Alfarama Journal of Basic & Applied Sciences

Faculty of Science Port Said University

https://ajbas.journals.ekb.eg
ajbas@sci.psu.edu.eg
http://sci.psu.edu.eg/en/

April 2024, Volume 5, Issue II (Special Issue)

DOI:https://doi.org/10.21608/AJBAS.2023.228593.1170

ISSN 2682-275X

Submitted: 16/08/2023
Accepted: 20/12/2023

Pages: 243 – 261

Enhancing Diagnostic Decision-Making with Image Mining Techniques: A Proposed Framework for Medical Images

Doaa E. Mossa[1,*], Mahmoud Y. Shams[2], Ahmed A. Salama[1]

[1]Department of Mathematics and computer Sciences, Faculty of Science, Port Said university, 42526, Egypt.
[2]Faculty of Artificial Intelligence, Kafrelsheikh University, 33516, Egypt.

***Corresponding author: daadooezzat@gmail.com**

ABSTRACT

The field of medical image mining has garnered significant attention from researchers and professionals alike. This paper delves into the challenges and issues associated with medical images, such as low accuracy, poor quality, and false features. In response, we propose a prototype framework that utilizes image processing and data mining to enhance diagnostic decision-making through the extraction of relevant features from medical images. Firstly, the framework implements image processing algorithms to address problems related to brightness and imaging environment, thereby improving the quality of targeted medical images. Secondly, image mining techniques, such as segmentation and clustering, are employed on the processed images to identify and extract pertinent indicators. The model is trained iteratively using reference images, and classification techniques are utilized to identify features in test medical images. The prototype, developed using MATLAB, was tested on medical images of patients suspected to have leukemia. Results demonstrate that the proposed framework outperforms many comparable models using the same dataset, with a maximum accuracy of 98% achieved using K-means segmentation and Support Vector Machine (SVM) classificatio, compared to the 85-95% accuracy of commonly used frameworks for leukemia diagnosis. Validation of the proposed model confirms its adequacy and highlights the value added by incorporating image mining after preprocessing medical images using typical image enhancement techniques.

Key Words:

Image Mining, Data Mining (DM), K-mean segmentation cluster classification and Support Vector Machine (SVM).

1. INTRODUCTION

Medical imaging is a valuable diagnostic tool for physicians and surgeons, but inaccurate or misinterpreted information from images can lead to serious consequences [1], [2]. Data Mining (DM) has

been widely recognized as a powerful approach to improving the quality of prediction, diagnosis, and disease classification using medical images [3], [4]. However, using medical imaging for diagnosis, particularly in dangerous cases like leukemia, presents significant challenges. Radiologists derive detailed information from high-resolution images to provide multi-oriented views for better clinical diagnosis. Despite the benefits of medical imaging, various problems can hamper the diagnosis process and lead to improper diagnosis or incomplete identifications. Medical images contain massive amounts of data that can be processed to extract valuable information, but noise and interference often degrade their quality [5], [6]. Image mining techniques can help enhance the accuracy of features extracted by radiologists and enable better computer-generated information to aid decision-making. Techniques such as image processing and segmentation can improve the quality of medical images, while recent research has focused on combining image processing and image mining techniques to improve diagnostic decision-making using medical images. This paper proposes effective image mining techniques to extract reliable information from medical images and address related challenges. The main objectives of the paper are listed as follows:-

1. **Framework Proposal:** Introduce a five-stage framework for medical image diagnosis, encompassing preparation, preprocessing, image transformation, image mining, and results visualization.
2. **Verification and Validation:** Verify and validate the proposed framework through experimentation, comparing results with real reference data and other existing systems, using commonly adopted ALL-DB datasets for leukemia diagnosis.
3. **Identification of Challenges:** Identify and articulate the challenges faced by physicians in medical imaging for diagnostics, including issues related to imaging environment, data extraction accuracy, and interpretation challenges.
4. **Improvement in Imaging Systems:** Highlight the drawbacks of current medical imaging systems, particularly film-screen radiography, and emphasize the need for improved digital imaging systems for better diagnostic capabilities.

The research addresses challenges and limitations in medical imaging for diagnostic purposes, focusing on problems related to imaging environments including:

- Accuracy in data extraction, interpretation challenges, and drawbacks in existing imaging systems, particularly film-screen radiography.
- The gap in existing research lies in the absence of a comprehensive framework that integrates preparation, preprocessing, image transformation, image mining, and results visualization for medical image diagnosis.
- Additionally, there is a need to address the challenges and drawbacks specific to medical imaging, especially in the context of leukemia diagnosis.

The paper addresses critical issues in medical imaging for diagnostic purposes, focusing on challenges related to accuracy in data extraction, interpretation, and drawbacks in existing systems, particularly film-screen radiography. The main issues that will be handled in this paper can be summarized as follows:

1. The paper proposes a five-stage framework for medical image diagnosis, filling the gap in existing research by integrating preparation, preprocessing, image transformation, image mining, and results visualization.
2. Verification and validation of the proposed framework are emphasized through experimentation, comparing results with real reference data and existing systems, using established datasets for leukemia diagnosis.
3. The identification and articulation of challenges faced by physicians in medical imaging provide insights into issues related to imaging environments, data extraction accuracy, and interpretation challenges.

4. Furthermore, the paper underscores the drawbacks of current medical imaging systems, advocating for improved digital imaging systems to enhance diagnostic capabilities, particularly in the context of film-screen radiography.

The main Contribution of the paper are presented as follows:

1. Identification the challenges faced by physicians in medical imaging, offering insights into issues related to imaging environments, data extraction, and interpretation challenges.
2. Recommendations are made for improving digital imaging systems, particularly in the context of film-screen radiography, to enhance diagnostic capabilities.
3. The application of the Support Vector Machine (SVM) data mining technique, combined with K-means segmentation, is highlighted for improved and more accurate identification of features in leukemia samples.
4. The paper demonstrates high accuracy results (98%) in leukemia diagnosis using the proposed framework and the SVM data mining technique, showcasing its effectiveness with benchmark datasets commonly used for identical leukemia patients (ALL-IDB1, ALL-IDB2).

The rest of the paper are structured as follows: Section 2 presents an overview of related work and ongoing endeavors in diagnosing medical images. Section 3 details the materials and methods employed in this study. Section 4 delves into an investigation of the proposed method, while Section 5 provides an illustration of the discussion and analysis of results.

2. RELATED WORK

Medical imaging is an important part of the contouring process during radiation preparation. Medical Imaging Techniques (MITs) are techniques that are minimally invasive and allow you to view inside the body without having to cut it open. It was originally used to help with the diagnosis or treatment of a variety of ailments. Medical imaging technology produces pictures that are used by a range of medical specialists to retrieve, diagnose, and investigate illnesses. The most common forms of medical scans include X-rays, CT (Computed Tomography), MRI (Magnetic Resonance Imaging), ultrasonography, PET (Positron Emission Tomography), and SPECT (Single Photo Emission Computed Tomography). These medical pictures, which have become an essential element of the patient's treatment, contain clinical information, physiological patterns, soft and hard tissues, as well as malignant and benign tumors of the body like brain, bone, teeth, soft tissues, blood vessels, breast, and stomach.

