Multimedia Tools and Applications (2022) 81:18595–18616
https://doi.org/10.1007/s11042-022-12271-x

# New techniques for efficiently k-NN algorithm for brain tumor detection

**Soobia Saeed[1] · Afnizanfaizal Abdullah[1] · N. Z. Jhanjhi[2] · Mehmood Naqvi[3] · Anand Nayyar[4,5]**

Received: 12 April 2021 / Revised: 2 January 2022 / Accepted: 14 January 2022 /
Published online: 9 March 2022


**Abstract**
The *k*-NN algorithm missing values is one of the current research issues, especially in 4D frequency. This study addresses the accuracy of the images, increases the efficiency of missing *k*-NN hybrid values, and constructs a research framework that can identify cancer-damaged areas isolated from non-tumors areas using 4D image light field tools. Additionally, we propose a new approach to detect brain tumors or cerebrospinal fluid (CSF) development in the early stages of formation. We apply a combination of the hybrid *K*-Nearest Neighbor (*k*-NN) algorithm, Fast Fourier Transform, and the Laplace Transform techniques on four-dimensional (4D) MRI (Magnetic Resonance Imaging) images. These approaches use a 4D modulation method that dictates the light field used for the Light Editing Field (LEF) tool. Depending on the user's input, an objective evaluation of each ray is calculated using the k-NN method to maintain the 4D frequency redundant light fields. We suggest that light field methods can improve the quality of images through LEF since the light field composite pipelines reduce the borders of artifacts.

**Keywords** Brain Tumor · MRI · Laplacian · K-Nearest Neighbor

✉ Soobia Saeed
saeed.soobia@graduate.utm.my

Afnizanfaizal Abdullah
afnizanfaizal@utm.my

N. Z. Jhanjhi
noorzaman.jhanjhi@taylors.edu.my

Mehmood Naqvi
syed.naqvi4@mohawkcollege.ca

Anand Nayyar
anandnayyar@duytan.edu.vn

Extended author information available on the last page of the article

## 1 Introduction

Medical images are an essential part of analyzing tumors as well as cerebrospinal fluid (CSF) leaks. In medical literature, several techniques have been proposed to analyze medical images [4]. A useful method, known as $k$-Nearest Neighbors ($k$-NN), is a part of the supervised learning algorithms family that can create a new category of data points from a dataset. The hybrid $k$-NN algorithm is a robust framework often used as a point of reference, such as for the artificial neural network (ANN). The hybrid $k$-NN algorithm combines the $k$-NN algorithm and a support vector machine (SVM) that detects the brain cancer and CSF distance values and then trains the given dataset of MRI images. The hybrid $k$-NN algorithm can be implemented with 4D Light Field Tool (LFT) methods. The hybrid $k$-NN algorithm reads from a complete dataset to find the nearest neighbors to classify the new data points.

Clustering is a handy technique used for identifying similarities among different clusters or groups of datasets. It is used as a research model, and there are many clustering algorithms in which the $k$-means algorithm is used to identify hidden patterns in the data. This algorithm is an unsupervised clustering algorithm that creates a specific number of disjoint level (non-dynamic) sets. The hybrid $k$-NN algorithm provides a clear and straightforward way to handle a particular dataset's request across a certain number of groups. The number of groups (expected $k$ clusters) is developed, and the $k$-mean algorithm randomly picks the $k$-mean calculated values from the beginning of the data. The first step is to take each point from a particular dataset and calculate complexity at the nearest location, taking into account the search's proximity. We use the Euclidean division and recalculate the new $k$-group location. The hybrid $k$-NN algorithm cluster technique is repeated until it finds the current $k$ collection location. This estimation minimizes the distance limit, known as the squared turn-up limit, filling in the empty spaces.

We investigated 4D images for the various cancer stages and how to recover the hidden information of the damaged cancer cells and CSF. We select 4D modeling to identify the tumor accurately and efficiently. The various cancer areas scattered in the vicinity with gaps are identified with high accuracy using the latest 4D image technology through LFT. The LFT increases the accuracy of the hybrid $k$-NN algorithm optimization to provide a better solution. The Laplace transformation and Fourier transformation reduce the missing values related to the Laplacian matrix and a diagonal matrix. Also, 4D LFT increases image resolution and improves the algorithm's accuracy after applying 4D image segmentation techniques. This paper progresses in the following manner: Section 1 provides an overview of brain cancer and CSF details. Section 2 presents the problem statement of the hybrid $k$-NN algorithm with brain cancer and CSF issues. In Section 3, we explain some of the previous related work done so far in this domain. Section 4 describes the methodology used to detect brain cancer through the MRI-4D interface of image segmentation. Also, sampling techniques of datasets and tools generated in the experimental results after performing statistical analysis are presented. Section 5 provides details of the results and discussions while mentioning the findings and limitations. Finally, Section 6 concludes and describes the contributions made by this study and suggests future directions.

## 2 Problem of statement

The problem associated with brain tumor and CSF fluid is difficult top identify in MRI images clearly. This research aims to examine whether the CSF leakage fluid and brain tumor images are helpful to improve the quality of images using the Laplace transformation and FFT algorithm in terms of both accuracy and quality. The datasets used in this research is a series of chronological poor MRI images of brain tumor (cancer) and CSF fluid leakage in patient's brain images. In addition, the previous research contribution is based on the 3D segmentation method that tried to enhance the quality of images to gain a higher resolution and remove the noise or existing errors. Therefore, the traditional method (3D) is not enough to enhance the accuracy of poor images of fluid as CT myelography or MR myelography is more significant than MRI in identifying the area of CSF leak as if it arises from a considerable tear [14]. The problem associated with the *k*-NN algorithm is that it cannot find the missing values of the *k*-nearest neighbors. We focus on this issue with the *k*-NN algorithm, which is highly sensitive to random samples because it only selects neighbors based on distance criteria. We also use 4D Images of MRI to investigate cancer stages and recover the hidden information of damaged cells of brain cancer and CSF with the hybrid *k*-NN algorithm and Laplacian transformation with a diagonal matrix Fig. 1.

## 3 Related works

In the literature, the retrieval of hidden information from brain cancer image datasets has become increasingly crucial. The *k*-NN approach has been widely used to fix the missing data problem [14]. However, past studies did not pay much attention to the significance of image features, which substantially affects the selection of the image's neighborhood data points.D. Bertsimas et al. [5] proposed a new way of recording the missing data values, called the featured weighted gray *k*-NN algorithm (FWG*k*-NN). Their approach measured the appropriateness of characteristics using the functionality of the FWG*k*-NN and missing values of the image data. Z. Liu et al. [19] provided an experimental evaluation of five University of

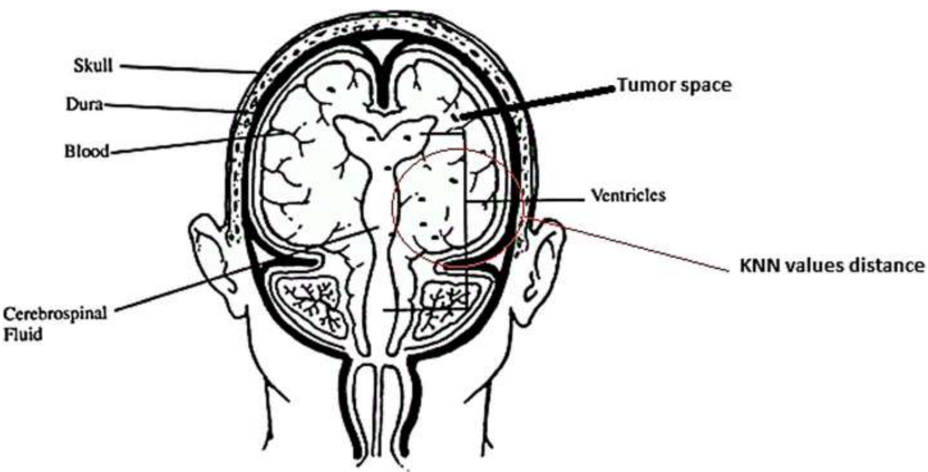


**Fig. 1** A brain cancer image showing the largest distance values using the *k*-NN algorithm

California Irvine (UCI) Machine Learning Repository datasets in three processes that fail in finding the distance values at different failure rates. Their empirical findings indicate that the types of datasets have an undeniable effect on the FWG*k*-NN assessment of missing information and the technique is better than the other four assessment methods.

