Mathematical Biosciences and Engineering

MBE, 19(9): 9039–9059.
DOI: 10.3934/mbe.2022420
Received: 22 February 2022
Revised: 15 May 2022
Accepted: 19 May 2022
Published: 22 June 2022

http://www.aimspress.com/journal/MBE

Research article

Improve correlation matrix of Discrete Fourier Transformation technique for finding the missing values of MRI images

Soobia Saeed[1,*], Habibollah Haron[1], NZ Jhanjhi[2], Mehmood Naqvi[3], Hesham A. Alhumyani[4] and Mehedi Masud[5]

[1] School of Engineering, Faculty of Computing, University Teknologi Malaysia, Skudai 81310, Malaysia
[2] School of Computer Science, SCS Taylor's University, Subang Jaya, Selangor Malaysia
[3] School of Electronics Engineering and Computer Science, Mohwak College, Canada
[4] Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia
[5] Department of Computer Science, College of Computer Science and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

* **Correspondence:** Email: saeed.soobia@graduate.utm.my.

Abstract: Missing values in the k-NN algorithm are a significant research concern, especially in low-grade tumours and CSF fluid, which are commonly identified in MRI scans. Missing values are usually ignored, but when data is mined, they can lead to bias and errors. In addition, the data is not missing at random. This study improves image accuracy, boosts the efficiency of missing k-NN hybrid values, and develops a research technique for detecting CSF fluid deposits in brain areas separated from non-tumor areas. We also offer a new method for detecting low-grade tumours or cerebrospinal fluid (CSF) formation in its early stages. In this study, we combine the hybrid K-Nearest Neighbor algorithm with the Discrete Fourier transform (DFT), as well as Time-Lagged analysis of four-dimensional (4D) MRI images. These dependencies exist in both space and time, but present techniques do not account for both sequential linkages and numerous types of missingness. To address this, we propose the DFLk-NN imputation method, which combines two imputation approaches based on a hybrid k-NN extension and the DFT to capture time-lag correlations both within and across variables. There are several types of missingness are enables the imputation of missing values across the variable even when all the data for a given time point is missing. The proposed method gives high accuracies of MRI datasets and retrieves the missing data in the images.

Keywords: missing values; hybrid k-NN; DFT; MRI; datasets

1. Introduction

When dealing with data analysis, missing values or data is a common occurrence. Human errors in data preparation, machine errors due to hardware failure, respondents' refusal to reply to particular inquiries, investigation drop-off, and the mixing of irrelevant data are all common causes of missing number [1,2]. Normally, the most ideal strategy to deal with missing qualities is to endeavor to keep away from the issue with cautious arranging. Missing data is very common and the social scientist are not often given the adequate attention, the missing problem either finessed or ignored. That's why the many researcher are working on missing data and focusing this problem why it is irrelevant for the particular studies. Sometimes, this missing data problem have no significant impact on the research studies but some times it will create the huge loss of data Many techniques and approaches are available for solving the missing imputation problem. Despite massive efforts to avoid the problem, some missing elements are usual and inevitable [3]. The problem affects all sectors that deal with data, including machine learning, artificial intelligence, measurements, and data-driven control, and it leads to a number of issues. Execution corruption, information examination issues, and one-sided results caused by differences in missing and complete values are among these issues [4]. Furthermore, the amount of missing data, the type of missing data, and the process underlying the missingness of the data all affect the severity of misplaced values [5]. Imputation [6] is a strategy for dealing with missing data that involves removing occurrences and replacing them with possible or estimated values [7–9]. The imputation strategy is observed as a more complex solution that reflects a number of aspects when dealing with misplaced values [10]. This is because it could lead to calculation and investigation errors, as well as data methodological discrepancies. Removal of instances with missing values is a path of least resistance and to some degree a methodology, yet the results would be the deficiency of valuable data. Values that aren't recorded in a dataset are known as missing data. It could be a single value missing in a single cell or an entire observation missing.Aside from deletion, imputation is the most embraced approach for handling missing values particularly on the pre-handling step in data analysis. This method fills in any missing values with the last observed value seen from the same subject. When a value is missing, it is replaced with the last observed value from the internal structure of the missing matrix and an auxiliary matrix of (nonmissing) respondent background data; they also account for individual variances in respondents' numerical rating scale usage. This methodology used the number of numerical and machine learning approaches have been developed in the literature, including mean imputation, regression, K closest neighbour, ensemble-based imputation, and others [11,12]. In any event, machine learning approaches have recently been proved to be more powerful in dealing with missing information when compared to statistical techniques. There are two types of machine learning approaches are used in this study that are supervised learning, unsupervised learning, and also Reinforcement Learning being particularly popular in missing data. Machine learning can be implement as an association analysis by the use of supervised learning. This research focuses on the advantages and disadvantages of the supervised learning classification algorithm. The supervised learning strategy mainly depends on labelled data and ignores the potential impact of unlabeled data in missing value prediction [13]. Supervised learning develop the class label for the purpose of distribution through coincide model using the predictor features. When the predictor features values are identified but the class label values are unidentified, the resulting classifier is used to test the assign class label instances.The unsupervised algorithm, on the other hand, learns to identify features and patterns for missing data completely on its own from unlabeled data [14]. To address the shortcomings of traditional supervised and unsupervised imputation methods, hybrid approaches have been used [15–19]. However, keep in mind that the main reasonable solution boils down to a moral plan and thorough investigation [20]. This is on the grounds that investigation of execution is dependent however not restricted to a few factors, for example, the algorithm selected, attribute selection and sampling techniques. Similarly, the time of big data approaches, it is difficult to manage just because of the process of training the datasets was not designed and also not work well with the large form of big data [21,22].

The research interest of this paper is to develop the unique technique to identify the missing data in MRI images. This technique is based on k-NN in classifying low-grade tumor and CSF. This technique focusses on detection method in which the aforementioned problems are considered in order to improve the classification and detection method for trained MRI datasets of low-grade tumor and CSF fluid. The technique work efficiently through the correlation method of time lagged with hybrid k-NN algorithm. The correlation matrices impute the missing data easily and find the delay in the data and reconstruct the new imputed data. DFT also perform the important role to retrieve the missing data in the same rows and column in sequence way. As a result, while dealing with missing data, the technique is consistently important as improper handling may result in incorrect deductions. This paper is organized as follows: Section 2 describes the past history of connected articles by presenting relevant work. Section 3 represent the proposed Correlation Matrix of Discrete Fourier Transformation (CM-DFT) Technique in the hybrid k-NN algorithm technique Section 4 summarizes the results and discussions, as well as their implications and accomplishments. Finally, in Section 5, the conclusion is presented, along with recommendations for future research.

2. Related works

In real-world datasets of image features, Missing imputation data in MRI images is one of the common problems nowadays. Due to this problem, numerous researchers and scholars try to overcome the missing data imputation through different techniques and improve the k-NN classification. Machhale et al. [23] the idea of the k-NN algorithm is not new, but to find the missing imputation values is mainly used in classification. There are two methods to cope with tumor and CSF detection in MRI datasets and missing imputation in k-NN algorithm classification. The first method is producing the hybrid k-NN algorithm for classification and the second overcome the missing imputation values in the k-NN algorithm. Developing and improving 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 images quality, better tumor detection, improve missing data extraction and classification performance. This is the main idea behind all proposed approaches in this research study. Dritsas et al. [24] Classification of k-NN algorithm is a well-known technique that uses for tumor and CSF detection to improve the accuracy and quality of an image and reduce the imputation missing data problem. Beretta and Santaniello [25] discuss the predictive models in their proposed technqiue, including k-NN, SVM, and decision tree for selection tailored of numerous modifications. Derrac et al. [26] propsed the fuzzy logic techniques which is used in the methodology to maximise the assigned values of the k-NN algorithm to retreive the misplaced data in fuzzy k-NN datasets by maximising the given value assigned in the k-NN algorithm. Pattern classification approaches were proposed by Armina et al. [27] to improve data sets and forecast models to handle missing value difficulties. Nezamabadi and Kabir [28] proposed and gather the data repositort datasets of UCI Machine Learning to improve the sample accuracy of 84 images. Alexander et al. [29] also optimised the sample output's given values in order to train learning algorithms that calculate dataset information. They also described an approach that was shown to be more accurate and efficient than previous k-NN algorithms in computer studies. Schievink et al. [30] proposed a multi-label classification technique for classifying the dependability of an incomplete pattern. S. Chowdhary and colleagues [31] discuss the attribute values from image databases were used to classify brain cancer. The algorithm were applied for detecting the misplaced data, binary transformation image datasets are classified. Missing data play an important role for locating an appropriate calculation if the images cannot be properly categorised. In this scenario, the misplaced values are estimated using k-NN and self-organizing approaches.

