Brain Tumor Diagnosis Support System: A decision Fusion Framework

by

Kalifa Mabrok Shantta

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Simon Yang Professor, University of Guelph
Supervisor(s)	Otman Basir Professor, University of Waterloo
Internal Member	Behzad Moshiri Professor, University of Waterloo
Internal Member	Hamid Tizboosh Professor, University of Waterloo
Internal-external	Mark Crowley Assistant Professor, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

An important factor in providing effective and efficient therapy for brain tumors is early and accurate detection, which can increase survival rates. Current image-based tumor detection and diagnosis techniques are heavily dependent on interpretation by neuro-specialists and/or radiologists, making the evaluation process time-consuming and prone to human error and subjectivity. Besides, widespread use of MR spectroscopy requires specialized processing and assessment of the data and obvious and fast show of the results as photos or maps for routine medical interpretative of an exam.

Automatic brain tumor detection and classification have the potential to offer greater efficiency and predictions that are more accurate. However, the performance accuracy of automatic detection and classification techniques tends to be dependent on the specific image modality and is well known to vary from technique to technique. For this reason, it would be prudent to examine the variations in the execution of these methods to obtain consistently high levels of achievement accuracy. Designing, implementing, and evaluating categorization software is the goal of the suggested framework for discerning various brain tumor types on magnetic resonance imaging (MRI) using textural features.

This thesis introduces a brain tumor detection support system that involves the use of a variety of tumor classifiers. The system is designed as a decision fusion framework that enables these multi-classifier to analyze medical images, such as those obtained from magnetic resonance imaging (MRI). The fusion procedure is ground on the Dempster-Shafer evidence fusion theory.

Numerous experimental scenarios have been implemented to validate the efficiency of the proposed framework. Compared with alternative approaches, the outcomes show that the methodology developed in this thesis demonstrates higher accuracy and higher computational efficiency.

Acknowledgments

Today is the last day to set the final touches on my research after a long and stressful time. Begin with, praise be to Allah; I have managed to reach this stage while being under the effects of cancer and radiation treatments at their early phases. However, it has been a great and eye-opening experience for me, and I would want to thank everyone who has assisted me along the way. First and foremost, I would like to thank my supervisor, Dr. Otman Basir, for his wisdom, patience, encouragement, passion, and vast knowledge.

I want to offer my heartfelt appreciation to the doctorate committee members, Dr. Mark Crowley, Dr. Hamid Tizhoosh, and Dr. Simon Yang, and Dr. Behzad Moshiri, for investing time in reading and providing many valuable comments on my thesis.

Dedication

To my mother's soul, may she rest in peace. To my father.

To my wife and my daughters and sons.

Table of Contents

List of Tables

List of Figures

Chapter 1

"Back in 2015, as I started my academic research, I was diagnosed with lung cancer, and after a year of complete recovery, unfortunately, the tumor moved to my brain. I was grateful for the diagnostic system that I went through, but what happened gave me a strong motivation to look for better methods to enhance medical image classification"

1 Introduction

Medical image analysis is an integral component of differential disease diagnosis. Correct diagnosis allows early treatment, while the incorrect diagnosis has repercussions that vary according to the type of improper diagnosis. Missing the disease altogether will result in delayed treatment, which might be fatal. In addition, misdiagnosis can be lengthy and costly and can lead to needless, painful, and often harmful treatment. These problems are even more significant concerning early brain tumor detection misdiagnosis. A substantial cause of the challenges associated with the medical image analysis process is the human factor. Different radiologists might have differing opinions. Menze et al. [1] showed that even expert classifiers exhibited considerable differences in locations where intensity gradients between tumor formations and surrounding tissue are smooth or masked by bias field aberrations that influence partial volume. On the other hand, automatic detection is prone to fluctuations in performance related to the technique chosen, and the imaging

modality analyzed. Therefore, a need exists for the design of a robust and reliable brain tumor detection support system. This research proposes to fuse a set of classifiers to detect and classify medical images to achieve reliable brain tumor detection. In addition, the attempt to answer what is the optimal technique for fusing such classifiers in the sense that the detection decision of the set is better than the decision of any single classifier of the set with respect to accuracy. The objective of combining the information and the decisions is to calculate the appropriate group of algorithms for a particular set of images and devise a relevant approach that can efficiently integrate the algorithm's decisions. The system developed for detecting brain tumors will be designed and implemented, and its performance will be analyzed.

This chapter presents an overview of medical imaging-based diagnosis, with an emphasis on brain tumor detection. The chapter discusses the motivation behind this research and the specific objectives of the work. The chapter concludes with an outline of this thesis.

1.1 Research Motivation

Human brain is a highly complicated organ, as detecting brain tumors is highly challenging. In general, the radiologist detects the brain tumor that takes a considerable time to examine MR images. The central concept is to enhance a computer-aided diagnostic system that would allow the radiologist to get a second opinion on whether or not there is a tumor present. As a primary diagnostic tool used in the healthcare system, medical imaging involves applying various technologies for visualizing the human body, which is employed as alternatives to surgery for monitoring, diagnosing, or treating a medical issue. The most frequently utilized diagnostic modalities for detecting several types of disease are magnetic resonance imaging and computed tomography [2].

However, the interpretation of MRI and CT results is not always straightforward. Scanning creates visual representations of the body's interior, but the performance of what the instruments have recorded depends on human eyes and radiologists' opinions. While the choice of treatment procedure is linked to the results obtained from scans, studies have demonstrated that not every radiologist classifies or explains MRI images in a similar way. In general, the classification of test outcomes can be affected by several influences, such as the radiologist's level of experience, workload, and fatigue. Besides, radiologists have differing opinions, and unfortunately, even the most experienced physicians can make mistakes. As mentioned in [3], even the most capable and highly skilled physicians can be in error, or their decisions can differ concerning many situations. For instance, discrepancies in the interpreting and reporting of medical pictures were discovered in 5,278 of 8,400 CT scan images at Vancouver General Hospital between October 2016 and January 2017[4]. Additionally, in Canada's Niagara region, groups of six hospitals have begun to recheck 4000 CT, MRI, and mammogram scans after discovering a radiologist's mistake in interpreting and classifying these images in the period from May 2014 to May 2015. They estimated that it would take up to four months to complete this revision [5,6]. Furthermore, in Mississauga, two hospitals from April 2012 to March 2013, as many as 3,500 CT scans and mammograms showed that the medical specialist had classified and misinterpreted a scan result [7,8]. The inadequate number of radiologists and large quantities of medical images such as MRI images that must be classified make human interpretation time-consuming and extremely expensive. Numerous imaging modalities, such as MRI, CT, and mammography, demand specialized capability for knowledgeable diagnosis. In addition, the demand for these imaging devices extremely increased, with the requirement to develop a system that

presents an alternative perspective. Besides, the detection precision can be enhanced. An advantage of applying and utilizing computers rather than another human opinion is that specialist radiologists' mistakes will be reduced. The sensitivity and specificity of tumor detection executed by an automatic system are assumed superior to those provided by radiologists. For this reason, computer system algorithms are recommended for application in the analysis and reading of a variety of medical images. The fundamental goal is to utilize advanced artificial intelligence to implement innovative image analysis tools to assist radiologists by offering more targeted and efficient detection of diseases such as brain cancer.

Key motivators of this research are:

1. Minimize the error of medical image classification.
2. The opportunity is to provide a second opinion for radiologists for analyzing medical images.
3. Reducing the time required to read and classify such images.

A further factor is that reliance on human knowledge for the organization of an MRI scan for evaluation, which characterizes the majority of current direct-detection methods, increases the chances of incorrect classification and identification of brain cancer [9].

Implementing an automatic brain tumor detection system can ensure the rapid and precise detection of a tumor and reduce observational oversights and the rates of the consequent false negatives reported by the radiologists who classify medical images [10].

The principal goal of this thesis is to achieve more precision in terms of sensitivity as well as specificity to the outcome and to reduce the time complexity by finding the best optimization technique.

1.2 Research Objectives

The research approach mainly focuses on devising intelligent algorithms to classify the tumor from the brain MR images. Besides, the goal of this research is to improve a system that can classify or detect brain tumors in a variety of medical images, such as those produced by CT and MRI. This system has been created to be a means of minimizing observational errors and the rates of the consequent false negatives that result from physician classification of medical images. The method introduced in this research can also be employed as additional input by the radiologists who make final decisions. The following are the specific objectives of the proposed research:

1. Investigate several pre-processing and segmentation techniques to implement the most recent medical image examination approaches. A variety of feature extraction methods will be employed to convert complex image content into content features. To create a novel method of significantly expanding the effectiveness of image feature selection, optimal features will be selected from a range of significant image features.
2. Since the selection of each classifier is the first stage of the fusion process, the developing methods for selecting the appropriate group of classifiers are essential. In this context, requiring is to evaluating and tuning each classifier individually to achieve optimal fusion performance.
3. Explore the potential of the Dempster-Shafer theory (DST) as a framework for facilitating classifier fusion. The DST represents the uncertainty inherent in the final decisions, and hence one can mitigate false negative and false positive

detection results. In addition, simple classifier fusion, such as majority voting, average probability, min rule-max rule, and product probability will be used for implementing the combination of classifier decisions.

4. Investigate comparatively the applicability and implementation of various medical images' algorithms in the feature extraction, and classification techniques in addition to multi-classifier combination techniques.
5. Conduct experiments to estimate the achievement of the system with regard to sensitivity, specificity, and classification accuracy.
6. The outcomes of the proposed classifier fusion framework are compared to that of a deep-learning based detection framework.

1.3 Thesis Statement

Medical imaging regularly demands experienced medical physicians to see best the information shown in the images. Nevertheless, because of different personal factors and short review time and tools, it is considered typical that various medical physicians may come up with different analyses or interpretations, heading to various examinations. This thesis proposes a new system for the detection and classification of brain tumors[3]:

"The 'human factor' in reviewing and classifying medical images is unavoidable; as a result, even obvious anomalies may go undetected; the mere fact that a radiologist misses an abnormality on a radiograph does not mean that he or she has committed malpractice; and not all radiographic misses are excusable. As a result, the emphasis of concentration should be on problems like the utilization of appropriate procedures to assist the radiologist in reaching the correct judgment or conclusion."

The literature on discovering and examining brain tumors is reviewed to realize the problem and study the topic. Several modalities for brain imaging are considered. Many classifications of brain cancer detection in CT and MRI were evaluated, and their performance was assessed. Multi-classifiers, all of which use the supervised training approach was used for image classification tasks. They differ actually in their approach on how to classify data. In addition, the best way to mix or fuse their outputs without considering the nature of the feature set and pattern representation of the input data. The confusion matrix and other classifier tools evaluate the output of the developed system.

1.4 Thesis Overview and Scope

This thesis is composed of the following chapters:

Chapter 1 presents the research motivation**,** objective, and research statement.

Chapter 2 presents the categories of brain tumors and medical-related information of brain cancers, MRI and CT images characteristics.

Chapter 3 presents a survey on brain tumor detection and segmentation.

Chapter 4 presents the Dempster-Shafer-based brain tumor detection algorithms.

Chapter 5 presents a Deep-Learning-based Tumor Characterization and Detection.

Chapter 6 presents the performance results of each classifier and that of the classifier fusion with respect to factors such as accuracy, true positive rate, true negative rate, and precision.

Chapter 7 summarizes the proposed brain tumor detection system, the contributions made in this thesis, and the recommendations suggested for future work.

Chapter 2

Brain Tumors and Medical Imaging

2.1 Introduction

The primary and most remarkable composite organ in the human body is the brain. It is constructed from more than a hundred billion nerves that interconnect trillions of connections named synapses [11]. It is the origin of all human behavior, thoughts, feelings, and understanding. It also integrates and controls relating to balance and autonomic functions in the body. The brain produces many hormones and regulates its processing, awareness, attention, and integration related to emotion. It is supported and protected by the surrounding skin, bones of the skull, and the meanings. It also holds a watery fluid called cerebrospinal fluid. This fluid flows through spaces between and within the brain spaces called ventricles. The skull is another special protector of the brain. It is a highly complicated structure, has compact and elastic types of bones. The brain is shaped by three major parts: the forebrain, midbrain, and hindbrain. The forebrain is formed by the cerebrum, thalamus, and hypothalamus. The cerebrum is the most significant portion of the brain. In fact, 85% of the brain's weight comes up from the cerebrum. The thinking portion of the brain and controls and maintains voluntary human

muscles also are from cerebrum functions. The midbrain has a tectum and tegmentum. The hindbrain forms the cerebellum, pons, and medulla. The brain is the supervisor for the displacement, dream, hunger, thirst, and other vital activities indispensable to survive. It controls the main five functions: receiving or taking the information by the senses, storing and recalling the information, analyzing and thinking about the information, force the controller, processing the functions simultaneously, or assigning all internal and external operations of the body.

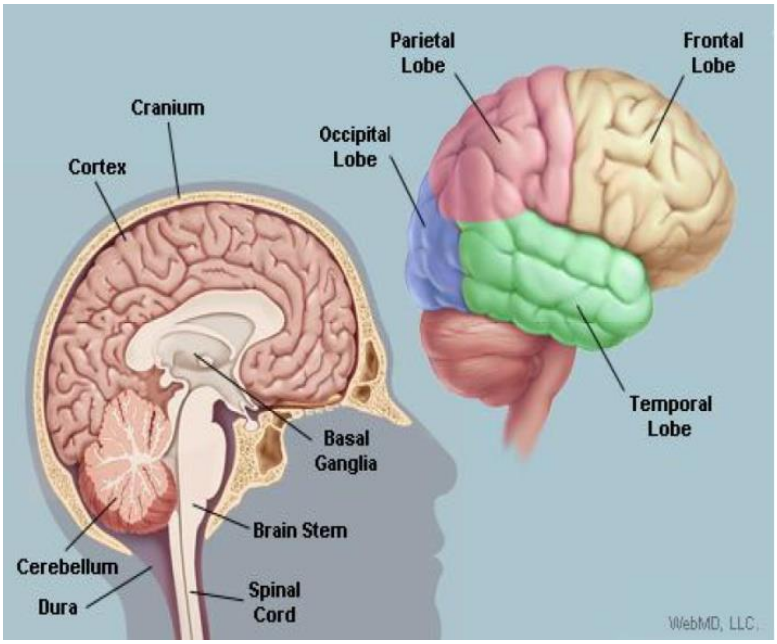

Figure 2.1: The structure of the brain [11].

This brain encloses an incredible number of neurons for the computational process in a particular unit. These neurons are attached within the brain, and those make direct connections to other neurons. The structure of the brain is shown in Figure 2.1. The brain is a smooth, sensitive, and soft form of tissue. It is the source of all human behavior, thoughts, feelings, and experience. It is supported and protected by the surrounding skin, bones of the skull. It also contains a watery fluid called Cerebrospinal fluid.

2.2 Brain Tumors

The tumors are the foremost reason for death in economically developed countries and the second most important reason for death in developing countries. Tumor recognized medically as a malignant neoplasm is an extensive group of a mixture of diseases, all concerning unregulated cell growth. In addition, a tumor is a great set of sicknesses that can begin in practically some organ or muscle of the body once irregular cells start to grow nonstop, spread outside their usual boundaries to attack attached parts of the body, and expanded to other organs. The latter process is called metastasizing and is a significant reason for death from cancer. A neoplasm and malignant tumor are other common names for cancer. Malignancy is the second reason for disease worldwide, accounting for approximately 9.5 million deaths, or one in six deaths, in 2018 [12].The grading system scales are utilized from grade 1 to grade 4, as stated by the world health organization. The benign and malignant cancer kinds are classified based on these grades. The low-level grade tumor is one and two, where the high-level grade is three and four. A brain tumor can affect a person at any age. The effects each person endures may not be the same. The diagnosis of the tumor area in the brain is challenging due to such a complex human brain structure. The fast-growing malignant tumor type is from grades three and four. It is also spread to adjacent portions of the brain or spinal, further hurtful, and may stay untreated. The classification, place, and dimension of the brain tumor in the early phase are essential in the medical field. Improving the novel imaging methods leads to support physicians to detect and observe the development of the tumor influence region at various phases [13].

Consequently, they can provide an appropriate diagnosis with this image scanning. Detecting a brain tumor in the early phases is a high issue; thus, appropriate therapy can be adopted. The correct treatment decision, such as radiation, chemotherapy, or surgery, is based on this information. As a result, the opportunity for survival of a tumor-infected person can safely rise if the swelling is discovered precisely in its early phase [14].

According to 2019, Canadian cancer statistics reports that it is predictable that about one in two Canadians will acquire tumors during their lifetime, and around one in four Canadians will lose their life of tumor. Predictably, 220,400 Canadians will be identified with tumors, and 82,100 will die from the illness in 2019 only, as shown in Figure 2.2 [15].

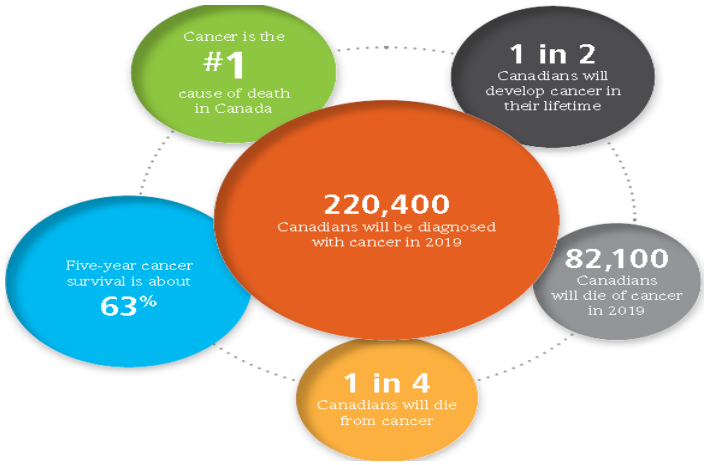

Figure 2.2: Canadian cancer statistics [15].

2.2.1 Benign and Malignant Brain Tumors

Benign tumors are non-cancerous growths in the body that cannot invade neighbouring tissue. They can be removed entirely and are unlikely to reappear. Benign brain tumors do not spread to adjacent tissue; however, they can cause significant pain, lasting brain damage, and death. Malignant brain tumors have no specific limits. They overgrow, create increasing pressure within the brain, and diffuse throughout the brain or spinal

cord beyond their point of origin. It is sporadic for malignant brain tumors to spread outside the brain.

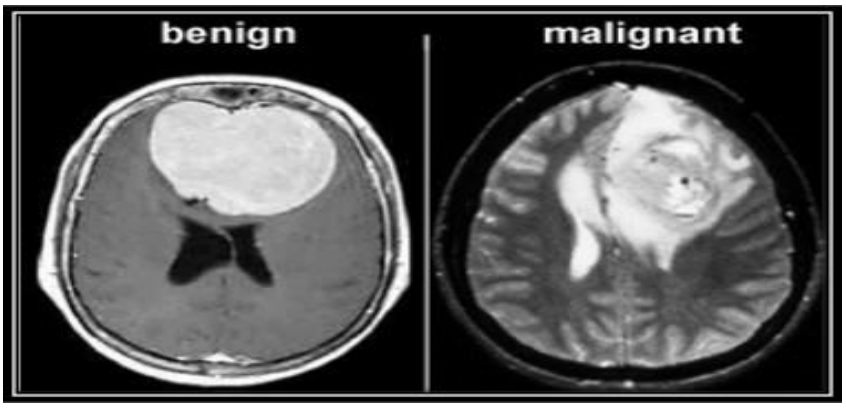

Figure 2.3: Benign tumor and malignant tumor [5].

A malignant brain tumor is either graded 3 or 4, whereas grade 1 or 2 tumors are usually classified as benign or non-cancerous.

2.3 Categories of Brain Tumor

Brain tumors are separated into dual clusters: primary (brain tumors) and secondary (metastatic) tumors, which arise from malignant cells that have moved from their principal place and invaded the central nervous system through the hematoencephalic barrier. Primary brain tumors can be benign or malignant, and they can be neuronal (brain cells) or Neuroepithelial in nature. The major feature of benign tumors is that they are made up of slow-growing cells arranged in well-defined patterns. Diagnosis might be difficult because of the reality that these cells mimic healthy cells once viewed under a microscope. It is worth noting that benign tumors account for roughly 41% of totally primary brain tumors. Once the tumor is not in a pivotal place and operating exclusion is possible, therapy is beneficial. Radiation therapy is an alternate treatment option, mainly when life-threatening illnesses are triggered by the benign growth's location [17]. Due to

their hostile and invading character and unlimited mass increase that ultimately drives major consequences such as pressure on key tissues, malignant tumors are life-threatening. Brain malignancies, dissimilar additional kinds of cancer tumors (lung, liver, breast, etc.), are typically confined and infrequently prevalent (metastasize) to different parts of the body. Furthermore, because of the sensibility of the adjacent brain tissue, surgical removal is considered exceedingly risky. Glioblastoma multiform (grade 4 astrocytoma) is the most common form of malignant primary brain tumor, accounting for around 20% of all primary brain tumors [18].

2.3.1 Gliomas

Glial cell tumors are a kind of development that starts in the brain or spine and develops from glial cells. The brain is an incredibly popular glioma website. Glioma is a broader categorization that includes two primary histologic subtypes: astrocytoma and oligodendroglia.

2.3.2 Meningioma

The second most frequent essential tumor of the focal nervous system is meningioma, arising in the arachnoid villi's arachnoid 'cap' cells. The majority of these tumors are typically benign, although they can also be cancerous [18].

2.4 Brain Tumor Diagnosis

A brain tumor is known as irregular cell development in the brain (National Cancer Institute, http://www.cancer.gov/). Tumors can be distinguished in many methods, including definite symbols and symptoms, screening examinations, or medical imaging. Once a possible tumor is detected, then it is identified by microscopic analysis of a tissue

sample. The tumor is typically treated with chemotherapy, radiation therapy, and surgery. The probability of surviving the disease fluctuates significantly by the type, location of the tumor, and the extent of the illness at the start of treatment. There are two types of brain tumors: benign and malignant. In opposition to healthy cells, tumor cells occur from unrestricted cell outgrowth and spread to the subsequent tissue. Even though benign tumors can develop enlarged and push on normal organs and tissue that can influence their operation. Prime brain tumors create in the brain, although secondary brain tumors start from other body sections. Typically, the definitive analysis of a cerebrum tumor is grounded on histological examination of tissue samples acquired by a biopsy. A biopsy is highly invasive and has vital dangers with evaluated morbidity of 2.4−3.5% and a death percentage of 0.2−0.8% [19,20]. For instance, a biopsy will not always be performed, incredibly old or infirm patients or patients with quite slow-increase tumors. Furthermore, surgery could be escaped for specific pathologies like lymphomas and brain abscesses. Imaging approaches, such as MRI, CT, and PET, could identify a brain tumor and bypass needless operation. The anatomical assessment of brain tumors and high-resolution information can be obtained by MRI, used widely in clinical practice.

Interestingly, MRS and MRSI can provide important information about metabolism. However, despite these promising potentials, MRS and MRSI are not yet widely applied in the clinical location since specialized processing, and examination of the obtained information are required [21]. This thesis focuses on the application of MRI.

2.5 Medical Imaging

Dzung et al. [22] demonstrated that an image gathers measurements in two-dimensional or three-dimensional space. In modern medicine, this process is called medical imaging, a

field in which significant advances have led to this technique for obtaining knowledge of the human body now having several beneficial clinical applications. Several new types of medical imaging have been produced in recent years, each characterized by particular benefits and drawbacks. Many types of analysis are employed with medical images: radiation absorption such as in X-rays, radio frequency (RF) as in magnetic resonance imaging, and acoustic pressure, like with ultrasound. In several medical imaging techniques, a different technology is utilized for formulating each type of image. Scalar imaging involves a single measurement made at each position in an image. Multichannel imaging refers to methods in which more than one measurement is made, such as MRI. In X-ray imaging, the image can be acquired in a continuous space and MRI in a discrete space. In images with two distinct dimensions, the position of every measurement is known as a pixel, and in three-dimensional images, it is called a voxel. Image types also vary regarding their display quality and their use for specific body tissues (e.g., bone, soft tissue, or tumors). These distinctions are a vital consideration for physicians selecting which imaging modality to use [23]. Since the beginning of this century, extensive developments in imaging technology have led to the invention of numerous additional imaging techniques. The imaging modalities frequently used are X-ray, computed tomography (CT), MRI, and the principal types of imaging used in modern medicine: radiography, MRI, and CT nuclear medicine. The following sections offer a whole discussion of the theory underlying both the MRI modality and CT imaging, which are the image techniques that have been applied for classification using the introduced system.

2.5.1 Goals of Medical Image Analysis Techniques

1. Quantification: measuring the attributes on medical images, assist radiologists to acquire assessment from medical photos.
2. To make the features measurable, it is necessary to extract objects from images by segmentation.

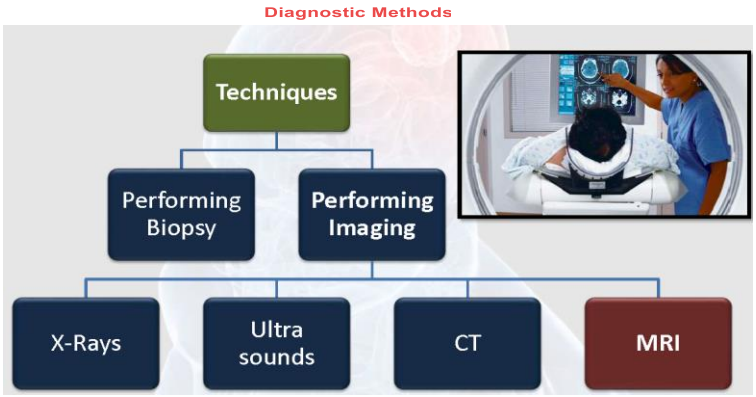

Figure 2.4: Different medical diagnostic methods.

3. Make a diagnosis using Computer Aided Diagnosis (CAD) based on measurements and attributes. Assist radiologists with their diagnostic procedures to ensure accuracy and efficiency.
4. Techniques for evaluation and validation.

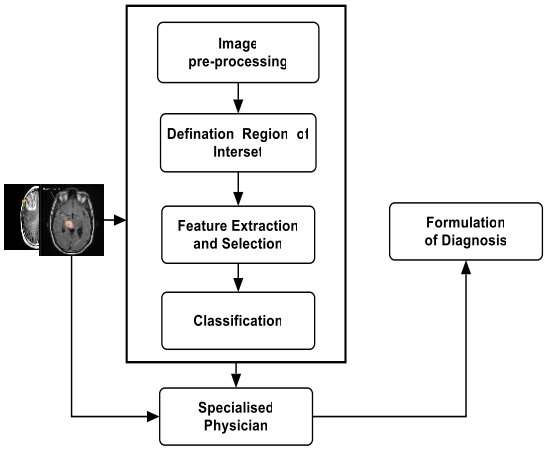

Figure 2.5: Computer aided diagnosis CAD.

2.6 Magnetic Resonance Imaging

Magnetic resonance imaging is a visual that steps extraordinary resolution photos from soft body tissue principally for clinical diagnosis. MRI utilizes a magnetic field and radiofrequency to produce a complete image of body organs and structures. It makes available a visual representation of the inside of a body for clinical examination and medical checking. Compared to other imaging formats, MRI can provide a remarkably high level of detail. For example, MRI images of white and grey brain matter can differentiate and diagnose aneurysms and tumors. The basic principle behind an MRI scanner is that, as a person is positioned in a strong static magnetic field, the protons in the body are aligned with the direction of the magnetic field. An RF pulse is generated in order to produce a magnetic resonance signal. The protons absorb the transferred power, and the radio wave is then switched off, and the RF energy absorbed is retransmitted at the resonance frequency. Simultaneously, the protons begin to realign, and during this realignment period, the protons begin to emit a radio signal. Figure 2.6 displays a graphic picture of an MRI machine and its utmost significant parts. Antennae (coils) are used for detecting these RF signals, which are then directed to a computer where the highly detailed images are rebuilt.

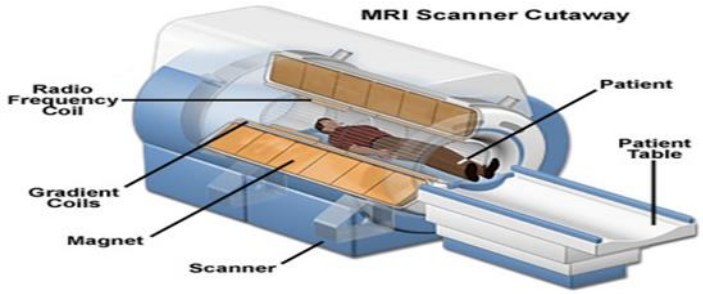

Figure 2.6: MRI scanner with its most important parts [24].

various sorts of magnetic resonance imaging are applied in this process, depending on necessity. The kind of series applied in magnetic resonance imaging on condition that as an input in the pre-processing phase is identical to T1, T2, and FLAIR. It is required to explain echo and repetition time to comprehend various kinds of MRI images. TE denotes the period from the focus of the radio frequency oscillation to the position of the TE. For oscillation progression with doubled echoes among every radiofrequency oscillation, some TE times may be described and are usually denoted TE1, TE2, TE3, etc. TR is the measurement of period among identical sequence points on a periodical chain of pulses and TE. Figure 2.7 shows a tumor's appearance in different images [23].

1. T1-weighted images comprise a black show of spinal fluid and liquid. White matter is brighter than gray matter. In cerebrum arrangement photos, T1 offers a better result, and fat shows shiny in this kind. Time of echo and time of repetition with 500 and 14 msec respectively.

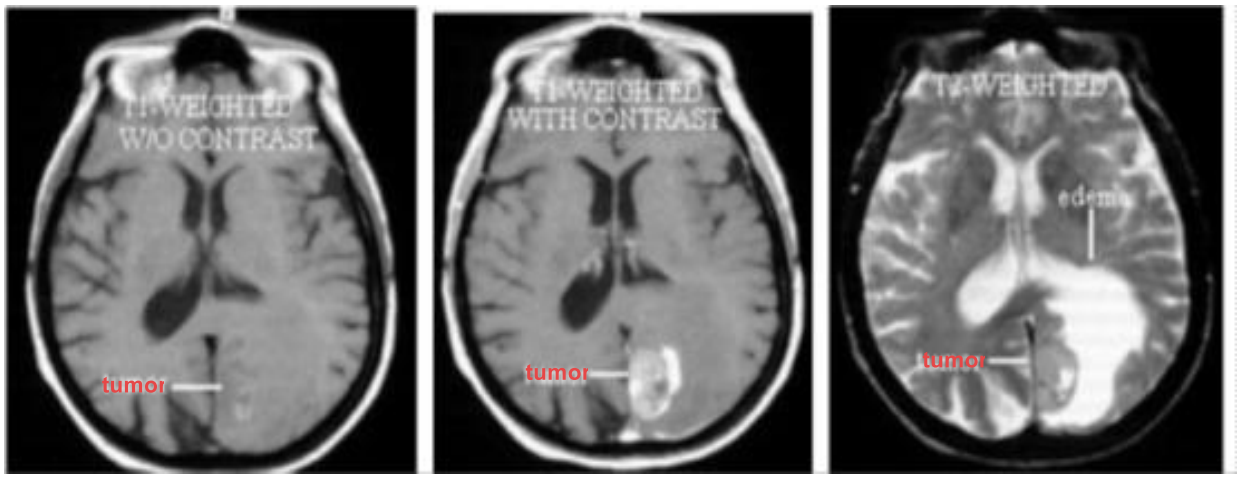

Figure 2.7: MRI brain images.