Missing data is a prevalent issue in real-world datasets of image features, which is why statistical literature has attracted significant interest. C. A. Bhardwaj et al. [6] suggested modification in existing techniques with constant blended and continuous variables based on the missing information's sequential increment. Latha, R. S., et al. [15] used many predictive models in their proposed framework, such as *k*-nearest neighbors, support vector machines, and decision tree selection tuned for various modifications of their techniques. Garg, Ginni, and Ritu Garg [9] described the detail of fuzzy logic techniques used in their algorithm to optimize the given value assigned in the *k*-NN algorithm to reduce the missing values in image datasets of fuzzy *k*-NN classifiers. R. Armina et al. [3] proposed pattern classification methods to enhance data sets to predict models to countermeasure missing value problems. In a large-scale computer study, S. Zhang et al. [32] proposed increasing the sample accuracy of 84 images of the UCI Machine Learning Repository datasets information. I. Altaf et al. [2] proposed a method for calculating missing values of *k*-NN data sets information and proposed some random missing information classification to optimize the assigned values for the highest distribution of five other techniques in the datasets: average allocation, closest neighbors *k* (*k*-mean), *k*-NN frequency, Bayesian PCA, and average predictive model. Furthermore, Nasor, Mohamed, and WalidObaid [21] optimized the assigned values of the sample's output to train learning algorithms used for calculating datasets' information. They also described a more accurate and efficient method of the existing *k*-NN methods demonstrated in ten subsequent applications by computational studies [27]. Wang, Lishan [31] suggested a technique to classify the reliability of an incomplete pattern based on multi-label classification with adaptive adherence to the missing values. S. Chowdhary et al. [8] Classified brain cancer based on attribute values used for image datasets. The datasets of binary transformation images are classified to implement the algorithm for identifying the missing information. However, if the images cannot be unambiguously categorized, missing values play a key role in obtaining an accurate assessment. In this case, the missing values are calculated based on the nearest *k*-NN and self-organizing techniques. L. S. Prahl et al. [22] suggested a model that is categorized according to each training class and the classification outcomes are integrated into the fundamental tasks with the suitable combination regulations for the principle classification. The *k*-NN algorithm belongs to certain classes and Meta classes (not equally distributed for various individual classes) with distinct types of images of brain cancer datasets. A. Kinaci et al. [12] suggested that the ratio of reliability captures the uncertainty, and accuracy and efficiency decreases the frequency of incorrect classifications. Several tests using real information datasets illustrate the efficiency of the proposed technique.

Huang et al. [10] proposed a deferred learning algorithm (nearest neighbor algorithm). Their algorithm attempts to round up projections in the training datasets using a support vector machine. Moreover, the *k*-Neighbor algorithm's forecasts are based on the average values of spatial neighbors. They used distance measures, such as Euclidean distance, Minkowski distance, and Mahalanobis distance, to choose neighbors in the datasets' hierarchical structure. N. R. Saunders et al. [26] proposed measuring characteristics, such as the Z-normalization technique and Min-Max used as standard methods to normalize the scale of variables that assign equal weights to all features evenly. M. Ueno et al. [29] proposed a new technique for assigning weights to individual features of datasets' images using output error when various decision tree models are constructed. A. M. Mendrik et al. [20] suggested a nearest neighbor computing algorithm (NNCA), which is an efficient method of completing missing information. Each missing value is replaced by the appropriate values selected from the datasets.

In addition to replacing missing data with reasonable values as soon as possible from the actual value, the integration algorithm should retain the original data structure and avoid distorting the distribution of variables. Although NNCA algorithms are practical, the impact of these techniques on the data structure is not well understood. Simulations were performed on synthetic datasets with distinct schemes and failure levels to assess the efficiency of NNCA within an individual iteration of 1NNCA and *k*-NN or weighted *k*-NN (w*k*-NN) into the frames. Regardless of the image, the *k*-NN generally exceeds 1NNCA and reduces errors of the *k*-NN missing values which are calculated in terms of integration precision. However, 1NNCA is the only way the data structure can be maintained when considering small values of 1NNCA. The use of the three algorithms, as mentioned earlier, provides the information from implicit error and image datasets. After receiving the missing information randomly, the same conclusions can be drawn when computed tomography studies are carried out.

B. Srinivas and G. Sasibhushana Rao [28] proposed a hybrid *k*-NN model (CNN-*k*-NN) for the classification of brain tumors by magnetic resonance imaging (MRI) that combines convolutional neural networks (CNN) with the *k*-NN algorithm. The CNN model is used for implementing its properties of feature extraction and then adding them to the *k*-NN algorithm to predict classes of an image dataset. The authors increased the performance accuracy of the proposed model up to 96.25%, which is the value that occurs as suggested in the testing of images and has proved to be higher in terms of precision, error rate, sensitivity, and specificity findings of the proposed model.

S. Sargolzaei et al. [25] presented a new strategy to the closest neighbor scheduling in their research using image datasets with empty values. The use of empty values and their implementation make the calculation more straightforward than the initial fuzzy *k*-NN technique. The capability of *k*-NN algorithms to adapt and various supervised learning issues are thus improved. Compared to the implementation of non-parametric statistical processes, empirical research was carried out. It determines the proposed algorithm of *k*-NN neighboring arrangements of the nearest position of the proposed *k*-NN is more accurate than *k*-NN fuzzy logic algorithm, and then implementation of the *k*-NN algorithm is better than the fuzzy logic algorithm. They concluded that *k*-NN is much more precise than other arranging processes of *k*-NN values. J. M. Cameron et al. [7] suggested a framework to identify cancer detection in images using the datasets with histological diagnosis of types of brain cancer. This technique incorporates the detection of cancer signs through vibrational spectroscopy of Fourier transformation. They suggested a few algorithms to manage images to identify cancer. However, this research focuses on increasing the frequency signal of images to increase the image accuracy using Fourier transform infrared (FTIR) spectroscopy. Komboet al. [13] proposed an effective technique which represents the allocation form of the k-NN algorithm, concentrating on the techniques used to assess the missing value and implementing the general or local datasets of information.

K. Usman and K. Rajpoot [30] highlighted the prospective improvement in the present technique. They anticipated that their research would provide the reader with a better knowledge of the modality direction. They proposed a comparative study with the *k*-NN algorithm on individual establishment methods such as “mean” and “standard deviation.” A training group with relevant research gathered information on different dimensions. In each group, the above techniques are applied, and the results are compared. It was concluded that the standard deviation was better than the mean value. F. Liang et al. [17] compared two attribution techniques based on *k*-NN and Fourier transform. They suggested calculation technique (FL-*k*-NN) incorporating time lag connections within and across the variables. H. Khotanlou et al. [11] suggested that the connection matrix must be distinguished to rebuild test information points by learning to assign different values to k and different test information points, called k-NN correlation matrix classification (CM-*k*-NN). In particular, the reduced square-loss function is used to minimize reconstruction error by rebuilding each test data point in all learning data points. After that, in the reconstruction phase, the Laplacian graphic organizer is recommended to keep the local data structure. Additionally, the authors' focus on matrix and mathematical object of datasets are applied to separate k values for distinct sample information, resulting in low dispersion to eliminate the reconstruction process's noisy property. *k*-NN techniques (including the suggested CM-*k*-NN technique) are used for regression and correction of missing information in relation to classification assignments. In L. Garg et al. [11] the authors performed sets of tests to demonstrate effectiveness, and experimental findings showed that the suggested technique was more precise and effective than the *k*-NN techniques found in process for data extraction, such as classification, regression and missing information classification.

