IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 February 2024, accepted 17 April 2024, date of publication 22 April 2024, date of current version 10 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3392032

RESEARCH ARTICLE

An Improved Robust Fuzzy Local Information K-Means Clustering Algorithm for Diabetic Retinopathy Detection

HUMA NAZ[1,2], TANZILA SABA[2], (Senior Member, IEEE), FATEN S. ALAMRI[3], AHMED S. ALMASOUD[2], AND AMJAD REHMAN[2], (Senior Member, IEEE)

[1]School of Computer Science, University of Petroleum and Energy Studies, Dehradun 248007, India
[2]Artificial Intelligence and Data Analytics Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia
[3]Department of Mathematical Sciences, College of Science, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Corresponding author: Faten S. Alamri (fsalamri@pnu.edu.sa)

ABSTRACT According to the International Diabetes Federation (IDF), roughly 33% of individuals affected by diabetes exhibit diagnoses encompassing diverse severity of diabetic retinopathy. In the year 2020, approximately 463 million adults within the age bracket of 20 to 79 were documented as diabetes sufferers on a global scale. Projections suggest a rise to 700 million by 2045. The proposed automated diabetic retinopathy detection methods aim to reduce the workload of ophthalmologists. The study presents the Robust Fuzzy Local Information K-Means Clustering algorithm, an advanced iteration of the classical K-means clustering approach, integrating localized information parameters tailored to individual clusters. Comparative analysis is conducted between the performance of Robust Fuzzy Local Information K-Means Clustering and Modified Fuzzy C Means clustering, which incorporates a median adjustment parameter to augment Fuzzy C Means for diabetic retinopathy detection. The results are evaluated on three datasets: IDRiD, Kaggle, and fundus images collected from Shiva Netralaya Center, India. Achieving a 94.4% accuracy rate and an average execution time of 17.11 seconds, the proposed algorithm aims to categorize a substantial volume of retinal images, thereby improving performance and meeting the crucial demand for prompt and precise diagnoses in diabetic retinopathy healthcare.

INDEX TERMS Diabetic retinopathy detection, unsupervised learning, Fuzzy C Means, clustering, healthcare.

I. INTRODUCTION

Diabetic retinopathy (D.R.), the predominant ocular complication associated with diabetes, impacts roughly one-third of the global diabetic population. Its prevalence is notably higher in regions where diabetes is more widespread. Approximately 60 million individuals are living with diabetes, with diabetic retinopathy impacting around 12% to 18% of diabetes patients in India [1]. Therefore, developing countries like India face a shortage of eye care facilities due to the substantial diabetic patient population. D.R. screening lacks a coordinated national strategy and is instead conducted sporadically. Challenges for hospital-based or outreach screening camps involve patient awareness, accessibility, and the scarcity of adequately trained ophthalmologists and clinical teams. If left untreated, this condition poses a substantial risk of vision loss, making it a leading cause of blindness among working-age adults in many developed nations. Regular eye exams are essential for detecting D.R. in its early stages [2]. The American Diabetes Association recommends annual dilated eye exams for people with diabetes. Ongoing research seeks to improve the understanding of D.R. and develop more effective treatments and early detection methods [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Hasan S. Mir.

The identification of D.R. involves the observation of various lesions, such as microaneurysms (M.A.) [4], hard exudates (EX) [5], and hemorrhages (H.M.) [6]. Red lesions are typically associated with M.A. and H.M. within this context, while bright lesions indicate soft and hard exudates. D.R. is categorized into five stages based on these lesions: no D.R., mild D.R., moderate D.R., severe D.R., and proliferative D.R., as shown in Figure 1. Deep learning excels in various domains but heavily relies on labeled data for supervised learning [7]. Collecting and labeling such data is costly and time-consuming, presenting a significant research opportunity to address the challenge of reducing this labeling burden for real-time applications [8].

The central challenge in Deep Neural Networks (DNN) is the issue of overfitting, even when clustering algorithms demonstrate excellent performance [9]. To mitigate this, feature selection can be employed to extract relevant information from the dataset, aligning with the application's requirements. These selected features offer valuable insights, allowing for a deeper exploration of possibilities [10], [11]. This research aims to enhance the synergy between deep learning methods and clustering algorithms, enabling the implementation of unsupervised learning techniques for efficient Dimensionality Reduction classification.

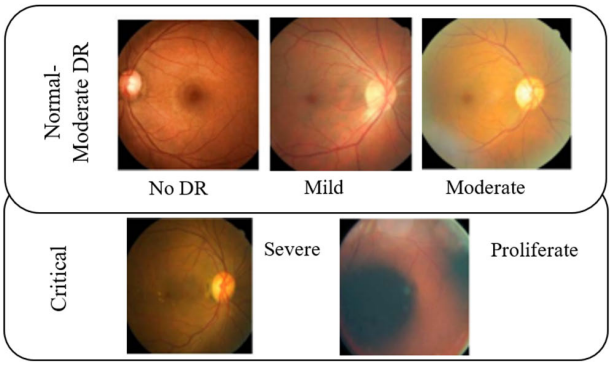

FIGURE 1. Diverse stages of D.R. [12].

The conventional k-means algorithm, a fundamental technique for vector quantization, exhibits certain limitations that impact its efficacy. Firstly, it operates linearly, implying that data elements assigned to one cluster are not considered for others [12]. In real-world scenarios, however, it is common for data points to have shared attributes across clusters. Secondly, the algorithm tends to converge towards local minima rather than the global minima due to its reliance on initial random centroid selection. This can adversely affect the final cluster configuration. Lastly, the k-means algorithm exhibits minimal adjustments toward cluster centers during each iteration, leading to extended convergence times. Thus, the proposed work combines k-means clustering with local cluster information, calculated using the inverse Euclidean distance and cluster centre. This function converges the cost function in local minima and provides better results (equation1).

$$Obj_{Mdkm} = \sum_{i=1}^{nc} \sum_{j}^{nx} M^f id_{ij} + lin_{ji} \tag{1}$$

The proposed algorithm is evaluated using fundus images related to D.R. detection, and a comparative analysis was conducted against the Modified Fuzzy C Means (MaFCM) algorithm [13]. This analysis also included Fuzzy C-Means (FCM), K-Means, and Autoencoder-based Deep Embedded Clustering (DEC) [14]. The outcomes of this comparison aim to validate the model's generalizability. Significantly, the performance of the proposed model exceeded that of the other algorithms.

The contributions of this research could be summarized as follows:

- The foundational k-means clustering algorithm is enhanced by incorporating local information parameters for specific clusters. This augmentation calculates the inverse Euclidean distance between data points and cluster centers. It serves as an adjustment factor, expediting the convergence of cluster centers and diminishing the iteration count needed.
- The proposed algorithm is compared with the modified Fuzzy C-means (FCM) algorithm, which is known for its ability to optimize global optima within the conventional FCM framework. This modified algorithm includes an additional median adjustment parameter in the objective function to enhance solution optimality further.
- The algorithms are assessed using the D.R. fundus dataset to evaluate the model's performance. Furthermore, the proposed models are compared with state-of-the-art methods to demonstrate the model's generalizability.

The remaining paper is organized as follows: The second section presents recent work on unsupervised deep learning methodologies. The next section of the paper discusses the proposed work in detail. The fourth section discusses the results, and the paper concludes with the fifth section.

II. LITERATURE REVIEW

Supervised learning methods offer great potential for solving various problems [15]. However, unsupervised learning methods can uncover many new opportunities [16]. Specifically, data mining techniques for clustering can organize unfamiliar data into meaningful structures without explicit guidance. Many of these clustering methods use distance measurements to do this. When you combine deep learning and clustering, you get what is known as deep clustering algorithms [17].

DEC is a method that excels in unsupervised learning because it is good at clustering data [18]. It can assign data points to clusters and learn valuable features from the data, which fills in gaps that cannot be done with supervised learning [7]. Several algorithms have evolved to improve DEC.

During DEC, a spatial space for features is created using Autoencoders. These Autoencoders transform the actual data

into different features in a hidden space [14]. The clustering process also affects how the Autoencoder training is performed by setting rules for clustering. DEC works in two steps: first, there is a pre-training phase where initial parameters like cluster centres and stopping criteria are set up. Then, in the fine-tuning phase, feature learning and clustering are combined. DEC prefers to apply Autoencoders because they are simple, reliable, and work well for reconstructing data [11].

