VOLUME - 12, ISSUE - 04, APRIL - 2023 • PRINT ISSN No. 2277 - 8160 • DOI : 10.36106/gjra

Original Research Paper

Mathematics

A NOVEL APPROACH TO LINEX MINIMAL SPANNING TREE USING MODIFIED LINEX INTUITIONISTIC FUZZY C-MEANS CLUSTERING INDUCED CANBERRA DISTANCE MEASURE FOR BREAST CANCER DETECTION

M. Nithya	Department of Mathematics, Mother Teresa women's University, Kodaikanal, Tamilnadu, India
Dr. K. Bhuvaneswari	Department of Mathematics, Mother Teresa women's University, Kodaikanal, Tamilnadu, India
Dr. S. Senthil*	Department of Economics and Statistics, ICDS, Collectorate, Dindigul, Tamilnadu, India. *Corresponding Author

ABSTRACT Breast cancer is one of the major causes of death among women, and early detection may decrease the aggressiveness of the disease. There are several factors known related to encourage an increased risk of breast cancer, One of the most common cancer types is breast cancer, and early diagnosis is the most important thing in its treatment. This research examines the determinant factors of breast cancer and measures the breast cancer patient data to build the useful classification model using data mining approach. In this paper, we have presented a novel approach to identify the presence of breast cancer lumps in mammograms. In this work, a fuzzy clustering algorithm is proposed based on the asymmetric loss function instead of the usual symmetric dissimilarities. Linear Exponential (LINEX) loss function is a commonly used asymmetric loss function, which is considered and LINEX Canberra measure using minimal spanning tree initialization method for selecting initial cluster centers is presented. The LINEX IFCM algorithm that indicates the spatial influence of the neighboring pixels on the center pixels plays a key role in this algorithm. Furthermore, the algorithm is extended for calculating membership and updating prototypes by minimizing the new objective function of LINEX_IFCM based intuitionistic fuzzy c-means. LIINEX Intuitionistic fuzzy c-means clustering algorithm for clear to identify of abnormalities for mammography images and Breast cancer patients symptoms used to predictive probability calculated by Pearson Chi-Square $\chi 2$ test at 0.05 significance level indicate a highly significant correlation between mammography performance and clinical symptoms of breast cancer. Finally, the developed algorithms are illustrated through conducting experiments on random dataset, partition coefficient and validation function are used to evaluate the validity of clustering also this paper compares the results of proposed method with the results of existing basic intuitionistic fuzzy c-means.

KEYWORDS : Intuitionistic fuzzy c-means, minimal spanning tree, initial cluster center, LINEX function, Canberra measure, Breast cancer, mammograms, Partition coefficient, Validation function

1. INTRODUCTION

Breast cancer is a severe risk to female health. Mammography is usually utilized in breast cancer screening due to its sensitivity to even small tumors [1-3]. Women are required to have a mammogram every year when they are above 40 years old. In developed and emerging countries, citizens shift their lifestyle from traditional to modern, increasing the incidence of breast cancers among women mainly between the ages of 35-55. It is possible to monitor the incidence of breast cancers by detecting breast cancers in their early stages Self and clinical breast checks, magnetic resonance imaging (MRI),ultrasound, and mammography are screening techniques used for breast cancer screening [4-5]. The image produced through mammography is called a mammogram, which consists of the background, breast region, fat tissue, breast masses, and microcalcifications with high intensities [6]. As the demand for mammography processing is growing, radiologists can produce errors that overlook important clues due to fatigue [7].

Nowadays, a variety of image segmentation techniques have been extensively explored for the purpose of image segmentation. In clinical diagnosis, magnetic resonance technology provides an excellent assistance to medical treatment, which has the ability to detect brain tumors [8]. In Medical image segmentation, we need to extract the necessary boundaries, surfaces in order to recognize the variation in an image from a normal image [9].

Among them Fuzzy C-means (FCM), a soft clustering process is considered as an extensively used image clustering algorithm [10]. In pixel based image segmentation, FCM is more effective with considerable amount of benefit [11]. But FCM clustering does not manage the spatial properties of an image rising to strong noise sensibility.

To handle this issue, many modified techniques [13-15] have been suggested. In these techniques, spatial information (useful property of an image) is combined with the membership function of the FCM to give new membership function. The spatial property of an image, allow to better handle the problems related to image segmentation. Atanassov [16] presented "Intuitionistic fuzzy sets" that considers vulnerability in the definition of the membership value. Different approaches have been designed with the Intuitionistic theory for the process of image segmentation. Chaira [17] has given a novel Intuitionistic fuzzy c-means (IFCM) clustering algorithm for the process of partitioning an image into segments. It includes hesitation degree and a new parameter termed as intuitionistic fuzzy entropy (IFE) in the objective function. Although this strategy utilizes the irregular initialization of cluster prototype which produces inaccurate results when chosen randomly.

This method does not include the local information and the spatial information, hence is sensitive to disturbances and other image artifacts [18-19]. In order to make clustering more robust, Intuitionistic Fuzzy C-means clustering [20] assume that a pixel belongs to multiple clusters with different membership degrees [21]. Parsian and Kirmani [22], stated that when the overestimation and the underestimation are not equally important, the symmetric loss functions are not appropriate. Varian [23] proposed an asymmetric loss function, called linear exponential (LINEX) loss function. A LINEX loss function is convex and asymmetric. On one side of zero, it is approximately exponential and on the other side, it is linear. To study more about the properties of LINEX loss function, we refer to [24].

