ENHANCED K-NEAREST NEIGHBOURS CLASSIFICATION PERFORMANCE BASED ON SEGMENTATION AND IMPUTATION OF MISSING DATA

SOOBIA SAEED

UNIVERSITI TEKNOLOGI MALAYSIA

ENHANCED K-NEAREST NEIGHBOURS CLASSIFICATION PERFORMANCE BASED ON SEGMENTATION AND IMPUTATION OF MISSING DATA

SOOBIA SAEED

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

Faculty of Computing
Universiti Teknologi Malaysia

NOVEMBER 2022

DEDICATION

This thesis is dedicated to my beloved Father **Sheikh Saeed Ahmed (late)** for his love, concern and support to make sure I achieve higher targets.

This thesis is dedicated to my my mother Mrs. Nargis Saeed and supervisor Prof. Dr. Habibollah Haron, who support me that the best kind of knowledge to have is that which is learned for its own sake.

No words can express my gratitude for my mentor and guiders Prof. Dr. Mehmood Naqvi who taught me that even the largest task can be accomplished if it is done one step at a time. He also always been in the forefront of love, concern, dua, advice and guidance to make sure I achieve the best of both worlds.

ACKNOWLEDGEMENT

First of all, all praise and glory to the Almighty ALLAH (Subhanahu Wa Ta'ala) who gave me the courage and patience to carry out this work. Peace and blessings of ALLAH be upon His Last Prophet MUHAMMAD (Sallulaho-Alaihe-Wassalam).

In the course of this research, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts throughout my research. In particular, I would like to express my gratitude to my supervisor, Prof. Dr. Habibollah Haron for his guidance, advice, patience, encouragement and support to complete this work. Prof. Dr. Habibollah Haron play an important role in my PhD journey. He accept my work especially that time when I was stuck in my research and guide me the right direction of research. I feel bless and great full that he is my supervisor. The things which I learn here with my supervisor are very beneficial in my career and I also acknowledge to my co-supervisor, Prof. Dr. Naweed. I. Syed for his guidance, advice, patience, encouragement and support to complete this work and I also acknowledge to my mentor Prof. Dr. Mehmmod Naqvi who guide and support me a lot for developing the understanding and learning criteria of my area of research that can I achieved my contributions.

I am extremely grateful to my family for their generous support throughout my PhD, and indeed the whole of my life. Their wisdom, kindness and excellent advice have made me who I am today and I hope that I have made them proud. Finally, I would like to thank all the persons who are not mentioned here, but have been a great help during my period of study.

ABSTRACT

Diagnosing data or object classification for magnetic resonance images is important in image segmentation especially data which is less effective to be identified namely low-grade tumors or cerebrospinal fluid (CSF).The aim of this thesis is to address the aforementioned problems associated with missing data in MRI images and noisy of MRI images that required more processing times. This thesis focus on segmentation of brain tumor and CSF classification of four-dimensional MRI images. Three datasets called Light Field Database (LFD) with improved accuracy of images and increased resolution have been created. A hybrid k-nearest neighbours (k-NN) framework with time complexity that consists of three techniques namely GrabCut support vector machine (GCSVM) and scale invariant feature transform (SIFT), hidden Markov model of k-mean clustering (HMkC) and k-NN, and correlation matrices of discrete Fourier transform (CM-DFT) have been proposed. Firstly, GCSVM and SIFT technique is a combination of three methods namely the GrabCut, Support Vector Machine and Scale Invariant Feature Transform. This result of the technique is 99.9% for SVM accuracy, 4606 for GrabCut segmentation of Maximum Flow, 50625 and 50168 for Nodes of Image Pixel and edges respectively, and 2.29 seconds for computational time. For SIFT by using LFD dataset, the performance of distance value in the segmentation is 1.464, 1.215 and 1.23 for dataset-I, dataset-II, dataset-III respectively. Meanwhile, computational time for dataset-I, dataset-II and dataset-III is 1.47 seconds, 1.88 seconds, and 1.35 seconds respectively. Secondly, HMkC and k-NN resolves the classification problem using the Iterated Condition Mode (ICMM) with k-mean clustering algorithm and k-NN algorithm. The classification result of the technique for the accuracy, sensitivity, specificity and computational time is 99.83%, 99.99%, 99.8%, and 14.9 seconds respectively. Thirdly, CM-DFT technique resolves the missing data imputation problem by using cross correlation of lagged hybrid k-NN with DFT (Hk-NN-DFT) to enhance the MRI images. The technique generates the not a non-missing values in terms of multiplication of 1100-3000 and 99.84% for the accuracy of missing data in the image. The missing ratio result of imputed missing data in the images after retrieving the missing ratio of dataset-I, II, and III is 0.9815 with the 1.533 second of computational time. These three techniques are useful to improve the proposed hybrid k-NN framework to ensure that the classification of brain tumor (low grade tumors) and CSF in images is conducted easily.

ABSTRAK

Mendiagnosis data dan klasifikasi objek bagi imej resonans magnetik (MRI) sangat penting dalam segmentasi imej terutama data yang kurang efektif untuk dikenalpasti iaitu tumor gred-rendah atau cecair serebrospinal (CSF). Matlamat tesis ini adalah untuk menangani masalah yang disebutkan di atas yang berkaitan dengan kehilangan data dalam imej MRI dan hingar dalam imej MRI yang memerlukan lebih masa pemprosesan. Tesis ini fokus kepada segmentasi tumor otak dan klasifikasi CSF bagi imej MRI empat dimensi. Tiga set data dipanggil Pangkalan Data Medan Cahaya (LFD) dengan ketepatan imej yang diperbaiki dan resolusi yang ditambah telah dihasilkan. Satu rangka kerja hibrid kejiranan k-terhampir (k-NN) dengan kompleksiti masa yang terdiri daripada tiga teknik iaitu mesin vektor sokongan GrabCut (GCSVM) dan transformasi ciri invarian skala (SIFT), model Markov tersembunyi bagi pengklusteran k-min (HMkC) dan k-NN, dan matriks korelasi transformasi Fourier diskrit (CM-DFT) telah dicadangkan. Pertama, teknik GCSVM and SIFT ialah gabungan tiga kaedah iaitu GrabCut, Mesin Vektor Sokongan dan Transformasi Ciri Invarian Skala. Keputusan teknik ini ialah 99.9% untuk ketepatan SVM, 4606 untuk segmentasi GrabCut bagi Aliran Maksimum, 50625 dan 50168 masing-masing untuk Nod Piksel Imej dan pinggir, dan 2.29 saat untuk masa pengiraan. Bagi SIFT dengan menggunakan set data LFD, prestasi nilai jarak dalam segmentasi ialah 1.464, 1.215 dan 1.23 masing-masing untuk set data t-I, set data-II, set data-III. Sementara itu, masa pengiraan untuk set data-I, set data-II dan set data-III masing-masing ialah 1.47 saat, 1.88 saat dan 1.35 saat. Kedua, HMkC dan k-NN menyelesaikan masalah klasifikasi menggunakan Mod Keadaan Berulang (ICMM) dengan algoritma pengkelasan k-min dan algoritma k-NN. Keputusan teknik pengkelasan bagi ketepatan, sensitiviti, spesifisiti dan masa pengiraan masing-masing ialah 99.83%, 99.99%, 99.8%, dan 14.9 saat. Ketiga, teknik CM-DFT menyelesaikan masalah imputasi data yang hilang dengan menggunakan korelasi silang k-NN hibrid tertinggal dengan DFT (Hk-NN-DFT) untuk meningkatkan imej MRI. Teknik ini menghasilkan nilai bukan tidak hilang dari segi pendaraban 1100-3000 dan 99.84% untuk ketepatan kehilangan data dalam imej. Keputusan nisbah kehilangan data bagi dalam imej selepas mendapatkan semula nisbah kehilangan data bagi set data-I, II, dan III ialah 0.9815 dengan masa pengiraan 1.533 saat. Ketiga-tiga teknik ini berguna untuk menambah baik rangka kerja hibrid k-NN yang dicadangkan untuk memastikan klasifikasi tumor otak (tumor gred rendah) dan CSF dalam imej boleh dijalankan dengan mudah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	-	One-Dimensional
2D	-	Two-Dimensional
3D	-	Three-Dimensional
4D	-	Four-Dimensional
6DoF	-	With Six Degrees of Freedom
AR	-	Autoregressive
ARMA	-	Autoregressive Moving Average
CM-DFT	-	Correlation Matrices of Discrete Fourier Transform
CSF	-	Cerebrospinal Fluid
CT	-	Computed Tomography
DFT	-	Discrete Fourier Transform
FCM	-	Fuzzy c-mean Algorithms
FFT	-	Fast Fourier Transform
GCDL	-	Globally Consistent Depth Labelling
GLCM	-	Gray Level Co-occurrence Matrix
GM	-	Grey Matter
HMM	-	Hidden Markov Model
HMRF	-	Hidden Markov Random Field
k-NN	-	K-Nearest Neighbour
TC	-	Time Complexity
LFD	-	Light Field Database
LFR		Linear Fractional Representation
LFT	-	Light Field Tool
MA	-	Moving Average
MRI	-	Magnetic Resonance Images
ODE	-	Ordinary Differential Equations
RF	-	Random Forest
SIFT	-	Scale Invariant Feature Transform