However, the Discriminately Boosted Clustering (DBC) algorithm closely matches the procedures of DEC, encompassing the training approach, clustering methodology, utilization of K.L. divergence, and the distribution of soft cluster assignments. The primary distinction lies in adopting a Convolutional Neural Network (CNN) based Autoencoder instead of a conventional Autoencoder (typically Feedforward). This alteration significantly enhances accuracy in the DEC algorithm, mainly when dealing with image-based datasets. [19].

Then, the Deep Clustering Network (DCN) approach tackles the challenge of integrating unsupervised learning into deep learning by creating a unified process. Instead of working with the more intricate latent feature space, DCN leverages the learned weights (typically denoted as "w") acquired after each training session for the clustering task [20]. This approach advocates a joint operation of dimensionality reduction and clustering, specifically employing k-means clustering assignment to achieve its objective. The optimization criteria employed in this process involve three practical steps: (1) reducing dimensionality, (2) reconstructing data, and (3) enhancing cluster-based regularization, as proposed by Min et al. in 2018. The latent representation that is clustering-friendly w is used instead of the latent feature space, as discussed below (equation 2).

$$h_i = f\left(x_i; w\right), \quad f\left(.; w\right) : R^M \rightarrow R^R \; where \; f\left(, ; w\right) \tag{2}$$

Here, equation (3) indicates the function of mapping the w weights associated with the particular input x in the respective hidden layer h_i.

While prior algorithms like DEC, DBC, and DCN have addressed specific problems, they exhibit limitations in terms of scalability and efficiency, particularly when handling extensive datasets. Deep Embedded Regularised Clustering (DEPICT) distinguishes itself through its unique approach [20]. Instead of following the conventional methods, DEPICT utilizes a multi-layer convolutional autoencoder and introduces a relative entropy-based objective function to regulate cluster assignments. DEPICT employs a data-dependent regularization strategy to enhance its robustness in computing the reconstruction loss, thereby guarding against overfitting during network training. This study presents a joint learning framework, which efficiently minimizes both the clustering loss and the reconstruction loss while concurrently training the network [20]. The initial stage of this combined process involves computing the probabilistic cluster assignment, as outlined in equation (3).

$$p_{ik} = \frac{\exp\left(\theta_k^T Z_i\right)}{\sum_{k'=1}^{k} \exp\left(\theta_{k'}^T Z_i\right)} \tag{3}$$

The above equation can also be defined as the Here, $\Theta = []$ represents the SoftMax function, which indicates the data point's current probability distribution. Further, p_{ik} it represents the probability calculation of data points in k_{th} the cluster (Maria Hauser).

Continuing with exploring unsupervised deep learning, researchers have delved into a novel approach, VaDE, which utilizes the Variational Autoencoder (VAE). Within VaDE, an unsupervised generative clustering method combines the Gaussian Mixture Model (GMM) and the Deep Neural Network (DNN). VaDE allows for interpolation within the latent representation space and the generation of new samples that adhere to the data distribution [21].

Data clustering in VaDE initially involves the application of the GMM, followed by generating latent embedding features denoted as "z." Subsequently, these features are directed towards the DNN for decoding into observable data. The joint operation of the VAE and the GMM is governed by the Evidence Lower Bound (ELBO), as demonstrated in the work [22].

The Joint Unsupervised Learning (JULE) approach's data processing encompasses deep learning and clustering. Through the utilization of a stacked Convolutional Neural Network (CNN), this method proves particularly well-suited for image datasets. A recurrent structure is employed within this framework, wherein agglomerative clustering is executed during the forward pass, while CNN-based learning occurs during the backward pass [22], [23]. In contrast to alternative approaches, JULE introduces a singular loss function aimed at optimizing the recurrent framework, which encompasses both CNN and agglomerative clustering components [18].

The data dimension is considered a critical factor that outlines the limitations and advantages of clustering algorithms [24]. In most cases, the distance between data points is measured by clustering algorithms to assess similarities, but this approach is less effective as data dimensionality increases in complexity. The utilization of deep learning algorithms can ameliorate this issue, but dimension reduction remains necessary, even in spectral clustering methods [25]. The deep clustering algorithm performance on high-dimensional data can be improved by establishing a well-defined objective function for training the deep learning model.

III. METHODOLOGY

The recent advancements in automated disease detection have been a tremendous help to society and overloaded medical and healthcare facilities. Many factors have contributed to this progress, such as high image acquisition quality, machine learning and artificial intelligence utilization, immense storage capacity, and improved computing facilities.

In this era, entities focused on development are increasingly demanding information and analytics to enhance their

outcomes, efficiency, and performance [26]. The importance of analytics and learning is increasing significantly in the context of progress, yielding promising results to generate highly efficient solutions for the betterment of the future. Despite utilizing traditional algorithms and their variations to advance outcomes, the emphasis on creating more optimal solutions is rising [27].

Based on previous research, it is evident that modifications are required in the processing methods of clustering/deep learning algorithms due to their limitations. The issue of overfitting is regarded as a central concern in the context of DNN, causing challenges even when the clustering algorithm's performance is exceptional [9]. Nevertheless, selecting features relevant to the application's requirements can be carried out to address the processing of irrelevant information in the dataset. Furthermore, the chosen features offer crucial data that enables a more profound understanding to explore additional possibilities [10], [11].

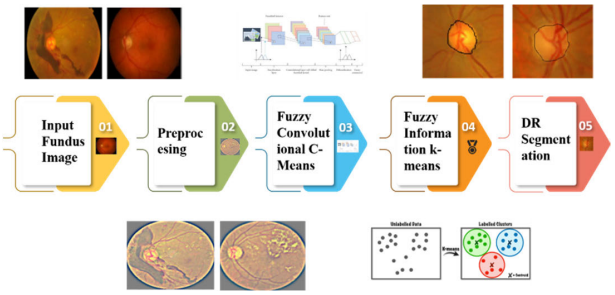

FIGURE 2. Proposed methodology.

This approach constructs an automated D.R. detection system, employing fuzzy information in conjunction with the K-means (FIKM) algorithm. The model is then compared with the MaFCM algorithm to demonstrate its generalizability. Statistical features are generated or recognized in this process to produce a distinctive combination of independent attributes. The proposed methodology includes a discussion of dataset description, preprocessing, and algorithm details, shown in Figure 2.

A. DATASET DESCRIPTION

In the study context, the utilized datasets include IDRiD, Kaggle dataset, and fundus images obtained from Shiva Netralaya in Uttar Pradesh, India. The fundus images, collected from Shiva Netralaya in Uttar Pradesh, India, were taken from 50-60 patients diagnosed with moderate to severe Diabetic Retinopathy (D.R.) affecting both eyes and captured using a Zeiss tabletop fundus camera with a 35mm focal length. The statistics of the dataset are shown in Table 1. The Indian diabetic retinopathy image (IDRiD) dataset was created using actual clinical eye exams from a clinic in Nanded, India. The IDRiD dataset focuses on individuals with diabetes. The captured images have a 50° field of view resolution of 4288 × 2848 pixels. They are stored in jpg format with a 50-degree field of view. They are categorized into five D.R. classes with

TABLE 1. Statistics of fundus images from shiva netralaya center (primary dataset).

Fundus Images Source	Location	No of Patients	Diagnosed Condition	Camera Used	Image Dimensions	Resolution	Bit Depth
Shiva Netralaya	Uttar Pradesh, India	50-60	Moderate to Severe Diabetic Retinopathy	Zeiss Tabletop Fundus Camera (35mm focal length)	2866 x 2866 pixels	96 dpi	24

severity levels from 0 to 4 and three Diabetic Macular Edema (DME) with severity levels from 0-2, as shown in Table 2.

The Kowa VX-10α fundus camera captures retinal images after dilating subjects' pupils with tropicamide. The dataset consists of 516 high-resolution (4288 × 2848 pixels) JPEG images following international clinical standards. Expert annotations for D.R. lesions, typical retinal structures, and severity levels for D.R. and DME are provided, making it a valuable resource for diabetic eye disease research [28].

