An Improvised Fuzzy-C Means Clustering Based Optimization Using Map Reducing Model in Big Data

Venkat Rayala
Research Scholar, Research Centre, Department of CSE, Cambridge Institute of Technology, Affiliated to Visvesvaraya Technological University, (VTU)-Belgaum, Bengaluru 560036, India, rayalavenkat534@gmail.com

Satyanarayan Reddy Kalli
Professor, Research Centre, Department of CSE, Cambridge Institute of Technology, Affiliated to Visvesvaraya Technological University, (VTU)-Belgaum, Bengaluru 560036, India

Abstract

Big data is gaining ground in many sectors, like industries, financial affairs, health etc. Since they can deal with huge volumes of data. Any real-world data may be appropriately organized by clustering the data using certain cluster methods, which in this clustering approach is a highly innovative & Improvised Fuzzy C-Means (IFCM) technique that can frame the data with excellent logic and very accurately. The MapReduce model is one of the most often and effectively utilized mining techniques for categorizing enormous amounts of data. In order to effectively process large amounts of data, this article combines the Stochastic Social Group Optimization (SSGO) with the MapReduce Model. The required outcomes were obtained by picking the best candidate solutions and arranging them into a reduction structure in order to acquire superior solutions. Ultimately, for each data sample, the recommended SSGO approach is used, which is based on categorizing with probable index values using succeeding possibility of data. For evaluation, the suggested technique is compared to three metrics: Sensitivity, Specificity, and Accuracy.

Key words: *IFCM, SSGO, Big Data, Map Reduce Model, Clustering.*

1 INTRODUCTION

Science and industrialization have enhanced the capacity for information in practically every subject of research and engineering, as well as in many applications [1]. Speed, diversity, and volume [2] are the three data attributes that big data [3] now possesses. The rate at which data is processed and created based on the applications required is referred to as speed, whilst the kind and type of data is referred to as variety. On identifies data volume for calculating data value [4]. Existing storing and processing technologies may be unable to handle the pace, variety, and volume of bulk data. The term "big data" describes this information. Analysing Big Data is a strategy for discovering important geometric and statistical patterns in enormous datasets. Besides of data storage and access, the massive rise in data causes a variety of processing issues. Since data collection is expensive, it is critical that the data be utilized properly so that more progress may be achieved by building more efficient algorithms.

Several commercial disciplines make use of massive amounts of data. Big data mining is

often difficult to handle with present technology and techniques due to the enormous and intricate data sets. Data mining on a single Computer involves high computational expenses for large datasets [5]. For the collection and processing of big data, more effective computer environments are consequently required. Big data necessitates the use of intelligent data analysis techniques that as image processing, automatic classification, and multi-time processes using combined data. Parallel strategies were developed to optimize available data in order to considerably accelerate computations. To solve the limits of the Big Data Analysis process, data mining approaches are adjusted to the evolving technology. Google created the MapReduce framework [6] to cope with Massive Data. The MapReduce approach, in combination with its distributed file system, provides a simple and robust framework for large-scale data analysis across a group of computers. MapReduce is a mechanism that maps and reduces functions. For Filtration and classifying the data are used for mapping, whereas the reduced based function performs a summarized function to get the output. Numerous research on the use of large-scale data mining to this technologies have been published, notably on characteristic reduction [7] class imbalance and case selection. Hence, large-scale data mining may be accomplished concurrently by using a variety of processor or computers terminals using traditional distributed techniques and MapReduce technologies. Each cluster contains things that are similar to elements from other groups in the cluster. For the grouping of literary topics, various clustering algorithms have been proposed. Conventional clustering techniques are divided into two categories: hierarchical and partial algorithms. In addition to traditional clustering algorithms, there are other clustering approaches. Overlapping clustering techniques differ from traditional clustering algorithms in that each item is treated as a separate cluster [8]. Each item may be in more than one cluster at the same time. Nevertheless, it has the problem of requiring non-deterministic present parameters. The advantage of an existing clustering algorithm is effective data processing.

