Science and Engineering, Vol.2, PP 115-119, October 2014.

[4]Dhingra, Gittaly, Vinay Kumar, and Hem Dutt Joshi, “Study of digital image processing techniques for leaf disease detection”, Multimedia Tools and Applications 77, No.15, PP 19951-20000, 2018.

[5]Zawish, M., Siyal, A. A., Ahmed, K., Khalil, A., & Memon, S. (2019). Brain tumor segmentation in MRI images using Chan-vese technique in MATLAB. 2018 International Conference on Computing, Electronic and Electrical Engineering, ICE Cube 2018, 1–6. https://doi.org/10.1109/ICECUBE.2018.8610987.

[6]Pushpa, B. R., & Louies, F. (2019). Detection and classification of brain tumor using machine learning approaches. In *International Journal of Research in Pharmaceutical Sciences* (Vol. 10, Issue 3, pp. 2153–2162). https://doi.org/10.26452/ijrps.v10i3.1442

[7]Hazra, A., Dey, A., Gupta, S. K., & Ansari, M. A. (2018). Brain tumor detection based on segmentation using MATLAB. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017, 425–430. https://doi.org/10.1109/ICECDS.2017.8390202.

[8]Ramana, M. V., Sreenivasa Reddy, E., & Satayanarayana, C. H. (2018). Enhanced curvlet transform based artificial neural network for brain tumor diagnosis. *International Journal of Engineering and Advanced Technology*, *8*(2), 245–252.

[9]Ahuja, N., A. Rosenfeld and R.M. Haralick, 1998. "Neighbor gray levels as features in pixel classification", Pattern Recognition, vol. 12, pp. 251-260.

[10]F. E. Ali, I. M. El-Dokany, A. A. Saad, F. E. Abd El-Samie, “Curvelet fusion of MR and CT images,” Progress in Electromagnitics Research, Vol. 3, pp. 215-224, 2008.

[11]M. Eltoukhy, I. Faye, B. Samir, “Using curvelet transform to detect breast cancer in digital mammogram,” 5th International Colloquium on Signal Processing & Its Applications (CSPA), 2009.

[12]Johnson, J. Alex Becker, “The whole brain: Atlas,” Harvard medical school, Second ed., 1999.

[13]Shen, Liran, and Qingbo Yin. “Texture classification using curvelet transform,” Proceedings of the International Symposium on Information Processing. 2009.

[14]Deepika Choudhary, Anjali Lather and Sandeep Yadav, “Overview of Image Processing”, Computer Science and Engineering, Vol.2, PP 115-119, October 2014.

[15]J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja. “Multiclass Brain Tumor Classification Using GA-SVM,” In Proceedings of the 2011 Developments in E-systems Engineering (DESE '11). IEEE Computer Society, Washington, DC, USA, 182-187, 2011.

[16]K. Deepa , K.S. Gokul, K. Hinduja, R. Rajkumar, “An efficient approach to predict tumor in Brain image using classification techniques,” Information Communication and Embedded Systems (ICICES), 2013 International Conference on , vol., no., pp.559,564, 21-22 Feb. 2013

[17]Joshi, Dipali M., N. K. Rana, and V. M. Misra. “Classification of brain cancer using artificial neural network,” Electronic Computer Technology (ICECT), 2010 International Conference on. IEEE, 2010

[18]R. Gonzalez, R. Woods and S. Eddins, “Digital Image Processing Using MATLAB” 2nd ed., Tata McGraw Hili New Delhi, 2010.

[19]E.I papageorgiou, P.P Spyridonos, “ Brain Tumor characterization using the soft computing technique of fuzzy maps”Applied soft computing 2008, vol 8,pp 820-828
[20]H. Rajini and Bhavani, “Automatic MR Brain Tumor Detection using Possibilistic C-Means and K-Means Clustering with Color Segmentation,” International Journal of Computer Applications 56(2):11-17, October 2012.

[21]Bhardwaj, Anant, and Kapil Kumar Siddhu. “An Approach to Medical Image Classification Using Neuro Fuzzy Logic and ANFIS Classifier,” International Journal of Computer Trends and Technology- volume4Issue3- 2013.

[22]Zhang, S. Wang, and L. Wu, “A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO,” Progress In Electromagnetics Research, Vol. 109, 325{343, 2010.

[23]Rathore, Vivek Singh, Messala Sudhir Kumar, and Ashwini Verma. “Colour based image segmentation using L* a* b* colour space based on genetic algorithm.” International Journal of Emerging Technology and Advanced Engineering 2, No. 6, PP156-162, 2012.

[24]Dr.S.Kannan,Vairaprakash Gurusamy & Subbu Kannan, “REVIEW ON IMAGE SEGMENTATION TECHNIQUES” Madurai Kamaraj University Conference Paper, October 2014.

[25]Runge V M, (2002). "Clinical MRI". W.B. Saunders. Philadelphia.