In this study, alongside the fundus images obtained from Shiva Netralaya Center and IDRiD, the dataset from the Kaggle competition titled "Diabetic Retinopathy Detection"

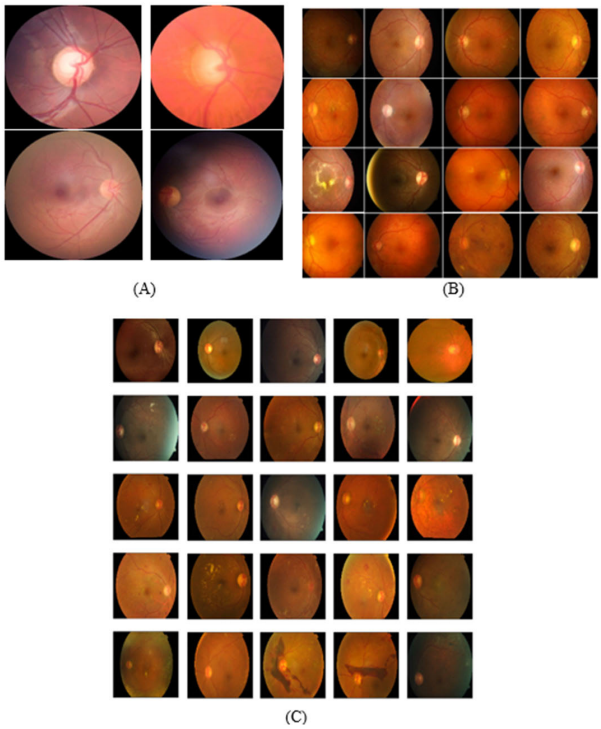

FIGURE 3. Retinal fundus images (A) Sample images from the primary dataset collected from the shiva netralaya center (B) Samples from IDRiD datasets showing dark, differently illuminated images (C) Kaggle dataset sample images.

is also incorporated. This Kaggle competition represented a significant milestone in medical image analysis and computer vision, focusing on the development of algorithms that can automatically detect and categorize D.R. in retinal images. The sample images of all applied datasets are shown in Figure 3.

TABLE 2. Statistics of kaggle IDRiD dataset.

Dataset	Type	No of Images	Train Set Size	Test Set Size	Labels Included
Segmentation	Fundus Images	81	51	30	None
	Ground truth Images (Lesions)	NA	51	30	M.A., H.M., Hard EX, Soft EX, Optic Disc
Disease Grading	Fundus Images (Color)	516	413	103	None
	Ground truth labels	NA	413	103	DR Severity Grade Diabetic Macular Edema Severity Grade
Localization	Fundus Images	516	413	103	None
	Ground truth Labels (Optic Disc Center)	NA	413	103	Optic Disc Center Location
	Ground truth Labels (Fovea Center)	NA	413	103	Fovea Center Location

The Kaggle competition generously provided a substantial dataset comprising 88,702 retinal images [29]. These images were meticulously gathered and annotated by EyePACS, with 35,126 images designated for training and 53,576 for testing, as clearly outlined in Table 3. Each retinal image was subjected to expert annotations, indicating the severity levels of D.R., which ranged from 0 to 4. Notably, the dataset encompassed images of varying dimensions, ranging from 320×211 pixels to 5184×3456 pixels. These images were captured under diverse imaging conditions, thereby adding complexity and diversity to the dataset. Notably, the dataset comprised images representing both left and right eyes, further enhancing its utility for developing and evaluating algorithms in the domain of D.R. detection and classification.

TABLE 3. Statistics of kaggle DR dataset.

	Normal	Mild	Moderate	Severe	-
Raw Images	18472	1870	1870	583	-
Augmentation	18472	28050	29869	27401	-
Total Number	36944	29920	34136	27984	-
Train Images	35844	28920	33136	26984	-
Test Images	1000	1000	1000	1000	-
Label	0-No DR	1-Mild NPDR	2-Moderete NPDR	3-Severe NPDR	4-Proliferative DR

B. PREPROCESSING

Preprocessing plays a pivotal role in optimizing image quality and facilitating accurate analysis. These images often suffer from challenges like noise, inconsistent lighting, and artifacts like reflections or dust spots. Preprocessing steps, including noise reduction, contrast enhancement, and artifact removal, ensure clear visualization of critical anatomical structures like blood vessels, optic disc, and macula. Normalization of pixel values helps maintain consistency across images. At the same time, segmentation techniques enabled by preprocessing are vital for delineating specific structures relevant to D.R. diagnosis, such as abnormalities in blood vessels [30]. Additionally, data augmentation during preprocessing enhances the robustness of machine learning models by introducing variations to the training dataset, ultimately improving the efficacy of automated analysis and detection of D.R. [31]. The preprocessed fundus images, obtained through circle cropping and Ben Graham's technique, are illustrated in Figure 4.

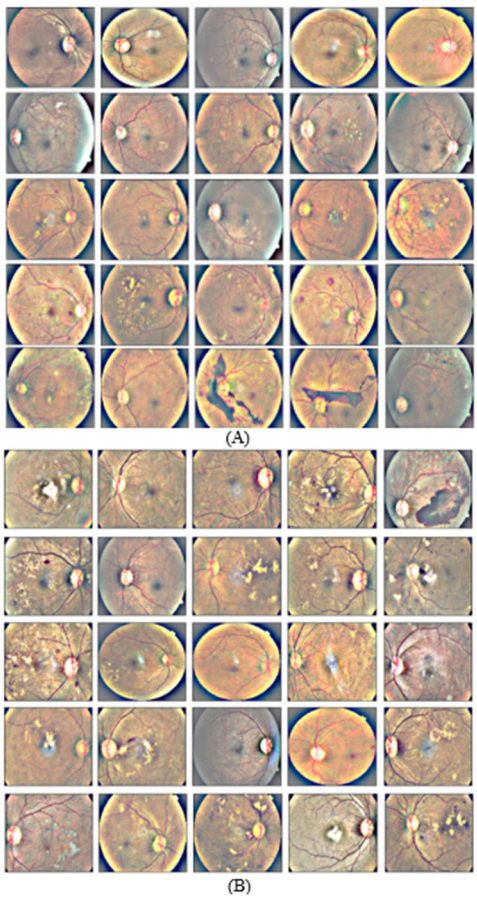

FIGURE 4. Preprocessed retinal fundus images (A) Circle-cropped and contrast-enhanced images from kaggle, IDRiD, and the primary dataset. (B) Clipped preprocessed fundus images.

C. FUZZY INFORMATION K-MEANS

The traditional k-means vector quantization approach stands out as an efficient clustering technique within the data mining realm. It employs a recursive technique to determine convergence, randomly selecting cluster centers and manipulating distances. The iterative process involves recalculating cluster centers by taking the mean of clusters from the previous iteration. This process repeats until convergence is achieved, gradually bringing the k centers closer to the converged state. The classic k-means algorithm is defined as follows: (1) Convergence function, (2) Cluster Centre computations, and (3) Cluster bin.

$$C_f = \sum_{i=1}^{nc} \left(\|C_c i - NC_c i\|^2 \right) \tag{4}$$

$$NC_c i = \left(\frac{1}{ndc_i} \right) \sum_{j=1}^{ndc_i} Cbin_i \tag{5}$$

$$Cbin_i = (Cbin_i \,|\, X_s i) \tag{6}$$

$$X_s i = \lim_{j=1_{tonx}} x_j; \quad if\ dist\ of\ x_j\ is\ \min\ to\ C_c i \tag{7}$$

$$dist = \left\| x_j - C_c i \right\|^2; \quad j = 1\ to\ nx,\ i = 1\ to\ nc \tag{8}$$

The Convergence function, denoted C_f,, calculates the variation between the cluster centers of the previous and current iterations (equation 4). Determining the stopping criterion and controlling recursive stages relies on the desired precision level specified by the user. The variable *nc* signifies the user-specified number of clusters, either existing or requested. Within this context, the variables $C_c iNC_c i$ are provided (equation 5). The calculation of the present cluster center, denoted as $C_c i$, and the new cluster center, represented by $NC_c i$, is executed by leveraging the cluster bin ($Cbin_i$) construction (equation 6).