Many of the algorithms utilized only intensity of pixel as the only feature for the segmentation of images and failed to classify noisy pixels accurately. The pixels in an image are exceedingly associated, i.e. the every pixel in the prompt vicinity has about the equivalent feature information unless there is some curve or contour. Subsequently, integrating spatial information along with the membership value results in more homogenous regions as compared to other methods. However, these techniques use the arbitrary initialization of cluster centers which give inaccurate outcomes and more time for optimization.

In this paper, to overcome the defects of the algorithms as mentioned, we propose LINEX function induced an intuitionistic fuzzy C-means algorithm using Canberra measure based on membership filtering and similarity measurements, named the MST based LINEX IFCM. We apply a similarity measurement method between pixels and cluster centers instead of simple distance measurements into the proposed minimal spanning tree initialization method based LINEX IFCM to modify the membership matrix. The local gray information and Euclidean distance information are both taken into consideration to integrate the location information and gray information. This property achieves a good enhancement of the segmentation ability. The intuitionistic fuzziness is embedded into the calculation process of similarity between the pixel and cluster centers, which achieves more accurate segmentation in the organization boundary. The algorithm is minimal spanning tree initialization method by a given LINEX function using LINEX IFCM induced Canberra measure, which helps to speed up the convergence of the algorithm.

Minimum spanning tree is a useful graph for detecting clusters of a given set of data points. LINEX Canberra measure MST has been well suited for clustering in the field of pattern recognition, image processing and computational biology. In this paper is presented a novel approach to automatically detect the breast cancer. The proposed approach utilizes initialization method based on LINEX Canberra measure MST is proposed to compute initial cluster centers for the Intuitionistic fuzzy c-means clustering algorithm for clear to identify of abnormalities for mammography images. We summarized the mammography results and evaluated the accuracy of mammography, specificity, sensitivity, positive likelihood ratio, negative likelihood ratios were initially calculated. In addition to all these performance evaluation measures, predictive probability of mammography screening was also evaluated through Pearson chi square analysis.

2. Quantitative Analysis of Mammograms

For 40 highly suspicious cases mammography were obtained in which area of lump was highlighted and specificity and sensitivity parameters were calculated. Calculations include total number of true positive (TP), true negative (TN), false negative (FN) and false positive (FP) cases were calculated. among these four classified categories True negative patients were those women in which no lump was identified and symptoms were due to normal breast cycle or clotting of fatty tissues were present. False positive (FP) were those with benign cancer while TP were cases in which malignant or invasive breast cancer was detected. False negatives cases where those who developed malignant breast cancer during the period of screening (12 months). Age of presentation of disease symptoms and mammography screening was also recorded. Predictive probability of breast cancer detection based on mammography screening is examined using chi square test 2 test at ≤ 0.05 significance level. Along with this percentage distribution of 60 selected cases based on obvious breast cancer clinical symptoms was also calculated shown in Table 11.

2.1 Sensitivity

The sensitivity is expressed as the ratio of number of true positive, to the sum of ratio of false negative and true positive. Purpose of calculating sensitivity is to measure the reliability of a diagnostic system at making positive and negative identification. Hence to calculate sensitivity for our system understudy, we applied following formula.

Sensitivity $= \frac{TP}{TP+FN} \times 100$

2.2 Specificity

Specificity is expressed as the ratio of the number of true negatives, to the sum of false positive and true negative. This value defines the probability of a screening test to identify true negative cases.

Specificity $= \frac{TN}{TN+FP} \times 100$

2.3 Positive and Negative Likelihood Ratio Calculation

In the next step, sensitivity and specificity values are used to calculate positive likelihood ratio and negative likelihood ratio. These calculations will further measure the accuracy of mammography based breast cancer detection. Statistical formula used for calculating positive and negative likelihood ratios based on our study sample is given below

Positive Likelihood Ratio $= \frac{\text{Sensitivity}}{1-\text{Specificity}}$

Negative Likelihood Ratio $= \frac{1-\text{Sensitivity}}{\text{Specificity}}$

2.4 Predictive probability

Predictive probability of first screen mammography in accurate detection of breast lumps is calculated through chi square ( ) at < 0.05 level of significance. Chi square ( )formula given below where O denotes observed values of TP,TN,FP and FN cases given in Table 12 and E denotes expected values calculated in Table 14.

Chi-square $\chi^2 = \sum \frac{(O-E)^2}{E}$

3. Minimal spanning tree algorithm

In this section, we proposed composite Hyper Tangent and Sigmoid distance measures for construct the minimal spanning tree.