The propsed CNN-k-NN algorithm technique for the classification of the MRI brain tumor images that integrates the k-NN algorithm and convolutional neural networks. Before incorporating them into the k-NN algorithm for class prediction in an image collection, the CNN model is utilised to implement its feature extraction properties. The authors increase the performance of the proposed model accuracy is 96.25%, that is precision, error rate, sensitivity, and specificity discoveries in image testing [32].

Po et al. [33] describe the effective classification methods of k-NN algorithm. It is a lazy learning method because of its less accuracy and as it depends on choosing a good value for "k", it cannot be used for large repositories like in dynamic web mining. In machine learning and k-NN algorithm even in medical areas, the existence of missing data is a main and serious issue. Knowledge or information is drawn out according to the quality of data, if some values are missing in the data, and then it will affect the descriptive along with inferential statistics and predictive analytics as well. Moreover, for data imputation k-NN serves as an effective approach to predict the misplaced values. A model is developed for every feature that has misplaced values and taken as input values or perhaps for all other input values. K-nearest neighbor 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. Nowadays, many researchers are struggling to handle the problem of data misplaced 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 misplaced values. When an algorithm is applied, these misplaced 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 misplaced values. The demand for a proper technique arises as the data is getting bigger periodically so that the chances of misplaced values reduction. This research explains the details of the existing models for solving the problems of the missing value, which helps in developing and applying a new method. As these are based on earlier research, they are not providing good outcomes in a 4D segmentation method. Many issues of misplaced values of the k-NN algorithm are attempted to enclose in this research. Therefore, the main purpose of this research is to counter this issue with minimum data loss, and tasks are performed under the demands.

Chu et al. [34] developed a new technique to deal with missing imputation problems in the k-NN algorithm. Misplaced imputation information of MRI images in k-NN algorithm with Fourier transformation (FT) is also employed. The k-NN algorithm locates a complete data set for the k number of neighbors 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. 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 this problem by suggesting an imputation method, namely Fourier k-nearest neighbours (Fk-NN), which is based on two imputation methods such as k-NN algorithm and Fourier transform. If all the data at a definite time point is misplaced and if different misplaced types appear within and across variables, this proposed method will allow the imputation of misplaced values.

Niranjana et al. [35] Based on the previous studies, a combination of the Fk-NN technique covers the imputation missing data problems. However, these techniques still have some problems. Although, Fk-NN performs better than other technique, but only find the missing data in one way at a time either rows or columns Fk-NN not only reduce the missing data problem but also a week to remove 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 instant, Fk-NN does not provide the best strategy for reducing the misplaced information in k-NN algorithm classification. Likewise, these methods do not provide a robust technique to identify misplaced information efficiently in the k-NN classification method. However, for the missing imputation data issues, most studies had constructed other techniques and classifiers using the Fk-NN technique without considering images quality improvement.

Zhang et.al. [36] discuss the new Fractional Fourier transform (FRFT) technique which is based on Fourier transform (FT) that is simplification of new linear transform. FRFT can be used unified time-frequency domain in to the signal transformation. This study provide a complete details of FRFT in this survey, The first step represent the three different variations such as weighted-type, sampling-type, and eigen decomposition-type and explain the process of FRFT. The second step conduct the technical studies with theoretical research such as software implementation, hardware operation, and optimal order selection. In the end, this FRFT application was able to work in the field of communication, cryptography, optimal engineering, radiography, remote sensing, fractional calculus, fractional wavelet transform, pseudo-differential operator, pattern recognition, and image processing.

DFT has been studied extensively, and study is still ongoing. Saeed et al. [37] suggested a DFT-based method for evaluating the correctness of medical image collections using two quick execution methods. This research study helps to analysis of brain tumor due to the usage of process of 4D image segmentation and Fourier transform. The size of the brain tumor and CSF range can be calculated by using 4D images and MATLAB software modelling methodologies, according to the scientists. By minimizing errors at boundaries, LFT can support the improve the quality of the light field editing application for segmentation and the LFT composite pipeline [38,39].

Wang et al. [40] describe the Enormous type of data that are generated the result to the development of biomedical imaging modalities such as Photoacoustic Tomography (PT), Computed Tomography (CT), Optical Microscopy and Tomography, and so on. While large biological data sets provide more information about diseases, they also pose significant obstacles in terms of adequately exploring the information. Data fusion methods combine various data observations to increase the consistency of the information, which leads to a better understanding of the data. However, data production is only the first step in the fusion process; noise, misplaced data, data shortage, and high dimensionality are all factors to study. This study provides an overview of data pretreatment achievements in biomedical data fusion, as well as insights gleaned from recent developments in the field.

Zhang et al. [41] discover neuroimaging as a technique for overcoming the inherent limits of individual imaging modalities by combining data from various imaging modalities. While also connecting physiological and cognitive data, neuroimaging fusion can improve temporal and spatial resolution, contrast, and imaging aberrations. This study present a review that includes an overview of current multimodal fusion issues and future research recommendations. The present medical applications of fusion for specific neurological illnesses, strengths and limitations of available imaging modalities, fundamental fusion rules, fusion quality assessment methodologies, and atlas-based segmentation and quantification fusion applications. Multimodal fusion has a lot of applications in clinical diagnostics and neurological research. Engineers, researchers, and doctors will benefit from more widespread education and study in the field of multimodal neuroimaging.

In conclusion, missing imputation data is one of the most challenging issue in the segmentation process especially in MRI images datasets. Many data are lost in the imputation application is common in various classification method. The misplaced i8mputation data problem can be solved using a variety of strategies and approaches in hybrid k-NN algorithm. The proposed CM-DFT technique is to develop to minimize the time delay and solve the missing values in the data. This research utilizes the proposed CM-DFT technique is used to combine the hybrid k-NN algorithm which is extended from the previous section to given that an improved solution for extracting the misplaced data in the MRI images and that's why we prefer to implement the correlation technique which is based on two time series function. The reason for selecting the correlation matrix is to extract the misplace values in the MRI images and rebuild the new data in the images that is minimize the computational time during extracting the data. Many important data and information either in software, hardware, sensors, or images are lost just because of irrelevant features or noise are presence in the data. This technique use DFT approach for reconstructing the new data in the same columns and rows that can researcher retrieve the reconstructed the information completely. DFT is a better method for locating misplaced data sequencailly. As a result, CM-DFT technique provides a more effective solution for misplaced imputation problem in MRI images. Hence the proposed technique provides improved results to enhance and identify the misplaced imputation data of trained MRI datasets using this technique by the combination of LHk-NN-DFT.

3. Trained datasets

This data gathered the original MRI images from three different hospitals, namely the University of Calgary, Canada, National Cancer Care Institute (NMI), and Medicare hospital, Pakistan, to investigate the improved performance of low-grade tumour and CSF detection in MRI and the missing data problem in a k-NN algorithm. We used human brain samples photos from the Cuming School of Medicine Lab at the University of Calgary in Toronto, Canada. The source of data acquired is https://sites.google.com/view/calgary-campinas-dataset/home. Cuming School of Medicine Lab is in charge of neurology and associated research. For the experiments, three different types of MRI LFD datasets is introduced in this research by utilizing the Light Field Toolkit namely CSF images, low-grade tumor, and CSF with low-grade tumor images including the raw data trained by the tools of Donald et al. [42] at Stanford University at https://Vincentqyw/Depth-Estimation-Light-Field/. These datasets were produced by the use of Lytro Illum software tools. Each types of datasets consists of images and the corresponding raw (decoded Lytro ESLF) files. To classify these LFD MRI images, output images, and metadata obtained using the Lytro command-line tool have also been included. The raw data imported from Lytro Illum is a light field image file (approximately 32M in size) in .lfp format. Silva [43] describe the file contains the following parts: light field image data as a raw data; metadata; light field image size. Next, decode the light field file to obtain the light field image, and then perform demosaicing and color correction on the light field image to obtain RGB color light field images to increase the pixel quality and improve the resolution of these MRI images. These trained MRI images show the behavior of the Lytro Illum light field process for MRI images.