2. T2-weighted images comprise the upper signal strength of cerebrospinal fluid and liquid related to tissue and cause it to appear luminous. The repetition of 4000 ms and echo of 19 ms is used in T2 to make photons-spin relaxation respectively . T2 is shiny showed at water and liquid, perfect with the edema muscle.
3. FLAIR is similar to T2, it has decreased cerebrospinal liquid, and so on, irregularities remain brightening. It is perfect for imaging cerebral edema. It uses precise length TR and TE times of 9000 ms and 114 ms for creating images respectively.

2.6.1 Physical Principles

The magnetic characteristics (nuclear spin) of the hydrogen nuclei, which are abundant in the human body, are used in MRI. An electromagnetic field is created when hydrogen nuclei rotate around their axis. Because the direction of the spins is random in nature, the overall magnetic field is null. The nucleus spins align with the external field (positive spin) or against it (negative spin) when put in a large magnetic field B_0 with a precession frequency or Larmor frequency. The spins precess at a rate of ω_0 proportional to the external field around the magnetic field axis:

$$\omega_0 = \gamma B_0 \tag{2.1}$$

The longitudinal component (*Mz*, parallel to B_0) and the transverse component of the magnetic vector of spinning nuclei may be expressed (*Mxy*, perpendicular to B_0). The number of positive spins outnumbers the number of negative spins by a little margin during run-through, resulting in a longitudinal magnetization *Mz*. The transverse magnetization is null because the spins do not precess in phase. When a radiofrequency RF pulse is delivered, energy is exchanged between the nuclei and the RF pulse B1 is injected at the resonance frequency ω_0; this changes the balance of spin. The net

magnetization vector tilts away from the linear axis, resulting in a transverse magnetization that a receiver coil may detect. In actuality, the pulse B_1 is perpendicular to B_0, resulting in a null longitudinal component at resonance and the highest possible transverse signal. The nuclei gradually return to equilibrium once the RF pulse is turned off (relaxation phenomenon). The electromagnetic energy that has been absorbed is retransmitted and forms the nuclear magnetic resonance NMR signal during relaxation. There are two types of relaxation processes: longitudinal and transverse relaxation. The spin-lattice (surrounding tissue) interaction correlates to the longitudinal relaxation. By providing energy to the lattice, the spin recovers to maintain its balance condition (the magnetic field's alignment B_0). The recovery of longitudinal magnetism is defined by After 63 percent of the final value is recovered, an exponential curve and a tissue a certain time constant T1 are used. M_0 is the net magnetization at equilibrium, which is decided by the proton density and the exterior magnetic fields strong point. The spin-spin interaction is described by transverse relaxation, which occurs when the spins go out of phase.

$$M_z = M_0(1 - \exp(\frac{t}{T_1}) \quad (2.2)$$

This results in an exponential decrease in the cross-sectional component, as measured by the a temporal constant that is unique to each tissue T2.

$$M_{xy} = M_{xy0} \exp(-\frac{t}{T_2}) \quad (2.3)$$

M_{xy0} is the transverse signal's amplitude after the RF pulse. M_{xy0} equals M_0 when the RF pulse is perpendicular to $B0$. T2 is usually smaller than T1 in practise. The picture contrast and definition of the various MRI sequences are determined by these two parameters [26].

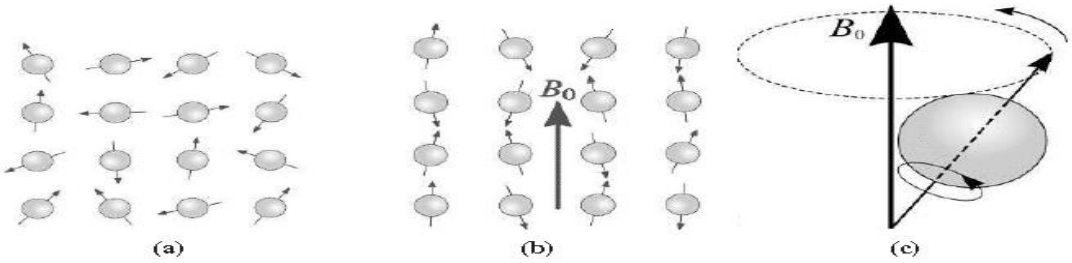

Figure 2.8: (a) Orientation of random spins, (b) alignment with the external field, (c) spin precession around B_0 [24].

The MRI offers a number of advantages that make this modality a preferred choice:

1. High degree of spatial resolution.
2. Excellent for soft tissue characterization.
3. Ability to provide functional brain measurements.
4. No lack of risk to the human body.

The drawbacks are that they are very noisy and movement can affect them. High static magnetic fields may induce nausea, vomiting, dizziness, and headaches in humans.

2.7 Computed Tomography

A CT scan (or CAT scan) has become a primary radiological method usable in a range of clinical applications. Tomography refers to imaging by sections or sectioning. A machine that operates based on tomography is called a tomographic, and the image it generates is a tomogram. CT is the process of scanning a patient in order to gather X-ray absorption coefficients obtained from thin body sections. Multiple measurements are acquired from these coefficients and are subsequently reconstructed into an image that displays the anatomy of the section that has been scanned. With CT, X-rays are applied to create a representation of the body in two-dimensional images. The X-ray transmitter rotates

through 360˚ around the patient at high speed, and a sensitive spherical radiation sensor positioned on the gantry around the patient measures the energy transferred.

The fundamentals of CAT scan processes consist of four steps:

1. With the patient properly located in the scanner, the correct protocols and technical factors are selected. The initialization of the scanning X-rays going from side to side the patient is attenuated according to tissue type. A detector system located opposite the X-ray tube measures the attenuation values as an analogue signal. This signal is transmitted to the ADC, which converts the attenuated signal from analogue to digital, preparing it for processing by the computer.
2. A computer reads the digital data and employs a mathematical formula called a reconstruction algorithm to generate a cross-sectional image. The image reconstruction, which involves millions of data points, is usually performed in less than a second by a group of array processors.
3. The operator displays the reconstructed image, still in its digital format, on an LCD monitor as an image suitable for manipulation.

A wide range of software is available for enhancing the screen image prior to storage. Enhancements include adjusting the density and brightness, changing the plane of the image from axial to sagittal or coronal, producing three-dimensional images, and presenting detailed angiography. The image can then be stored on the computer hard drive or using an external medium such as a versatile optical disk or a hard copy on photosensitive film. Figure 2.9 provides examples of CT images [28].

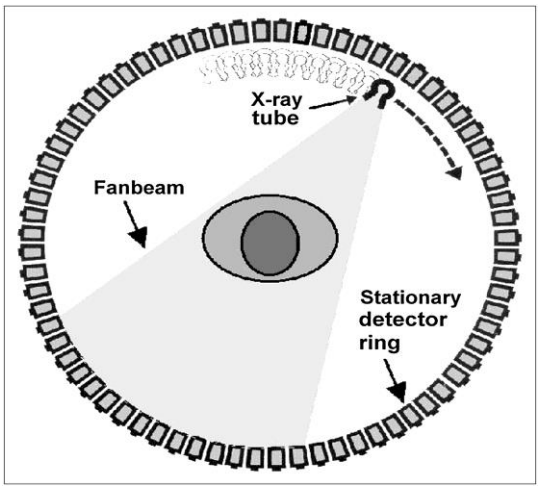

Figure 2.9: Fourth-generation scan geometry fixed detector ring [27].

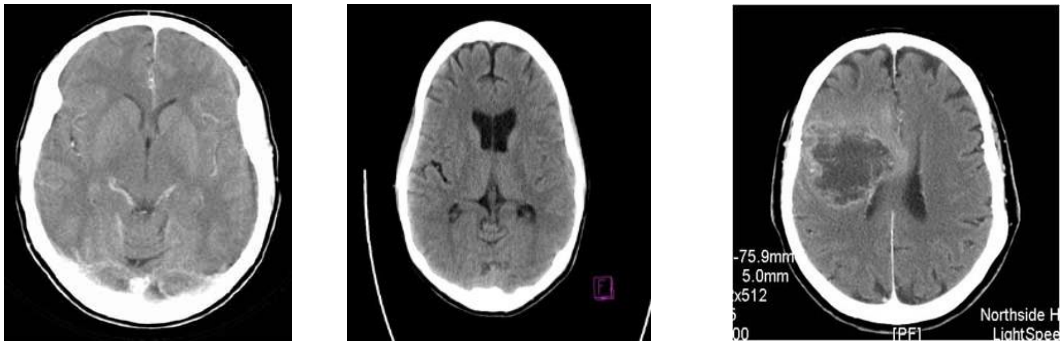

Figure 2.10: CT images.

2.8 Other Medical Imaging Modalities

This section provides a brief introduce to other medical modalities that are beyond the scope of this research. [29]:

1. The X-ray technique is the earliest and most regularly utilized form of medical imaging. A beam of X-rays is sent over the body onto a sensitive plate, and the ensuing unabsorbed X-rays cause the darkening of photographic material. The developed film displays a shadow image of the patient that provides a measure of the weakening of the X-ray within the tissue. X-ray images are frequently

employed for examining conditions such as broken bones, lung disorders, swallowed items that must be retrieved, and blocked blood vessels.

2. The ultrasound modality is a technique based on the application of ultra-frequency sound waves to produce visible photos that represent the interior of the body for clinical examination and therapeutic intervention. A pulse of ultrasonic energy is diffused into the patient from a sensor located on the skin, and the same sensor receives the backscattered echo signal, from which it creates the image. Ultrasound imaging is utilized for large-scale diagnostic imaging of body organs and soft tissues.

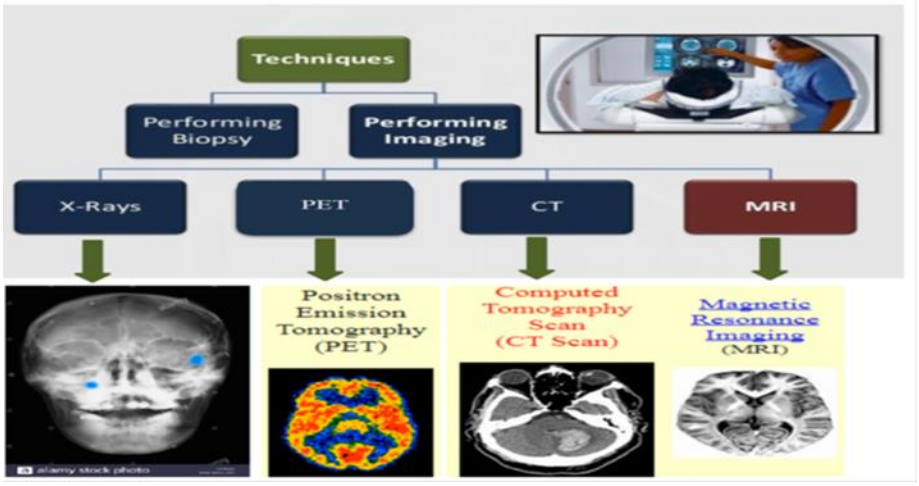

Figure 2.11: Various method of brain tumor imaging.

3. Positron emission tomography, also known as PET imaging, is a form of nuclear medicine. A patient's vein emits gamma radiation as they decay, while a gamma camera scans the radiation area and creates the image. This modality is usually employed to detect cancer, discover whether a tumor has expanded within the body, determine blood flow to the heart muscle, and map normal human brain and heart functions.

Various medical imaging used for brain tumor diagnosis shows in Figure 2.11.

Chapter 3

Background and Literature Review

3.1 Survey on Brain Tumor Detection and Segmentation

Abstract- Despite significant advancements in medical technology, clinicians still face a difficult and time-consuming challenge in detecting brain tumors. Early and accurate identification of brain tumors allows for more powerful and effective treatment, leading to higher survival rates. The ability to automatically detect and classify brain tumors will lead to increased efficiency and predictability.

Automatic detection and classification approaches, on the other hand, differ from method to method and are often image modality reliant. This thesis looks at current detection methods and weighs them on their benefits and drawbacks.

3.2 Introduction

The cancer cells multiply and create a tumor, which is a mass of tissue. In a normal situation, bodily cells die and are replaced with new ones. The presence of malignant and other tumors disturbs this phase to some extent. Tumor cells proliferate despite the fact that the body does not require them, and they do not die like healthy cells. As a result of

this mechanism, cancer continues to spread as new cells are added to the mass. Glioma is a term that refers to a quickly expanding primary brain tumor. Gliomas are tumors that arise from glial tissue, which supports and maintains the cells that transport information from the brain to all areas of the body. Swelling of the cerebrum can be benign or malignant. Non-cancer tumors are non-cancerous growths in the body that do not infiltrate nearby tissue. They can be completely eliminated and are unlikely to return. Although benign brain tumors cannot spread to surrounding tissue, they can cause severe discomfort, long-term brain damage, and death. Brain tumors that are malignant have no defined boundaries. They spread rapidly throughout the brain and spinal cord, increasing pressure and spreading beyond their place of origin. Malignant brain tumors that spread beyond the brain are rare.

3.3 Brain Tumor Detection

As per the survey, one of the highest death rates in the world is brain tumors. Symptoms include changes in the hormones, blood clots, weakness, uncontrolled walking, muddled speech, mood swings, and vision loss . The tumor location defines its type, and its proper diagnosis can save the life of the patient [30]. Benign tumors are non-cancerous growths in the body that cannot invade neighbouring tissue. They can be excluded entirely and are doubtful to appear again. Benign brain tumors do not diffuse to neighbouring tissue; they can create meaningful pain, enduring brain damage, and death. Techniques like MRI or CT scan give the complete structure of a brain tumor as it directs into the intracranial cavity producing a clear tumor image. MRI scan scans by using strong magnetic fields and high radio frequencies to provide detailed information of soft tissues. Computed tomography scan scans by sending X-ray beams. Stages concerned with recognizing

cerebrum tumors are the initializing of an image, feature extraction, segmentation, and post-processing. The needed stages for any automatic brain tumor detection are demonstrated in Figure 3.1

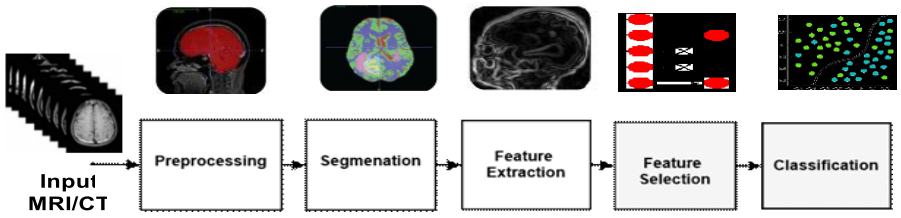

Figure 3.1: Stages in brain tumor detection system.

3.3.1 Image Pre-processing

Pre-processing images is an important part of every image-based application. For the following reasons, a pre-processing stage is required:

1. Pre-processing prepares the images for higher-level processing such as segmentation and feature extraction.
2. Removes the marks or labels such as name, date, and other details (film artifacts) in the image that can affect the classification task.
3. The image quality must be improved.
4. Reduces any types of noise in the image.

3.3.2 Image Segmentation

Image extraction aims to divide a medicinal photo into diverse sections and extract the ROI. In particular, it is used for separating components from the remainder of the image so that they can be observed or recognized as objects. Alireza et al. [31] divided image segmentation into several groups, as illustrated in Figure 3.2. Alireza et al. [31] and Yu Jin [32] published general surveys of image segmentation. Some reviews targeted the

segmentation of MRI images in particular [33,34,35]. Zhang et al. [36] and Clarke et al. [37] provided direct comparisons of diverse approaches for segmenting MRI images.

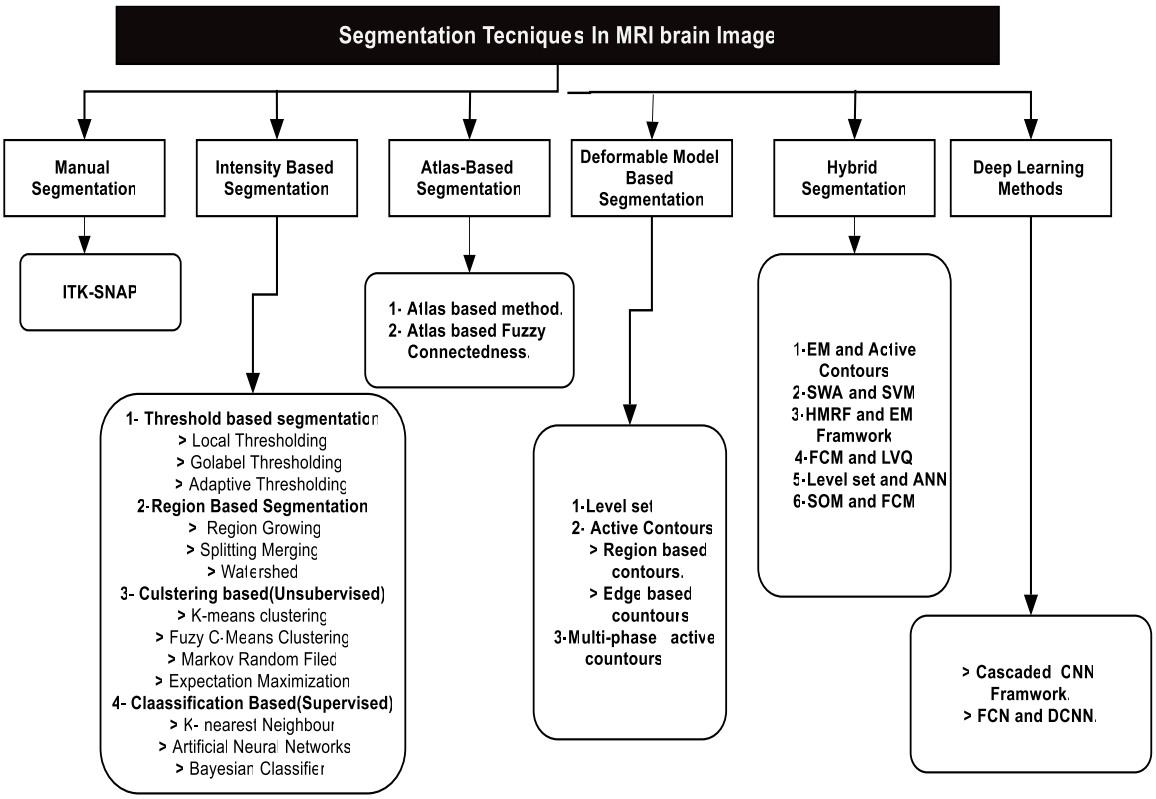

Figure 3.2: Several segmentation methods in MRI brain image examination.

3.3.3 Feature Extraction and Selection

The defined objective of attribute extraction is to reduce the primary information grounded on computing particular components or attributes [38]. The dimensionality reduction is one of the goals from the feature extraction stage, which accurately identifies interesting components of an image as a compressed attribute trajectory. The technique is valuable for applications with large images, for which feature representation must be reduced to enable the fast completion of jobs like image identical and recovery. Ziedan et al. [39] demonstrated that the best-known feature extraction approaches are local binary pattern and spatial dependence matrices. Many research documents have been undertaken

to discuss the features of available extraction techniques [40,41]. The objective of feature selection is to eliminate unrelated and unnecessary features from input records to choose a subgroup of related features for the construction of robust classifiers. This step will probably increase the building speed and precision of the latest algorithm. From a theoretical viewpoint, it can be claimed that the ideal attribute chosen for supervised knowledge tasks needs a comprehensive examination of entirely probable feature subsections. Nevertheless, for a considerable number of features or models, conducting a complete examination of all features to create an ideal attribute set is unrealistic. For this reason, a supervised learning algorithm is employed for analysing an appropriate estimate of the best conventional of attributes for a specific algorithm rather than for determining an optimal set.

3.3.4 Classification Algorithms

In the field of machine learning and taxonomy, researchers have developed new approaches and computer programs to achieve a certain goal. Their research points to the creation of specific learning strategies for improving standard accomplishment based on the use of model data or past experiences [42]. The acceptance of the training during supervised learning is based on patterns with output labels. The usual supervised learning is seen in Figure 3.3. The training assignment is called "classification" when the output values indicate the various classes to which the samples belong.

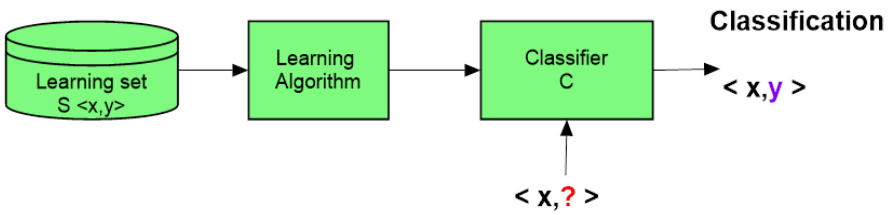

Figure 3.3: Supervised learning model.

Because they were derived from model data, the classifier components accurately characterized the training data. The training dataset for the taxonomy method that will be used to build the classifier is made up of N data points that are formally denoted as $\{x_i, y_i\}^N n=1$, with $x_i \in R^d$ being an input example of dimension d and $y_i \in \{-1, +1\}$, the corresponding class label for a two-class classification task.

3.4 Literature Review

Adel Kermi et al. [43] propose an automated brain cancer segmentation technique in three-dimensional magnetic resonance imaging by using brain and standard group similarity analysis. To minimize noise, the image is pre-processed. The FBB method is both efficient and unsupervised. Tumor detection is done automatically using the FBB technique. In each scenario of tumor form and volume, a geodesic level set-based third deformable model is used to differentiate the tumor's boundaries. Detecting and segmenting tumors takes an average of five minutes to calculate. The results were 38.04% accuracy and 89.01% sensitivity, respectively.

Anithia, S. Murugavali [44] introduce utilizing an algorithm to do a detailed and systematic assessment. Tumor information is unique because MRI segmentation is based on anatomical characteristics and possibly aberrant tissue data. A two-tier method uses K-means algorithms to accomplish successful segmentation and classification. The feature extraction is achieved afterward by implementing the discrete wavelet transform and learning the neural network's SOM. The KNN algorithm then learns the outcome filter features. The testing procedure is similarly done in two phases. It outperforms standard classification approaches, and the results of the trial show that it does. Regular and irregular MRIs are efficiently organized using two-tier classification segmentation

algorithms. The method is implemented using the MATLAB R2013a platform. A statistical measure of this two-level classifier technique is subjected to sensitivity and specificity expressions. The stated result shows that it outperforms the SVM-based classification approach and may be used in medical imaging applications such as image classification and CAD. The accuracy factor was 85%, while the sensitivity factor was 100%.

Singh, A. [45] propose data mining techniques for the classification of magnetic resonance imaging photos. Pre-processing, partition, attribute extraction, and grouping are the four phases of classification. Improvement and skull stripping are done in the main stage to boost speed and precision. During the segmentation step, a fuzzy C-means collecting approach is used. A grey level matrix is used to extract characteristics from magnetic resonance imaging photographs. In the last stage, SVM is used to categorize the pictures. The findings of this study showed that MRI image categorization may be done with a high degree of accuracy and efficiency.

Daniele Ravi et al. [46] present a new dimensionality reduction and processing approach to creating a comprehensive structural map for operation margin. Tissue categorization is hampered by manifold embedding and conflicting findings from various dimensionality reduction approaches. While this technique does the same task in two phases: first, tissue categorization is performed using a distributed stochastic neighbour process, which is then followed by a semantic segmentation method based on semantic texton forests. The proposed system can assist in the development of cancer. The techniques' real-time nature can improve clinical accuracy by providing additional information that can reduce the risk of incorrect sectioning of healthy tissue. The generated tumor maps can be of

exceptional quality and precision by using a well-established method that is shared with a classifier. The accuracy and sensitivity achieved were 81.90% and 80.91%, respectively. Lami Salem et al. [47] suggest a convivial procedure for Glioblastoma medialization. The tumor area is retrieved using a quick-spreading identical method based on global pixel wise-data. To analyse cancer development, the new model uses a cellular automata-inspired algorithm and a fast marching approach. This technique has an optimised runtime of less than 0.7 seconds for each image and does not require much training. When compared to healthy cells, glioblastoma has a distinct grey level potency. The brain picture is divided into two areas using this information. Because of the intensity levels, regions with Glioblastoma are then matched with the estimated model. The tumor is extracted in real-time using the suggested technique.

Mukambik S. Um Ran. [48] propose the following processes to manage four phases in an MRI image: pre-processing, segmentation, pattern extraction, and pattern identification. The skull is eliminated from the MRI image during the pre-processing phase using a twofold thresholding approach; the suggested research provides comparative learning of dual algorithms for cancer identification in MRI images. The first technique is based on using non-parametric deformable models with an active contour to segment brain cancer from magnetic resonance imaging brain images. Another approach used is the K-means segmentation algorithm. Following segmentation, decision making occurs in two stages. DWT was used to extract features and create a Gray Level Co-occurrence Matrix. Finally, in the taxonomy step, SVM is used. There are seventeen noncancerous and twenty-four malignant MRI pictures in the collection. For the segmentation job, the K-means method was used, while the SVM was used for the classification step. In benign

picture classification, the accuracy is 94%, while in malignant image classification, it is 82%. When compared to K-means segmentation, the level set gives the best results.

K. Sudharani, et al. [49] histogram, K-NN method, and distance matrix are some of the approaches used in the proposed procedure. Histogram, for the most part, provides the total amount of the given number of strengths dispersed in a certain snapshot. For proper geometrical representation, the image size was adjusted to 629×839. After adjusting the K value, the KNN algorithm is used to classify and identify the brain tumor. To categorize, the distance is computed using the Manhattan metric. The algorithm was implemented in LabVIEW. All of the photos evaluated had a categorization score of about 95%.

Rasel Ahmed et al.[50] propose method that includes stages such as primary photo treating, segmentation, attributes abstraction, and the last phase is tumor taxonomy employing an ANN algorithm. In pre-processing, adjusted adaptive threshold and histogram imaging were applied to employ together weiner2 and median2 filter. Attribute abstraction is done in dual stages. In the Initial stage, statistical features and in the second-order area, property grounded statistical attribute is obtained. Formerly support vector machine classifies brain MRI pictures into ordinary or cancer brains. The ANN algorithm categorizes the brain tumor. The dataset contains 39 images represents benign and malignant tumors.

Keitan Machale et al. [51] propose an intelligent system that classifies MRI brain images as ordinary and malignant. In this suggested method, four stages were used. The feature extraction step is followed by the pre-processing stage, and ultimately the classification

stage. Diverse classification approaches were applied, such as SVM and KNN, to distinguish 50 images. They conclude that overall accuracy of 98% was obtained.

Cail et al. [52] and Vermaa et al. [53] suggest the intensity command was used to give the function vector in magnetic resonance imaging images. In the categorization step, the SVM was employed. They are also prepared to distinguish the tumor and non-tumor sub-areas and split the healthy tissue.

James Tilton [54] the technique for producing a better hierarchically linked picture subdivision has been defined. Specific area-merging algorithms might provide such segmentations at various levels of lower detail. Following that, the area combining-based hierarchical subdivision, as well as its recursive hierarchical segmentation, was made available. This was utilized to apply the information from the segmentation hierarchy to the area characteristics based on transformations. Furthermore, in this technique, seed point selection in hierarchical extraction and recursive hierarchical extraction remained a difficulty.

Sumitra and Saxena [55] suggest that to categorize the MR brain pictures, a neural network approach should be employed. These are divided into three stages: attribute extraction, dimensionality classification, and reduction. Absolute key features, such as estimations of median, mean, and variance, as well as the greatest and lowest intensity, are eliminated from MRI images using principal component analysis (PCA). In addition, the pattern classification process used a backpropagation neural network. A method for segmenting MRI images that is automated has been discovered.

Xiao et al.[56] propose to calculate tumor and lateral ventricular deformation characteristics in the brain The proposed method is divided into four stages: pre-

processing, extraction, segmentation, and classification. The surrounding picture's non-consistency and repeated features were evaluated in the initial step. Unsupervised segmentation methods were used to estimate the deformation properties of the tumor extraction, and lateral ventricular deformation was used to mine the features. The most often utilized approaches are KNN and pattern matching, which are both employed by conventional fuzzy connected methods C-means (FCM). The primary drawback is that the cluster CSF is incorrectly allocated to a non-CSF pixel, however throughout the extraction procedure, a global mask is used to eliminate this unwanted pixel.

Dahshan et al.[57] suggest a three-stage mixing technique. The factor of MRI pictures has been lowered by using principal components, and dual methods have been expanded by using DWT in feature abstraction. The main classifier is based on the (FP-ANN), whilst the one used to categorize natural or unnatural MRI individual images is based on the (FP-ANN). The drawbacks of this approach are that it necessitates the creation of new learning datasets as well as changes to picture databases. Furthermore, future research may be broadened to include the processing of diseased brain tissue.

Amia Hald et al. [58] propose automatically segment grey-scale pictures with an unmonitored dynamic picture segmentation using fuzzy with a genetic algorithm. This method divides a picture into sections using an unmonitored spatial grey-scale image segmentation approach. By combining intensity information with neighbouring connections, this method promises to give reliable picture segmentation. Furthermore, by automatically segmenting the pictures in high quality, the Fuzzy Hopfield Neural Network collecting supports the formation of the population of a genetic algorithm.

Liwg et al. [59] present a novel training-based multi-source mixing structure for segmenting newborn infants' brain pictures, with the goal of assimilating characteristics from random forest multi-source picture extraction. The multi-source in this scenario includes images that are then repeatedly analyzed including possibly refined tissue, such as CSF, WM, and GM. A second examination was carried out on the primary hurdle of clinical image computing and computer secondary participation. The approach to be employed is rather limited, and it needs a large number of training sets and physical segmentation results. For each of the five-time periods, fifty training data are available, and considerable effort is required to achieve manual segmentation; moreover, the limits will be investigated in future studies.

Yunlang Ceai et al. [60] applied detection, structure, and grouping inference for regular repetitive patterns with the pictures. These are caused by repetitive frameworks, color patterns, or repetitive reflections. The segmentation algorithm advocated in this work conformed to the traditional area growing picture segmentation project, which used a mean-shift-like operational procedure to cluster local picture bits into groups. Furthermore, it utilized an uninterrupted combined alignment to correspond to resembling bits and modified the subspace grouping. The outcome of a greater-level group of picture arrangements may be applied to deduce the calculation of things and assess the basic arrangement of a congested area.

Mohmad Awead et al. [61] a genetic algorithm and an artificial neural network were utilized to investigate a multi-component segmentation of images. Various techniques were applied to separate the multi-component pictures. The multi-component picture division technique is performed using a non-parametric unmonitored ANN known as a

self-organizing map (SOM) and mixture genetic algorithm. Subsequently, the image's principal features were recognized using the SOM with no previous information to cluster the picture into homogeneous areas.

