International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Jul-24)

http://dx.doi.org/10.12785/ijcds/160111

Neutrosophic Clustering: A Solution for Handling Indeterminacy in Medical Image Analysis

Sitikantha Mallik[1], Suneeta Mohanty[2] and Bhabani Shankar Mishra[3]

[1,2,3]*School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India*

Received 17 May. 2023, Revised 11 Apr. 2024, Accepted 16 Apr. 2024, Published 1 Jul. 2024

Abstract: The need for additional innovation in the healthcare industry has become more apparent as the world begins to recover from the ravages of the pandemic. While computational intelligence has quietly become integrated into more and more fields, its applications were not something the average person discussed until recently. Computational Intelligence is becoming more and more applicable in several sectors around the world like health, industrial, business and commercial sectors. A.I.'s ability to provide faster and improved functionality is what healthcare workers at healthcare centers believe will be a significant implication in the strife towards improving healthcare delivery and patient care. One of the major applications of A.I. in healthcare is pattern mapping of medical images which mainly involves image processing. It seeks to extract significant things from the image through clustering. Therefore, choosing a suitable clustering method for a specific data set is a crucial step in the process of image segmentation. Numerous modifications to the clustering algorithm, such as the fuzzy k-mean algorithm, have been presented up to this point. All of the data mining techniques currently in use are capable of handling the uncertainty brought on by numerical deviations or unpredictable phenomena in the natural world. But, present data mining challenges in the real world may include indeterminacy components. In this article, we propose a new clustering approach for the segmentation of dental X-ray images that is based on neutrosophic logic. The authentic dental patients' dataset from KIDS(Kalinga Institute of Dental Science) Hospital is used to validate the proposed approach. The experimental findings demonstrated the proposed method's superiority in terms of clustering quality over the existing ones.

Keywords: Image processing, Uncertainty, Silhouette coefficient, Neutrosophic logic, Fuzzy k-means, Indeterminacy

1. Introduction

A.I.(Artificial Intelligence) is handling everything from working on projects and emails to answering frequently asked questions and educating us for the upcoming stages of technological progress. The adoption of telemedicine, virtual care solutions, and electronic health records created using A.I. is gaining prominence across the globe. Medical providers are trying to leverage these digital transformation journeys. Despite all medical advancements, global cases of deadliest diseases like cancer, and strokes continue to rise. The word "cancer" still carries a death sentence for the majority of people despite decades of research and scientific advancements. Deadly conditions like cancer are the result of a multi-stage process that turns benign or normal cells into malignant ones. This means that catching cancer early gives patients the highest chance of surviving. Medical practitioners state that if the condition is discovered during the first stage, there are greater percent chances of cure. However, the chances of survival are lesser if the disease is detected at later stages. In India, the most common reason for cancer-related mortality is late diagnosis. To make matters worse, the majority of patients have cancers that are already advanced when they are diagnosed. Low patient survival and high fatality rates are a result of delayed detection [1].

Cancer does not have to be a death sentence since cancer-related mortality can be significantly decreased by early detection of the disease and routine screenings for the disease. Early disease diagnosis, according to WHO (World Health Organisation), concentrates on symptomatic patients as soon as possible so that treatment can start right away. But screening is distinguishable. The goal of screening is to identify warning signs in persons who appear to be healthy so that prompt treatment can be given to stop the spread of the disease. A.I. has the potential to revolutionize healthcare by providing a faster and more efficient diagnosis of medical conditions for patients. It can also enhance diagnostic capabilities which are not possible through the doctors' naked eyes and help make healthcare more accurate to people for efficient treatment. A.I. in healthcare has the ability to increase healthcare access, lower costs, and better clinical care. Modern A.I. based medical scans can assist in discovering aberrant body parts that the conventional methods might miss. Cluster analysis is a crucial topic of study in pattern recognition and machine learning that aids

E-mail address:sitikanthamallik@gmail.com, smohantyfcs@kiit.ac.in, bsmishrafcs@kiit.ac.in

in the comprehension of a data structure for subsequent applications.

In most cases, the clustering process involves dividing the data into various groupings where there is a high degree of similarity within a cluster and of dissimilarity between clusters. In contrast to how distinct data objects are categorised into different groups, an aggregation of related data pieces is referred to as a cluster and these data pieces are placed into the same group [2]. The concept of neutrosophy introduces an entity into the set as a consideration. Smarandache contributed indeterminacy as a brand-new concept to the fuzzy set. As a result, the neutrosophic set can be defined as an ordered triple N = (T, I, F), where T denotes the degree of truthiness, F denotes the degree of falsity, and I is the degree of indeterminacy. Also referred to as neutrosophic components here are T, I, and F.[3]

The related works for the dental segmentation from X-ray pictures can be divided into two primary steps i.e. image processing and machine learning approaches. Segmentation, which separates a medical image into sections for more research on disorders, is one of the most important tasks in medical imaging. Many clustering algorithms have been created with the intention of segmenting medical data. The k-means clustering approach is the first technique. The simplicity and low cost of this method are its benefits, but it has the disadvantages of being unable to deal with inconsistent and indeterminate values [4]. Fuzzy k-means clustering, which can be used for diagnostic medical objectives, is the second technique [5]. In standard k-means, the membership values are either zero or one, and cannot take values between 0 and 1, unlike fuzzy k-means. Fuzzy relations have some drawbacks, including an inability to manage inconsistent data and a high processing cost when running a nonlinear algorithm. The final technique, called neutrosophic clustering, organises the data into groups and identifies the pixels that belong to each group. In order to compute the similarity score, it turns each pixel of the input image into a set of neutrosophic values.