Alhroobet al. [1] proposed a method of calculating missing values in most clinical and biomedical datasets. Registration of patients can be divided into several institutions; devices can fail, and sensors cannot always be used. While these often-lost values are ignored, biases and mistakes can occur when extracting information. J. Liu et al. [18] suggested a method using convolutional neural networks to identify the blood glucose, and on the other hand, it focuses on missing values of the *k*-NN algorithm. This method predicts the various kinds of errors that have not yet been incorporated with the present techniques of *k*-NN algorithms. S. Saeed and A. Abdullah [23] proposed an algorithm used for the detection of brain cancer which has been used in image segmentation and MRI images datasets in data mining and machine learning. This research focuses on computerized image processing to detect brain cancer using image segmentation techniques and histograms. Lavanya, S. R., and Dr. R. Mallika [16] suggested a comparison of techniques in three biological data sets. These techniques are implemented for identifying the missing values in time series method. This research aims to build a research framework that can define or isolate the area of cancer damage from tumors and non-tumors using Fourier transform. Another research tool, based on the Fourier transform, is the main mathematical method for frequency analysis and has broad scientific and engineering applications. Fast Fourier transformation (FFT) is a contemporary transformation method and computing FFT are extensively studied and active research continues. S. Saeed et al. [24] proposed a method based on FFT for two fast execution algorithms to evaluate the accuracy of medical datasets of images. This study helps in detecting brain cancer due to the interconnection process of 4D image segmentation and Fourier transform. The authors implement the 4D images on AI software modeling techniques to measure the volume of brain damage cells within a CSF. LFT can be instrumental in improving the light field editing application for segmentation and LFT composite pipeline quality by reducing defects at boundaries.

## 4 Methods

The main objective of our study is to develop a system that can detect CSF leakage in the brain's tumor-region, or it can separate patients who have tumors. We collected data through various sources and analyzed through the necessary research tools. We use a hybrid *k*-NN algorithm to determine the objectivity of the classifications. The hybrid *k*-NN algorithm (clustering algorithm) is used as our research framework. We used supervised machine learning hybrid k-NN algorithms. Clustering is a very effective technique used to identify similarity among different clusters or groups (to measure the huge data at the sample time). There are many clustering algorithms in which the k-means algorithm is used to identify hidden patterns from the data. In our research k-means algorithm explores the unseen information by taking the attributes such as space of value, sign of object, stages of cancer and demographics, etc. The Laplacian model identifies the cancer at a very early stage. Moreover, the scientific quality of the relevant literature was checked through critical appraisal tools for better analysis as shown in Fig. 2.

The above figure shows the 4D images with the measurement of histogram values of CSF leakage in brain cancer. This research uses the spatial (phonological images) to identify the CSF leakage with brain cancer images. However, this research contributes the 4D segmentation with missing values of imputation which is mentioned by given below Fig. 3:

### 4.1 Trained datasets

This segment addresses in-depth analysis of data sets. Rich and high-quality data are used for a successful multi-image segmentation method to increase the resolution of

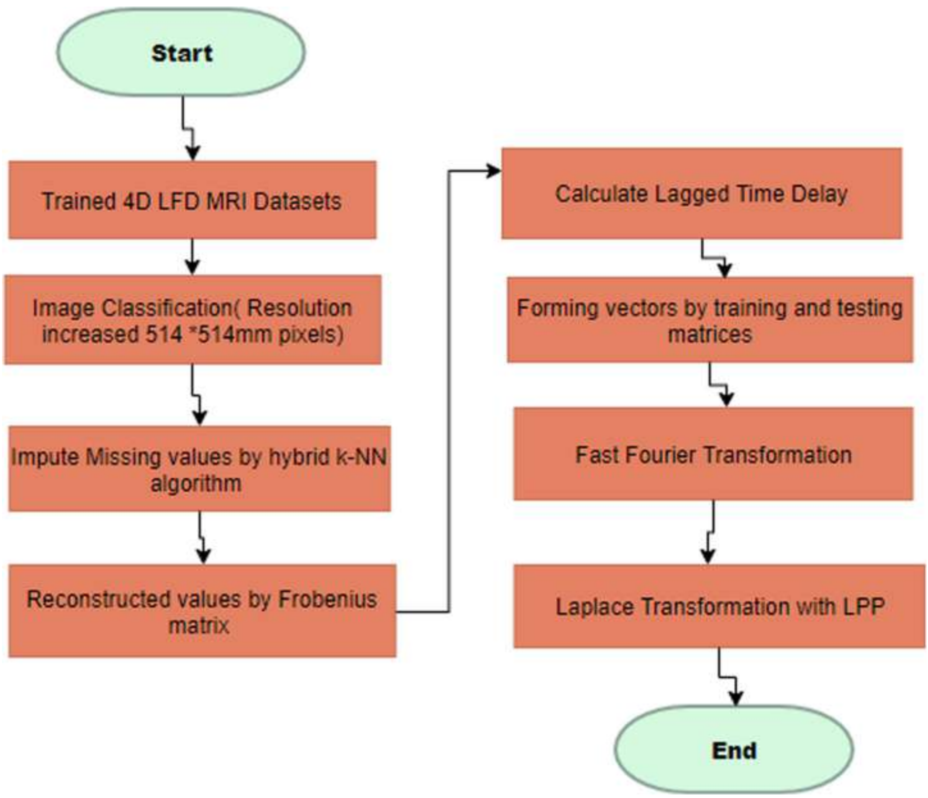


**Fig. 2** Structure of proposed framework that shows the proposed framework used in different types of techniques

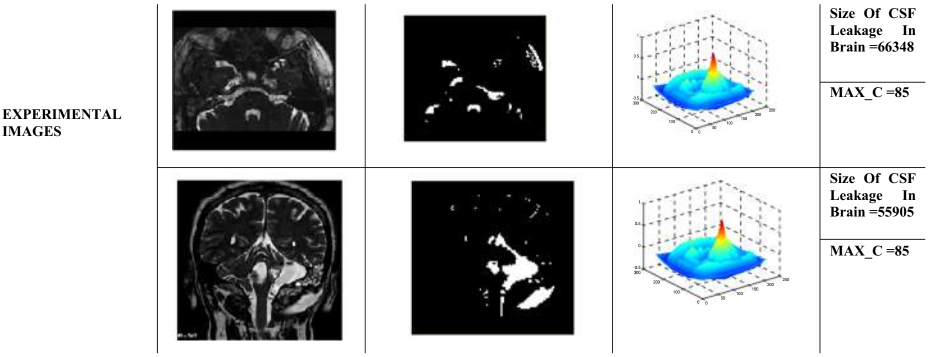


**Fig. 3** 4D image segmentation with histogram

the medical image dataset. The main motivation for creating the Light Field Database in medical MRI images is the fact that there is no database publicly available in MRI and CT scan images for the diagnosis of brain tumor (low-grade tumor) and cerebrospinal fluid (CSF) in the initial stage of cancer that is in liquid form which is not clearly visible in the initial growth of cancer. The research in this section focuses on creating LFD to conduct experiments in this emerging field. The external standardized light fields are extracted using the Lytro command-line tools to help recreate the 14 × 14 shifted views, which can then be used for various applications. The Lytro light field software tool is generally considered to improve the quality of MRI datasets. Trained LFD consists of more than 3000 MRI images, and the datasets are divided into different five categories:

1. Healthy Brain
2. Cerebral Spinal Fluid Leak Images.
3. Low-Grade Tumor Images.
4. CSF with Brain Tumor images.
5. High-Grade Tumor Images.

The sample MRI images was collected by National Cancer Care Institute (NMI)"https://www.ncci.org.pk/ and "Medicare hospital" in Pakistan, from the source of data gathered, is https://www.medicarehospital.pk/. This stage serves as an empirical analysis to evaluate the effectiveness of the proposed technique which is related to the proposed method.

### 4.2 Missing values imputation

Missing data imputation is one of the biggest problems in segmentation methods, especially in MRI datasets. Imputation application problems are common in various classifications where a few data are missing in the datasets. However, many techniques and approaches exist to solve this missing imputation data problem in the hybrid k-NN algorithm. CM-FFT technique at the first level is used for reducing the missing values and minimizing the time lag (delay). This research is related to finding the missing information in MRI images, which is somehow difficult to locate the position of tumor or fluid in the brain. This research utilizes the extended hybrid k-NN algorithm from the previous section with the novel proposed CM-FFT technique, providing a better solution for retrieving the missing data in the MRI datasets. The research in this section is applied to a correlation matrix with Fast Fourier Transformation. The reason for selecting the correlation matrix is to combine the hybrid k-NN algorithm with time-lagged (delay) to extract the missing values to reconstruct the data in the datasets and reduce the execution time in retrieving the information process. In addition, FFT is the better approach to find the missing information in the same sequence of variables. Therefore, the proposed CM-FFT technique is a better solution for solving the missing imputation problem in the MRI datasets.