SVM	-	Support Vector Machine
SWT	-	Stationary Wavelet Transform
T1	-	Training Set 1
T2	-	Training Set 2
WM	-	White Matter

LIST OF SYMBOLS

Ci	-	Cluster Centre
dω	-	Differential Fourier Transform
V	-	voxel
$Xt \mid Xt - 1$	-	Transition Model
$P(Et \mid Xt)$	-	Sensor Model
$\omega_0=2\pi/T$	-	Fundamental Frequency

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

Digital image processing involves the use of digital computers to enhance images by employing algorithms (Chakravorty, 2018). Digital image processing has many benefits over analog image processing. It allows a vast range of algorithms to be applied to the input data and avoids noise production or distortion problem during processing. Several digital image processing applications exist in the military, agriculture, industry, and medical sciences (Rafael C, and Gonzalez, 2018). Medical images are captured and stored digitally (Elnakib *et al.*, 2011), and the use of digital image processing in medical sciences has become necessary for identifying different types of diseases such as brain tumors and cerebral spinal fluid leakage.

Image segmentation is one of the most exciting and challenging image processing techniques used in medical imaging applications, especially in magnetic resonance imaging (MRI), computed tomography (CT) scans, and X-rays (Khedaskar *et al.*, 2018). Image segmentation is a vital part of image processing and computer vision applications. In recent years, the application of image processing techniques has rapidly increased in the computing and medical fields. However, diagnosing an image for a disease is a tedious and time-consuming task, especially in the identification of an object's regions, edges, and abnormal shapes and colors in MRI or CT scans by radiologists. Image segmentation divides the image into different region forms (Elnakib *et al.*, 2011). Medical image segmentation plays a vital role in many medical applications such as surgery, post-surgical assessment, classification, detecting abnormalities in the human body, etc. (Zhang, 2007).

There are multiple methods of automatic and semi-automatic image segmentation, but most of them have weaknesses because of unknown noise presence, noisy images, poor image contrast, inhomogeneity, and fewer identified boundaries in medical images (Balafar *et al*., 2010). Typically, medical images contain complex structures, and their accurate segmentation results are necessary for clinical diagnosis (Balafar *et al.*, 2010). The precise segmentation of an image is most important for classifying tumor, cancer, edema, cerebrospinal fluid (CSF), and necrotic tissues. One such complicated and challenging procedure is brain image segmentation. MRI is a unique imaging technique that helps radiologists classify the abnormalities in different brain areas in the early stages of their development. However, classifying tumor development in its initial stages or CSF in the brain is difficult to achieve by using MRIs.

Classifying CSF location is one of the biggest challenges in medicine and neurosurgery (Prosser *et al.*, 2011). MRI is less effective in tracking the location of CSF leakage and has difficulty showing where it is deposited in the brain. All MRI noisy image scans are susceptible to artifacts; the previous traditional three dimensional MRI image segmentation method is insufficient to find the location of CSF and low-grade tumor accurately (Kranz *et al.*, 2016). Primarily health professionals suggest nasal fluid to identify the CSF leak by detecting a protein, but still, there is a need for MRI and CT scans so that the position and depth of the tumor and CSF leakage may be found. Due to this problem of low-grade tumor and CSF identification, numerous researchers have tried to find reliable techniques for tumor classification.

Another approach used in digital image processing is machine learning (ML). ML is the study of computer algorithms often characterized by how algorithms learn through experience to make more accurate predictions. There are four basic approaches used in ML: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. The type of algorithm data scientists choose to use depends on what type of data they want to predict. ML algorithms develop a model based on sample data known as training data to make predictions or decisions without being explicitly programmed to do so (Angra and Ahuja, 2017). Several applications of ML algorithms exist in situations where it is challenging to develop conventional algorithms and use them to perform tasks (Diksha *et al.*, 2017).

Tumor classification in MRI images is one of the common issues in the medical field, especially in neurosurgery (Rajasekaran and Gounder, 2018). MRI is a widely used medical technology for diagnosis of various tissue abnormalities and classification of tumors. Various techniques and classification methods have been developed to identify tumors in MRI images but these approaches have weaknesses when used to find a small or low-grade tumor. The active development in computerized medical image segmentation has played a vital role in scientific research. This helps doctors to quickly and easily decide the necessary treatment to give. Brain tumor and CSF leak segmentation is a hot point in the research field of information technology. The process of segmentation for brain tumor segmentation is motivated by assessing tumor growth, treatment responses, computer-based surgery, radiation therapy treatment, and developing tumor growth models. Therefore, a computer-aided diagnostic system is meaningful in medical treatments to reduce the workload of doctors and provide accurate results (Rajasekaran and Gounder, 2018). Due to the problem of accurate result, this thesis tries to improve the classification method. There are three different methods used in this thesis. The first method is producing the support vector machine (SVM) and GrabCut for segmenting low-grade tumors and CSF. The second method is used for feature extraction, matching the image distance and scaling the images using the scale invariant feature transform

(SIFT). The third method is producing a combination of the hidden Markov model (HMM) and k-mean clustering algorithm for classification of tumor and CSF in MRI images. The combination of these methods will create a hybrid technique to classify low-grade tumors and CSF in MRI images.

The k-nearest neighbours (k-NN) is a supervised machine learning algorithm used to interpret non-parametric machine learning algorithms. It produces relatively high quality and competitive results. The algorithm is multipurpose and can be used to figure out classification and regression issues. The k-NN algorithm is used to determine and apply good classifiers to classify tumor or CSF present or absent in MRI images. This thesis applied the k-NN algorithm for classifying segmented MRI images and identifying low-grade tumor pixel and non-tumor pixels as well; the same method was applied for CSF MRI segmented images (Maleki, 2021). HMM is also a good approach for classification although this thesis applied k-NN algorithm to achieve accurate results for low-grade tumor and CSF. The basic idea of this approach is the consecutive use of the HMM and KNN algorithms. To check for the existence or absence of a tumor, the probabilities of both transition states are first calculated using the HMM classifier. The KNN algorithm is used to choose between the two most likely possibilities of classification by HMM. In HMM, the difference between the maximum probability and the second is smaller than the threshold determined by HMM and training samples (Wang and Ju, 2008).

Missing data imputation in MRI images is one of the common problems nowadays. Due to this problem, numerous researchers and scholars try to overcome the issue of missing data through different techniques and improve the k-NN classification. The idea of the k-NN algorithm is not new, but finding the missing values is mainly used in classification (Machhale *et al.*, 2015). There are two methods to cope with tumor and CSF classification in MRI datasets and missing values in k-NN algorithm classification. The first method is producing a hybrid k-NN algorithm for classification and the second is overcoming the missing values in the k-NN algorithm.

Developing and improving a hybrid k-NN algorithm for classification can be helpful to overcome these problems since the k-NN classifier is always easy to obtain for classification. Therefore, improving the k-NN classifier in the learning process is expected to result in higher image accuracy, better tumor and CSF classification, and improved missing data extraction and classification performance. This is the main idea behind all proposed approaches in this thesis. Classification of a hybrid k-NN algorithm is a well-known technique used for tumor and CSF classification to improve the accuracy of an image and reduce the imputation of missing data problem (Dritsas *et al.*, 2018). Hybrid k-NN algorithm for classification aims to classify and remove noisy, nonlinear, and irrelevant data while other techniques are used to extract the missing data in MRI datasets. This thesis imputes the missing data of MRI datasets and increases the accuracy of images in the investigation of low-grade tumor and CSF classification. It improves the performance of the proposed hybrid k-NN based classification framework techniques to classify the missing data efficiently.