[26]R. Gonzalez and R. Woods, “Digital image processing” 2nd ed. upper saddle river new jersey: prentice hall.
[27]R. Gonzalez, R. Woods and S. Eddins, “Digital Image Processing Using MATLAB” 2nd ed., Tata McGraw Hili New Delhi, 2010
[28]Khan, Shehroz S., and Amir Ahmad, “Cluster center initialization algorithm for K-means clustering,” Pattern recognition letters 25, No. 11, PP 1293-1302, 2004.

[29]Dhanachandra, Nameirakpam, Khumanthem Manglem, and Yambem Jina Chanu, “Image segmentation using K-means clustering algorithm and subtractive clustering algorithm”, *Procedia Computer Science* 54, PP 764-771, 2015.

[30]K. Rohini and G. Suseendran, “Aggregated K Means Clustering and Decision Tree Algorithm for Spirometry Data”, Indian Journal of Science and Technology, Vol 9(44), November 2016.

[31]Chandhok, Chinki, Soni Chaturvedi, and A. A. Khurshid, “An approach to image segmentation using K-means clustering algorithm”, International Journal of Information Technology (IJIT) 1, No. 1, PP 11-17, 2012.

[32]Betke, Margrit, and Nicholas C. Makris, “Fast object recognition in noisy images using simulated annealing”, In Proceedings of IEEE International Conference on Computer Vision, PP 523-530. IEEE, 1995.

[33]Saravanan, C, “Color image to grayscale image conversion”, In 2010 Second International Conference on Computer Engineering and Applications, Vol. 2, PP 196-199, IEEE, 2010.

[34]S. T. Rasheed, “Counting and Classification Moving Objects in Digital Video”, M.Sc thesis, University of Technology, Department of Computer Sciences, 2017.

[35]Kumar, Gaurav, and Pradeep Kumar Bhatia, “A detailed review of feature extraction in image processing systems”, In 2014 Fourth international conference on advanced computing & communication technologies, PP 5-12, IEEE, 2014.

[36]Duanmu, Xiaoyin. "Image retrieval using color moment invariant." In *2010 Seventh International Conference on Information Technology: New Generations*, pp. 200-203. IEEE, 2010.

[37]Pradeep, N., H. Girisha, and K. Karibasappa, “Segmentation and feature extraction of tumors from digital mammograms”, Computer Engineering and Intelligent Systems 3, No. 4, PP 37-46, 2012.

[38]Aggarwal, Namita, and R. K. Agrawal, “First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images”, Journal of Signal and Information Processing, Vol 3, PP 146-153, 2012.

[39]Er. Kanchan Sharma, Er. Priyanka , Er. Aditi Kalsh, and Er.Kulbeer Saini, “GLCM and its Features” International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Vol 4, Issue 8, August 2015.

[40]Azawi, Raghad Majeed, Dhahir Abdulhade Abdulah, Jamal Mustafa Abbas, and Ibrahim Tareq Ibrahim. "Brain Tumors Classification by Using Gray Level Co-occurrence Matrix, Genetic Algorithm and Probabilistic Neural Network." *Diyala Journal of Medicine* 14, No. 2 (2018): 138-151.

[41]Theodoridis and K. Koutroumbas, “Pattern recognition”, 4th ed. Elsevier/Acad. Press, 2008.

[42]Mohanaiah, P., P. Sathyanarayana, and L. GuruKumar, “Image texture feature extraction using GLCM approach”, International journal of scientific and research publications 3, No. 5, PP 1-5, 2013.

[43]Richert, Willi, “Building Machine Learning Systems with Python”, Packt Publishing Ltd, No. 1200713, 2013.

[44]Ayodele, Taiwo Oladipupo, “Types of machine learning algorithms”, new advances in machine learning, PP 19-48, 2010.

[45]Russel, Stuart, and Peter Norvig, “Artificial Intelligence: A Modern Approach”, Third Edition, Prentice Hall, 2010.

[46]Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, “Foundations of Machine Learning”, the MIT Press ISBN 9780262018258, 2012.

[47]Jamal. M. AL-Tuwaijari and S. I. Mohammed, “Face Image Recognition Based on Linear Discernment Analysis and Cuckoo Search Optimization with SVM”, International Journal of

Computer Science and Information Security. ISSN 1947 5500, IJCSIS, Vol. 15, No. 11, November 2017.

[48]Kolluru, Pavan Kumar, “SVM Based Dimensionality Reduction and Classification of Hyperspectral Data”, University of Twente Faculty of Geo-Information and Earth Observation (ITC), 2013.

[49]Lasse Alfredsson, “Feature extraction for image selection using machine learning”, M.s.C Thesis in Electrical Engineering Department of Electrical Engineering, Linkoping University, 2017.

[50]Ghaidaa Waleed., and Jamal Mustafa Al-Tuwaijari. "Satellite Images Scene Classification Based Support Vector Machines and K-Nearest Neighbor." *Diyala Journal For Pure Science* 15, No. 03, PP. 70-87, 2019.