Medical image mining poses several challenges that are specific to this field. These challenges can be summarized as follows:

- Enhancing and restoring medical images to improve their quality and clarity
- Developing automated and accurate segmentation techniques to identify features of interest within medical images
- Registering and fusing multimodality images to provide a comprehensive view for better diagnosis
- Classifying image features and characterizing and typing structures to aid in diagnosis and treatment
- Quantitatively measuring image features and interpreting the measurements to extract valuable information
- Developing integrated systems for the clinical sector to improve diagnosis and treatment
- Developing representation of images that can encode the textual information hidden within them for successful image mining
- Overcoming the obstacle of automatic decision criteria derivation for clustering
- Proposing a suitable indexing method and standardizing indexing and retrieving procedures for better knowledge extraction from images
- Developing a query language that can request visual patterns and textual metadata related to an image and unifying it
- Utilizing the vast amount of information available on the World Wide Web by treating it as an image database containing a huge volume of images.

Addressing these challenges is crucial for successful medical image mining and extracting valuable information from medical images to aid in diagnosis and treatment diagnosis [7].

As a subset of data mining, medical image mining focuses on analyzing the features and added data of the images. The main challenge in this situation is to correctly identify and adopt the most effective mining method, maybe combining a number of these techniques over the course of several rounds and based on comparing outcomes. Image mining techniques, like data mining techniques, involve a variety of algorithms. The fact that medical images have high grey scale, clear meaning, easy reconstruction, and other properties makes the situation more difficult. Such features as the medical image's grey scale, density distribution, color features, and others must be defined using an effective algorithm. While waiting, it should think about the data mining algorithm based on image feature extraction, and to explore the gray levels and its density distribution along with clustering analysis of normal human tissues and organs.

Context-based image retrieval (CBIR) is a technique for recovering digital photographs from vast databases based on their content. CBIR, as opposed to manual image mining, employs machine vision to analyse pictures and extract characteristics in order to accomplish a certain pattern. Query by Image Content (QBIC) and content-based visual information retrieval are other terms for CBIR (CBVIR).

The CBIR approach takes a query picture as input and ranks target photos in a database based on their resemblance to the query image. The result is a collection of photos ordered by their similarity to the query image. CBIR systems frequently employ techniques such as feature extraction, indexing, searching, and ranking to achieve accurate and efficient image retrieval.

The process of evaluating and extracting features from photographs, such as colour, texture, shape, and spatial information, is known as feature extraction. The CBIR system then indexes and searches for photos in the database using these features. The ranking procedure entails sorting the obtained photos according to their similarity to the query image. CBIR is a goal-oriented image-to-image search engine. For retrieval, a database of target photos is necessary. The target photos that are closest to the query image are returned. We can use the image directly for similarity, however there are several issues:

- The picture is massive in size.
- Pixels have a lot of redundancy.
- The semantic information is not carried by a pixel [8].

Table 1 presents a comprehensive overview of various methodologies and techniques employed in different studies for the diagnosis of medical images, particularly focusing on leukemia. Granero et al. [9] utilized a methodology based on four-moment statistical features and Artificial Neural Network (ANN). The study, conducted on the ALL-IDB dataset, employed techniques such as K-Means and SVM, achieving a segmentation accuracy of 97% in binary classification. Sallam et al. [10] presents Various supervised classifiers, including Random Forest (RF), K Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naïve Bayes (NB), were subjected to comparison. The presented methodology attains an impressive level of accuracy, registering at 99.22%, accompanied by precision and sensitivity rates both reaching 99% using a dataset from Kaggle Lymphoblastic Leukemia (ALL) image.

Boldú et al. [11] employed mathematical morphology and color clustering in their methodology for peripheral blood images. Using various machine learning techniques, the study achieved an overall accuracy of 94%, incorporating assessments based on color, texture, and geometric features. Rehman et al. [12] implemented a Convolutional Neural Network (CNN) and robust cell segmentation for stained bone marrow images. The deep learning approach resulted in an impressive accuracy of 97.78%, with a processing time of 163.63 seconds. Al Mamun et al. [13] implemented advanced classifiers and features for the analysis of peripheral blood smear images. The study reported an impressive accuracy of 99.22%, emphasizing an automated recognition system. Tusar and Anik [14] employed Deep Neural Networks for the analysis of microscopic images, achieving an accuracy of 98% in identifying shapes of ALL blast

cells. Jusman et al. [15] applied GoogleNet and CNN methodologies to human blood samples, achieving an accuracy of 96.06% for GoogleNet and 94.69% for CNN.

Overall, these studies collectively contribute to the diverse landscape of methodologies and techniques in medical image diagnosis, offering varying levels of accuracy and efficiency for leukemia diagnosis. The results provide valuable insights into the evolving state of the art in medical image analysis for this specific application.

Table 1. The comparative study between the recent approaches and the proposed model.

Reference	**Methodology**	**Dataset**	**Performance Metrics**
Granero et al. [9]	Four-moment statistical features, Artificial Neural Network (ANN)	ALL-IDB	Segmentation Accuracy: 97%, Binary Classification
Sallam et al. [10]	Random Forest (RF), K Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naïve Bayes (NB)	from Kaggle Lymphoblastic Leukemia (ALL) image	Accuracy: 99.20%
Boldú et al. [11]	Mathematical morphology, color clustering	Peripheral Blood Images	Overall Accuracy: 94%, Color, Texture, Geometric Features
Rehman et al. [12] i	Convolutional Neural Network, robust cell segmentation	Stained Bone Marrow	Accuracy: 97.78%, Processing Time: 163.63 s
Al Mamun et al. [13]	Advanced classifiers, features	Peripheral Blood Smear	Accuracy: 99.22%, Automated Recognition System
Tusar and Anik [14]	Deep Neural Networks	Microscopic Images	Accuracy: 98%, Identifying Shapes of ALL Blast Cells
Jusman et al. [15]	GoogleNet, CNN	Human Blood Samples	Accuracy (GoogleNet): 96.06%, Accuracy (CNN): 94.69%
Proposed Work	SVM	ALL DB1 and ALL DB2	Accuracy: 98.00%

3. MATERIALS AND METHODS

The proposed model is organized into five phases while combining both image processing algorithms and image mining techniques as shown in the following workflow shown in Figure 1.

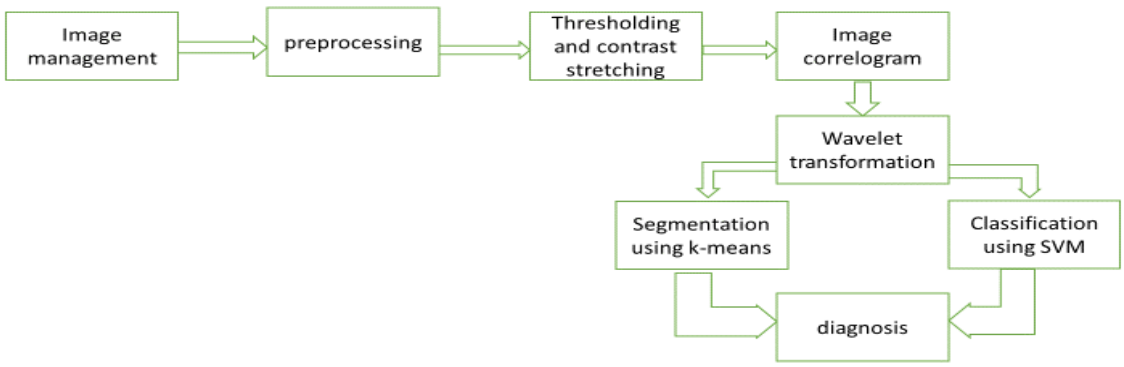


Fig 1. Image mining workflow diagram.