3.1. LINEX function MST

Given the grayscale point set D, the hierarchical methods starts by constructing a minimal spanning tree (MST) from the points in D. $x = (x_1, x_2, \ldots\ldots x_n)^T$ In and y= $y = (y_1, y_2, \ldots\ldots y_n)^T$ are two points of a MST and e(x,y) is an edge between x and y then LINEX distance function between x and y is denoted by and calculated using equation (1)

$$d_{LINEX}(x,y) = \exp\left(\frac{1}{K} \sum_{i=1}^{n} \frac{|x_i - y_i|}{x_i + y_i} \right) - \left(\frac{1}{K} \sum_{i=1}^{n} \frac{|x_i - y_i|}{x_i + y_i} \right) - 1 \quad (1)$$

3.2. Cluster Separation (CS)

The definition of CS between cluster centers is given by the following

$$CS = \frac{\delta(T_i)}{\Delta(T_i)} \quad (2)$$

where is the minimum standard deviation of and is the maximum standard deviation of . The value of CS ranges from 0 to 1. A low value of CS means that the two centroids are too close to each other and the corresponding MST Separation not valid. A high CS value means the MST separation of the data is even and valid. If the CS is greater than the threshold, the MST partition of the dataset is valid. Then, we increase the number of cluster by and test the CS again.

This process continuous until the CS is smaller than the threshold. The value setting of the threshold for the CS will be practical and is dependent on the dataset. The higher the value of the threshold the smaller the number of clusters would be, generally the value of the threshold will be .

3.3. Algorithm for determining the initial cluster centers

Algorithm: LINEX MST

Input: Data points

Output: optimal number of cluster centers

Let e1 be an edge in the LINEX Canberra measure MST constructed from Data points

Let be the set of disjoint subtrees of LINEX Canberra measure MST.

1. Create a node v, for each data points.
2. Compute the edge weight using equation (1)
3. Construct an LINEX Canberra measure MST1 from 2
4. $S_T = \phi$, $n_c = 1$, $C = \phi$
5. Repeat.
6. For each $e1 \in HSMST1$
7. Current longest edge or $\bar{T} + \sigma(T)$. e remove e1 from LINEX_MST1.
8. $S_T = S_T \cup \{T'\} // T'$ is new disjoint subtrees(regions).
9. $n_c = n_c + 1$
10. Compute the center using average of points.
11. $C = \cup_{T_i} \in S_T \{c_i\}$
12. $\delta(T_i)$ = Minimum standard deviation of .
13. $\Delta(T_i)$ = Maximum standard deviation of .
14. $CS = \frac{\delta(T_i)}{\Delta(T_i)}$
15. Until CS
16. Update the clusters points, repeat step 7 to step 14.
17. Finally we obtain the cluster centers.

4. Formulation of proposed method LINEX Canberra measure based intuitionistic fuzzy c means clustering

4.1. Intuitionistic fuzzy c-means (IFCM) algorithm

Intuitionistic fuzzy set given by Atanassov [16] considers both membership $\mu(x),\ x \in X$ and non-membership $\nu(x),\ x \in X$. An intuitionistic fuzzy set A in X, is written as

$$A = \{ x, \mu_A(x), \nu_A(x) \setminus x \in X \}$$

where $\mu_A(x) \to [0,1], \nu_A(x) \to [0,1]$ are the membership and non-membership degrees of an element in the set A with the condition $0 \le \mu_A(x) + \nu_A(x) \le 1$ when $\nu_A(x) = 1 - \mu_A(x)$ for every x in the set A, then the set A becomes a fuzzy set. Also indicated a hesitation degree $\pi_A(x)$ which arises due to lack of knowledge in defining the membership degree of each element x in the set A and is given by

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \text{ , } 0 \le \pi_A(x) \le 1$$

In [16] intuitionistic fuzzy c-means, minimizes the objective function as:

$$J_{IFCM} = \sum_{i=1}^{C} \sum_{k=1}^{N} u_{ik}^{*m} \| x_i - c_k \|^2 + \sum_{i=1}^{C} \pi_i^* e^{1-\pi_i^*} \qquad 1 < m < \infty \tag{3}$$

$u_{ik}^* = u_{ik} + \pi_{ik}$,where u_{ik}^* denotes the intuitionistic fuzzy membership and

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left[\frac{\| x_i - c_k \|}{\| x_i - c_j \|} \right]^{\frac{2}{m-1}}} \tag{4}$$

$$\pi_{ik} = 1 - u_{ik} - \left(\frac{1 - u_{ik}}{1 + \lambda u_{ik}} \right), \lambda > 0 \tag{5}$$

$$c_k = \frac{\sum_{i=1}^{N} u_{ik}^m x_i}{\sum_{i=1}^{N} u_{ik}^m} \tag{6}$$

$$\pi_i^* = \frac{1}{N} \sum_{k=1}^{N} \pi_{ik} \tag{7}$$

This iteration will stop when

$$\max_{ij} \left\{ \left| u_{ik}^{*k+1} - u_{ik}^{*k} \right| \right\} < \in ,$$

where $\in$ is a termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_{IFCM} .

4.2. LINEX Canberra measure based Intuitionistic fuzzy c-means (IFCM) algorithm

In this section, we want to use the LINEX Canberra measure in Intuitionistic fuzzy c-means (IFCM) algorithm when the overestimating and the underestimating are not of the same importance. The procedures are the same as a Intuitionistic fuzzy c-means algorithm. All the entities are assigned to their nearest centroid from MST, using a LINEX loss function as the dissimilarity distance. The procedure continues until there is no change in clusters. Now consider the following optimization problem,

objective function: ,

$$J_{IFCM}(U,C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u^m_{ij} L_{LINEX}(x_j, c_i) + \sum_{i=1}^{c} \pi_i^* e^{1-\pi_i^*} \qquad 1 < m < \infty \tag{8}$$

where $L_{LINEX}(x_j, c_i) = \exp\left(\frac{1}{K} \sum_{i=1}^{n} \frac{|x_j - c_i|}{x_j + c_i} \right) - \left(\frac{1}{K} \sum_{i=1}^{n} \frac{|x_j - c_i|}{x_j + c_i} \right) - 1$

