Acute Lymphoblastic Leukemia (ALL) Prediction using Fuzzy-Based Mathematical Clustering Feature Extraction Algorithm

Mrs.Vidhya. B[1]
Department of Information Technology, Sri Ramakrishna College of Arts & Science, Coimbatore, India, vidhya.b@srcas.ac.in

Dr.A.Kumar Kombaiya[2]
Department of Computer Science, Chikkana Government Arts College, Tiruppur, India, kumar_kombaiya@rediffmail.com

Abstract:

Leukemia's are cancers of the hematopoietic stem cells that cause the bone marrow. Acute and chronic are the two broad categories of leukemia. Acute leukemia is classified into two subtypes namely, Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML). ALL is a childhood cancer that primarily affects children from 5 to 14 years old. If leukemia is predicted and identified in its early stage, appropriate treatment can be given to the patient. In this way, the disease can be cured, and patient's recovery can be expedited and the lives of many leukemia patients can be saved. Leukemia cells are complex. Diagnosing leukemia disease through medical images is a difficult task. The primary goal of this research work is to extract the significant features from leukemia microscopic ALL images. The Fuzzy based Mathematical Clustering (FMC) algorithm is proposed in this work for feature extraction. From the experimental analysis, it is found that a proposed FMC has properly extracted GLCM, shape, and color features from leukemia images with a higher performance ratio than existing Gray Level Difference Method (GLDM), Gray Level Run Length Matrix (GLRLM), and Mean-shift feature extraction algorithms.

Keywords – Medical Images, Leukemia, Acute Lymphoblastic Leukemia (ALL), Feature Extraction and GLCM.

DOI Number: 10.14704/nq.2022.20.9.NQ44447 **Neuro Quantology 2022; 20(9):3915-3929**

1. INTRODUCTION

In image processing, feature extraction is a method for redefining a bigger set of redundant data into a smaller number of features (or feature vectors). Feature extraction [3] [5] is the process of transforming raw data into a set of features. The percentage of lymphoblasts present in bone marrow samples or peripheral blood is used to diagnose and follow up on ALL. This is the criterion that is employed in the diagnosis and following observation [8]. A peripheral blood or bone marrow sample with a lymphoblast cell count of greater than 20% reveals the existence of acute lymphoblastic leukaemia. If there are more than 20% lymphoblasts in the peripheral blood or bone marrow, the sample is classified as ALL [8]. The nucleus and cytoplasm of lymphoblasts, which differ from lymphocytes in terms of morphological and functional abnormalities, are crucial in identifying malignancy in blood samples. A lymphoblast is characterized by a large nucleus with an uneven form and size [9]. It's tough to spot cytoplasm in blood sample photos because it appears infrequently and, when it does, it's brightly colored. Lymphocytic nuclei and cytoplasm reflect morphological and functional changes [15]. They are crucial in the detection of leukemia in blood samples. Medical photos in digital format take a lot of memory. Compiling, storing, accessing, and retrieving them will take more time. Salient features are extracted from a picture to decrease data space and processing time.

Most the hematologists for recognition of lymphoblast's features are Nucleus size and shape, Nucleoli, Cell Size, Nucleus color, Nuclear

chromatin, Nucleus boundary, Amount of cytoplasm, Cytoplasmic color, Nucleus to Cytoplasm ratio, and Cytoplasmic boundary. A Feature is called the subset of one or more measurements of the image where each value specifies some measurable property of the Image [1] [17]. Low-level features and high-level features are two main categories for features. While high-level features must be estimated from the low-level features, low-level features can be measured straight from the image. Instead of processing the entire image, it is desirable to perform Feature extraction which is the formation of variables called feature vectors from the given set of variables. Figure 1 illustrates the different features in leukemia images.

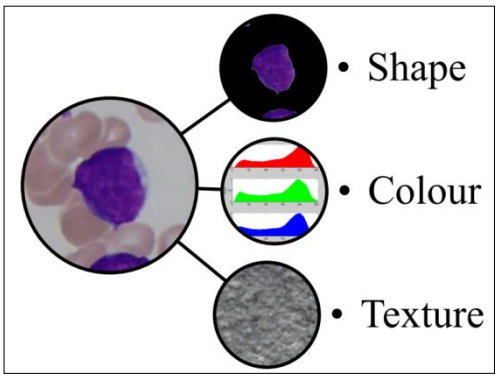

Figure 1 Different Features in Leukemia Images

a) Shape-based Features

One essential element that describes an image's physical properties is its form. Noise, arbitrary distortion, and obstruction can all taint it. This makes the process of image recognition more difficult [12]. The representation of shape is distinguished into two foremost categories, region-based and boundary-based representations. In a boundary-based approach, the features are extracted from the Region of Interest (RoI). Whereas, in the region-based method the features are extracted from the internal details as well as the external details such as area, perimeter, circularity, etc. In this research, the following shape-based features are extracted from the segmented image of the nucleus [14].

b) Color-based Features

The colour elements of an image are represented by color-based characteristics. The RGB colour scheme used for leukaemia images makes it possible to distinguish between blood and bone marrow cells [13].

c) Texture-based Features

The arrangement of an image's fundamental components is described by its texture feature. Therefore, separating the photos solely based on color-based attributes is not desired. Texture properties help with picture recognition, image retrieval, etc. in certain situations. The Gray-Level Co-occurrence Matrix (GLCM) is one of the often used measurements that may be used to describe the texture properties [15].

The remainder of the paper is structured as follows: The literature review on ALL image feature extraction is presented in Section II. The proposed methodology for the feature extraction technique is discussed in Section III. The outcomes of the experiments are discussed in Section IV. The conclusion is found in Section V.

2. LITERATURE REVIEW

Sahlol et al. (2018) examined ALL images using SSOA. In the preprocessing stage, the entire RGB color leukemia images were converted into CMYK color space. Different types of features such as color, texture, and shape-based features were extracted. The redundant features were removed using SSOA, leukemia and non-leukemia cells were identified in the images.

Samant et al. (2018) investigated Type-2 diabetes using soft computing techniques. Various features like statistical, texture, and wavelet features were extracted from the iris region. Popular filter-based feature selection approaches were used to remove the redundant features. Classification models such as Binary Tree (BT) Model, SVM, Adaptive Boosting (AB) model, Generalized Linear (GL) model, NN, and RF were utilized to evaluate the feature selection techniques. The experimental results proved that T-test based feature selection method performed better than other methods.

Jothi et al. (2018) designed a Graphical User Interface (GUI) to classify the normal and leukemia images. Various statistical features, texture, wavelet, color, and morphology were extracted. STRSPSORR, STRSFFQR and STRSPSOQR were different feature selection hybrid algorithms, which was developed on the strengths of rough set theory and evolutionary

algorithms were applied and their performance was evaluated. The results showed that STRSFFQR provided better results than other related algorithms.

Srisukkham et al. (2017) introduced two Bare-bones PSO (BBPSO) algorithms for selecting the most significant features. This approach combined the accelerated food chasing mechanism. The experimental results showed that BBPSO performed better than other optimization algorithms.