Specifically, the contributions of the paper are summarized as follows:

1) Propose a recommendation system for early dental disease detection using neutrosophic method by classifying malignant and benign tumors
2) Model the medical image segmentation problem in the form of neutrosophic clustering
3) Apply neutrosophic clustering to segment dental X-rays belonging to genuine dataset of patients from KIDS Hospital during the period 2022–2023
4) Suggest the most appropriate clustering indices that should be opted for the algorithm
5) Conduct several comparisons of our approach with other ML classifiers and Deep Learning approaches using clustering indices

The rest of the paper is organized as follows: Section 2 illustrates the role of neutrosophic logic in medical imaging. Section 3 describes about the overview of different clustering methods used in image segmentation and their algorithms. Section 4 presents different clustering indices used for measuring accuracy. Section 5 implies about experimental result analysis among different clustering methods. Section 6 contains the conclusion and future work.

2. Neutrosophic Logic In Medical Imaging

Neutrosophic logic is a generalization of classical logic that allows for degrees of truth, falsity, and indeterminacy to coexist in a proposition. Because of issues including image noise, artefacts, anatomical variability, uncertainty, and imprecision are frequently present in medical imaging, making this technique well-suited to handle them [6]. In the context of medical imaging, neutrosophic logic can be used in various ways. Here are a few examples:

A. *Image segmentation*

Neutrosophic logic can be used to perform image segmentation, which involves separating an image into different regions that correspond to different anatomical structures or pathological conditions. Neutrosophic segmentation methods can handle uncertainties in the image data and produce more accurate and robust results compared to classical segmentation methods [7].

B. *Image fusion*

Medical images from various approaches, such as CT/CAT Scan, Ultrasound scan, and magnetic resonance imaging (MRI) scan, can be combined using neutrosophic logic. The distinct qualities and uncertainties of each modality can be taken into account via neutrosophic fusion approaches, which can result in images that are more thorough and informative [8].

C. *Diagnosis and decision-making*

Neutrosophic logic can be used to model the uncertainty and imprecision in medical diagnosis and decision-making. For example, a neutrosophic expert system can take into account multiple sources of information, such as patient history, laboratory results, and imaging findings, and provide a more comprehensive and accurate diagnosis or treatment recommendation.

Overall, neutrosophic logic has the promise of enhancing the precision and dependability of medical imaging analysis and interpretation as well as assisting physicians in making more educated decisions based on the information at hand.

3. Methodology/ Image Segmentation By Clustering Methods

Image segmentation is the division of a picture into several segments to transform the representation of an image into something more relevant and understandable. The technique of segmenting an image is frequently required and should be considered a crucial stage in the processing and analysis of an image. For this study field to advance, repeatable studies with documented benchmarks are necessary. The process of selecting an effective segmentation model is challenging, and the model has superior segmentation with faster computation. The issue is reformatted to save calculation time while maintaining high-quality outcomes [9]. The toughest unsupervised learning challenge is clustering, which, similar to all other issues in this category, entails finding a pattern in a group of unlabeled data. The process of arranging things into groups whose constituents are associated in some way is known as clustering. A set of things that have similar characteristics to one another and are dissimilar to those found in other clusters are thus considered to be in a cluster [10].

A. Types of k-means clustering algorithm

There are mainly three types of k-means clustering algorithms: standard k-means, fuzzy k-means, and hierarchical k-means.

1) Standard k-means clustering

The most fundamental and often used form of k-means clustering is called standard k-means clustering. It is a partition-based clustering technique where focus is to partition a set of objects into clusters in such a way that the sum of squared distances between data points and their nearest centroid remains minimum. Each data point in a standard k-means cluster is determined by the distance to the closest centroid.

2) Fuzzy k-means clustering

Fuzzy k-means clustering is a variant of k-means that allows for the soft assignment of data points to clusters. In fuzzy k-means, each data point is assigned a degree of membership to each cluster based on the distance to the centroids of all clusters. The degree of membership indicates the likelihood of a data point belonging to each cluster, allowing for overlapping clusters.

3) Hierarchical k-means clustering

A hierarchical clustering approach is called hierarchical k-means clustering when it performs k-means clustering iteratively over data partitions until the necessary number of clusters is reached or a stopping criterion is reached. In hierarchical k-means, the initial partition is often the result of a standard k-means clustering, and the partitions are merged based on their proximity [11]. There are also other variants of k-means clustering, such as density-based, spherical, and kernel k-means, which incorporate additional constraints or assumptions on the data distribution or shape. The features of the data and the particular clustering issue at hand determine whether a variant of the k-means method is used.