1.2 Problem Background

This section attempts to present low-grade tumor and CSF classification with their trends of development over time, and using the same approach of k-NN algorithm for classification. The other approach used and described in this thesis is missing data imputation using the k-NN algorithm. In this regard, this section is divided into three subsections to present the background of classification of low-grade tumor or CSF in MRI through image segmentation approaches, k-NN algorithm classification and their missing values, and removal of irrelevant features and nonlinear data in MRI images.

1.2.1 MRI Segmentation of Tumor and CSF

Various methods have been developed to handle the tumor and CSF segmentation method issues. Several methods already exist for efficient brain tumor segmentation but there is still a need to improve segmentation methods as few of the tumor diagnosing methods is still critical from MRI images to diagnose. The segmentation method extracts different types of tumor tissues like active tissues, tumor, necrosis, and edema from normal brain tissues such as white matter (WM), grey matter (GM), and CSF. Brain tumor can easily be segmented from MRI images but there is a present need for accurate results, reproducible segmentation, and classification of abnormalities which are not anticipated and not visible in MRI. Brain tumor segmentation is composed of multiple stages and the manual segmentation process of a brain tumor in MRI images is a time consumin and tedious task that presents many challenges. Therefore, there is a need for an automated segmentation process for a brain tumor. Multiple techniques exist to evaluate the performance of automated computerized brain tumor segmentation methods in the medical field (Deshmukh and Jadhav, 2014); these are as follows.

Ibrahim *et al.* (2021) proposed a technique that was comprehensive in solving the k-NN classification of tumor in MRI datasets. The most common technique used for solving the classification is the SVM in k-NN algorithm. ML, k-NN and SVM illustrate various crucial transactions (Ibrahim *et al.*, 2021). Even with a list of input variables, the SVM approach predicts exactly for all new observations. Hence, SVM is an instant and ready-to-use approach that many technicians use for brain imaging. SVM is a frequently used technique that fulfils the fundamentals of the k-NN algorithm. It enhances the gap or difference between the categories and leads to efficient overall performance. Therefore, SVM is used in this thesis to achieve accuracy and clarity in the classification.

Kalvakolanu (2021) developed a deep learning approach to classify tumor segmentation and classified tumors into meningioma, glioma, and pituitary tumors. In this research, the author used a skull stripping segmentation-based instrument from the MRI images to remove the skull and GrabCut method to check and verify the accurate classification for the retained feature of tumor from MRI skull stripped masks. This method uses GrabCut segmentation to diagnose tumors and various tissue abnormalities and is extensively used in medical technology worldwide as magnetic resonance imaging. The uses of medical imaging technologies are not only limited to visualizing and examining anatomic structures but they also render their services as tools for surgical and radiotherapy planning, simulation, locating the growth of diseases, and many more. This gives a preoperative plan to remove the tumor safely. To resolve this diagnosing issue, several researchers are suggesting various automated segmentation methods for treating and classifying brain tumor and cerebrospinal fluid leaks.

This thesis applies a supervised four dimensional (4D) light field tool (LFT) segmentation method to the MRI datasets to increase the resolution of images and uses a graph cut algorithm for solving the segmentation of brain tumor. This method employs a graph cut algorithm and has comprehensive details and redundancy. This analysis evaluates spatial and angular neighbouring rays, in order to conserve redundancy in the MRI light field database (LFD) .The other part of this technique is the k-means clustering algorithm which is also adopted for image storing during segmentation. When the MRI images are taken, the clustering k-mean algorithm stores the images and makes the neighbouring values effortlessly visible in the k-NN algorithm which is not possible in other algorithms. The efficiency of the k-NN algorithm is also increased by using MRI datasets with increased accuracy as well (Valdés and Jesús, 2019).

The aforementioned approaches have two issues in common for tumor segmentation. First, they do not entirely segment the tumor in the MRI. Second, in most cases, their performance is highly affected by noisy data which is addressed in this section.

1.2.2 Tumor and CSF Classification from MRI Images

Song *et al.* (2020) proposed a new algorithm to deal with and generate better results of classification for noisy and inherently complex images which is still an existing and challenging problem in the segmentation method. This research has developed a new adaptive hidden markov model to explain the spatial and semantic relationship among pixels of an image. The method uses the HMM for tumor classification based on previous and subsequent segmentation results. This method uses probabilistic reasoning over time and space for brain tumor segmentation of MRI datasets. This spatio-temporal model improves the sensitivity and specificity of segmentation with an image-based transition model but still needs some improvement to accurately find a tumor in the segmentation method as it is still not properly visible in the images due to image noise.

Li *et al.* (2021) discussed an important problem in medical image processing and proposed a new classification technique for brain tumors based on brain magnetic resonance imaging (MRI) results. A computer programme that can quickly and accurately classify tumors in patient brain MRI images and evaluate the data in real time may raise the chance that patients will survive by reducing the time it takes to make a diagnosis. The pituitary tumor, glioma, and meningioma are the three types of brain tumors accurately classified by this research's new statistical method, which is based on MRI images. The feature pixels of MRI images are achieved by the implementation of a collection of convolutional operators to a pixel's neighbouring area in an MRI image that yields the features for that pixel. Furthermore, pixels that are in a tumor's background are also given a statistical profile. Using a dynamic programming approach for classification, the trained HMM is used to assign labels to individual pixels in an MRI image, and the tumor region's labels are analysed to determine the image's classification outcome. There is no need for a large number of computational resources for training and classification processes to be conducted efficiently. MRI images used in the proposed method were analysed to show that it is capable of providing accurate classification results for all three types of brain tumors. A comparison of the proposed method with current methods for classifying brain tumors suggests that it can be more accurate than other methods. Furthermore, real-time analysis indicates that the proposed approach can probably be used to classify brain tumors in real-time.

Huang *et al.* (2011) proposed a new method to deal with and generate better results of classification based on biomedical research and clinical applications which show that water and fat decompose during magnetic resonance imaging (MRI). Using a two-phase approach to solve the three-point water-fat decomposition problem, this research proposes a new approach. The research contribution consists of two main components: The local smoothness of field inhomogeneity is formulated using a background-masked Markov Random Field (MRF) energy model, and the MRF energy model is then optimized using a new Iterated Conditional Modes (ICM) approach. In order to prevent the incorrect propagation of background estimations and to increase efficiency, background masking is integrated with the MRF energy model. This research proposed a new ICM algorithm's Stability Tracking (ST) mechanism, which dynamically monitors pixel iterative stability so that computation of each iteration is only carried out on unstable pixels, which is its key element. ICM efficiency is greatly increased by the ST mechanism. In this research, a median-based initialization approach as well as an adaptive gradient-based method for parametric setup of the MRF model were developed. This technique uses high resolution mouse datasets obtained from 7-Tesla MRI to assess the robustness of the proposed methodology.

The aforementioned approaches have two issues in common for tumor classification. First, they do not entirely classify a tumor in the MRI. Second, in most cases, their performance is highly affected by noisy data which is addressed in this section.

1.2.3 k-NN Algorithm and Its Missing Data Imputation

One of the simplest and effective classification methods is the k-NN algorithm. It is a lazy learning method because of its lower accuracy and as it depends on choosing a good value for "k", it cannot be used for large repositories like in dynamic web mining. Machine learning and k-NN algorithm are used in medical areas where the existence of missing data is a main and serious issue. Knowledge or information is drawn out according to the accuracy of data; if some values are missing in the data, it will affect the descriptive statistics along with inferential statistics and predictive analytics as well. Moreover, for data imputation, k-NN serves as an effective approach to predict the missing values. A model is developed for every feature that has missing values and taken as input values or perhaps for all other input values. K-nearest neighbour model is one of the popular techniques in which a new sample is imputed by determining the closest training set samples and the nearby points are averaged to complete the value (Po *et al.*, 2020).