Jothi et al. (2016) examined MRI brain tumor images using rough set theory and optimization algorithms. Region growing and threshold methods were applied to segment the tumor region in the images. Shape, intensity, and texture-based features were extracted from the segmented MRI image. STRSFFQR algorithm was employed to select the most significant features. The performance of the STRSFFQR algorithm was compared with the existing feature selection algorithms such as STRSPSO-RR, Cuckoo Search (CS), ABC, and STRSPSO-QR.

Moradi et al. (2016) made use of the Fuzzy C-Means (FCM) clustering algorithm to segment the image of the nucleus. Geometric and statistical features were extracted from the segmented image and PCA was employed to remove the redundant features. The classification based on SVM was used to classify cells into four groups, namely noncancerous, L1, L2, and L3.

Applications of the ABC algorithm was examined by Agrawal et al. (2015) for the study of cervical cancer images. Haralick features were extracted from the region of interest and ABC with KNN algorithm and ABC with SVM were applied for feature selection. The results showed that ABC with SVM performed better than the ABC with KNN.

An earlier diagnostic method for acute lymphoblastic leukemia images was proposed by Mohapatra et al. (2014). In this approach, morphological, texture, and colour-based features were extracted and the popular filter-based feature selection algorithm namely, T-test was utilized to select significant features. The author(s) developed an ensemble of classifiers to diagnose the disease.

Jothi et al. (2013) presented a hybrid feature selection approach based on tolerance rough set and PSO algorithm for the extraction of salient features from Digital Mammogram Images. GLCM was applied to extract the relevant features. STRSPSOQR and STRSPSORR algorithms were employed to select the most significant features. The experimental results show that the proposed approaches increased the diagnostic accuracy when compared to the existing feature selection algorithms.

The soft set-based feature selection technique as applicable to lung cancer images was investigated by Jothi et al. (2012). Region growing algorithm was applied to segment the lung region in the image and nineteen Haralick features were extracted. These features were utilized for feature selection processes. The Soft Set-based Unsupervised Quick Reduct (SSUSQR) algorithm selected the most prominent features which were related to the decision attribute.

3. METHODOLOGY

The key objective of this research work is to extract the significant features from leukemia microscopic images. Figure 1 shows the system methodology of the proposed research work.

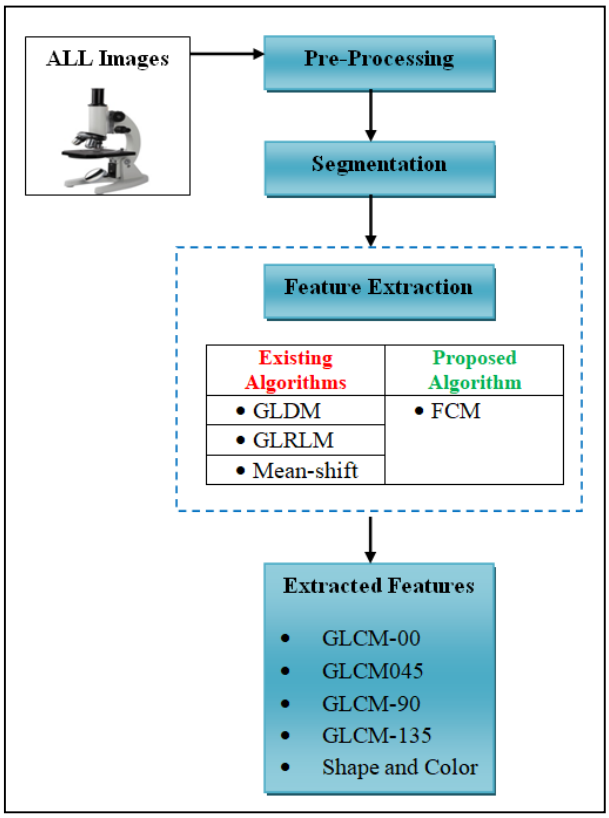

Figure 2 System Methodology

3.1 PREPROCESSING

The acute lymphoblastic leukemia image databases namely, ALL-IDB1 and ALL-IDB2 are the datasets of microscopic ALL images of blood samples. It is extensively used for evaluating and comparing of feature extraction algorithms. Medical images are obtained from different modalities such as MRI, X-Ray, CT scan, and microscope, which is very hard to handle directly. Therefore, preprocessing techniques are used to remove the noise and enhance the quality of an image. In this research work, the pre-processing techniques include noise removal and contrast enhancement. The digital microscopic images are often corrupted by noise and speckles due to the fault of electronic equipment, environmental factor, and human error during the process. Inference from noise-affected images may lead to imperfect diagnosis. In such instances, medical image processing helps a lot in producing clear images. In the current research, popular filtering techniques are employed like median filter, average filter, and Gaussian filter to enhance the image of the leukemia cell.

3.2 SEGMENTATION

Covering-oriented rough set is an expansion of existing theory which alters rough approximation, partition or similarity relation. It differs from the classical rough set, which uses covering of the universal set instead of partitions or equivalence class. The covering created by symmetric and reflexive relations and it finds the similarity of the unary covering. The property of covering is the intersection of any two features is the blending of finite features. This algorithm incorporates the Covering-oriented rough approximation for computing the upper and lower estimations of the clustering algorithm. The CRKM clustering algorithm is outlined in Algorithm.

3.3 FEATURES EXTRACTION

In the proposed method, a set of features are extracted from the segmented nucleus. The main categories of features are extracted and classified as GLCM features. Features are classified into different types, which are shape-based features, color-based features, and texture-based features.

GLCM Feature

The mean and the standard deviation of an Image are used to obtain the first-order statistics of an image, which is concerned with the properties of individual pixels. The second orders statistics are derived from the GLCM, which is based on the co-occurrence or spatial inter-dependency of two pixels at specific relative positions in an image. These matrices are measured for the angles of 0, 45, 90, and 135. In the co-occurrence matrix, the fourteen features such as Contrast, Sum Average, Correlation, Difference Entropy, Sum Variance, Entropy, Sum of variance or squares, Angular second moment, Sum Entropy, Cluster Tendency, and Inverse Different Moment, and Difference Variance are measured for the segmented Nucleus Image. The orientation or the angle decides the value of the homogeneity, contrast, entropy, and energy. The homogeneity and the entropy give information about the dominant values of the main diagonal to the basis of the frequencies of the problem. The information about the randomness of the spatial distribution of the image is given by the energy value.