In previous years, meta-heuristic approaches have been frequently employed to cope with clustering. Clustering issues may potentially be seen as NP-hard grouping problems from an optimum standpoint. The cutting-edge optimization purports to demonstrate that swarm intelligent algorithms may efficiently handle many design concerns, including exceedingly complex NP-hard problems. Yet, the majority of the case studies in the current literature deal with optimization issues with a few to a several hundred variables. In comparison to real-world applications, the dimensionality investigated is fairly minimal. Nevertheless, it is unclear if these algorithms can be directly applied to large-scale real-world problems. The full scalability has yet to be shown. Hence, high-dimensional optimization performance is a common hurdle to all present optimization approaches. Even with a non-complex cost function, the conclusions are far from ideal for the global function as dimensionality increases.

This work offers a novel optimization approach called social group optimization (SGO) focused on the human behavior of learning and solving complicated issues to solve a few obstacles such as computing efforts, optimum solutions, and consistency in giving optimal answers. Additionally, numerous behavioral attributes, such as caring, honesty, compassion, dishonesty, bravery, fear, fairness, and respect,

are desired to be cultivated in order to solve difficult situations [9]. By examining the varied features of everyone in the group, group solving ability may produce more active answers than solo capacity.

2 RELATED WORKS

This section presents the literary assessment, in which different approaches for the classification of enormous data in works are discussed along with problems. Also included in this section is an analysis of any flaws. They developed a distribution-based nearest neighbor classification model based on clusters for the purpose of doing various analyses more quickly and presented it in [10]. Map Reduced Method was established to execute process employing computing mechanism sample for optimizing the sampling process. The problem with dimensions continued to exist throughout the procedure, despite the fact that the approach delivered excellent accuracy decrease rates. Utilizing the method of multiplier optimization, the authors of [11] presented Echo State Networks as a means of carrying out neighborhood exchanges between components that are located closer together. In addition, training patterns for new nodes were not required in any way. The experiment performed on synthetic data showed improved performance; nonetheless, adding weights without first taking into consideration inaccurate value estimates is a drawback.

In this study, we present a method known as social group optimization (SGO), which is an optimization strategy that uses populations [12]. It derives its motivation from the notion of social conduct shown by humans in the process of resolving a difficult issue. In this article, a flowchart is used to describe both the conceptual framework of the SGO algorithm as well as its mathematical formulation.

The clever actions that are shown by groups of insects or animals in nature have ensured the continued existence of their species over the course of thousands of years. In this study, a new swarm intelligence method for addressing optimization problems dubbed the social group entropy optimization (SGEO) algorithm is developed [13]. SGEO stands for the social group entropy optimization. The primary contributions of this research are the social group model, the status optimization model, and the entropy model. These models serve as the foundation for the algorithm that has been suggested.

Identifying the best solutions to technical applied issues is necessary due to the presence of financial and physical restrictions; yet global optimization algorithms are unable to provide these solutions [14]. It is required to move between known various local and global solutions in order to achieve optimization that is both precise and quick. In the work [15] that was done recently, a social group optimization, or SGO, was offered as a means of tackling issues involving multimodal functions and data clustering.

Malignant is now one of the most prevalent forms of severe cancer in the human population. Melanomas are cancers that begin in the skin. As a direct result of this, there is a growing need for methods that are both automated and resilient in order to provide accurate and prompt clinical identification and detection of skin cancer [16]. A social group optimization (SGO) assisted automated technique was created for the purpose of analyzing dermoscopy pictures for signs of skin melanoma in this present body of work.