B. K-means Clustering

One of the efficient unsupervised learning methods to address the prominent clustering issue is K-means. The technique classifies a predefined data set using a predetermined number of fixed-a priori clusters (let's use k clusters as an example). Since the sum of squares error or intra-cluster variance is a measurement of error, the goal of k-means is to reduce this value.

The algorithm's stages are described:

1) Place K pointers in the region where the objects that are being grouped, are located. These points show the first set of centroids.
2) Each item should be positioned in the group according to its nearest centroid.
3) Recompute the coordinates of the K centroids after all the entities have been allocated. Replicate the second and third steps only after centroids cease migrating. In order to calculate the parameter that needs to be minimized, the items are then separated into groups.

C. Fuzzy K-Means Clustering

Fuzzy clustering method (FCM) uses membership degrees between zero and one rather than making clear allocations of the data to clusters and allows the data points to belong to several clusters. FCM which was introduced by James Bezdek in 1981, is the most popular clustering algorithm. It is commonly used in pattern recognition and is based on the minimization of an objective function.[12]

The clustering algorithm which has been extended using fuzzy k-means clustering, allows different data points to be part of numerous clusters with different membership levels. It is a form of the k-means method that enables the depiction of ambiguity and uncertainty in data. Each data point in fuzzy k-means clustering is given a membership grade for each cluster based on its similarity to the centroid of that cluster. The membership grade indicates how much the data item is a part of the cluster. For each data point across all clusters, the sum of the membership grades is 1.

The fuzzy k-means technique is composed of the following steps:

1) Define the number of clusters (k) and the fuzzy exponent (m).
2) Randomly initialize the centroids for each cluster.
3) Calculate the membership grades for each data point for each cluster using the following formula:

$$c_j = \frac{\sum_{i=1}^{n} \left(u_{ij}\right)^m x_i}{\sum_{i=1}^{n} \left(u_{ij}\right)^m} \tag{1}$$

Where, u_{ij} is the membership grade of data point i for cluster j

4) Update the centroids for each cluster using the following formula:

$$u_{ij} = \left(\frac{1}{\sum_{k=1}^{k} \left(\frac{d_{ij}}{d_{ik}}\right)^{2/m-1}} \right)^{-1} \tag{2}$$

where, d_{ij} is the Euclidean distance between data point i and cluster centroid j

d_{ik} is the Euclidean distance between data point i and cluster centroid k, and m is the fuzzy exponent

c_j is the new centroid for cluster j, x_i is the data point i, and n is the number of data points

5) Repeat steps 3 and 4 until convergence or a maximum number of iterations is reached.
6) Output the clusters and their membership grades for each data point.

Fuzzy k-means clustering is used in various functions, such as data analysis, pattern recognition, segmentation of medical images, and data mining. It allows for more flexible clustering than standard k-means, especially in cases where data points may belong to multiple clusters or where the boundaries between clusters are ambiguous.

D. Neutrosophic k-means Clustering

Since the classic k-means clustering algorithm has been extended to accommodate ambiguous and incomplete data, therefore it's known as neutrosophic k-means clustering. Here, each data point is represented by three values instead of only one in standard k-means clustering.First part refers to the degree of belongingness of members in a cluster membership, 2nd part refers to the non-belongingness of members in a cluster and the final part is indeterminacy.

The degree of membership indicates how much a data point is part of a cluster, whereas the degree of non-membership indicates how a data point is not part of a cluster. The degree of indeterminacy is a measure of how uncertain or ambiguous a data point is.[13]

The neutrosophic k-means clustering algorithm works by first initializing the cluster centroids, then iteratively updating the degree of membership, degree of non-membership, and degree of indeterminacy for each data point based on its distance to each cluster centroid. The cluster centroids are then updated based on the updated degrees of membership, non-membership, and indeterminacy for each data point.

The basic idea is to allocate each data point to a cluster based on its degree of membership in each cluster, which can range between completely belonging to a cluster and not belonging to it, with an intermediate degree of indeterminacy.

Here are the steps of the neutrosophic k-means clustering algorithm:

1) Initialization: Randomly select k initial centroids for the clusters.
2) Membership assignment: For each data point x, calculate its degree of membership in each cluster u_i , $\bar{u}_i$ & $\tilde{u}_i$, which correspond to the degree of truth, indeterminacy, and falsity, respectively. These degrees are calculated based on the distance of x to each centroid using the neutrosophic distance measure.
3) Centroid update: Calculate the new centroids for each cluster based on the arithmetic mean of weights of the data points, where each weight refers to the degree of membership.
4) Iteration: Repeat steps 2 and 3 until convergence, i.e., until the centroids no longer change significantly.
5) Result: The data points with the highest degree of membership in each cluster define the final clusters.

The neutrosophic k-means clustering algorithm can be applied to various types of data, such as numerical, categorical, or mixed data, and can handle missing or incomplete data. It has been used in various applications, such as image segmentation, text clustering, and medical diagnosis. However, it requires careful selection of the neutrosophic distance measure and tuning of the number of clusters, which can be challenging in practice.

4. Clustering Evaluation Metrics

A validity index plays a crucial part in clustering. It aids in determining the right number of clusters to include in a data set. Clustering indices are measures utilised to evaluate the performance of the clustered results obtained by a clustering algorithm. These indices provide quantitative metrics for comparing different clustering algorithms or different parameter settings of the same algorithm.