A co-occurrence matrix is a matrix that is defined over an image reflecting the distribution of co-occurring values at a given offset. A co-occurrence matrix of an image (HH, aa) is mathematically defined as follows. A GLCM comprises the details about the pixel positions that have similar gray level values. In this matrix, the rows and columns represent the range of the image gray levels. The value $PP(HH,j)$ stored at the position (HH,j) is the frequency with which the gray levels HH and j occur, at a given distance and a given direction such as 00, 450, 900 and 1350 degree. Table 1 shows the GLCM Matrix.

$$G_{\Delta m,\Delta n}(x,y) = \sum_{a=1}^{i}\sum_{b=1}^{j} \begin{matrix}\{1, if\ I(a,b) = x\ and\ I\ (a+\Delta i, b+\Delta j) = y \\ 0, otherwixe\ \}\end{matrix} \quad (1)$$

Table 1 GLCM Matrix

Degree	0^0				45^0				90^0				135^0			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	2	1	0	1	0	1	0	3	0	2	1	3	0	3	0	1
2	1	6	3	5	1	4	3	3	2	4	4	4	3	0	4	4
3	0	3	6	1	0	3	4	1	1	4	2	2	0	4	0	3
4	1	5	1	4	3	3	1	2	3	4	2	2	1	4	3	2

The objective of the feature selection technique is to select a small number of features, such as class labels in classification, which minimizes repetition while maximizing relevance to the aim. In this study, 108 features—108 shape-based features, 88 GLCM texture characteristics (each direction 22 features), and 12 color-based features—were recovered from each channel of the segmented nucleus image. The selected features are presented in Table 2 presents Selected Features using MDSSA.

Table 2 Selected Features using MDSSA

Dataset	No. of classes	No. of Features Extracted	No. of Objects	No. of Selected Features
GLCM_0	2	22	368	11
GLCM_45	2	22	368	13
GLCM_90	2	22	368	11
GLCM_135	2	22	368	12
Shape and Color	2	20	368	12

3.3.1 Fuzzy based Mathematical Clustering (FMC)

Fuzzy based Mathematical Clustering (FMC) model is proposed in this work for extracting features from microscopic ALL images. Features of cancer cells are recorded in the knowledge base in the training section. In this work taken GLCM and SC features are extracted from ALL images. The fuzzy based feature clustering mines the pixel values of color features from mean, standard deviation from the mathematical model. The FCM method evaluates every object's region (i.e. area), GLCM feature with different angels such as 0, 45, 90, and 135 and SC feature to attain roundness and boundary pixels of candidate object which is mathematically formulated as follows,

$$Area = Diff(Boundary) \quad (2)$$

$$Conv = Sum\sqrt{(Area + 2)} \quad (3)$$

$$Feature\ ratio = \frac{(4*\pi*Area)}{Conv^2} \quad (4)$$

From the above equations (2), (3), and (4) is used for extract the object area, convolution and feature ratio are obtained by using Mathematical Model. It is extensively mines the affected blood cells from ALL images. In FMC method, Mathematical model is represented as dissimilarity between training and testing ALL input image. Let $Image_i$ signifies the original image in which the structuring portions are applied with Mathematical Model, which accomplishes an opening function OF on the original image $Image_i$, which is mathematically expressed as follows,

$$OI = Max\{OF_{IT}(image_i)\} where\ IT = 1,2,\dots n \quad (5)$$

$$RI = OF_{image_i}^{Recon}(OI) \quad (6)$$

The FCM has the benefit of minimizing the sum of square errors by performing repeated partitions in an iterative manner and hence final extraction is attained. The mathematical representation to

perform repeated partitions through iteration is formulated as follows,

$$VE_{Cluster} = (m_{ab})^{e}*\text{Dist}(O_b CC_a) \quad (7)$$

From the equation (7), the ALL image feature extraction through cluster $VE_{Cluster}$ is the product of membership degree of O_b for cluster a and e and that represents the exponent for each membership degree with a distance between object O_b and cluster center CC_a respectively. The cluster center determined for twelve iterations is mathematically expressed as,

$$CC_a^{IT} = \frac{(m_{ab}^{IT})*o_b^{IT}}{m_{ab}^{IT}}, where\ IT = 1,2,\dots n \quad (8)$$

The FMC mines the pixel values of color features from the mean, standard deviation of the opening image from mathematical model step. This in turn supports in improving the clustering efficiency in an effective manner. Finally, FMC method offers accurate feature extraction and improving the performance rate of ALL image diagnosis.

Algorithm 1: Pseudo-code for Fuzzy based Mathematical Clustering (FMC)

Input: Peripheral Blood Microscope ALL Images

Output: Feature extracted ALL images.

Procedure

Step 1. Initialize partition vector $|M = m_{ab}|$ and threshold value

Step 2. For each $Image_i$

Step 3. Compute area using $Area = Diff(Boundary)$

Step 4. Compute convolution using $Conv = Sum\sqrt{(Area + 2)}$

Step 5. Compute feature ratio using $Feature\ ratio = \frac{(4*\pi*Area)}{Conv^2}$

Step 6. Perform Mathematical Model for feature extraction

Step 7. Fuzzy based Feature Clustering

Step 8. Evaluate fuzzification

Step 9. Obtain Feature Clustering Step

Step 10. Evaluate cluster center using $CC_a^{IT} = \frac{(m_{ab}^{IT})*o_b^{IT}}{m_{ab}^{IT}}, where\ IT = 1,2,\dots n$

Step 11. End for

4. EXPERIMENTAL RESULTS

All the experiments are implemented by Matlab 2018b and running on Windows 8.1.

4.1 Dataset

An illustration of the algorithm is applied to five benchmark datasets which are analyzed by diverse characteristics. The datasets are ALL_IDB1 and ALL_IDB2.

- **ALL_IDB1 -** It contains 39000 blood elements, with professional oncologists labeling the cells. This collection contains 510 potential lymphoblasts, or lymphoblast-like cells. In this dataset, only fully described lymphoblasts are taken into account and categorised.
- **ALL_IDB2 -** There are 260 photos total, and lymphoblasts are depicted in 50% of them. Images from the ALL-IDB2 and ALL-IDB1 show similar gray-level characteristics.

4.2 Performance Measures

Several performance measures are used to analyze the efficiency of the proposed and existing algorithms. The Recall, Precision, and F-score measure, Accuracy, and Mean Absolute Error (MAE) have been used to evaluate the performance of the proposed algorithm.

Precision

It measures the ratio of correctly predicted leukocytes to the number of correctly and incorrectly predicted leukocytes as follows,

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

Recall

It is the ratio of the number of correctly identified leukocytes was predicted from images.

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

F-Measure

The weighted harmonic mean of the precision and recall values is used to calculate it.

$$F - measure = 2\ X\ \frac{Precision\ X\ Recall}{Precision+Recall} \quad (11)$$

Mean Absolute Error (MAE)

It is the average of all absolute errors.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x| \quad (12)$$

Accuracy

It calculates the global prediction rate by calculating the ratio of correct results to the total number of leukemia image as follows,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