The key phases of the suggested model, which are also illustrated in Figure (2), are as follows:

- Phase one: preparation phase
- Phase two: preprocessing phase
- Phase three: transformation phase
- Phase four: applying image mining techniques
- Phase five: diagnosis

The suggested prototype will be used to diagnose leukaemia using tiny blood pictures. Based on the target microscopic images, a colour and shape-based segmentation algorithm is initially used to detect and segment WBCs. The nucleus and cytoplasm are then segmented using K-means clustering and region growth. Following that, many characteristics reflecting the nucleus and cytoplasm sub-images' shape, texture, colour, and statistical-based information are retrieved. Following that, a Support Vector Machine (SVM) classification is used to distinguish between healthy (normal) and unhealthy (abnormal) cells. Meanwhile, the system considers the nucleus and cytoplasm in both segmentation and feature extraction to provide a high-quality diagnostic. The key steps of the processing operations are described briefly below.

3.1 The Transformation phase: In this phase the core idea is to apply a number of transformation tasks on the preprocessed target RGB medical image into a number of calculated/processed transformations. The considered transformation tasks include: HCV histogram, Color moment calculations, Correlogram, wavelet transformation and segmentation. Following is and following is a brief description preceded with spot lightening the main challenges induced by medical image to gain an insight about the importance of the transformation phase.

3.2 HSV histogram and Correlogram: In this transformation option, the RGB picture is converted into the HSV color space. A color histogram is then computed based on the HSV values to analyze different hues in the image. From the HSV histogram, a Correlogram is obtained, which is a table indexed by color pairs and their probabilities of occurrence at a certain distance **[16]**. The auto-correlogram focuses on spatial correlation between similar hues. However, both correlograms and auto-correlograms are computationally expensive. To address this, a minimal number of colors and distance values are used in the correlogram to achieve good results without increasing the computational cost. Color moments are utilized as characteristics to identify images based on color. The color distribution of a picture is viewed as a probability distribution, and moments describe different aspects of this distribution **[17]**. In this case, three key moments—mean, standard deviation, and skewness—are employed for each of the three color channels (Hue, Saturation, and Brightness in the HSV system). Each picture channel is associated with

three moments, resulting in a total of nine moments to define a picture. The mean represents the average color value, the standard deviation measures the distribution's variance, and the skewness indicates the asymmetry of the distribution. This approach allows the identification of images based on their color characteristics **[18]**.

Following that, in the transformation, the colour moments are obtained, which are metrics that may be used to distinguish pictures based on their colour attributes. After calculating colour moments, it is feasible to discover colour similarity between photos. Similarity values can then be compared to picture index values in a database for applications like image retrieval.

3.3 Gray Scale Image: The grey value of an image is calculated using an equation involving the red, green, and blue values. These color channels are then processed using a bank of two-dimensional Gabor filters. A Gabor filter combines a sinusoidal wave with a specific frequency and direction, modulated by a two-dimensional Gaussian envelope. The filter's frequency and orientation selectivity are more apparent in its representation in the frequency domain. In the frequency domain, the Fourier representation of the filter describes how each frequency component of the input image is modified or affected. This representation is referred to as the modulation transfer function (MTF). To perform texture segmentation, it is necessary to analyze both the spatial characteristics and the spatial-frequency characteristics of the image simultaneously **[19]–[21]**.

3.4 Wavelet transformation: The goal is to apply a discrete wavelet transform and extract moments from it. The Discrete Wavelet Transform (DWT) offers more flexibility and provides richer information compared to other transformation methods. It is a technique that decomposes data into subbands or frequency components. In contrast, Fourier transform breaks down a signal into sine waves of different frequencies. One advantage of wavelets over Fourier is that wavelets are not limited by a fixed period but extend from negative to positive infinity. In the Fourier transform domain, we lose complete information about the original signal. On the other hand, wavelet coefficients are localized and easier to interpret. Wavelets are designed for adaptive systems as they can adapt to changes in the signal, whereas the Fourier transform is suitable when the signal consists of only a few fixed components.**[22]**.

3.5 Segmentation process: It's worth noting that, despite the numerous segmentation methodologies proposed in the research and in practise, no one strategy has been advocated that is effective for all sorts of applications. The properties of a picture are used to split it into pieces. Each zone is homogeneous in terms of brightness, colour, texture, and reaction. This helps identify lumps, tiny calcium deposits, and possible lesions. Furthermore, by breaking up the thick tissue parts, it assists in roughly determining breast density. The purpose is to continue processing the previous stage's image and convert it to L*a*b* (CIELAB) colour space, which captures all of the colours experienced by the human eye and may be used as a reference across applications. The Smoothing filter is then used to eliminate the noise and artefacts from the picture. Low pass filters and high pass filters are the two types of filters. A smoothing filter is a low pass filter. It is used to eliminate high spatial frequency noise in digital photographs. The smoothing filter's moving window operator alters the value of each pixel in the picture one at a time by some function of a particular region of the image.

3.6 Image Mining Phase:

3.6.1 Main Steps of Image Mining: The Smoothing filter is then used to eliminate the noise and artefacts from the picture. Low pass filters and high pass filters are the two types of filters. A smoothing filter is a low pass filter. It is used to eliminate high spatial frequency noise in digital photographs. The smoothing filter's moving window operator alters the value of each pixel in the picture one at a time by some function of a particular region of the image. The Smoothing filter is then used to eliminate the noise and artefacts from the picture. Low pass filters and high pass filters are the two types of filters. A smoothing filter is a low pass filter. It is used to eliminate high spatial frequency noise in digital photographs. The smoothing filter's moving window operator alters the value of each pixel in the picture one at a time by some function of a particular region of the image. The steps of such a chosen criterion are as follows:

- One computes the squared distance from the appropriate cluster centre for each pixel.
- The total of these distances for all points in the data set is calculated. The cluster's centre is chosen, and each pixel in the picture is assigned to it.
- The cluster centre is recalculated using the average of all pixels.
- This loop continues, and the next step is to connect each point in a given data set with the closest centroid **[23]**.

3.6.2 K means Algorithm for Segmentation: Classifier-based segmentation use pattern recognition algorithms to separate feature vectors extracted from a photograph into an expected set of classes **[24]**. These classes are likely to be used whenever features are produced as a result of reference segmentations, such as manual segmentation. The characteristics employed in the categorization process are various, and might be related to the intensity, texture, or other properties of a picture. A simple technique is used to randomly seek for appropriate segmentation to highlight the potential of categorization. The k - nearest neighbours (KNN) classification is a simple classifier in which each object is categorised based on a majority vote of the circumstances k - closest training samples. In image segmentation, each pixel in the feature space is labelled with the most common class among the k - closest neighbours of its consistent feature vector. Class-based segmentation has been effectively used to medical picture segmentation, such as pulmonary nodule detection in the chest, reduction in erroneous assumptions for breast cancer diagnosis, and brain tissue segmentation in MR data. The following pseudo code summarises the K-means Algorithm (Algorithm 1).