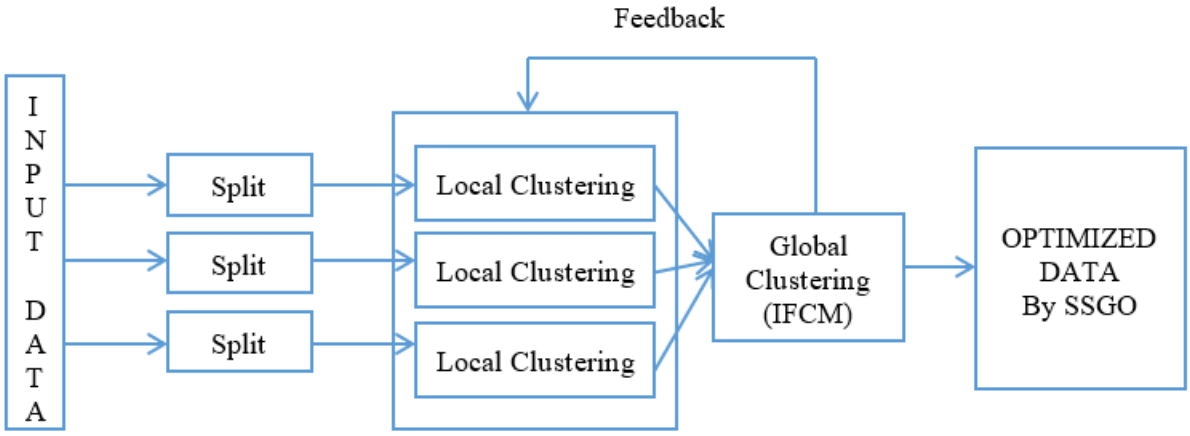
The Social Group Optimization Algorithm, is a meta-heuristic optimization method that was presented in the year 2016 for the purpose of

tackling issues involving global optimization [17]. In the research that has been published, it has been shown that SGO is successful in comparison to other optimization techniques.

3 PROPOSED METHODOLOGY

The proposed architecture is shown in fig 1. By beginning with pre-processing, the input data was first trained and evaluated. An image has been scaled and dimensionally reduced, as well as noise removed. The dimensional reducing data was then forwarded to map reductions to extract the best data. The acquired picture is then grouped using IFCM [18], the grouped data is optimised using a Meta heuristic approach using the SSGO, and the best data is categorized.

Figure 1: Proposed Architecture



3.1 Improvised Fuzzy C-Mean's approach

In IFCM, the probability value of every data point is assigned for the corresponding CC depending on the data point and the clustering distance. IFCM also yields exceptional results in cases when the data overlap exists. While computing time and precision are required, it also needs the execution of many iterations, and Eudoxus' Euclidean distance assesses uneven weight. As a result, this may be accomplished by combining CNN with encoder-decoder [19].

Let us consider the dataset $Z = \{z_1, z_2, \dots, z_q\}$ with cluster set $X = \{x_1, x_2, \dots, x_p\}$ and probability set $W = \{w_{kl} \mid 1 \le k \le e, 1 \le 1 \le p\}$.

We advance this FCM as an Optimized Auto-Encoder

$$min: \sum_{k=1}^{e} \sum_{l=1}^{p} w_{kl}^{o} \|z_l - x_k\|^2$$
$$\sum_{l=1}^{e} w_{kl} = 1, w_{kl} \ge 0 \quad (1)$$

In order to overcome limitations, we introduce an Advanced FCM approach

$$L_o(W, X) = \sum_{k=1}^{e} \eta_i \sum_{k=1}^{p} (1 - u_{kl}^{o})^{\circ} + \sum_{k=1}^{e} \sum_{k=1}^{p} w_{kl}^{m} \|z_l - z_i\|^2 \quad (2)$$

Optimizing this gives:

$$x_k = \sum_{l=1}^{p} w_{kl}^{o} z_l / \sum_{l=1}^{p} w_{kl} \quad (3)$$

Membership matrix

$$w_{kl} = \left(1 + \left(\frac{e_{kl}}{\eta_k}\right)^{-1/(o-1)}\right)^{-1} \quad (4)$$

Here e_{kl} is the distance from cluster to membership matrix.