Table 3 Performance Analysis of Feature Extraction - ALL_IDB1

Dataset	Algorithms	Pre.	Rec.	F-Sc.	MAE	Acc.
GLCM-0	GLDM	68.35	70.55	69.43	0.149	68.16
	GLRLM	73.49	76.11	74.77	0.133	79.79
	Mean-shift	80.26	82.74	81.48	0.091	83.12
	FMC	82.51	84.61	83.54	0.063	84.44
GLCM045	GLDM	69.62	71.94	70.73	0.151	69.26
	GLRLM	75.94	78.24	77.07	0.135	80.89
	Mean-shift	81.95	84.25	83.08	0.092	84.22
	FMC	84.45	86.75	85.58	0.064	85.54
GLCM-90	GLDM	72.37	74.29	73.23	0.155	70.96
	GLRLM	78.64	80.54	79.57	0.139	82.59
	Mean-shift	84.65	86.55	85.58	0.096	85.92
	FMC	87.15	89.05	88.08	0.068	87.24
GLCM-135	GLDM	74.09	76.19	75.12	0.158	72.16
	GLRLM	80.34	82.44	81.37	0.142	83.79
	Mean-shift	86.35	88.45	87.38	0.099	87.12
	FMC	88.85	90.95	89.88	0.071	88.44
SC	GLDM	75.49	77.59	76.52	0.161	73.56
	GLRLM	81.74	83.84	82.77	0.145	85.19
	Mean-shift	87.75	89.85	88.78	0.102	88.52
	FMC	90.25	92.35	91.28	0.074	89.84

Table 3 represents the Performance Analysis of Feature Extraction of the ALL_IDB1 Dataset. From the experimental results, it is noticed that the proposed FMC Feature Extraction algorithm produces higher Precision, Recall, F-Score, Accuracy values and minimum MAE rate than other Feature Extraction algorithms with respect to ALL_IDB1 Dataset.

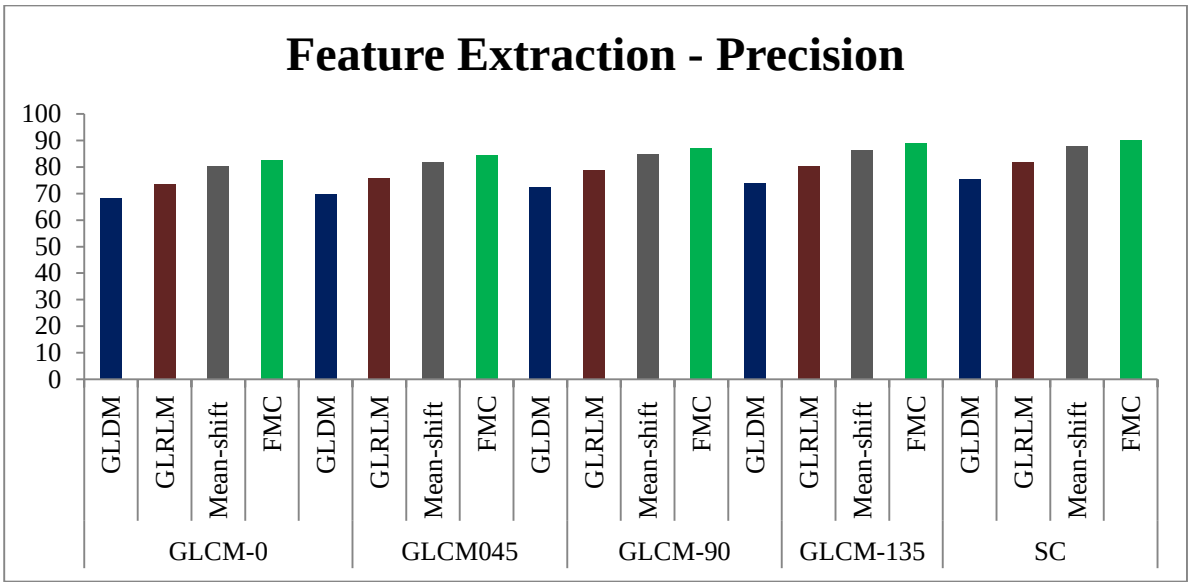

Figure 3 Performance Analysis – Precision for ALL_IDB1

Figure 3 illustrates the Performance Analysis of Precision of the ALL_IDB1 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher Precision values than other Feature Extraction algorithms with respect to ALL_IDB1 Dataset.

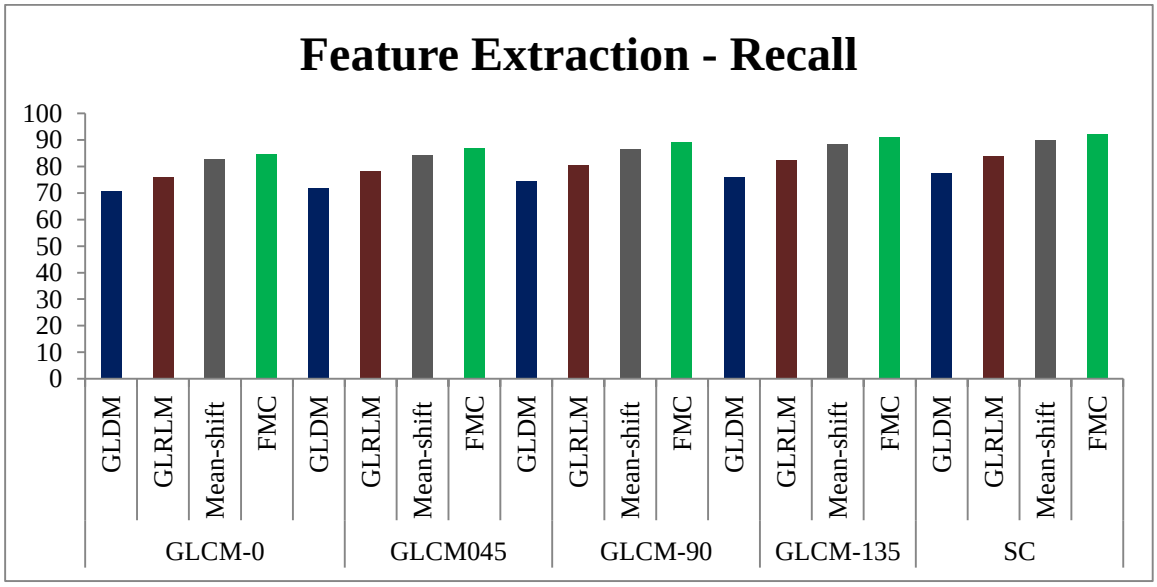

Figure 4 Performance Analysis –Recall for ALL_IDB1

Figure 4 shows the Performance Analysis of Recall of the ALL_IDB1 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher Recall values than other Feature Extraction algorithms with respect to ALL_IDB1 Dataset.