Here are some commonly used clustering indices:

A. PBM-index

The PBM index is computed from the distances between points and their barycenters as well as the distances between the individual barycenters, which is an acronym made up of three authors' initials, Pakhira, Bandyopadhyay, and Maulik. The PBM index is described as follows:

$$PBM(K) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K\right)^2 \tag{3}$$

$$E_K = \sum_{k=1}^{K} E_k \tag{4}$$

such that

$$E_k = \sum_{j=1}^{n} u_{kj} \left\| \mathbf{x}_j - \mathbf{z}_k \right\| \tag{5}$$

and

$$D_K = \max_{i,j=1}^{K} \left\| \mathbf{z}_i - \mathbf{z}_j \right\| . \tag{6}$$

Where:

- E_k is the total distance from each cluster's points to their barycenter
- E_1 is the total distance from each point in the data set to its barycenter
- k is the number of clusters

The PBM (Probabilistic Binary Matrix) index, whose greatest value denotes the proper partitioning, is used to link a measure to several dataset partitions. As a result, if k, the cluster count, is modified within a certain span & the data is partitioned using an underlying clustering technique, the appropriate number of clusters present in the data can be calculated. Number of clusters is determined by the value of k, which matches the highest value of the PBM index, as seen in [12, eq. (3)]. PBM index is a clustering evaluation metric that is used to measure the similarity between two partitions obtained by clustering algorithms [14]. It is based on the probabilistic interpretation of Rand index. Rand index mainly compares the number of collaborations and differences between the clusters in the two partitions to determine how similar the two partitions are.

The PBM index takes into account the probability of assigning an object to a cluster, which makes it more suitable for evaluating partitions obtained by probabilistic clustering algorithms. It is more robust than other clustering evaluation metrics, such as the Rand index and the Jaccard index, in noisy and overlapping clustering scenarios. [15]

The PBM index has been used in various applications, such as image segmentation, document clustering, and community detection in social networks. However, it may not be suitable for evaluating partitions obtained by non-probabilistic clustering algorithms, such as k-means and hierarchical clustering.

B. *Silhouette Score*

The term "silhouette" describes a technique for analyzing and verifying consistency among data clusters. Analysis of silhouettes could be performed to investigate the separation distance between clusters created by the algorithm. The average silhouette coefficient applied to all the samples is indicated by the silhouette score.

$$Silhouette\ Score = \frac{\text{b - a}}{\text{max (b, a)}} \tag{7}$$

where, b = average intra-cluster distance
a = average inter-cluster distance

The silhouette coefficient is calculated using the average of the distances between the nearest and intra-clusters for all samples,explained in Figure 1. The range of the silhouette coefficient is [-1,1]. If there is more separation between clusters, the higher the silhouette coefficients (the nearer to +1) and values close to 0 indicate overlapping or poorly separated clusters. The sample is closest to or on the decision boundary between the two adjacent clusters if the value is zero, whereas samples with a value of negative one may have been accidentally assigned to another cluster [16].

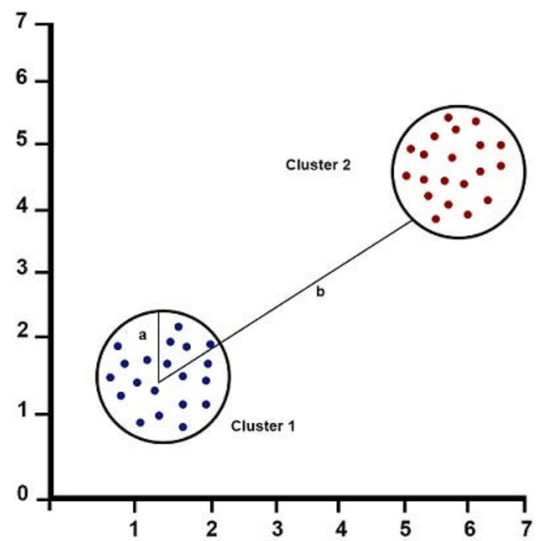

Figure 1. Silhouette Score

C. *IVF Index*

With the use of vector similarity search, we can quickly and accurately browse through a wide variety of media, including articles and GIFs, in only a few seconds. By storing vector representations of the data we want to search in an index, vector similarity search can find matches between sets of data. Clustering is used to narrow the search window for the Inverted File Index (IVF) [17]. The IVF (Inverted File) clustering algorithm is a technique used for large-scale image retrieval. It involves indexing features of images and then clustering them for fast and efficient retrieval. It is described as follows:

$$IFV = \frac{1}{C}\sum_{j=1}^{C}\left\{\frac{1}{N}\sum_{k=1}^{N}\mu_{kj}^{2}\left[\log_2 c - \frac{1}{N}\sum_{k=1}^{N}\log_2 \mu_{kj}\right]^2\right\}\frac{SD_{\max}}{\overline{\sigma_D}} \tag{8}$$

Where:

- μ_{kj}^2 = The membership of data point in the cluster
- N = Number of data points (records)
- c = Number of clusters
- $SD_{\max}$ = Maximal distance between centers
- $\overline{\sigma_D}$ = Even deviation between each object and the cluster center

Inverted file indexes are useful in clustering because they allow for efficient retrieval of documents that contain specific terms, which can aid in the process of selecting features or terms for use in the clustering algorithm. Because of its simplicity, great search quality, and manageable search rate, it is a widely popular index. Due to their stability and resilience, IFVs are frequently utilised in fuzzy clustering for geographical data [18]. The dataset is supposed to produce the most ideal results when IFV is at its maximum value.