The iterative updating of the cluster centers and membership functions for minimizing the objective function follows:

$$c_{ij} = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \quad , i = 1, 2, \ldots\ldots\ldots\ldots C \tag{9}$$

Here $u_{ij} = \frac{L_{LINEX}(x_j, c_i)^{\frac{-1}{(m-1)}}}{\sum_{k=1}^{c} \left(L_{LINEX}(x_j, c_i) \right)^{\frac{-1}{(m-1)}}}$, $i = 1, 2, \ldots\ldots C; j = 1, 2, \ldots\ldots n$ (10)

where $u_{ik}^* = u_{ik} + \pi_{ik}$ where u_{ik}^* denotes the intuitionistic fuzzy membership and

$$\pi_{ik} = 1 - u_{ik} - \left(\frac{1 - u_{ik}}{1 + \lambda u_{ik}} \right), \lambda > 0 \tag{11}$$

$$\pi_i^* = \frac{1}{N} \sum_{k=1}^{N} \pi_{ik} \tag{12}$$

The LINEX Canberra measure is suitable for clustering in which it can actually induce the necessary conditions.

This iteration will stop when, $\max_{ij} \left\{ \left| u_{ik}^{*k+1} - u_{ik}^{*k} \right| \right\} < \in$

The MST based LINEX_IFCM algorithm iteratively optimizes J_{IFCM} by continuous updating u_{ik}^* and until the difference in successive u_{ik}^* values is very small $\le \in$ where $\in$ is a small positive value between 0 and 1.

5. Efficient LINEX Canberra measure induced IFCM based MST [LINEX_IFCM]

5.1. Efficient LINEX_FCM Algorithm

Stage 1: Set the cluster centroids $\{c_j^*\}_{j=1}^{c}$ by using LINEX Canberra measure MST initialization method.

Stage 2: Compute the membership function using (10)

Stage 3: Update the cluster centroids using (9)

Stage 4: Go to stage (3)-(5), repeat until convergence.

Stage 7: Image segmentation after defuzzification and then a region labeling procedure is proposed.

Stage 8: The termination criterion is as follows $|J_m - J_{m-1}| \prec \in$, where m is the iteration count, $\in$ is a small number that can be set by the user.

The proposed efficient LINEX MST obtained cluster centers; the LINEX_IFCM algorithm continues iteratively updates, membership and centroids with these values. When this improved, Efficient LINEX_IFCM algorithm has converged, another defuzzification process takes place in order to convert the fuzzy partition matrix to a crisp partition matrix that is segmented.

5.2. Validation function based on feature structures

Two representative functions for the fuzzy partition namely; Partition coefficient V_{pc} and Validation function V_p are used to evaluate the validity of clustering [25-26].

$$V_{pc} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{*\,2} \tag{14}$$

$$V_p = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} u^m_{ij} L_{LINEX}(x_j, c_i) + \sum_{i=1}^{c} \pi_i^* e^{1-\pi_i^*}}{N \times \min\left\{\left\| c_i - c_j \right\|^2\right\}} \quad (15)$$

5.3 Fisher's Discriminant Ratio Concept

Discriminant function analysis or the original dichotomous discriminant analysis was developed by Sir Ronald Fisher in 1936. This method is a statistical analysis to predict a categorical dependent variable, and more continuous or binary independent variables. Fisher's Discriminant Ratio (FDR) is commonly used to measure the strength of discrimination the individual features in separate two classes based on its value. The expression of $\delta(\overline{T_i})$ and $\Delta(\overline{T_i})$ are the minimum and maximum average value of two classes respectively, while $\delta(\sigma^2(T_i))$ and $\Delta(\sigma^2(T_i))$ are minimum and maximum variance of the two classes in the feature to be measured respectively. FDR formula is defined as following equation:

$$FDR = \frac{\Delta(\overline{T_i}) - \delta(\overline{T_i})}{\Delta(\sigma^2(T_i)) - \delta(\sigma^2(T_i))} \quad (16)$$

Where : FDR = Fisher's Discriminant Ratio

$\Delta(\overline{T_i})$ = Maximum average edge weight of

$\delta(\overline{T_i})$ = Minimum average edge weight of

$\Delta(\sigma^2(T_i))$ = Minimum variance edge weight of

$\delta(\sigma^2(T_i))$= Minimum variance edge weight of

The result is given by FDR are the features that have a maximum difference in the average of the class and minimum variance of each class, therefore the high FDR value will be obtained. If two features have the average absolute difference that is equal but the number of variance is different, therefore the feature with the minimum number of variance will get the high FDR value. In other hand, if two features have the number of variance that is equal, but the average absolute difference is different, therefore the feature with the maximum average absolute difference that will get the high FDR value [27].