The collection MRI datasets images of CSF leak and low-grade tumor are more than 500 patient data from the hospitals records and research lab. Datasets from three hospitals and lab are used in this research work and they are converted into images from using the LFT kit to become the improved images. The resolution of these images is based on 9 × 9 × 512 × 512 × 3 light fields as individual PNGs and depend upon the size of images. The Configuration files with software settings and disparity range are 512 × 512 and 5120 × 5120 depth, and disparity are 512 × 512 and 5120 × 5120 evaluation masks as PNGs. Thus the accuracy obtained from the original datasets is increased.

4. Material and methods

Time-lagged and DFT have two modules of proposed CM-DFT technique which play an important role in the hybrid k-NN algorithm. Figure 1 shows the correlation matrix of the time lagged and proposed hybrid k-NN technique as well as the DFT. The inputs of this proposed technique are multiple phases of MRI images of CSF and low-grade tumor that is sufficient for finding the missing data imputation. The first module of the proposed technique consists of two phases: 1) correlation variable of lagged time delay, and 2) proposed hybrid k-NN algorithm. In this phase, operations such as finding the missing values, reconstructing the new data into a numeric form, or replacing the empty space with non empty space(retrieve the numeric values) are performed with the help of correlation matrix of a time-lagged with the hybrid k-NN algorithm.This research work focus on correlation coffiecent variable of two time series function which is mentioned in detail in the next subsection. In the second phase, an operation such as applying the DFT technique in the application of Discrete-Time Fourier Transform (DTFT) to identify the missing data in the same sequence of signal length in time series data. The details of the proposed technique is given below in detail in the following section.

4.1. Proposed method

Missing imputtaion data in MRI is one of the most challenging problem in segmentation method. Few data points missing issues in various classfication of MRI datasets is very common in imputation application. This research used various strategies and approaches to be solve the missing data impuation in hybrid k-NN algorithm. The CM-DFT technique is utilised to decrease misplaced data and minimise time lag (delay). This study focuses on the location of tumor and CSF in MRI which is difficult to identify. CM-DFT technique is used to provide the improved solution for retreveing the values that is misplaced in the MRI datasets using the combination of hybrid k-NN algorithm from the extension of previous section. Discrete Fourier Transformation (DFT) is used to apply the findings in this part to a correlation matrix. Correlation matrix was implement with the combination of time-lagged and hybrid k-NN algorithm for reconstruct the missing data in the MRI datasets and decrease the computational time of the proposed technique. In the end, DFT is used to extract the values that are not present in the same rows and column sequencailly in this research. Hence, CM-DFT provid the better solution of resolving the misplaced imputation data problem in MRI images.

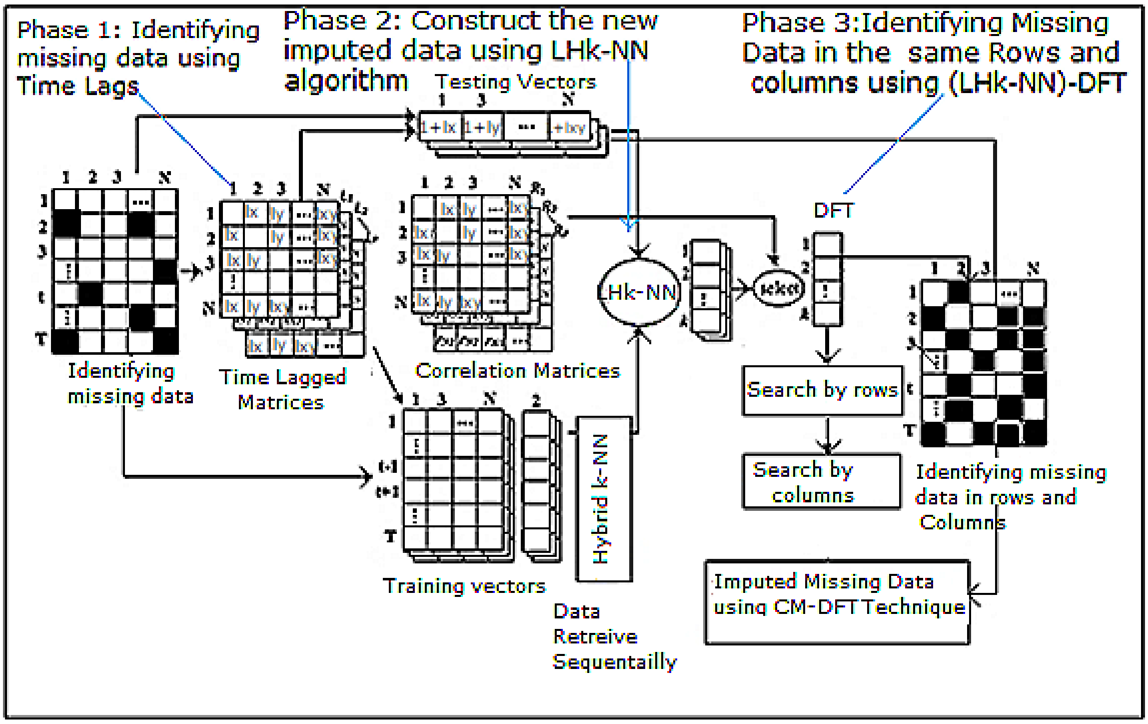


Figure 1. Block diagram of Imputation of missing data using the DFT.

4.2. Time lags technique

This section discuss the solution time delay in reconstructing missing imputation data in the hybrid k-NN algorithm. We employ the hybrid k-NN algorithm and time-lagged to solve the missing data problem. This algorithm consists of two parameters, such as "k", for the nearest locations, and the number of time delays. These parameters are implemented in an organized manner to select the nearest neighbor values to make the strongest correlation for each variable and k near-est location points across all lags.

This technique is applied to minimize the time delay for testing and training vectors of MRI datasets, as this research is the focus to identify the variables which are correlated with time lags. The technique in this research uses the cross-correlation method to create the similarity measurement of the two-time series functions, and the time delay is implement to both of them. The correlation function is defined by the variable of u and v for the time delay(d) which is represented by the formulation of Eq (1):

$$C_{uv}(e)=\frac{Fcuv(d)}{\sqrt{Fcuu(0)Fcvv(0)}} \tag{1}$$

whereas, Fc represents the correlation function, d is the time delay, and u and v are the cross-correlation variable of the proposed datasets in Eq (2)

$$Fc_{\,uv}(d)=\{\frac{1}{C-d}\sum_{a=1}^{T}(ut-\acute{u}\ \ )(vt+d-\grave{v}), if\ \ d\geq o$$

$$\frac{1}{C+d\sum_{a=1-d}^{T}(ut-\acute{u}\ )(vt+d-\grave{v}\ )} \qquad otherwise\ \ 0\ \ , \tag{2}$$

whereas, T is represented the series of length of ứ and ȕ variable that is indicate the u and v average values in Eq (2).

Furthermore, the correlation matrices are constructed by the number of time lag (p) values in the ordered form of 1, 2,...,p, by reducing the strength of delay, which is the number of time lags that indicates the future event of time series function. Thus each pair of the variables holding the lag (time delay) make the correlation of the strongest variable (maximum size) and the lag with the weakest in reconstructing the missing data in the images. Each L variable makes the correlation between the numbers of time lag with the M variable. L and M are the generated variables of each pair which makes the correlation function with time lag indicated by M × M in the trained MRI datasets.

An element of L can be positive or negative based on two situations are given below:

1) L is positive when the number of the variable v have a response delayed in terms of unit time Luv to values of u in datasets

2) L can be negative when the number of the variable u have a response delayed in terms of unit time Lvu for the values of v in the datasets can be represented by Luv=- Lvu. As a result, all correlation values are recorded in matrices, which are used in the hybrid k-NN algorithm's nearest location selection of k.