D. Davies Bouldin index

DB (Davies Bouldin) index, normalized by the sum of cluster diameters, calculates the average similarity between each cluster and its most similar cluster. DB Index is built on the premise of the with-cluster and between cluster distances. The number of clusters the data points should be labelled in is usually chosen using this method. Therefore, lowering the DB index is the key goal. The result is determined by averaging the distances between each cluster and its closest neighbour, where proximity is measured by the ratio of within-cluster to between-cluster lengths. Consequently, clusters with more space between them and less dispersion will receive higher scores. Lower DB Index numbers imply better grouping with a minimum score of 0.

$$\delta_k = \frac{1}{n_k} \sum_{i \in I_k} \left\| M_i^{\{k\}} - G^{\{k\}} \right\| \tag{9}$$

denoted by

$$\Delta_{kk'} = d\left(G^{\{k\}}, G^{\{k'\}}\right) = \left\| G^{\{k'\}} - G^{\{k\}} \right\| \tag{10}$$

Where: δ_k = the mean distance of the points belonging to cluster C_k to their barycenter, G_k

$\Delta_{kk'}$ = the distance between the barycenters $G^{\{k\}}$ and $G^{\{k'\}}$ of clusters C_k and C_k'

The mean value of all the clusters for the quantities, M_k is calculated in DB index as seen in (11). DB Index is denoted by C. [19]

$$C = \frac{1}{K} \sum_{k=1}^{K} M_k = \frac{1}{K} \sum_{k=1}^{K} \max_{k' \neq k} \left(\frac{\delta_k + \delta_{k'}}{\Delta_{kk'}} \right) \tag{11}$$

E. Contrast to Noise Ratio

In medical imaging, Contrast to Noise Ratio (CNR) measures the contrast between the background and the target tissue (i.e., the neighbouring tissue). ROI refers to region of interest. The Signal to Noise Ratio (SNR) measures how well the picture signal stands out against the background in a particular area. The size and contrast of the objects affect our capacity to distinguish them from a noisy background. ROI will choose every pixel in the chosen area of the image. The picture viewing program will next normally present at least a few significant metrics inside each ROI. the ROI's average picture signal as well as its noise standard deviation.[20]

We can determine the SNR and CNR for a given dataset by measuring the signal level and noise level in certain ROIs.

$$SNR = \frac{\text{Avg pixel values in Signal ROI}}{\text{Std Background ROI}} \tag{12}$$

$$CNR = \frac{\text{Avg Signal ROI} - \text{Avg Background ROI}}{\text{Std Background ROI}} \tag{13}$$

It is apparent that the CNR differs from the SNR and depends heavily on the local contrast. The objects will be easier to visualize in relation to the background as the CNR is raised. The measurement of the imaging signal in a tissue area in relation to the background tissue refers to the Signal to Noise Ratio (SNR). The contrast between the typical image values in an interest tissue and the background in an MRI image refers to the contrast to noise ratio (CNR) (i.e., the surrounding tissue). CNR is frequently used instead of SNR in tomographic imaging, such as CT and MRI because it is simpler to immediately identify the tissue of interest.

5. Experimental Results Analysis And Discussion

Since there's no properly accepted algorithm for segmenting photographs, it can be difficult to assess the concept of image segmentation. MATLAB has been used to test the proposed image processing system for the identification of dental infections. MATLAB was used to carry out the steps of image processing such as preprocessing, segmentation, feature extraction, and clustering. To extract these characteristics, the ROI is segmented after the picture has been transformed to greyscale. The final clusters are produced using additional techniques. To distinguish between normal and malignant images, the medical images are clustered into different groups based on retrieved features.

A. Dataset

In order to verify the clustering performance, the experiment is conducted using the dataset consisting of dental affected patients. This dataset includes 1,000 unidentified and anonymous panoramic dental X-rays from the Kalinga Institute of Dental Studies in Bhubaneswar. The topics span a broad spectrum of dental disorders, including both healthy and partially and completely edentulous patients. These datasets were used to evaluate how well the clustering technique performed. The linked dataset contained 1000 patients aged 18 years and over who had done dental X-rays within six months of diagnosis during the period from 1 January 2022 to 31 December 2022. The datasets include patients' profiles containing sex, age, dental X-rays, medications, appointments, clinic visits, and other patient data. The dataset's objective is to diagnostically determine whether a patient has dental issues or not so that patients can get earlier treatment. These instances were chosen from a bigger database under several restrictions. During testing, a training set which is made up of 80 percent of the dataset and a test set which is made up of 20 percent of the dataset, are chosen at random.