6. RESULTS AND DISCUSSION

Table 1. Random Data

Data			Intensity	Data			Intensity
S.No	X	Y	I(v)	S.No	X	Y	I(v)
1	1.80	2.00	0.50	11	12.00	4.00	0.80
2	2.00	2.20	0.91	12	11.50	3.50	0.45
3	2.00	1.80	0.12	13	12.50	3.50	0.55
4	2.00	3.50	0.40	14	21.00	10.00	0.65
5	8.80	3.00	0.50	15	21.00	11.00	0.25
6	9.00	3.20	0.38	16	20.50	10.50	0.35
7	9.00	2.80	0.60	17	21.50	10.50	0.75
8	9.20	3.00	0.12	18	2.00	4.00	0.70
9	7.00	2.80	0.80	19	19.00	20.00	0.60
10	12.00	3.00	0.90	20	11.00	12.00	0.40

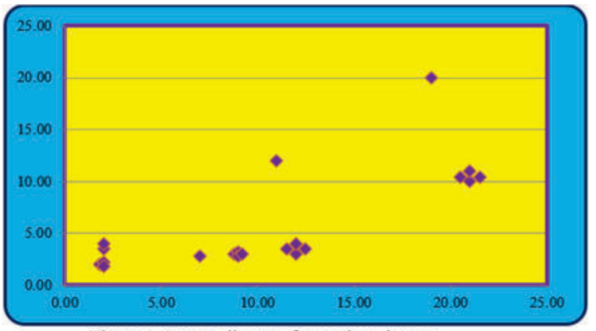

Figure 1. Scatter diagram for random dataset.

Table 2. Dissimilarity matrix

	Co-ordinate		intensity											
S.No	x	y	I(v)	vertex	1	2	3	4	5	6	7	8	9	10
1	1.80	2.00	0.50	1	0.0000	0.0081	0.0265	0.0101	0.0375	0.0531	0.0429	0.1047	0.0488	0.0731
2	2.00	2.20	0.91	2		0.0000	0.0380	0.0198	0.0571	0.0739	0.0463	0.1150	0.0281	0.0386
3	2.00	1.80	0.12	3			0.0000	0.0374	0.1056	0.0983	0.1092	0.0402	0.1081	0.1383
4	2.00	3.50	0.40	4				0.0000	0.0340	0.0257	0.0450	0.0768	0.0499	0.0677
5	8.80	3.00	0.50	5					0.0000	0.0018	0.0010	0.0209	0.0077	0.0102
6	9.00	3.20	0.38	6						0.0000	0.0046	0.0166	0.0157	0.0176
7	9.00	2.80	0.60	7							0.0000	0.0261	0.0039	0.0076
8	9.20	3.00	0.12	8								0.0000	0.0416	0.0405
9	7.00	2.80	0.80	9									0.0000	0.0068
10	12.00	3.00	0.90	10										0.0000
11	12.00	4.00	0.80	11										
12	11.50	3.50	0.45	12										
13	12.50	3.50	0.55	13										
14	21.00	10.00	0.65	14										
15	21.00	11.00	0.25	15										
16	20.50	10.50	0.35	16										
17	21.50	10.50	0.75	17										
18	2.00	4.00	0.70	18										
19	19.00	20.00	0.60	19										
20	11.00	12.00	0.40	20										

Table 3. LINEX Canberra measure based minimal spanning tree edges

S.No	Edges	LINEX Canberra measure	S.No	Edges	LINEX Canberra measure
1	(1,2)	0.0081	11	(11,10)	0.0022
2	(1,4)	0.0101	12	(10,9)	0.0068
3	(1,18)	0.016	13	(10,8)	0.0405
4	(1,3)	0.0265	14	(10,17)	0.0435
5	(1,5)	0.0375	15	(17,14)	0.0006
6	(5,7)	0.001	16	(14,16)	0.0061
7	(5,6)	0.0018	17	(16,15)	0.0022
8	(5,12)	0.0037	18	(16,20)	0.01
9	(12,13)	0.0011	19	(16,19)	0.0124
10	(13,11)	0.004			

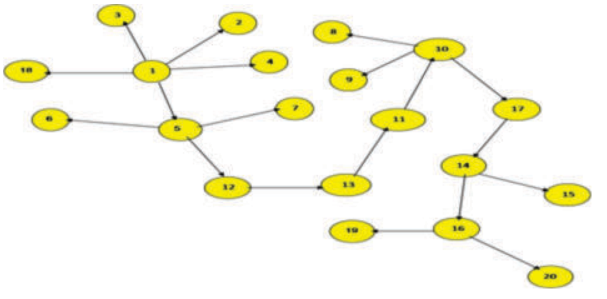

Figure 1. LINEX induced Canberra distance based Minimal spanning tree connected through points

Figure 1 shows a typical example of LINEX Canberra measure MST1 constructed from point set (from Dissimilarity matrix), in which inconsistent edges are removed to create subtree (clusters/regions).our algorithm finds the center of each clusters, which will be useful in many applications. Generally in most of the clustering algorithm data points can be represented as dissimilarity matrix representation. It contains the distance values between the data points represented as lower or upper triangular matrix. Our LINEX induced Canberra distance based minimal spanning tree algorithm constructs LINEX_MST1 from the dissimilarity matrix is shown figure 1. First to identify the longest edge in the LINEX_MST to generate subtree (clusters). Table 3, the longest edge weight 0.0435 connecting the data points 10 and 17 is find to be inconsistent one. By removing the inconsistent edge from the LINEX_MST, data points in the LINEX_MST partitioned into two subtrees or clusters and namely

$T_1 = \{1,2,3,4,5,6,7,8,9,10,11,12,13,18\}$ and $T_2 = \{14,15,16,17,19,20\}$

Next to find minimum and maximum standard deviation of T_i then Cluster Separation value less than 0.5 is valid. Here standard deviation of T_1=0.0139 and Standard deviation of T_2=0.0050 then $CS = \frac{0.0050}{0.0139} = 0.3610 < 0.5$ is valid. Secondly to remove next longest edge weight is 0.0405. Now minimal spanning tree produced outlier is {8}. Third to remove next longest edge, LINEX_MST partitioned into Three subtrees or clusters T_1, T_2 and T_3 namely T_1={1,2,3.4,18}, T_2 ={5,6,7,9,10,11,12,13}and.