IFCM-SSGO Algorithm

Input: Dataset, M, n, e

Output: optimized cluster member and membership vec

Step1: Initialization of membership matrix V

Step 2: for k=1 to M do

Step 3: fork=1 to e do

Step 4: Cluster center updation $\eta_k = \sum_{l=1}^{p} w_{kl}^{l} f_{TD(kl)} \; / \; \sum_{l=1}^{0} w_{kl}^{o}$

Step 5: for k=1 to e do

Step 6: for l=1 to p do

Stpe7: $w_{kl} = \left(\left(1 + \left(\frac{f_{TD\,(kl)}}{\eta i}\right)^{-1/(0-1)}\right)^{-1}\right)$

Step 8: end of for loop (step 6)

Step 9: end of for loop (step 5)

Step 10: end of for loop (step 2)

In general, FCM techniques utilise the unsupervised analysis to place unknown data components in the best appropriate cluster or organization. Through the data, it accumulates N- (biggest) clusters. This technique assigns every data element to one of N classes based on its functionality and distance. Routine statistics components are grouped according to similarities and a centroid is created for each cluster reflecting function values. Since each centroid is linked to a cluster component, choosing the most exact and suitable centroid is critical. Calculating the mean in cluster seems to be an iterative method that takes as inputs the number of N clusters and the data items and groups them based on their similarity [20]. It works by starting with a random centroid and updating it based on each detail. The sections that follow discuss IFCM clustering.

i. Assigning Data: Data gathering is the first step in clustering. The initial cluster centroid produced is one-of-a-kind. Each fact is assigned to its nearest centroid using Euclidean, Manhattan, and City-block distances. Using Cj as the jth centroids of C, each element x may be assigned to a cluster using the equation below.

$$\text{argmin dist}(\mathrm{C_j}, \mathrm{x})^2 \qquad (5)$$

In the equation given above, dist(.) represents distance (L2). Suppose si represents a collection of data points allocated to each centroid of a cluster, and then it should be updated repeatedly.

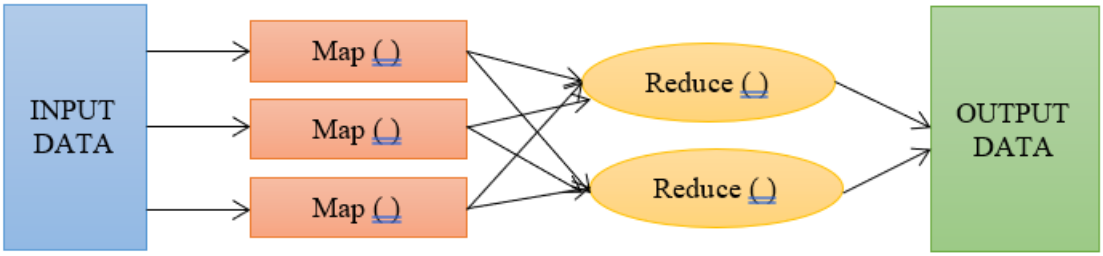
ii. Updating Centroid Distance: By calculating the mean of all associated data items, IFCM updates centroids repeatedly. In principle,

$$C_i = \frac{1}{|S_i|} \sum x_i \qquad (6)$$

This approach is repeated until the termination conditions are satisfied. We employed a feature-adaptive (or performance-adaptive) halting scenario, as opposed to typical IFCM algorithms, that use iterations as stopping criteria. Since we developed SSGO for IFCM clustering, which cluster all data components depending on best model characteristics (i.e., similarity or distance metrics with the greatest centroid selection), our recommended SSGO is set termination criterion to obtain optimum overall performance.

3.2 Map Reduction Model

Figure 2: MapReduce Model for Data Clustering



The memory scarcity problems in massive data analysis might be handled using the MapReduce paradigm, MRM. By separating and processing data, the MapReduce paradigm may be implemented by using mapper and reduced functions. The data and attributes input matrix is shown [21]. The class amount is stored in the vector, which is considered a training data sample simultaneously with the input matrix. The mappings and reduction events in the procedure are detailed in Figure 1's architectural categories.