[51]Ramon Huertaa, Fernando Corbachob ,and Charles Elkanc, “Nonlinear support vector machines can systematically identify stocks with high and low future returns”, Algorithmic Finance 2, PP 45-58, 2013

[52]Achirul Nanda, Muhammad, Kudang Boro Seminar, Dodi Nandika, and Akhiruddin Maddu, “A comparison study of kernel functions in the support vector machine and its application for termite detection”, Information 9, No. 1 , P 5, 2018.

[53]Soentpiet, Rosanna, “Advances in kernel methods: support vector learning”, MIT press, 1999.

[54]Vapnik, Vladimir, and Vlamimir Vapnik. “Statistical learning theory”, 1998.

[55]Wang, Zhe, and Xiangyang Xue.,“Multi-class support vector machine.”, In Support Vector Machines Applications, Springer, PP 23-48, 2014.

[56]Nuha Jameel Ibrahim, “Intelligent Human Detection and Recognition System, electronic thesis, Computer Sciences, University of Technology, 2014.

[57]I. Arel, D. C. Rose, and T. P. Karnowski, “Deep machine learning-a new frontier in artificial intelligence research [research frontier]”, Computational Intelligence Magazine, IEEE, vol. 5, PP 13-18, 2010.

[58]Glauner, Patrick O. “Deep convolutional neural networks for smile recognition”, arXiv preprint arXiv: No.1508.06535, 2015.

[59]Benuwa, Ben Bright, Yong Zhao Zhan, Benjamin Ghansah, Dickson Keddy Wornyo, and Frank Banaseka Kataka, “A review of deep machine learning,” In International Journal of Engineering Research in Africa, Vol. 24, PP 124-136. Trans Tech Publications Ltd, 2016.

[60]Bengio, Yoshua, Aaron Courville, and Pascal Vincent, “Representation learning: A review and new perspectives”, IEEE transactions on pattern analysis and machine intelligence 35, No. 8, PP 1798-1828, 2013.

[61]LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document recognition.” Proceedings of the IEEE 86, No. 11, PP 2278-2324, 1998.

[62]Guo, Yanming, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew, “Deep learning for visual understanding: A review.”, Neurocomputing 187, PP 27-48, 2016.

[63]Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I. Sánchez, “A survey on deep learning in medical image analysis”, Medical image analysis, PP 60-88, 2017.

[64]Anna Gummeson “Prostate Cancer Classification using Convolutional Neural Networks” M.S.C thesis Lund Institute of Technology, 2016.

[65]Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “Imagenet classification with deep convolutional neural networks”, In Advances in neural information processing systems, PP 1097-1105, 2012.

[66]Zeiler, Matthew D, “Hierarchical convolutional deep learning in computer vision”, Ph.D Thesis, New York University, 2013.

[67]Yamashita, Rikiya, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi, “Convolutional neural networks: an overview and application in radiology”, Insights into imaging 9, PP 611-629, 2018.

[68]Iftene, M., Q. Liu, and Y. Wang. “Very high resolution images classification by fine tuning deep convolutional neural networks” In Eighth International Conference on Digital Image Processing (ICDIP 2016), International Society for Optics and Photonics, vol.10033, PP 100332D. 2016.

[69]Demuth, Howard B., and Mark H. Beale. , Orlando De Jess, and Martin T Hagan. *“Neural Network Design”, Martin Hagan*, 2014

[70]LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, “Deep learning”, Nature 521, No. 7553, PP 436-444, 2015.

[71]Ioffe, Sergey, and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.”, *arXiv preprint arXiv:1502.03167,* 2015.

[72]Bottou, Léon, and Olivier Bousquet, “The tradeoffs of large scale learning”, In Advances in neural information processing systems, PP 161-168, 2008.

[73]Kantardzic, Mehmed. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.

[74]Reem Majeed Ibrahim, “Classification of Face Image Based on Gender Using Intelligent Method”, M.s.C electronic thesis, Computer Sciences, University of Technology, 2015.

[75]Ismael, M. R., & Abdel-Qader, I. (2018). Brain Tumor Classification via Statistical Features and Back-Propagation Neural Network. IEEE International Conference on Electro Information Technology, 2018-May, 252–257. https://doi.org/10.1109/EIT.2018.8500308.

[76]Pei, L., Vidyaratne, L., Hsu, W. W., Rahman, M. M., & Iftekharuddin, K. M. (2020). Brain tumor classification using 3D convolutional neural network. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11993 LNCS(Icasert), 335–342. https://doi.org/10.1007/978-3-030-46643-5_33.

[77]Pashaei, A., Sajedi, H., & Jazayeri, N. (2018). Brain tumor classification via convolutional neural network and extreme learning machines. 2018 8th International Conference on Computer and Knowledge Engineering, ICCKE 2018, Iccke, 314–319. https://doi.org/10.1109/ICCKE.2018.8566571.

[78]Seetha, J., & Raja, S. S. (2018). Brain tumor classification using Convolutional Neural Networks. Biomedical and Pharmacology Journal, 11(3), 1457–1461. https://doi.org/10.13005/bpj/1511.