Algorithm 1: The steps of K-means Algorithm

1 Input: K, a collection of x1.....xn points.
2 Distribute the centroids C1.....Ck at random.
3 Repeat until convergence is reached:
For each point x_i:
Find nearest centroid Cj.arg min D(xi,Cj) Assign the point
x_i to cluster j For each cluster $j=1.....K$:

$$c_j(a) = \frac{1}{n_j} xi - c_j \sum_{i=1}^{n} x_i(a) \quad (1)$$

For each cluster $j=1.....K$, create a new centroid C_j
by taking the mean of all the points x_i assigned to cluster j in the previous step.
4 When none of the cluster allocations change, stop.

The k-Means clustering algorithm is used to generate three distinct segments in an image, and then characteristics are extracted from these segments. Texture analysis provides information about the spatial distribution of intensities in the image. One approach to extract texture features is using the Gray Level Co-occurrence Matrix (GLCM), which describes the relationships between neighboring pixels.

The GLCM calculates second-order statistics, such as the probability of two pixels having specific grey levels at certain spatial connections. These statistics are represented as 2-dimensional matrices, which can be computed for different distances and orientations. Various established statistical metrics in medical practice can be used to extract textual features from the GLCM. Some of these features include:

a. Energy (Uniformity): This measures the visual homogeneity or uniformity within the image.
b. Contrast: It captures the local variances or differences in intensity within the image.
c. Entropy: It quantifies the level of disorganization or randomness in the image's texture.

d. Correlation: This feature measures the linear dependencies or patterns within the image.
e. Homogeneity: It indicates how close the distribution of GLCM elements is to the GLCM diagonal, reflecting the uniformity of texture.
f. These features provide valuable information about the form and texture characteristics of the image, particularly in analyzing nuclei or other structures of interest **[25]**.

Skewness and kurtosis are added to the texture descriptors in addition to the GLCM-induced features. Skewness describes the frequency of extreme values in a distribution, whereas Kurtosis depicts the imbalance and asymmetry from the mean.

3.6.3 Image Mining Using Support Vector Machine (SVM) Classification: Support vector machines (SVM) are machine learning algorithms that analyse data used for categorization and regression analysis. SVM is used to detect changes in brain structure using magnetic resonance images, which aids in the diagnosis of neurological diseases. It is used to aid in the diagnosis of cancer images using a genetic algorithm that classifies tumours as benign or malignant **[26]**.

Following the application of the aforementioned segmentation job, the model moves on to classification using the Support Vector Machine (SVM) technique. Because of its ease of use, the SVM is a popular data mining method. It is regarded as an alternative to the SVM algorithm depicted in the preceding block diagram. In this stage, the input is further classified into three categories:

- A statistical method in which the vectors have a multidimensional feature space as well as a complicated frequency pattern.
- Structural method, in which the texture is ordered to complete the pattern at the texture primitive level, and then a formal picture is made between the structural pattern and the syntax of language. In the meanwhile, the frequencies outperform the statistical and spectral approaches.
- The photos are autocorrected in a spectral way, however the frequencies are less efficient.

Furthermore, in order to obtain a better understanding of how the SVM works, we will briefly demonstrate it next. In other words, the SVM model is a representation of the instances as points in space, mapped such that the examples of the different categories are separated by as wide a distance as feasible. New instances are then mapped into the same space and classified based on which side of the gap they land on. SVM is excellent in analysing the separating planes and identifying the biggest margin to offer support to the data points. So we'd want to learn how to map: X->Y, where x' belongs to X and y' belongs to Y and is a class label **[27]**.

It's also worth noting that this algorithmic strategy is based on different features analysis by examining predicted error reduction. It takes into account the empirical risk in order to enhance the training technique. The risk calculation in this case is based on structural analysis, which reduces generalisation error. The error margin is assessed under class deviation and closest training patterns are produced based on it. Furthermore, the proposed framework is built on the polynomial kernel representation, which allows for more effective learning to the elements and higher accuracy. The following is a fictitious explanation of the major steps in classification and regression:

1. First, the root is chosen.
2. The root is separated into two nodes to decrease impurity.
3. The node, which has been separated into two children, then chooses the independent and explanatory data.
4. Repeat the preceding two steps with the resultant node as parent until the maximum data size is found.
5. The tree is then eliminated by a set of nodes based on cross validation and cost complexity.

4. PROPOSED METHODS

The proposed model includes the main phases that aid the preprocessing stages to prepare the data for diagnosis and classification. The main goal of preparing stage is to collect the medical images targeted for analysis via a number of commonly used criteria and methods including color images, blood elements,

resolution and texture. Meanwhile, this phase is concerned with proper preparation and labeling of collected images as well as identifying and proper organizing to the collected data set keeping in mind some of the standard commonly used data sets according to the domain of diagnosis.

The most frequent kind of juvenile cancer is acute lymphoblastic leukaemia (ALL). ALL develops when bone marrow stem cells get DNA abnormalities that cause them to overproduce. As a result, immature stem cells known as lymphoblasts proliferate throughout the body, causing a slew of aseptic symptoms such as bruising, bleeding from the mouth or nose, infections, bone pain, fever, enlarged lymph nodes, shortness of breath, and weakness. Thrombocytes, erythrocytes (red blood cells, RBCs), and leukocytes are examples of typical blood components (white blood cells, WBCs). A stain is put to a peripheral blood smear to distinguish them under a microscope. Thrombocytes are tiny bluish-purple pieces. RBCs are more numerous than other components and appear as greyish-pink biconcave discs. WBCs have a dark blue-purple nucleus and are divided into five types: neutrophils, lymphocytes, eosinophils, basophils, and monocytes. WBC subclassification necessitates the evaluation of several morphological characteristics such as cell size, cytoplasm colour, the presence of blue or red staining granules, the number of nuclear lobes, the cytoplasm to nucleus ratio, and the presence of subcellular components such as vacuoles and nucleoli **[28]**.

Based on a variety of characteristics and signs, lymphocytes are seen in the microscopic examination in the case of ALL as follows:

- Lymphoblasts can differ in appearance from regular lymphocytes in addition to having a distinct shape. Normal lymphocytes have a spherical, blue-purple stained nucleus that is about the same size as an RBC. The nucleus lacks nucleoli, has smooth borders, and is packed with closed chromatin. Depending on how reactive the lymphocyte is, the cytoplasm is sparse but could also be copious. It also has a pale blue stain.

- In contrast, a lymphoblast may have a larger nucleus that stains sparse red purple. The nucleus may contain distinct nucleoli, have open chromatin, and be indented with rough boundaries. The cytoplasm may stain deep blue but is otherwise scanty **[29]**. The steps of the proposed model are shown in details in Figure 2 and Algorithm 2.