3.2.1 Mapper Process

The following is a map reduction model with a number of mappers: S = {S1, S2, ... SV }; 1 <$a \leqslant V$, V represents the absolute mapping number. Data will be divided, and the probability index table will be built by the mapper parts. Data partitioned according to input

$$B = \{A_1, A_2, \ldots, A_a, \ldots, A_v\}; 1 < a \leq V \quad (7)$$

Using the data value as the partition information, Aa is the partition information,

$$S_a = \{\mu_l^i(a), \sigma_l^{2i}(a).c_i(a).m_a\} \quad (8)$$

Where, μil is the median, $\sigma 2\ il$ is Variant value and class mark are present. Calculate the average mapper value as follows:

$$\mu_l^i = \frac{\sum_{a=1}^{V} \mu_l^i(a) \times m_a}{\sum_{a=1}^{V} m_a} \quad (9)$$

Where, $\mu il(a)$ is the average value of the ith class and ma is the number of the a[th] mapper. Variance is also computed as follows:

$$\sigma_l^{2i} = \frac{\sum_{a=1}^{V} \sigma_l^{2i}(a) \times m_a}{\sum_{a=1}^{V} m_a} \quad (10)$$

Where, $\sigma 2\ (a)$ is Data variance for the i[th] class.

3.2.2 Reducer Process

It combines the probability index-based tables of the mapper with those of the reducer. A technique is used to build a single table from indices-based tables using the below reducer method.

$$P = \frac{\sum_{a=1}^{V} P(a)}{V} \quad (11)$$

$P(a)$ is the single probability index-based database based on reducing-based methods.

3.3 Stochastic Social Group Optimization (SSGO)

The SSGO technique is separated into two sections. The first section is the "improving phase," while the second part is the "acquiring phase." The skill level of each member in the group is increased during the 'improving phase,'

thanks to the impact of the best individual in the group. The best person in the group is the one with the greatest degree of knowledge and ability to solve the issue, and during the 'acquiring phase,' each individual improves his or her knowledge by mutual contact with other participant in the organization and the best person in the group at the moment. The following is a rudimentary mathematical understanding of this notion. In Stochastic SGO (SSGO), improving phase remains the same, and the acquiring phase only modified.

Let $X_j; j = 1, 2, 3, \ldots . N$ be the social group members, $X_j = (X_{j1}, X_{j2}, X_{j3}, \ldots . X_{jD})$ D denotes the number of traits that define an individual's dimensions and $f_j; j = 1, 2, 3, \ldots . N$ are the values correspond to their respective fitness levels.

Improving phase: The best individual (gbest) in every social group seeks to spread information among all members, which helps others in the group enhance their knowledge.

Hence, $g_{bestg} = \min\{X_j; j = 1, 2, 3, \ldots . N\}$ at generation g in order to solve the minimization issue

During the improvement phase, each individual receives information from the group's best (gbest) member. Each person's updating may be calculated as follows:

$$\begin{aligned} &for\ i = 1{:}N \\ &for\ j = 1{:}D \\ X_{newij} = C * X_{oldij} &+ r \\ &* (gbest(j) - Xoldij) \\ &End\ for \\ &End\ for \end{aligned}$$

where r as a random number, $r \sim U(0, 1)$Take X_{new} if It provides a better level of fitness than X_{old} The parameter c represents self-introspection.

Acquiring phase: During the acquisition phase, members of a social group interact with the best member (Gbest) of that group and with other members randomly in order to acquire knowledge. When another person is more knowledgeable than the individual, he or she acquires new knowledge. A knowledgeable individual has a greater impact on others in terms of influencing them to learn from him or her. A person learns something new from another person if the other person has more information than he or she does and he or she has a greater Identity Possibility (IP) of gaining that knowledge. As a result, the modified acquiring phase is written as

for j=1:D

Randomly select one person P , where $i \neq r$

If f(X_i)< f(X_r)

If r>IP

for j=1:n

X_(newi,j)=X_(oldi,j)+r1(X_(i,j)- X_(r,j))+r2 * (〖gbest〗_j-X_(i,j))*

End for

End for

Else

For j=1:D

$$X_{newi} = X_{oldi} + r1 * (X_r - X_i) + r2 * (gbest_j - X_{ij})$$

End for

End If

Accept X_{new} if it gives a better fitness function value.

End for

where r1 and r2 are the sequences are independent of one another, $r1 \sim U(0, 1)$ and $r2 \sim U(0, 1)$. Equation above illustrates how these sequences have an impact on stochastic nature of the algorithm. A flowchart has been developed to make the entire process easier to comprehend and implement.