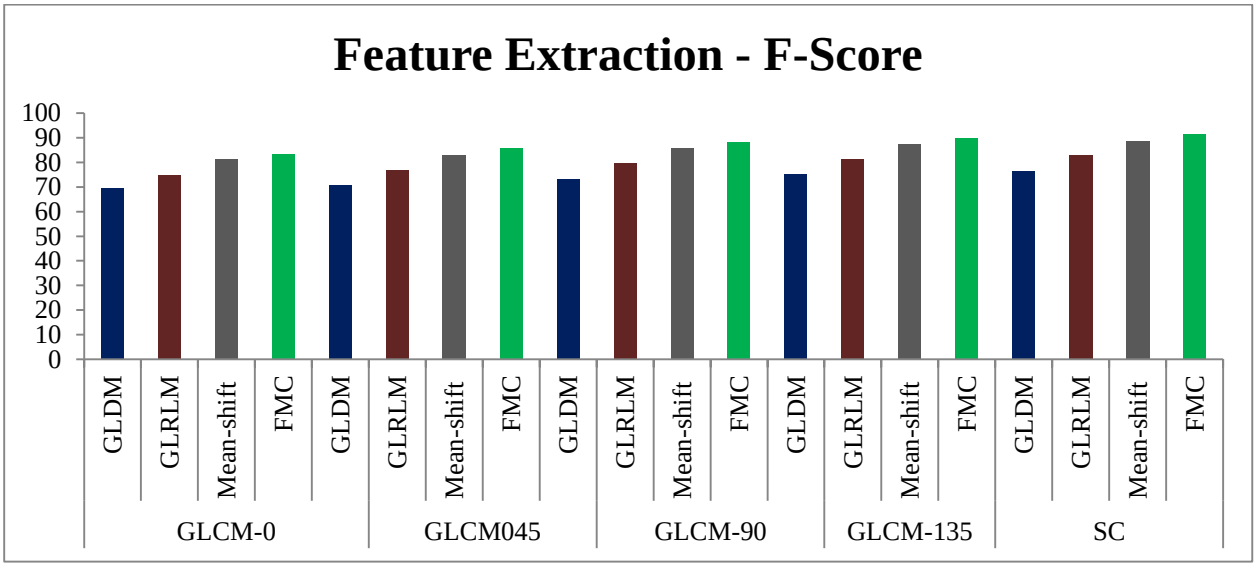

Figure 5 Performance Analysis – F-Score for ALL_IDB1

Figure 5 demonstrates the Performance Analysis of F-Score of the ALL_IDB1 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher F-Score values than other Feature Extraction algorithms with respect to ALL_IDB1 Dataset.

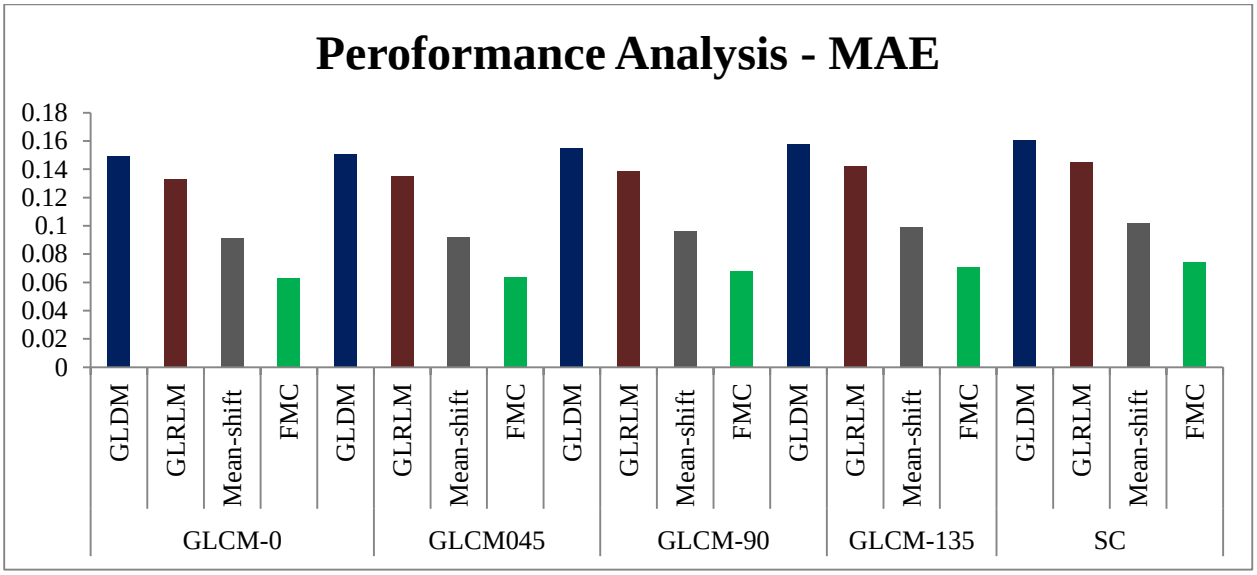

Figure 6 Performance Analysis of MAE for ALL_IDB1

Figure 6 presents the Performance Analysis of MAE of the ALL_IDB1 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces minimum MAE value than other Feature Extraction algorithms with respect to ALL_IDB1 Dataset.

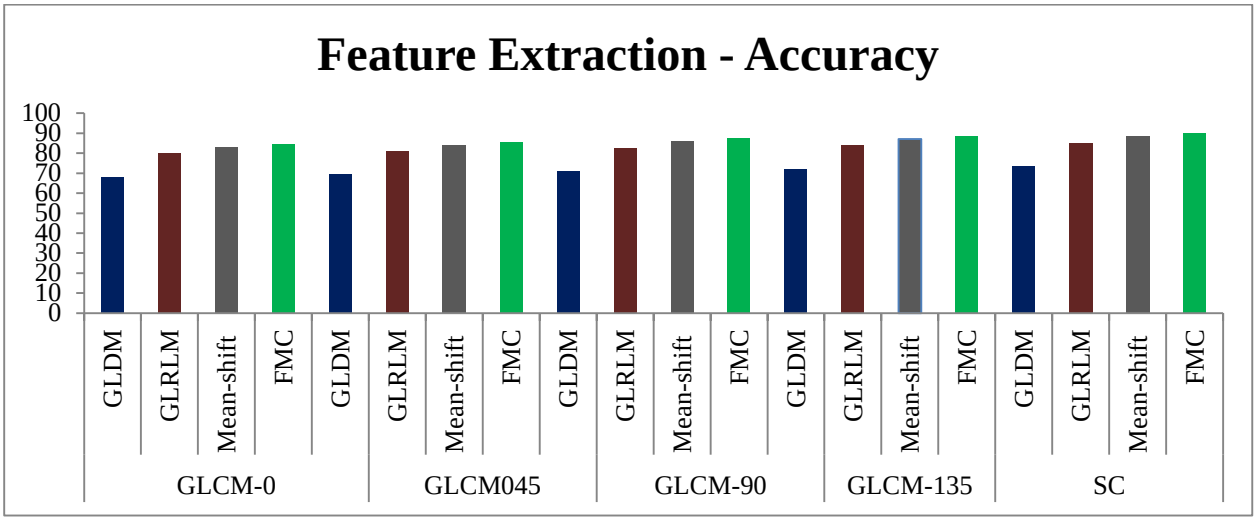

Figure 7 Performance Analysis – Accuracy for ALL_IDB1

Figure 7 gives the Performance Analysis of Accuracy of the ALL_IDB1 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher accuracy than other Feature Extraction algorithms with respect to ALL_IDB1 Dataset.