This research is based on reducing the missing values which are used in the imputation techniques are given below:

### 4.3 Experimental factors of reconstruct the structure

Suppose that U ϵ Wn × d stands for training data point where n = number of training data points and suppose V ϵ W n × m stands for matrix test data, where m = number of test data. We use U training data points to recreate all test data. Va is the distance between U T Wa and Va, where Va × Wn represents the performing weights for training data points. Similarly, we use training X data points in the number of test data to recreate all Va test data among U T Va and W [30].

$$\min\sum_{a=1}^{m}||\mathrm{U\ T\ Va}-\mathrm{va}||22 = \min \parallel \mathrm{U\ TVa}-\mathrm{va}\parallel 2 \tag{1}$$

Where F = Frobenius matrix and W × Wn × m mrepresents the performance weight matrix and the relationship between the training data points and the test data. The function of improving the Eq. 3 is [30]:

$$\mathrm{Vi} = (\mathrm{UUT})-1\mathrm{X\ V} \tag{2}$$

However, UUT is not always reversible in real applications. For this purpose, the traditional interception function is added to the term's simple arrangements in Eq. 4.

$$\mathrm{minV} \parallel \mathrm{U\ TV}-\mathrm{V} \parallel 2\ \mathrm{F} + \rho\parallel \mathrm{V}\parallel 22 \tag{3}$$

The Eq. 5 is called edge regression, where p is the turning parameter [30].

$$\mathrm{Va} = (\mathrm{UUT} + \rho\mathrm{I})-1\mathrm{X\ V} \tag{4}$$

We need to use non-zero values so we use correlation between training and testing data points and the data test will become partial training data points. We can use norm parameter term to propose and replace the term from 1-norm to 2-norm in Eq. 6 [30].

$$\mathrm{F1\ (S)} = \parallel \mathrm{S} \parallel 1 \tag{5}$$

In eq. 6, we need to remove the noisy data points that are almost irrelevant to all test data during the reconstruction process. It is defined in Eq. 7 as follows [30]:

$$\mathrm{F2\ (S)} = \parallel \mathrm{S} \parallel 21 \tag{6}$$

### 4.4 Calculating the lagged time

We need to identify the correlated variables and the time lags for the vector of testing and training. The cross-correlation is used to measure two time series as a function of a time delay applied to one of them. The cross-correlation, rUy, between variables, u and y, for time delay (e) is [17]:

$$\mathrm{ruv}(\mathrm{e}) = \frac{\mathrm{Fuy}(\mathrm{e})}{\sqrt{\mathrm{Fuu}(0)\mathrm{Fvv}(0)}} \tag{7}$$

$$\mathrm{F\ uy}(\mathrm{d}) \begin{cases} \dfrac{1}{\mathrm{r-e}} \sum_{\mathrm{a}=1}^{\mathrm{T}} (\mathrm{ut-u})(\mathrm{vt} + \mathrm{e-y}), \text{if } \mathrm{e} \geq \mathrm{o} \\ \dfrac{1}{\mathrm{r} + \mathrm{e} \sum\limits_{\mathrm{a}=1-\mathrm{d}}^{\mathrm{T}} (\mathrm{ut-u})(\mathrm{yt} + \mathrm{e-v})} \quad \text{otherwise } 0 \end{cases}, \tag{8}$$

Where T represents the length of the chain, U and V represent the average of u and v respectively, d is compared from -(E - 1) to +(E - 1), and E is the maximum time delay of missing values where both are present in this calculation in eq. 8.

Matrices are generated by decreasing intensity for each of the p lags, with the ordered correlations from 1 ... p. Thus, the delay, d, is the highest correlation (max) and Lp is the lag for each pair of variables L1 with the lowest correlation. Each L is an N × N matrix, where elements show the time lags for each correlation between the N variables. A luv portion may be positive (variable v values have a delayed response to u values in the time unit luv) or negative (variable u values have a delayed response to v values in the time unit) and luv = −lvu. The matrix's diagonal elements are not calculated as these elements give the signal's auto-correlation and are not used in this algorithm. The corresponding values for all |ruv|, are stored in the matrices R1 … Rp, which are used in the neighbor collection steps.

### 4.5 Forming vectors

L*k*-N N is more complex than the *k*-NN hybrid system, with different delays depending on the set of factors to be considered. For each step, we have created a set of testing and training vectors. We construct the function u over time t with some missing values. The function u has a delay relationship with variables v and z where luv lags slightly. The test vector is constructed with the t + luv and t + slightly lagged values of v and z. The training vectors are produced similarly, and u values are stored separately, which are needed for inclusion. Training tests are provided by the current values of u, where the resulting period must be within 1 to T (data length) after the accumulation of errors. This limits instances of vector time training to a missing value:

$$[\max(1, 1-\min(\mathrm{lu1}, \ldots..\mathrm{luN})), \min(\mathrm{lu1} \ldots\ldots\ldots \mathrm{luN}))]$$

Where lu1… luN are the time lags of relationships between u and all N variables for the recent lag matrix.

The distance between instances u and y in eq. 9 are given below [11]:

$$\mathrm{e}\,(\mathrm{U},\mathrm{Y})=\frac{\sum\limits_{a=1}^{n}(\mathrm{ui}\wedge\mathrm{va})\times(\mathrm{ua}-\mathrm{va})2\times\mathrm{Va}}{\sum\limits_{i=1}^{n}(\mathrm{ua}\wedge\mathrm{va})} \tag{9}$$

In Eq. 9, we develop a calculation method based on Fourier conversion to use the previous values of the missing values for each variable included with the correlations between the variables of the L*k*-NN calculations. We form a data segment from the raw data to indicate the last missing data points. When values from s1 to sp. - 1 are present (or calculated), and sp. ... sq. does not exist, Fourier descriptors are achieved by using Eq. 10 [11].

$$\mathrm{FK}=\sum_{c=1}^{P-1}\mathrm{Sb}\times exp-\left(\frac{2\pi i}{\mathrm{p}-1}\right)(\mathrm{C}-1)(\mathrm{h}-1) \tag{10}$$

### 4.5.1 Experimental results

This section identifies and discusses the simulation’s computational results, which were developed using statistical significance analysis, in order to evaluate the performance of the proposed Correlation Matrix of Fast Fourier Transformation (CM-FFT) technique in the hybrid k-NN algorithm for finding missing imputation data extracted from MRI images in time series data. Figures 4 and 5 show the extension of the Correlation Matrix of time-lagged with the hybrid k-NN algorithm, cross-correlation variable of coefficient, and Fast Fourier Transformation for extracting the data of missingness values in images for a combination of hybrid Lk-NN algorithm.

The above figure shows Fast Fourier conversion to use the previous missing values for each included parameter comparison between the L*k*-NN calculation variables such as (b-1) (g-1) (h-1). It forms a data segment from the raw data to indicate the last missing data point.