Tian Lan et al. [62] suggest using the kernel FCM technique to segment a brain picture. The FCM, spatial FCM, and kernelled FCM methods are all computed using the brain picture, but the accuracy element is verified using the error rate. As a result, the FCM approach has a higher accuracy rate than the other techniques, according to this research.

Agarwal et al. [63] suggest to divide the brain MRI picture into two clusters, GM and WM, they used a bias field correction approach coupled with fuzzy c-means segmentation. Finally, the level set segmentation is performed to these regions, and the results indicated a higher precision than the previous methods, according to this article.

Ping and Honglei Wang [64] this paper presents a modified FCM method for MRI brain picture segmentation. The technique is implemented by integrating geographical neighbourhood information into the standard FCM algorithm and altering each cluster's membership weighting. Both artificially produced and authentic images are subjected to the suggested algorithm. The suggested approach outperforms the traditional FCM algorithm in terms of noise resistance on synthetic images and MRI brain images degraded by Gaussian noise and salt-pepper noise.

Muhammad et al.[65] this paper proposes using convolutional neural networks (CNN) for feature extraction from food pictures. On a publicly accessible Pittsburgh fast-food picture dataset, a linear support vector machine classifier was trained using a 3-fold cross-validation technique. For categorization, features from three distinct, fully linked layers of CNN were employed. Two categorization tasks have been established. The first

job was to sort photos into 61 categories, and the second was to sort them into seven categories. For 61 and 7 class tasks, the best results were achieved utilizing 4096 features with an accuracy of 70.13% and 94.01%, respectively.

Lu xiaojun et al. [66] propose a method that uses attribute fusion to better describe pictures for face recognition using DCNN attribute extraction. They utilized PCA to reduce the combined attribute's capacity. For two classes, the SVM machine classifier is used. This method can detect faces with extreme occlusion, substantial confusion, and size discrepancies, according to test results. On FDDB, this approach achieves an 89% recall rate and a 97% average accuracy, according to the conclusion.

Er-Yang Huan et al. [67] propose CNN-based body constitution recognition technique that can distinguish different types of people's constitutions based on facial images. The determined model first extracted the facial picture characteristics using CNN and then merged the abstracted features with the hue attributes. To get the grouping result, the aggregated information is sent through the Soft-max classifier. They claim that the approach proposed in this study can achieve a precision of 65.3%.

Bhandari et al.[68] propose covariance matrices represent the characteristics of deep convolutional neural networks (DCNN) for face expression identification. The covariance matrices have the same spatial geometry as symmetric positive definite (SPD) matrices. They show that the covariance descriptors generated on DCNN features are more efficient than the usual classification with fully connected layers and soft max by performing facial expression classification using Gaussian kernel on SPD manifold. They demonstrate that the suggested technique achieves performance by using the VGG-face and exponent architectures and doing extensive tests on the Oulu-CASIA and SFEW

datasets. They conclude that the proposed method provides state-of-the-art facial expression recognition performance.

Hong Liang et al. [69] they describe the common methods used in paragraph attribute extraction, then expands on the often-used DL process in paragraph attribute extraction and its implementation, and last, it anticipates the use of deep learning in attribute abstraction. They come to the conclusion that, in comparison to other machine learning techniques, DL can detect complicated relationships from the attribute and train lower-level characteristics from practically unprocessed source data.

Hawaii et al. [70] suggest a deep neural network-based brain tumor classification approach that is fully automated. Glioblastoma sickness pictures of various grades were subjected to the proposed algorithm. The construction of a convolutional network is shown in a unique way.

Matilda Lorentzon [71] machine learning is proposed to extract features for picture selection. Histograms of directed gradients, discrete cosine transform domain features, and characteristics derived from a pre-trained CNN are among the feature extraction methods utilized. For content categorization, the attributes that were excluded from CNN yielded the best results.

Ronjian Li et al. [72] the deep learning imaging data achievement for better brain sickness analysis is the title of their paper. They proposed a deep learning based framework for assessing multi-modality imaging data. Images of Alzheimer's disease from two modalities, PET and MRI, were utilized to assess this method. They came to the conclusion that their technique outperformed previous methods substantially.

Pingping Zhu [73] propose the CNN algorithm to abstract target attributes from the sonar photos. The SVM is implemented in the recognition stage, which was trained based on manually labeled data. The result demonstrates that deep learning attribute extraction provides good achievement compared to applying various feature abstraction approaches.

Heba Mohsen [74]propose brain tumors classification using learning neural networks. The deep neural network is applied to categorize the MRI dataset of 66 brain tumor images. The conclusion is that using the deep neural network algorithm demonstrates great accuracy related to standard classifiers.

Pereira et al. [75] propose an auto distinguish technique to classify cancer utilizing a convolutional network within three 3×3 kernels. The use of small kernels allows designing a deeper architecture, besides having a positive effect against overfitting, given the fewer number of weights in the network. Also, they investigated the use of intensity normalization as a pre-processing step, which though not common in CNN-based segmentation methods, proved together with data augmentation to be very effective for brain tumor segmentation in MRI images. The technique achieved the entire core's initial location.

Siar et al.[76] introduce a convolutional neural network algorithm to identify multiple sclerosis and normal tumors concurrently. The convolutional network successfully classified 96% of the pictures into one of three group.

Szilagyi et al. [77] propose a divide brain tumors from MRI images, recommend using a Fuzzy C-Means algorithm. The proposed approach makes use of prior data at two implementation points: the fuzzy C-means-formed groups of voxels are classified as maybe cancers or non-cancers based on data extracted from train sizes; and the selection

of fuzzy-means factors (amount of groups, fuzzy exponent). They conclude that preliminary evaluation results demonstrated the procedure's capacity to extract abstract information about the presence and location of cancer.

Xu et al. [78] propose an effective and operative way, utilizes a convolutional network for grouping and segmentation. The suggested approach applied Image-Net for abstract attributes. The outcomes achieved 97% and 84% precision for grouping and segmentation, respectively.

Pan et al. [79] introduce a deep learning building, and base neural systems for cancer categorizing by MRI images have been considered and evaluated. The outcomes demonstrate that the system routine grounded on the Sensitivity and Specificity of the convolutional network better by 19% associated with ANN.

Basheera et al. [80] propose a new method that utilizes CNN for categorizing brain tumors into benign and three various kinds—applying an improved ICA composite model to extract the tumor from an MRI photo. After the extracted photo, profound attributes are removed and systematic. The outcomes are evaluated by determining the algorithm's achievement on a data record free with Harvard Medical School.

Menegola et al. [81] propose a scenario transfer learning appears as a prominent solution. The aim is to clarify how transfer learning schemes may influence classification results. Particularly focused in the automated melanoma screening problem, a case of medical imaging in which transfer learning is still not widely used. In addition, explored transfer with and without fine-tuning, sequential transfers and usage of pre-trained models in general and specific datasets They conclude that the experimental design is sensitive to

the image annotation, that is, small changes in the fold assembling can cause huge impacts in the final results.

Bhandari et al. [82] examine the importance of CNNs in segmenting brain tumors by first learning about CNNs and then performing dissertation research to find an example segmentation pipeline. Additionally, look into the future efficacy of CNNs by looking into a new field called radionics. This study looks at the quantitative characteristics of brain tumors including form, texture, and signal strength in order to predict clinical outcomes like the presence of the tumor and treatment response.

Naima Otberdout et al. [83] suggest a correlation matrix is employed to encrypt the DCNN attribute discrimination for facial appearance. The space geometry of the correlation matrix of symmetric is positive. Employ the Gaussian kernel on the SPD manifold to accomplish face expression classification. They demonstrate that characteristics calculated by DCNN are more powerful and efficient than traditional classification methods. They implemented the VGG-face and ExpNet model with large practices on the Oulu-CASIA and SFEW datasets. They conclude that the suggested technique reaches achievement for face appearance identification.

Hashemzeh et al. [84] propose a brain tumor identification, a novel hybrid model combining neural autoregressive distribution estimation (NADE) and CNN has been proposed. Three types of brain tumors were used to test the system with a total of 3064 T1 spin-lattice relaxation time MRIs. The results reveal that the CNN-NADE combination has a substantial categorization, indicating that clinical pictures are difficult to produce.

Seetha et al. [85] suggest brain tumor recognition by applying convolutional neural networks (CNN) classification. Small kernels designed, and neuron weight to obtain more in-depth architecture. The weight of the neuron is specified as small. Investigational outcomes display that the CNN records 96% precision with slight complication and the recent approaches.

Khan et al. [86] present the CNN algorithm and data expansion and image handling to classify malignant and non-malignant brain MRI scan pictures. They compare the scratched CNN algorithm's performance to pre-trained VGG-16, ResNet-50, and Inception-v3 models. The results showed that model accuracy was 100%, whereas VGG-16 was 96% accurate, ResNet-50 was 89% accurate, and Inception-V3 was 75% accurate.

3.5 Conclusion

The brain tumor examination is a precise and difficult task that requires constant attention to precision and correctness. As a result, a thorough technique focusing on the innovative framework for expanding more robust picture cancer segmentation and detection methodologies is desperately needed. This chapter provides an overview of currently recommended techniques for detecting brain tumors using MR brain imaging. This is due to brain tumor detection being complicated and sensitive, precision and consistency will be critical components of the selected approach.

Chapter 4

Brain Tumor Detection System

4.1 Introduction

A brain tumor is considered a dangerous illness, which requires early and accurate detection. However, the principal detection methods depend on radiologists who may apply a misdiagnosis due to human error; furthermore, more effort and a more significant amount of time are necessary to decide on detection.

This research work addresses automated detection of brain tumors and seeks to discover a new but powerful resolution for this. The formal problem definition is listed below:

1. Low classification performance in expressions of accuracy, sensitivity, and reliability.
2. Radiologists required high classification time.
3. Manual classification of medical images is never 100% accurate.

It is indispensable to accept, as a priority, the truth that all radiologists, even highly-skilled ones, do commit errors. The caseload, case complexity, speedy reporting requirements, and radiologist fatigue all play a part in a radiologist's report quality. Different individual factors and limited analysis periods and tools different medical

doctors may reasonably often produce diverse classifications or interpretations, leading to different diagnoses. Brady et al. [3] demonstrate that radiology includes decision making in a situation of uncertainty, and consequently, cannot on every occasion create a successful explanation or opinion. Additionally, in many diseases, doctors and emergency care physicians trust radiology test reports to conclude their patients' diagnoses and the course of their treatment. Furthermore, the image capacity and complex data offered to radiologists for classification have risen dramatically in the last few years.

This thesis suggests a novel brain tumor detection support system technique, which introduces machine learning techniques to classify images more quickly and accurately. The following investigation methodology was performed. The literature on detecting and analyzing brain tumors is studied to describe the problem and study topic. Numerous techniques of brain tumor recognition in MRI was studied, and their performance was evaluated. Moreover, the multi-classifier utilizes the supervised method for the image classification task. They apply different methodologies to classify data with the inspiration of developing the taxonomy of benign and malicious lesions in brain MRI images. Therefore, all of the outputs of the algorithms are combined in a confident approach to accomplish the last decision. New detection and classification of tumors are offered. A confusion matrix weighs the output of the developed technique. The aim is to promote a system that increases disease detection by reducing the false negative and positive rates due to observational oversights

.

4.2 Brain Tumor Detection System Flowchart

The proposed solution introduces decision fusion frameworks that coordinate decisions from multi-classifier into a single decision, as shown in Figure 4.1. The brain tumor detection support system is implemented to classify digital MRI and CT images to detect brain tumors as benign or malignant. The introduced system block diagrams consist of two phases, a training phase and a testing phase. Both phases consist of the following stages: initialization, segmentation, feature abstraction depending on discrete wavelet transform DWT, feature decrease by PCA, multi-classifier, and training accuracy assessment; finally, the decision fusion is the last stage in the testing phase, which is responsible for combining different decisions resulting from the multi-classifier stage.

In the decision fusion, the elementary combining classifier, such as majority voting, minimum likelihood, maximum likelihood, product likelihood, and average probability, is employed to fuse considerable evidences from the classification stage. The Dempster-Shafer theory is applied for the principal technique to fuse multi decision to arrive at one final decision and express any uncertainty regarding this decision.

4.2.1 Training Phase

The first stage is the pre-processing, applying median filter followed by image segmentation, applying the threshold technique, and feature extraction by applying the discrete wavelet transformation DWT technique. The length of the feature vectors is decreased through implementing the PCA method. The set of compact feature vectors and the class label are utilized to learn the multi-classifier group. A cross-validation technique is applied for a successful generalization capability of the system. Besides, the training accuracy assessment offers a confidence score for the performance of each

algorithm in the training phase. This confidence of each classifier will be used later in the decision fusion and allows robust classifiers to participate with greater involvement in the decision fusion task.

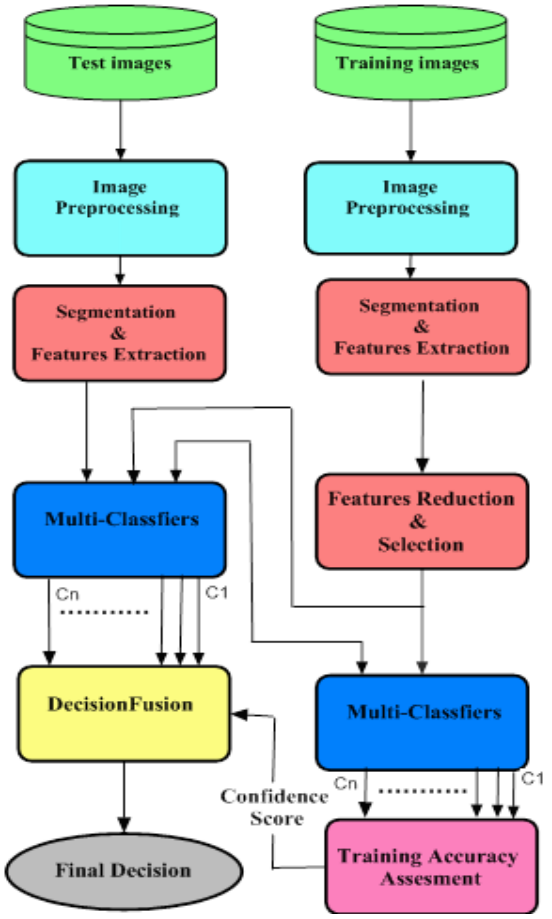

Figure 4.1: Flowchart of brain tumor detection system.

4.2.2 Testing Phase

The user or the radiologist is inputting the brain MRI image, which the supporting system must classify. The segmentation and feature extraction is applied, and the PCA decreases the attribute dimension. This decreased attribute pattern of dimension is applied to the multi-algorithm stage. The multi-algorithm group is a set of algorithms that produce “local” choices by utilizing a suitable decision combination rule to obtain one final trusted decision. The essential stage is the decision fusion, where all the local

decisions are fused into one decision, namely, normal or abnormal (benign or malignant) using different elementary combinations similarly majority voting, weighting averaging, minimum, maximum, and probability schemes. The primarily combination method is Dempster Shafer theory of evidence DST.

4.3 Brain Tumor Detection Support System

The block diagram presented in Figure 4.2 demonstrates the seven phases of constructing the brain tumor detection support system. In the following sections, each stage and the applied techniques will be discussed in detail.

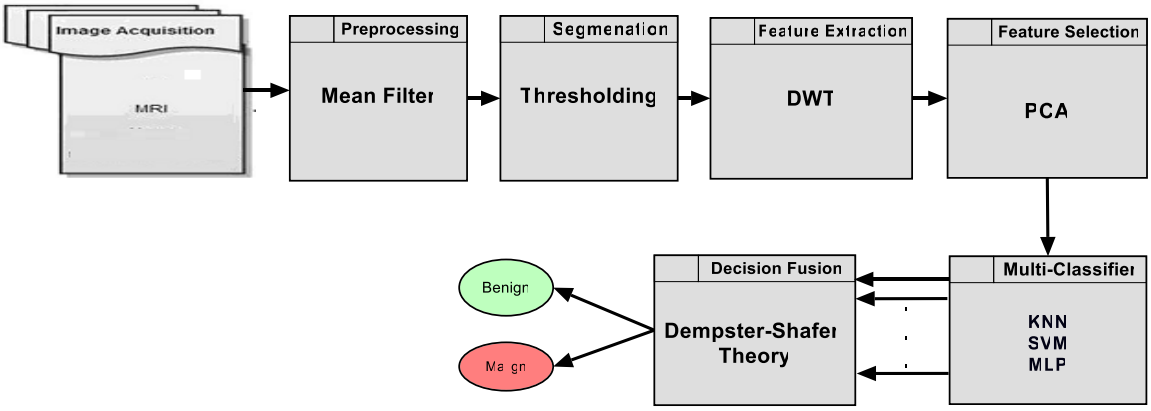

Figure 4.2: System blocks diagram.

4.3.1 Image Acquisition

At this stage, the features are removed from the area of attention of the image and classed as either benign or malignant and are assigned to the extracted features. The features extracted from the normal MRI (class 0) images are assigned as a normal class, and those from the abnormal MRI (class1) images are assigned as an abnormal class. The assigned features are kept in a database. This stored data is essential for the classifier to assign the class to the image under test. The image acquisition steps are presented in Figure 4.3.

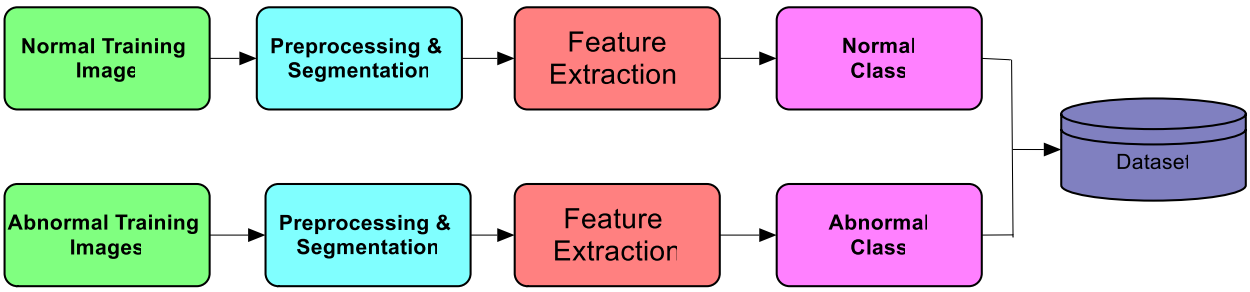

Figure 4.3: Dataset collections and labeling.

4.3.2 Pre-processing Stage

Image initialization is a significant aspect of any photo-based application. In a wide variety of image handling techniques, particularly for the assignment of cancer partition, pre-processing is critical. The following two stages, i.e., the segmentation and extraction stages, produce accurate results. Their high noise levels render raw MRI/CT images collected from laboratories inappropriate for direct use, and a pre-processing stage is necessary for the following aims:

1. Any noise in the image must be reduced.
2. Pre-processing prepares the images for higher-level processing such as segmentation and feature extraction.
3. Marks or labels such as name, date, and other details (film artifacts) that can affect the classification task must be eliminated.
4. Image quality requires to be improved.

The MRI image contains artifacts or marks similar to the patient's name, label, and tags. It is substantial to smooth an image, whereas keeping its edges. The overlap of grayscale is a factor that causes followed phases tasks such as feature extraction, segmentation, and classification to further complications [87]. In numerous biomedical image-processing applications, filtering is likely the generally essential operation, wherever it decreases the

noise scale and increases the fineness of the image. Picture marks are detached using a tracking procedure. We can use a tracking procedure to eliminate marks surrounding the MR photos. Nevertheless, the tracking procedure applies the maximum threshold for the grayscale photo that is 255. Several pre-processing methods are shown in Figure 4.4.

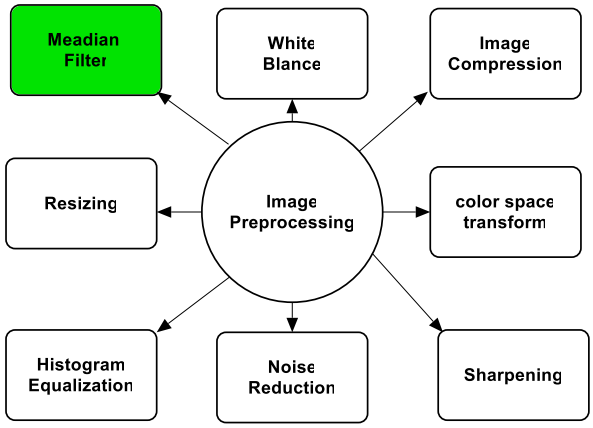

Figure 4.4: Various pre-processing techniques and highlighted the used technique.

4.3.2.1 Removal of Film Artifacts

The MRI/CT images contain film artifacts or labels such as surname, age, and symbols. In such a broad diversity of image-treating requests, it is essential to smooth the photo whereas maintaining its ends. The grey points usually interfere, which causes any of the following stages similarly—segmentation, feature extraction, and labeling more challenging. The tracking algorithm is applied to eliminate film artifacts. Figure 4.5 shows the removal of film artifacts. The tracking procedure is employed to eliminate image artifacts. First, the tracking procedure is applied to take off film artifacts like tags, names, and tags out of the MRI photo. Letter artifacts are present in most brain MRI images due to patient's information being embedded in them. The high quality of MRI

machines ensures metal-related and susceptibility artifacts are very few [88]. The steps applied to perform the tracking algorithm:

1. Read the MRI image and save it into a two-dimension array.
2. Choice the highest threshold value for eliminating white tags
3. Flag value set to 255.
4. Select pixels that are the strength value is equal to 255.
5. If the strength value is 255 now, the pointer amount is set to zero and therefore the tags are detached from MRI.
6. Else, skip to the following pixel.

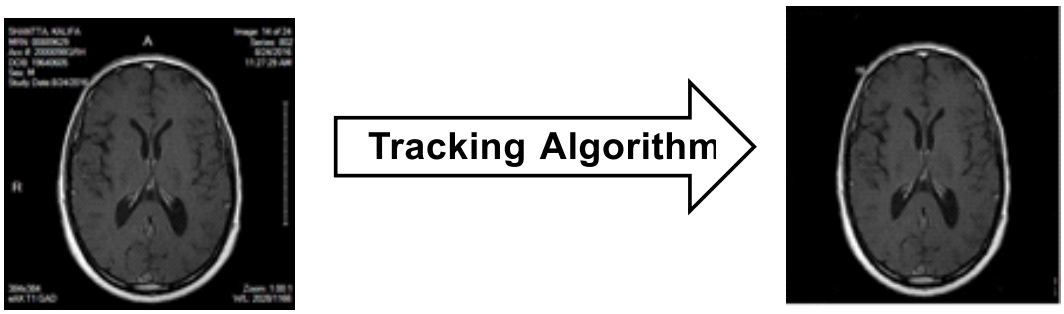

Figure 4.5: Original MRI image (left) after tracking algorithm applied (right).

4.3.2.2 Enhancement

Filtering is possibly the primary procedure in various medical image classification assignments. The function is to decrease the noise level and raise the condition of the image. In image processing, it is regularly desired to implement some noise reduction on an image or signal. One of the nonlinear digital filters is the median filter, which is usually implemented to reduce noise. Median filtering is quite extensively applied in digital image treating because it preserves edges although removing noise[89]. This

method calculates the neighboring pixels' median to define the novel amount of the picture element. A median is computed by arranging picture element amounts wholly by their magnitude, at that point choosing the median amount as the pixel's novel value. Median filtering is useful in reducing noise from two-dimensional signals without blurring boundaries [90]. The median value is employed in its place of the strength value of the midpoint pixel. When the process is repeated, a high-resolution image is obtained.

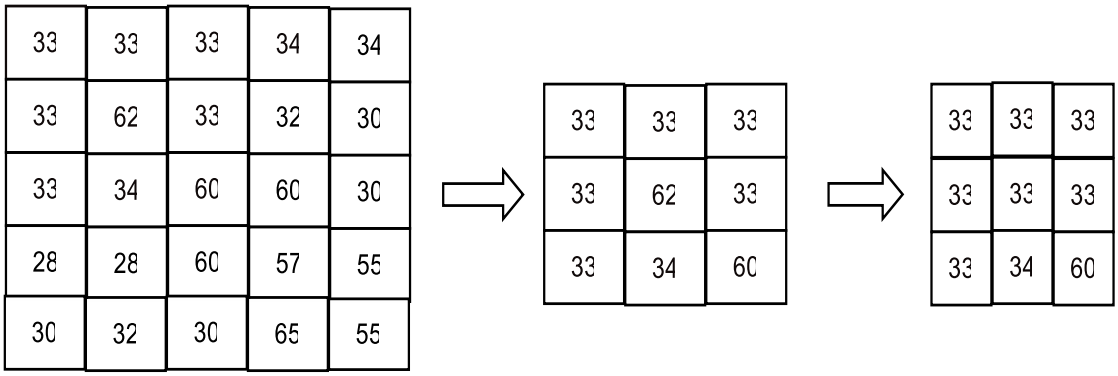

Figure 4.6: Median filtering for 3×3 sliding window on the MRI.

The white pixel in the image foreground represents the tumor and their pixel strength value from zero to 255, where the black pixel with strength from zero to 10 is the background. The disparity of the area is calculated by:

$$C = \frac{f-b}{f+b} \tag{4.1}$$

C is the contrast ,where f and b is the average gray-level rate of the foreground and the background respectively.

4.3.3 Image Segmentation

Image segmentation aims to group a volumetric medical image into partitioned areas, typically into anatomic structures essential for a particular task. This is specifically utilized to separate regions from the remainder of the image to observe or recognize them as objects. The thresholding approaches are applied in the proposed system, and further

details of this technique will present in this section. In the literature, several segmentation techniques are present. However, no one technique can be assumed perfect for all images. In addition, all approaches similarly are not suitable for a specific kind of image. Figure 4.7 shows various segmentation techniques with a highlighted applied method in this research. Thresholding is the manageable and most frequently utilized technique of image segmentation. A binary region map or binary image is obtained with only one threshold after converting a greyscale or color image. The binary map contains two areas, which are possibly separated; the first of these has pixel intensity with input data and is of less value than a threshold background; the second area belongs to the input values which are equal to or exceed the threshold foreground [91].

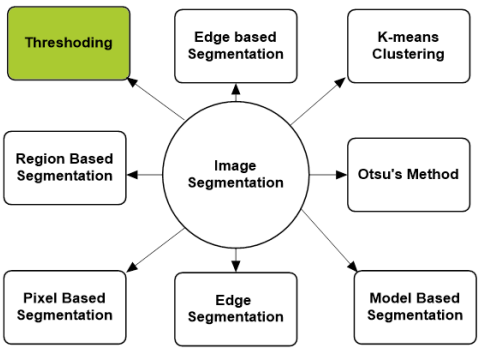

Figure 4.7: Various segmentation methods.

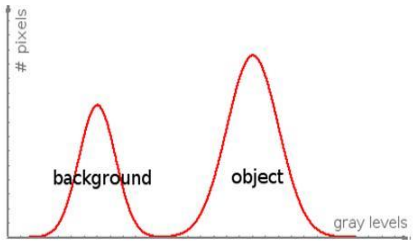

Figure 4.8 : Background and object over grey level.

The distribution of object (foreground) and background over the grey level shown in Figure 4.8 . The threshold is a value in a grey level that splits pixel strengths into binary portions:

$$k(x,y) = \begin{cases} background\ if\ k(a,b) < T \\ foreground\ \ if\ k(a,b) \geq T \end{cases} \quad (4.2)$$

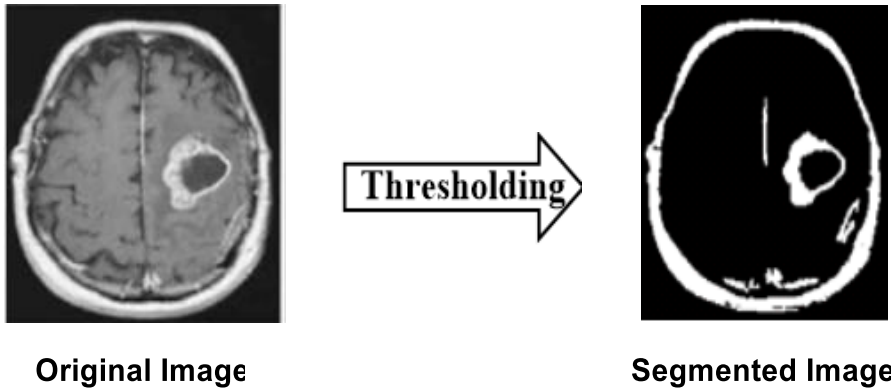

Figure 4.9: MRI before applied thresholding and after.

An incorrect threshold value results in an imperfect segmentation process [92,93]. If more than one region with different grey levels is extracted, it must have more than one multi thresholding. Figure 4.9 demonstrates the original and segmented image after the thresholding method is applied.

4.3.4 Feature Extraction Scheme Using DWT

The process of defining a set of features, or image properties, that will most effectively or usefully represent the information needed for analysis and classification is known as feature extraction. This process substantially affects the classification outcome and shows an integral part in the effectiveness of any image classification. Attribute extraction is likewise a kind of size decrease, which professionally identifies motivating components of an image as a compact feature vector. This technique is beneficial for applications with

large images, for which feature representation must be reduced to enable the fast completion of jobs such as image matching and retrieval. Figure 4.10 shows various feature extraction techniques. This method is used to abstract the attributes from the MRI image. The wavelet transform decomposes a signal into a collection of basic functions. These functions are known as wavelets. The basis can be obtained by applying translations and scaling (stretch/compress) on the mother wavelet ψ(t). Therefore, the wavelet transform offers information in both time and frequency [94].

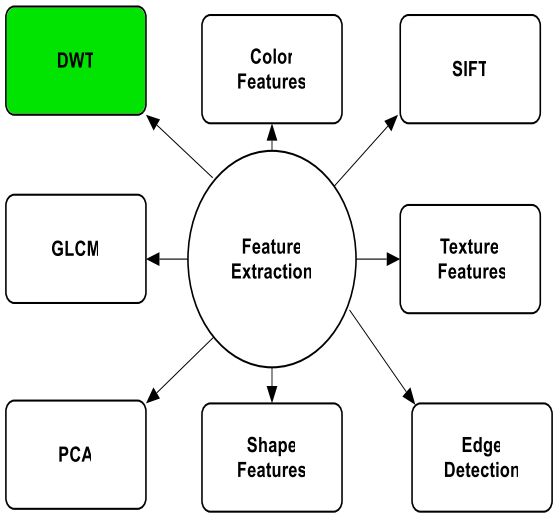

Figure 4.10: Various feature extraction.