4.2.1. Lagged hybrid k-NN technique

The proposed CM-DFT technique is combined with the LHk-NN technique, which is more efficient than the k-NN algorithm. This proposed technique is applied to define the multiple lags that vary across the variable pairs and create the training and testing vectors of datasets images of every single lag (p). This technique constructs the function u over time t with some missing values and creates the relationship delay with v and z variables of MRI datasets where u has a time-lagged and produces the term of lags (p) Luv and Luz separately. The testing vector of datasets is constructed using the values of v and z at lag (p) T + Luv and T + lag (p) Luz. The training vectors, which are the applicant values of imputation saved individually, are also created using the same manner. After removing the accumulation of mistakes, the training vectors generate the outcomes of the current values of u and time T after adding the lags (p) inside the data length (1 to T). The unit time of training vectors of missing values is developed using this approach, which is represented by:

[max (1,1-min(lu1,…..luM)),min(lu1……… luM))]

4.2.2. Finding missing imputation data using LHk-NN technique

This research extends the previous work on the LHk-NN technique for identifying the imputation of misplaced values in the given MRI images. One of the problems with low-grade tumors is that the malignant area is not properly visible in MRI images.

For ease, This research uses the weighted modification distance approach in the k-NN algorithm for finding the missing values which are nearly similar to Mahalanobis distance. The weighted modification distance can handle the missing values of both the training and the testing vectors of datasets. The strength of correlation between the variables and time lags (p) are in contrast to each other. This technique needs to incorporate measurement of weight distance in the current datasets for calculating each neighbor or correlated variables of missing values if it is receiving any missing information. This technique confirms that the correlated variables and their connected lags (p) are more weighted than the weakly correlated variables or nearest values of "k" in time. The new constructed data are imputed due to the correlation matrix of time lagged with hybrid k-NN algorithm and retrieve the misplaced information in the MRI images in terms of replaced the empty space into the non-empty space and find the ne generated numeric values by the use of Eq (3).

$$\mathrm{d}\ (u,v)=\frac{\sum_{i=1}^{n}(ui\wedge vi)\times(ui-vi)2\times wi}{\sum_{i=1}^{n}(ui\wedge vi)} \tag{3}$$

The correlation numbers of missing variables and w_i variables are normalized, where n is the number of variables and w_i is their weight. Here is the wj, which denotes the non-missing variables u and v, confirming that both values are contained in the coherent values. The datasets of the variables is computed between the two vectors of non-missing pairs by the average weighted Euclidean distance. This means that highly correlated values have a bigger influence on distance than less correlated variables. Finally, for each L matrix, the outcome is expressed in terms of lag (p) sets of k-nearest neighbors. This research calculated the average values of k neighbors with the lowest weighted distance neighbors (remove the sets of lag (p) in k neighbor's values).

4.2.3. DFT approach

This approach combines patterns with variables to deal with datasets of images that need to be reconstructed. This approach utilizes the (DFT) from the previous section of the LHk-NN technique, which is implemented to each variable to impute the misplaced information. Firstly, the data segment is constructed with the starting data of the indication of the last non-missing values. Whereas the given values in the previous section are imputed through $lu1, \ldots, lM$ are missing values in the datasets. CSF is a leak in only one direction in the brain so that's why this research uses DFT for extracting the misplaced information in the MRI datasets. For the same limited sequence of rows and columns, DFT is used to determine the location and degree of leakage in the brain in one direction. DFT translates a finite sequence of data into an equally spaced variable of the function into the same length of signal of the DTFT, which converges to the continuous Fourier transform in time series. In DFT, the complex-valued function of the variable is also dealt with via the DTFT. DFT works in the sequence form of the frequency signal to detect the missing or hidden data by following the direction of CSF fluid or tumor. In addition, DFT generates the Fourier descriptors variable v, and IDFT redevelops the length of the signal (F, T) in terms of Fourier descriptors to impute the k-NN misplaced information in the same series of the sequence. DTFT is continuous (and periodic), and DFT gives discrete samples of one cycle, whereas DFT provides discrete variables of one series and turns all data (non empty space) in the same sequence into one series. The DFT returns all non empty space values from one DTFT cycle if the original sequence is one cycle of a periodic function. In this study, the suggested hybrid k-NN method is employed to solve the misplaced information imputation problem for MRI datasets by taking into account the correlation matrix. Using each of the approaches given, this study imputes one value and then combines them. In comparison to prior procedures, this unique technique is more stable and produces unbiased data. The two methods estimate these values, and generate the result in terms of combined DFT-LHk-NN approach.

5. Experimental results

This Section describe the computational and statistical outcomes of simulations formulated by statistical significance analysis, that is represent the performance of the proposed CM-DFT technique in the hybrid k-NN approach to identifying the misplaced information imputation. Figures 2 and 3 shows the statistical results of time-lagged with a correlation coefficient of misplaced imputation for two-time series values with the help of the extension of k-NN imputation with the lagged cross-correlation matrix. Thus, simulation has been performed with various calculated lags values for positive or negative values of variable in unit time such as 0, −1, 1, 2, −2, −100, 100, −3, 3, 4 and calculated Lags weighted using correlations such as 0, 29,837,625, −29,837,225, −57,825,750, 57,716,900, 2,875,815,000.00000, −2,826,305,000.00000, 84,300,525, −84,145,350, −109,525,700. The section of this research mentioned the details of our proposed method for imputed misplaced information in two-time series data with lags and various types of missingness within a variable using the cross-correlation method, devise a strategy for hybrid k-NN algorithm extension with time-lag correlations. Because correlations can persist for long periods and time measurements can be indeterminate when used to find missing imputation data. This section shows the simulated results that generate the lagged with correlation matrices which are constructed by the number of time lag (p) values and reducing the strength of delay in the time series function. The simulation results generate the graphs in the form of testing and training vectors which shows the holding variable of each pair of lag (time delay) that is made the correlation of the strongest variable with the maximum size and the lag with weakest in the procedure of reconstructing the missing data in the form of numeric values of hybrid k-NN algorithm. It also represents the correlation between the number of time lag with each pair of variables that shows the generated values of each variable makes the correlation function with a time lag in the proposed MRI datasets.

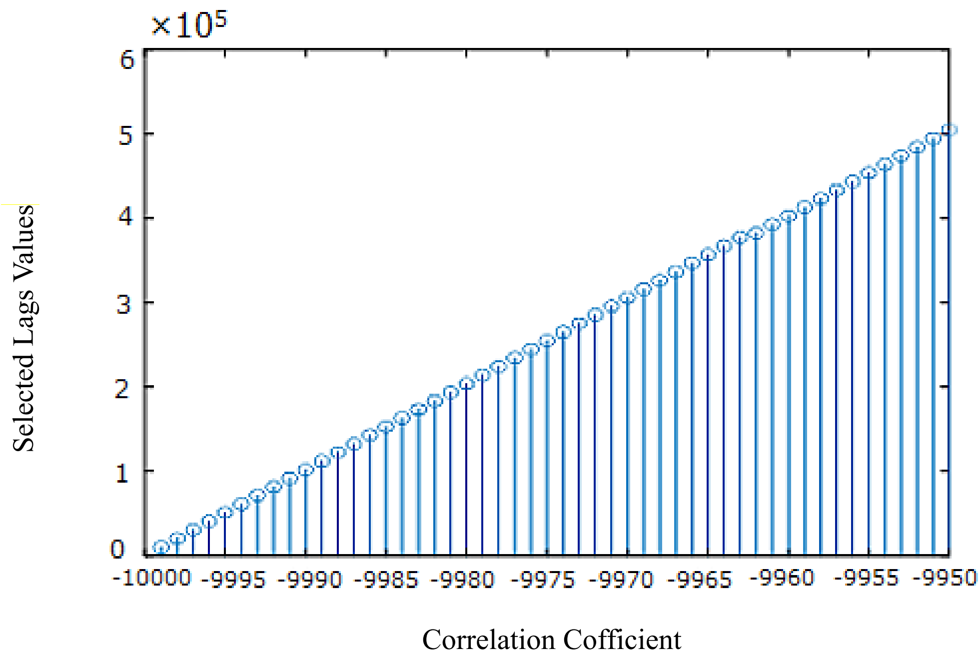


Figure 2. Time lagged with a correlation coefficient of missing imputation of time series data.