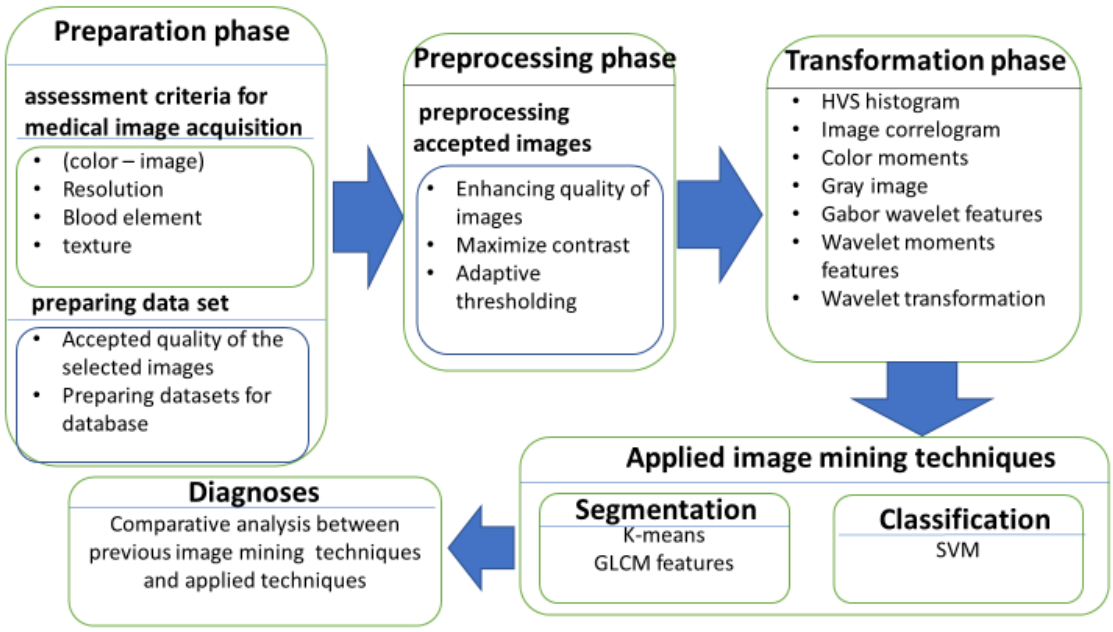


Fig 2. Main Phases of the proposed framework

Algorithm 2: The Pseudo code for the proposed main phases

Phase 1: Preparation

Collect Images: Collect a large dataset of images relevant to the task at hand.

Preprocess Images: Preprocess the images to ensure they are of a consistent format and quality. This may involve resizing, normalizing, and converting to grayscale or another desired format.

Build Database: Organize the preprocessed images into a database for efficient storage and retrieval.

Phase 2: Preprocessing

Noise Removal: Apply noise reduction techniques to remove unwanted artifacts from the images.

Segmentation: Segment the images into meaningful regions or objects of interest.

K-means for Segmentation:

Initialization: Randomly choose k centroids (cluster centers) within the data space.
Assignment: Assign each data point to the closest centroid, forming initial clusters.
Recomputation: Update the centroids by calculating the mean of each cluster.
Iteration: Repeat steps 2 and 3 until the centroids converge (no significant change in cluster assignments).
Output: Each data point is assigned to a cluster, effectively segmenting the image into different regions.
Additional details:
K-means and SVM can be combined in a pipeline where K-means segments the image into regions and SVM classifies each region based on its features.
The choice of k in K-means and the kernel and parameters in SVM depend on the specific data and task and require experimentation to find the optimal settings.
Other algorithms can be used for segmentation (e.g., mean-shift, watershed) and classification (e.g., decision trees, random forests) depending on the specific needs.

Feature Extraction: Extract features from the segmented images that are relevant to the task at hand. These features could be color, texture, shape, or other properties.

Phase 3: Transformation

Feature Selection: Select a subset of the extracted features that are most relevant and informative for the task.

Dimensionality Reduction: Reduce the dimensionality of the feature space to improve computational efficiency and avoid the curse of dimensionality.

Data Normalization: Normalize the selected features to ensure they are on a common scale.

Phase 4: Data Mining

Model Building: Apply data mining algorithms to the transformed data to discover patterns and relationships between the features. This could involve clustering, classification, association rule mining, or other techniques.

Model Evaluation: Evaluate the performance of the built model on a separate set of test data.

Refinement: Refine the model based on the evaluation results to improve its accuracy and generalizability.

SVM for Classification:

Input: Training data with labeled examples (x_i, y_i), where x_i is an image feature vector and y_i is the corresponding class label.

Initialize: Choose values for C and kernel.

Optimize: Solve the quadratic programming problem to find the optimal decision boundary hyperplane that maximizes the margin between classes. This typically involves iterative optimization algorithms.

Classification: For a new data point x_new, calculate its distance to the hyperplane and predict its class based on the side of the hyperplane it falls on.

Phase 5: Interpretation

Interpret the Results: Interpret the discovered patterns and relationships in the context of the specific task.

Draw Conclusions: Draw conclusions and make informed decisions based on the insights gained from the image mining process.

Validate Findings: Validate the findings by testing them on real-world data or applications.

Acute lymphoblastic leukaemia (ALL) is manually categorised as follows: Despite these classifications, ALL subtypes show substantial variation. It is difficult to distinguish between normal lymphocytes and abnormal lymphoblasts manually in an objective and accurate manner. The accuracy of the findings vary according to the observer's ability and attentiveness, as well as the calibre of the blood sample. Furthermore, manual microscopy is a time-consuming technique that takes several minutes to complete.

The purpose of automating microscopy in cell biology is to improve patient diagnosis through enhanced speed and accuracy. Image processing software is developed to extract essential information from medical photographs. This research might pave the way for the development of new techniques such as colour imaging, segmentation, classification, and clustering. Because the ambiguity of the symptoms and indications of ALL frequently leads to a false diagnosis, haematologists struggle with picture categorization, which is time-consuming and prone to mistake. As a result, rapid, accurate, and automated picture recognition is required. Pre-processing techniques are used to improve the quality of photos by enhancing contrast and eliminating faults in order to obtain more improved and/or error-free images. The picture quality is critical for the outcome of any medical study since it impacts both the capacity to detect characteristics under analysis and the precision of subsequent measurement. Imaging issues are fairly widespread, resulting in photos with flaws like noise or distortions like scratches, lapping tracks, and comet tails **[30]**. Image Collection Transformation Segmentation Classification Preprocessing. Pre-processing techniques are used to improve the quality of photos by enhancing contrast and eliminating faults in order to obtain more improved and/or error-free images. The picture quality is critical for the outcome of any medical study since it impacts both the capacity to detect characteristics under analysis and the precision of subsequent measurement. Imaging issues are fairly widespread, resulting in photos with flaws like noise or distortions like scratches, lapping tracks, and comet tails. Image Collection Transformation Segmentation Classification Preprocessing.

The ALL-IDB dataset, on the other hand, is the foundation for the testing experiments. Throughout our study and model validation, we employ the widely available ALL_IDB dataset. A Canon Power Shot G5 camera and an optical microscope were used to take pictures of the adopted dataset. Every image that was recorded was a JPG file with a resolution of 2592 x 1944 and a 24-bit color depth. The ALL_IDB1 version 1.0, which is utilized for testing both segmentation and classification systems as well as conventional image preprocessing techniques, is the foundation for testing. The dataset utilised consists of

108 images obtained in September 2005 that comprise about 39000 blood components; the lymphocytes in these blood components were identified by trained oncologists. The images are then collected at various microscope magnifications ranging from 300 to 500.The ALL-IDB2 image collection has also been created to evaluate the effectiveness of categorization algorithms. The dataset utilised consists of 108 images obtained in September 2005 that comprise about 39000 blood components; the lymphocytes in these blood components were identified by trained oncologists. The images are then collected at various microscope magnifications ranging from 300 to 500. Figures 3, and 4 investigates the examples of the images contained in ALL-IDB1, and ALL-IDB2, respectively. Throughout, our testing procedures we used 44 healthy images and 33 patient images from IDB1 and extracted the transformation features and the segmentation features explained before using the proposed algorithm tool and saved it in a mat file that stores these features. Also, we used 99 healthy cell image and 99 patient lymphoblast image from ID21 and extracted the transformation features and the segmentation features explained before using the proposed algorithm tool and saved it in the same mat file that stores these features. So, the whole mat file contains the features of 275 images. ALL-IDB dataset is used for training and testing the proposed system.