Where Fk is the k-Fourier descriptor with the calculated time values of m, where p, g, and q can be calculated from Fourier descriptors using Eq. 11 as given below:

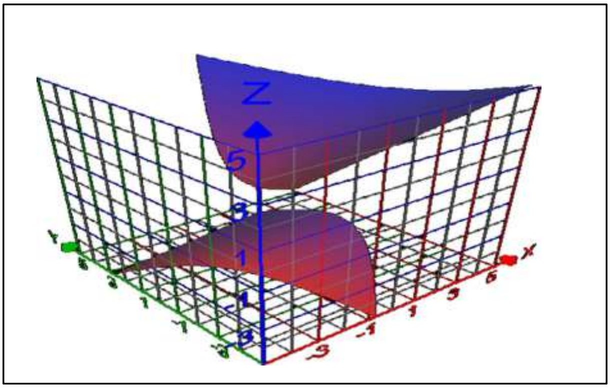


**Fig. 4** Fast Fourier transformation of identifying the missing values calculation

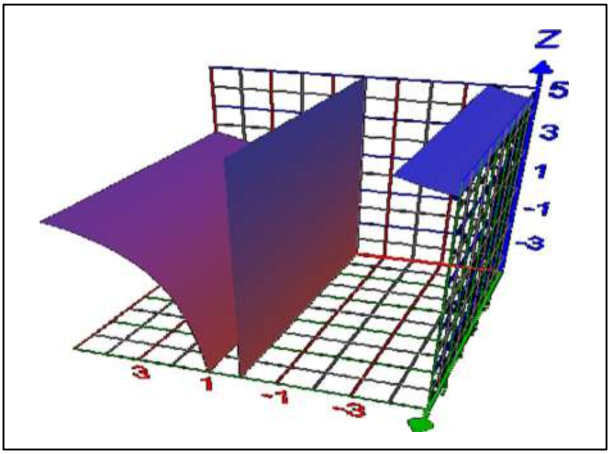


**Fig. 5** Implementation of Laplacian to calculate the missing values

$$\mathrm{Sm} = \frac{1}{\mathrm{p}-1} \sum_{\mathrm{c}=1}^{\mathrm{P}-1} \mathrm{Sb} \times exp-\left(\frac{2\pi \mathrm{i}}{\mathrm{p}-1}\right)(\mathrm{C}-1)(\mathrm{g}-1)(\mathrm{h}-1) \tag{11}$$

### 4.6 Laplace transformation

We apply Laplacian transformation in Eq. 12, which is given below:

$$\mathrm{Sm} = \frac{1}{\mathrm{p}-1} \sum\nolimits_{\mathrm{c}=1}^{\mathrm{P}-1} \mathrm{Sb} \times exp-\left(\frac{2\pi i}{\mathrm{p}-1}\right)(\mathrm{C}-1)(\mathrm{g}-1)(\mathrm{h}-1) \tag{12}$$

The above figure shows the Laplacian and the diagonal matrices, where $L \in W_{d \times d}$ is the Laplacian matrix and D W d × d is the slant matrix. LPP must ensure that the closest neighbors of the original data are kept in the new space after dimension reduction in order to minimize nonlinear dimensions.

$$\mathrm{Sm} = \min\left[\rho\mathrm{n} \sum\nolimits_{\mathrm{c}=1}^{\mathrm{P}-1} \mathrm{Sb} \times exp-\left(\frac{2\pi i}{\mathrm{p}-1}\right)(\mathrm{c}-1)(\mathrm{g}-1)(\mathrm{g}-\mathrm{n})(\mathrm{h}-1)\right] \tag{13}$$

The above figure shows that the LPP is a nonlinear dimensionality reduction method and the *k*-NN ensures that the original LPP data after preserving the new space performs corresponding dimensionality reduction. To improve the equation using the optimization method, we calculate the optimal Sm equation through rebuilding weight or correlation between training data points and the second test data point. A positive relationship between the data point of the ith learning data set and the test shows the positive weight (i.e., wa, c > 0), while a negative relationship means that (i.e., wa, b < 0) there is no correlation between data learning point (ath) and data test (cth), which means zero weight (i.e., wa, c = 0). In this case, to forecast the test data, data point "a" is not used. Instead of using all learning data points to predict test data, we use training data points only with a non-zero coefficient. Suppose that Sg∈ Wn × m is optimal because we better understand the characteristics of the proposed method Figs. 6, 7 and 8.

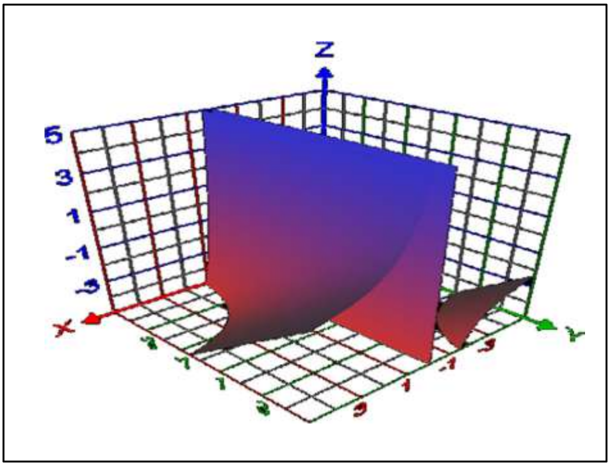


**Fig. 6** Nonlinear diagonal and the Laplacian matrices

The above figure shows the iteration method and derivative turning point parameters to show the nonlinear dimension reduction method and the original LPP data after the new space has been preserved, where P is the turning point of derivative parameters. So, the predicted value of the given term in equation is:

$$\text{Predicted value} = \sum_{b=1}^{n}\left(\frac{\text{va, c}}{\sum_{c=1}^{n}\text{va, c}} \times \text{Sa, c} \right. \tag{14}$$

This method of iteration and the number of training data points represents the true value of the $a^{th}$ training data, where n is the number of data points for training and Va is a training point that represents the actual value of data for training.

The above figure shows the method of iteration and the number of training data points representing the true value of $a^{th}$ training data, where n is the number of data

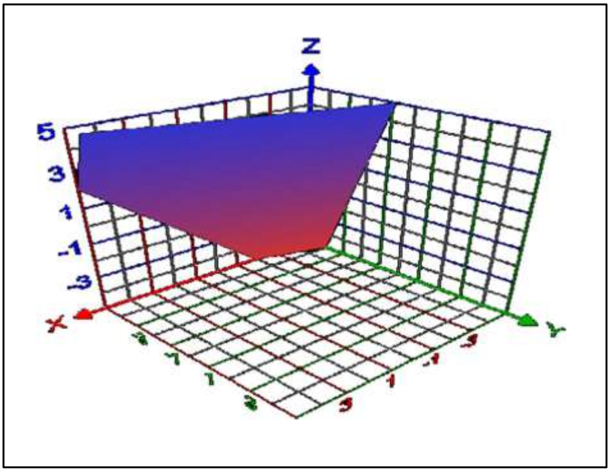


**Fig. 7** Derivative parameters

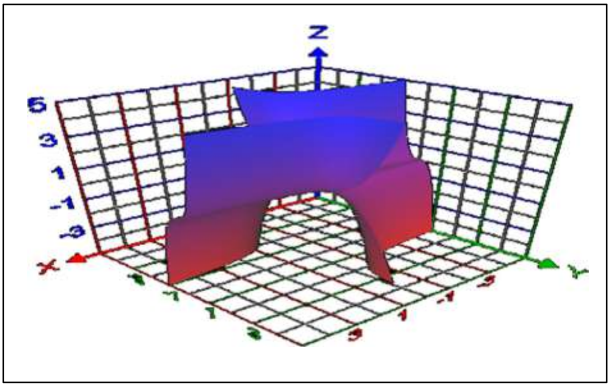


**Fig. 8** Iteration values of trained data points

points for training and Va is a training point that represents the actual value of data for training.