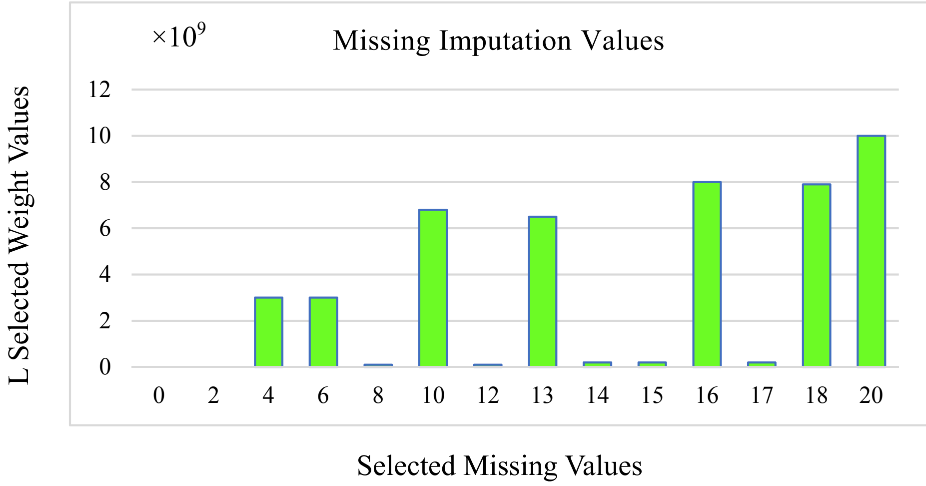


Figure 3. Calculated values of Time lagged in hybrid k-NN algorithm.

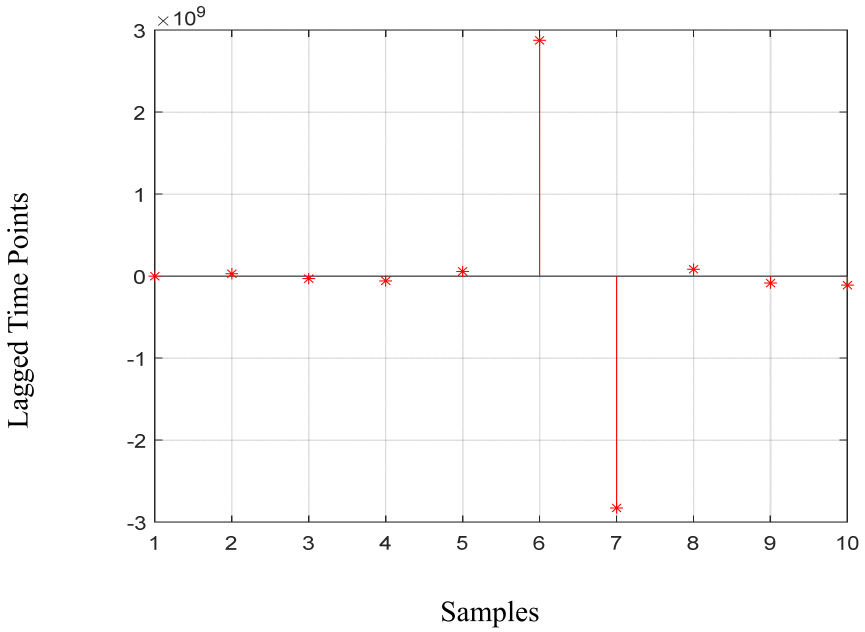


Figure 4. Lagged time point of missing data.

Figures 4–6 show the average lagged calculated values which is -5.9969×10^6 and cross-correlation matrix of calculated values such as 1.0000, 0.8927, 0.8927, 1.0000 which indicate the lagged time point with a correlation coefficient of the hybrid k-NN algorithm at the point k = 1–10 of training and testing vectors for trained MRI datasets. We take the values of hybrid-NN algorithm 1–10 for identifying the missing values till 1–10 for calculate average lagged and try to reduce the maximum delay with the help of our proposed technique. This shows the strongest correlation of hybrid k nearest neighbours across all lags, weighted by the strength of the correlation using lagged time point averaging data. Here are the simulation results show the weight distance of the given datasets for calculating the values of each neighbor or correlated variables of missing values which represent the connection of time lag and correlated variables in the hybrid k-NN algorithm. Table 1 shows the missing values that are generated by replacing original values with not a number (NaN) taking the values of converting the testing vectors into a single variable in time series data. Calculate the cross-correlation of time-lagged into 1.0 second as the correlations between the variables during the imputation, and it can deal manifold types of missingness happening in a single variable by taking the maximum values of k (k = 1–10) to fill the missing values are performed empty space into non empty space (numeric form of values). These experimental results generate high accuracy in the case of empty values that would be able to impute the missing values for MRI datasets of low-grade tumor and CSF fluid in the images. These results obtained the better accuracy of MRI datasets of tumor or CSF fluid, especially for non-visible or hidden data. The results of the experiments show that missing values can impute the initial phase of a tumor that is not in a sizable form and deposit CSF fluid data in MRI images. These findings are increase when the dataset is misplaced, with numerous sequentially misplaced values and totally empty instances in the collection after retrieving the missing values.

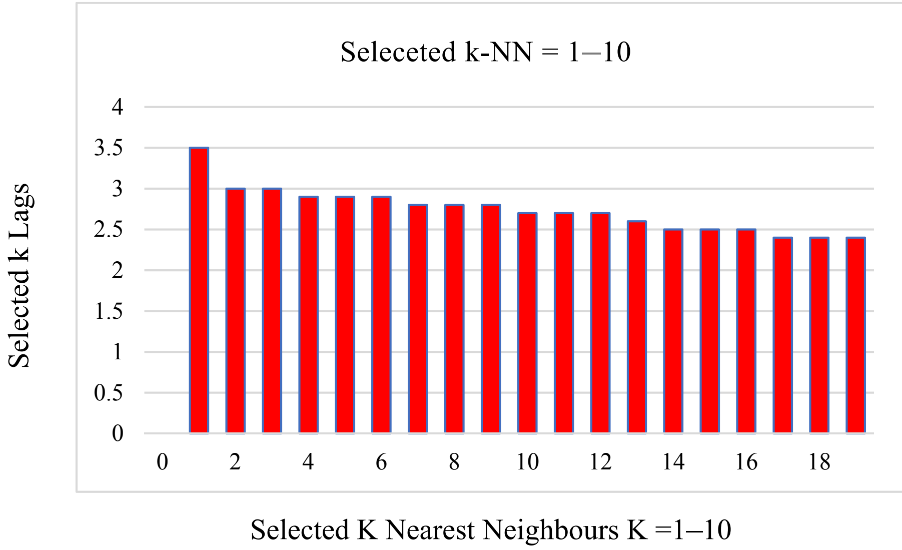


Figure 5. Lagged time point of missing data in the selected values of k = 10.

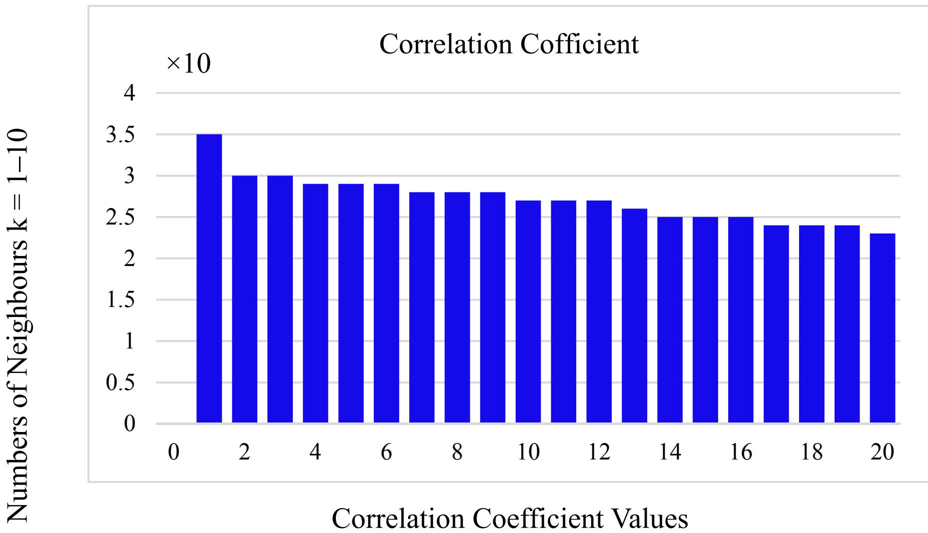


Figure 6. Lagged Time point with a correlation coefficient of the hybrid k-NN algorithm with k = 10.