B. Experimental Setup

The K-means, fuzzy clustering, and neutrosophic clustering methods are implemented in Matlab 2021a programming language performed on a PC with Intel Core i7 processor, 8GB RAM, and the operating system used here is Windows 10 (64 bits).

C. Parameters

Four evaluation criteria are used to compare the effectiveness of clustering algorithms. They are as follows: performance is assessed using Davies Bouldin Index (DB), Visibility Metric (VM), Silhouette Coefficient (SC), and Inverted File Index (IVF) index. The likelihood that a viewer would be able to distinguish differences between two images is determined by visibility metrics (VM), which are then calculated to determine the Contrast-to-Noise Ratio (CNR) of the noise image. Better image segmentation is correlated with higher levels of SC, VM, and IVF. However, segmentation is considered better if DB Index values are lower.[21]

D. Comparison among clustering methods

To demonstrate the effectiveness of the suggested strategy for data clustering, we demonstrated three segmentation algorithms in terms of validity indices based on the aforementioned experimental setup. The initialised parameters are c = 3, $\alpha = 0.9$, $\delta = 0.9$ and $\gamma = 0.95$. Using methods like standard k-means, fuzzy k-means, and neutrosophic k-means, the experiment was run on five randomly selected dental X-ray medical photos from the dataset. The results are expressed in Figure 2 along with the essential numerical bounds, and the insights are enumerated in Table I.

In Table II, the median of the results is displayed. If IVF, VM, and SC values are higher and the DB value is lower, then that segmentation strategy is preferable. It is highly desirable if the silhouette coefficient is also positive. The average performance analysis chart in Figure 2 shows that the neutrosophic k-means clustering has a lower DB value than the others and has higher VM, and IVF values. Figure 3 provides a full analysis of the Silhouette Score statistical parameter.We computed the silhouette score for each data point in a clustering solution with several clusters and then took the average of these scores to determine the average silhouette score.

The time evaluation of the suggested algorithms is shown in Table III. Figure 4 predicts how long each

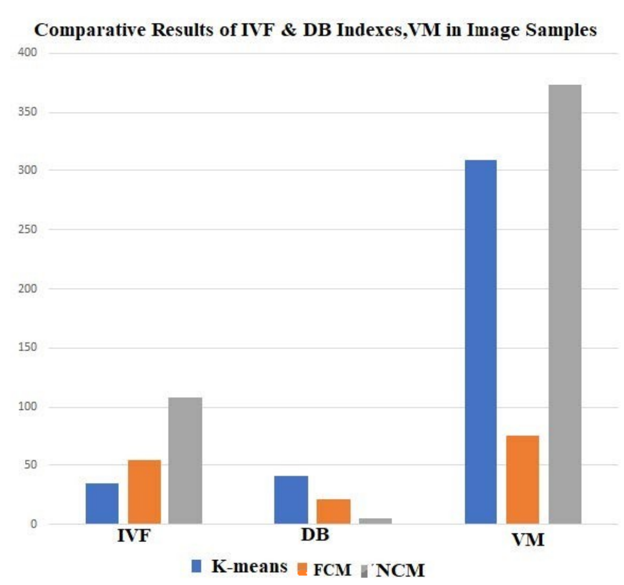

Figure 2. Performance Analysis Chart

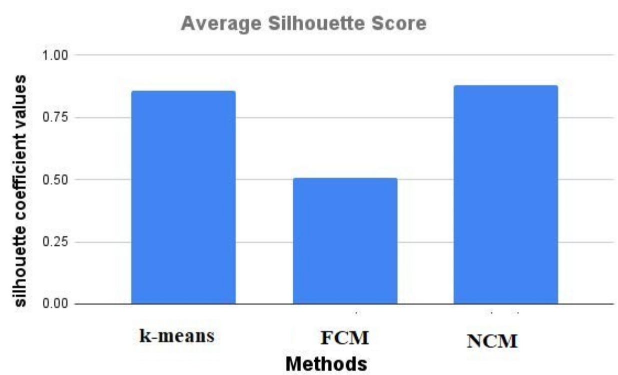

Figure 3. Silhouette Score Chart

procedure will take on average. It illustrates that, in contrast to other methods, the K-means method takes the least time but produces subpar data. When compared to K-means, the Fuzzy clustering method(FCM) yields good results, but it takes longer. When compared to the fuzzy clustering method, the neutrosophic clustering method takes a lot less time on average to process five images. The computational results demonstrate that the neutrosophic clustering technique outperformed the alternative methods in terms of clustering performance. Therefore, it was found that the neutrosophic method outperforms other approaches.