Figure 9 depicts the combination of missing values ratio, which is replaced by non-zero values after imputing the missing values, LHk-NN algorithm imputes missing data points, and FFT in terms of calculated values of Imputation using FFT for one variable with simulated missing data points, actual data (in blue), and imputed data (in red). To test this, MRI datasets were used to generate simulated missing values. The missing values were extracted into rows and columns in the same order for each subject, and all imputed methods were used in the cross-correlation of time-lagged values using this proposed novel technique. In Lk-NN, time instances with missing values are completely absent. Because of the use of time lags, the test vector may not be empty because lagged values may be present. In this case, Lk-NN was able to impute all missing values, so this did not occur. The experimental results show that the FFT computed values are where the missing values are found and replace the missing values with single values in the same sequence form, whereas FFT works in the frequency signal sequence form to detect missing or hidden data in the sequence form to detect the range of CSF fluid or a brain tumor. These simulation results show that the FFT generates discrete variables of a single series and converts all data (non-zero values) from the same sequence into a single series. Furthermore, FFT is used to determine the length of a signal. It converts one frequency-domain function to another, or simply FFT of the original time-domain function. Finally, the nearest point of the decomposed sequential values found in the same direction of the sequence of a variable achieves the better solution of the imputed data.

| Name | Imputed Empty Instance | Cross Correlation of Time-Lagged | Missing values of the variable |
|---|---|---|---|
| Hybrid k-NN | No | No | No |
| Fast Fourier | Yes | No | No |
| LHk-NN | Yes | Yes | Yes |
| LH(FFk-NN) | Yes | Yes | Yes |

| MRI Datasets Images | Missing Rations | LHk-NN | FFT | Execution Time (Second) |
|---|---|---|---|---|
| D-Images-I | 0.9% | 0.97 | 0.97 | 2.20 |
| D-Images-II | 10% | 0.92 | 0.94 | 2.11 |
| D-Images-III | 11% | 0.99 | 0.95 | 1.42 |

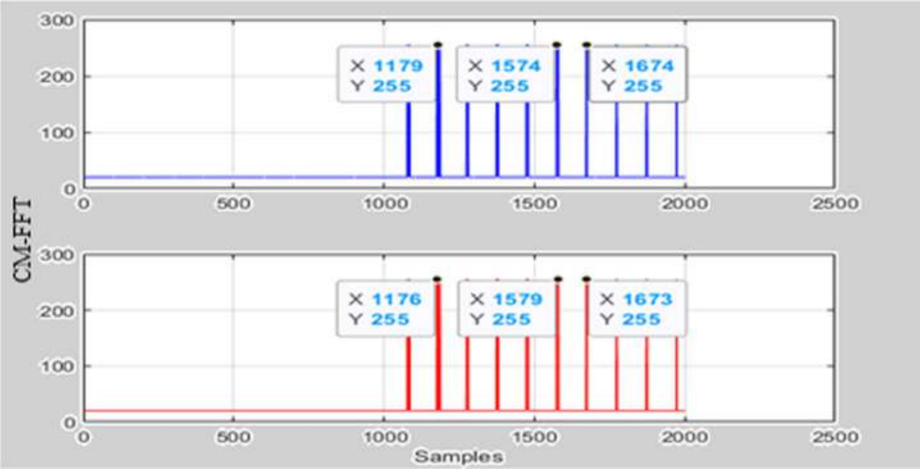


**Fig. 9** Missing values identification through FFT in the sequence form of rows and column to replace the zero values into non-zero values

### 4.6.1 Root mean square error (RMSE)

The root mean square error (RMSE) was used to calculate the results and actual ratings to assess the implementation's performance. The RMSE formula is shown below. A high RMSE value indicates a less accurate prediction of the rating, whereas a low value indicates a more accurate prediction of the real value. The difference between the actual and predicted ratings is denoted by the Sa − Sa∽, where Sa∽ is the predicted rating and Sa is the actual rating. The total number of ratings is n as shown by the Eq. (15):

$$\text{So,}\quad \text{RMSE} = \frac{1}{n}\sqrt{\sum_{b=1}^{n}(\text{Sa}-\text{Sa}\sim)2} \tag{15}$$

**Table 1** Comparison of Previous Result to Current Results

| Features | Decision Tree | Linear Discriminate | Quadratic Discriminate | Logistic regression | SVM<br>Linear | KNN Fine<br>K=1 | KNN Medium (Euclidean)<br>K=10 | KNN<br>K=100 |
|---|---|---|---|---|---|---|---|---|
| Accuracy (C) | 90.90% | 86.40% | 90.90% | 86.40% | 93.20% | 96.90% | 93.20% | 70.50% |
| Sensitivity (C) | 93% | 87% | 73% | 73% | 92% | 92% | 93% | 92% |
| Specificity (C) | 28% | 70% | 70% | 70% | 60% | 60% | 30% | 60% |
| Accuracy (P) | Nil | Nil | Nil | Nil | 0.881 | 71.43% | Nil | Nil |
| Sensitivity (P) | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil |
| Specificity (P) | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil |
| Accuracy (P) | Nil | Nil | Nil | Nil | 98.34% | 93.34% | Nil | Nil |
| Sensitivity (P) | Nil | Nil | Nil | Nil | 95.23% | 90% | Nil | Nil |
| Specificity (P) | Nil | Nil | Nil | Nil | 100% | 95% | Nil | Nil |
| Accuracy (P) | 100% | 61.73% | Nil | 100% | 100% | 91.36% | Nil | Nil |
| Sensitivity (P) | 100% | 84.91% | Nil | 100% | 100% | 96.23% | Nil | Nil |
| Specificity (P) | 100% | 17.86% | Nil | 100% | 100% | 82.14% | Nil | Nil |

Sa = $a^{th}$ test data point.
Sa ∽ = Predicted value.

### 4.7 PSEUDO code of missing values

The pseudo code of missing values of *k*-NN algorithm and combination of Fourier transform and LPP data transformation are given below:

**Algorithm 1: MISSING VALUES OF DATA IMPUTATION**

**Input:** (i,j)
**Output:** [K] ≠0, K=1-10
1. **If** (the number of non-zero value in x)// denoting missing values or empty space
2**. Initialize** []; i=1, j=1, k=1-10, where ms=missing values
3. For K= 1-100
4. If (pos[k] ≠0 or ms[k] =”*”)
5. Then [k] = “*”;
6. ***X, Y) = Image // Object function***
7. Where, ***X=(x0, x1 x2……,xn) and Y=(y0, y1,y2…..yn)***
**8. D= {(x0, c0),….(xn, cn)}**
**9.** ***x=(x1,…..,xn)*** *new instance to be classified*
10. For each labelled instance (xi,ci)
11. Calculate d(xi, x)
12. Order **d(xi, x)** from lowest to highest, ***(i=0,1,….,N)***
13. Select the K nearest instances to ***x: Dx<sup>K</sup>***
14. Assign to x the most frequent class in ***Dx <sup>K</sup>*** // keep the best solution for images in k-NN classification
15. *If (the number of non-zero value in missing values ≥1);*
16. Then missing values [K+1] ++;
17***. Else;***
18. Missing values [K=1-100] ++;
19***. While K≥100; do***
***20.*** Compute Fast Fourier Transformation *// New solution*
*21. If (the number of non-zero value in diagonal matrices of Fast Fourier transformation)*
22. Apply IDFT // Get new solution to identify missing values
23. Compute Laplace transformation with diagonal matrix // Get new approach use for Laplace technique
24. Initialize with Laplace and iteration to reconstruct the new data
25. Compute LPP // Get save the new reconstructed data
26. **End while**
27. **End if**
28. **End function**

The above the proposed algorithm show the combination of k-NN algorithm with

## 5 Comparison

The comparison of previous work and latest studies are given below:

The Table 1 and 2 above shows the longer values of k-NN (K = 1–10) and shows its performance for the new proposed model with higher dimensions. This investigation increases the accuracy of missing values for the k-NN hybrid algorithm. This paper also proposes a high-dimensional modulation method that monitors which field of illumination can be used in the light-emitting field. Using the light field segmentation of a 4D image in two different ways, the MRI imaging pre-prepared method treated the final destination settings of the image specified for the rest of the procedures Table 3.

This study addresses the classification results of the k-NN algorithm, which is based on semi-supervised machine learning, in this section. Using 4D photographs, this study focused on four segmentation techniques for brain cancer and CSF leakage. This model illustrates the measurement of a thin tumor shell and a small volume of CSF leakage within the brain. The model is created by software tools that use major k-NN techniques on a spread sheet to cover the patient information obtained from the UK National Cancer Research Institute (NCRI) database.