Table 1. Generating the missing values.

Retrieving Missing Values	NaN (not a non-missing values)
X (1100)	= NaN;
X (1200)	= NaN;
X (1400)	= NaN;
X (1600)	= NaN;
X (1800)	= NaN;
X (2000)	= NaN;
X (2200)	= NaN;
X (2400)	= NaN;
X (2600)	= NaN;
X (2800)	= NaN;
X (3000)	= NaN;

Table 2. Simulation Parameters of hybrid k-NN algorithm and DFT.

MRI Datasets Image	Hybrid k-NN Parameters	DFT Parameters
CSF Images	K = 1–10	n = 2 points
Low-Grade Tumor Images	K = 1–10	n = 6 points
CSF with Low-Grade Tumor Images	K = 1–10	n = 8 points

Time-lagged and Discrete-Time of Fourier Transform matrix parameters are used to improve the missing data in the images and reduce the imputation problem of misplaced information using the CM-DFT technique particularly for MRI/CT scans. The performance measures are divided into three categories of datasets in this research, as shown in Table 2.

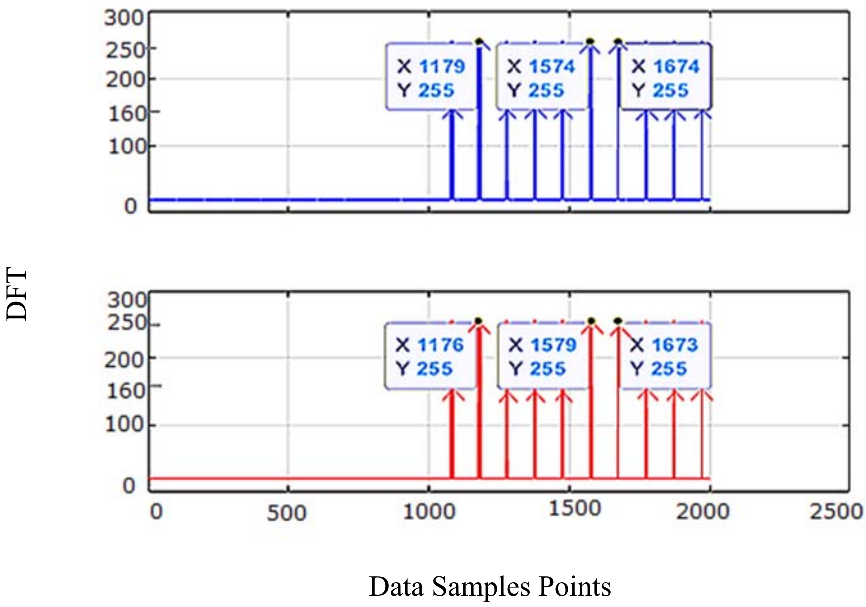
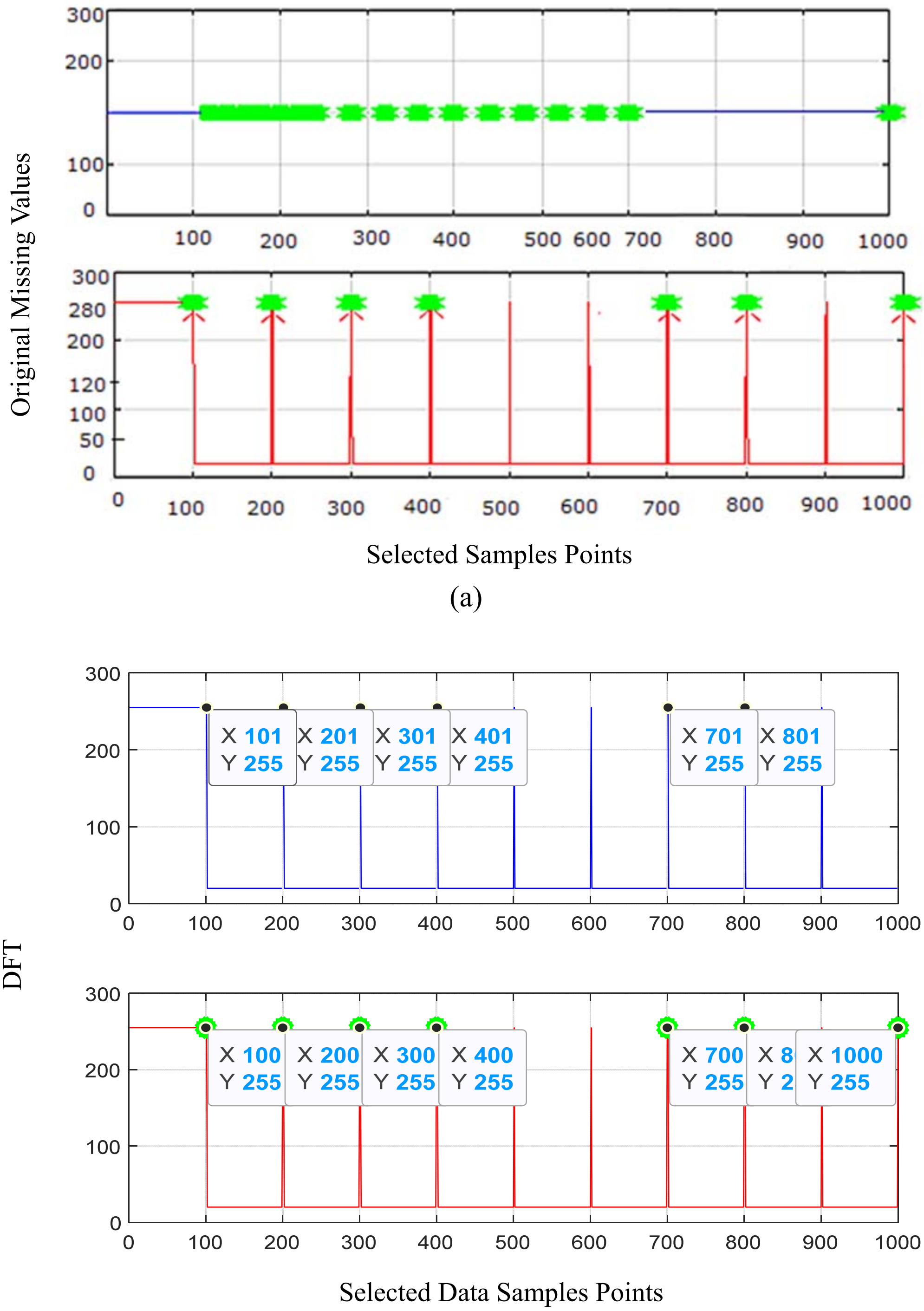


Figure 7. Missing values extraction through DFT and DTFT in the same sequence of rows and column and replace the empty space into non empty space (numeric values).



(a)

(b)

Figure 8. (a) Graph with missing values are invisible form. (b) Graph with missing values are visible form.

After imputing the missing values, Figures 7 and 8(a) demonstrate the combination of missing values that are substituted by non-empty space values. Figure 8(b) shows the imputed missing data after retrieve the data; LHk-NN impute the missing data points where as the DFT calculate the imputed values for one variable with the experimental misplaced data points in terms of DTFT, the original data is indicated the blue color in the given Figure and red color indicate the imputed data. Figure 8(a),(b) shows the difference of both before and after imputation of retrieving the missing values results. This study, MRI datasets is used for finding the misplaced information and time lagged used for correlation values which is mentioned in details in Section 4. From this research, the time instances are completely absent that have the misplaced information in LHk-NN and the missing values were reconstructed in to the same rows and column due to the non empty test vector are used in the time lags. In this situation, Lk-NN was impute the missing values and DFT calculate the missing imputed data in terms of DTFT where the missing values were reconstructed and single values in the same arrangement form are substituted for the missing values. DFT also used for find the hidden data as well as DFT generates the discrete variables of one series and convert all data in the non empty space (numeric values) with in the same sequence. DFT also used for identifying the signal length due to the change of one function in time domain to another function of frequency domain. Finally, DFT decomposed the imputed data that is obtained by the finding the nearest location of the sequential values in the same direction of the variable. Tables 3 and 4 show the accuracy values of missing data and correlated values of time-lagged also which indicate the accuracy of all three datasets and generate the non-missing values.