Karibasappa et al. [95] state that the major interest of wavelets is their presentation of time and frequency representation, which is particularly useful for classification processes. Suppose $x(t)$ is a square-integrable function, then the continuous WT of $x(t)$ relative to a given wavelet $\varphi(t)$ is defined as:

$$w_{\varphi}(c,d) = \int_{-\infty}^{\infty} x(t) * \varphi_{c,d}(t)dx \tag{4.3}$$

$$w_{\varphi c,d}(t) = \frac{1}{\sqrt{c}}\varphi(\frac{t-c}{d}) \tag{4.4}$$

The wavelet $\Psi_{c,d}$is determined by the mother wavelet Ψ by scaling beside shifting; c is the scaling parameter as well as d is the shifting factor, and together c and d are

constants. Furthermore, DWT is a method that splits the image into some frequency resolutions, and every section with a resolution similar to its scale is studied. DWT is stated as [96]:

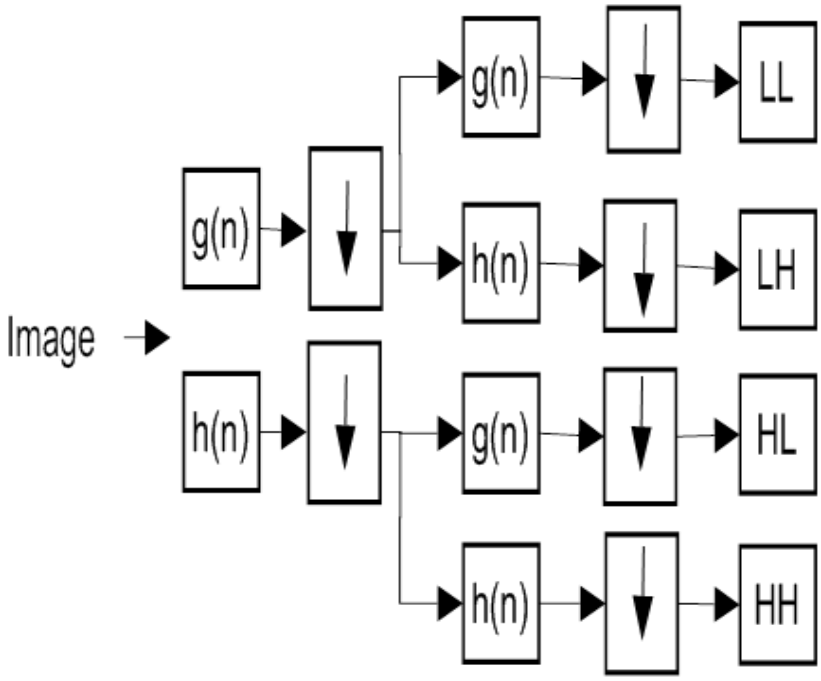

Figure 4.11: 2D DWT block diagram.

$$DWT_{x(n)} = \begin{cases} d_{i,j} = \sum x(m)\, h_i^*(n-2ij) \\ a_{ij} = \sum x(m) g_i^*(n-2ij) \end{cases} \quad (4.5)$$

The detailed components in signal *x(n)* are represented by *di,j*, a term, where the approximation components are by the *ai,j* term. The low and high pass functions are indicated by $h(n)$ and $g(n)$ individually. The terms *i* and *j* indicate the wavelet scale and shifting parameters. The central point of discrete wavelet transform is a various scale illustration from a task. Through applying the wavelets, the particular purpose may be inspected by different scales of resolution. The new photo is a procedure lengthways the

horizontal and vertical paths by applying the high and low pass filters in a different direction. The following result is from low-pass *g(n)* and high-pass *h(n)* filters. First, the LL sub-band represents the approximation coefficients following the execution of the low-pass filter on both the row and column of the image. The LH sub-band is following the execution of the low and high pass filters on the column and row. The HH sub-band following the execution of the high-pass filter both on the column and row, and the HL sub-band following the execution of the high and low-pass filters both on the column and row, respectively. The three groups LH, HH, and HL, represent the detailed coefficients. The sub-band photo LL that represents the approximation coefficients is used to calculate the next level of DWT.

In this work, three levels of decomposition of DWT are applied to a given photo, and the Haar wavelet transform is executed. The simple, as well as more straightforward, wavelet is the Haar wavelet. It is principally a square wave having one period. The image size in phase one is each 128×128, second phase 64×64, and third phase 32×32. Figure 4.12 illustrates the four bands, LL, LH, HL, and HH, of the initial level of DWT decomposition. Half of the signal's size is reduced by half at the individual decomposition stage compared to the earlier period.

Consequently, if the image size is k × k, then the first decomposition stage size is k/2 × k/2, and the next level is k/4 × k/4, and so on. Therefore, the dimension of the calculation element achieved as of the initial level decomposition of a k × k photo is k/2 × k/2, where the subsequent level is k/4 × k/4, and so on. Additionally, as the degree of the divide is extended, a compacted, coarse approximation of the image is achieved.

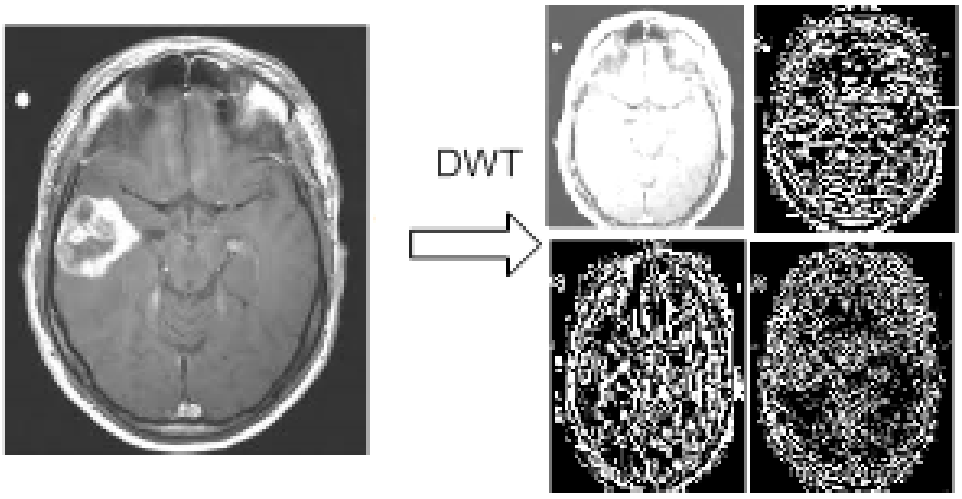

Figure 4.12: MRI image and first level of DWT.

4.3.5 Feature Reduction Scheme Using PCA

A well-known method is principal component analysis PCA for remodeling the current input attribute into a novel and smaller-sized feature space. The PCA is a orthogonal linear conversion that assigns the data to a novel dimension scheme. The highest variation by any projection of the information appears to fall on the initial coordinate first component; the following highest variance comes on the next coordinate second component, and so forth. PCA executes size reduction while preserving as much of the randomness in the high-dimensional space as possible. The input attribute size is converted toward a smaller-sized attribute size utilizing the highest eigenvectors of the correspondence matrix. Given a collection of data, the principal component analysis obtains the linear lower-sized illustration of the information. Thus, the variance of the new information is maintained [97]. The reduction attribute depends on PCA restricting the attribute vector to the elements carefully chosen by the principal component analysis, leading to an effective taxonomy process. Therefore, the principal goal of applying the

primary component analysis is to decrease the DWT coefficients size that consequence in a further helpful and high-performance classification. Following this are the stages required to perform the principal components of the entered matrix containing the DWT coefficients. The dimension of the entry array is reduced to only 13 from 1024. Figure 4.13 depicts the necessary phases for performing the PCA of the entry path.

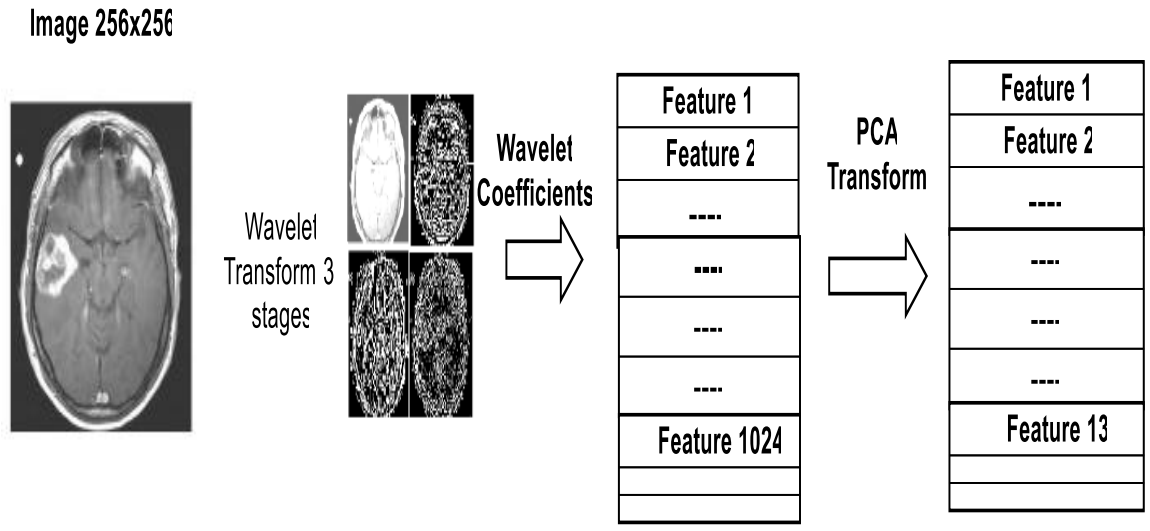

Figure 4.13: Feature extraction and reduction scheme.

Consequently, the attribute abstraction procedure was produced by executing two actions; initially the wavelet elements were obtained by applying the discrete wavelet transform, and following this, the significant coefficients were compact by the principal component analysis.

4.4 Supervised Classification Algorithms

Supervised learning indicates that the model learns from previously identified outcomes and changes their internal factors to fit themselves to the input information. As the paradigm is appropriately learned, it can give precise guesses around invisible or unseen information. During supervised learning, the approval of the training is dependent on

patterns that carry output labels. When the output values identify the various classes to which the samples relate, the training assignment is called classification. The classifier factors have described the training data consistently because they have been collected from the model data. The training dataset to be given to the separation classifier for constructing the classifier consists of N data points that can be formally denoted as {xi, yi}Nn = 1, with xi $\epsilon\mathbb{R}d$ being an input example of dimension d and yi $\epsilon\{-1, +1\}$, the corresponding class label for a two-class classification task. The following sections describe different algorithms that can produce classifiers for MRI and CT data. In this thesis, three algorithms KNN, SVM, and MLP, are applied to solve the tumor classification problem. Some of the classifiers are shown in Figure 4.14.

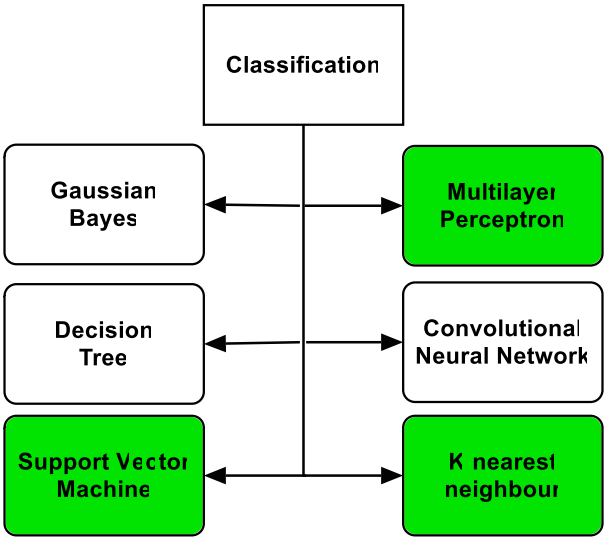

Figure 4.13: Some of image classifiers and highlight the used classifiers.

4.4.1 K-Nearest Neighbours

The k-nearest neighbours algorithm is grounded on a distance function that measures the correlation or differences among data points [98]. A distance function frequently used for this determination is the Euclidean distance:

$$d(p,q) = \sqrt{\sum_i^n (q_i - p_i)^2} \tag{4.6}$$

In the nearest neighbour rule, a novel testing point x is consigned to the class of the data point xi closest to x Figure 4.15. This approach might produce an unreliable classification result since the classification rule assigns the new data point based on only a single N point, which might, for example, be unrepresentative of its class. When k neighbours are considered rather than one, a more reliable result is possible. The k-nearest neighbours algorithm determines the k data closest to x and classifies them according to the majority of equal classifications in this group. The use of data normalization can avoid the variable that has the most significant scale dominating the distance measure. The advantage of the KNN algorithm is easy to use and does not create supposition regarding the information.

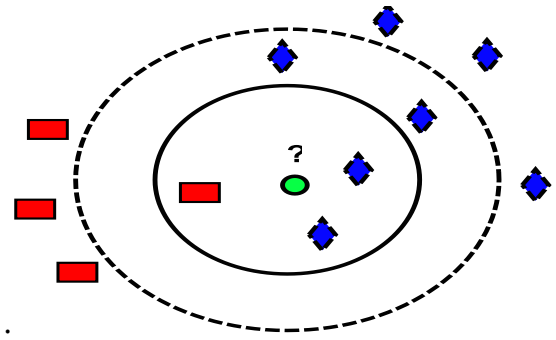

Figure 4.14: Three nearest neighbours.

4.4.2 Multilayer Perceptron

The MLP is a primary category of ANN model, known to be a universal approximates [99]. In the case of one hidden layer, the outcome is achieved by feeding a data point $x(\epsilon \mathbb{R}^d)$ to the hidden layer, as illustrated in Figure 5.16. Also, Figure 4.17 shows a

mathematical model of a discrete perception. Wherever the weight denoted by *w*, the output and the input is indicated by y and *x* , respectively. In the case of MLP with one unseen layer the outcome is achieved by input the data point $x(\in \mathbb{R}^d)$ to the hidden layer, consisting of H hidden units, which are themselves connected to the outcome unit(s).

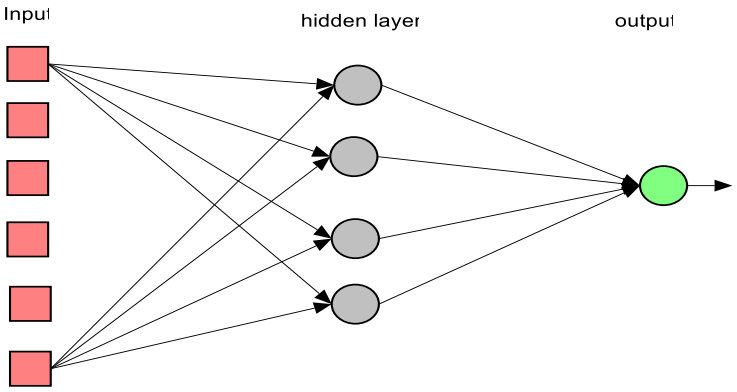

Figure 4.16: Architecture of an MLP, containing four hidden units and one output.

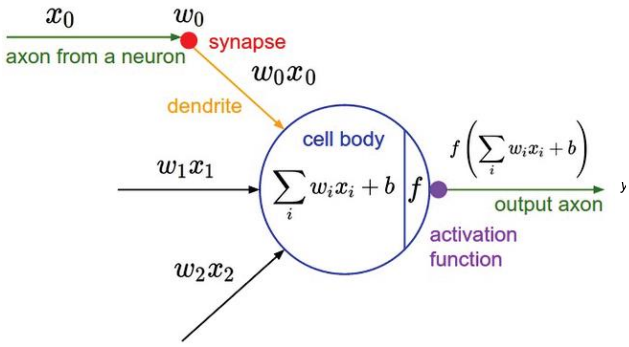

Figure 4.17: Mathematical model of a discrete perceptron or neuron.

For hidden neuron *h* this can be represented by:

$$z_h = f\left(\sum_{j=1}^{d} w_{jh}x_j + b_h\right), h = 1, \ldots, H \tag{4.7}$$

with *f*(·) an activation function (e.g. tanh), w_{jh} the weight among the input *j* and unseen neuron *h* and *b* the bias term of unseen neuron *h*. The output unit activation $y_{k'}^*$ is modelled as:

$$y^*_{k'} = \mathrm{g}(\sum_{h=1}^{H} w_{hk'} z_h + b_{k'}) \quad (4.3)$$

where $w_{hk'}$ represents the weight between unseen neuron *h* and the output unit, *b* the bias expression of the output unit and $g(\cdot)$ the activation function. For dual taxonomy tasks, there is one output unit, $k' = 1$, and $g(\cdot)$ can be represented by a logistic sigmoidal function. The perceptron learning process follows these steps [100]:

1. Each weight is multiplied by the input, and computes the total sum. The backpropagation is one of techniques applied to tune the weights.
2. The activation function could the output up and down by this technical step. . It causes likely to adjust the value yield of the neuron.

The stages of neural network learning:

1. Initialization: primary weights are implemented to all the neurons.
2. The inputs of the learning set are brought over the neural network and an output is calculated.
3. The goal of backpropagation is to modification the weightiness of the node to reduce a mistake to the smallest possible value.
4. Weight update: weights are altered to the best values based on the outcomes of the backpropagation procedure.

The paradigm is set to create estimates for anonymous input information. New data can be entered into the paradigm, a forward pass is done, and the paradigm produces its guess.

4.4.3 Support Vector Machine

The linear SVM algorithm attempts to discover between all hyperplanes that reduce the learning fault in the distinguishable information. This line splits the learning information

with outer space from their nearby points (the max distance plane). SVM classifier is a well-known non-probabilistic algorithm designed to seek linear separability of the classes. The most excellent significant training points are the support vectors that state a hyperplane. The central concept underlying SVM classifiers, which were developed principally for two-class classification tasks, is using a hyperplane technique for determining the decision limits for the classification of the data points of various groups. The records are linearly divisible in such a situation, and the separating hyperplane can be determined for various directions Figure 4.18. The kernel SVM algorithm is considered for non–linear divisible information. The elementary idea includes two sequential phases: plan the input attribute space into a greater size attribute space, by a non-direct conversion kernel. The mapping is executed to catch a attribute domain, wherever the information can be lined separate [101]. In that attribute domain, the system builds the maximal boundary hyperplane as defined earlier. However, SVMs work to maximize the boundary, and the goal is to build a hyperplane with a maximal length separating both classes. The formulations for an SVM classifier are as follows:

Data: $<\mathbf{x}_i, y_i>$, i = 1,..,I

$\mathbf{x}_i \in R^d$, $y_i \in \{-1,+1\}$

planes in R^d are parameterized by a vector (w) and a constant b, which This may be able to be represented as $x_i.w + b = 0$.

The hyperplane h can be stated as:

$$w^T \varphi(x_i + b \leq -1 \; for \; y_i = -1 \tag{4.8}$$

$$w^T \varphi(x_i + b \geq +1 \; for \; y_i = +1 \tag{4.9}$$

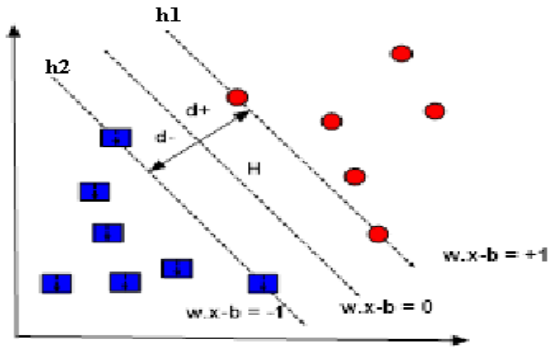

Figure 4.18: SVM representation for linearly separable dataset

Equivalent to:

$$y_i(w^T\varphi(x_i) + b) \geq 1, i = 1, \ldots, N, \tag{4.10}$$

In addition, the algorithm is represented as :

$$y(x) = sign(w^T\varphi(x) + b) \tag{4.11}$$

But the data of classes are overlapping in the most real life application, which causes a good linear sorting out of question. Consequently, a limited feature of the wrong classification would be permitted about the border. The resulting optimization issue, where the infringement of the restricted is punished, for SVMs is formed as

$$min_{w,\xi,b}\mathcal{J}_1(w, \xi) = \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_{i,} \tag{4.12}$$

such that

$$y_i(w^T\varphi(x_i) + b) \geq 1 - \xi_{i,} \quad i = 1, \ldots, N, \tag{4.13}$$

$$\xi_i \geq o, \quad i = 1, \ldots, N, \tag{4.14}$$

C is soft margin cost function parameter that regulates the effect of individually separate support point; this procedure includes trade-off fault punishment for stabilization. SVMs are established on the standard of basic hazard minimization that matches paradigm complication and experimental error. The primary group of limitations agrees Eq.4.10,

while another group requires slack variables ξ_i , afford the wrong classification. The amount of ξ_i specifies the space of x_i corresponding to the decision border:

$\xi_i \geq 1$; $y_i(w^T\varphi(x_i) + b) < 0$ indicate a different sign between the decision purpose and the goal, showing t x_i is wrongly classified.

1. $0 < \xi_i < 1 : x_i$ is appropriately categorized, nevertheless, falls interior of the boundary.
2. $\xi_i = 0 : x_i$ is properly categorized and falls exterior of the boundary. Normally, the problematic statement in Eq.4.9 and 4.11 is mentioned as original optimization problematic. Consistently, the optimization problem for this algorithm could be formed in the double dimension by the intermediary of the Lagrangian with Lagrange multipliers $\alpha_i \geq 0$ for the primary set of limitations, Eq.(4.10).

A quadratic programming problem applied to obtain the resolution for the Lagrange multipliers. Lastly, the SVM algorithm can represent in this form:

$$y(x) = sign\left(\sum_{i=1}^{N} \alpha_i\, y_i K(x, x_i) + b\right) \quad (4.15)$$

anywhere the kernel function $K(\cdot,\cdot)$ it is symmetric and all its values are positive. It fulfils Mercer's condition formerly, significance that $K(x, x_i)$ equals $\varphi(x)^T \varphi(x_i)$. It is frequently named the kernel trick as no obvious building of the mapping $\varphi(\cdot)$ is required. This allows support vector machine to effort in a high-dimensional attribute space, short of really execution. Several categories of kernel purposes:

1. Linear SVM: $K(x,y) = x^T z$.
2. Polynomial SVM of degree$d$: $K(x, y) = (\tau + x^\tau z)^d, \tau \geq 0$.
3. Radial basis function RBF: $K(x, y) = \exp(-\frac{\|x-z\|_2^2}{\sigma^2})$.

where,

1. ‘σ’ is the variance and hyper parameter

2. $||X_1 - X_2||$ is the Euclidean (L_2-norm) distance among two points X_1 and X_2.

Equation (4.12), α_i should be above all nonzero values, support points, in it is place of exclusively data points. The identical trajectories x_i are mentioned the as support points. These support vectors are positioned near the resolution border and share in the building of the splitting hyperplane [102].

4.5 Multi-Classifier Architecture

In image processing, classification is crucial since it influences the quality of image interpretation. Many existing approaches produce acceptable results but are limited to a certain image type or need previous knowledge that is frequently unavailable. If the classification outcomes of many algorithms can be efficiently combined, such algorithm fusion can be preferred to the results of all individual algorithms. The main principle underlying algorithm fusion is that while obtaining a choice, one should not depend just on one classifier's perspective; instead, algorithms should collaborate in decision production by merging their outputs. As a result, the most difficult challenge to solve when fusing various algorithms is solving disagreements among them. The challenge is figuring out how to integrate the findings of various algorithms to achieve a superior outcome.

Multiple classifier systems can belong to one of the following configurations: cascaded serial, sequential), parallel, or hierarchical, as illustrated in Figures 4.19, 4.20, and 4.21 [96]. In a cascading architecture, as shown in Figure 4.19, the yield of one classifier is utilized as entered to the following classifier. The latest algorithm in the chain produces

the final prediction. The disadvantage of this configuration is that the final classifier is unable to correct a previous one. In a parallel configuration, all classifiers independently generate results combined in a unified output, as indicated in Figure 4.20. The selection of a suitable combination method contributes to the performance of this topology.

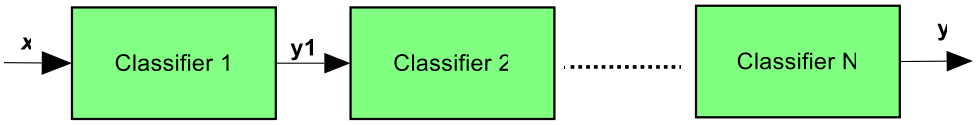

Figure 4.19: Sequential classifier combination.

This feature becomes significant if the base algorithms have varied performance so that low-performance classifiers can disturb the total performance of the multi-algorithm. In this research, this configuration is applied when multi-classifier are used.

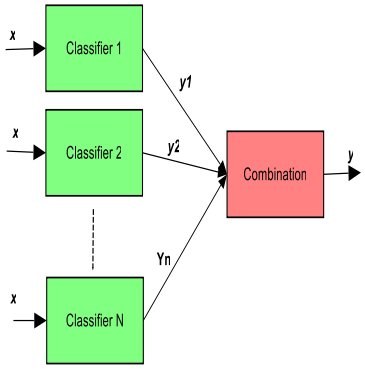

Figure 4.20 : Parallel classifier combination.

In a hierarchical technique, to achieve ideal performance, together cascading and parallel formations are employed to fuse the algorithms, as depicted in Figure 4.21. This method has the advantage of overcoming the drawbacks associated with cascading and parallel configurations.

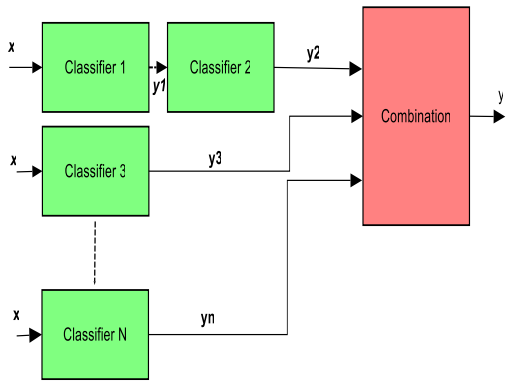

Figure 4.21: Hybrid classifier combination.

4.6 Levels of Fusion

One of the most critical considerations in the combination process is the type of data that should be merged. Data combination can generally occur at two levels: before the classification step, feature levels, and after the classification step, at the decision level. The several popular information fusion and combination approaches are listed in table 4.1.

4.6.1 Feature Fusion

The combination of data from different references might happen at the data level or the attribute level. Data level fusion combines information from several sources before subjecting it to attribute removal or selection. This assumes that the similarity and connection between raw data points are either known in advance or can be accurately inferred. This is known as pixel level fusion in the image processing literature. Combining distinct feature sets taken from various sources is related to feature layer combinations. In reality, raw data, such as hyperspectral data, contains the most knowledge content. Additional processing, in particular attribute removal and choice, lowers the quantity of data accessible to the combination procedure.

Combination technique	Application
Dempster-Shafer theory of evidence	Decision making, beliefs intervals
Bayesian theory	Decision making between multiple hypotheses
Neyman-Pearson criteria	Decision making
Pixel level fusion	Image processing, image segmentation
Neural Network	Signal interpretation
Fuzzy logic	Handle vagueness
Markov random field	Image processing
Knowledge based system	Pattern recognition

Table 4.1 : Common data fusion and integration techniques.

A unique resultant pattern vector can be computed as a weighted mean of the particular pattern vectors where the attribute sets are homogenous, features of the same kind of data. When feature sets are heterogeneous, such as attributes derived from various types, connecting them to produce a singular pattern vector is possible. Attribute removal and choice techniques are utilized to decrease the capacity of the resultant attribute vector. If the feature sets are inconsistent, the connecting is not available. A feature-level representation offers several benefits over a data layer representation. Fusion procedures that integrate information early in the processing are more successful than combination processes that later integrate information. It is important to remember that feature-level fusion might combine the same raw data sets or multiple data sources reflecting the same photographed scene. Furthermore, feature level fusion should be simple if the attribute sets are developed from the equal attribute removal or choice technique used to the identical information. Nevertheless, attribute layer combination is challenging if the

attribute sets are obtained from various methods and data sources. The following factors contribute to the difficulty. For starters, the connection between attribute scopes produced from various procedures may be unknown. Attribute groups from various origins may be inconsistent. Supplementing them to produce a new feature group is viable if the attribute groups are feature vectors of constant length. However, this series procedure may end in the curse of dimensionality, in which growing the number of attributes degrades system performance, mainly when the quantity of training examples is insufficient. Also, a dimensionality reduction technique can be used with a feature concatenation scheme. Before initiating the classification/matcher process, the feature extraction or selection method guarantees that increased or associated feature amounts are discovered and discarded. This is most likely one of the most significant advantages of doing fusion at the feature level. The substantial variations in the range and the shape (distribution) of the different feature vectors are another issue produced by the concatenation process. This challenge might be handled by using the feature normalization process that moves feature values into a common domain. The mean and variance of feature data are modified using a transformation function. A proper normalization scheme can also be used to handle outliers in feature values, essentially the min-max and median normalization methods. On the other hand, feature normalization may not be essential if the feature values from many origins are by now similar.

4.6.2 Classification Fusion

The fusion after classification can be divided into four categories: dynamic classifier selection, fusion at the decision level, fusion at the rank level and fusion at the match score level. The following section will discuss some of these strategies of combination.

4.7 Decision Fusion

Any decision support system (DSS) aims to develop a model to make proper judgments with a small quantity of input data/information. Quite frequently, especially in medical-sensitive systems, the accuracy of the judgments made is essential. In such circumstances, the minimal information restriction is less relevant as long as the final judgement can be reached in an acceptable amount of time. According to one viewpoint, DSS advancement should be built on the continual development of current techniques and the discovery of new ones. Another viewpoint indicates that when the limitations of current individual techniques are reached and it becomes difficult to produce a better one, the answer to the problem may be as simple as combining existing high-performing methods in the hopes of achieving more significant outcomes. In terms of reducing uncertainty, such information fusion appears to be worthwhile. Each technique generates its own set of mistakes, not to mention the possibility that the input data is incorrect or incomplete. However, different techniques working on different data should create different mistakes. Given that all individual methods function well, a combination of many such experts should minimize total classification error and, as a result, emphasize correct conclusions. In recent years, information fusion approaches have been extensively researched, and their application in the classification area has been extensively explored.

Fusion Algorithm approaches can be clustered into two groups: dynamic classifier selection and classifier fusion grounded on combination by the rank level and combination grounded on combination by the resulting level. In this research, the combination approaches that are grounded on fusion at the decision level were

implemented. A hierarchy of approaches used in combining classifiers is shown in Figure 4.21.

In addition, there are many techniques for classifier fusion. Two significant approaches can obtain the concluding decision:

1. Integration (or combination): totally, algorithms participate to the concluding result, supposing competitive algorithms. This approach used in our decision fusion stage.
2. Selection: in this category only one algorithm, produce the result for every feature set. It considers that algorithms are complementary.

Besides, the form of information created by the members of classifiers can be assorted into three stages [103]:

1. Abstract: every classifier outputs the class tag for every input outline. Every algorithm *Di* yields a class tag $rj \epsilon \Omega$,j =1,...,M. Therefore, for any object $x \in \mathbb{R}^n$ to be grouped, the M algorithm outcomes describe a trajectory $r = [r_1, \ldots\ldots, r_M]^T \in \Omega^L$; $\Omega = \{w_1, w_2, \ldots, w_c\}$ is a collection of target class.
2. Rank: every algorithm yields a ranking list of likely labels for every input vector. The output of individually D_j is a subgroup of Ω, with the option in ascending order of plausibility of being the exact tag
3. Measurement: individually algorithm outcomes a score, likelihood or sureness level for every input form. Every classifier D produces a c-dimensional.