The outcomes of the experiments are performed by a dataset named "Malignant Brain Cancer with CSF Leakage," and the original information collection of 4-dimensional information is then randomly split into four segmentation approaches and datasets, resulting in a precision of 96.9%, as seen in the results. This thesis trained many supervised machine-learning models using the training dataset, including KNN Fine (K = 1), KNN Medium (Euclidean) (K = 10), and KNN Coarse (K = 100). We assess the utility of the suggested strategies by measuring four assessment metrics: Sensitivity (Sen), Specificity (Spe), and accuracy (Acc). Table II summarizes the study findings. KNN Medium (Euclidean)(K = 10), KNN Cosine (K = 10), KNN Cubic (K = 10), KNN Weighted (K = 10), and Ensemble Bagged Tree achieved an accuracy of 93.2% among twenty-four machine-learning methods models. The computing time of each classifier is also measured to determine its complexity, and Table 2 displays the processing time of each classifier. This study also compared our findings to those published in the literature in the form of related works. The comparison reveals that the k-NN (K = 1–10) models in the current analysis outperform previous works with 96.9% accuracy. These approaches outperform previously published works in the literature due to their shorter measurement time and use of the far-distant values. However, it does not reach 100% precision with all approaches, but in 4 dimensions, the k-NN outperforms previous work. This analysis is based on LLP since it eliminates the missed values of the hybrid k-NN algorithm while also minimizing the gap of far values. In this analysis, the researchers used laplacain transformation in terms of Fourier transformation and value optimization to quickly explain and detect missing values. This thesis also addresses Lk-NN, which is more complex than k-NN and must account for all of the delays that range from one variable pair to the next. Using the training vector, this research study created a set of test and training vectors for each of the delays.

## 6 Result and discussion

We made it clear that the time delay of the experiment and the training vector arising from the time delay are linked with improving the calculated information to determine the validation

**Table 2** Hybrid K-NN Algorithm with Missing Imputation Data

| Techniques | Reference | Objectives | Performance Metrics | Tools | Results Compared | Evaluation | Achievements |
|---|---|---|---|---|---|---|---|
| k-NN algorithm | (Lorenzo Beretta et al., 2016) | Retrieve Missing Data | Improve Missing Data Imputation | MATLAB Programming | Nil | k-NN Missing Data<br>K = 10 = inacc = 0.11750<br>SD = 0.52735 | Improve the Performance |
| K-NN algorithm with most correlated features (KNN–MCF) | (Sanjar. k et al., 2020) | To overcome the missing data | Improve k-NN classification Method | MATLAB Programming | Nil | k-NN Classification of prediction accuracy = 92.01% | Improve the Performance |
| (FLk-NN) Imputation method | (Rahman. SA, 2019) | Missing Data Imputation | Improve k-NN classification Method | MATLAB Programming | Compare with EM, BPCA,Hot deck, Fourier, inpaint, MEI, MICE | FLk-NN = 0.043<br>Accuracy of imputed data = 50% | Improve the Performance |
| K-NN algorithm | (Muhammad. MB et al., 2021) | Missing Data Imputation | Improve Mean and *k* nearest neighbour (KNN) Algorithm | MATLAB Programming | Compare with mean imputation, median imputation, *k* nearest neighbours, sample imputation, and multiple MICE | Classification Accuracy = 92% | Improve the Performance |
| Multiple imputation method using LSSVM (Least Squares Support Vector Machine) | (Sivapriya. TR et al., 2020) | Missing Data Imputation | Improve Imputation and Classification of Missing Data | MATLAB Programming | Compare with LSSVM-PSO classifier models | k-mean: sen = 78%, spec = 79%, accu = 80%<br>SVM: sen = 94%, spec = 94%, accu = 97%<br>CZLSSVM: sen=94%, spec = 96%, accu = 99%<br>BPN: sen = 89%, spec = 88%, accu = 90%<br>C4.5: sen = 85%, spec = 86%, accu = 77% | Improve the Performance |
| | | | | **Our Current Work** | | | |
| Hybrid k-NN Algorithm | 2021 | Missing Data Imputation in MRI image reconstruction | Improve hybrid k-NN Method | MATLAB Programming | All of Above | Generate the NaN values = x (1200–4300)<br>Accuracy = 99.9%<br>LH(Fk-NN) =0.97<br>Accuracy of imputed data =50% | Improve the Performance |

**Table 3** Overall Result of Accuracy

| Sensitivity (%) | Specificity (%) | Overall Accuracy (%) |
|---|---|---|
| 92% | 60% | 96.90% |
| 94% | 30% | 93.20% |
| 92% | 60% | 70.50% |

process factors. Here we also define L*k*-NN, which is more complicated than the hybrid *k*-NN that takes the delays that vary from one variable pair to another into consideration. For each of the delays using the learning vector, we created a set of vectors with testing and training. LPP is a way to reduce nonlinear dimensions and the LPP ensures that the nearest neighbors of the original data are maintained in the new space after calculating the dimensions to mitigate the loss of allocation values for hybrid *k*-NN. Another objective of this research is to enhance the tumor-based image resolution applied on segmented areas of the tumor to determine the cause of CSF leakage. These equations and structures mainly represent the size of the tumor and need to calculate the missing values of hybrid *k*-NN with high accuracy with the usage of the latest 4D technology using LFT. We calculated the tumor growth area of cancer cells and brain cells damaged due to CSF leakage from the first step to the last step. Using different software tools, we analyzed the forming vector that applies to self-correlation. The main area of our research is brain tumors and accuracy is the key tool for success, which is why the study suggests that MRI gets the best pictures and the best results. This research is based on LLP due to the reduction of the missing values of the *k*-NN algorithm and to minimize the distance of far values as well. Using Laplacian transformation in terms of Fourier transformation and optimization of the values, we can easily understand and detect the missing values. We also describe L*k*-NN, which is more complex than the hybrid *k*-NN and must take into account the delays varying from one variable pair to another. Using the training vector, we created a set of test and training vectors for each of the delays.

## 6.1 Findings

We summarize our contributions in this research work as below:

1. Firstly, we contributed a high-dimensional segmentation process with the graph cutting technique to calculate the accuracy of image datasets. All images have the same process implemented in graph cutting, but the quality is different in 4D. After implementing the 4D techniques on images, we can find some uniqueness for images' selection. Note that the images have improved quality as compared to 3D, but if we use color images in 4D segmentation, it provides better results as compared to black and white images.
2. Secondly, with general improvement, we address the missing data issue. The framework creates a predictive value optimization problem that explicitly addresses the missing inclusion values and can be used to generate multiple impressions. We provide the nearest neighbor functions derived from the *k*-NN algorithm. This optimization provides a new perspective on the classic data loss problem and leads to developing new algorithms to improve data accuracy.
3. Thirdly, we develop solutions to the missing data problem for each calculation model by following the general method of calculation of the Frobenius matrix, standard Frobenius matrix, reconstruction weight matrix, the relationship between training data points and test

data with rotation parameter optimization function. These methods manage non-zero values easily, and the relationship between the training and the data test points become partial data points for the training.

4. Finally, we evaluated the methods for overcoming the value of missing data for the hybrid *k*-NN algorithm in the computational experiments of the new proposed method. We use the Fast Fourier Transform method to include missing hybrid *k*-NN algorithm values with delayed correlation and multiple missing data types in time series data. To create the double optimization method, we use the FK-*k*-NN method:

1. Fast Fourier Transformation (FFT)
2. Extension of hybrid k-NN algorithm with lagged correlation
3. Improvement of the calculated data to determine the validation method
4. Variables associated with the time delay of the test and the training vector resulting from the time delay
5. Describing L*k*-NN, which is more complex than the hybrid *k*-NN and must consider delays varying from one variable pair to another using the training vector
6. Building a matrix array with a diagonal matrix D equivalent to $R_{d \times d}$.

### 6.2 Limitations

Due to the lack of resources in terms of both time and cost, some features on this research work are compromised that can add more strength to its functionality.

1. High-dimensional segmentation needs to be improved with using another algorithm which will help to find more accuracy of the brain segmentation process.
2. Applying high-dimensional segmentation for original MRI imaging, which can create ease in finding the CSF leakage problem? As a result, the health care practitioner can easily understand the initial stages of cancer in the MRI samples.

The 4D techniques applied to the human brain scanning would help medical technicians performing MRI scans to easily detect the disease.