The number of Cluster bins available for use is denoted as ndc_i equivalent to the user-specified number of clusters (*nc*), as the generated bins align with the required number of clusters. The $Cbin_i$ algorithm is developed utilizing a set of selected data points, denoted as $X_s i$, which represent the minimal distant points or elements between x_j and $C_c i$ (equation 7). These data points are crucial in determining the cluster center during clustering. The variable *nx* indicates the dataset's cardinality, signifying the total number of elements within the dataset. The traditional k-Means algorithm exhibits several limits in properly handling data. One such restriction is its reliance on linear processing, whereby data elements contained inside one cluster are not accessible in another. However, in real-time, the likelihood of mutual occurrences is significant. The k-means method often converges to local minima rather than global minima, influencing the final cluster by the initial random selection. The minimal migration of data points towards the cluster centers in each iteration directly impacts the time complexity of the entire process [32]. However, by incorporating local information through fuzzy clustering, k-means can capture complex cluster structures and accurately identify clusters even in overlapping or irregularly shaped clusters. The proposed method, Fuzzy Informative k-means, is presented as follows.

$$lin_{ji} = \frac{1}{d_{ij}} \left(Max\,(x) - C_c i \right)^f \tag{9}$$

The equation (9) denotes the Local Information parameter lin_{ji} of a specific cluster. The calculation involves utilizing the inverse Euclidean distance between the data points and the cluster centers. The primary purpose of this function is to address the global optimum issue. It is additionally utilized as an adjustment factor to facilitate the movement of cluster centers, hence reducing the overall number of iterations required.

$$NC_c i = \left(\frac{1}{ndc_i} \right) \sum_{j=1}^{ndc_i} M^f i + lin_{ji} \tag{10}$$

The center of the previous membership function in FIKM, represented by equation (10), aligns with the standard k-means algorithm mentioned in equation (4). The algorithm introduces the parameter to enhance the selection of cluster centres and prevent entrapment in local optima.

$$M^f k = X_s k \quad \forall\ W_k \tag{11}$$

where $W_{ij} = [0|\frac{1}{nc}|\frac{2}{nc}|\frac{3}{nc}|\frac{4}{nc}, \ldots \frac{n-1}{nc}]$, *n* and *nc* represent, the number of clusters determined by the user's request. The weight, denoted as W, is assigned to each cluster member to quantify the distinction between individual elements. Furthermore, equation 11 represents the membership function in FIKM, encompassing clusters categorized by the predetermined $C_c i$. Furthermore, the computation $NC_c i$ relies on $C_c i$. An inherent deviation from the conventional k-means algorithm lies in the fuzzy attribute incorporated into the FIKM framework. The membership function $M^f k$ represents a cluster containing members, some of which might have mutual membership with other clusters. In the given model, each element within a cluster holds the potential to be a member of multiple clusters.

$$X_s k = \lim_{j=1_{tonx}} x_j;\ if\ d_{ij}\ of\ x_j\ is\ \min\ to\ C_c k \tag{12}$$

$$d_{ij} = \left\| x_j - C_c i \right\|^2;\ j = 1\ to\ nx, \quad i = 1\ to\ nc \tag{13}$$

Equations (12) and (13) are employed to determine the data selection process within each cluster. This involves measuring the distance between the data elements and the cluster centers. Once the data is identified as belonging to a specific cluster, it is highly prioritized. Conversely, non-members are assigned a lower priority.

$$Obj_{fiter} = \sum_{i=1}^{nc} \sum_{j=1}^{nx} M^f id_{ij} + lin_{ji} \tag{14}$$

Equation (14) represents the objective function, assessing the degree of convergence attained through the recursive process. This function employs the current membership functions and the local information parameter to forecast the convergence state. Hence, the fuzzy approach is executed on the k-means algorithm, demonstrating the algorithm's systematic procedure below.

1. Initialize the required operative inputs, such as the number of clusters c, convergence criteria, and data elements.
2. The cluster centers are randomly from the data points for the first iteration. These cluster centers are indicated using $C_i = [C_1, C_2, C_3 \ldots . C_{nc}]$. Once the random cluster centers are chosen from the data vector x_j, they will be utilized to calculate the d_{ij} distance, where measuring methods vary depending on the precision and type of the cluster.
3. The membership data, denoted as $M^f k$ and containing fuzzy weights, is constructed using random clusters from preceding iterations. The cluster priority influences the determination of (fuzzy weights). The algorithm yields only one output- the membership function- specifically when it reaches the convergence stage.
4. Assess convergence after each iteration to compute the objective function using equation 15.

$$Obj_{fiter} - Obj_{fiter-1} < C_{gnc_{cri}} \tag{15}$$

   where $C_{gnc_{cri}}$ shows the convergence criteria for the D.R. detection.
5. Conclude the iteration upon reaching the convergence state and choose the final membership function from the clusters. Alternatively, if convergence is not achieved, repeat the iteration from Step 2.

The reason for choosing the Fuzzy Information K-Means over alternative methods is its capacity to effectively manage noisy and high-dimensional data while offering robustness against outliers. Its resilience to outliers ensures that noisy data points do not disproportionately impact clustering outcomes. Moreover, this method balances the interpretability of traditional k-means and the flexibility of fuzzy clustering, rendering it suitable for diverse real-world applications where data often demonstrates intricate patterns and noise.

D. MODIFIED FUZZY C-MEANS

The FCM method employs a soft clustering approach, distinguishing itself from k-means by its capability to identify mutual clusters where data points may belong to more than one cluster. FCM is an iterative process aimed at minimizing the objective function. Convergence is assessed by analyzing the reduction in mean square error. As FCM operates under the unsupervised learning paradigm, it initiates the clustering process with random selections [33], [34]. The FCM process involves three stages: (i) Cluster Centre, (ii) Membership Function, and (iii) Objective Function shown below. The objective function of the FCM is given by equation 16.

$$J_{iter} = \sum_{i=1}^{N} \sum_{j=1}^{nc} \left[M^f \left\| D_{x_i - c_j} \right\| \right] \tag{16}$$

where N is the number of data points in the data vector and nc the number of clusters (the user can specify this)? M^f denotes a membership function provided using equation 17.

$$M^f \left(\frac{d_{ji}}{d_{ki}} \right) = M_{iter} = \frac{1}{\left[\sum_{k=1}^{nc} \left(\frac{d_{ji}}{d_{ki}} \right)^{\frac{2}{f-1}} \right]} \tag{17}$$

where d_{ji} and d_{ki} represent the distance measurements between the data point D_{xi} and the cluster center C_j, which can be represented explicitly using the Euclidean distance method. The fuzzification factor f is a parameter used in the process of fuzzification. The cluster center C_j (equation 18) is often obtained from the M^f, which represents the specified cluster.

$$C_j = \frac{\sum_{i=1}^{N} M^f D_{xi}}{\sum_{i=1}^{N} M^f} \tag{18}$$

Despite FCM yielding better results, the limitation arising from local optima amplifies the time complexity required to achieve convergence. In certain instances, convergence may not be attained. Given the nature of the problem, numerous algorithms have been proposed based on FCM to achieve global optima with improved time complexity. In order to improve the robustness of the clustering method, various improvements have been proposed for existing versions of FCM. These improvements include facilitation techniques such as fuzzification, local optima minimization, contour region similarity assessment, noise tolerance, reduction in execution time, and enhanced parameter selection. The proposed MaFCM algorithm incorporates the median adjustment parameter. Adding a median adjustment parameter can help achieve the global optima by introducing a penalty term F_{mmfi} that encourages cluster centroids to move towards the median of the cluster instead of being strictly determined by the data points. Moreover, incorporating the median adjustment parameter influences the centroids by the cluster median rather than just the mean. Thus, adding the median adjustment parameter in the FCM objective function enhances the algorithm's robustness, stability, and convergence properties, ultimately leading to better clustering results, especially in noisy or sparse data scenarios.

$$J_{iter} = \sum_{i=1}^{N} \sum_{j=1}^{nc} \left[M^f \left\| D_{x_i} - C_j \right\|^2 + F_{mmfi} \right] \tag{19}$$

$$F_{mmfi} = \sum_{i \in N_i} \frac{Med(W_{ij})}{Med\left(W_{ij}\right) + 1} \left\| D_{x_i} - C_j \right\|^2 \tag{20}$$

$$C_{pre_j} = \frac{\sum_{i=1}^{N} M^f D_{x_i}}{\sum_{i=1}^{N} M^f} \tag{21}$$

$$C_{adjM} = Med(\left\| C_{old_j} - C_{pre_j} \right\|) \tag{22}$$

$$C_{adj} = \frac{C_{adjM}}{Max(P_v)} \tag{23}$$

$$C_j = (C_{pre_j})^{\frac{1}{C_{adj}}} \tag{24}$$

$$M_{iter} = \frac{1}{\sum_{k=1}^{nc} \left(\frac{d_{ji} + F_{mmfi}}{d_{ki} + F_{mmfi}} \right)^{\frac{1}{f-1}}} \tag{25}$$

As demonstrated in equations (19)-(25), the Median adjustment parameter modifies the derivative form of an objective function, cluster center, and membership function F_{mmfi}. This F_{mmfi} is the only factor accountable for making the algorithm optimal and effective. The following are the steps for running this algorithm.