In this study, cluster-based methods, or the unsupervised approach, were suggested for picture segmentation. Different photos were examined using different clustering techniques such as k-means, fuzzy k-means, and neutrosophic k-means. The segmentation parameters DB, SC, PBM, CNR, and IVF are used to assess how well the suggested algorithms perform. The experimental results

TABLE I. Performance Evaluation

Sl.no.	Methods	IVF	DB	SC	VM
Image1	K-means	35.79	31.049	0.85	338
	FCM	47.05	22.8	0.126	79.113
	NCM	112.36	0.367	0.927	379.11
Image2	K-means	30.47	43.34	0.87	308
	FCM	83.67	19.86	0.5401	79.116
	NCM	110.78	0.4268	0.6755	371.084
Image3	K-means	37.89	41.814	0.88034	294.4
	FCM	48.07	25.846	0.8676	72.972
	NCM	105.31	0.286	0.9735	372.97
Image4	K-means	32.61	40.016	0.818	308.4
	FCM	48.79	18.062	0.8176	75.47
	NCM	106.37	0.246	0.96	375.471
Image5	K-means	37.61	46.497	0.695	294.53
	FCM	44.59	20.8	0.1887	71.02
	NCM	107.38	0.881	0.873	371.02

TABLE II. Average Calculation of Clustering Indices

Sl. No.	Method	Clustering Indices			
		IVF	DB	SC	VM
1	K-means	34.874	40.54	0.822	308.66
2	FCM	54.434	21.47	0.508	75.5382
3	NCM	108.44	0.441	0.881	373.931

TABLE III. Comparative results in all samples

Sl. No.	Image	K-means	NCM	FCM
1	Image1	10.92	12.837	19.67
2	Image2	9.61	11.63	16.9
3	Image3	10.769	11.608	17.55
4	Image4	11.41	15.97	17.43
5	Image5	10.503	12.595	14.17

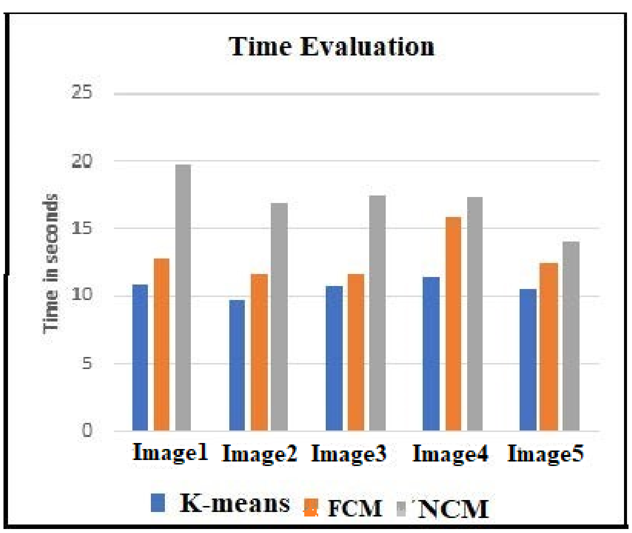

Figure 4. Average Time Evaluation of all methods

demonstrated that visual segmentation using k-means yields poor results despite consuming less time. In comparison to k means, the modified variant k means algorithm requires fewer iterations. The neutrosophic method is faster and yields more accurate results than the traditional FCM, which requires comparatively more time. Therefore, it is concluded that the proposed algorithm i.e. neutrosophic algorithm worked well, according to the computational results.

6. Conclusions And Future Work

In this study, we focused on the neutrosophic clustering method for segmenting medical images. Neutosophic logic, clustering, indeterminacy, medical applications of neutrosophic logic, clustering indices,and orthogonal neutrosophic matrix are among theoretical contributions. Neutrosophic sets which are an extension of intuitionistic fuzzy sets, become a noticeable subject in computer vision and image processing.These concepts are useful for building a new clustering method that converts an image into the neutrosophic domain and determines how similar pixels expressed in neutrosophic elements are to one another. Neutrosophic sets are used in other applications such as segmentation, noise reduction, and image retrieval. Neutrosophic usage in medical images brings forward improvement in accurate diagnosis through more accurate segmentation. Even though neutrosophic is utilized for visual restoration and segmentation, the input images are typically grayscale or medical images. The interesting thing is that there aren't many studies that consider neutrosophic logic while segmenting or restoring images. Therefore, colour image restoration and segmentation are novel techniques in a neutrosophic context.

According to experimental results done on X-ray dental records from the hospital, the suggested approach surpasses the pertinent fuzzy clustering method. It has been shown that the suggested method outperforms fuzzy k-means in terms of DB, SC, and VM validity index values. The results of this research indicate a number of potential avenues for further research (i) applying different types of nuetrosophic sets,(ii) reducing the processing time,(iii) segmentation of 3D medical images and (iv) proposing hybrid method of neutrosophic logic with other deep learning techniques.

References

[1] T. Gaber, G. Ismail, A. Anter, M. Soliman, M. Ali, N. Semary, A. E. Hassanien, and V. Snasel, "Thermogram breast cancer prediction approach based on neutrosophic sets and fuzzy c-means algorithm," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4254–4257, 2015.

[2] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, "Advances in knowledge discovery and data mining," *American Association for Artificial Intelligence*, 1996.

[3] F. Smarandache, "A unifying field in logics : neutrosophic logic : neutrosophy, neutrosophic set, neutrosophic probability . infinite study ." *American Research Press*, 1999.