Nowadays, many researchers are struggling to handle the problem of missing data values in the field of image segmentation. It is one of the most challenging problems in the field of research. Many reasons give rise to missing values. When an algorithm is applied, these missing values provide insufficient and irrelevant data thus leading to invalid results and influencing the performance of any method. There are several imputation techniques based on the nature of the missing values. The demand for a proper technique arises as the data is getting bigger periodically so there is a higher chance of missing values (Po *et al.*, 2020). This research explains the details of the existing models for solving the problem of the missing value, which helps in developing and applying a new method. As these are based on earlier research, they do not currently provide good outcomes in a 4D segmentation method. Many issues of missing values of the k-NN algorithm will be included in this research. Therefore, the main purpose of this research is to counter these missing values with minimum data loss, and tasks are performed under these demands.

Chu *et al.* (2008) developed a new technique to deal with imputation problems in the k-NN algorithm. Missing data imputation of MRI datasets in the k-NN algorithm with Fourier transformation (FT) was also employed. The k-NN algorithm locates a complete data set for k number of neighbours or more identical cases that have patterns similar to the missing data row and column. There are many missing values in biomedical and clinical data of patients due to many circumstances like disassociation among various institutions, failure of images and sensor devices, etc. (Niranjana *et al.*, 2021). Biased, invalid, and wrong outcomes will result if these missing values are not taken into notice. These dependencies occur over time, but the existing methods eventually have to integrate these secular connections and numerous samples of missingness. This research deals with the problem by suggesting an imputation method, namely Fourier k-nearest neighbours (Fk-NN), which is a combination and extension of two imputation methods, k-NN algorithm and Fourier transform. If all the data at a certain time point is missing and if different missing types appear within and across variables, this proposed method will allow the imputation of missing values.

Based on previous studies, a combination of the Fk-NN technique covers the imputation of missing data problems. However, these techniques still have some problems. Although Fk-NN performs better than other techniques, it only finds the missing data in one way at a time either rows or columns (Niranjana *et al.*, 2021). Fk-NN not only reduces the missing data problem but also removes the irrelevant features and nonlinear data which have been created due to incorrectly using Fk-NN. The presence of irrelevant features and nonlinear data is harmful for k-NN algorithm classification as it is creating errors during the reconstruction process. For instance, Fk-NN does not provide the best strategy for reducing the missing data in k-NN algorithm classification. Likewise, this classification method does not provide a robust technique to identify missing data efficiently in the k-NN classification procedure. However, for the missing data imputation issues, most studies had constructed other techniques and classifiers using the Fk-NN technique without considering image accuracy improvement.

1.3 Problem Statement

The images used in classification play an important role in achieving a higher accuracy of classification. Therefore, segmentation is important so that a higher accuracy of classification can be achieved for low-grade tumor and CSF MRI images. Since the structural changes of the tumor interact with other normal tissues, separate segmentation of each target of the tumor images would not be efficient for creating the maximum margin distance for the traditional segmentation method (Gessert and Nil Thorben, 2020). Normally, CT myelography or MR Myelography is more significant than MRI for segmenting the area of CSF leak and low-grade tumor, but there is still a need for MRI and CT scans so that the position and depth of the tumor and leakage may be found (Kranz *et al.*, 2016). The issue of maximum margin distance is a major concern of this thesis to increase the resolution and also remove the noise in MRI images. This thesis aims to segment the CSF leakage and low-grade tumor (the initial stage of brain tumor) in MRI images and improve the resolution of images using the GrabCut algorithm to enhance the maximum margin distance. When using segmentation, improving the accuracy of MRI images requires classification. The previous method cannot identify the exact location of a tumor and CSF. Therefore, identification of location is important based on segmented images. The accuracy of classification can be improved using the Hidden Markov Model with k-mean and k-NN. Therefore, to validate the result of classification, k-NN is used in terms of accuracy, since k-NN does not deal with missing data imputation. This issue can be solved by the correlation method of time-lag with k-NN and DFT. The correlation matrices reduce the missing data imputation problem efficiently and replace the empty space with numeric values using the time-lagged with k-NN.

Enhanced the performance of hybrid k-NN based classification framework by employing to improve the accuracy of MRI LFD datasets of low-grade tumor and CSF leak and impute the missing values of classified MRI images.

The Research Questions (RQ) are as follows:

i. How to improve the result of segmented MRI images and how to identify exact the location of CSF and low-grade tumor?
ii. How to enhance the accuracy of classification of MRI images and how to validate the result of classification of low-grade tumor and CSF area?
iii. How to reduce the missing data imputation in classified MRI images?

These questions formed the basis for undertaking this thesis and were simultaneously investigated. RQ (i) is linked with objective (i), while RQ (ii) is linked with objective (ii) and RQ (iii) is linked with objective (iii).

1.4 Research Aim and Objective

The aim of this thesis is to develop an enhanced classification method based on k-NN that uses pre-processed segmented data and post-processed imputed missing data of low-grade tumor and CSF. The segmented data is a result of hybrid GCSVM and SIFT technique while the imputed missing data is based on HMkC and k-NN techniques. The following objectives are designed to achieve the aim.

The objectives of the thesis are:

i. To propose a GrabCut support vector machine and scale invariant feature transform (Hybrid GCSVM and SIFT) technique for segmentation of MRI images and identification of location of CSF and low-grade tumor.

ii. To design a Hidden Markov Model with k-mean cluster (HMkC) model for classification of CSF and Hidden Markov Model with k-mean and k-NN model to validate the result of HMkC.

iii. To develop a Correlation Matrix of Discrete Fourier Transform (CM-DFT) technique in reducing the missing data imputation of the improved image in hybrid k-NN based classification framework.

1.5 Research Scope

The research documented in this thesis is within the confines of the following concepts:

i. The MRI images are taken from three different hospitals, namely National Cancer Care Institute (NMI) and Medicare hospital in Pakistan and the Cumming School of Medicine Lab at the University of Calgary, Canada (the source of gathered data is https://cumming.ucalgary.ca/).

ii. Implementations of algorithms are done by MATLAB.

iii. Evaluation of the performance of the models is based on sensitivity, specificity, and accuracy.

1.6 Research Contribution

As mentioned earlier, the available data is related to classify the low-grade tumor and CSF in MRI images. It is difficult to identify low-grade tumor and CSF in the starting phase of disease inside of the brain. However, MRI can classify the tumor or cancer in images easily, but in the beginning phase, a tumor inside the brain is difficult to find and the location of CSF deposit inside the brain is still unknown. This is one of the biggest challenges in the neurosurgery field. Good data with multiple techniques are required to solve this problem. In the medical field, doctors and pathologists use the traditional method of CT myelography or MR Myelography instead of MRI because they provide better results if it is an in-sizable form of tear arises.

Therefore, this thesis creates the improved MRI LFD to solve the lower accuracy MRI images and improve the resolution of MRI images through Lytro Illum Light Field Tool (LFT). This improved LFD helps doctors and pathologists so that they can see a low-grade tumor or find the location of CSF in MRI images. This thesis also solves the problem of missing data of MRI which is an urgent need and removes the irrelevant features and nonlinear data in k-NN algorithm.

The main contributions of this research are achieved after implementing the proposed techniques in hybrid k-NN based classification framework given below:

1. Hybrid GCSVM and SIFT technique is used to improve the segmentation method of low-grade tumor and CSF in MRI images to increase the accuracy of improved LFD MRI datasets. This technique also increases the resolution of the images by developing the maximum margin distance space. SIFT is also included in this technique to identify the location of CSF and tumor as well.
2. A hybrid HMkC technique is used to improve the classification method of low-grade tumor and CSF in images and the validation of classified images using HMkC and k-NN techniques.
3. CM-DFT technique is used to reduce the missing data values efficiently and impute the missing data in the same numbers of rows and columns in sequence. Time Complexity (TC) is used to calculate the execution time of proposed hybrid k-NN based classification framework.

1.7 Thesis Outline

Chapter 1, Introduction, starts with an introduction to the thesis topic. Thereafter, it presents the problem background, problem statement, research aims and objectives, scope and significance of the research. The chapter also describes the organization of the thesis.

Chapter 2, Literature Review, interprets the research done previously, the current body of literature on this research's main topics, and presents the related work.

Chapter 3, Research Methodology, provides the thesis approach, methodology, and a brief review of the framework adopted. It explicates six phases; the first phase explains the problem formulation, research question, objectives and scope. The second phase shows data collection and data preparation, and performance measurement. The third, fourth and fifth phases show the details of the hybrid GCSVM and SIFT techniques, hybrid HMkC and k-NN techniques, and CM-DFT techniques. The sixth phase shows the testing, analysis and evaluation.