1. The FCM algorithm begins by initializing parameters, such as the desired number of clusters, the fuzzy factor, convergence criteria (indicating the convergence status), and the maximum allowable iterations.
2. In order to streamline the clustering process, the data utilized is referred to as the Data vector D_{x_j}. This vector is subjected to fuzzification and afterward reorganized according to a random membership function.
3. The dimension *Nxnc* is utilized to generate random memberships, and its variants exhibit a soft clustering nature, implying the mutual association of data points in two or more clusters. Here, N the total number of data points available is represented, and the user-determined number of clusters is utilized.
4. As the procedure is iterative, the membership function M^f is calculated at each iteration using the latest cluster center C_j. A new is computed if the convergence state is not achieved based on this calculation.
5. The proposed version of the FCM algorithm distinguishes itself from other iterations by employing a distinct method for calculating its cluster center C_j in contrast to C_{prej} wherein the membership function M^f is a crucial factor. One of the primary objectives of this approach is to minimize iterations, thereby achieving the shortest possible convergence time.
6. The distance between the prior cluster center and the present cluster is computed as C_{adjM} per equation (18) to promote better parameter selection and faster convergence.
7. The C_{pre_j} is then rectified to generate the final cluster center C_j. The F_{mmfi} article is updated with new information, stimulating the nature of the ideal result. The procedure is repeated until the convergence state is attained.

IV. RESULT AND DISCUSSION

In this research, the focus is given to the improvements of existing clustering algorithms to find a perfect match for unsupervised methodology. The results section showcases the performance evaluation of the proposed clustering algorithms and provides a comparative analysis among them. Moreover, these algorithms (MaFCM and FIKM) are compared with existing clustering algorithms or older versions of FCM and K-Means to stimulate the results better. The findings are evaluated using standard IDRiD, Kaggle competition dataset, and primary dataset fundus images obtained from Shiva Netralaya Centre Saharanpur, India.

The available techniques are compared by presenting numerical results with varying noise levels. Traditional algorithms such as k-means, FCM, EnFCM (Enhanced Fuzzy C-Means), FGFCM (Fast Generalised Fuzzy C-means), and FLICM (Fuzzy Local Information C Means) are tested in this part under noisy and original conditions to evaluate the proposed approach. The parameter for analysis is chosen based on the limitations of existing clustering algorithms. Parameters such as accuracy and time complexity are the main constraints necessary for the decision to enhance the proposed algorithm in the context of achieving better segmentation and efficiently reaching global optima.

The proposed algorithms can extract maximum information from fundus images as per the results. This examination is tailored for around 850 fundus Images sourced from diverse databases. Figure 3 shows the input fundus images from the functional dataset, as illustrated in the description. The evaluation is conducted on original fundus images. The results indicate that MaFCM, followed by EnFCM and FIKM, can extract maximum information from the fundus image. This assessment was conducted on 25% of the medical images from various databases. A subset of the results (S1-S15) is presented in Table 4, and a graphical representation of the performance analysis is shown in Figure 5.

TABLE 4. Comparison of clustering algorithms for accuracy under noiseless condition.

Sample	k-means	FIKM	FCM	EnFCM	FLICM	FGFCM	MaFCM
S1	83.86	90.03	80.52	93.69	85.57	87.62	96.57
S2	82.38	89.40	79.53	90.23	85.68	87.38	97.97
S3	83.26	89.63	79.42	81.11	84.19	86.60	95.02
S4	84.28	91.55	80.63	83.38	84.60	89.40	97.41
S5	81.57	90.23	79.26	82.97	83.21	86.77	96.03
S6	82.40	88.70	79.02	81.32	83.27	88.02	98.34
S7	82.50	89.03	78.63	85.20	85.10	86.52	99.03
S8	82.73	88.90	77.63	79.27	83.64	85.14	95.99
S9	83.91	91.51	81.34	79.30	86.96	87.54	96.73
S10	82.54	88.12	79.45	82.63	84.27	85.61	98.14
S11	81.26	90.65	80.26	81.53	84.78	87.67	96.32
S12	82.28	89.72	81.02	83.54	85.29	88.02	98.47
S13	81.76	90.03	78.83	81.98	83.60	86.62	97.03
S14	82.24	88.55	79.23	82.23	84.21	86.14	98.64
S15	83.65	91.87	81.34	86.20	85.61	87.54	98.03

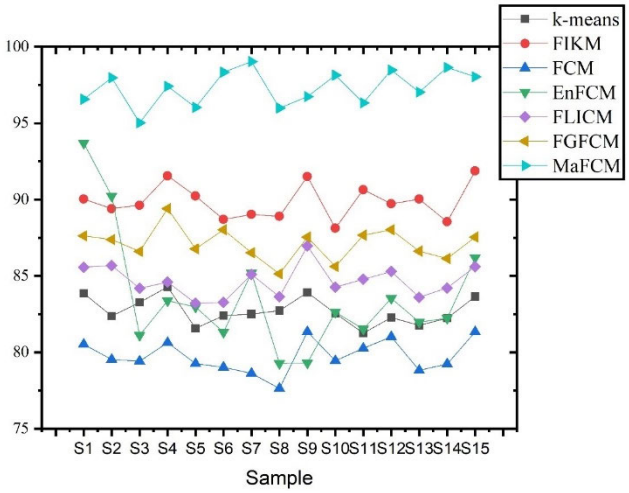

FIGURE 5. Graphical representation of performance analysis of clustering algorithms under noiseless conditions, indicating that MaFCM outperforms other state-of-the-art algorithms.

The k-means algorithm is excluded from the comparison because it performs hard clustering, unlike the fuzzy version (FCM). The FIKM, incorporating a fuzzy median strategy, stands out for its unique ability to obtain a median value from neighboring pixels, improving average accuracy. Table 5 compares various clustering algorithms in terms of accuracy under noisy conditions. To assess algorithm characteristics, [35] manipulated the image's noise level (variance), denoted as σ, representing the intensity of noise affecting the data. Handling pixel contamination in an image poses a challenge for categorizing or clustering intricate data, as noises can manifest at different frequencies—high or low—impacting real-time data. This experiment focuses on analyzing high-frequency noise. The study includes four adjustable noise levels from 0 to 1: 0.2, indicating a 20% infection rate of information; 0.4, indicating a 40% infection rate; 0.6, indicating a 60% infection rate; and 0.8, indicating an 80% infection rate. The graphical representation of performance analysis with noise variance is shown in Figure 6.

TABLE 5. Comparison of clustering algorithms for accuracy under noisy condition.