[4] H. Hu, J. Liu, X. Zhang, and M. Fang, "An effective and adaptable k-means algorithm for big data cluster analysis," *Pattern Recognition*, vol. 139, p. 109404, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003132032300105X

[5] H. Yadav, J. Singh, and A. Gosain, "Experimental analysis of fuzzy clustering techniques for outlier detection," *Procedia Computer Science*, vol. 218, pp. 959–968, 2023, international Conference on Machine Learning and Data Engineering. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923000765

[6] A. Ansari, R. Biswas, and S. Aggarwal, "Neutrosophic classifier: An extension of fuzzy classifer," *Applied Soft Computing*, vol. 13, no. 1, pp. 563–573, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494612003419

[7] D. Ko and B. Sharma, "Challenges and future directions in neutrosophic set-based medical image analysis," pp. 313–343, 01 2019.

[8] A. R. Hawas, A. S. Ashour, and Y. Guo, "8 - neutrosophic set in medical image clustering," *Neutrosophic Set in Medical Image Analysis*, pp. 167–187, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128181485000084

[9] K. K. Singh and A. Singh, "A study of image segmentation algorithms for different types of images different types of images," *Int. J. Comput. Sci*, no. 7, p. 5, 2012.

[10] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, p. 264–323, sep 1999. [Online]. Available: https://doi.org/10.1145/331499.331504

[11] R. T. Ng and J. Han, "Efficient and effective clustering methods for spatial data mining," *Very Large Data Bases Conference*, 1994. [Online]. Available: https://api.semanticscholar.org/CorpusID:13136810

[12] M. G. C. A. Cimino, B. Lazzerini, and F. Marcelloni, "A novel approach to fuzzy clustering based on a dissimilarity relation extracted from data using a TS system," *Pattern Recognition*, vol. 39, no. 11, pp. 2077–2091, Jan 2006.

[13] S. Mallik, S. Mohanty, and B. S. Mishra, "Neutrosophic logic and its scientific applications," *Biologically Inspired Techniques in Many Criteria Decision Making*, pp. 415–432, 2022.

[14] M. K. Pakhira and A. Dutta, "Finding number of clusters using vat image, pbm index and genetic algorithms," *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia*, p. 217–221, 2011. [Online]. Available: https://doi.org/10.1145/1963564.1963601

[15] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, pp. 487–501, 2004. [Online]. Available: https://api.semanticscholar.org/CorpusID:36291827

[16] K. Aoyama, K. Saito, and T. Ikeda, "Inverted-file k-means clustering: Performance analysis," 2020.

[17] L. H. Son, B. C. Cuong, P. L. Lanzi, and N. T. Thong, "A novel intuitionistic fuzzy clustering method for geo-demographic analysis," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9848–9859, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417412004307

[18] B. Desgraupes, "Clustering indices," *University of Paris Ouest-Lab Modal'X*, vol. 1, no. 1, p. 34, 2013.

[19] A. Gallo-Bueno, M. Reynaud, M. Casas-Cabanas, and J. Carrasco, "Unsupervised machine learning to classify crystal structures according to their structural distortion: A case study on li-argyrodite solid-state electrolytes," *Energy and AI*, vol. 9, pp. 100–159, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666546822000180

[20] "Ecr 2011 book of abstracts - b - scientific sessions," *Insights into Imaging*, vol. 2, pp. 135–324, 2011.

[21] T. Sahoo, S. Panda, and S. R. Biswal, "Application of hierarchical clustering approach for prediction of grain size in heat-treated en9 steel," *ECS Advances*, vol. 2, no. 2, p. 021001, jun 2023.

Sitikantha Mallik is Research Scholar in Computer Science and Engineering at KIIT Deemed to be University, Odisha. He received the B.Tech. in Computer Science and Engineering from College of Engineering & Technology, Odisha in 2015, M.Tech. degree in Computer Science and Engineering from KIIT University, in 2019. He has published three research publications in reputed conferences, book chapters, and journals indexed in Scopus/SCI/Web of Science.

Dr. Suneeta Mohanty is Associate Professor in Computer Science and Engineering at KIIT Deemed to be University, Odisha. She has 14 years of teaching experience in Computer Science & Engineering. She earned her M.Tech degree in Computer Science & Engineering from College of Engineering & Technology which is a constituent college of BPUT, Odisha. She earned her Ph.D. from KIIT Deemed to be University, Odisha. She has been a life member of ISCA, ISTE, and IET. Her research interests include Cloud Computing, Network Security, WSN, and IoT. Her research contribution includes 02 co-edited books published by Springer Nature, more than 45 research publications in reputed conferences, book chapters, and journals indexed in Scopus/SCI/Web of Science.

Dr. Bhabani Shankar Prasad Mishra was born in Talcher, Odisha, India in 1981. He received the B.Tech. in Computer Science and Engineering from Biju Pattanaik Technical University, Odisha in 2003, M.Tech. degree in Computer Science and Engineering from KIIT University, in 2003, Ph.D. degree in Computer Science from F.M.University, Balasore, Odisha, India, in 2011, and Post Doc in 2013 from Soft Computing Laboratory, Yansei University, South Korea . Currently he is working as a Professor at School of Computer Engineering, KIIT University, Bhubaneswar, Odisha, India. His research interest includes Pattern Reorganisation, Remote Sensing, Data Mining, Soft Computing, Big Data and Machine Learning. He has published more than 90 research articles in reputed Journal and Conferences, one patent and has edited more than five books of current importance. Under his guidance, 20 M.Tech and 5 PhD scholars are already been awarded.