Chapter 4, the hybrid GCSVM and SIFT technique, addresses the first objective of the thesis. A Hybrid GCSVM and SIFT technique is proposed which is based on the segmentation method of low-grade tumor and CSF in MRI. This chapter uses three different methods that make up the hybrid technique and also increase the resolution of the proposed LFD MRI images.

Chapter 5, the hybrid HMkC and k-NN technique, addresses the second objective of the thesis. A hybrid HMkC and k-NN technique is proposed which is based on the classification method of low-grade tumor and CSF in MRI. This chapter uses two different methods that make up the hybrid technique and also improve the hybrid k-NN model for classification.

Chapter 6, the CM-DFT technique, addresses the third objective of this thesis. CM-DFT is proposed, which is based on a combination correlation matrix of time-lagged with hybrid k-NN algorithm and DFT. CM-DFT solves the missing imputation problem in the hybrid k-NN framework.

Chapter 7, Conclusion and future works, displays the outcomes and contributions of this analysis, and future enhancement. It also highlights the achievements of the objectives along with the relative performance analysis.

REFERENCES

Abbott, N. J., Pizzo, M. E., Preston, J. E., Janigro, D., and Thorne, R. G. (2018). The role of brain barriers in fluid movement in the CNS: is there a ‘glymphatic’system?. *Acta neuropathologica*, 135(3), 387-407.

Abd-Ellah, M. K., Awad, A. I., Khalaf, A. A., & Hamed, H. F. (2018). Two-phase multi-model automatic brain tumor diagnosis system from magnetic resonance images using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 1(1), 1-10.

Ahmed, M. M., and Mohamad, D. B. (2008). Segmentation of brain MR images for tumor extraction by combining kmeans clustering and perona-malik anisotropic diffusion model. *International Journal of Image Processing*, *2*(1), 27-34.

Aibinu, A. M., Salami, M. J. E., Shafie, A. A., & Najeeb, A. R. (2008). MRI reconstruction using discrete Fourier transform: a tutorial. *World Academy of Science, Engineering and Technology*, 42(1), 179.

Al-Amri, S. S., and Kalyankar, N. V. (2010). Image segmentation by using threshold techniques. *Computer Vision and Pattern Recognition*, 2(5),12-18.

Altaei, M. S. M., and Kamil, S. Y. (2020). Brain tumor detection and classification using SIFT in MRI images. In *AIP conference proceedings*, AIP Publishing LLC, 2292 (1), 1-9.

Al-Qahtani, F. H., and Crone, S. F. (2013). Multivariate k-nearest neighbor regression for time series data—A novel algorithm for forecasting UK electricity demand. In *The 2013 international joint conference on neural networks (IJCNN)*. IEEE 1-18.

Angra, S., and Ahuja, S. (2017). Machine learning and its applications: A review. In *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, IEEE, pp. 57-60.

Angulakshmi, M., and Lakshmi Priya, G. G. (2017). Automated brain tumour segmentation techniques—a review. *International Journal of Imaging Systems and Technology*, 27(1), 66-77.

Avinash S, K. Manjunath, Senthil K.S. (2019). A New Algorithm for Feed Forward Neural Network and K-Nearest Neighbors Classifier for Accurate Detection of Lung Cancer. *Journal of Critical Reviews*, 6(5), 8-17.

Balafar, M. A., Ramli, A. R., Saripan, M. I., & Mashohor, S. (2010). Review of brain MRI image segmentation methods. *Artificial Intelligence Review*, 33(3), 261-274.

Banerjee, S., and Mitra, S. (2020). Novel volumetric sub-region segmentation in brain tumors. *Frontiers in computational neuroscience*, 14, 3-5.

Bauer, S., Wiest, R., Nolte, L. P., and Reyes, M. (2013). A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology*, 58(13), 97-102.

Baumgartner, J., Flesia, A. G., Gimenez, J., & Pucheta, J. (2013). A new approach to image segmentation with two-dimensional hidden Markov models. In *2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence*. IEEE, 213-222.

Beevi, S. Z., Sathik, M. M., and Senthamaraikannan, K. (2010). A robust fuzzy clustering technique with spatial neighborhood information for effective medical image segmentation. *IEEE Publication format, International Journal of Computer Science and Information Security, IJCSIS*, 7(3), 1-10.

Berent, J., and Dragotti, P. L. (2007). Unsupervised Extraction of Coherent Regions for Image Based Rendering, *The British Machine Vision Conference* (*BMVC)*, 1-10.

Beretta, L., and Santaniello, A. (2016). Nearest neighbor imputation algorithms: a critical evaluation. *BMC medical informatics and decision making*, 16(3), 197-208.

Bhide, A. S., Patil, P., and Dhande, S. (2012). Brain Segmentation using Fuzzy C means clustering to detect tumor Region. *International Journal of Advanced Research in Computer Science and Electronics Engineering*, 1(2), 85-90.

Bishop, T. E., and Favaro, P. (2011). The light field camera: Extended depth of field, aliasing, and super resolution. *IEEE transactions on pattern analysis and machine intelligence*, 34(5), 972-986.

Blanchet, L., Krooshof, P. W. T., Postma, G. J., Idema, A. J., Goraj, B., Heerschap, A., and Buydens, L. M. C. (2011). Discrimination between metastasis and glioblastoma multi-forme based on morphometric analysis of MR images. *American journal of neuroradiology*, 32(1), 67-73.

Blumenfeld, H. (2010). Neuroanatomy through clinical cases. *Sunderland: Sinauer Associates*, 976-982.

Bohlen, M., Busch, C. J., Sehner, S., Forterre, F., Bier, J., Berliner, C., and Münscher, A. (2019). Tumor volume as a predictive parameter in the sequential therapy (induction chemotherapy) of head and neck squamous cell carcinomas. *European archives of oto-rhino-laryngology*, 276(4), 1183-1189.

Bolan, J., Hall, E., Clifford, C., & Thurow, B. (2016). Light-field imaging toolkit. *Software X*, *5*, 101-106.

Boykov, Y. Y., and Jolly, M. P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *Proceedings eighth IEEE International conference on computer vision. ICCV 2001*, *IEEE*, 1, 105-112.

Boykov, Y., and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence*, 26(9), 1124-1137.

Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11), 1222-1239.

Brinker, T., Stopa, E., Morrison, J., and Klinge, P. (2014). A new look at cerebrospinal fluid circulation. *Fluids and Barriers of the CNS*, 11(1), 1-16.

Calabrese, C., Poppleton, H., Kocak, M., Hogg, T. L., Fuller, C., Hamner, B., and Gilbertson, R. J. (2007). A perivascular niche for brain tumor stem cells. *Cancer cell*, 11(1), 69-82.

Cerri, S., Hoopes, A., Greve, D. N., Mühlau, M., and Van Leemput, K. (2020). A longitudinal method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. In *Machine Learning in Clinical Neuroimaging and Radio genomics in Neuro-oncology*, Springer, Cham, 119-128.

Chakravorty, P. (2018). What is a Signal?. *IEEE Signal Processing Magazine*. 35 (5): 175–177.

Chang, M. M., Sezan, M. I., Tekalp, A. M., and Berg, M. J. (1996). Bayesian segmentation of multi slice brain magnetic resonance imaging using three-dimensional Gibbsian priors. *Optical Engineering*, 35(11), 3206-3221.

Chatfield, C., and Xing, H. (2019). The analysis of time series: an introduction with R. *Chapman and hall/CRC*. 6(1), 153-204.

Chinnu, A. (2015). MRI brain tumor classification using SVM and histogram based image segmentation. *International Journal of Computer Science and Information Technologies*. 6(2), 1505-1508.

Christ, M. J., and Parvathi, R. M. S. (2012). Segmentation of medical image using K-means clustering and marker controlled watershed algorithm. *European Journal of Scientific Research*, 71(2), 190-194.

Chu, E. (2008). *Discrete and continuous Fourier transforms: analysis, applications and fast algorithms*. Chapman and Hall/CRC, 424-458.

Clarke, L. P., Velthuizen, R. P., Clark, M., Gaviria, J., Hall, L., Goldgof, D., and Brem, S. (1998). MRI measurement of brain tumor response: comparison of visual metric and automatic segmentation. *Magnetic resonance imaging*, 16(3), 271-279.