σ	k-means	FIKM	FCM	EnFCM	FLICM	FGFCM	MaFCM
0.2	91.65	96.64	92.05	92.29	94.13	87.62	97.63
0.4	85.56	92.45	86.53	86.82	89.12	87.64	93.4
0.6	77.13	87.70	81.28	81.31	85.93	82.52	90.19
0.8	62.86	81.76	67.73	67.82	72.37	67.87	83.62

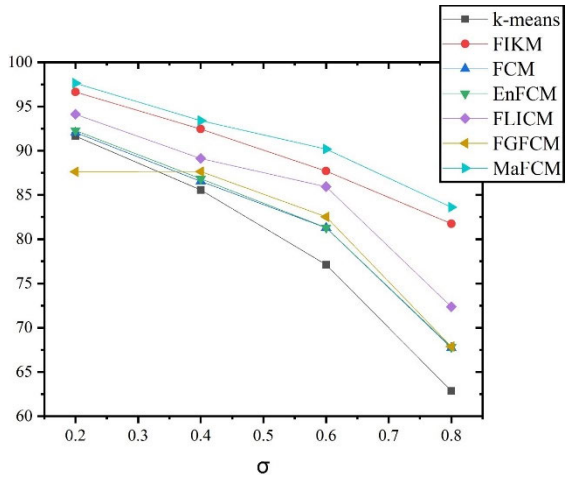

FIGURE 6. Graphical representation of performance analysis of clustering algorithms under noisy conditions, indicating that MaFCM outperforms other state-of-the-art algorithms.

In the given configuration featuring an Intel i7 fifth-generation processor with a 32GB VRAM GPU and 49GB RAM, it is crucial to consider the computational resources available for efficient deep learning algorithm simulations. Deep learning algorithms rely heavily on GPUs or high-computation facilities to expedite learning. Given this context, it becomes imperative that the clustering algorithm employed for data processing operates within acceptable time constraints. Table 6 provides insights into the average time taken by the algorithm during the simulation. Typically, clustering algorithms make critical decisions in the first iteration, such as selecting their initial points. It is paramount that the processing time of the clustering algorithm aligns with the system's capacity, ensuring that the overall learning process remains within acceptable time limits. Consequently, a comparative analysis of these metrics becomes essential to establish a standardized measure, as shown in Figure 7.

TABLE 6. Average time taken by clustering algorithms.

Run no.	k-means	FIKM	FCM	EnFCM	FLICM	FGFCM	MaFCM
1	15.8	17.32	21.45	20.69	17.52	17.67	17.11
2	15.34	18.45	21.11	21.13	17.18	17.33	17.92
3	16.81	18.12	20.81	20.54	18.23	18.20	18.02
4	15.84	17.43	21.23	20.92	18.60	18.11	16.93
5	15.87	18.20	20.89	20.28	17.92	18.43	17.86

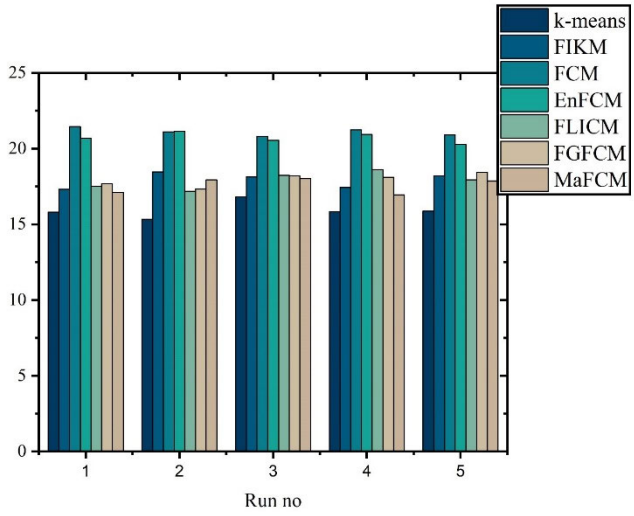

FIGURE 7. Graphical representation of comparative analysis of time complexities for clustering algorithms, indicating that MaFCM followed by FLICM outperforms other state-of-the-art algorithms.

The primary objective of this study is to enhance the capabilities of unsupervised learning methodologies. Several challenges include feature corruption due to the insecurity of clustering loss, model complexity leading to implementation and training issues, hyper-parameter tuning, and overlapping clusters.

The comparison focuses on the accuracy of unsupervised clustering, denoted as ACC and expressed by equation (26). Additionally, the discussion involves the Normalized Mutual Information (NMI), represented by equation (27), and

the utilization of the Adjusted Rand Index (ARI) as per equation (28). This approach facilitates a robust evaluation framework, enabling conclusive findings in pursuing improved unsupervised learning outcomes.

A. UNSUPERVISED LEARNING ACCURACY

Unsupervised clustering accuracy (ACC) is a measure that assesses the performance of clustering algorithms by comparing the obtained clusters with the ground truth without using any labeled information. The ACC is calculated using the following equation (26). The ACC ranges from 0 to 1, where 1 indicates a perfect clustering, meaning all data points are correctly assigned to their respective clusters. A lower ACC value suggests a less accurate clustering.

$$ACC = \frac{Number\ of\ correctly\ clustered\ data\ points}{Total\ number\ of\ data\ points} \quad (26)$$

In the given equation (26), "Number of correctly clustered data points" refers to the count of data points that are assigned to the correct clusters based on the ground truth, and "Total number of data points" is the overall number of data points in the dataset.

B. NORMALIZED MUTUAL INFORMATION (NMI)

Normalized Mutual Information (NMI) is a measure that quantifies the similarity between two clustering's while considering the entropy of the clusters [36]. It is often used in unsupervised learning to evaluate the quality of a clustering algorithm. The NMI is calculated using the equation (27). The NMI ranges from 0 to 1, where 0 indicates no mutual information (no similarity between the clustering), and 1 indicates perfect agreement between the clustering. A higher NMI value implies a better agreement between the algorithm's clustering and the ground truth.

$$NMI = \frac{2 \times I(C, K)}{H(C) + H(K)} \quad (27)$$

The given equation (27) represents the mutual information between the two clustering's K. It measures the amount of information shared between the two clustering's.

$H(C)$ and $H(K)$ are the entropies of clustering C K used to quantify the uncertainty or disorder within each clustering.

C. UNSUPERVISED LEARNING ACCURACY

Adjusted Rand Index (ARI) is another measure to assess the similarity between two clustering's. It considers pairs of samples and measures the agreement between the actual and predicted cluster assignments, correcting for chance [37]. The ARI is calculated using the equation (28). The ARI ranges from -1 to 1, where -1 indicates perfect disagreement between the clustering's, 0 indicates the agreement expected by chance, and 1 indicates perfect agreement between the clustering's.

$$ARI = \frac{RI - Expected\_RI}{\max(RI - Expected_{RI}) - Expected\_RI} \quad (28)$$

The given equation (28) stands for the Rand Index, which is the ratio of the number of samples $Expected\_RI$ to the expected value of the Rand Index under a null hypothesis of random cluster assignments and max $(RI - Expected_{RI})$ shows the maximum possible value of the ARI.

The MaFCM stands out due to its incorporation of a fuzzy median technique, providing a unique capability to derive a median value from a window of neighboring pixels, thereby achieving superior average accuracy. The proposed technique is designed to extract information even from minor changes in the data neighborhood, enabling the efficient construction of clusters. Moreover, the FIKM algorithm demonstrates resistance to segmentation and is primarily focused on global optima. Thus, as time progresses, the significance of these diverse approaches becomes evident.

Compared to existing methods such as K-Means, FCM, EnFCM, and FLICM, the proposed algorithms are evaluated as shown in Table 4 and Table 5. The findings indicate that these algorithms specialize in various areas and operations; some excel in rapid clustering, while others excel in cluster deletion, addressing noise during their processes.

Tables 7,8, and 9 present detailed results generated parameter-wise (i.e., ACC, NMI, and ARI). The findings consistently favor the proposed methods, with MaFCM demonstrating promising results: ACC = 94.4, NMI = 92.7, and ARI = 91.9. MaFCM outperforms FIKM in clustering data points and achieves a convergence state faster, as shown in Table 6.

TABLE 7. Comparative analysis of clustering algorithms for unsupervised learning accuracy on functional dataset.

Dataset	FIKM	MaFCM
Kaggle	83.8	94.4
IDRiD	88.4	89.2
Primary	89.3	91.3

TABLE 8. Comparative analysis of clustering algorithms for NMI on functional dataset.

Dataset	FIKM	MaFCM
Kaggle	91.4	92.7
IDRiD	83.6	84.6
Primary	90.7	96.9

TABLE 9. Comparative analysis of clustering algorithms for ARI on functional dataset.