3.4 Implementation of SSGO with IFCM

This section provides a step-by-step description of the implementation of SSGO-IFCM.

Step 1: Identifying the problem and initializing the parameters

Determine the size of the population (N), the generation number (g), the parameters (D), and the variables' limits (UL, LL). Here are the definitions of the optimization problem: Minimalize the cost utility f(X). Subject to $= (X_1, X_2, X_3, \ldots . X_D)$ so that $= (X_{j1}, X_{j2}, X_{j3}, \ldots . X_{jD})$ where f(X) is that, X is a vector of design variables, then the objective function can be computed$L_{Li} \leq x_i \leq U_{L,i}$.

Step 2: Initiate the population

The size of the population and the features of the population are used to generate a random population. For SSGO, population size refers to the number of individuals. The population can be categorized as follows:

$$population = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \ldots & x_{1,D} \\ \vdots & & & \ddots & \vdots \\ X_{N,1} & X_{N,2} & X_{N,3} & \cdots & X_{N,D} \end{bmatrix}$$

Step 3: Improving Phase

Determine gbest after that. It is important to determine which solution is most appropriate for that iteration. The improving phase is similar to the learning phase in that each participant gains knowledge from their group's best.

$$For\ i = 1 : N$$

$$For\ j = 1 : D$$

$$X_{newij} = C * X_{oldij} + r * (gbest(j) - X_{oldij})$$

End for

End for

The value of c represents the self-introspection factor. For a particular situation, the value of c may be determined experimentally. In this work, we set it at 0.2 after a comprehensive examination of our examined difficulties, and r is a random value.

Step 4: Acquiring phase

As previously stated, during the acquiring phase, a member of a social group connects with the best individual, i.e., the best of the group, and then engages at random with other members of the group in order to acquire information. "Acquiring phase" defines the mathematical phrase.

Step 5: Termination criterion

If the max iteration numbers is reached, the simulation is terminated; otherwise, redo Steps 3-4.

4 RESULTS & DISCUSSION

4.1 Dataset Selection and Pre-processing

This data set is linked to "https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000". This data set contains a stable assortment of images of malignant and benign skin moles. The dataset is divided into two records, more or less every containing 1800 images (224x244) of 2 different molarity. The finite size as well as wide range of available histologic set of images inhibit artificial neural training for rapid recognition of skin lesions with pigmentation. To resolve this concern, the HAM10000 data - set ("Human Against Machine with 10000 training pictures") has been used. Dermatoscopic images from diverse demographics were collected, conquered, as well as archived. The final set of data includes 10015 dermatoscopic images that can be utilized for academic purposes.

4.2 Performance metrics

4.2.1 Normalized Mutual Information (NMI)

Mutual data is characterized as the measurement of mutual dependency among two parameters in general. NMI ranges from 0 to 1, with 0 indicating no mutual information and 1 indicating perfect correlation. A higher NMI value suggests a more effective clustering model.

$$NMI = \big(h(e) + h(a)\big)\big(h(e,a)\big)^{-1} \quad (12)$$

4.2.2 Adjusted Rand Index (ARI)

Random Index is essentially more than a measurement of similarity among two separate data clustering. Rand Index values range from 0 to 1, with 0 indicating that two distinct data clustering exist at any moment and 1 indicating that data grouping are absolute. A higher ARI value suggests that the model is more efficient.

$$AverageRand\ index = \frac{(Rand\ Index - true\ negative)}{\big(Max(Rand\ Index) - E(Rand\ Index)\big)^{-1}} \quad (13)$$

$$Clustering\ Accuracy = P\left(\sum_{k=1}^{P} 1(A_k = max(d_k)\ )\right)^{-1} \quad (14)$$

The clustering assignment is denoted by dk in the above equation.