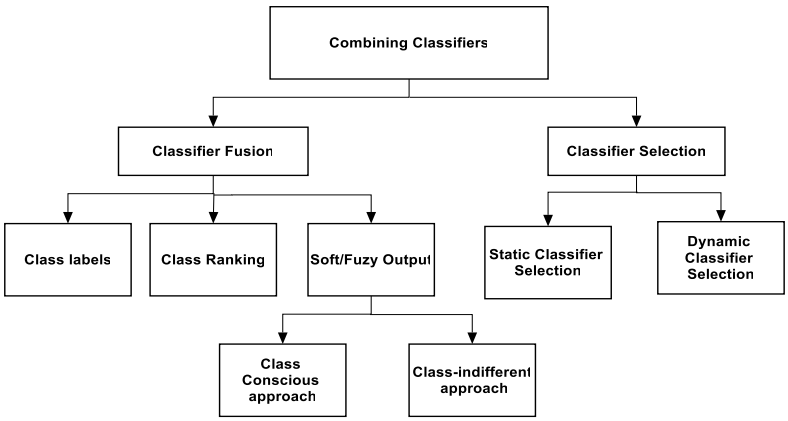

Figure 4.22: A hierarchy of methods used in combining classifiers.

The objective of all decision fusion systems is to create a model, which given a minimum amount of input data/information, is capable of producing proper decisions. Figure 4.22 illustrates the applied techniques utilized to combine the multi-classifier based on the classifier output type. For instance, if the algorithms' outcome of rank or abstract kind applies, specific fusion techniques may be applied, similarity majority vote, weighted majority vote, and Bayesian fusion. However, the probability schemes such as Min, Max, product, and the average probability and the scale of decision fusion organizations have been classified as shown in Figure 4.23. There are two kinds of classifier fusion planning: classifier fusion and classifier selection [104]. In classifier combination, each algorithm is given all data on the attribute domain, and the outcomes from various algorithms are fused. Each algorithm participates in making the last result. In contrast, in algorithm selection, each algorithm is specialized in an exact domain of the attribute domain, and the local professional only chooses the outcome of the ensemble. Dempster-Shafer theory could be applied if the classifier output from the measurement type applies.

In algorithms combination models, classifier outputs are fused to obtain a set conclusion. The majority vote, likelihood schemes, weighted averaging, and Bayes method are used for classifiers fusion. The DST is employed to define the calculated trust of algorithms,

SVM, K nearest neighbour, and ANN. The classifiers output is used as the source of proof for calculating the beliefs.

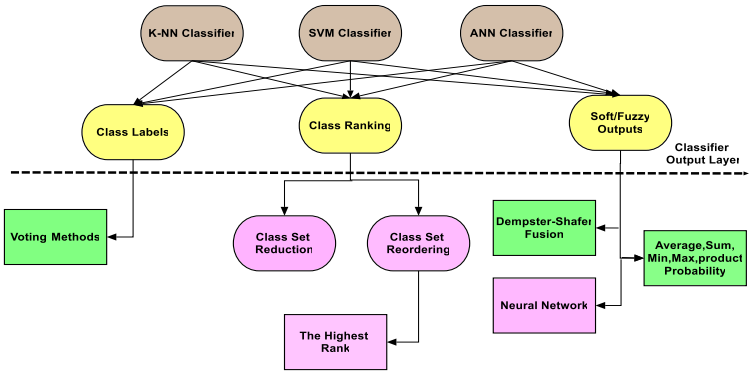

Figure 4.23: Fusion techniques category.

4.7.1 Elementary Combiners

In this research work, the combination rules such as Min, Max, average, and product probability with the majority vote rule are applied to fuse the multi-classifier as a first technique [105,106,107]. The decision on the majority vote applies when all of the classifiers vote for one class or more than 50%, plus one of the classifiers vote for the same class. Therefore, in order to formulate the concluding classification, decision outputs from every classifier were combined. Following this, there is a majority vote regulation that fulfills:

$$R_r(A) = \sum_{j=1}^{k} d_{c,j}\,(A) \tag{4.16}$$

Where c is the label, A is the testing pattern vector, j = 1, . . . , *K*, *k* amount of algorithms and the selection of an odd number to avoid a tie in the majority vote method; $d_{c,j}$ is the paired resolution value [0, 1], 0 matches the incorrect classification and 1 the true classification.

Ponti Jr [108] shows that a combination of classifiers on the measurement level of all these rules can be applied:

1. Min: from among the classifiers, computes the minimum score of each class and sets the unknown testing attributes to the class, which has the maximum grade.

$$R_c^{min}(A) = \min p\left(\frac{c}{A}\right). \quad (4.17)$$

2. Max: from among the classifiers discoveries the maximum result of every class and classifies the unknown sample to the class, which has the maximum grade between the maximum score.

$$s_{R_c}^{max}(A) = \max p\left(\frac{c}{A}\right). \quad (4.18)$$

3. Product: multiply the scores created from every classifier and set the class label of the maximum score to the unknown input attribute.

$$R_c^{product}(A) = \prod_{i-1}^{k} p\left(\frac{c}{A}\right). \quad (4.19)$$

4. Sum: adds the grade created by every single classifier and sets the class label of the maximum result to the unknown input attribute.

$$R_c^{sum}(A) - \sum_{i=1}^{k} p\left(\frac{c}{A}\right). \quad (4.20)$$

4.7.2 Majority Voting

The majority vote is one of simplest combinations rules, which works upon the abstract level. Kitler [109] have verified that in majority voting, the decision indicates a class whichever of the subsequent declarations is measured:

1. All of the classifiers vote for one class.
2. Half the number of classifiers plus one vote for one class (simple majority).

This collects most of the classifiers, which vote for one class even the total number of votes did not exceed half. Majority voting is an ideal fusion technique under the small

hypotheses that the total of classifiers, *R* is odd in order to avoid a tie situation and independent of the classifier outputs.

4.8 Dempster-Shafer Theory of Evidence

The DST was implemented to manage ambiguity management and incomplete reasoning. Dissimilar to the Bayes method, the DST can model unknown information [110,111]. The accumulation of proofs is utilized to diminish a set of hypotheses. The DST method permits the demonstration of uncertainty due to the ambiguity of the proof. If ambiguity becomes zero, the DST model is identical to a typical Bayes paradigm [112]. *X* is represented by basic belief *m(X)* delivered by the source of evidence under consideration and has the next features:

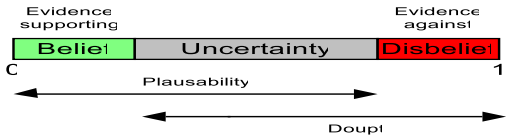

Figure 4.24: Different Measurements over a unit Interval.

1. The green area over the unit interval represents the belief.
2. The red area represents the disbelief.
3. The grey area represents the ignorance, which indicates that neither the belief nor the disbelief range is selected.
4. Plausibility, Pls (B) is the addition of uncertainty and belief measures. This indicates that most stretches of belief are in ignorance, and do not demonstrate trust nor disprove it.
5. The doubt is entire uncertainty and disbelief.

$$\sum_{x \subseteq \Theta} m(X) = 1 \qquad and \quad m(\emptyset) = 0 \tag{4.21}$$

When ϕ is empty, it shows the certainty that an empty set is always equal to zero and that Θ characterises the entire frame of discernment. The trust function for the occurrence of D is specified by:

Bel *(D)* = $\sum m(X)$, $X \subseteq D \text{ and } D \subseteq \Theta$

DST manages the opinions' rating from a total number of proofs or evidence and their fusion. In this research, three experts provide the evidence: the group of classifiers, which we call: E1, E2, and E3. These three classifiers represent the evidence, which offers measures for the event space. Let the event space be Θ = {B, M}. In our case, benign or malignant represents the event space.

Furthermore, there are belief and uncertainty measures for each event. These measures contain a belief for each occurrence and ambiguity. Therefore, BelE1 (M), BelE1 (B), and BelE1 (uncertainty) are measures produced from evidence E1 where:

BelE1 (M) + BelE1 (B) + BelE1 (uncertainty) = 1; similarly for E2, and E3 measures.

This explains how the DST is applied to this work, assuming that a person has a brain tumor and that there are two possible diagnoses, the patient has either a malignant or a benign tumor. In this case, the total outcome of the event space is four opportunities which are Θ = {M, B,[M, B], ϕ}. This indicates that the patient is either M (malignant), or B (benign), whereas {M, B} is one of two classes, either malignant or benign that is a symptom of ambiguity, or ∅ is neither malignant nor benign which demonstrates an unusual case. The sources of evidence are an ANN, KNN, and SVM classifier. The framework of event space is defined as Θ, and every classifier, D_i, would participate in its result by determining its opinions along Θ.

The classifier output provides the evidence in this work, and the Dempster-Shafer combination rule can process this evidence. The source of proof is not dependent; neither does the intersection set presuppose the empty rule. Dempster's fusion rule may be used to fuse any two beliefs such as Bel1 and Bel2 to produce a unique confidence function. Dempster's rule of fusion is a technique, which fuses proof from several independent sources. Furthermore, the probability mass functions are fused employing the Dempster decree, on the assumption that Bel A and Bel B are two belief assignments over the event space Θ, with probability masses m_A and m_B, respectively. Therefore, the total possibility mass proposes that c is:

$$m(c) = K \sum_{a_i \cap b_j} m_A(a_i) * m_B(b_j) \quad (4.22)$$

where, K is the normalizing factor. This function is called the orthogonal sum of Bel_A and Bel_B, indicated as $Bel_A \oplus Bel_B$. This sum can also be denoted as $m_A \oplus m_B$ which is

$$m(c) = K[m_A(a_i) * m_B(b_j)] \quad (4.23)$$

The normalising factor, K assists as a amount of the contradiction among the two certainty functions which is specified by[113] :

$$K = \frac{1}{1-k} = \frac{1}{1-\sum_{a_i \cap b_j}[m_A(a_i * m_B(b_j)} \quad (4.24)$$

Furthermore, *k* is the so-called amount of conflict between the two belief functions. If Bel_A and Bel_B do not conflict, then $k = 0$. If $K = 1$ then the functions totally contradict and $Bel_A \oplus Bel_B$ does not exist [114,115].

This conflict factor will be monitored during the evidence combination, and if it exceeds the preset levels, it will contradict the judgement, indicating that more testing is necessary. In most cases, a proposition has more than two sources of proof. Dempster's

Rule is used to pairs of functions repeatedly in order to fuse multiple certainty functions. There are three belief functions in this work:

Bel_{ANN}, Bel_{SVM}, and Bel_{KNN}.. The final sum is $Bel_{ANN} \oplus Bel_{SVM} \oplus Bel_{KNN}$.

Nevertheless, the order of fusion is of no consequence give the calculated mathematics of the rule[116,117]:

$$Bel(C) = \frac{\sum_{Ai \cap Bi = C; C \neq \emptyset} Bel(Ai) \times Bel(Bi)}{1 - \sum_{Ai \cap Bi = \emptyset} Bel(Aj) \times Bel(Bj)} \quad (4.25)$$

The approach of pairwise fusion is used. It combines, for example, the views of KNN classifiers (K) and supports the vector machine(S) method in the first step. The output of the final combination of the SVM and KNN classifiers is fused in the following stage, with the evidence provided by the ANN classifier (A). Let $bel_{KNN(B)}$ and $bel_{KNN(M)}$, represent the beliefs from the KNN classifier for both classes as benign (B) and malignant (M). Likewise, for the SVM classifier, evidence is given as $bel_{SVM(B)}$ and$bel_{SVM(M)}$, where U_{SVM} and U_{KNN} are the unbelief or uncertainty of the two classifiers. Bel (M) is a trusted mass specified to classify a malignancy. This is calculated by a product of benign trust of SVM and KNN, considering the independence of the pieces of evidence sources. The multiplication of the benign belief of KNN and the uncertainty of SVM and the uncertainty of KNN and the benign belief of SVM is added, and all these essential opinions are summed. Therefore,

$$bel_{comb(B)} = bel_{SVM(B)} \times bel_{KNN(B)} + U_{SVM} \times bel_{KNN(B)} + bel_{SVM(B)} \times U_{KNN} \quad (4.26)$$

4.8.1 Steps of Combination

Dempster's combination includes an estimation of belief and unbelief or uncertainty resulting from each classifier. Figure 5.25 illustrates the inputs and outputs of each

classifier. Output ‘K’ indicates the belief values obtained from the k-Nearest Neighbour, while, ‘S’ indicates the belief values from the support vector machine, and ’A’ indicates the belief values from the artificial neural network. The first phase combines evidence from the KNN classifier and supports the vector machine classifier outcomes. The uncertainty and beliefs are applied to Dempster’s rule as input. S and K are the evidence, which provide beliefs from the support vector machine and the KNN classifier respectively. The beliefs $bel_{KNN(B)}$ and$bel_{KNN(M)}$, where bel_{KNN} indicate the belief provided by the KNN and two classes, benign (B) and malignant (M) under study. Likewise, the support vector machine classifier beliefs are given as $bel_{SVM(B)}$ and $bel_{SVM(M)}$. The uncertainties for two classifiers are U_{KNN} and U_{SVM}; benign and malign classes are achieved from SVM and k-nearest neighbour classifiers respectively. Figure 4.26 shows combinations of SVM and KNN Classifiers.

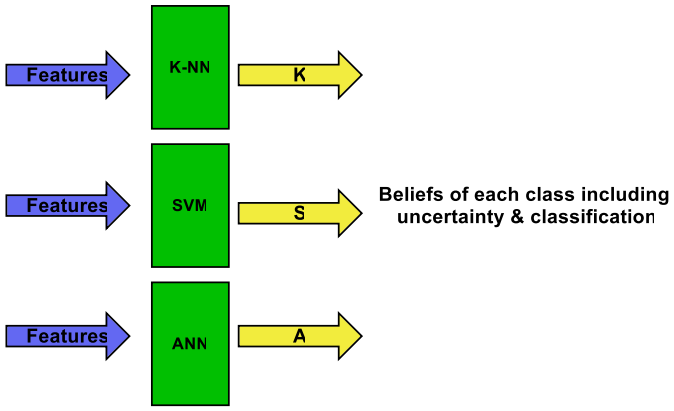

Figure 4.25: Block diagram of individual classifiers.

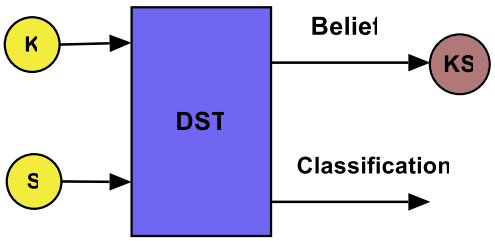

Figure 4.26: Combinations of SVM and KNN classifier.

Bel (B) indicates a trust, which belongs to the benign group. Firstly, in order to calculate the benign belief, multiply benign trust masses of KNN and SVM, multiply the uncertainty for KNN and benign trust of SVM, and finally multiply benign belief of KNN and ambiguity or uncertainty of SVM. Secondly, in order to complete the benign belief all of the fundamental beliefs are summed. Consequently, KNN and SVM combined beliefs are shown in Table 4.2:

$$bel_{comb(B)} = bel_{SVM(B)} \times bel_{KNN(B)} + U_{SVM} \times bel_{KNN(B)} + bel_{SVM(B)} \times U_{KNN} \quad (5.27)$$

The factor 1 - Σ A ∩ B is used to normalize this combined belief, as A ∩ B= Φ. Therefore, we can calculate the final combined belief as follows:

		KNN classifier beliefs		
		Bel B	*Bel M*	U_{KNN}
SVM Classifier beliefs	*Bel B*	*Bel B*	*Bel Ø*	$U_{KNN} \times bel_{SVM(B)}$
	Bel M	*Bel Ø*	*Bel M*	$U_{KNN} \times bel_{SVM(M)}$
	U_{SVM}	$bel_{KNN(B)} \times U_{SVM}$	$bel_{KNN(M)} \times U_{SVM}$	*Bel U*

Table 4.2: Fusion of KNN and SVM classifiers outputs

$$bel_{comb(B)} = \frac{bel_{svm(B)} \times bel_{kNN(B)} + U_{svm} \times bel_{kNN}(B) + bel\_svm(B) \times UkNN}{1 - bel_{svm(M)} \times bel_{kNN(B)} + bel\_svm(B) \times bel_{kNN(M)}} \quad (4.28)$$

$$bel_{comb(M)} = \frac{bel_{svm(M)} \times bel_{kNN(M)} + U_{svm} \times bel_{kNN(M)} + bel\_svm(M) \times U_{kNN}}{1 - bel_{svm(M)} \times bel_{kNN(B)} + bel\_svm(B) \times bel_{kNN(M)}} \quad (4.29)$$

$$bel_{comb(U)} = \frac{bel(U)}{1 - bel_{svm(M)} \times bel_{kNN(B)} + bel\_svm(B) \times bel_{kNN(M)})} \quad (4.30)$$

The highest fusing belief value determines the final assigned class; similarly, the result of combination of KNN and SVM is combined with the ANN classifier Figure 4.27.

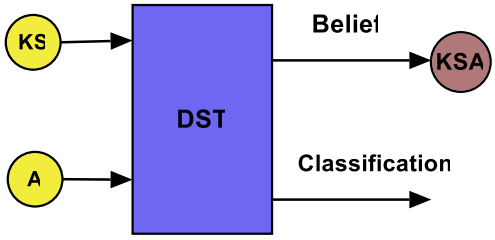

Figure 4.27: Combination of KS(KNN and SVM fusion) and ANN classifier.

$$bel_{comb(B)} = \frac{bel_{KS(B)} \times bel_{ANN(B)} + U_{KS} \times bel_{ANN(B)} + bel\_KS(B) \times U\_ANN}{1 - bel_{KS(M)} \times bel_{ANN(B)} + bel\_KS(B) \times bel\_ANN(M)} \quad (4.31)$$

$$bel_{comb(M)} = \frac{bel_{KS(M)} \times bel_{ANN(M)} + U_{KS} \times bel_{ANN(M)} + bel\_KS(M) \times U\_ANN}{1 - bel_{KS(M)} \times bel_{ANN(B)} + bel\_KS(B) \times bel\_ANN(M)} \quad (4.32)$$

$$bel_{comb(U)} = \frac{bel(U)}{1 - bel_{KS(M)} \times bel_{ANN(B)} + bel\_KS(B) \times bel\_ANN(M)} \quad (4.33)$$

Chapter 5

Deep Learning Based Feature Extraction for Tumor Characterization and Detection

Deep Learning is a rapidly expanding subject of artificial intelligence that has become a hot issue in a variety of fields. Deep Learning's concrete accomplishments in numerous areas, such as education, manufacturing, transportation, healthcare, military, and automotive, are now visible in our daily lives. Deep Learning is a branch of machine learning that evolved from Artificial Neural Networks, in which a cascade of layers is used to extract higher-level characteristics from raw data and generate predictions about incoming data. The impact of attribute extraction, which is deeply embedded in training techniques such as Convolutional Neural Networks, will be discussed in this study (CNN). In addition, the article offers to do research on recent advances in deep learning techniques and attribute extraction approaches. As demand grows, extensive attribute extraction assignment research has become increasingly more important. The characterization and detection of brain tumors will be utilized as a case study to illustrate Deep Learning CNN's capacity to accomplish successful tumor characterization and representational learning.

5.1 Introduction

For many centuries, wise people have dreamt of constructing a machine with the power to imitate the human brain. Rina Dechter introduced the expression deep learning, and its origin dates from 1986. The term “neocogitron,” which was introduced in 1980, inspired

the convolutional neural network [125]; however, recurrent neural networks (RNNs) were suggested in 1986 [126]. Following this, in the 1990s, LeNet enabled deep neural networks (DNNs) to apply in practice, although this was not generally accepted [127]. The hardware restriction causes the LeNet structure to be simplistic; therefore, it cannot be used with large datasets. Deep belief networks (DBNs) and a layer-wise pre-training structure were introduced in about 2006 [128]. Their principal objectives were to learn a basic two-layer non-learning paradigm, such as limited Boltzmann machines (RBMs), freeze every parameter, place a new layer at the highest level, and train only these variables for the new layer. This method enabled researchers to train more profound neural networks than was previously possible, thereby motivating a rebranding of neural networks to deep learning. After several decades of advancement, deep learning originating from artificial ANNs became one of the most effective instruments compared with other well-performing machine-learning algorithms. It has been observed that some deep-learning techniques developed from the first ANNs, which included RNNs, RBMs, DBNs, and convolutional neural networks (CNNs) [129]. Although graphics processing units (GPUs) are famous for their action in the computation of large-scale frameworks in organizing structures on one mechanism, numerous distributed deep-learning networks have been developed to accelerate the training of deep-learning paradigms [130]. Large volumes of data have labels or even noisy labels; therefore, specific research attaches greater importance to enhancing training modules' noise strength by applying unmonitored or semi-supervised deep-training methods. The max of the existing deep-learning paradigms concentrates exclusively on one modality, resulting in a restricted representation of real world data. At present, researchers are giving a greater

concentration to a cross-modality framework, which can be a massive advancement in deep learning.

Feature extraction defined as the translation of some input data into a set of features. This commences with an opening series of reliable data in machine learning, subsequently developing the adopted standards (aspects)[118,119]. It is anticipated that such features are non-redundant and descriptive. It simplifies the experimental stages and meaningful learning in a situation where the data input into an algorithm is too significant to be managed and thought to be redundant (a significant volume of data, but little information). The features will be converted into a smaller representation group of factors, known as the features vector. Consequently, the extraction of discriminatory features in the signal improves the reduction of the data vector length by disregarding redundancy in the data and by concentrating the appropriate information into a feature vector of a considerably smaller dimension. It is anticipated that the preferred features will define feature extraction as the translation of some input data into acquiring the relevant information from the provided dataset. The resulting task is well executed by recognizing the reduced volume of data rather than being supplied with big data. Feature extraction corresponds to decreasing the number of advantages necessary to describe a massive sequence of detail. In conducting an analysis inquiry of complex data, the convoluted variable's sum is the principal difficulty. It is usual that analysis with several variables requires considerable computational power and a large volume of memory and access to the categorization algorithm to surfeit the training method to compute the new one. The expression feature extraction is a frequently applied expression for a type of planning set of parameters that emerges from these difficulties while still defining the

data with sufficient accuracy[120,121]. Performing an organic image is a reduced paradigm for advocating decision-making, for example, recognition, classification, and detection of the pattern. A technique that reduces the provided data volume by removing the comprehensive characteristics presupposes various features from the previously supplied features. This aims to decrease the feature analysis cost, allow greater categorization efficiency, and develop classifier precision. Another particular facet is achieved from such obtainable input data through this procedure. Classification is performed by applying another set of aspects used to attain categorization, such tasks being utilized to achieve unique properties [122]. Such a mechanism is to extract reliable information from the image. This dimensionality contraction method recognizes a reduced group of features, being the series of the primary ones and the procedure of acquiring suitable characteristics that are enclosed inside the supplied input data.

Eventually, the magnitude of such data will be reduced to retain only the principal information. The main stage in numerous PC vision and item classification assignments is the extraction of useful features; consequently, several scientists have concentrated on designing robust features for multiple image grouping jobs. Considerable attention is now paid to convolutional neural networks (CNN) and feature-learning algorithms. Every photo is entered straight into the CNN algorithm. Subsequently, the procedure abstracts its most significant features [123,124]. In one of the following sections, more details for CNN. This chapter’s layout is divided into deep-learning history, Deep Learning background, a Literature review, feature extraction, deep learning analysis, a convolution neural network, followed by the methodology and experiment result, and finally, the conclusion.

5.2 Methodology

The datasets contain T2-weighted 256 x 256 in-plane resolution. The first benchmark dataset comprises 66 (18 benign and 48 malign) brain MRI images. The second dataset is 160 brain MRI pictures, where 20 are Benign and 140 malign. The third dataset download contains 47 brains, benign MR images. The total dataset is 85 and 188 brain MRIs, benign and malign images, respectively. Since this is a small dataset, There were not enough MRI images to learn the convolutional neural network. Therefore, data augmentation was proper to overcome the information imbalance matter in the dataset. Besides artificially expand the size of a real dataset, data augmentation techniques produce various copies of it. Machine learning models can benefit from data augmentation methods. An experiment found that a deep learning model with picture augmentation performs better in terms of training loss. However, data augmentation may be utilized in a wide range of fields. The following are some of the applied in this work prevalent image data augmentation techniques: Position augmentation, such as scaling, translation, cropping, flipping, and padding. Afterward data augmentation, the dataset now contains 850 benign and 1128 malign MRI images, 1978 MRI images in total. The Alex net builder was applied to recognize and categorize the images. The data was split in the following way: 70% of the data for training, 15% for validation, and 15% for testing. The Alex Net CNN was constructed to be applied to distinguish the MRI brain images as benign or cancerous (malign), which involved five convolutional layers and three layers of pooling layers, and wholly linked layers followed by the classification layer. The 3.6 Python programming language under Anaconda platform software was used to implement the CNN algorithm.

There are two classes labelled as 0 and 1, representing the benign and malign brain image. The Python programming software is used to implement the CNN algorithm. Besides, comparing the CNN algorithm results with the results obtained in chapter four. The classifiers K-NN, SVM, and ANN, were applied to distinguish the cerebrum tumor. The classic classifiers' features were 13 features after applying discrete wavelet transform and principal component analyses for each MRI image, respectively. The 3.8 Weka platform is used to apply these classifiers.

5.3 Some Insight

Machine learning is being used to evaluate a growing amount of data that has gotten significantly more sophisticated over time. Indeed, the rise of deep learning over the last decade has aided in the development of more effective learning paradigms. Several machine-learning tasks aim to categorize problems, with systems similar to the one shown in Figure 5.1. For starters, a characteristic retrieved from the supplied input data might result in a fresh exemplification of the data that is unique to the present task. Following that, a classification technique is learned, as well as the above-mentioned features, to accomplish the assignment. When the training is complete, this approach should be used to data that was not observed during the training stage and should accurately anticipate the response, as well as the class label in this situation. The extracted characteristics of the input were commonly handmade, especially until recently, implying that such features are intentionally intended for the input data and the present job. These are usually related to a specific subset of the data type, such as photos of handwritten English words in ink, rather than the data type itself, such as images of handwritten English words in ink. Most such features struggle to cope with change; nevertheless, machine learning is a technique for extracting features from data to develop a feature extractor. Nonetheless, rather than creating a technique to categorize individual photos, a learning system is built to extract qualities from the input. This

indicates that the network is learning directly from the input pixels, which is a more relevant characteristic when it comes to pictures.

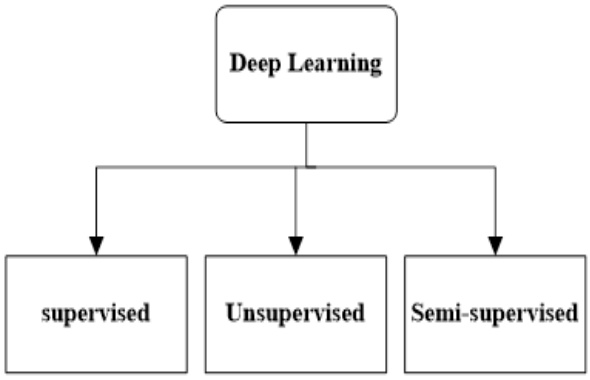

Figure 5.1: Hierarchical structure of learning.

As a result, this is a preferable approach of leveraging handmade features for a variety of reasons. By performing such training to each dataset, the trained paradigm may be adjusted to multiple input types; however, each dataset may need to be hand-tuned. Furthermore, this technique should not need a professional grasp of the pictures being analyzed.

5.4 Feature Extraction

Machine learning is applied to analyze a growing volume of information, which is becoming increasingly complicated. The emergence of deep learning during the past decade has undoubtedly helped generate learning paradigms that are becoming more effective. Numerous machine-learning jobs seek to categorize difficulties. Since features are extracted from the input data, we can consider this a novel representation of the data, particularly for this task. Subsequently, in addition to these features, which complete the job, a categorization method is learned. We should apply this method to unobserved data in the training stage period upon completion of the training. We ought to precisely forecast its response precisely, and in this situation, also the class label. Frequently, and particularly until recent times, the features extracted from the input were handcrafted, implying that they are specially designed for the input

data and the present task. It is standard practice for these not to be exclusively tied to the data type, for instance, handwritten images of words' pictures, but rather to a specific subset, such as English words handwritten in ink on parchment. Usually, most such features cannot manage change well; nevertheless, ML is a different procedure of removing attributes from the information to learn a feature extractor. The learning system is built to extract attributes from the input instead of planning a classifying image. This shows that the network is learning, from the input pixels directly, a greater level of features concerning ideas. Consequently, for many reasons, we regard this as an improved method of utilizing handcrafted features. We can adjust the trained paradigm to several input types by applying this training to each dataset; nevertheless, we may need to hand-tune each dataset for handcrafted features. However, a specialist understanding of the images being analyzed is not necessary for this approach. Concerning information retrieval and image analysis, feature extraction is a significant and fundamental issue. Although a considerable amount of time is needed to hand-design a useful feature, deep-learning allows the acquisition of such attributes whose objective is new applications. Deep learning has attained much as a new feature extraction technique. Traditional systems and deep learning principally differ in that deep learning automatically acquires aspects from huge information rather than handcrafted attributes. This is generally dependent on the previous knowledge of designers, and it is certainly not possible to obtain an advantage from big data. It is possible for deep learning to automatically get attribute representation from big data, including millions of variables. Deep learning's principal benefit is that it is unnecessary to abstract attributes from the photo manually. During training, the system learns to pull elements by feeding the image to the system (pixel values) [130].

Convolution neural network construction is typically a series of feedforward layers that apply and pooling layers and convolutional filters. After the final pooling layer, CNN employs numerous completely associated layers that convert the previous layers' 2D feature maps of 1D vector for categorization. Although one advantage of CNN architecture is that it does not need a feature

extraction procedure before training, a CNN from first principles occupies much time. Moreover, it isn't easy because it requires a significantly sizeable categorized dataset for training and constructing earlier the paradigm being prepared for grouping, which is not possible in every instance. Furthermore, the hardware needed to process numerous filters for greater-sized images is 256×256. This is a standard feedforward network where the information moves from input to the output layers through numerous invisible layers, being a minimum of two [132].

5.5 Deep Learning Analysis

Deep structured learning is another name for deep learning, and it's a group from machine learning. Deep learning requires two substantial factors; one is a massive amount of classified data and powerful computing. It appoints the group of ML methods, which define several layers of performance in deep- designs. Next, a detailed examination of developed deep learning designs: Deep-Belief Networks (DBNs), Multi-Layer Perceptron (MLPs), and Convolutional Neural Networks. The central design block is a bipartite undirected photographic representation named the Restricted Boltzmann Machine (RBM). The (RBM) was original proposed by Hinton [133]. More technically, RBM is a stochastic neural model. Neural network purpose: We have neuron-like units with zero or one activation function based on the neighbours they're linked to; stochastic, meaning these activations have a like hood element. Graphical demonstration of the Restricted-Boltzmann Machine is displayed in Figure 5.2.