Table 4 Performance Analysis of Feature Extraction - ALL_IDB2

Dataset	Algorithms	Pre.	Rec.	F-Sc.	MAE	Acc.
GLCM-0	GLDM	67.85	70.05	68.93	0.142	67.86
	GLRLM	72.99	75.61	74.27	0.126	79.49
	Mean-shift	79.76	82.24	80.98	0.084	82.82
	FMC	82.01	84.11	83.04	0.056	84.14
GLCM045	GLDM	69.12	71.44	70.23	0.144	68.96
	GLRLM	75.44	77.74	76.57	0.128	80.59
	Mean-shift	81.45	83.75	82.58	0.085	83.92
	FMC	83.95	86.25	85.08	0.057	85.24
GLCM-90	GLDM	71.87	73.79	72.73	0.148	70.66
	GLRLM	78.14	80.04	79.07	0.132	82.29
	Mean-shift	84.15	86.05	85.08	0.089	85.62
	FMC	86.65	88.55	87.58	0.061	86.94
GLCM-135	GLDM	73.59	75.69	74.62	0.151	71.86
	GLRLM	79.84	81.94	80.87	0.135	83.49
	Mean-shift	85.85	87.95	86.88	0.092	86.82
	FMC	88.35	90.45	89.38	0.064	88.14
SC	GLDM	74.99	77.09	76.02	0.154	73.26
	GLRLM	81.24	83.34	82.27	0.138	84.89
	Mean-shift	87.25	89.35	88.28	0.095	88.22
	FMC	89.75	91.85	90.78	0.067	89.54

Table 4 represents the Performance Analysis of Feature Extraction of the ALL_IDB1 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher Precision, Recall, F-Score, Accuracy values and minimum MAE rate than other Feature Extraction algorithms with respect to ALL_IDB1 Dataset.

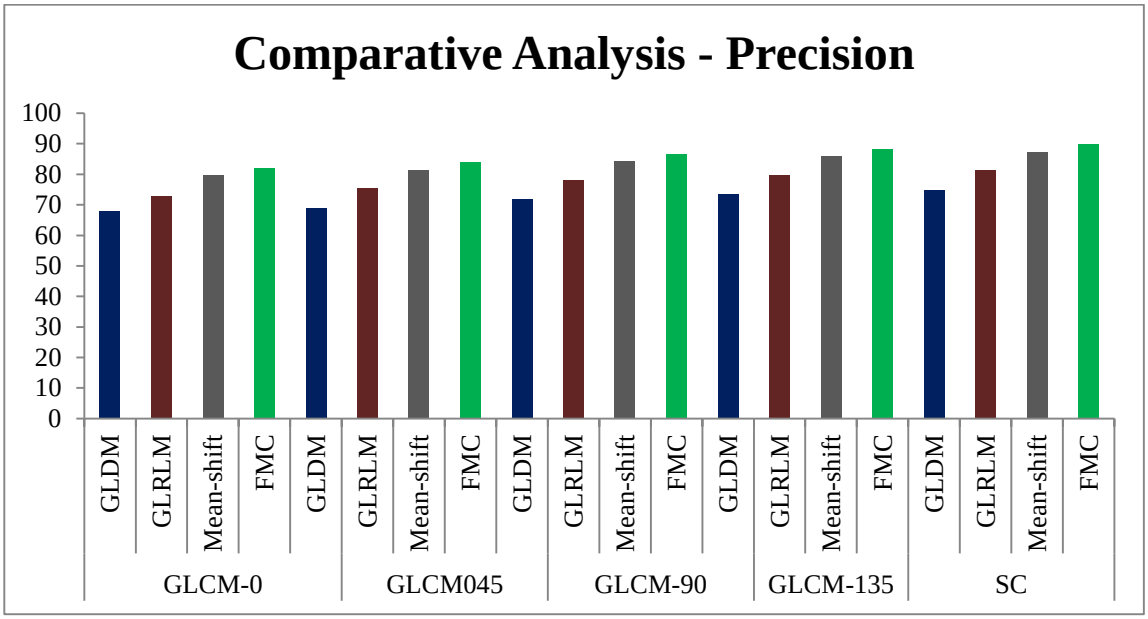

Figure 8 Comparative Analysis – Precision for ALL_IDB2

Figure 8 illustrates the Comparative Analysis of Precision of the ALL_IDB2 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher Precision values than other Feature Extraction algorithms with respect to ALL_IDB2 Dataset.

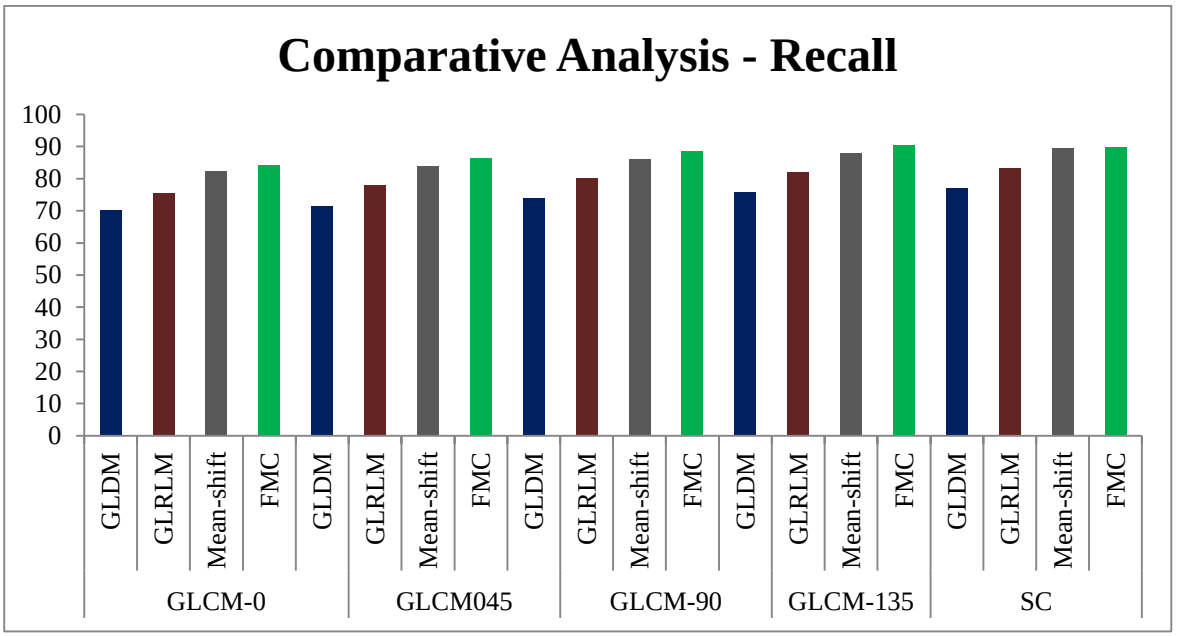

Figure 4 Comparative Analysis –Recall for ALL_IDB2

Figure 9 demonstrates the Comparative Analysis of Recall of the ALL_IDB2 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher Recall values than other Feature Extraction algorithms with respect to ALL_IDB2 Dataset.