Table 3. Accuracy values vs. images.

Accuracy Values	Dataset-Image-I	Dataset-Image -II	Dataset-Image -III
	0.9993	0.9988	0.9984

Table 4. Correlation values.

Cross-Correlation Values	Dataset-Image -I		Dataset-Image -II		Dataset-Image -III	
	R = 1.0000	0.9468	R = 1.0000	0.9201	R = 1.0000	0.9509
	0.9468	1.0000	0.9201	1.0000	0.9509	1.0000

Table 5. Cross-correlation of Lk-NN algorithm with DTFT.

MRI Datasets Images	Missing Ratios	LHk-NN	DFT	Execution Time (Second)
D-Images-II	0.9%	99.17%	98.92%	2.1093
D-Images-III	10%	95.52%	93.58%	2.0753
D-Images-IV	12%	99.78%	95.52%	1.5331

Table 6. Calculated values of (LH *k*-NN)-DFT.

Name	Imputed Empty Instance	Cross Correlation of Time-Lagged	Missing Values of the ariable
Hybrid k-NN	No	No	No
Discrete Fourier	Yes	No	No
LHk-NN	Yes	Yes	Yes
LHDFT(k-NN)	Yes	Yes	Yes

Table 5 shows the retrieving missing values which are hidden in the MRI images. The section of this research is to select the target the missing ratio is reached. The ratios are 0.9 to 10 to 12% for missing ratios. The maximum length of consecutively misplaced values is retrieved by calculating the LHk-NN and DTFT values and calculate the average values in terms of combining the DFT and LHk-NN algorithm as shown in Table 4.

Table 6 show the results of our calculated values of LH (DFk-NN) values using this proposed novel technique where the method is compared in terms of calculating impute the missing values including the missing time instance, the inclusion of correlation of time lags, handle the ability to all of three trained MRI datasets information.

6. Results and discussion

We made it clear that the resulting training vector are linked with calculated information of the validation variable process in order to develop the experiment's time delay method. We use LHk-NN which is more complex than hybrid k-NN as it proceeds the accounts for different delays among variable pairs and built the series of training and testing vectors for each delays with the help of learning vector. The motive of this research is to detect the missing data of low-grade tumor and CSF in the images. These finding indicate the location of CSF and in sizeable form of low-grade tumor in the MRI images in early stages that is need to calculate the misplaced values of hybrid k-NN algorithm missing data with high accuracy. We examine the forming vectors that is implemented by the self-correlation method with the help of graphs and retrieval accuracies. The main purpose of this research is to identify the misplaced information in the same rows and column sequencailly. We develop the sets of testing and training vectors for each of the delays using the NaN values and generated the graphs after retrieving the misplaced information. In order to investigate the effect of the proposed CM-DFT technique for handling the misplaced information values or lost data in the proposed hybrid k-NN algorithm, this Section extend the previous work for retrieving the missing imputation values in the proposed technique. Based on the algorithm and experimental work have achieves the improved and efficient results and compare than the existing work in the previous studies. Tables 2–5 show the improve results of proposed technique. By inspecting the results, the following facts can be observed. It was providing the improve results due to the use of DFT to extract the missing values in the proposed technique. This will help to retrieve the missing values in the same rows and column in the same sequence.

7. Conclusions

In this study, we used the CM-DFT technique to offer new methods for classifying MRI images in the early phases of tumour or CSF growth. The correctness and efficiency of missing k-NN hybrid values were the focus of our investigation. One of the most widely studied topics, especially in medical MRI pictures, is K-NN missing values. LHk-NN is a more sophisticated version of k-NN that accounts for delays that differ between variable pairs.We used the training vector to generate a set of testing and training vectors for each delay. The DFT ensures that the closest neighbours of the original data are kept in the same order after extracting missing values to reduce the loss of hybrid k-NN allocation values. For clear-cut tumour recognition, we applied this unique technique fairly early in the development of low-grade tumours.We used a combination of LHk-NN, (Lk-NN)-DFT, and less execution time to extract missing data, missing ratios, and a combination of LHk-NN, (Lk-NN)-DFT, and less execution time to improve picture quality for a better cancer diagnostic prediction. This suggested technique uses the CM-DFT technique to detect missing imputation data in medical images and generates better and more efficient outcomes to reduce missing data in datasets. This proposed technique presents the CM-DFT technique for finding the missing imputation data in the proposed technique and generate to improve and efficient results to reduce the misplaced information in the medical images of datasets. The advantage of this technique is to provide a huge platform for both computing and the medical field to utilize this technique to find the data in extremely noisy images. The purpose of this research work is to investigate the misplaced imputation problem within the hybrid k-NN algorithm and to develop a new technique for using CM-DFT to impute missing values.

i. The proposed technique has been applied to the hybrid k-NN algorithm and the performance was evaluated with time-lagged correlation matrices of hybrid k-NN combined with DFT.

ii. The proposed technique provides improved results to enhance and identify the mispalced imputation data in the trained MRI datasets using this technique by the combination of LHk-NN-DFT.

iii. In order to improve the efficiency and minimize the execution time, the proposed technique provides the better results and generates the NaN values are shown in terms of x (1100–3000), Hybrid k-NN accuracy is 99.84%, LHk-NN accuracy is 0.9973, Missing Ratio is 0.9, 10, 12%. Accuracy of imputed data with less execution time is 1.533.

In future, this research will implement to lung cancer and small non-smell lung cancer and other lung diseases to identify the missing details tumor in the early phase of tumor. This current research is very useful in the society of computing engineering, neurosurgery, medical software engineering, and electronic engineering. The benefits of this technique can be implemented in the medical industry, especially in neurosurgery, physician and doctors and also it is helpful for non-medical professionals person who can use this technique to identify the tumors. The researcher also uses this technique for their further research as well.

Acknowledgements

All authors contributed equally. All authors have read and agreed to the published version of the manuscript. We would like to give special thanks to Taif University Researchers supporting project number (TURSP-2020/10), Taif University, Taif, Saudi Arabia.

Conflict of interest

The authors declare no conflict of interest

References

1. B. Suthar, H. Patel, A. Goswami, A survey: classification of imputation methods in data mining, *Int. J. Emerging Technol. Adv. Eng.*, **2** (2012), 309–312.
2. R. Houari, A. Bounceur, A. K. Tari, M. T. Kecha, Handling missing data problems with sampling methods, in *2014 International Conference on Advanced Networking Distributed Systems and Applications*, (2014), 99–104. http://doi.org/10.1109/INDS.2014.25
3. P. E. McKnight, K. M. McKnight, S. Sidani, A. J. Figueredo, *Missing Data: A Gentle Introduction*, Guilford Press, 2007.
4. O. F. Ayilara, L. Zhang, T. T. Sajobi, R. Sawatzky, E. Bohm, L. M. Lix, Impact of missing data on bias and precision when estimating the change in patient-reported outcomes from a clinical registry, *Health Qual. Life Outcomes*, **17** (2019), 106. http://doi.org/10.1186/s12955-019-1181-2