Davis, J. V., Kulis, B., Jain, P., Sra, S., & Dhillon, I. S. (2007). Information-theoretic metric learning. *In ICML '07: Proceedings of the 24th international conference on Machine learning*, ACM, New York, NY, USA, 209–216.

Deshmukh, R. D., & Jadhav, C. (2014). Study of different brain tumor MRI image segmentation techniques. *International Journal of Science, Engineering and Computer Technology*, *4*(4), 133.

Diksha, S. and Neeraj, K. (2017). A Review on Machine Learning Algorithms, Tasks and Applications, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET).* 6(1), 1548- 1551.

Dong, L., Ogunbona, P., Li, W., Yu, G., Fan, L., and Zheng, G. (2006). A fast algorithm for color image segmentation. In *First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06)* IEEE, 2, 685-688.

Dritsas, Elias, Maria Trigka, Panagiotis Gerolymatos, and Spyros Sioutas. (2020). Trajectory clustering and k-NN for robust privacy preserving spatiotemporal databases. *Algorithms*. 11(12), 207-2010.

Dunlavy, D. M., Kolda, T. G., & Acar, E. (2010). Poblano v1. 0: A MATLAB toolbox for gradient-based optimization. *Sandia National Laboratories, Tech. Rep. SAND2010-1422*.

Dansereau, D. G., Girod, B., and Wetzstein, G. (2019). Supplementary Material LiFF: Light Field Features in Scale and Depth. *In Proceeding: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* Nashville, TN, USA, 1-10.

Elnakib, A., Gimel'farb, G., Suri, J. S., & El-Baz, A. (2011). Multi-Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies. *Springer Science & Business Media*, 1(1), 1-39.

Fang, Z., Tian, R., Jia, Y. T., Xu, T. T., and Liu, Y. (2017). Treatment of cerebrospinal fluid leak after spine surgery. *Chinese Journal of Traumatology*, 20(2), 81-83.

Garg. G., and Garg. R. (2021). Brain Tumor Detection and Classification based on Hybrid Ensemble Classifier. *Computer Vision and Pattern Recognition*, 6(1), 1-16.

Gao, P., Shan, W., Guo, Y., Wang, Y., Sun, R., Cai, J., and Wu, Z. (2022). Development and Validation of a Deep Learning Model for Brain Tumor Diagnosis and Classification Using Magnetic Resonance Imaging. *JAMA Network Open*, *5*(8), 1-11.

Gessert, Nils Thorben. (2020). Deep learning with multi-dimensional medical image data. *PhD diss., Technische Universität Hamburg.*

Globerson, A., and Roweis, S. (2005). Metric learning by collapsing classes. *Advances in neural information processing systems*, *18*, 451-458.

Gonzalez, R.C, and Woods, R.E. (2006). Digital Image Processing. *Upper Saddle River*, NJ 07458, USA: Prentice-Hall, Inc.

Gortler, S. J., Grzeszczuk, R., Szeliski, R., and Cohen, M. F. (1996). The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 43-54.

Greig, D. M., Porteous, B. T., and Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2), 271-279.

Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K. (2020). KNN model-based approach in classification. In *OTM Confederated International Conferences On the Move to Meaningful Internet* Systems Springer, Berlin, Heidelberg, 986-996.

Haney, S. M., Thompson, P. M., Cloughesy, T. F., Alger, J. R., and Toga, A. W. (2001). Tracking tumor growth rates in patients with malignant gliomas: A test of two algorithms. *American Journal of Neuroradiology*, 22(1), 73-82.

Hasan, S. M., & Ahmad, M. (2018). Two-step verification of brain tumor segmentation using watershed-matching algorithm. *Brain informatics*, *5*(2), 1-11.

He, X., & Niyogi, P. (2004). Locality preserving projections. *Advances in neural information processing systems*, *16*(16), 153-160.

Horn, D. R., and Chen, B. (2007). Light shop: interactive light field manipulation and rendering. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games*, 121-128.

Hrydziuszko, O., and Viant, M. R. (2012). Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics*, 8(1), 161-174.

Htun, S. N. N., Zin, T. T., and Tin, P. (2020). Image Processing Technique and Hidden Markov Model for an Elderly Care Monitoring System. *Journal of Imaging*, *6*(6), 49.

Huang, F., Narayan, S., Wilson, D., Johnson, D., & Zhang, G. Q. (2011). A fast iterated conditional modes algorithm for water–fat decomposition in MRI. *IEEE transactions on medical imaging*, *30*(8), 1480-1492.

Ibrahim, and Abdul azeez. A. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends*, 2(1), 10-19.

Ibrahim, W. A., and Morcos, M. M. (2002). Artificial intelligence and advanced mathematical tools for power quality applications: a survey. *IEEE Transactions on Power Delivery*, 17(2), 668-673.

Jafri, M. Z. M., Abdul baqi, H. S., Mutter, K. N., Mustapha, I. S., and Omar, A. F. (2017). Measuring the volume of brain tumor and determining its location in T2-weighted MRI images using hidden Markov random field: expectation maximization algorithm. In *Digital Optical Technologies, International Society for Optics and Photonics*, 10335-103351.

Jarabo, A., Masia, B., Bousseau, A., Pellacini, F., and Gutierrez, D. (2014). How do people edit light fields? *ACM Trans. Graph*, 33(4), 4.

Jinhwan, J. (2020). Data Processing Techniques for Real-Time Traveler Information: Use of Dedicated Short-Range Communications Probes on Suburban Arterial. *The Open Transportation Journal*, *14*(1), 95-112.

Kalvakolanu, A. T. S. (2021). Brain Tumor Detection and Classification from MRI Images. *Master Thesis*, *California Polytechnic State University*, 1-63.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, 24(7), 881-892.

Kaur, A., Kaur, L., and Singh, A. (2021). State-of-the-art segmentation techniques and future directions for multiple sclerosis brain lesions. *Archives of Computational Methods in Engineering*, 28(3), 951-977.

Khalil, H. A., Darwish, S., Ibrahim, Y. M., and Hassan, O. F. (2020). 3D-MRI brain tumor detection model using modified version of level set segmentation based on dragonfly algorithm. *Symmetry*, 12(8), 1256-1268.

Khedaskar, S. V., Rokade, M. A., Patil, B. R., & Tatwadarshi, P. N. (2018). A survey of image processing and identification techniques. *VIVA-Tech International Journal for Research and Innovation*, *1*(1), 1-10.

Kranz, P. G., Luetmer, P. H., Diehn, F. E., Amrhein, T. J., Tanpitukpongse, T. P., and Gray, L. (2016). Myelographic techniques for the detection of spinal CSF leaks in spontaneous intracranial hypotension. *American Journal of Roentgenology*, *206*(1), 8-19.

Kolmogorov, V., and Zabih, R. (2002). Multi-camera scene reconstruction via graph cuts. In *European conference on computer vision*. Springer, Berlin, Heidelberg, 82-96.

Kowdle, A., Sinha, S. N., and Szeliski, R. (2012). Multiple view object co segmentation using appearance and stereo cues. In *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 789-803.

Latha, C., and Perumal, K. (2017). Improved probability based fuzzy C-means and active contour using brain tumor images. In *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI,* IEEE, 880-887.

Lee, C. H., Schmidt, M., Murtha, A., Bistritz, A., Sander, J., and Greiner, R. (2005). Segmenting brain tumors with conditional random fields and support vector machines. In *International Workshop on Computer Vision for Biomedical Image Applications*. Springer, Berlin, Heidelberg, 469-478.

Leo, M. J. (2019). MRI brain image segmentation and detection using K-NN classification. In *Journal of Physics: Conference Series*. 362(1), 1-12.

Levoy, M., and Hanrahan, P. (1996). Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 31-42.

Lalitha, and Ramashri. (2019). Brain MRI Image Segmentation using Grab cut Algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(2), 1418-1421.

Li, G., Citrin, D., Camphausen, K., Mueller, B., Burman, C., Mychalczak, B., and Song, Y. (2008). Advances in 4D medical imaging and 4D radiation therapy. *Technology in cancer research & treatment*, 7(1), 67-81.

Li, G., Sun, J., Song, Y., Qu, J., Zhu, Z., and Khosravi, M. R. (2021). Real-time classification of brain tumors in MRI images with a convolutional operator-based hidden Markov model. *Journal of Real-Time Image Processing*, *18*(4), 1207-1219.