Table 1: Comparison of existing and suggested methods for a given database

Methods	Accuracy	Precision	Sensitivity	Specificity	F1-Score
FCM	78.24	72.21	88.32	68.32	82
IFCM	82.36	72.34	93.28	71.23	84
Proposed	91.7	81	99	78	91

Figure 3: Performance Evaluation of different algorithms

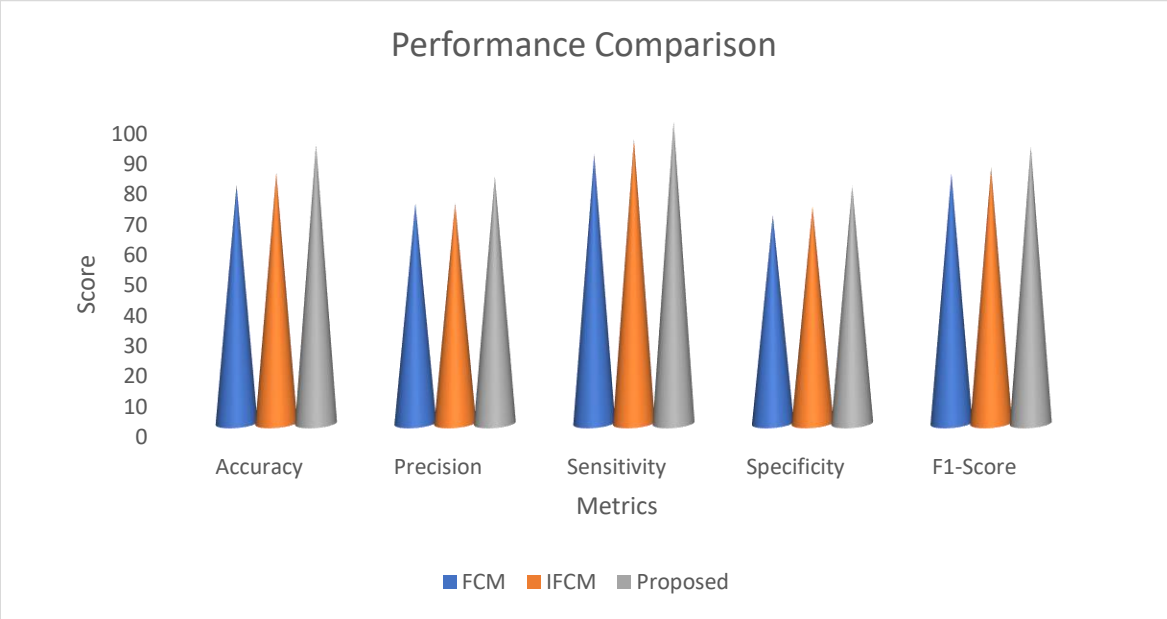


Table 1 and fig 3, illustrating some observations of Training Statistics and Test Sets, based on the case data sets, determines the results for the classifier, classes instances based upon the same observation, compares performance measures of various FCM, IFCM techniques are compared to proposed techniques.

5 CONCLUSIONS

IFCM is a conventional FCM with an extra function variable for determining the distance among the instances and the Cluster; moreover, we incorporate SSGO to improve the efficiency metrics. Moreover, SSGO aids in effective and speedier model training; when paired with fuzzy C-Means, IFCM has a fine clustering model. In order to test IFCM, well-known machine learning datasets are used: MNIST. In addition, a full comparative study is performed using performance metrics such as accuracy, normalized mutual index, and adjusted rand index; in each of these metrics, IFCM with SSGO outperforms several state-of-the-art approaches like as FCM and K-means. Clustering is regarded a rookie mechanism for data analysis in the machine learning sector; yet, IFCM with SSGO contains a superb clustering mechanism with minimal growth in comparison to other existing models. There are various more areas that must be prioritized for real-time data clustering.

REFERENCES

Xu, Z. (2022). Computational intelligence based sustainable computing with classification model for big data visualization on map reduce environment. Discover Internet of Things, 2(1), 2.

Sardar, T. H., & Ansari, Z. (2022). Distributed big data clustering using mapreduce-based fuzzy C-medoids. Journal of The Institution of Engineers (India): Series B, 1-10.

Bashabsheh, M. Q., Abualigah, L., & Alshinwan, M. (2022). Big data analysis using hybrid meta-heuristic optimization algorithm and MapReduce framework. In Integrating meta-heuristics and machine

learning for real-world optimization problems (pp. 181-223). Cham: Springer International Publishing.