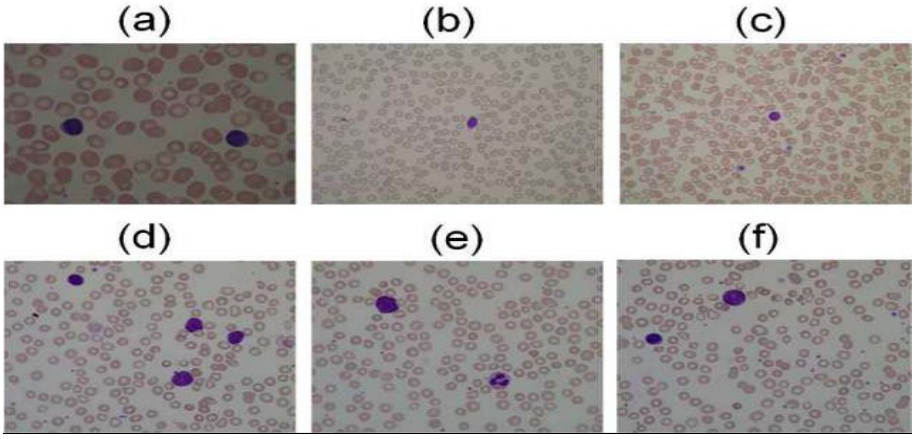
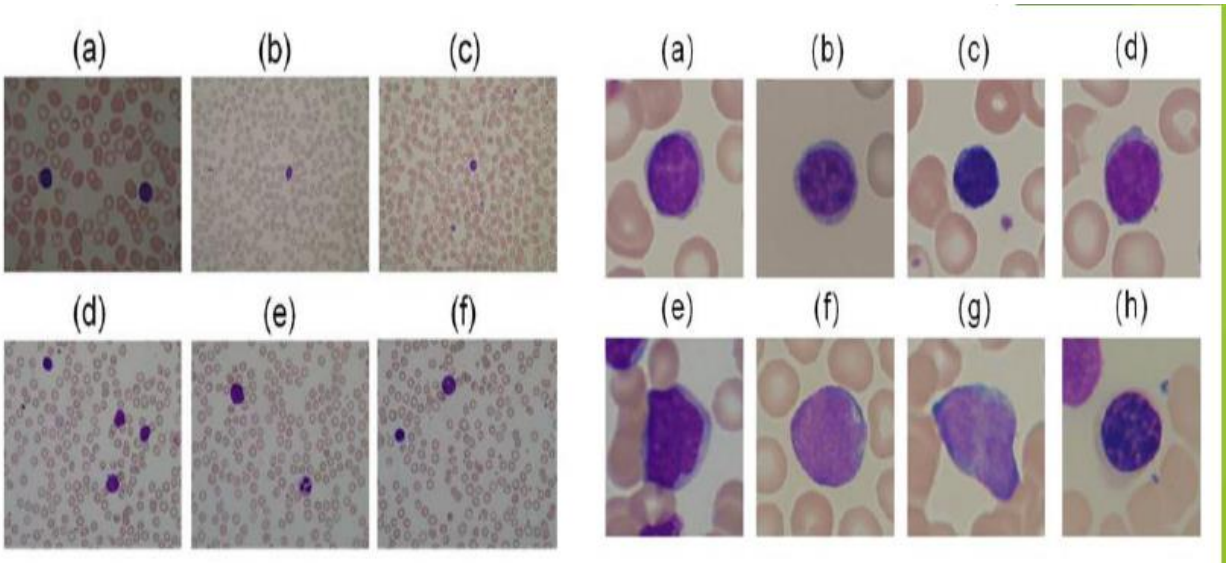


Fig 3. Instances from ALL-IDB1 include images depicting healthy cells from non-ALL patients (a-c) and potential lymphoblasts from ALL patients (d-f).



(I) (II)

Fig 4. Examples of the images in ALL-IDB2. (I) Healthy cells for non-ALL patients (a-f), and (II) probable lymphoblast's from ALL patients (a-h).

Throughout, our testing procedures we used 44 healthy images and 33 patient images from IDB1 and extracted the transformation features and the segmentation features explained before using the proposed algorithm tool and saved it in a mat file that stores these features. Also, we used 99 healthy cell image and 99 patient lymphoblast images from ID21 and extracted the transformation features and the segmentation features explained before using the proposed algorithm tool and saved it in the same mat file that stores these features. So, the whole mat file contains the features of 275 images. ALL-IDB dataset is used for training and testing the proposed system.

5. DISCUSSION AND RESULTS ANALYSIS

The experiments were carried out on a PC with a 3.30 GHz Core i5 processor and 16 GB of RAM, running under the Windows 8.1 operating system. As mentioned before the experiments used ALL-IDB detests for training and Testing. The following table shows how the dataset as organized (Table 1).

Table1. ALL-IDB detests for training and Testing.

Utilized Dataset	**Training Images**	**Testing Images**
ALL-IDB1	Healthy: 44 Patient: 33	Healthy: 15 Patient: 16
ALL-IDB2	Healthy cell: 99 Patient lymphoblast: 99	Healthy cell: 31 Patient lymphoblast: 31

The proposed system classification accuracy is compared in the following table with systems using ALL-IDB2 dataset, the proposed system has higher accuracy (Table 2).

Table2. The proposed system classification accuracy

Authors	Model	Accuracy
Mishra ,s, et al [31].	KNN	85%
Mishra ,s, et al [31].	CNN	88%
Siew Chin et al [32].	SVM	90%
Siew Chin et al [32].	MLP	95%
Proposed System	SVM	98%

Accuracy is a major performance measure considered here to compare results of the proposed system as compared to a number of the published systems. Comparison is made using same data set according to the following equation:

$$Accuracy = \frac{Tp+Tn}{Tp+Fp+Tn+Fn} * 100 \quad (2)$$

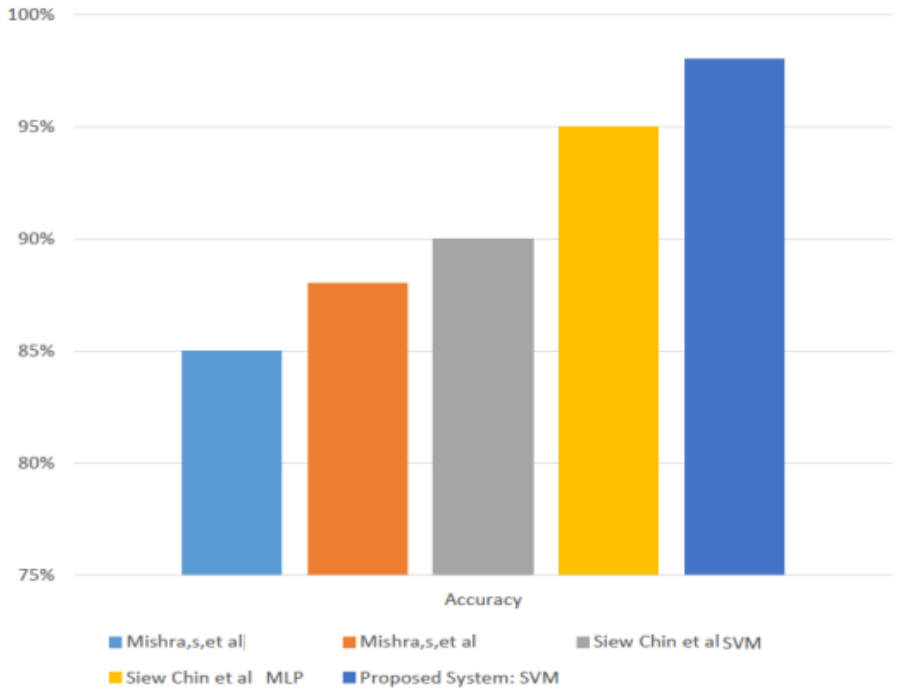


Fig 5. Comparison Chart of the Accuracy Performance Measures

Table 3 shows a detailed comparison which indicates that the proposed framework outperformed the alternative ones.

Table3. Comparative of Results of different approaches as Vs. the Proposed one.

System	Proposed Methodology	Accuracy
Mishra et al. [31]	The KNN was utilised, and it had the lowest, which is due to the model keeping the training set for measuring nearest neighbours. As a result, it was prone to data segmentation in situations where unusually noisy pictures or cell populations were unequally distributed in training or test sets. Despite this, the KNN performed well on the classification challenge, demonstrating the predictive potential of raw pixels from noisy cell-centered pictures.	85%
Mishra et al. [31]	Because of a single convolutional layer, the CNN utilised demonstrated higher accuracy. Anatomic pathology error, which includes cytology, has been found to have a mean error rate of 1-5%, with broad range.	88%
Chin Neoh et al. [32].	They presented an ALL detection decision support system. It incorporates a suggested SDM-based clustering algorithm that accounts for both within- and between-cluster scatter variations for robust nucleus and cytoplasm segmentation. The segmented nucleus and cytoplasm yield a total of 80 feature descriptors. These characteristics are fed into the SVM to identify lymphocytes and lymphoblasts.	90%
Chin Neoh et al. [32].	They presented an ALL detection decision support system. It incorporates a suggested SDM-based clustering algorithm that accounts for both within- and between-cluster scatter variations for robust nucleus and cytoplasm segmentation. The segmented nucleus and cytoplasm yield a total of 80 feature descriptors. These characteristics	95%

	are sent into the MLP to identify lymphocytes and lymphoblasts.	
Proposed System	The approach that is suggested in this paper uses multiple scientific aspects to classify and grade leukaemia samples using SVM. The photographs of the leukaemia samples are acquired and processed using various algorithms and filters. In order to extract various aspects, such as colour, texture, etc., created algorithms are used. K-means is used for segmentation and SVM is used for classification. SVM seeks to fit an optimal hyperplane between the classes and only uses training samples that lie at the edge of the class distributions in feature space (support vectors). This approach allows for the definition of the most informative training samples prior to the analysis, thereby improving the accuracy of the classification. The use of features and segmentation approaches, along with the classification algorithm that minimizes error margins as much as possible, make the proposed approach highly effective.	98%

6. CONCLUSION AND FUTURE WORK

This paper proposed a five-stage framework including preparation, preprocessing, image transformation and visualization for diagnosis leukemia disease. The proposed framework models have been verified and validated based on results compared to some real reference data as well as to a number of other systems using the same data set. Experiments made were based on the commonly adopted ALL-DB data sets which are commonly used for leukemia diagnosis [33]–[36]. A number of important findings and conclusions are summarized below:

- Medical Imaging in practice suffer from a number of problems related to equipment and environment but still present a common effective tool for many diseases' diagnostics. Such problems and challenges facing physicians from perspectives of diagnostics using medical images can be identified as: (i) problems associated with medical images due to imaging environment such as contrast problems, poor imaging and noise pickup, (ii) challenging facing extracting accurate justified data and information out of the medical images and (iii) challenges related to interpretation of extracted data at acceptable degree of certainty.
- Enhancing medical pictures is sometimes difficult, particularly for film screen radiography, due to a lack of sufficient image density and contrast, necessitating repeated imaging for acceptable image quality.
- Furthermore, there are drawbacks to film-screen radiography since the film-screen image receptor is not a suitable medium for performing the tasks of radiation detection, picture presentation, and image preservation. Because of their potential for picture post-processing, the aforementioned constraints give impetus for improving and inventing better digital imaging systems for use in digital radiology.
- Assessment of the proper image mining techniques whether selecting the right ones or combining and identifying their sequence of application represents the key ideas towards optimality of the medical imaging from perspectives of diagnosis decisions.

On the other hand, a detailed examination of the usage of picture mining has confirmed a number of crucial conclusions:

- Acute leukemia is a set of diverse fatal illnesses that afflict people of all ages, and early detection is critical to reducing morbidity and death.
- Acute leukemia is a set of diverse fatal illnesses that afflict people of all ages, and early detection is critical to reducing morbidity and death.
- The application of the SVM data mining technique following the K-means enabled improved and more accurate identification of features reflecting shape, texture, color, and statistical-based

information of the nucleus and cytoplasm sub-images. To distinguish between healthy (normal) and unhealthy (abnormal) cells, a Support Vector Machine (SVM) is used. To provide a high-quality diagnostic, the system considers the nucleus and cytoplasm during segmentation and feature extraction. Furthermore, the SVM-based classification and grading of leukemia samples utilizing several scientific criteria has shown to be highly useful. The classification employs k-means segmentation and SVM classification. SVM attempts to fit an ideal hyperplane between the classes and employs only those of the training samples that are on the edge of the class distributions in feature space (support vectors). Prior to the analysis, this should allow for the selection of the most informative training samples.

- The characteristics utilized, as well as the segmentation technique and classification algorithm, succeeded in lowering error margins as much as possible, valuing the accuracy results, which reached 98 percent utilizing the benchmark data sets ALL-IDB1, ALL-IDB2 typically used for identical leukemia patients.

As illustrated in Algorithm 1, we utilized K-mean in segmentation in preprocessing stage and then used SVM as a classifier in classification stage and the over all accuracy achieved is 98.00%.

In the future, a deep learning model can be utilized in order to expand the dataset with large scale database and for classification and feature extraction to boost the results obtained and tackle the overfitting problem.

7. REFERENCES

[1] M. Y. Shams, O. M. Elzeki, M. Abd Elfattah, T. Medhat, and A. E. Hassanien, "Why Are Generative Adversarial Networks Vital for Deep Neural Networks? A Case Study on COVID-19 Chest X-Ray Images," in *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*, Springer, 2020, pp. 147–162.

[2] M. Y. Shams, O. M. Elzeki, L. M. Abouelmagd, A. E. Hassanien, M. A. Elfattah, and H. Salem, "HANA: A Healthy Artificial Nutrition Analysis model during COVID-19 pandemic," *Computers in Biology and Medicine*, vol. 135, p. 104606, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104606.