Liang, Z.P. and Lauterbur, P.C. (2000). Principles of Magnetic Resonance Imaging. *A signal processing perspective*, IEEE Press, New York.

Lindeberg, T. (2012). Scale invariant feature transform. *Scholar pedia*, pp.1-12.

Liu, Y. T., Zhang, H. X., & Li, P. H. (2011). Research on SVM-based MRI image segmentation. *The Journal of China Universities of Posts and Telecommunications*, 18, 129-132.

Liu. J and Guo. L. (2015). A new brain MRI image segmentation strategy based on k-means clustering and SVM. In *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, IEEE, 2(1), 270-273.

Logeswari, T., and Karnan, M. (2010). An improved implementation of brain tumor detection using soft computing. In *2010 Second International Conference on Communication Software and Networks*, IEEE, 147-151.

Lowe, D. G. (2004). Distinctive image features from scale-invariant key points. *International journal of computer vision*, *60*(2), 91-110.

Lu, C. Y., Xu, Z. S., and Ye, X. (2019). Evaluation of intraoperative MRI-assisted stereotactic brain tissue biopsy: a single-center experience in China. *Chinese Neurosurgical Journal*, 5(1), 1-8.

Lumsdaine, A., and Georgiev, T. (2009). The focused plenoptic camera. In *2009 IEEE International Conference on Computational Photography (ICCP)*, IEEE, 1-8.

Luo, M., Ma, Y. F., and Zhang, H. J. (2003). A spatial constrained k-means approach to image segmentation. In *Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint*, IEEE, 2, 738-742.

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

Maleki, N., Zeinali, Y., and Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. *Expert Systems with Applications*, 164 (1), 113981-11981.

Maeno, K., Nagahara, H., Shimada, A., and Taniguchi, R. I. (2013). Light field distortion feature for transparent object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2786-2793.

Marsh, M. (2006). Implementing the “GrabCut” Segmentation Technique as a Plug-in for the GIMP. *Rhodes University*, 1-59.

Martinez, F., Charte, F., Frías, M. P., & Martínez-Rodríguez, A. M. (2021). Strategies for time series forecasting with generalized regression neural networks. *Neurocomputing*, 1-13.

Martínez-Más, J., Bueno-Crespo, A., Khazendar, S., Remezal-Solano, M., Martínez-Cendán, J. P., Jassim, S., and Timmerman, D. (2019). Evaluation of machine learning methods with Fourier Transform features for classifying ovarian tumors based on ultrasound images. *PLoS One*, *14*(7), e0219388.

Mashiat, A., Akhlaque, R. R., Fariha, F. H., and Patwary, M. S. H. (2020). *Comparison of different CNN architectures for brain tumor detection using fMRI* (Doctoral dissertation, Brac University), 1-155.

Mild, H. K., Lundström, R., and Wilén, J. (2019). Non-Ionizing radiation in Swedish health care—exposure and safety aspects. *International journal of environmental research and public health*, 16(7), 1186.

Mustaqeem, A., Javed, A., and Fatima, T. (2012). An efficient brain tumor detection algorithm using watershed & thresholding based segmentation. *International Journal of Image, Graphics and Signal Processing*, 4(10), 34-42.

Nagahashi, T., Fujiyoshi, H., and Kanade, T. (2009). Video segmentation using iterated graph cuts based on spatio-temporal volumes. In *Asian Conference on Computer Vision*, Springer, Berlin, Heidelberg, 655-666.

Natarajan, P., Krishnan, N., Kenkre, N. S., Nancy, S., and Singh, B. P. (2012). Tumor detection using threshold operation in MRI brain images. In *2012 IEEE* *international conference on computational intelligence and computing* research, IEEE,1-4.

Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. (2005). Light field photography with a hand-held plenoptic camera, *Doctoral dissertation, Stanford University*, 1-51.

Niranjana. G., and Chatterjee. D. (2021). Security and Privacy Issues in Biomedical AI Systems and Potential Solutions. In *Handbook of Artificial Intelligence in Biomedical Engineering*, Apple Academic Press, 289-309.

Orczyk, T., Doroz, R., and Porwik, P. (2021). Missing Value Imputation Method Using Separate Features Nearest Neighbors Algorithm. In *International Conference on Computational Science*, Springer, Cham, 128-141.

Osher, S., and Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, 79(1), 12-49.

Pan, W., Gu, W., Nagpal, S., Gephart, M. H., and Quake, S. R. (2015). Brain tumor mutations detected in cerebral spinal fluid. *Clinical chemistry*, 61(3), 514-522.

Pascal, N. E., Pierre, E., and Emmanuel, T. (2014) Hybrid Method Segmentation for Medical Image Based on DWT, FCM and HMRF-EM. *International Journal of Computer and Information Technology*, 3(2), 275-282.

Po, Laura, Nikos Bikakis, Federico Desimoni, and George Papastefanatos. (2020). Linked Data Visualization: Techniques, Tools, and Big Data. *Synthesis Lectures on Semantic Web: Theory and Technology*, 10(1), 1-157.

Popoola, J. J., Godson, T. E., Olasoji, Y. O., and Adu, M. R. (2019). Study on capabilities of different segmentation algorithms in detecting and reducing brain tumor size in magnetic resonance imaging for effective telemedicine services. *European Journal of Engineering and Technology Research*, 4(2), 23-29.

Pradhan, S. (2010). Development of Unsupervised Image Segmentation Schemes for Brain MRI using HMRF model. *National Institute of Technology Rourkela, Orissa*, India, 1-113.

Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., and Gerig, G. (2003). Automatic brain tumor segmentation by subject specific modification of atlas priors1. *Academic radiology*, 10(12), 1341-1348.

Prosser, J. D., Vender, J. R., & Solares, C. A. (2011). Traumatic cerebrospinal fluid leaks. *Otolaryngologic Clinics of North America*, 44(4), 857-873.

Pushpa, B. R., and Athira, P. R. (2021). Plant Species Recognition Based on Texture and Geometric Features of Leaf. In *2021 3rd International Conference on Signal Processing and Communication (ICPSC)* IEEE, 315-320.

Rafael C, and Gonzalez. (2018). Digital image processing. *University of Tennessee Richard E. Woods, Med Data Interactive New York, NY: Pearson*. 4(1), 120-118.

Rahman, C. M., and Rashid, T. A. (2019). Dragonfly algorithm and its applications in applied science survey. *Computational Intelligence and Neuroscience*, 2(3),1-12.

Rahman, S. A., Huang, Y., Claassen, J., Heintzman, N., and Kleinberg, S. (2015). Combining Fourier and lagged k-nearest neighbor imputation for biomedical time series data. *Journal of biomedical informatics*, 58, 198-207.

Raja, P. S., and Thangavel, K. (2020). Missing value imputation using unsupervised machine learning techniques. *Soft Computing*, 24(6), 4361-4392.

Raja, S.G. and Nirmala, K. (2019). Detection of brain tumor using K-nearest neighbor (KNN) based classification model and self-organizing map (SOM) algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE).* 8(8), 787-791.

Rajalakshmi, N., and Prabha, V. L. (2012). Brain tumor detection of MRI Images based on color-converted hybrid pso+ k-means clustering segmentation. *European Journal of Scientific Research*, 5-14.

Rajasekaran, K. A., and Gounder, C. C. (2018). Advanced Brain Tumor Segmentation from MRI Images. *Basic Physical Principles and Clinical Applications, High-Resolution Neuroimaging*, 83-108.

Roa, C., Du Le, V. N., Mahendroo, M., Saytashev, I., and, Ramella-Roman, J. C. (2021). Auto-detection of cervical collagen and elastin in Mueller matrix polarimetry microscopic images using K-NN and semantic segmentation classification. *Biomedical Optics Express*, *12*(4), 2236-2249.

Ray, S., and Turi, R. H. (1999). Determination of number of clusters in k-means clustering and application in color image segmentation. In *Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*, 137-143.

Rother, C., Kolmogorov, V., and Blake, A. (2004). “GrabCut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, 23(3), 309-314.

Roy, S., Nag, S., Maitra, I. K., and Bandyopadhyay, S. K. (2013). A review on automated brain tumor detection and segmentation from MRI of brain. *Computer Vision and Pattern Recognition*, 5(2), 1-9.