T_3={14,15,1617,19,20}Next to find minimum and maximum standard deviation of T_1 then Cluster Separation value less than 0.5 is valid. Here standard deviation of T_1=0.0082, Standard deviation of T_1= 0.0020 and Standard deviation of T_3= 0.0050 then $CS = \frac{0.0020}{0.0082} = 0.2500 < 0.5$ is valid. Fourth to remove next longest edge weight is 0.0265. Now minimal spanning tree produced outlier is {3} Here standard deviation of T_1=0.0041, Standard deviation of T_2=0.0020 and Standard deviation of T_3=0.0050 then $CS = \frac{0.0020}{0.0041} = 0.5028 > 0.5$ is not valid. Finally LINEX minimal spanning tree produced three cluster center and outliers. Then the center of the cluster and its convergence of standard FCM and IFCM are determined under successive interactions of experiments using data points. The standard FCM algorithm and the numbers of updated centers are high under the objective function of Euclidean distance measures. This takes more iteration to converge the termination value of algorithm. With the new efficient objective function based LINEX induced Canberra measure the termination value is achieved, with very less iteration and with much better performance in getting membership (Table 8) than standard FCM. Table 9 gives the number of iteration to achieve the results of cluster on the data points by standard FCM and LINEX_IFCM. It is clear from the final cluster, membership (Table 8), scatter diagram (Fig 2), that our proposed LINEX_IFCM is much faster than the standard FCM and the method is converged fast to terminate condition with less run time. To test the effectiveness of LINEX_IFCM, the weighted minimal spanning tree based IFCM is used as center. This is done to find out the fuzzy membership and appropriate number of clusters. Thus, we have concluded the final optimal clusters formed as 3. This algorithm has also reduced the number of iterations. Best result is achieved by this measure fuzzy partition coefficient V_{pc} maximum and validation function V_p minimum (Table 10). The LINEX_IFCM clustering algorithm has the following membership value intimacy (Table 8),

Table 4. Final membership of three clusters of Intuitionistic FCM method and object allocation

Co-ordinate (x,y)			intensity				appropriate cluster
S. No	x	y	I(v)	Mem-1	Mem -2	Mem-3	
1	1.80	2.00	0.50	0.9429	0.0413	0.0158	1
2	2.00	2.20	0.91	0.8696	0.0976	0.0328	1
3	2.00	1.80	0.12	0.7255	0.1820	0.0925	1
4	2.00	3.50	0.40	0.8930	0.0801	0.0269	1
5	8.80	3.00	0.50	0.0462	0.9299	0.0240	2
6	9.00	3.20	0.38	0.0682	0.8869	0.0449	2
7	9.00	2.80	0.60	0.0394	0.9434	0.0172	2
8	9.20	3.00	0.12	0.1901	0.6772	0.1327	2
9	7.00	2.80	0.80	0.1995	0.7341	0.0664	2
10	12.00	3.00	0.90	0.0864	0.8407	0.0729	2
11	12.00	4.00	0.80	0.0799	0.8109	0.1092	2
12	11.50	3.50	0.45	0.0493	0.8939	0.0568	2
13	12.50	3.50	0.55	0.0304	0.9337	0.0360	2
14	21.00	10.00	0.65	0.0276	0.0774	0.8950	3
15	21.00	11.00	0.25	0.0446	0.0917	0.8638	3
16	20.50	10.50	0.35	0.0260	0.0589	0.9151	3
17	21.50	10.50	0.75	0.0346	0.0888	0.8766	3
18	2.00	4.00	0.70	0.8398	0.1204	0.0398	1
19	19.00	20.00	0.60	0.0305	0.0787	0.8908	3
20	11.00	12.00	0.40	0.0485	0.1794	0.7720	3

Table 5. Comparison of iteration count

	No. of iterations	No. of clusters
Standard FCM	7	3
IKFCM	6	3
MST based Intuitionistic FCM	3	3

Table 6. Cluster validity function

Standard FCM	0.8250	0.2554
IKFCM	0.8905	0.2234
LINEX Intuitionistic FCM	0.9012	0.0457

Table 7. FDR = Fisher's Discriminant Ratio

No of Clusters	FDR
2	0.1653
3	1.6036

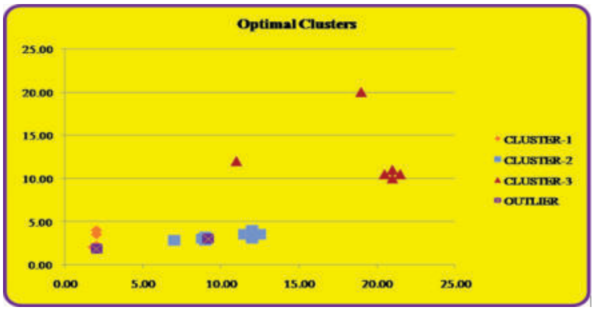

Figure 2. Initialization method based Intuitionistic FCM, final cluster three.