Ali, S. A. G., Al-Fayyadh, H. R. D., Mohammed, S. H., & Ahmed, S. R. (2022, June). A Descriptive Statistical Analysis of Overweight and Obesity Using Big Data. In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-6). IEEE.

Jain, D. K., Boyapati, P., Venkatesh, J., & Prakash, M. (2022). An intelligent cognitive-inspired computing with big data analytics framework for sentiment analysis and classification. Information Processing & Management, 59(1), 102758.

Mehta, B. B., & Rao, U. P. (2022). Improved l-diversity: Scalable anonymization approach for privacy preserving big data publishing. Journal of King Saud University-Computer and Information Sciences, 34(4), 1423-1430.

Jayasri, N. P., & Aruna, R. (2022). Big data analytics in health care by data mining and classification techniques. ICT Express, 8(2), 250-257.

Venkateswarlu, Y., Baskar, K., Wongchai, A., Gauri Shankar, V., Paolo Martel Carranza, C., Gonzáles, J. L. A., & Murali Dharan, A. R. (2022). An Efficient Outlier Detection with Deep Learning-Based Financial Crisis Prediction Model in Big Data Environment. Computational Intelligence and Neuroscience, 2022.

Naik, A., & Satapathy, S. C. (2021). A comparative study of social group optimization with a few recent optimization algorithms. Complex & Intelligent Systems, 7, 249-295.

Chidambaram, S., & Gowthul Alam, M. M. (2022). An integration of archerfish hunter spotted hyena optimization and improved ELM classifier for multicollinear big data classification tasks. Neural Processing Letters, 54(3), 2049-2077.

Ramakrishnan, U., & Nachimuthu, N. (2022). An Enhanced Memetic Algorithm for Feature Selection in Big Data Analytics with MapReduce. Intelligent Automation & Soft Computing, 31(3).

Dey, N., Rajinikanth, V., Ashour, A. S., & Tavares, J. M. R. (2018). Social group optimization supported segmentation and evaluation of skin melanoma images. Symmetry, 10(2), 51.

Srinivasareddy, S., Narayana, Y. V., & Krishna, D. (2021). Sector beam synthesis in linear antenna arrays using social group optimization algorithm. National Journal Of Antennas And Propagation, 3(2), 6-9.

Satapathy, S., & Naik, A. (2016). Social group optimization (SGO): a new population evolutionary optimization technique. Complex & Intelligent Systems, 2(3), 173-203.

Feng, X., Wang, Y., Yu, H., & Luo, F. (2016). A novel intelligence algorithm based on the social group optimization behaviors. IEEE Transactions on systems, man, and cybernetics: systems, 48(1), 65-76.

Naik, A., Satapathy, S. C., Ashour, A. S., & Dey, N. (2018). Social group optimization for global optimization of multimodal functions and data clustering problems. Neural Computing and Applications, 30, 271-287.

Naik, A., & Satapathy, S. C. (2021). A comparative study of social group optimization with a few recent

optimization algorithms. Complex & Intelligent Systems, 7, 249-295.

Rayala, V., & Kalli, S. R. (2020). Big Data Clustering Using Improvised Fuzzy C-Means Clustering. Rev. d'Intelligence Artif., 34(6), 701-708.

Venkat, R., and K. Satyanarayan Reddy. "Dealing Big Data using Fuzzy C-Means (FCM) Clustering and Optimizing with Gravitational Search Algorithm (GSA)." 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2019.

Venkat, R., and K. Satyanarayan Reddy. "Clustering of Huge Data with Fuzzy C-Means and applying Gravitational Search Algorithm for Optimization." International Journal of Recent Technology and Engineering (IJRTE). Volume-8 Issue-5, January 2020

Venkat, R., and K. Satyanarayan Reddy. “A Clustering based Hybrid Optimization approach using Evolutionary Computing and Map Reduction Architecture on Big Data.” NeuroQuantology, Volume 20, Issue 6, 2022, pp. 7584-7603.