[3] O. M. Elzeki, M. Shams, S. Sarhan, M. Abd Elfattah, and A. E. Hassanien, "COVID-19: a new deep learning computer-aided model for classification," *PeerJ Computer Science*, vol. 7, p. e358, 2021.

[4] O. M. Elzeki, M. Abd Elfattah, H. Salem, A. E. Hassanien, and M. Shams, "A novel perceptual two layer image fusion using deep learning for imbalanced COVID-19 dataset," *PeerJ Computer Science*, vol. 7, 2021.

[5] E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, "Covid-19 diagnosis-based deep learning approaches for covidx dataset: A preliminary survey," *Artificial Intelligence for Disease Diagnosis and Prognosis in Smart Healthcare*, p. 107, 2023.

[6] E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, "The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16591–16633, 2023.

[7] E. V. Ramana, S. Sapthagiri, and P. Srinivas, "Data mining approach for quality prediction of injection molding process through statistica SVM, KNN and GC & RT techniques," *Inject. molding*, vol. 1, p. 9, 2016.

[8] Y. Liu, J. Yu, and Y. Han, "Understanding the effective receptive field in semantic image segmentation," *Multimedia Tools and Applications*, vol. 77, pp. 22159–22171, 2018.

[9] M. A. Granero, C. X. Hernández, and M. E. Valle, "Quaternion-Valued Convolutional Neural Network Applied for Acute Lymphoblastic Leukemia Diagnosis," in *Intelligent Systems*, A. Britto and K. Valdivia Delgado, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 280–293. doi: 10.1007/978-3-030-91699-2_20.

[10] N. M. Sallam, A. I. Saleh, H. Arafat Ali, and M. M. Abdelsalam, "An efficient EGWO algorithm as feature selection for B-ALL diagnoses and its subtypes classification using peripheral blood smear images," *Alexandria Engineering Journal*, vol. 68, pp. 39–66, Apr. 2023, doi: 10.1016/j.aej.2023.01.004.

[11] L. Boldú, A. Merino, S. Alférez, A. Molina, A. Acevedo, and J. Rodellar, "Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis," *Journal of Clinical Pathology*, vol. 72, no. 11, pp. 755–761, Nov. 2019, doi: 10.1136/jclinpath-2019-205949.

[12] A. Rehman, N. Abbas, T. Saba, S. I. ur Rahman, Z. Mehmood, and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," *Microscopy Research and Technique*, vol. 81, no. 11, pp. 1310–1317, 2018, doi: 10.1002/jemt.23139.

[13] A. Al Mamun, M. J. Hossen, A. Tahabilder, A. Musha, R. Hasnat, and S. K. Saha, "Acute lymphoblastic leukemia detection approach from peripheral blood smear using color threshold and morphological techniques," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 4, p. 3692, 2022.

[14] M. T. H. K. Tusar and R. K. Anik, "Automated Detection of Acute Lymphoblastic Leukemia Subtypes from Microscopic Blood Smear Images using Deep Neural Networks," *arXiv preprint arXiv:2208.08992*, 2022.

[15] Y. Jusman, R. O. Ningrum, and M. A. Fawwaz Nurkholid, "Leukemia Cell Image Classification Using CNN: AlexNet and GoogLeNet," in *2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, Sep. 2023, pp. 1–6. doi: 10.1109/ICEEIE59078.2023.10334867.

[16] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, "Human skin detection using RGB, HSV and YCbCr color models," *arXiv preprint arXiv:1708.02694*, 2017.

[17] F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, and H. Soltanian-Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," *Pattern recognition*, vol. 36, no. 8, pp. 1725–1736, 2003.

[18] M. K. I. Rahmani, S. Khan, J. Al-Sultan, and D. B. Ojha, "A Content-Based Medical Image Retrieval Algorithm," in *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, IEEE, 2022, pp. 73–78.

[19] M. Y. Shams, M. Z. Rashad, O. Nomir, and R. M. El-Awady, "Iris recognition based on LBP and combined LVQ classifier," *arXiv preprint arXiv:1111.1562*, 2011.

[20] M. Y. Shams and S. H. Sarhan, "A vision system for multi-view face recognition," *arXiv preprint arXiv:1706.00510*, 2017.

[21] R. Hammouche, A. Attia, S. Akhrouf, and Z. Akhtar, "Gabor filter bank with deep autoencoder based face recognition system," *Expert Systems with Applications*, vol. 197, p. 116743, 2022.

[22] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.

[23] E. Bair, "Semi-supervised clustering methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 5, pp. 349–361, 2013.

[24] S. Sarhan, A. A. Nasr, and M. Y. Shams, "Multipose Face Recognition-Based Combined Adaptive Deep Learning Vector Quantization," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–11, Sep. 2020, doi: 10.1155/2020/8821868.

[25] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," *International journal of scientific and research publications*, vol. 3, no. 5, pp. 1–5, 2013.

[26] O. Elzeki, M. Shams, N. Hikal, and M. Elaraby, "Signature recognition based on support vector machine and deep convolutional neural networks for multi-region of interest," *Journal of Theoretical and Applied Information Technology*, vol. 98, pp. 3887–3897, Dec. 2020.

[27] F. Ardjani, K. Sadouni, and M. Benyettou, "Optimization of SVM multiclass by particle swarm (PSO-SVM)," in *2010 2nd International Workshop on Database Technology and Applications*, IEEE, 2010, pp. 1–4.

[28] J. Liu *et al.*, "Quantitative imaging of Candida utilis and its organelles by soft X-ray Nano-CT," *Journal of Microscopy*, vol. 270, no. 1, pp. 64–70, 2018.

[29] D. Bor, E. Unal, and A. Uslu, "Comparison of different phantoms used in digital diagnostic imaging," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 795, pp. 160–166, 2015.

[30] A. M. Elshewey, M. Y. Shams, N. El-Rashidy, A. M. Elhady, S. M. Shohieb, and Z. Tarek, "Bayesian Optimization with Support Vector Machine Model for Parkinson Disease Classification," *Sensors*, vol. 23, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/s23042085.

[31] S. Mishra, L. Sharma, B. Majhi, and P. K. Sa, "Microscopic image classification using DCT for the detection of acute lymphoblastic leukemia (ALL)," in *Proceedings of International Conference on Computer Vision and Image Processing: CVIP 2016, Volume 1*, Springer, 2017, pp. 171–180.

[32] S. Chin Neoh *et al.*, "An intelligent decision support system for leukaemia diagnosis using microscopic blood images," *Scientific reports*, vol. 5, no. 1, p. 14938, 2015.

[33] F. M. Talaat and S. A. Gamel, "Machine learning in detection and classification of leukemia using C-NMC_Leukemia," *Multimedia Tools and Applications*, pp. 1–14, 2023.

[34] S. Chand and V. P. Vishwakarma, "Applications of deep learning in acute leukemia detection-a review," in *AIP Conference Proceedings*, AIP Publishing, 2023.

[35] A. Ramagiri *et al.*, "Image Classification for Optimized Prediction of Leukemia Cancer Cells using Machine Learning and Deep Learning Techniques," in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, IEEE, 2023, pp. 193–197.

[36] W. Rahman, M. G. G. Faruque, K. Roksana, A. S. Sadi, M. M. Rahman, and M. M. Azad, "Multiclass blood cancer classification using deep CNN with optimized features," *Array*, vol. 18, p. 100292, 2023.