Ruan, S., Lebonvallet, S., Merabet, A., and Constans, J. M. (2007). Tumor segmentation from a multispectral MRI images by using support vector machine classification. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 1236-1239.

Rustin, G. J., Quinn, M., Thigpen, T., Du Bois, A., Pujade-Lauraine, E., Jakobsen, A., and Vermorken, J. (2004). Re: New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). *Journal of the National Cancer Institute*, 96(6), 487-488.

Sajid, S., Hussain, S., and Sarwar, A. (2019). Brain tumor detection and segmentation in MR images using deep learning. *Arabian Journal for Science and Engineering*, 44(11), 9249-9261.

Sakka, L., Coll, G., and Chazal, J. (2011). Anatomy and physiology of cerebrospinal fluid. *European annals of otorhinolaryngology, head and neck diseases*, 128(6), 309-316.

Sanjar, K., Bekhzod, O., Kim, J., Paul, A., & Kim, J. (2020). Missing Data Imputation for Geolocation-based Price Prediction Using KNN–MCF Method. *ISPRS International Journal of Geo-Information, 9*(4), 227-246.

Sattar, S., Alibhai, S. M., Spoelstra, S. L., and Puts, M. T. (2019). The assessment, management, and reporting of falls, and the impact of falls on cancer treatment in community-dwelling older patients receiving cancer treatment: results from a mixed-methods study. *Journal of geriatric oncology*, 10(1), 98-104.

Sedghi, A., O'Donnell, L. J., Kapur, T., Learned-Miller, E., Mousavi, P., and Wells. (2021). Image registration: Maximum likelihood, minimum entropy and deep learning. *Medical Image Analysis*, 69(1), 101-112.

Selvy, P. T., Palanisamy, V., and Purusothaman, T. (2011). Performance analysis of clustering algorithms in brain tumor detection of MR images. *European Journal of Scientific Research*, 62(3), 321-330.

Sharma, S., & Rattan, M. (2019). An improved segmentation and classifier approach based on hmm for brain cancer detection. *The Open Biomedical Engineering Journal*, 13(1), 1-18.

Shen, S., Sandham, W. A., and Granat, M. H. (2003). Preprocessing and segmentation of brain magnetic resonance images. In *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, IEEE, 149-152.

Shirly, S., and Ramesh, K. (2019). Review on 2D and 3D MRI image segmentation techniques. *Current Medical Imaging*, 15(2), 150-160.

Shokouhifar, M., and Abkenar, G. S. (2011). An artificial bee colony optimization for mri fuzzy segmentation of brain tissue. In *2011 International Conference on Management and Artificial Intelligence IPEDR*, 6-10.

Silva, A. (2016). Light field processor: a lytro illum imaging application. *University Institute of Lisbon*, 1-180.

Sivapriya, T. R., Kamal, A. N. B., & Thavavel, V. (2012). Imputation and classification of missing data using least square support vector machines–a new approach in dementia diagnosis. *International journal of advanced research in artificial intelligence*, 1(4), 29-33.

Sokołowski, A., and Pardela, T. (2014). Application of Fourier transforms in classification of medical images. In *Human-computer systems interaction: backgrounds and applications,* Springer, Cham, 3, 193-200.

Solomon, J., Butman, J. A., & Sood, A. (2006). Segmentation of brain tumors in 4D MR images using the hidden Markov model. *Computer methods and programs in biomedicine*, 84(2-3), 76-85.

Song, Y., Adobah, B., Qu, J., & Liu, C. (2020). Segmentation of ordinary images and medical images with an adaptive hidden Markov Model and Viterbi Algorithm. *Current Signal Transduction Therapy*, *15*(2), 109-123.

Spector, R., Snodgrass, S. R., and Johanson, C. E. (2015). A balanced view of the cerebrospinal fluid composition and functions: focus on adult humans. *Experimental neurology*, 273, 57-68.

Taheri, S., Ong, S. H., and Chong, V. F. H. (2010). Level-set segmentation of brain tumors using a threshold-based speed function. *Image and Vision Computing*, 28(1), 26-37.

Tatiraju, S., and Mehta, A. (2008). Image Segmentation using k-means clustering, EM and Normalized Cuts. *Department of EECS*, 1, 1-7.

Telano, L. N., and Baker, S. (2020). Physiology, Cerebral Spinal Fluid (CSF). *Stat Pearls*, 12-24.

Toufiq, D. M., Sagheer, A. M., and Veisi, H. (2021). A Review on Brain Tumor Classification in MRI Images. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(14), 1958-1969.

Tzika, A. A., Astrakas, L., and Zarifi, M. (2011). Pediatric Brain Tumors: Magnetic Resonance Spectroscopic Imaging. *Diagnostic Techniques and Surgical Management of Brain Tumors*, 205-212.

Ullah, Z., Farooq, M. U., Lee, S. H., and An, D. (2020). A hybrid image enhancement based brain MRI images classification technique. *Medical hypotheses*, *143*, 1-8.

Valdés. V and Alejandro. J. (2019). Deep Learning in Medical Image Analysis: Alzheimer's Disease Classification & Whole Fetal Brain Segmentation. *Institute of Technology de monetary*, 1-112.

Vasuda, P., and Satheesh, S. (2010). Improved fuzzy C-means algorithm for MR brain image segmentation. *International Journal on Computer Science and Engineering*, 2(5), 12-18.

Vicente, S., Kolmogorov, V., and Rother, C. (2008). Graph cut based image segmentation with connectivity priors. In *2008 IEEE conference on computer vision and pattern recognition*, 1-8. IEEE.

Wang, D. (2019). Efficient level-set segmentation model driven by the local GMM and split Bregman method. *IET Image Processing*, 13(5), 761-770.

Wang, L. (2019). Research and Implementation of Machine Learning Classifier Based on KNN. In *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 677(5), 1-12.

Wang, Q., and Ju, S. (2008). A mixed classifier based on combination of HMM and KNN. In *2008 Fourth International Conference on Natural Computation, IEEE*, 4(1), 38-42.

Wanner, S., and Goldluecke, B. (2012). Globally consistent depth labeling of 4D light fields. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 41-48.

Wanner, S., Straehle, C., and Goldluecke, B. (2013). Globally consistent multi-label assignment on the ray space of 4d light fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1011-1018.

Xu, Y., Nagahara, H., Shimada, A., and Taniguchi, R. I. (2019). TransCut2: Transparent Object Segmentation from a Light-Field Image. *IEEE Transactions on Computational Imaging*, 5(3), 465-477.

Yin, L., Sun, J., Tang, C. K., and Shum, H. Y. (2004). Lazy snapping. *ACM Transactions on Graphics*, 23(3), 303–308.

Zhang, D. Q., and Chen, S. C. (2004). A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial intelligence in medicine*, 32(1), 37-50.

Zhang, J., Ma, K. K., Er, M. H., and Chong, V. (2004). Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. In *International Workshop on Advanced Image Technology (IWAIT'04)*, 207-211.

Zhang, M., Zhou, Y., & Wang, S. (2020). A modified nonlinear conjugate gradient method with the Armijo line search and its application. *Mathematical Problems in Engineering*, 1-14.

Zhang, S., Cheng, D., Deng, Z., Zong, M., and Deng, X. (2018). A novel kNN algorithm with data-driven k parameter computation. *Pattern Recognition Letters*, 109, 44-54.

Zhang, S., Li, X., Zong, M., Zhu, X., and Cheng, D. (2017). Learning k for knn classification. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(3), 1-19.

Zhang, S., Zong, M., Sun, K., Liu, Y., & Cheng, D. (2014). Efficient k-NN algorithm based on graph sparse reconstruction. In *International Conference on Advanced Data Mining and Applications* Springer, Cham, 356-369.

Zhang, Y. (2007). A novel medical image segmentation method using dynamic programming. *In Proc: International conference on medical information visualization –bio Medical visualization*. USA, pp 69–74.

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

Zhou, J., Chan, K. L., Chong, V. F. H., and Krishnan, S. M. (2006). Extraction of brain tumor from MR images using one-class support vector machine. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, 6411-6414.

Zhuge, Y., Krauze, A. V., Ning, H., Cheng, J. Y., Arora, B. C., Camphausen, K., and Miller, R. W. (2017). Brain tumor segmentation using holistically nested neural networks in MRI images. *Medical physics*, 44(10), 5234-5243.