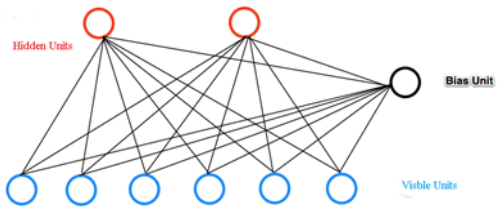

Figure 5.2: Graphical description of restricted-Boltzmann machine.

The network is restricted because the no-hidden unit is connected to any other visible unit, and no invisible unit is connected to any other invisible unit. DBNs can be observed as a structure of

single, unsupervised systems, i.e., RBMs + Sigmoid Belief Networks. DBNs' most significant success is their ability of learning attributes, which is obtained by a 'layer-by-layer training procedure where the higher-level attributes are learned from the preceding layers. Figure 5.3 displays a stack of RBM, and single or multiple layers are summed for discrimination jobs.

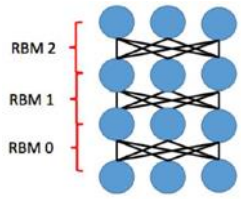

Figure 5.3: A typical stack of restricted-Boltzmann machine.

1. DBNs can be formed by stacking RBMs.

2. Each layer is trained as a Restricted Boltzmann Machine.

3. Train layers sequentially starting from the bottom (observed data) layer. (Greedy layer-wise).

4. Each layer learns a higher-level representation of the layer below.

5. The training criterion does not depend on the labels. (Unsupervised)[134].

It contains visual units, any 0 or 1 or real-number, and a collection of binary invisible junction. A structure through visual path v and unseen path h is given with energy specified :

$$E(s,h) = -\sum_{i\in visible} \sigma_i\, s_i - \sum_{j\in hidden} b_j\, h_j - \sum_{ij} s_i\, h_i w_{ij} \quad (5.1)$$

where α_i, b_j, w_{ij} are the system factors. Assumed this energy, the system assumes to each couple *s*, *h* likelihood:

$$P(s,h) = \frac{1}{Y}\, e^{-E(s,h)} \quad (5.2)$$

All likely pairs of visual and invisible paths are summed to produce *Y* as separation function.

It's fast way to achieve unbiased pair as no direct link between visible and invisible units

$$P(h_j = \frac{1}{s} = \sigma(b_j + \sum_i s_i w_{ij}) \quad (5.3)$$

the logistic sigmoid function is σ(·). Likewise, assumed an invisible path *h* the likelihood of a visual junction s_i to be consigned to one is specified:

$$P(s_i = \frac{1}{h} = \sigma(b_j + \sum_j h_j w_{ij}) \quad (5.4)$$

Obtain as large as the probability of the visual and invisible vectors team $\{s, h\}$. The network parameters are tuned using the training data. RBMs are limited in the complication of the information they can appear as it is single two layers deep system. To mitigate this problem, several more profound paradigms constructed on Restricted Boltzmann Machine are proposed. The deep belief network (DBN) is the most ordinary model derived from RBMs and the DBM [135,136,137]. These algorithms are several layers of likelihood systems that implement a non-linear conversion to the data. The greedy wise manner is used to train DBNs somewhere every layer learned as its RBM. Except for the top two that remain undirected, the last paradigm holds single the top-down connections of the layers. In all layers, DBMs have undirected weights, unlike DBNs. Similar to DBNs, the weights are learned in a greedy mode. To assessment and maximize the probability straightway is hugely computationally costly.

In summary, Deep Belief Networks employs an eager and active layer-by-layer method to train the latent variables in each invisible layer and a backpropagation technique for adjustment. This practical learning approach increases the generative achievement and the discriminative power of this network.

The MLP is a deep, artificial neural network. It contains more than one perceptron. The input layer to collect the signal, the decision produced by the outcome layer, and the Qualitative amount of invisible between input and output layers represent the computing tools of MLP. MLP is regularly useful for supervised learning difficulties. They practice collecting input-output sets and training to parade the correspondence (or dependencies) among the inputs and outputs. The practice comprises tuning the weights and biases of the model to decrease deviation. Tuning both the weight and bias relative to the fault is done by the backpropagation technique, and the error is calculated in different methods. Several artificial neurons constructed the Neural networks. Each input into the neuron has its weight related where weight is the real number required to tuning

once we finally reach to learn algorithms. No limit input number to a neuron, from one to *i;* the total input number denoted by *i*. Figure 5.4 demonstrates a mathematical model of a discrete perceptron or neuron.

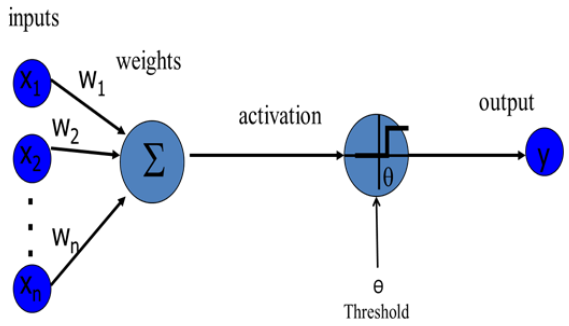

Figure 5.4: A mathematical model of a discrete perceptron or neuron.

The inputs may be denoted : x_1, x_2, … x_i. And the matching weights for the inputs denoted w_1, w_2, .. w_n , and the output a = x_1w_1 + x_2w_2... +x_nw_i . In feedforward network: The neurons in every layer feed their output forward to the following layer till we get the last output from the neural network. There can be any number of invisible layers inside a feedforward network, and the number of neurons can be random. Another MLP properties of architecture :

1. There is no direct connection from the input to the output layers.
2. Completely connection among the layers.
3. Often more than three layers.
4. It is not required output and input units be equal.
5. No connections within a layer.
6. No limit to the invisible units in each layer, and could be greater or less than input or output units.

The data flow starts at the input to the output layer after passing through the invisible layer. The output layer produced the output and compared it with the actual output value. Using the backpropagation method and the series basis of calculus, fractional by products of the error

function w.r.t. the different scales and biases are back-propagated out of the Multi-layer perceptron. This differential action offers a gradient of error, over which the parameters could be tuned and proceed with the MLP one stair nearer to the lowest error. Any gradient-based optimization algorithm can be applied to perform this task, such as stochastic gradient descent. The model keeps doing this optimization until the error reaches the lowest. This case is famous as convergence. MLPs are connected, each junction with a single layer join by the special weight *wij* to each junction in the following layer. The two activation functions are both sigma functions, and are described by $x(si)$ - $\tanh(si)$ and $x(vi)$ – $\tanh(si)$. The primary piece is a hyperbolic tan function that dimensions from minus one to one. Simultaneously, the other is the log function, which is alike in configuration but dimensions from zero to one. Here *si* is the result of the *jth* junction and vi the weighted sum of the given network.

5.5.1 Convolutional Neural Network

Convolutional neural networks are constituents of appropriate neural network technique classification. As well as having the capacity to learn image feature representations automatically, CNN has exceeded several traditional hand-crafted feature methods [138]. A convolutional neural network (CNN) depends on following one another locally-connected convolutional layers, each of which has an equal number of filters, down-sampling layers, and the fully connected layers that serve as classifiers [139]. Figure 5 depicts CNN's comprehensive architecture. Local receptive fields, weight sharing, and down-sampling procedures are three aspects of convolutional neural network architecture that make it efficient. With a convolution filter of the same magnitude, every neuron recognizes input from a minor piece of the preceding layer, as shown by the local receptive field. Local receptive fields are also employed in the convolutional and down sampling layers. The convolutional layer uses weight sharing to control the capacity of the paradigm and reduce its difficulty. Finally, non-linear down-sampling is used to lower the spatial dimensions of the image and the free factors of the paradigm. These notions enable CNN

to perform well in recognition tasks. The Convolutional Neural Networks layers, on the other hand, are:

1. The Convolutional Layer: A convolutional layer comprises a large number of channels, each of which has its own set of parameters that must be learnt. The channels have a smaller height and weight than the information volume. To process an actuation map comprised of neurons, each channel is convolved with the info volume. The speck items between the info and channel are logged at each spatial place, and channel is slid over the width and stature of the information. Six feature maps were created by the first convolutional layer, which had six filters. Every feature map represents distinct aspects of a picture, such as represented points or specified vertical edges. The complication operation is described:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_i^l + b_j^l\right) \tag{5.5}$$

Here, j is a particular convolution feature map, M_j denotes a set of input maps, k_j^l denotes the filter, b_j denotes feature map bias, l denotes the CNN layer, and f denotes the activation function. The ReLU is a commonly used activation function for adding non-linearity to a system [140].

2. The Pooling Layer performs a down sampling operation on the convolutional layers to decrease their locative aspect. The size of the pooling mask and the type of pooling operation must be specified first, and then applied to the pooling layer [141]. The pooling process multiplies the pixel values recorded by the pooling mask by a trainable coefficient, then adds the result to a trainable bias. The following is a description of the pooling procedure:

$$x_j^l = f\left(B_j^l pool(x_j^l)\right) + b_j^l \tag{5.6}$$

Where x_j^{l-1} is the j^{jt} area of interest caught by the pooling mask in the previous layer, pool is the specific operation done on the region (max or average), B_i is a trainable coefficient, b_j^l is a trainable bias, and f is a trigger function[142].

3. Fully linked layers: which performed categorization using the extraction attributes from preceding layers. As in an original neural network, the result of the last convolutional or pooling layer is used to produce fully linked layers.

Figure 5.5 shows primary CNN architecture.

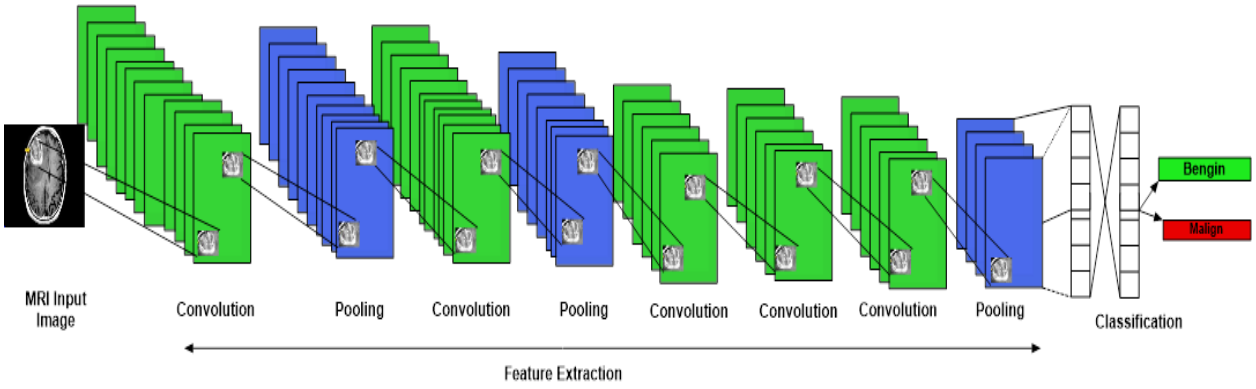

Figure 5.5: Figure 5.5: Brain tumor detection CNN algorithm.

5.6 Experiment Results and Discussion

The classification conclusions are arranged for the following algorithms: KNN, SVM, and ANN. Table 5.1 demonstrates a confusion matrix of every classifier. Class 0 represents the normal (benign); class 1 represents abnormal (malignant), The KNN algorithm offers the highest accuracy in categorizing classes related to class 0 (benign), where 75 images are correctly classified as benign, and 13 are misclassified as malignant. The artificial neuron network achieved the highest performance in the categorization related to class 1 as 170 images were correctly categorized as malignant, while 18 images were misclassified as benign. Figure 6.6 and 6.7 shows

ANN			SVM			KNN		
	0	1		0	1		0	1
0	70	15	0	74	11	0	75	13
1	19	169	1	18	170	1	21	167

Table 5.1: Algorithms confusion matrixes.

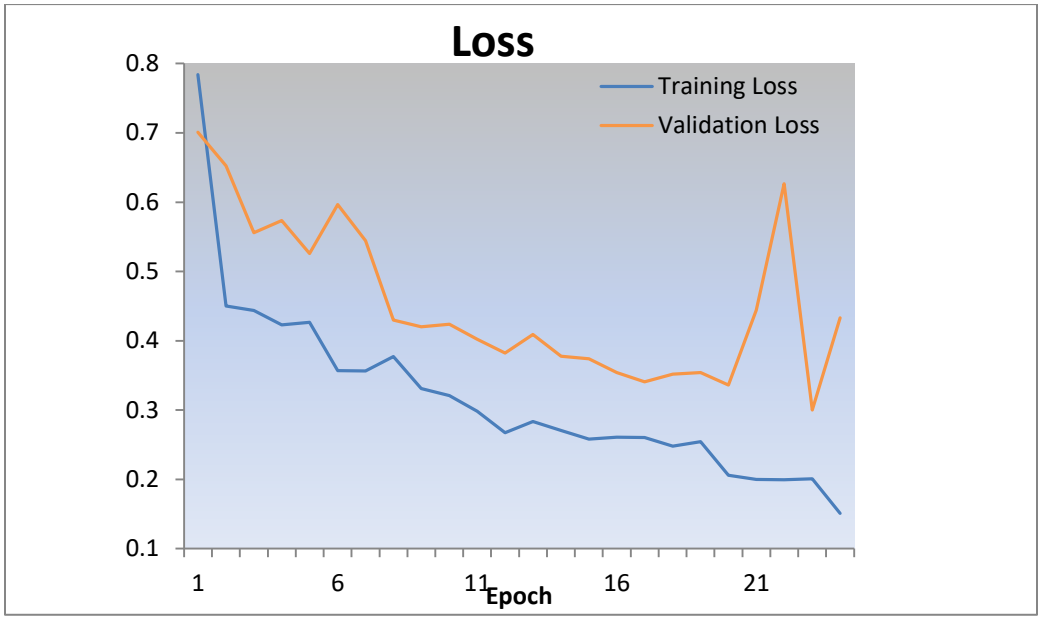

Figure 5.6: Training and validation loss.

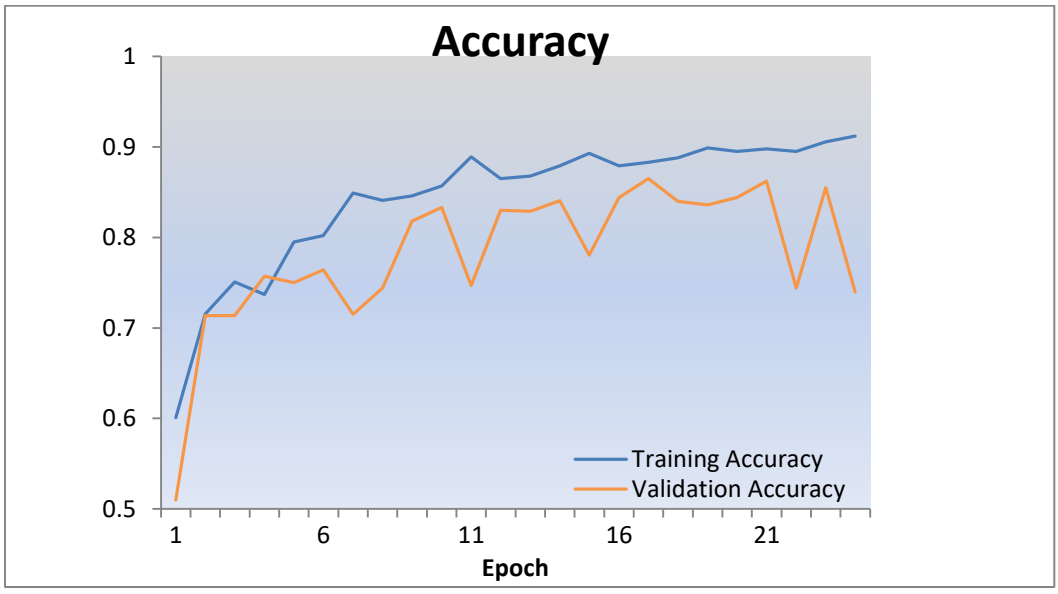

Figure 5.7: Training and validation accuracy.

The convolution neural network achieved high accuracy in the MRI brain tumor taxonomy as benign or malicious scans at a precision of 91%. The sigmoid fully Connected layer classifier in the CNN building has been applied to assess the suggested method's effectiveness. The classifier's performance is validated by multi measuring factors similarly accuracy and F1 score. The F1 score combines the model's precision and recall, and is required when you want to investigate a balance between Precision and Recall. Also, it is applied to assess binary classification systems. The CNN accuracy and F1 score obtained is 90% and 91%, respectively. However, the CNN algorithm has a great calculation cost and required the right quantity of data for the learning phase. Table 5.2 illustrates the accuracy and F1 score of each algorithm.

Classifier	**ANN**	**SVM**	**KNN**	**CNN**
Accuracy %	87	89	88	90
F1 Score	90	91	90	91

Table 5.2: Algorithms comparison using F1 score and the accuracy.

5.7 Conclusion

Feature extraction involves converting training data and authenticating it with additional aspects to produce adequate machine learning. The CNN algorithm was applied for the brain tumor detection task. Besides, three classical classifiers, KNN, SVM, and the ANN classifier, were used to classify the same datasets to compare the CNN algorithm's performance among these classifiers. The CNN is precisely suitable for choosing an auto-feature in medical images such as magnetic resonance imaging. The datasets contain two classes, normal and abnormal. Physicians classified these scans. Besides, this study has presented a survey of locating the effect of automatic feature extraction and dissemination, which is applied in deep learning; for example, a CNN. The objective of this is to demonstrate the emanating study of existing configuration on

attribute extraction techniques during recent years. As the demand for the application grew, significant analysis and research of the feature extraction platform changed into being particularly active. It would help conduct more experiments to validate the multi-modal architecture's possible advantages against applying a specific convolution neural network.

Chapter 6
Result and Analysis

6.1 Model Performance Evaluation

One of the significant parts of constructing classifiers is assessing their performance [143]. Apart from the measure that is used to quantify the classification performance, careful planning is desired to obtain this estimate. In addition, one is often interested in comparing. The ratio of the state that is appropriately classified represents by the accuracy [144]. In contrast, the relation of states that are not correctly categorized indicates the classification error rate. For multi-class problems, accuracies can be summarized in a confusion matrix [145]. This matrix structure relates the true classification (on its horizontal axis) and the predicted classification (on its vertical axis) for a data point. Accuracies can be calculated such that the whole sum for each actual class result becomes 100%. These factors are not perfect as they can be misguiding if the data is unbalanced. For example, in a situation, If one class label represents 90% of the data, an algorithm that at all times expects this class already achieves an accuracy of 0.90 (the majority rule). The mean error ratio for every independent set, unbalance in the data can be considered. True positive rate and true negative rate: are other statistical

procedures used to assess classifiers predication. The sensitivity and specificity formula are presented in section 6.3.1.

6.2 Evaluation Strategies

In this thesis, the three algorithms are supervised approaches, meaning that a training data set is required to construct the classifier. However, the classifier or algorithm's performance is evaluated on a set of new, unseen examples during the testing phase. If the algorithm on the training data achieves high performance, the algorithm capability on invisible data may be low. This behavior is named overfitting [146]. Thus, the generalization capability of the algorithm is specified through this testing stage. Generalization is a significant feature and overwhelmingly the final aim of the algorithm since it gives information about the quality of the developed classifier or algorithm. Suppose one aims to evaluate a classifier's performance and rely on the number of existing data points. In that case, it could be essential to divide the primary data between learning and testing sets. This strategy can, for instance, be applied to determine hyperparameter, i.e., C, for SVM-based classifiers. [147,148]. In this work, the data set is split into 5 folds. In the first iteration, the first fold is used to test the model and the rest are used to train the model. In the second iteration, 2nd fold is used as the testing set while the rest serve as the training set. This process is repeated until each fold of the 5 folds have been used as the testing set.

6.3 Experimental Work

Several experiments were carried out to assess the success of the developed framework in brain tumor diagnostic assistance systems. The datasets are two-weighted 256 x 256 in-plane resolution. One dataset was obtained from the Harvard Medical College, and the

second was downloaded from the OASIS website (http://oasis-brains.org /). These are benchmark datasets utilized in brain MRI image analysis tasks and contain benign and malignant MRI brain images. The first benchmark dataset includes 66 brain MRI images (18 benign and 48 malignant). The second dataset consists of 160 MRI brain pictures, of which 20 are benign and 140 malignant. The third dataset contains 47 benign MRI images, as shown in table 6.1. Experts have labelled these images. The experiment was conducted, and every algorithm was evaluated individually by the same datasets and the number of features required to achieve maximum accuracy. The combination of the multi-classifier is undertaken by different methods such as minimum and maximum probabilities, majority vote, and average likelihood. The Dempster-Shafer is the primary method to combine the classifiers' outputs. The two-class confusion matrix is shown in Figure 6.2. A confusion matrix was obtained to define the execution of the classifiers both in individual and combination classifiers' tasks. The benign and malignant images are correctly classified as true negative and true positive, respectively.

Datasets	Benign	Maliganat
DS-160	20	140
DS-66	18	48
DS-37	47	-
Total 226	85	188

Table 6.1: MRI Datasets.

A False Positive represents the classification of all incorrect results as being malignant where they are benign. False Positive is the false signal in the recognition task. A False negative represents the classification of all incorrect results as benign where they are malignant.

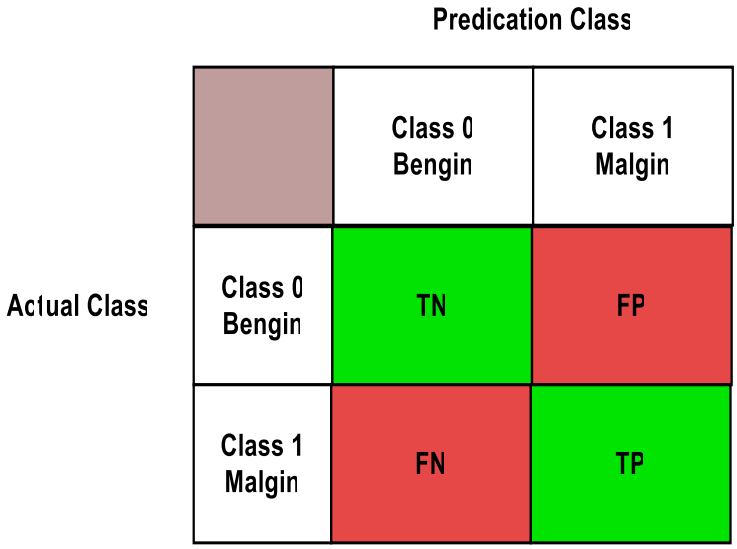

Table 6.2: Two class confusion matrix.

6.3.1 Sensitivity versus Specificity

The sensitivity, specificity, and precision are metric measure, which can be utilized to measure the performance of the taxonomy paradigm. These measurement are grounded on the regard that a predication point continuously drops into one of the next 4 classes [149]:

1. Sensitivity: the possibility is that detection examination is positive when the patient has a tumor.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{6.1}$$

1 Specificity: the possibility is that a detection examination is negative when the patient does not have a tumor.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{6.2}$$

2 Accuracy: The possibility that all a detection examination is correctly classified.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \tag{6.3}$$

3 Precision: the fraction of abnormal images with correct results.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{6.4}$$

where:

TP= Amount of malignant images accurately classified.

TN= Amount of benign images accurately classified.

FP= Amount of benign images classified as malignant.

FN= Amount of malignant images classified as benign.

There is another statistical measure called F-score or F-measure (weighted harmonic mean) used when determining the trade-off between recall (sensitivity) and precision and calculated by this formula [150]:

$$F1 = 2.\frac{precision.recall}{precision+recall} \tag{6.5}$$

Numerous diverse experimentations were executed to assess the achievement of the construct system in light of attribute decrease effectiveness, classification precision, contrasts with other systems and calculation difficulty examination. The size of the feature after three level of DWT is 32 × 32. principal components analysis is used to decrease the attribute path size to only 13, which is the first 13 principal components, preserving 94.5% of whole variance of the decomposed attributes. Table 6.3 shows confusion matrix with extra class 2 represents the uncertainty.

	Predication Class		
	Class 0 Bengin	Class 1 Malgin	Class 2 Uncertainty
Actual Class			
Class 0 Bengin	TN	FP	Class 0 Bengin
Class 1 Malgin	FN	TP	Class 1 Malgin

Table 6.3: Confusion matrix include class 2 represents the uncertainty.

6.4 Comparison with State-of-the-Art Classifiers

There are several reasons for preferring a multi classifier system over a single classifier. It is mainly done to improve the accuracy and efficiency of the classification system. The success of classifier fusion depends on two factors: a pool of diverse individual classifiers to be fused, and the proper combining method.

As a result, a baseline model may be utilized to determine the foundation of brain tumor detection performance against which all other models can be compared. Given the stochastic nature of data and algorithms, all prediction models incorporate mistakes, and achieving a perfect score is impossible in practice. Besides, the actual purpose of applied machine learning is to investigate the space of potential models and determine what a good model score looks like compared to the baseline on a given dataset. Because machine learning model performance is relative, establishing a solid baseline is essential. Also, a baseline is a straightforward and well-understood technique for forecasting your predictive modelling challenge. This model's skill serves as the foundation for determining the minimum acceptable performance of a machine learning model on a

given dataset. The baseline model's outcomes serve as a. benchmark against which all other data-driven models may be measured In addition, A baseline serves as a benchmark against which other machine learning algorithms can be measured.

The Zero Rule algorithm in the Weka machine learning platform predicts the class value with the most observations in the training dataset for a classification predictive modelling issue where a categorical value is anticipated. The dataset brain tumor detection is applied to the Zero Rule algorithm and achieves an accuracy of 67%. Therefore, any machine learning method must obtain an accuracy higher than this Figure to show that it can solve this issue. Several machine learning algorithms were examined, such as RJ48, Naive Bayes, and Random Forest. However, the chosen algorithms in this research are support vector machine, k-nearest neighbours, and artificial neural network, which achieved the highest accuracy between the others by 89%, 88%, and 87%, respectively.

6.5 Algorithms Tuning

Machine learning procedures can be adjusted to extract various conduct. The objective of classifier fine-tuning is to catch the most significant point or points in that hypercube for a given task. This is regularly called classifier tuning or classifier hyper-parameter development. It is a kind of trial and error experimental operation. An additional solid process is planning a controlled empirical to adjust various predefined classifier arrangements and prepare tools to check and contrast the outcomes [151].

The WEKA 3.8 platform is used for the classification task, and WEKA offers four different options for implementing the data mining process. The WEKA knowledge explorer is easy to use a framework with a graphical user interface that offers all the WEKA package facilities. Another framework is Weka experiment environment that

permits to creation, runs, and modifies an experiment. The experiment can be described in a text file and tested with the WEKA framework. WEKA knowledge flow environment permits the description of the experiment as a flow of steps with some visual connections between them. The WEKA Workbench contains much states-of-the-art data pre-processing and machine learning algorithms. In this framework, Weka experiments interface applied to tune each classifier for the best operational parameters.

6.5.1 Cross-Validation

The statistical approach of cross-validation is used to assess the performance of machine learning models. It is frequently used in applied machine learning to compare and choose a model for a particular predictive modelling issue. It is simple to comprehend, implement, and produce skill estimates with lower bias than other approaches. In addition, cross-validation is a strategy for determining how well the model worked. It is always necessary to evaluate the precision of the model to ensure that it has been adequately trained with data and is free of overfitting and under fitting. The data is separated into k subsets in K-fold cross-validation, and train the model on k-1 subsets while keeping the last one for testing. This procedure is done k times, with one of the k subsets serving as the test set/validation set each time and the remaining k-1 subsets as the training set. The model is then averaged against each of the folds before being finalized.

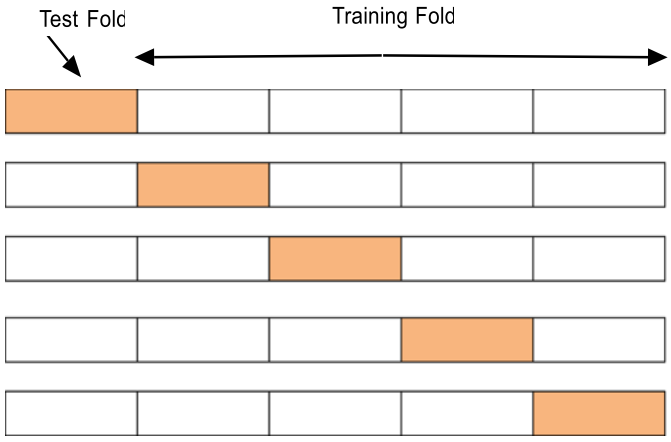

Figure 6.1: 5 K cross-validation.

A common choice for K-fold cross-validation is K = 5. After trying different values of k empirically, the result end with k = 5 is the best choice. If choosing a large number for the k-fold, then model slight bias, significant variance, and computation time increased; however, choosing k as a small number causes slight variance, considerable bias, and reduced computation time. Accuracy versus the 5 folds for each classifier is shown in Figure 6.2. The total dataset is 273 images containing two classes, benign or malignant tumor. The dataset is divided into five folds. Each fold contains 55 images after adding two images in order to have equal folds. Each sub-fold contains 11 images. After testing each algorithm five times, average accuracy was calculated for each classifier. Besides, this process is repeated to each algorithm during the parameters tuning task.

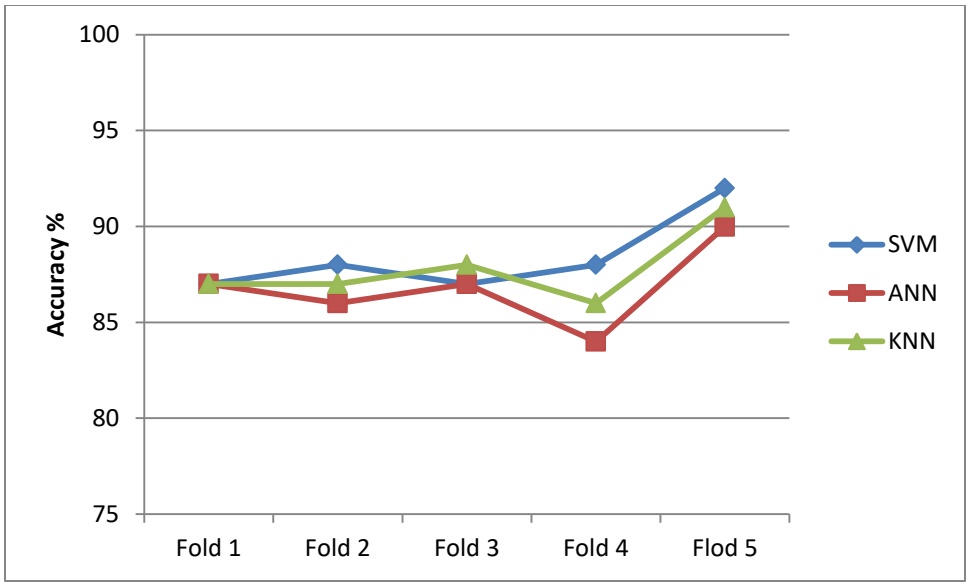

Figure 6.2: Accuracy versus folds of cross-validation for applied classifiers.