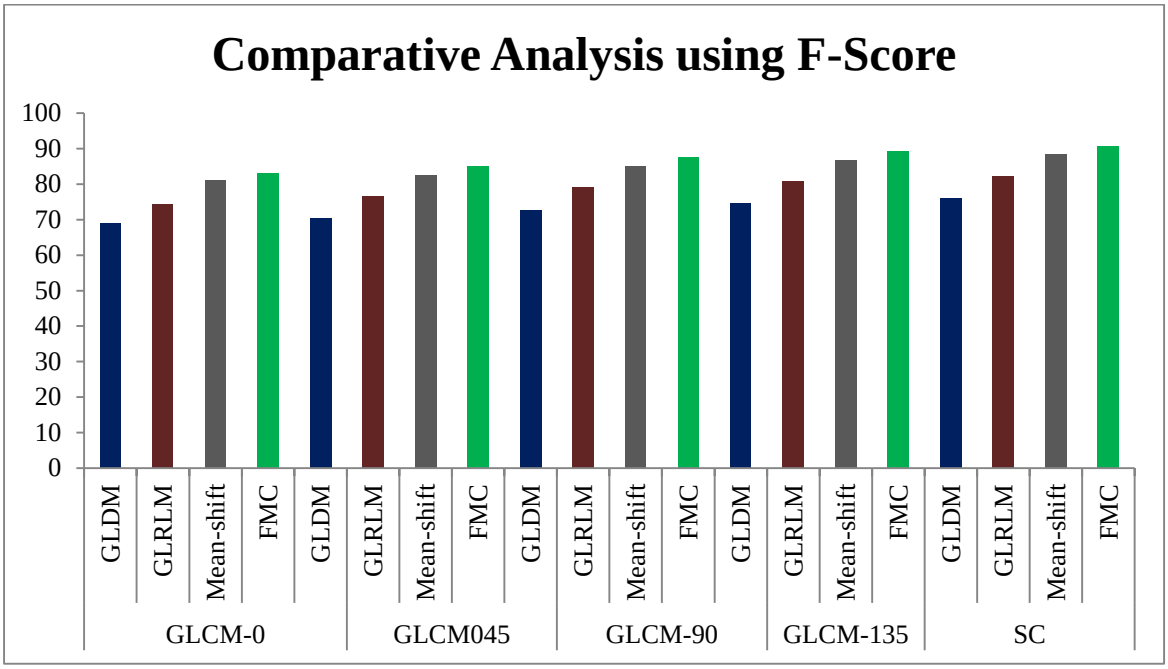

Figure 10 Comparative Analysis – F-Score for ALL_IDB2

Figure 10 presents the Comparative Analysis of F-Score of the ALL_IDB2 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher F-Score values than other Feature Extraction algorithms with respect to ALL_IDB2 Dataset.

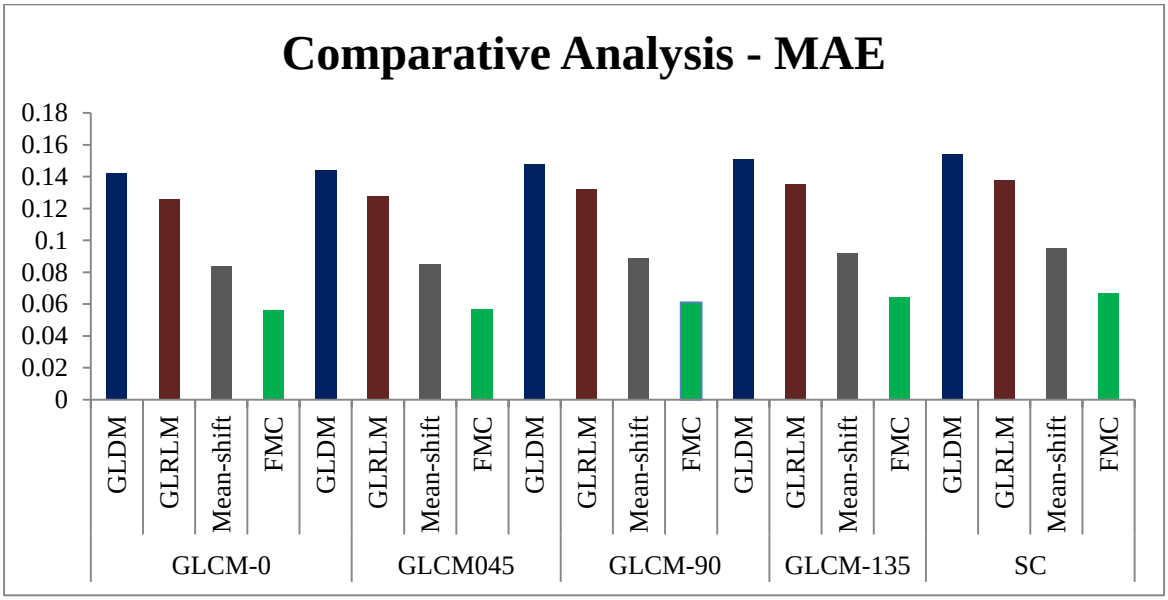

Figure 11 Comparative Analysis of MAE for ALL_IDB2

Figure 11 depicts the Comparative Analysis of MAE for the ALL_IDB2 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces minimum MAE value than other Feature Extraction algorithms with respect to ALL_IDB2 Dataset.

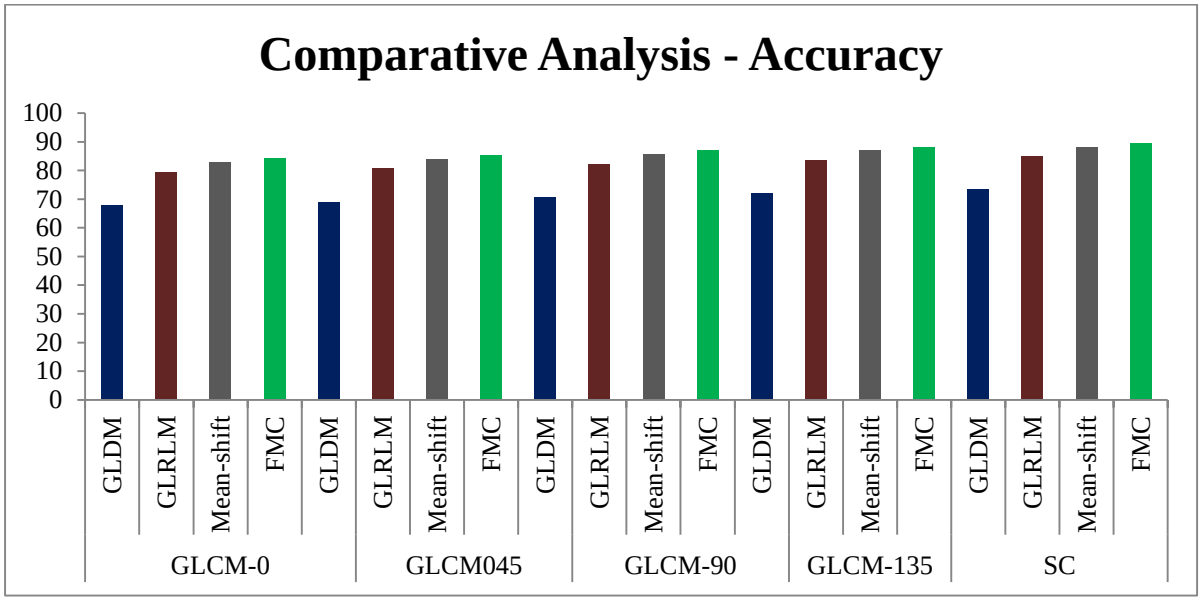

Figure 12 Comparative Analysis – Accuracy for ALL_IDB2

Figure 12 gives the Comparative Analysis of Accuracy of the ALL_IDB2 Dataset. From the experimental results, it is observed that the proposed FMC Feature Extraction algorithm produces higher accuracy than other Feature Extraction algorithms with respect to ALL_IDB2 Dataset.