Dataset	FIKM	MaFCM
Kaggle	82.8	91.9
IDRiD	79.5	90.2
Primary	80.2	82.5

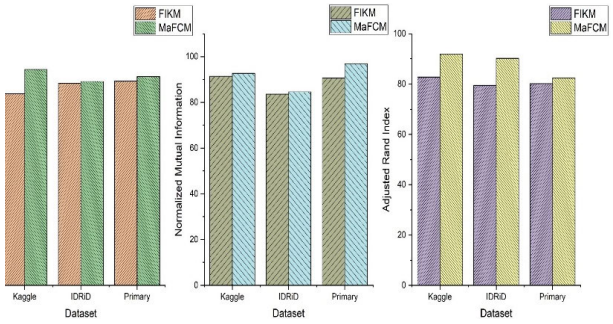

FIGURE 8. Graphical representation of comparative analysis of ACC, NMI, and ARI indicating that FIKM outperforms the MaFCM for D.R. detection.

The Kaggle dataset, a well-known dataset with many data points, demonstrates the highest performance for both algorithms. Additionally, the primary dataset yields superior results for both proposed methods. Figure 8 presents the graphical comparison of proposed methods for ACC, NMI, and ARI performance measures.

During experimentation with the proposed algorithms, several outcomes and potential limitations emerge. The algorithm typically yields more flexible and accommodating clusters to complex data structures, especially in scenarios with noisy or high-dimensional data. However, challenges may arise in determining the optimal number of clusters and selecting appropriate fuzzification parameters, which can significantly impact the clustering results. Additionally, the algorithm's performance may be sensitive to the initialization of cluster centers and the choice of distance metric, requiring careful tuning for optimal outcomes. Furthermore, while the algorithm offers robustness against outliers, excessively skewed distributions or heavily overlapping clusters may pose challenges. Addressing these limitations often involves iterative experimentation, parameter tuning, and robust validation techniques to ensure the algorithm's effectiveness across diverse datasets and scenarios.

V. CONCLUSION

DR is a severe chronic disease affecting approximately one-third of diabetic patients worldwide. Developing an automatic D.R. detection system based on retinal fundus images is imperative to address this widespread concern. However, labeling data for processing through supervised learning methods in medical imaging can be time-consuming. Consequently, unsupervised clustering methods offer an alternative, revealing hidden patterns and relationships in the data. The proposed work introduces an optimized clustering algorithm by incorporating fuzzy local parameters into K-Means algorithms to enhance performance and achieve global optima. The proposed algorithm demonstrates an accuracy rate of 94.4% and an average execution time of 17.11 seconds. Evaluation under noisy and noiseless conditions on different image samples indicates superior performance in noisy conditions, achieving 97% accuracy.

FUTURE SCOPE AND OPEN CHALLENGES

There is still a need for improvement in the existing approach to enhance performance in the field of D.R. detection. Thus, this section addresses the future scope and open challenges researchers should focus on to enhance D.R. classification performance. The challenges that need to be addressed include-

1. Real-time Image Analysis: Developing algorithms capable of analyzing retinal images in real-time, allowing for faster diagnosis and intervention, particularly in telemedicine and remote healthcare settings.
2. Automated Screening in Primary Care: Implementing automated D.R. screening tools in primary care settings to improve access to early detection and intervention, especially in underserved communities with limited access to eye care specialists.
3. Deep Learning on Small Datasets: Addressing the challenge of developing accurate deep learning models for D.R. detection with limited data availability by exploring transfer learning, data augmentation, and semi-supervised learning techniques.
4. Ethical A.I. Deployment: Ensuring ethical deployment of AI-based DR detection systems by addressing bias, fairness, privacy, and transparency in algorithmic decision-making processes to maintain trust and equity in healthcare delivery.
5. Mobile Health Solutions: Leveraging mobile technologies, such as smartphone-based retinal imaging devices and mobile apps, for scalable and cost-effective D.R. screening and monitoring, enabling early detection and timely intervention in resource-constrained settings.

DATA AVAILABILITY STATEMENT

DIARETdb1 dataset used in this study can be accessed at https://paperswithcode.com/dataset/diaretdb1. The IDRiD dataset is available at https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid. Interested researchers can request access from the corresponding author for the primary dataset used in this study.

FUNDING

This research was funded by Princess Nourah bint Abdulrahman University and Researchers Supporting Project number (PNURSP2024R346), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ACKNOWLEDGMENT

This research was supported by Princess Nourah bint Abdulrahman University and Researchers Supporting Project number (PNURSP2024R346), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