5. H. Kang, The prevention and handling of the missing data, *Korean J. Anesthesiol.*, **64** (2013), 402–409. http://doi.org/10.4097/kjae.2013.64.5.402
6. T. Donders, G. Van Der Heijden, T. Stijnen, K. Moons, A gentle introduction to imputation of missing values, *J. Clin. Epidemiol.*, **59** (2006), 1087–1091. http://doi.org/10.1016/j.jclinepi.2006.01.014
7. J. Ludbrook, Outlying observations and missing values: how should they be handled?, *Clin. Exp. Pharmacol. Physiol.*, **35** (2008), 670–678. http://doi.org/10.1111/j.1440-1681.2007.04860.x.
8. Z. Zhang, Missing values in big data research: some basic skills, *Ann. Trans. Med.*, **3** (2015), 75–84. https://doi.org/10.3978/j.issn.2305-5839.2015.12.11
9. D. Langkamp, A. Lehman, S. Lemeshow, Techniques for handling missing data in secondary analyses of large surveys, *Acad. Pediatr.*, **10** (2010), 205–210. https://doi.org/doi:10.1016/j.acap.2010.01.005
10. J. T. Shawe, N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, United Kingdom, 2004.
11. J. W. Graham, Missing data analysis: Making it work in the real world, *Ann. Rev. Psychol.*, **60** (2009), 549–576. http://doi.org/10.1146/annurev.psych.58.110405.085530
12. A. N. Baraldi, C. K. Enders, An introduction to modern missing data analyses, *J. Sch. Psychol.*, **48** (2010), 5–37. https://doi.org/10.1016/j.jsp.2009.10.001
13. I. Muhammad, Z. Yan, Supervised machine learning approaches: A survey, *ICTACT J. Soft Comput.*, **5** (2015), 115–119. http://doi.org/10.21917/ijsc.2015.0133
14. C. Nithya, V. Saravanan, A study of machine learning techniques in data mining, *Int. Sci. Refereed Res. J.*, **1** (2018), 31–38.
15. I. B. Aydilek, A. Arslan, A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm, *Inf. Sci.*, **233** (2013), 25–35. https://doi.org/10.1016/j.ins.2013.01.021
16. J. Lin, N. Li, M. A. Alam, Y. Ma, Data-driven missing data imputation in cluster monitoring system based on deep neural network, *Appl. Intell.*, **50** (2020), 860–877. http://doi.org/10.1007/s10489-019-01560-y
17. A. Choudhury, M. Kosorok, Missing data imputation for classification problems, preprint, arxiv: 2002.10709.
18. B. Al-Helali, Q. Chen, B. Xue, M. Zhang, A new imputation method based on genetic programming and weighted K-NN for symbolic regression with incomplete data, *Soft Comput.*, **25** (2021), 5993–6012. http://doi.org/10.1007/s00500-021-05590-y
19. D. Peng, M. Zou, C. Liu, J. Lu, RESI: A region-splitting imputation method for different types of missing data, *Exp. Syst. Appl.*, **168** (2021), 114425. https://doi.org/10.1016/j.eswa.2020.114425
20. J. Fan, F. Han, H. Liu, Challenges of big data analysis, *Nat. Sci. Rev.*, **1** (2014), 293–314. https://doi.org/10.1093/nsr/nwt032
21. J. Qiu, Q. Wu, G. Ding, Y. Xu, S. Feng, A survey of machine learning for big data processing, *EURASIP J. Adv. Signal Proc.*, **2016 (**2016), 67. https://doi.org/10.1186/s13634-016-0355-x
22. T. Hao, Z. Huang, L. Liang, B. Tang, Biomedical and health information processing and analysis, *Math. Biosci. Eng.*, **17** (2020), 5147–5149. http://doi.org/10.3934/mbe.2020278
23. K. Machhale, H. B. Nandpuru, V. Kapur, L. Kosta, MRI brain cancer classification using hybrid classifier (SVM-KNN), in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, (2015). https://doi.org/10.1109/SCEECS.2014.6804439

24. E. Dritsas, M. Trigka, P. Gerolymatos, S. Sioutas, Trajectory clustering and k-NN for robust privacy preserving spatiotemporal databases, *Algorithms*, **11** (2020), 207–210. https://doi.org/10.3390/a11120207
25. L. Beretta, A. Santaniello, Nearest neighbour imputation algorithms: a critical evaluation, in *The 5th Translational Bioinformatics Conference (TBC 2015)*, (2015), 1–7. http://doi.org/10.1186/s12911-016-0318-z
26. J. Derrac, F. Chiclana, S. Garcia, F. Herrera, An interval-valued k-nearest neighbor's classifier, in *16th World Congress of the International Fuzzy Systems Association (IFSA)*, (2015), 378–384. https://doi.org/10.2991/ifsa-eusflat-15.2015.55
27. R. Armina, A. Zain, N. Ali, R. Sallehuddin, A review on missing value estimation using imputation algorithm, in *International Conference on Materials Physics and Mechanics*, (2017), 892–895. http://doi.org/10.1088/1742-6596/892/1/012004
28. H. Nezamabadi-pour, E. Kabir, Concept learning by fuzzy k-NN classification and relevance feedback for efficient image retrieval, *Exp. Syst. Appl.,* **36** (2009), 5948–5954. https://doi.org/10.1016/j.eswa.2008.07.008
29. A. L. Green, A. Arnaud, J. Batiller, S. Eljamel, J. Gauld, P. Jones, et al., A multi-centre, prospective, randomized controlled study to evaluate the use of a fibrin sealant as an adjunct to sutured dural repair, *Br. J. Neurosurg.*, **29** (2015), 11–17. http://doi.org/10.3109/02688697.2014.948808
30. W. I. Schievink, M. M. Maya, S. Jean-Pierre, M. Nuño, R.S. Prasad, F. G. Moser, A classification system of spontaneous spinal CSF leaks, *Neurology*, **87** (2016), 673–679. https://doi.org/10.1212/WNL.0000000000002986
31. S. Chowdhary, S. Damlo, M. C. Chamberlain, Cerebrospinal fluid dissemination and neoplastic meningitis in primary brain tumors, *J. Moffitt Cancer Center***,** **24** (2017), 1–16. https://doi.org/10.1177/107327481702400118
32. B. Srinivas, R. G. Sasibhushana, A hybrid CNN-KNN model for MRI brain tumor classification, *Int. J. Recent Technol. Eng.,* **8** (2019), 5230–5235. https://doi.org/10.35940/ijrte.B1051.078219
33. L. Po, N. Bikakis, F. Desimoni, G. Papastefanatos, Linked data visualization: Techniques, tools, and big data, *Synth. Lect. Semantic Web Theory Technol*., **10** (2020), 1–157. https://doi.org/10.2200/S00967ED1V01Y201911WBE019
34. E. Chu, Discrete and continuous Fourier transforms: analysis, applications and fast algorithms, in *Chapman and Hall/CRC*, (2008), 424–458.
35. G. Niranjana, D. Chatterjee, Security and privacy issues in biomedical AI systems and potential solutions, in *Handbook of Artificial Intelligence in Biomedical Engineering*, Apple Academic Press, 2021.
36. Y. Zhang, S. Wang, J. Yang, Z. Zhang, P. Phillips, P. Sun, et al., A comprehensive survey on fractional Fourier transform, *Fundam. Inf.*, **151** (2017), 1–48. https://doi.org/10.3233/FI-2017-1477
37. S. Saeed, A. Abdullah, N. Z. Jhanjhi, Implementation of Fourier transformation with brain cancer and CSF images, *Indian J. Sci. Technol.*, **12** (2019), 1–9. https://doi.org/10.17485/ijst/2019/v12i37/146151.
38. A. Q. Khan, I. Ahmad, H. S. Alayachi, M. S. Noorani, A. Khaliq, Discrete-time predator-prey model with flip bifurcation and chaos control, *Math. Biosci. Eng.,* **17** (2020), 5944–5960. http://doi.org/10.3934/mbe.2020317

39. Alsisi A, Eftimie R, Trucu D. Nonlocal multiscale modelling of tumour-oncolytic viruses interactions within a heterogeneous fibrous/non-fibrous extracellular matrix, preprint, arXiv:2112.09732.
40. S. Wang, M. Celebi, Y. Zhang, X. Yu, S. Lu, X. Yao, et al., Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects, *Inf. Fusion*, **76** (2021), 376–421. https://doi.org/10.1016/j.inffus.2021.07.001
41. Y. Zhang, Z. Dong, S. Wang, X. Yu, X. Yao, Q. Zhou, et al., Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation, *Inf. Fusion*, **64** (2020), 149–187. https://doi.org/10.1016/j.inffus.2020.07.006
42. D. G. Dansereau, B. Girod, G. Wetzstein, Supplementary material LiFF: Light field features in scale and depth, in *Proceeding: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019). https://doi.org/10.1109/CVPR.2019.00823
43. A. Silva, *Light field processor: a lytro illum imaging application*, Ph.D thesis, University Institute of Lisbon, 2016.