Table 3 Execution Time for Feature Extraction Algorithms

Datasets	Feature Extraction Algorithms	Execution Time (Milliseconds)
ALL-DB1	GLDM	1972
	GLRLM	1953
	Mean-shift	1620
	FMC	**1326**
ALL-DB2	GLDM	1956
	GLRLM	1703
	Mean-shift	1597
	FMC	**1279**

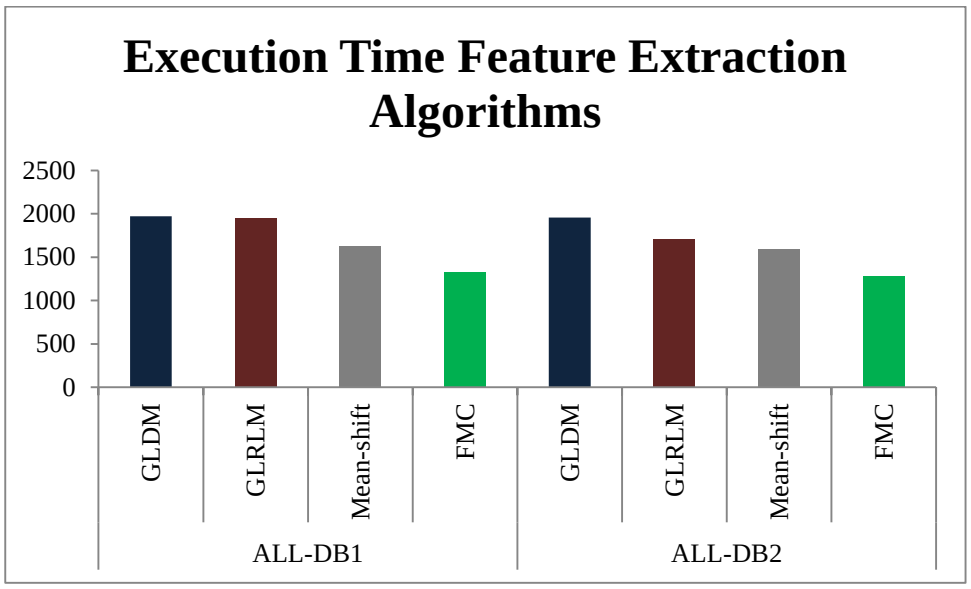

Figure 13 Execution Time for proposed Feature Extraction Algorithms

Table 5.3 and Figure 13 present the execution time for existing and proposed feature extraction algorithms. From the performance analysis, it is found that the proposed FMC proposed feature extraction algorithm required a minimum execution time than existing algorithms.

5. CONCLUSION

The primary aim of this research work is to extract the significant features from leukemia microscopic images. The FMC algorithm is proposed in this work for feature extraction. In this research, three different types of features were employed to address the segmented nucleus image. In texture features, GLCM was computed for the dimensions 0, 45, 90, and 135. From the experimental analysis, it is found that a proposed FMC has properly extracted GLCM, shape, and color features from leukemia images with a higher performance ratio than existing GLDM, GLRLM, and Mean-shift feature extraction algorithms.

References

1. Agrawal, V., Chandra, S. (2015). Feature selection using Artificial Bee Colony algorithm for medical image classification. In Contemporary Computing (IC3), 2015 Eighth International Conference on IEEE. 171-176.
2. Ashwini Rejintal and Aswini N, "Image Processing based Leukemia Cancer Cell Detection", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, ICACT - 2016 Conference Proceedings, Special Issue – 2016.
3. Banday SA, Mir AH (2016) Statistical textural feature and deformable model based brain tumor segmentation and volume estimation. Multi-media Tools Applications.
4. Chris A, Mulyawan B (2012) A combination of feature selection and co-occurrence matrix methods for leukocyte recognition system. J Softw Eng Appl 5:101
5. Haralick, R. M., Shanmugam, K., Dinstein, I. H. (1973). Textural features for image classifcation. IEEE Transactions on systems, man, and cybernetics, 3(6), 610-621.
6. Hinge CS et al (2013) Classification of RBC and WBC in peripheral blood smear using KNN. Indian Journal of Research 2:2250–1991
7. Jothi, G., Inbarani, H. H. (2012). Soft set based feature selection approach for lung cancer images. arXiv preprint arXiv:1212.5391
8. Jothi, G., Inbarani, H. H. (2016). Hybrid Tolerance Rough Set–Firefly based supervised feature selection for MRI brain tumor image classification. Applied Soft Computing, 46, 639-651.

9. Jothi, G., Inbarani, H. H., & Azar, A. T. (2013). Hybrid Tolerance Rough Set: PSO Based Supervised Feature Selection for Digital Mammogram Images. International Journal of Fuzzy System Applications (IJFSA), 3(4), 15- 30.
10. Jothi, G., Inbarani, H. H., Azar, A. T., Devi, K. R. (2018). Rough set theory with Jaya optimization for acute lymphoblastic leukemia classification. Neural Computing and Applications, 1-20.
11. Khan S, Hussain M, Aboalsamh H (2015) A comparison of different Gabor feature extraction approaches for mass classification in mammography. Multimedia Tools Application 76:33–57.
12. Mohapatra, S., Patra, D., Satpathy, S. (2014). An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Computing and Applications, 24(7-8), 1887-1904.
13. Moradi Amin, M., Memari, A., Samadzadehaghdam, N., Kermani, S., Talebi, A. (2016). Computer aided detection and classification of acute lymphoblastic leukemia cell subtypes based on microscopic image analysis. Microscopy research and technique, 79(10), 908-916
14. Putzu, L., Caocci, G., Di Ruberto, C. (2014). Leucocyte classification for leukaemia detection using image processing techniques. Artificial intelligence in medicine, 62(3), 179-191
15. Sahlol, A. T., Abdeldaim, A. M., Hassanien, A. E. (2018). Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm. Soft Computing, 1-16.
16. Samant, P., Agarwal, R. (2018). Machine learning techniques for medical diagnosis of diabetes using iris images. Computer methods and programs in biomedicine, 157, 121-128.
17. Srisukkham, Worawut, Li Zhang, Siew Chin Neoh, Stephen Todryk, and Chee Peng Lim. (2017). Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization. Applied Soft Computing 56, 405-419.
18. Wahhab, H. T. A. (2015). Classification of acute leukemia using image processing and machine learning techniques (Doctoral dissertation, University of Malaya).