6.5.2 k-Nearest Neighbour Tuning

In Machine learning, KNN is a significant classification algorithm. It is based on similar cases with same class labels are near each other. The output can be set as the class with the maximum frequency from the K-most like instances; if KNN is applied for the classification, each instance, a vote for their class, and the class with the most choices is taken as the prediction. This is will investigate the significance of k, which is the number of neighbors to query to make a prediction. To select the K that is right for given data. The KNN classifier runs several times with different K values and selects the K that decreases the number of errors encountered while continuing the algorithm's ability to make predictions when given previously unseen data. The Weka 3.8 platform provided experiment graphical user interface GUI where the user can run the algorithm with different choices numbers of K parameters and distance measure:

IBk, k=1,3,5 distance Function=Euclidean.

IBk, k=3, distance Function=Euclidean, Manhattan.

In addition, the experiment uses the default 5-fold cross-validation as described in pervious section. The K value equal to five with Euclidean distance measure gives the highest accuracy for a given dataset.

6.5.3 MLP Tuning

The tuning parameters that configure this component are listed, where each one shows the optimal value applied to obtain high performance.

1. Seed: applied to begin the random numbering generator. Random numbers are utilized for tuning the primary weights of the connections through nodes, and for shuffling the learning data. Seed(default) = 0.
2. Momentum: is the value that is implemented to the weights throughout updating. Momentum = 0.2.
3. Hidden Layers: This describes the number unseen layers of the neural network as well as the numbers of cell from every layer.
4. Validation Threshold: used to dismiss the learning process. The value set how many times in a row the validation set error can get worsen before learning is ended. Validation Threshold = 20.
5. Batch Size: The preferred number of vector instances kept in cache once if the batch prediction is being complete. Extra or fewer features may be provided, however, this gives implementations a way to assign a choose batch size. Batch size = 100.
6. Learning rate: The amount the weights are updated. Learning rate(default) = 0.3.

6.5.4 SVM Tuning

Support vectors machine goes to catch the top hyperplane to detached the dissimilar labels. This is achieved by maximizing the distance amongst instance points and the hyperplane. The algorithm looks at each twosome of data points until it finds the nearby pair in an individual class and draws out a straight line (or plane) in the middle. When the input data is linearly divided, finding the hyperplane is straightforward. However, categorization areas frequently overlay, and no particular conventional line can function such as a border. The SVM with RBF kernel is applied and definite by dual factors: C, or the misclassification cost, and γ, which is relational to the inverse of a support vector's radius of influence. The adjusted C and γ are by a search through among 0.0001 and 100000.

6.6 Result and discussion

First, evaluates the number of features that produce the classification process high accuracy performance and feature reduction effectiveness. The introduced system is based on the DWT decomposition for feature extraction. After three levels of decomposition configuration, the size of the LL is 32 × 32. PCA is employed to decrease the attribute vector size from 65536 to 1024; however, these attributes are still extensive. Thus, principal components are used to reduce these features. The variances against the principal components are from one to twenty, showing at the latest nineteen principal components, which are simply 1.85% of the principal attributes. Determine the correct principal components, which provide the most excellent outcome; the multi-classifier's performance was investigated with altered principal components up to 13 components.

The offered scheme achieved the highest accuracy, with only 13 principal components for input images. Figure 8.1 displays the achievement of the algorithms in terms of accuracy versus the number of principal components.

This section also offers the results of each classifier and the classifier fusion by the elementary combiner and DST. The outcomes of each algorithm and the classifier fusion are presents by the DS theory. The classes are normal (benign) and abnormal (malignant), which are indicated as zero (benign) and one (malignant), respectively.

The class indicated by two identifies uncertainty and, consequently, more testing required. This classification is invaluable when the case of false negatives is exceptionally high. Moreover, it may be the best option to warn an expert of uncertainty rather than of an unconvinced and possibly incorrect decision.

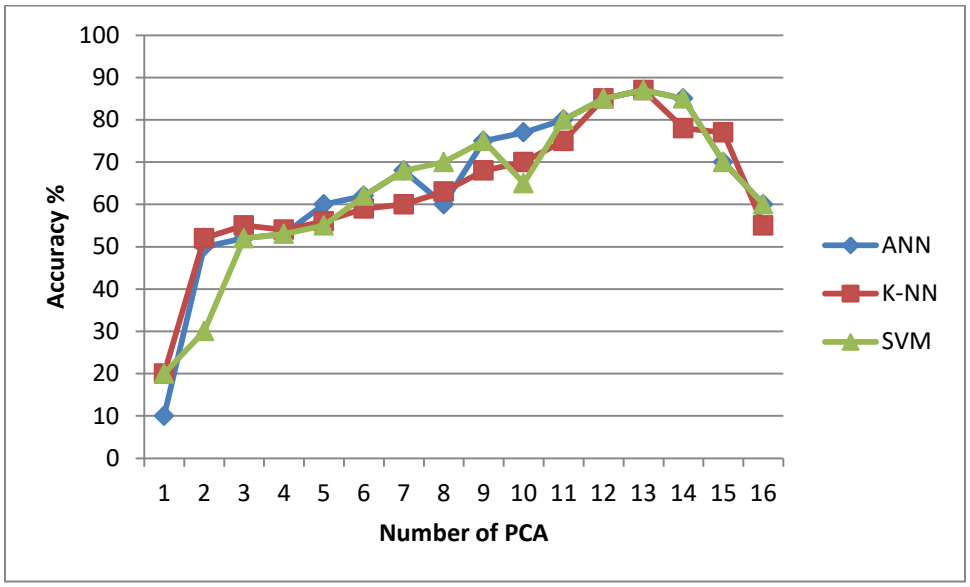

Figure 6.3: KNN, SVM, and ANN performance vs No. of principle components.

The outcomes are given in the configuration of a confusion matrix. The classification outcomes are prepared for three algorithms: KNN, ANN, SVM, and the Dempster-Shafer

Theory. Class zero represents the normal (benign), class one represents the abnormal (malignant), and class two corresponds to the uncertainty classification.

ANN				SVM				KNN			
	0	1	2		0	1	2		0	1	2
0	70	15	0	0	74	11	0	0	75	13	0
1	19	169	0	1	18	170	0	1	21	167	0

Table 6.4: SVM, ANN, and KNN confusion matrix.

The confusion matrix is shown in table 6.4 for three classifiers SVM, KNN, and ANN. The KNN algorithm offers the highest accuracy in categorizing classes related to class 0 (benign), where 75 images are correctly classified as benign, and 13 images are misclassified as malignant. The SVM algorithm achieves the highest performance in the categorization related to class 1(malignant). There were 170 and 18 images correctly classified by SVM as malignant and benign, respectively. The DST confusion matrix is shown in table 6.5, where 77 benign and 175 malignant images are correctly classified as benign and malignant, respectively; however, eight false-positive and 13 false-positive images are misclassified. The uncertainty (more test required) is represented in class 2 (DST scenario one), where no monitoring for conflicting beliefs during the combination process. The DST achieve a better result in both classes 77 and 175 in benign and malignant classes correctly classified, respectively. Figure 6.2 displays the performance of each classifier and DST. Combining the algorithms through DST improves the classification result regarding the accuracy, sensitivity, specificity, and precision.

DST Fusion			
	0	1	2 Uncertainty (More Test Required)
0	77	8	0
1	13	175	0

Table 6.5: DST confusion matrix with no conflicting factor setup.

The false-negative rate is when a classified image is shown as benign where it is, in fact, malignant, and the false positive rate is when an image is classified as malignant when it is benign. A false positive can result in unnecessary therapy, and a false negative can result in an inaccurate diagnosis, which is particularly dangerous since an illness has been neglected. These results' impact means an improvement is necessary for the tool applied to combine the decisions; this helps decrease the records of false-positive and false-negative effects. Introducing DST allows extra class when there is no certainty about the last decision to be rendered, as more tests are needed. DS's planned system gives us the capacity to decrease false negative and false positive when there is no trust decision (DST scenario two). This will be done by monitoring the evidence's conflict factor while fusion the evidence and defining a threshold for this factor. If the conflict factor is the same or greater than the threshold, the decision will be more tested classified as class 2, which shows that more tests are required to arrive at a trusted decision.

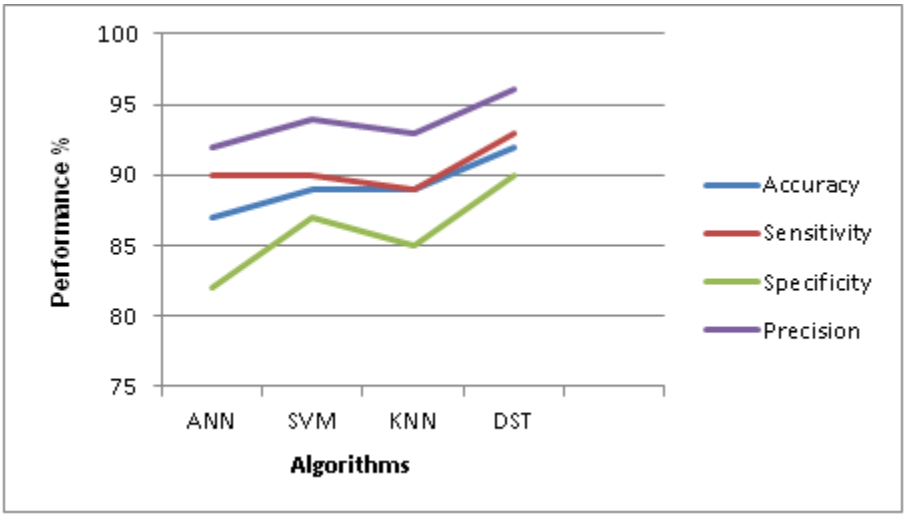

Figure 6.4: Comparison of several multi-classifier, and DST results.

The elementary combination rules include average, product, minimum, maximum probability, and the majority vote applied to combine classifiers outputs. These combining results show in Figure 6.4. As seen, the average likelihood gives a better outcome than the other rules regarding accuracy, sensitivity, specificity, and precision.

The DST represents the ambiguity or uncertainty of contradiction within the evidence to be combined. In addition, if the conflict is equal to the threshold, the scans or images classified as class 2 represent uncertainty in the decision.

Applying the DST helps to minimize both the false positive and negative cases to increase sensitivity and specificity measures. Table 6.6 illustrates the decision matrix for DST with determined thresholds. It contains only four images and three images as false negative and false positive, respectively.

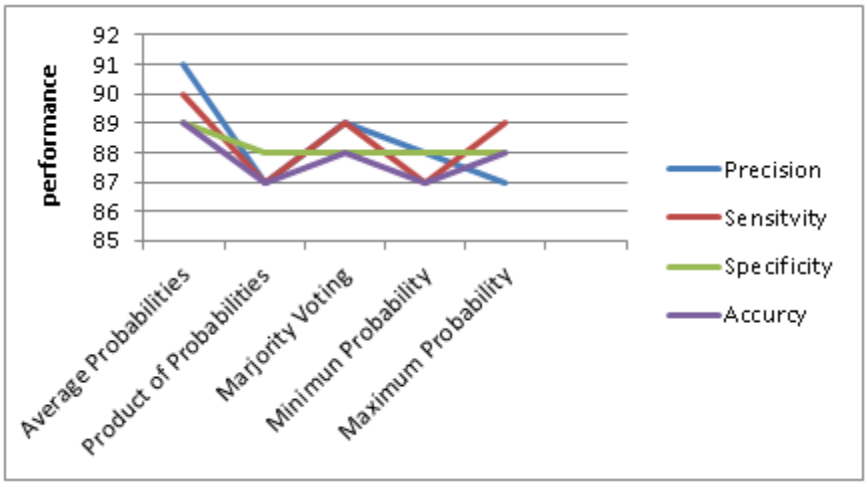

Figure 6.5: Comparison of several multi-classifier fusion procedure.

Moreover, class 2 includes four images from class 0 (benign) and three images from class 1 (Malignant). The Dempster-Shafer Theory recorded fewer classification errors in both classes, as shown in Table 6.6. This indicates that four images from class 0 (benign) and three images from class 1(Malignant) require further tests (uncertainty), meaning that the

contradiction between the evidence attains the predetermined threshold and cannot produce a confident decision.

DST Fusion			
	0	1	2 Uncertainty (More Test Required)
0	78	3	4
1	5	180	3

Table 6.6: DST confusion matrices with set up k conflicting factor.

This is possibly a helpful mechanism for evaluating uncertainty and reliability in brain tumor detection when it is undesirable to achieve a correct measurement from experiments. A significant feature of this theory is the fusion of proof achieved from various algorithms and opposition modeling. The achieved result demonstrates that the Dempster-Shafer Theory enhances the classification process and that the overall accuracy reached 96%. Moreover, the sensitivity and specificity obtained are 97% and 96%, respectively, as shown in Table 6.7. It is essential to resolve the false negative and positive rates by more test required or uncertainty class It also shows that the DST technique has the best accuracy, specificity, sensitivity, and precision.

Classifier	Accuracy	Specificity	Sensitivity	Precision
KNN	88	88	89	92
SVM	89	87	90	93
ANN	87	82	89	92
Average Probability	91	89	90	91
DSET	96	96	97	98

Table 6.7: Comparison of multi-classifier, majority vote and DST.

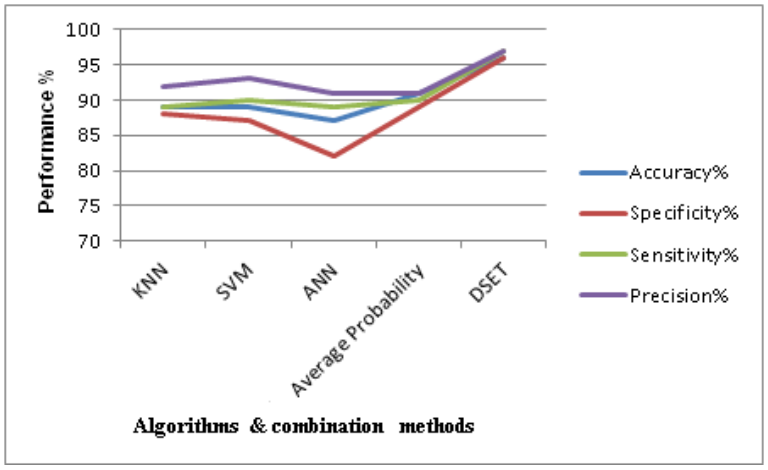

Figure 6.6: Comparison of elementary combination methods and DST fusion.

Table 6.4 and Figure 6.4 illustrated that DST produces high accuracy, sensitivity, specificity, and precision. Figure 6.5 shows the significant increase in these statistical measures when monitoring conflicting factor while combining the algorithm's output.

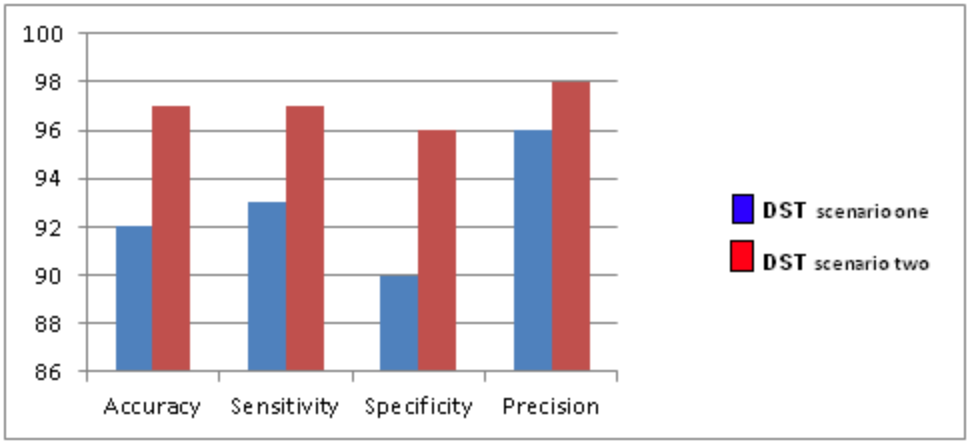

Figure 6.7: Chart of comparison the two scenarios of DST.

The method presented in this thesis utilizes these specific algorithms. It combines the outcomes and attempts to expand upon the outcomes of the particular methods. Several algorithms perform well for a specific dataset but showing weak outcomes for some other datasets. In this situation, trusting only one algorithm could lead to wrong classifications as the datasets are different. Dempster Shafer's approach enhances this shortage of an algorithm to act on different datasets. The algorithms, which can perform their

classification outcomes related to the beliefs to several labels, can be used as specific algorithms and combined with the other classifiers providing improved precision. The advantage of the Dempster Shafer fusion method is that it also takes into attention uncertainty. The combination classifier classifies records as uncertain if there is a conflict in individual classifier results. The uncertain classification does not qualify as misclassification but demands for more detailed medical investigation. The above results conclude that DST formalism has the distinct advantage of modelling uncertainty in conflicting information.

Classifier	ANN	SVM	KNN	CNN	DST
Accuracy %	87	89	88	91	96
F1 Score	90	91	90	92	97

Table 6.8: Algorithms comparison in terms of F1 score and the accuracy.

6.6.1 Time Analysis

Another essential element to consider while evaluating the brain tumor support system is the computation time. The experiments were carried out using a Dell Desktop computer with a 3 GHz processor and 8 GB of RAM, running under the Windows 10 operating system. All 273 images were applied to the support system. The corresponding computation time was recorded and the average value was computed. The spent time of the different phases is represented in Figure 6.8. For each 256 × 256 image, the averaged computation time for segmentation, feature extraction, feature reduction, multi classifiers, and DST fusion was 0.02 s, 0.04 s, 0.025 s, 0.015 s, and 0.08 s, respectively.

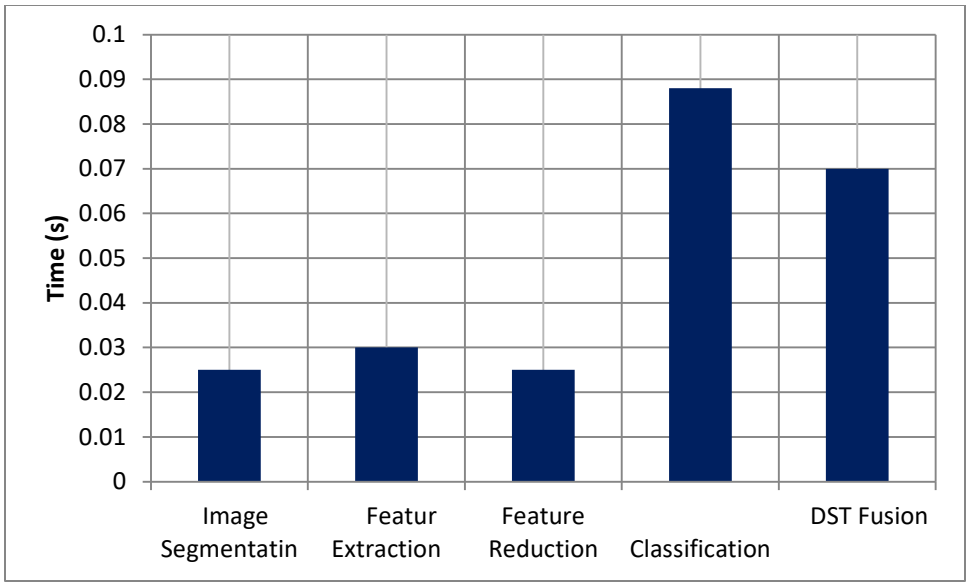

Figure 4.9: Computation times at different stages.

The multi classifier stage was the most time-consuming at 0.088 s. The averaged calculation time on each 256 x 256 image was 0.23 seconds, which is rapid enough for a real time diagnosis.

6.6.2 Performance of Classifiers Fusion by DST Versus CNN

The CNN model is utilized in various applications and domains, but it is most common in image and video processing tasks. The convolutional layer is the main building component of the CNN in terms of architecture. A CNN employs 2D convolutional layers to combine learned features with incoming data, making it ideally suited to processing 2D data, such as images. Because CNN's do not require human feature extraction, there is no requirement to pick the necessary features to categorize the images. However, the classical classifiers require a separate feature extraction stage. In addition, straight feature engineering interested in classical machine learning, these classifiers are easy to infer and understand. As well, regulation hyperparameters and altered paradigm designs are extra means to directly comprehend the information and the implied classifiers better.

Compared to the performance of classic classifiers versus the Alex Net CNN algorithm introduced in Chapter Five, deep networks need enormously significant datasets in the training phase to achieve a high performance. In addition, deep networks need a fast graphics processing unit to become skilled in a realistic period with important information. Large memory and a high speed central processing unit (CPU) are prerequisites to using graphics processing units (GPUs) successfully. However, classical classifiers can learn relatively quickly with a decent CPU, and the most potent hardware is unnecessary. Since these classifiers do not require a high calculation time, they could try out multiple methods in a reasonably short period of time. A comparison among the algorithms in terms of accuracy and F1 score is shown in Figure 6.9. The classical classifiers' computation time is minimal compared with the considerable amount of computation time required by the CNN.

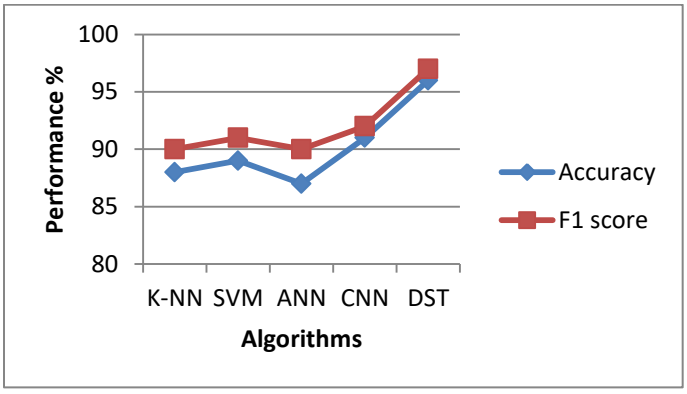

Figure 6.9: Algorithms assessment using accuracy & F1 score.

The introduced system, with the classical classifiers SVM, KNN, and ANN combined with a decision fusion using DST, obtained a high accuracy of 96% and an F1 score of 97%, as shown Figure 6.8.

6.7 System Performance and Trust

It is critical to pick the right metrics to assess the machine learning model. The measures used to evaluate and compare the performance of the machine learning algorithms have an impact on how they are measured and compared. The confusion matrix is one of the most intuitive classification problems, in which the output can be of two or more types of classes. The confusion matrix given and the terms associated with it were explained in Section 6.3.2. In brain tumor detection, which intends to capture all cancer cases, it might classify a person who does not have cancer as cancerous. This might be less dangerous than not identifying/capturing a cancerous patient since doctors will send the cancer cases for further examination and reports. However, missing a cancer patient would be a vast error as no further examination would be done. In a perfect scenario, the model would produce no false positives and no false negatives. Nevertheless, that is not the case in real life situations, as no model will be 100% accurate.

In this research work, three machine learning algorithms were used: k-nearest neighbors, support vector machine, and artificial neural network. These algorithms were applied to detect two classes of brain tumor, and they produced their predication for each input feature vector. Therefore, choosing robust classification algorithms is the first step to building a successful brain tumor detection system. The three applied classifiers were selected empirically based on several experiments. Their prediction results were then compared with state of the art machine learning algorithms, such as decision tree, ZeroR, random forest, Naive Bayes, and PART in the multi-classifier group.

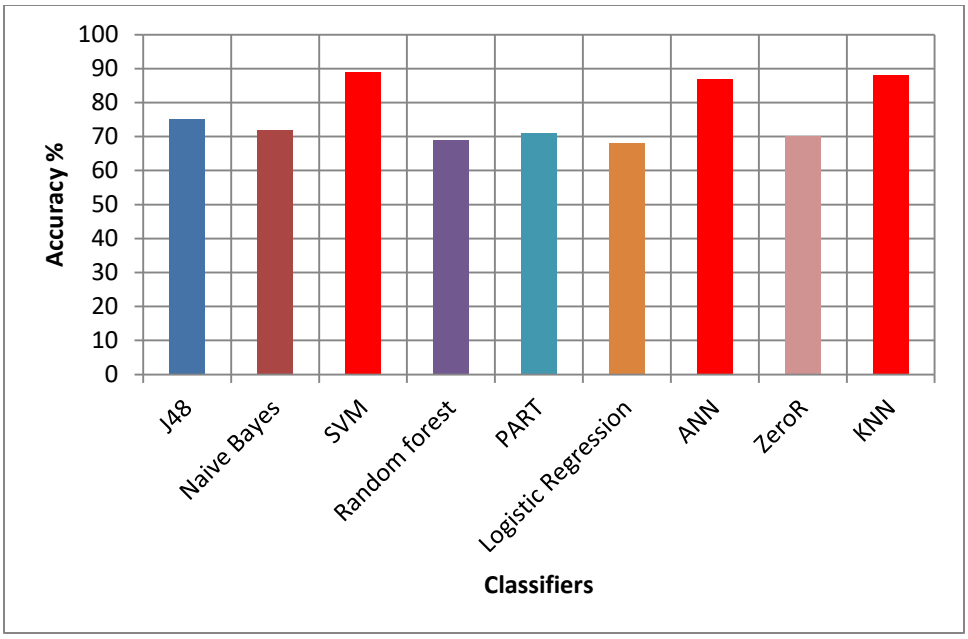

Figure 6.10: Various algorithms performance in terms of accuracy. Figure 6.10 shows the accuracy performance of each algorithm for the same dataset. The detection trust score for the chosen brain tumor algorithms was higher than the others in terms of sensitivity, specificity, and precision. However, the final prediction in this work came from the fusion of these classifiers by DST to achieve a final trusted decision. The trust spectrum zooms out and measures the model's trustworthiness across different features when tested on a finite set of inputs. When presented visually, the trust spectrum provides an excellent overview of when one can and cannot trust the brain tumor detection support model. Figure 6.11 shows the spectrum of the first 50 samples' beliefs produced by the DST

fusion technique from the output of the multi classifier.

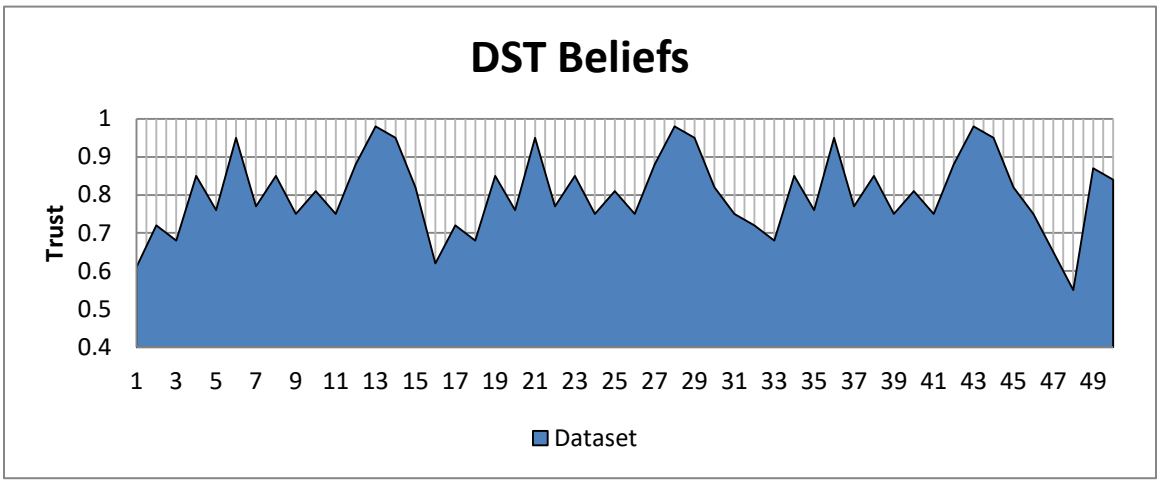

Figure 6.1 The trust spectrum of DST fusion.

The threshold to accept these beliefs as the trusted result is equal to 0.74 or greater. Therefore, the system will reject any decision with a lower threshold trust level. In addition, the evidence from the classifiers' outputs, conflicting factor k, is monitored during the fusion process. The factor k in the DST combination rule indicates the amount of evidential conflict. If k = 0, this shows complete compatibility; if k =1, it shows complete contradiction; and if $0 < k < 1$, it shows partial compatibility.

The system will classify any testing image as being in class 2 if the conflict value reaches the preset value. This procedure will reduce the total of the false positive and negative rates, and increase the sensitivity, specificity, and overall classification accuracy. The DST combination rule was explained in Section 4.8.1.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The thesis aimed to design an automatic system to classify brain tumors from MRI images by fusing decisions of multi-classifier. This has been successfully achieved. In addition, this thesis introduces a systematic procedure to MRI brain image pre-processing, segmentation, feature extraction, feature selection, classification, and decision combination. Based on the combination of median filter, threshold technique, discrete wavelet transform, multi-classifier, and Dempster-Shafer theory of evidence. Initially, the MRI brain images were smoothened and enhanced by a nonlinear median filter to exclude any existing artifacts surrounding the image, such as name and date. In brain MRI segmentation, the thresholding technique implemented to extract attention is the brain tumor in this task. This segmentation is usually straightforward for brain MRI studies. In the MRI image, features are extracted through implementing the DWT. In addition, as the decomposition level increased, a compactor but coarse approximation complex was attained. Therefore, wavelets afford an easy ranked frame for the photo information translates. In this work, up to three breakdown levels through Harr wavelet were used to extract features. In attribute selection or decrease, principal component

analyses algorithm PCA applied to reduce the feature from 1024 to 13 components, which could preserve 95.4% of the overall variance. The compact attributes are used as input to three algorithms. The multi-classifier stage consists of three algorithms: ANN, SVM, and KNN. All of which are used supervised techniques. They differ in their approach to classifying and predicting the classes. Two algorithms are non-parametric, and one is a parametric classifier. The results from the classifier were analyzed using a confusion matrix. The multi-classifier combination was designed and implemented using different decision fusion basics such as maximum, minimum, product, average probability, and majority vote, where the primary multi-fusion technique is the Dempster-Shafer theory. The DST outperforms the other techniques in combining the multi-decision with sensitivity (97%), specificity (96%), and overall accuracy up to 96%. The combination technique has displayed optimal classification precision.

Moreover, the combination method stills powerful in the existence of very various algorithm performances. The capability to manage such conditions powerfully and classify samples as ambiguous in algorithm uncertainty makes this procedure bright for medical image processing. This is a high speed, rather good recall or sensitivity, and specificity. The system can execute great brain image databases and afford fast results in medical settings. The overall performance and the outcomes display that the DST algorithm in the fusion task achieves better than the other present approaches.