6.1. Statistical Evaluation of Diagnostic Performance of Mammography for Breast Cancer

Our study of random sample in terms of reported breast cancer associated symptoms and patient's age group reveals that 10% of the patients had dense calcification, 10% had watery discharge from breast, 40% were complaining of lump, 30% had pain in breast tissues, 5% cases were having both lump and pain. While 5% were suffering from pain as well as discharge from breast tissues given in table 1. Patient's data is categorized into two age groups; 25% cases belong to age group of 30-40 years while majority (75%) belongs to age group of 41-50 years. Mammography details revealed 33% were having benign tumor while malignancies were reported in 50% cases and 17% cases were diagnosed as normal shown in Table 8.

To evaluate the performance of diagnostic procedure for primary screening of breast cancer, initially specificity and sensitivity was calculated. Randomly select the sample dataset out of 40 patients subjected to mammography for detection of lump in mammary tissues, diagnosis reports analysis revealed 21 cases as TP, as disease was present in them while 2 false negative cases were observed in which diseases was present but symptoms or clinical presentation could not be evaluated through mammography. Likewise, 1 false positive case were reported through mammography and 16 true negative cases were also identified in which no indication of disease was observed. All the cases in terms of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are properly summarized in Table 9.

Table 8. Percentage distribution of study sample (n=60) based on age of disease presentation (years), breast cancer associated symptoms and nature of tumor.

Patient's Factors	%age
Age at diseases presentation (years)	
30-40	25% (10)
41-50	75% (30)
Breast Cancer Associated Symptoms	
Watery Discharge	10% (4)
Calcification	10% (4)
Lump	40% (16)
Pain	30% (12)
Lump & Pain	5% (2)
Pain and watery Discharge	5% (2)
Nature of Tumor	

Benign	33% (13)
Malignant	50% (20)
Normal	17% (7)

Table 9. Summary of total no of cases diagnosed through mammography as true positive (TP), false positive (FP), false negative (FN) and true negative (TN) in selected study sample.

Test	Breast Lump Present (n)		Breast Lump Absent (no)		Total
Positive	True Positive (TP)	21	False Positive (FP)	1	TP +FP= 22
Negative	False Negative (FN)	2	True Negative (TN)	16	FN + TN= 18
Total	TP +FN = 23		TN +FP=17		40

Table 10.Statistical analysis of Sensitivity, Specificity, Positive and Negative Likelihood ratio at 95% CI to evaluate diagnostic accuracy of mammography for breast cancer detection.

No.	Statistic Evaluation	Formula	Value
1	Sensitivity	$\frac{TP}{TP+FN}\times 100$	91.30%
2	Specificity	$\frac{TN}{TN+FP}\times 100$	94.11 %
3	Positive Likelihood Ratio	$\frac{\text{Sensitivity}}{1-\text{Specificity}}$	15.52
4	Negative Likelihood Ratio	$\frac{1-\text{Sensitivity}}{\text{Specificity}}$	0.09

Table 11.Chi square test analysis for evaluation of diagnostic accuracy of mammography for breast cancer detection

Test	Observed value (O)				Row Total
	Breast Lump Present (n)		Breast Lump absent (n)		
Positive	True Positive (TP)	21	False Positive (FP)	1	Sum of row=22
Negative	False Negative (FN)	2	True Negative (TN)	16	Sum of row = 18
Column Total		23	17		N = 40

Chi square () Test					
Observed cases (O)	Expected value (E)	$(O-E)$	$(O-E)^2$	$\frac{(O-E)^2}{E}$	p-value
True Positive	[(22*23)/40] =12.65	8.35	69.7225	5.51	<0.00001** ** highly significant as p-value is <0.05
False Negative	[(18*23)/40] =10.35	-8.35	69.7225	6.74	
False Positive	[(22*17)/40] = 9.35	-8.35	69.7225	7.46	
True Negative	[(18*17)/40] = 7.65	8.35	69.7225	9.11	
Chi-square $\chi^2 = \sum \frac{(O-E)^2}{E}$				28.82	

Formula of Expected value (E) [sum of row * sum of column /(n)]

6 CONCLUSION

In this article, is introduced new alternative approach for breast cancer disease diagnosis and classifying benign and malignant breast cancer using LINEX Canberra meausre MST initialization based modified LINEX Intuitionistic fuzzy c-means clustering algorithm for clear to identify of abnormalities for mammography images.. we have successfully applied Fisher's Discriminant Ratio and Chi-square as feature selection alogorithm to breast data. Finally we summarized the mammography results and evaluated the accuracy of mammography, 91.30% sensitivity, 94.11% specificity, 15.52 positive likelihood ratio, 0.09 negative likelihood ratios were calculated.

The resulting classification process is perfomed using the Statistical learning method and determines whether the tumor is normal, benign or malignant with the output image of the segmented part of the breast. It would be helpful to health professionals for making timely decisions for disease management in breast cancer patients.