## 7 Conclusion

In this research, we have proposed four new approaches to classify MRI images (4-dimensional) in the early stages of tumor or CSF development using the hybrid *k*-Nearest Neighbor (*k*-NN) algorithm technique. Our research addressed the issues of accuracy and the efficiency of missing *k*-NN hybrid values. *k*-NN missing values is one of the most popular research issues especially in 4D frequency. We defined L*k*-NN, which is more complicated than the *k*-NN and takes into account delays that vary from one variable pair to another. We created a set of testing and training vectors for each of the delays using the training vector. The LPP is a means of reducing non-linear dimensions. LPP ensures that the closest neighbors to the original data are maintained in the new space after reducing the dimensions, to reduce the loss of hybrid *k*-NN allocation values. We used the 4D model process approach at the beginning of the tumor development for clear-cut tumor recognition. We implemented various image enhancement techniques to improve image quality for a better prediction of cancer diagnosis. This research proposes the concept of a research framework that will identify tumors from non-tumors by using a 4D image light field model. By interfacing 4-dimensional models followed by software modeling techniques, we measure the range, frequency, ratio, masses and size of brain damage cells deep inside the CSF and also improve the performance of hybrid *k*-NN algorithms compared to the previous one.

## Declarations

**Conflicts of interests/competing interests** We all authors declare that we have no conflicts of interests/ competing interests of any sort associated with the manuscript.

## References


1. Alhroob A, Alzyadat W, Almukahel I, Altarawneh H (2020) Missing data prediction using correlation genetic algorithm and SVM approach. Population 11(2)
2. Altaf I, Vohra AH, Shams S (2016) Management of cerebrospinal fluid leak following posterior cranial fossa surgery. Pakistan J Med Sci 32(6):1439–1144
3. Armina R, Zain AM, Ali NA and Sallehuddin R (2017) "A review on missing value estimation using imputation algorithm," In Proc. International Conference on Materials Physics and Mechanics, Langkawi, Malaysia, pp. 892–895
4. Bahadure NB, Ray AK, Thethi HP (2017) Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. Int J Biomed Imaging 17(1):1–12
5. Bertsimas D, Pawlowski C, Zhuo YD (2018) From predictive methods to missing data imputation: an optimization approach. J Mach Learn Res 18(196):1–39
6. C. A. Bhardwaj, M. Mishra and K. Desikan, "Dynamic feature scaling for k-nearest neighbor algorithm," In Proc. International Conference on Mathematical Computer Engineering, India, Pp.1–10, 2018.
7. Cameron JM, Rinaldi C, Butler HJ, Hegarty MG, Brennan PM, Jenkinson MD, Syed K, Ashton KM, Dawson TP, Palmer DS, Baker MJ (2020) Stratifying brain tumor histological sub-types: the application of ATR-FTIR serum spectroscopy in secondary care. Cancers 12(7):1710–1715
8. Chowdhary S, Damlo S, Chamberlain MC (2017) Cerebrospinal fluid dissemination and neoplastic meningitis in primary brain tumors. J Mofitt Cancer Center 24(1):1–16
9. Garg, Ginni, and Ritu Garg (2021) "Brain Tumor Detection and Classification based on Hybrid Ensemble Classifier." arXiv preprint arXiv:2101.00216
10. Huang S, YuejunLyu YP, Huang M (2019) Analysis of factors influencing rockfall runout distance and prediction model based on an improved KNN algorithm. IEEE Access 7:66739–66752
11. Khotanlou H, Colliot O, Atif J, Bloch I (2016) 3d brain tumor segmentation in mri using fuzzy classification, symmetry analysis, and spatially constrained deformable models. Fuzzy Sets Syst 60(10):1–25
12. Kinaci A, Algra A, Heuts S, Donnell DO, Zwan AVD et al (2018) Effectiveness of dural sealants in prevention of cerebrospinal fluid leakage after craniotomy: a systematic review. World Neurosurg 118(1): 368–376
13. Kombo OH et al (2020) Long-term groundwater level prediction model based on hybrid KNN-RF technique. Hydrology 7.3, 59
14. Ks SR, Murugan S (2017) Memory based hybrid dragonfly algorithm for numerical optimization problems. Expert Syst Appl 83:63–78
15. Latha RS et al (2020) Brain tumor classification using SVM and KNN models for smote based MRI images. J Crit Rev 7(12):1–4
16. Lavanya SR, Mallika R (2020) A mode fuzzy weight based canonical Polyadic (Mfwcp) and adaptive neuro fuzzy Interface system (Anfis) for missing value imputation in breast cancer prediction. Eur J Mole Clin Med 7(11):2053–2066
17. Liang F, Qian P, Su KH, Baydoun A, Leisser A et al (2018) Abdominal, multi-organ, auto-contouring method for online adaptive magnetic resonance guided radiotherapy: an intelligent, multilevel fusion approach. Artif Intell Med 90(1):34–41

18. Liu J, Chen F, Pan C, Zhu M, Zhang X, … Liao H (2018) A cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas. Trans Biomed Eng IEEE 65(9):1943–1952
19. Liu Z, Pan Q, Dezert J, Martin A (2018) Adaptive imputation of missing values for incomplete pattern classification. Pattern Recogn 52:85–95
20. Mendrik AM, Vincken KL, Kuijf HJ, Breeuwer M, Bouvy WH et al (2015) MR brains challenge: online evaluation framework for brain image segmentation in 3t mri scans. Comput Intell Neurosci 1(1):1–16
21. Nasor M, Obaid W (2020) Detection and localization of early-stage multiple brain tumors using a hybrid technique of patch-based processing, k-means clustering and object counting. Int J Biomed Imaging 2020
22. Prahl LS, Stanslaski MR, Vargas P, Piel M, Odde DJ (2018) Glioma cell migration in confined microchannels via a motor-clutch mechanism. Biorxiv:1–30
23. S. Saeed and A. Abdullah, "Investigation of a brain cancer with interfacing of 3-dimensional image processing," In Proc. International Conference on Information Science and Communication Technology (ICISCT), Karachi, Pakistan, PP.1–12, 2019.
24. Saeed S, Abdullah A, Jhanjhi NZ (2019) Implementation of fourier transformation with brain cancer and CSF images. Indian J Sci Technol 12(37):1–9
25. Sargolzaei S, Sargolzaei A, Cabrerizo M, Chen G, Goryawala M et al (2015) Estimating intracranial volume in brain research: an evaluation of methods. Neuroinformatics 13(4):42
26. Saunders NR, Dziegielewska KM, Møllgard K, Habgood MD (2018) Physiology and molecular biology of barrier mechanisms in the fetal and neonatal brain. J Physiol 596(3):5723–5756
27. Schievink WI, Maya MM, Jean-Pierre S, Nuño M, Prasad RS, … Moser FG (2016) A classification system of spontaneous spinal CSF leaks. Neurology 87(7):673–679
28. Srinivas B, Sasibhushana Rao G (2019) A hybrid CNN-KNN model for MRI brain tumor classification. Int J Recent Technol Eng (IJRTE) 8(2):5230–5235
29. Ueno M, Chiba Y, Murakami R, Matsumoto K, Kawauchi M, … Fujihara R (2016) Blood-brain barrier and blood-cerebrospinal fluid barrier in normal and pathological conditions. Brain Tumor Pathol 33(1):89–96
30. Usman K, Rajpoot K (2017) Brain tumor classification from multi-modality mri using wavelets and machine learning. Patt Anal Appl 20(1):871–881
31. Wang, Lishan. Research and implementation of machine learning classifier based on KNN." IOP Conf Series: Mat Sci Eng. Vol. 677. No. 5. IOP Publishing, 2019.
32. Zhang S, Xuelong L, Zong M, Zhu X, Cheng D (2017) Learning k for k-nn classification. ACM Trans Intell Syst Technol 8(3):1–19




## Affiliations


### Soobia Saeed[1] · Afnizanfaizal Abdullah[1] · N. Z. Jhanjhi[2] · Mehmood Naqvi[3] · Anand Nayyar[4,5]

[1] Department of Software Engineering, Universiti Teknologi Malaysia-UTM, 81310 Skudai, Malaysia

[2] School of Computer Science and Engineering, SCE Taylor's University, Subang Jaya, Malaysia

[3] School of Engineering Technology, Mohawk College, 135 Fennell Ave W, Hamilton, Canada

[4] Graduate School, Duy Tan University, Da Nang 550000, Viet Nam

[5] Faculty of Information Technology, Duy Tan University, Da Nang 550000, Viet Nam