The authors would also like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

[1] M. R. Shoaib, H. M. Emara, J. Zhao, W. El-Shafai, N. F. Soliman, A. S. Mubarak, O. A. Omer, F. E. A. El-Samie, and H. Esmaiel, "Deep learning innovations in diagnosing diabetic retinopathy: The potential of transfer learning and the DiaCNN model," *Comput. Biol. Med.*, vol. 169, Feb. 2024, Art. no. 107834.
[2] S. Akhtar, A. Ali, S. Ahmad, M. I. Khan, S. Shah, and F. Hassan, "The prevalence of foot ulcers in diabetic patients in pakistan: A systematic review and meta-analysis," *Frontiers Public Health*, vol. 10, Oct. 2022, Art. no. 1017201.
[3] I. A. Manarvi and N. M. Matta, "Investigating information needs of Saudi diabetic patients," *Current Diabetes Rev.*, vol. 15, no. 2, pp. 149–157, Jan. 2019.
[4] A. C. Maritim, R. A. Sanders, and J. B. Watkins, "Diabetes, oxidative stress, and antioxidants: A review," *J. Biochem. Mol. Toxicol.*, vol. 17, no. 1, pp. 24–38, Jan. 2003, doi: 10.1002/jbt.10058.
[5] S. Joshi and P. T. Karule, "Detection of hard exudates based on morphological feature extraction," *Biomed. Pharmacol. J.*, vol. 11, no. 1, pp. 215–225, Mar. 2018, doi: 10.13005/bpj/1366.
[6] D. Nagpal, S. N. Panda, M. Malarvel, P. A. Pattanaik, and M. Zubair Khan, "A review of diabetic retinopathy: Datasets, approaches, evaluation metrics and future trends," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7138–7152, 2021, doi: 10.1016/j.jksuci.2021.06.006.
[7] J. Enguehard, P. O'Halloran, and A. Gholipour, "Semi-supervised learning with deep embedded clustering for image classification and segmentation," *IEEE Access*, vol. 7, pp. 11093–11104, 2019, doi: 10.1109/ACCESS.2019.2891970.
[8] A. Kumar, L. Bi, J. Kim, and D. D. Feng, *Machine Learning in Medical Imaging*. Amsterdam, The Netherlands: Elsevier, 2019.
[9] N. Hatipoglu and G. Bilgin, "Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships," *Med. Biol. Eng. Comput.*, vol. 55, no. 10, pp. 1829–1848, Oct. 2017, doi: 10.1007/s11517-017-1630-1.
[10] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Inf. Sci.*, vol. 250, pp. 113–141, Nov. 2013, doi: 10.1016/j.ins.2013.07.007.
[11] S. Fogel, H. Averbuch-Elor, D. Cohen-Or, and J. Goldberger, "Clustering-driven deep embedding with pairwise constraints," *IEEE Comput. Graph. Appl.*, vol. 39, no. 4, pp. 16–27, Jul. 2019, doi: 10.1109/MCG.2018.2881524.
[12] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep learning techniques for diabetic retinopathy classification: A survey," *IEEE Access*, vol. 10, pp. 28642–28655, 2022, doi: 10.1109/ACCESS.2022.3157632.
[13] H. Naz, R. Nijhawan, and N. J. Ahuja, "An automated unsupervised deep learning–based approach for diabetic retinopathy detection," *Med. Biol. Eng. Comput.*, vol. 60, no. 12, pp. 3635–3654, Dec. 2022, doi: 10.1007/s11517-022-02688-9.
[14] M. Kampffmeyer, S. Løkse, F. M. Bianchi, L. Livi, A.-B. Salberg, and R. Jenssen, "Deep divergence-based approach to clustering," *Neural Netw.*, vol. 113, pp. 91–101, May 2019, doi: 10.1016/j.neunet.2019.01.015.
[15] T. Saba, S. T. F. Bokhari, M. Sharif, M. Yasmin, and M. Raza, "Fundus image classification methods for the detection of glaucoma: A review," *Microsc. Res. Technique*, vol. 81, no. 10, pp. 1105–1121, Oct. 2018, doi: 10.1002/jemt.23094.
[16] S. Larabi-Marie-Sainte, L. Aburahmah, R. Almohaini, and T. Saba, "Current techniques for diabetes prediction: Review and case study," *Appl. Sci.*, vol. 9, no. 21, p. 4604, Oct. 2019, doi: 10.3390/app9214604.
[17] J. Ganesan, A. T. Azar, S. Alsenan, N. A. Kamal, B. Qureshi, and A. E. Hassanien, "Deep learning reader for visually impaired," *Electronics*, vol. 11, no. 20, p. 3335, Oct. 2022, doi: 10.3390/electronics11203335.
[18] Y. Ren, K. Hu, X. Dai, L. Pan, S. C. H. Hoi, and Z. Xu, "Semi-supervised deep embedded clustering," *Neurocomputing*, vol. 325, pp. 121–130, Jan. 2019, doi: 10.1016/j.neucom.2018.10.016.
[19] W. Wang, D. Yang, F. Chen, Y. Pang, S. Huang, and Y. Ge, "Clustering with orthogonal AutoEncoder," *IEEE Access*, vol. 7, pp. 62421–62432, 2019, doi: 10.1109/ACCESS.2019.2916030.
[20] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: From the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39501–39514, 2018, doi: 10.1109/ACCESS.2018.2855437.
[21] T. Melo, A. M. Mendonça, and A. Campilho, "Microaneurysm detection in color eye fundus images for diabetic retinopathy screening," *Comput. Biol. Med.*, vol. 126, Nov. 2020, Art. no. 103995, doi: 10.1016/j.compbiomed.2020.103995.
[22] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy C-means algorithms for very large data," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 6, pp. 1130–1146, Dec. 2012, doi: 10.1109/TFUZZ.2012.2201485.
[23] T. P. Lillicrap and K. P. Kording, "What does it mean to understand a neural network?" 2019, *arXiv:1907.06374*.
[24] A. T. Azar, Z. I. Khan, S. U. Amin, and K. M. Fouad, "Hybrid global optimization algorithm for feature selection," *Comput., Mater. Continua*, vol. 74, no. 1, pp. 2021–2037, 2023, doi: 10.32604/cmc.2023.032183.
[25] G. Atteia, E. S. M. El-kenawy, N. A. Samee, M. M. Jamjoom, A. Ibrahim, A. A. Abdelhamid, A. T. Azar, N. Khodadadi, R. A. Ghanem, and M. Y. Shams, "Adaptive dynamic dipper throated optimization for feature selection in medical data," *Comput., Mater. Continua*, vol. 75, no. 1, pp. 1883–1900, 2023, doi: 10.32604/cmc.2023.031723.
[26] D. C. Hoang, R. Kumar, and S. K. Panda, "Realisation of a cluster-based protocol using fuzzy C-means algorithm for wireless sensor networks," *IET Wireless Sensor Syst.*, vol. 3, no. 3, pp. 163–171, 2013, doi: 10.1049/iet-wss.2012.0132.
[27] E. Dupont, "Learning disentangled joint continuous and discrete representations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 710–720.
[28] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, Jul. 2018, doi: 10.3390/data3030025.
[29] R. Pires, S. Avila, J. Wainer, E. Valle, M. D. Abramoff, and A. Rocha, "A data-driven approach to referable diabetic retinopathy detection," *Artif. Intell. Med.*, vol. 96, pp. 93–106, May 2019, doi: 10.1016/j.artmed.2019.03.009.
[30] H. Liu, L. Teng, L. Fan, Y. Sun, and H. Li, "A new ultra-wide-field fundus dataset to diabetic retinopathy grading using hybrid preprocessing methods," *Comput. Biol. Med.*, vol. 157, May 2023, Art. no. 106750, doi: 10.1016/j.compbiomed.2023.106750.
[31] M. Nahiduzzaman, M. Robiul Islam, M. Omaer Faruq Goni, M. Shamim Anower, M. Ahsan, J. Haider, and M. Kowalski, "Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119557, doi: 10.1016/j.eswa.2023.119557.
[32] M. Hassan, A. Chaudhry, A. Khan, and J. Y. Kim, "Carotid artery image segmentation using modified spatial fuzzy C-means and ensemble clustering," *Comput. Methods Programs Biomed.*, vol. 108, no. 3, pp. 1261–1276, Dec. 2012, doi: 10.1016/j.cmpb.2012.08.011.
[33] W. Zhou, C. Wu, D. Chen, Y. Yi, and W. Du, "Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method," *IEEE Access*, vol. 5, pp. 2563–2572, 2017, doi: 10.1109/ACCESS.2017.2671918.
[34] C. Birtolo and D. Ronca, "Advances in clustering collaborative filtering by means of fuzzy C-means and trust," *Expert Syst. Appl.*, vol. 40, no. 17, pp. 6997–7009, Dec. 2013, doi: 10.1016/j.eswa.2013.06.022.
[35] K. S. Chuang, H. L. Tzeng, S. Chen, J. Wu, and T. J. Chen, "Fuzzy C-means clustering with spatial information for image segmentation," *Computerized Med. Imag. Graph.*, vol. 30, no. 1, pp. 9–15, 2006, doi: 10.1016/j.compmedimag.2005.10.001.
[36] M. Bennasar, Y. Hicks, and R. Setchi, "Feature selection using joint mutual information maximisation," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8520–8532, Dec. 2015, doi: 10.1016/j.eswa.2015.07.007.
[37] M. Hoffman, D. Steinley, and M. J. Brusco, "A note on using the adjusted Rand index for link prediction in networks," *Social Netw.*, vol. 42, pp. 72–79, Jul. 2015, doi: 10.1016/j.socnet.2015.03.002.

HUMA NAZ received the master's degree from the Department of Computer Science and Engineering, Chitkara University. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, University of Petroleum and Energy Studies (UPES), Dehradun. Her research interests include unsupervised deep learning, data analytics, and machine learning the healthcare applications. She has published more than eight research articles in reputable SCI and Scopus-indexed journals. Other than that, she has filed five patents.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently a Full Professor with the College of Computer and Information Sciences, Prince Sultan University (PSU), Riyadh, Saudi Arabia, and also a Leader of the AIDA Laboratory. She has published over 300 publications in high-ranked journals. Her primary research interests include bioinformatics and data mining and classification using AI models. She received the Best Student Award from the Faculty of Computing, UTM, in 2012; and the Best Researcher Award from PSU, from 2013 to 2016. She is an editor of several reputed journals and on a panel of TPC of international conferences.

FATEN S. ALAMRI received the Ph.D. degree in system modeling and analysis in statistics from Virginia Commonwealth University, USA, in 2020. Her Ph.D. research included Bayesian dose-response modeling, experimental design, and nonparametric modeling. She is currently an Assistant Professor with the Department of Mathematical Sciences, College of Science, Princess Nourah Bint Abdulrahman University. Her research interests include spatial area, environmental statistics, and brain imaging.

AHMED S. ALMASOUD received the degree from the University of Technology Sydney. Since 2014, he has been with Prince Sultan University (PSU), Riyadh, Saudi Arabia, where he is currently an Assistant Professor with the College of Computer and Information Sciences. He has published original articles in the finest journals in the area of his studies. His research interests include (but not limited to) artificial intelligence, machine learning, security architecture, and the Internet of Things.

AMJAD REHMAN (Senior Member, IEEE) received the Ph.D. degree from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, specializing in information security using image processing techniques, in 2010. He is currently an Associate Professor with CCIS, Prince Sultan University, Riyadh, Saudi Arabia. He is also a PI in several projects and completed projects funded by MoHE Malaysia, Saudi Arabia. His research interests include bioinformatics, the IoT, information security, and pattern recognition. He received a Rector Award for the 2010 Best Student from UTM Malaysia.