7. Acknowledgement

We would like to thank the reviewers for their constructive comments. We thank to Dr.R.David Chandrakumar, Professor, Department of Mathematics, Vickram College of Engineering for his encouragement and support given. I also thank Dr.A.Sathya, Assistant Professor, Department of Mathematics, Government College of Daman, Daman & Diu for her timely help throughout this work.

REFERENCES

[1] Siegel R, Desantis C and Jemal A, "Colorectal cancer statistics", Cancer Journal of Clinicians, vol. 64(2), pp.104-117(2014).
[2] Ertosun M and Rubin D, "Probabilistic visual search for masses within mammography images using deeplearning" In Proc. IEEE Int. Conf. of Bio Informatics and Biomedics (BIBM), Paris, France, pp.1310-1315(2019).
[3] Tzikopoulos S, Mavroforakis M and Theodoridis S, "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry" Computing Methods Programs Biomedical, vol. 6(1), pp. 47-63 (2020)
[4] Punitha S, Amuthan A and Joseph KS, "Benign and malignant breast cancer segmentation using optimized region growing technique" Future Computing and Informatics Journal, vol. 3, no. 2, pp. 348-358 (2018).
[5] J. Dheeba J and Selvi ST, "A Swarm Optimized Neural Network System for Classification of Microcalcification in Mammograms" Journal of Medical Systems, vol.36(5), pp. 3051-3061(2012).
[6] Pawar MM and Talbar SN, "Genetic Fuzzy System (GFS) based wavelet co-occurrence feature selection in mammogram classification for breast cancer diagnosis" Perspectives in Science, vol.8, pp. 247-250 (2016).
[7] W. Qian, F. Mao, X. Sun, Y. Zhang, D. Song, and R. A. Clarke, "An improved method of region grouping for microcalcification detection in digital mammograms" Computerized Medical Imaging and Graphics, vol. 26(6), pp.361-368(2002)
[8] Rehman ZU, Naqvi SS, Khan TM, Khan MA, Bashir T,"Fully automated multi-parametric brain tumour segmentation using superpixel based classification" Expert Syst. Appl,vol.118, pp.598–613 (2019).
[9] Pham D L, Xu C and Prince J. L, " Current methods in Medical Image Segmentation" Annual Review of Biomedical Engineering, vol (2),pp.315-337 (2000).
[10] Bezdek JC, "Measuring fuzzy uncertainty" Fuzzy Systems 2 (1994).
[11] Jawahar CV and Ray A K, "Fuzzy statistics of digital images" IEEE Signal Process. Lett.vol.3(8), pp.225-227 (1996).
[12] Li. BN, Chui CK, Sihang and Ong SH,"Integrating Spatial fuzzy clustering with Level Set methods for automated medical image segmentation" Computers in Biology and Medicine, vol.41, pp. 1-10 (2011).
[13] Cai W, Chen S and Zhang D,"Fast and Robust Fuzzy c-means clustering algorithm incorporating local information for image segmentation." Pattern Recognition, vol.40, pp.825-838 (2007).
[14] Shenton M.E, Koerte IK, "Automated versus manual segmentation of brain region volumes in former football players" Neuroimage Clin, vol.18, pp.888–896 (2018).
[15] Guenette JP, Stern RA et al., "Active contours driven by edge entropy fitting energy for image segmentation" Signal Process, vol.149, pp.27–35 (2018).
[16] Atanassov K, "Intuitionistic fuzzy sets" Fuzzy Sets Systems, vol.20(1), pp.87–96 (1986).
[17] Chaira T, "A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images" Applied Soft computing, vol.11(2), pp. 1711-1717 (2011).
[18] Szmidt E and Kacprzyk J, "Distances between intuitionistic fuzzy sets" Fuzzy sets Systems, vol.114, pp. 505-518(2000).
[19] Arora J and Tushir M,"Robust spatial intuitionistic fuzzy C-means with city-block distance clustering for image segmentation" Journal of Intelligent & Fuzzy Systems, vol.35, pp. 5255-5264 (2018).
[20] Li G, Long C, Chen CLP and Zhou J, "Integrating guided filter into fuzzy clustering for noisy image segmentation" Digit. Signal Process, vol.83,pp. 235–248 (2018).
[21] Xu Z and Wu J, "Intuitionistic fuzzy C-means clustering algorithms" J. Syst. Eng. Electron, vol. 21, pp.580–590 (2010).
[22] Parsian and Kirmani SNUA, "Estimation under LINEX Loss Function", Handbook of Applied Econometrics and Statistical Inference, vol.165, pp. 53-76 (2002).
[23] Varian HR, "A Bayesian Approach to Real Estate Assessment" Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage (Eds S. E. Fienberg and A. Zellner), pp.195-208 (1975).
[24] Zellner A, Bayesian Estimation and Prediction Using Asymmetric Loss Functions, Journal of the American Statistical Association, vol. 81, pp.446-451(1986).
[25] Xie X.L and Beni G, A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell., Vol. 3(8), pp-841-847(1991).
[26] Bezdek J.C.,Hall L.O.,Clarke L.P, Review of MR image segmentation techniques using pattern recognition, medical physics 20(4), pp. 1033-1048 (1993).
[27] EkoPrasetyo, "Data Mining-Mengolah Data Menjadi Informasi Menggunakan Matlab", Andi Offset, Ed.1, pp.28-30, (2014).