7.2 Research Contributions

This study's main contribution is to expand the classification performance of detecting brain abnormality grounded on multi-classified MRI brain scans and decision fusion accuracy. In this research, multi-classifier were applied to the classification task, three

algorithms, two non-parametric, and one parametric classifier. They are different in their theory of classification to improve the final decision after combining their outputs. The Dempster-Shafer theory DST was applied in the fusion task increases the discriminative accuracy and represents any uncertainty in the final determination. The DST theory of evidence provides the ability to model various partial ignorance types, limited or conflicting evidence. Initially, there are two classes representing the output classes 0 as benign and 1 as benign and malign. Besides, class 2 was added to represent the uncertainty in the combining process when there is undoubtedly conflict between the two beliefs, i.e., classifiers output. These three classes representing accept, reject, or more tests required by applying this approach. The other methods have only two types, as there is a tumor or not. Introducing DST offers us another class with no confidence about the final decision to be represented, as more tests are required. DS's proposed system allows us to reduce false negatives and false positives when there is no firm decision for the correct class. This will be done by controlling the conflict factor between the evidence while combining the decisions and specifying two thresholds. First, suppose the conflicting factor reaches the first threshold and is less than the next threshold. In that case, the final decision will be considered by the classifier, which has a high confidence rate of fault acceptance in this category. If the conflicting equal or greater than the second threshold, the decision will be more test. This procedure gives us to decrease fault acceptance and fault rejection. The two-threshold factors are tuned until obtaining the optimal values. The contribution of this research is listed in following:

1. A speedy and clinically oriented technique has been advanced to detect brain tumors and normal tissues from multi-modal images.
2. Capabilities of improving a robust system have been explored to distinguish brain tumors grounded on MRI brain photos and determine the tumor absence or presence. In addition, it will represent any uncertainty that could exist while combining tumor classifiers.
3. The proposed and filtering algorithms are employed to eliminate high-frequency components (noise) and undesirable MRI areas.
4. Dempster-Shafer's theory DST is applied in the fusion task to increase the discriminative accuracy and represent any uncertainty in the final determination. The DS theory of evidence provides the ability to model partial ignorance, limited or conflicting evidence.
5. Powerful algorithms were applied in the classification and decision-fusion stages, which can be used for other diseases with small or no modification.

7.3 Future Works

Identification of brain tumors is reliant on the detection of abnormal brain tissues. However, several aspects of brain tumors' successful treatment are determined by the tumor category, location, and dimensions. Identifying a tumor is a sensitive and challenging assignment; consequently, precision and consistency are permanently consigned considerable significance. For future work, this thesis research can be extended to increase detection and segmentation precision. Here is a list of recommendations for possible extensions of the works of this thesis research:

1. The detection accuracy can be increased further by using Magnetic Resonance Spectroscopic Imaging (MRSI), which provides the brain regions' chemical composition. The tumor region's chemical composition is different from the non – tumors regions.
2. The segmentation and feature extraction stages can be developed by different techniques to enhance the classification stage and obtain higher efficiency.
3. The system can be further used to classify images with dissimilar diseased cases, categories, and disease statuses using other types of image modalities (e.g., CTS canner, PET, MRS, and mammogram) for cancer classification.

Bibliography

[1] B. Menze, A. Jakab, S. Bauer, et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. In: *IEEE Transactions on Medical Imaging* (2014), p. 33. doi: 10.1109/TMI.2014.2377694.

[2] Lakshmi, A., and T. Arivoli. "COMPUTER AIDED DIAGNOSIS SYSTEM FOR BRAIN TUMOR DETECTION AND SEGMENTATION." *Journal of Theoretical & Applied Information Technology* 64.2 (2014).

[3] Brady, Adrian, et al. "Discrepancy and error in radiology: concepts, causes and consequences." *Ulster Med J* 81.1 (2012): 3-9.

[4] Thousands of radiology exams under review at terrace hospital.[online]. www.pressreader.com/canada/the-province/20170226/281621010111879

[5] Three errors found in NHS radiology review.[online]. https://www.niagarafallsreview.ca/news/niagara-region/2016/04/20/three-errors-found-in-nhs-radiology-review.html.

[6] Thousand of medical scans to be reviewed after Niagara radiologist’s mistake.[online].https://www.theglobeandmail.com/life/health-and-fitness/health/niagara-health-system-radiologists-work-under-review/article25073441/

[7] Toronto radiologist's 3,500 CT scans, mammograms reviewed.[online]. https://nationalpost.com/news/canada/ontario-hospitals-waited-five-months-to-publicize-fears-about-possible-errors-in-reading-of-3500-mammograms-and-ct-scans

[8] Ontario hospitals alerting patients of possible errors in 3,500 CT scans, mammograms.[online].https://www.macleans.ca/news/ontario-hospitals-alerting-patients-of-possible-errors-in-3500-ct-scans-mammograms/

[9] El-Dahshan, El-Sayed A., et al. "Computer-aided Diagnosis of human brain tumor through MRI: A survey and a new algorithm." Expert systems with Applications 41.11 (2014): 5526-5545.

[10] Soltaninejad, Mohammadreza, et al. "Automated brain tumor detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI." International journal of computer assisted radiology and surgery 12.2 (2017): 183-203.

[11] Garis, Hugo. "An artificial brain ATR's CAM-Brain Project aims to build/evolve an artificial brain with a million neural net modules inside a trillion cell Cellular Automata Machine." *New Generation Computing* 12.2 . 215-221.

[12] World Health Organization. (2018). Noncommunicable diseases country profiles 2018.

[13] Almog, N., Ma, L., Schwager, C., Brinkmann, B. G., Beheshti, A., Vajkoczy, P., ... & Abdollahi, A. (2012). Consensus micro RNAs governing the switch of dormant tumors to the fast-growing angiogenic phenotype. *PloS one*, *7*(8), e44001.

[14] Vigneswaran, K., Neill, S., & Hadjipanayis, C. G. (2015). Beyond the World Health Organization grading of infiltrating gliomas: advances in the molecular genetics of glioma classification. Annals of translational medicine, 3(7), 95. https://doi.org/10.3978/j.issn.2305-5839.2015.03.57.

[15] Cancer.ca/en/research/cancer-statistics/canadian-cancer-statistics.

[16] Shapiro W, Shapiro J. Biology and treatment of malignant gliomas. Oncology 998;12:233-240.

[17] Levin V, Leibel S, Gutin P. Cancer: Principles and Practice of Oncology. Lippincott Williams & Wilkins, 2001.

[18] Kitange GJ, Templeton KL, Jenkins RB. Recent advances in the molecular genetics of primary gliomas. Current Opinion in Oncology 2003;15:197-203.

[19] Favre J., Taha J.M., and Burchiel K.J. An analysis of the respective risks of hematoma formation in 361 consecutive morphological and functional stereotactic procedures. Neurosurgery, 50:48–57, 2002.

[20] Field M., Witham T.F., Flickinger J.C., Kondziolka D., and Lunsford L.D. Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. Journal of Neurosurgery, 94:545–551, 2001.

[21] Luts J., Poullet J.-B., Garcia-Gomez J.M., Suykens J.A.K., Heerschap A., and Van Huffel S. The effect of feature extraction for MRS classification. In Proceedings of the 15th Scientific Meeting and Exhibition of International Society for Magnetic Resonance in Medicine (ISMRM) − European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), Berlin, 2007.

[22] Pham, Dzung L., Chenyang Xu, and Jerry L. Prince. "Current methods in medical image segmentation 1." Annual review of biomedical engineering 2.1 (2000): 315-337.

[23] Understanding Medical Radition.[Online] https://www.medicalradiation.com/types-of-medical-imaging .

[24] Kristen Coyne. MRI: A Guided Tour. Florida State University, Magnet Lab, 2012. Available from: http://www.magnet.

[25] Kato, Hiroyuki, et al. "Silent cerebral microbleeds on T2*-weighted MRI correlation with stroke subtype, stroke recurrence, and leukoaraiosis." Stroke 33.6 (2002): 1536-1540.
[26] Mike Puddephat. Principles of magnetic resonance imaging. Easy Measure, 2002.

[27] Goldman, Lee W. "Principles of CT and CT technology." *Journal of nuclear medicine technology* 35.3 (2007): 115-128. [28] Schick, Fritz. "Whole-body MRII at high field: technical limits and clinical potential." European radiology 15.5 (2005): 946-959.

[29] Elangovan, A., & Jeyaseelan, T. (2016, February). Medical imaging modalities: a survey. In *2016 International Conference on emerging trends in engineering, technology and science (ICETETS)* (pp. 1-4). ieee.

[30] Chandana, S. R., Movva, S., Arora, M., & Singh, T. (2008). Primary brain tumors in adults. American family physician, 77(10), 1423.

[31] Norouzi, A., Rahim, M. S. M., Altameem, A., Saba, T., Rad, A. E., Rehman, A., & Uddin, M. (2014). Medical image segmentation methods, algorithms, and applications. IETE Technical Review, 31(3), 199-213.

[32] Zhang, Yu Jin. (2001).A survey on evaluation methods for image segmentation. Pattern recognition 1335-1346.
[33] Işın, Ali, Cem Direkoğlu, and Melike Şah. "Review of MRI-based brain tumor image segmentation using deep learning methods." *Procedia Computer Science* 102 (2016): 317-324.

[34] Wells, William M., et al. "Adaptive segmentation of MRI data." IEEE transactions on medical imaging 15.4 (1996): 429-442.

[35] Sinha, A., Pratik, R., & Panchal, M. (2014). MRI IMAGE SEGMENTATION–A REVIEW.

[36] Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE transactions on medical imaging, 20(1), 45-57.

[37] Clarke, L. P., et al. "MRII segmentation: methods and applications." Magnetic resonance imaging 13.3 (1995): 343-368.

[38] Nassiri, M. J., Vafaei, A., & Monadjemi, A. (2007). Texture feature extraction using Slant-Hadamard transform. International Journal of Applied Science, Engineering and Technology, 3, 4.
[39] Ziedan, R. H., Mead, M. A., & Eltawel, G. S. (2016). Selecting the Appropriate Feature Extraction Techniques for Automatic Medical Images Classification. International Journal

[40] Li, J., Tseng, K. K., Zu Yi Hsieh, C. W. Y., & Huang, H. N. (2014). Staining pattern classification of antinuclear autoantibodies based on block segmentation in indirect immunofluorescence images. PloS one, 9(12).

[41] Medjahed, S. A. (2015). A comparative study of feature extraction methods in images classification. International journal of image, graphics and signal processing,

[42] Luts, J., Laudadio, T., Idema, A. J., Simonetti, A. W., Heerschap, A., Vandermeulen, D., & Van Huffel, S. (2009). Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo, 22(4), 374-390.

[43] Kermi, A., Andjouh, K., & Zidane, F. (2018). Fully automated brain tumor segmentation system in 3D-MRI using symmetry analysis of brain and level sets. IET Image Processing, 12(11), 1964-1971.

[44] Anitha, V., & Murugavalli, S. (2016). Brain tumor classification using two-tier classifier with adaptive segmentation technique. IET computer vision, 10(1), 9-17.

[45] Singh, A. (2015, February). Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM. In 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 98-102). IEEE.

[46] Ravì, D., Fabelo, H., Callic, G. M., & Yang, G. Z. (2017). Manifold embedding and semantic segmentation for intraoperative guidance with hyperspectral brain imaging. IEEE transactions on medical imaging, 36(9), 1845-1857.

[47] Sallemi, L., Njeh, I., & Lehericy, S. (2015). Towards a computer aided prognosis for brain glioblastomas tumor growth estimation. IEEE transactions on nanobioscience, 14(7), 727-733.

[48] Mukambika, P. S., & Uma Rani, K. (2017). Segmentation and classification of MRI brain tumor. Int. Res. J. Eng. Technol.(IRJET), 4(07), 683-688.

[49] Sudharani, K., Sarma, T. C., & Rasad, K. S. (2015, December). Intelligent Brain Tumor lesion classification and identification from MRI images using k-NN technique. In 2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) (pp. 777-780). IEEE.

[50] Ahmmed, R., Swakshar, A. S., Hossain, M. F., & Rafiq, M. A. (2017, February). Classification of tumors and it stages in brain MRI using support vector machine and artificial neural network. In 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 229-234). IEEE.

[51] Machhale, K., Nandpuru, H. B., Kapur, V., & Kosta, L. (2015, May). MRI brain cancer classification using hybrid classifier (SVM-KNN). In 2015 International Conference on Industrial Instrumentation and Control (ICIC) (pp. 60-65). IEEE. Segmentation_paper

[52] Cai, H., Verma, R., Ou, Y., Lee, S. K., Melhem, E. R., & Davatzikos, C. (2007, April). Probabilistic segmentation of brain tumors based on multi-modality magnetic resonance images. In 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 600-603). IEEE.

[53] Verma, R., Zacharaki, E. I., Ou, Y., Cai, H., Chawla, S., Lee, S. K., ... & Davatzikos, C. (2008). Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of MR images. Academic radiology, 15(8), 966-977.

[54] Tilton, J. C., Tarabalka, Y., Montesano, P. M., & Gofman, E. (2012). Best merge region-growing segmentation with integrated nonadjacent region object aggregation. IEEE Transactions on Geoscience and Remote Sensing, 50(11), 4454-4467.

[55] Sumitra, N., & Saxena, R. K. (2013). Brain tumor classification using back propagation neural network. International Journal of Image, Graphics and Signal Processing, 5(2), 45.

[56] Xiao, K., Liang, A. L., Guan, H. B., & Hassanien, A. E. (2013). Extraction and application of deformation-based feature in medical images. Neurocomputing, 120, 177-184.

[57] El-Dahshan, E. S. A., Hosny, T., & Salem, A. B. M. (2010). Hybrid intelligent techniques for MRI brain images classification. Digital Signal Processing, 20(2), 433-441.

[58] Halder, A., Pramanik, S., & Kar, A. (2011). Dynamic image segmentation using fuzzy c-means based genetic algorithm. International Journal of Computer Applications, 28(6), 15-20.

[59] Wang, L., Gao, Y., Shi, F., Li, G., Gilmore, J. H., Lin, W., & Shen, D. (2015). LINKS: Learning-based multi-source IntegratioN frameworK for Segmentation of infant brain images. NeuroImage, 108, 160-172.

[60] Cai, Y., & Baciu, G. (2013). Detecting, grouping, and structure inference for invariant repetitive patterns in images. IEEE Transactions on Image Processing, 22(6), 2343-2355.

[61] Awad, M., Chehdi, K., & Nasri, A. (2007). Multicomponent image segmentation using a genetic algorithm and artificial neural network. IEEE Geoscience and remote sensing letters, 4(4), 571-575.

[62] Lan, T., Xiao, Z., Hu, C., Ding, Y., & Qin, Z. (2014, July). MRI brain image segmentation based on Kerneled FCM algorithm and using image filtering method. In 2014 International Conference on Audio, Language and Image Processing (pp. 511-515). IEEE.

[63] Mishro, P. K., Agrawal, S., Dora, L., & Panda, R. (2017, October). A fuzzy C-means clustering approach to HMRF-EM model For MRI brain tissue segmentation. In 2017 6th International Conference on Computer Applications In Electrical Engineering-Recent Advances (CERA) (pp. 371-376). IEEE.

[64] Wang, P., & Wang, H. (2008, December). A modified FCM algorithm for MRI brain image segmentation. In 2008 International Seminar on Future BioMedical Information Engineering (pp. 26-29). IEEE.

[65] Farooq, M., & Sazonov, E. (2017, April). Feature extraction using deep learning for food type recognition. In International conference on bioinformatics and biomedical engineering (pp. 464-472). Springer, Cham.

[66] Lu, X., Duan, X., Mao, X., Li, Y., & Zhang, X. (2017). Feature extraction and fusion using deep convolutional neural networks for face detection. Mathematical Problems in Engineering, 2017.

[67] Huan, E. Y., Wen, G. H., Zhang, S. J., Li, D. Y., Hu, Y., Chang, T. Y., ... & Huang, B. L. (2017). Deep convolutional neural networks for classifying body constitution based on face image. Computational and Mathematical Methods in Medicine, 2017.

[68] Otberdout, N., Kacem, A., Daoudi, M., Ballihi, L., & Berretti, S. (2018). Deep covariance descriptors for facial expression recognition. arXiv preprint arXiv:1805.03869.

[69] Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. EURASIP journal on wireless communications and networking, 2017(1), 1-12.

[70] Menegola, A., Fornaciali, M., Pires, R., Avila, S., & Valle, E. (2016). Towards automated melanoma screening: Exploring transfer learning schemes. arXiv preprint arXiv:1609.01228.

[71] Lorentzon, M. (2017). Feature extraction for image selection using machine learning.

[72] Li, R., Zhang, W., Suk, H. I., Wang, L., Li, J., Shen, D., & Ji, S. (2014, September). Deep learning based imaging data completion for improved brain disease diagnosis. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 305-312). Springer, Cham.

[73] Zhu, P., Isaacs, J., Fu, B., & Ferrari, S. (2017, December). Deep learning feature extraction for target recognition and classification in underwater sonar images. In

2017 IEEE 56th Annual Conference on Decision and Control (CDC) (pp. 2724-2731). IEEE.

[74] Mohsen, H., El-Dahshan, E. A., El-Horbaty, E. M., & Salem, A. M. (2017). Brain tumor type classification based on support vector machine in magnetic resonance Bibliography 177 images. Annals Of "Dunarea De Jos" University Of Galati, Mathematics, Physics, Theoretical mechanics, Fascicle II, Year IX (XL), (1).

[75] Pereira, Sérgio, et al. "Brain tumor segmentation using convolutional neural networks in MRI images." *IEEE transactions on medical imaging* 35.5 (2016): 1240-1251.

[76] Siar, Halimeh, and Mohammad Teshnehlab. "Diagnosing and classification tumors and MS simultaneous of magnetic resonance images using convolution neural network." *2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*. IEEE, 2019.

[77] Szilagyi, Laszlo, Laszlo Lefkovits, and Balazs Benyo. "Automatic brain tumor segmentation in multispectral MRI volumes using a fuzzy c-means cascade algorithm." *2015 12th international conference on fuzzy systems and knowledge discovery (FSKD)*. IEEE, 2015.

[78] Xu, Yan, et al. "Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation." *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2015.

[79] Pan, Yuehao, et al. "Brain tumor grading based on neural networks and convolutional neural networks." *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015.

[80] Basheera, Shaik, and M. Satya Sai Ram. "Classification of brain tumors using deep features extracted using CNN." *Journal of Physics: Conference Series*. Vol. 1172. No. 1. IOP Publishing, 2019.

[81] Menegola, Afonso, et al. "Towards automated melanoma screening: Exploring transfer learning schemes." *arXiv preprint arXiv:1609.01228* (2016).

[82] Bhandari, Abhishta, Jarrad Koppen, and Marc Agzarian. "Convolutional neural networks for brain tumor segmentation." *Insights into Imaging* 11 (2020): 1-9.

[83] Otberdout, Naima, et al. "Deep covariance descriptors for facial expression recognition." *arXiv preprint arXiv:1805.03869* (2018).

[84] Hashemzehi, Raheleh, et al. "Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE." *Biocybernetics and Biomedical Engineering* 40.3 (2020): 1225-1232.

[85] Seetha, J., and S. Selvakumar Raja. "Brain tumor classification using convolutional neural networks." *Biomedical & Pharmacology Journal* 11.3 (2018): 1457.

[86] Khan, Hassan Ali, et al. "Brain tumor classification in MRI image using convolutional neural network." *Mathematical Biosciences and Engineering* 17.5 (2020): 6203-6216.

[87] Pan Lin ,Yong Yang, Chong-Xun Zheng, Jian-Wen Gu,.An Efficient Automatic Framework for Segmentation of MRI Brain Image, Proceedings of the Fourth.

[88] Krupa, Katarzyna, and Monika Bekiesińska-Figatowska. "Artifacts in magnetic resonance imaging." Polish journal of radiology 80 (2015): 93.

[89] Roy, A., Singha, J., Manam, L., & Laskar, R. H. (2017). Combination of adaptive vector median filter and weighted mean filter for removal of high-density impulse noise from colour images. *IET image processing*, *11*(6), 352-361.

[90] Thangavel, K & Karnan, M 2005, "CAD System for pre-processing and Enhancement of Digital Mammograms", GVIP Journal, vol.5, no.9, pp.69- 74, 2005.J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[91] P. Sahoo, S. Soltani, and A. Wong, "Survey of thresholding techniques," Comp. Vis. Graph. Image Proc., Vol. 41, no. 2,pp. 233_60, 1988.

[92] Arifin, Agus Zainal, and Akira Asano. "Image segmentation by histogram thresholding using hierarchical cluster analysis." *Pattern recognition letters* 27.13 (2006): 1515-1521.

[93] A. Rehman, and D. Mohamad, "A simple segmentation approach for unconstrained cursive handwritten words in conjunction of neural network," Int. J. Image Process., Vol. 2. no. 3, pp. 29_35, 2008.

[94] Staining Pattern Classification of Antinuclear Autoantibodies Based on Block Segmentation in Indirect Immunofluorescence Images, Jiaqian Li, Kuo-Kun Tseng, Zu Yi Hsieh, Ching Wen Yang, and Huang-Nan Huang, PLoS One. 2014; 9(12): e113132., Published online 2014 Dec 4. doi: 10.1371/journal.pone.0113132.

[95] K. Karibasappa, S. Patnaik, Face recognition by ANN using wavelet transform coefficients, IE (India) J. Computer Eng. 85 (2004) 17–23.

[96] P.S. Hiremath, S. Shivashankar, Jagadeesh Pujari, Wavelet based features for color texture classification with application to CBIR, Int. J. Computer Sci. Network Sec. 6 (9A) (2006) 124–133.

[97] Al-Haj, Ali. "Combined DWT-DCT digital image watermarking." Journal of computer science 3.9 (2007): 740-746.

[98] Duda R., Hart P., and Stork D. Pattern classification. Wiley, New York, 2001.

[99] Bishop C.M. Neural networks and pattern recognition. Oxford University Press, Oxford, 1995

[100] Rady, El-Houssainy A., and Ayman S. Anwar. "Prediction of kidney disease stages using data mining algorithms." Informatics in Medicine Unlocked 15 (2019): 100178.

[101] Kecman V. Learning and Soft Computing, Support Vector Machines, Neural Networks, and Fuzzy Logic Models. Cambridge: MIT Press, 2001.

[102] Luts, Jan, et al. Classification of brain tumors based on magnetic resonance spectroscopy. Diss. PhD thesis. Faculty of Engineering, KU Leuven, Leuven, Belgium, 2010

[103] Yi Lu. Knowledge integration in a multiple classifier system. Applied Intelligence, 6(2):75–86, 1996.

[104] Ranawana, Romesh, and Vasile Palade. "Multi-classifier systems: Review and a roadmap for developers." *International journal of hybrid intelligent systems* 3.1 (2006): 35-61.

[105] L.I. Kuncheva, J.C. Bezdek, R.P.W. Duin, Decision templates for multiple classifier fusion: an experimental comparison, Pattern Recognition 34 (2001) 299–314.

[106] L.A. Alexandre, A.C. Campilho, M. Kamel, On combining classifiers using sum and product rules, Pattern Recognition Lett. 22 (2001) 1283–1289

[107] Ponti Jr, Moacir P. "Combining classifiers: from the creation of ensembles to the decision fusion." Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2011 24th SIBGRAPI Conference on. IEEE, 2011. Bibliography 175

[108] J. Kittler, M. Hatef, R.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 226–239.

[109] Bloch, Isabelle. "Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account." *Pattern Recognition Letters* 17.8 (1996): 905-919.

[110] Rancesco Gargiulo, Claudio Mazzariello, and Carlo Sansone. Multiple classifier systems:theory, applications and tools. In Handbook on Neural Information Processing, pages 335–378. Springer, 2013.

[111] Voorbraak, Frans. "A computationally efficient approximation of Dempster-Shafer theory." *International Journal of Man-Machine Studies* 30.5 (1989): 525-536.

[112] Foley, Bethany G. "A Dempster-Shafer Method for Multi-Sensor Fusion." (2012).

[113] Monney, Paul-Andre, and Moses Chan. "Modelling dependence in Dempster-Shafer theory." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 15.01 (2007): 93-114.

[114] Al-Ani, Ahmed, and Mohamed Deriche. "A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence." Journal of Artificial Intelligence Research 17 (2002): 333-361.

[115] Quost, Benjamin, Marie-Hélène Masson, and Thierry Denoeux. "Classifier fusion in the Dempster–Shafer framework using optimized t-norm based combination rules." *International Journal of Approximate Reasoning* 52.3 (2011): 353-374. [153] Altman D.G. and Royston P. What do we mean by validating a prognostic model? Statistics in Medicine, 19:453–473, 2000.

[116] Valente, Fabio, and Hynek Hermansky. "Combination of acoustic classifiers based on Dempster-Shafer theory of evidence." 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. Vol. 4. IEEE, 2007.

[117] Dezert, Jean, Pei Wang, and Albena Tchamova. "On the validity of Dempster-Shafer theory." *2012 15th International Conference on Information Fusion*. IEEE, 2012.

[118] Mingqiang, Yang, Kpalma Kidiyo, and Ronsin Joseph. "A survey of shape feature extraction techniques." *Pattern recognition* 15.7 (2008): 43-90.

[119] Chen, Yushi, et al. "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks." *IEEE Transactions on Geoscience and Remote Sensing* 54.10 (2016): 6232-6251.

[120] Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review." *Neural computation* 29.9 (2017): 2352-2449.

[121] Zhou, Yu, et al. "Polarimetric SAR image classification using deep convolutional neural networks." *IEEE Geoscience and Remote Sensing Letters* 13.12 (2016): 1935-1939.

[122] Minaee, S., Abdolrashidi, A., & Wang, Y. (2015, August). Iris recognition using scattering transform and textural features. In 2015 IEEE signal processing and signal processing education workshop (SP/SPE) (pp. 37-42). IEEE.

[123] Wang, H., & Raj, B. (2017). On the origin of deep learning. arXiv preprint arXiv:1702.07800. Bibliography 176

[124] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

[125] Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295-2329.

[126] Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Greedy layer-wise training of deep networks. In Advances in neural information processing systems (pp. 153-160).

[127] Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models With Big Data. Front. Artif. Intell, 3(4).

[128] Deng, L. (2018). Artificial intelligence in the rising wave of deep learning: The historical path and future outlook [perspectives]. IEEE Signal Processing Magazine, 35(1), 180-177.

[129] Strogatz, S. (2018). One giant step for a chess-playing machine. New York Times.

[130] Siar, Halimeh, and Mohammad Teshnehlab. "Diagnosing and classification tumors and MS simultaneous of magnetic resonance images using convolution neural network." *2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*. IEEE, 2019.

[131] Szilagyi, Laszlo, Laszlo Lefkovits, and Balazs Benyo. "Automatic brain tumor segmentation in multispectral MRI volumes using a fuzzy c-means cascade algorithm." *2015 12th international conference on fuzzy systems and knowledge discovery (FSKD)*. IEEE, 2015.

[132] Gong, Yunchao, et al. "Compressing deep convolutional networks using vector quantization." *arXiv preprint arXiv:1412.6115* (2014).

[133] Pan, Yuehao, et al. "Brain tumor grading based on neural networks and convolutional neural networks." *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015.

[134] Fiore, Ugo, et al. "Network anomaly detection with the restricted Boltzmann machine." *Neurocomputing* 122 (2013): 13-23.

[135] Zhao, Zhiqiang, et al. "Discriminant deep belief network for high-resolution SAR image classification." *Pattern Recognition* 61 (2017): 686-701.

[136] Bhandari, Abhishta, Jarrad Koppen, and Marc Agzarian. "Convolutional neural networks for brain tumor segmentation." *Insights into Imaging* 11 (2020): 1-9.

[137] Otberdout, Naima, et al. "Deep covariance descriptors for facial expression recognition." *arXiv preprint arXiv:1805.03869* (2018).

[138] Hashemzehi, Raheleh, et al. "Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE." *Biocybernetics and Biomedical Engineering* 40.3 (2020): 1225-1232.

[139] Seetha, J., and S. Selvakumar Raja. "Brain tumor classification using convolutional neural networks." *Biomedical & Pharmacology Journal* 11.3 (2018): 1457.

[140] Wu, Jianxin. "Introduction to convolutional neural networks." *National Key Lab for Novel Software Technology. Nanjing University. China* 5.23 (2017): 495.

[141] Oyedotun, O., & Khashman, A. (2017). Iris nevus diagnosis: convolutional neural network and deep belief network. Turkish Journal of Electrical Engineering & Computer Sciences, 25(2), 1106-1115.

[142] Scherer, D., Müller, A., & Behnke, S. (2010, September). Evaluation of pooling operations in convolutional architectures for object recognition. In International conference on artificial neural networks (pp. 92-101). Springer, Berlin, Heidelberg.

[143] Adams, Niall M., and David J. Hand. "Improving the practice of classifier performance assessment." *Neural computation* 12.2 (2000): 305-311.

[144] Foody, Giles M. "Status of land cover classification accuracy assessment." *Remote sensing of environment* 80.1 (2002): 185-201.

[145] Parker, J. R. "Rank and response combination from confusion matrix data." *Information fusion* 2.2 (2001): 113-120. [146] Hawkins, Douglas M. "The problem of overfitting." *Journal of chemical information and computer sciences* 44.1 (2004): 1-12.

[147] Schaffer, Cullen. "Selecting a classification method by cross-validation." *Machine Learning* 13.1 (1993): 135-143.

[148] Guyon, Isabelle, and André Elisseeff. "An introduction to feature extraction." *Feature extraction*. Springer, Berlin, Heidelberg, 2006. 1-25.

[149] Trevethan, Robert. "Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice." *Frontiers in public health* 5 (2017): 307.

[150] Yacouby, Reda, and Dustin Axman. "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models." Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems. 2020.

[151] Subramaniam, Tharani. "Tuning algorithms for pid controller using so computing techniques." *IJCSNS* 8.4 (2008): 278.

List of Publications

1. Shantta, Kalifa, & Otman Basir. "Brain Tumor Detection and Segmentation: A Survey." IRA-International Journal of Technology & Engineering (ISSN 2455-4480) [Online], 10.4 (2018): 55-61. Web. 30 Apr. 2020.
2. Shantta, Kalifa, & Otman Basir. "Brain Tumor Diagnosis Support System: A Decision Fusion Framework." *I*RA-International Journal of Applied Sciences (ISSN 2455-4499) [Online], 15.3 (2020): 30-47. Web. 30 Apr. 2021.
3. Kalifa Shantta , & Basir, Otman . "Deep Learning Feature Extraction for Brain Tumor Characterization and Detection." IRA-International Journal of Applied Sciences (ISSN *2455-4499)* [Online], 16.1 (2021): 1-15. Web. 30 Apr. 2021.
4. Kalifa Shantta , & Basir, Otman . "Automatic MRI Brain Tumor Segmentation Techniques: A Survey." IRA-International Journal of Applied Sciences (ISSN 2455-4499) [Online], 16.2 (2021): 25-38. Web. 30 Apr. 2021.