CLAHE enhanced hybrid feature descriptors for classification of acute lymphoblastic leukaemia in blood smear images

Renuka Veerappa Tali*

Department of ECE,
K.S. School of Engineering and Management,
VTU,
Bengaluru, Karnataka, India
Email: renukavtali@gmail.com
*Corresponding author

Surekha Borra

Department of ECE,
K.S. Institute of Technology,
Bengaluru, Karnataka, India
Email: borrasurekha@gmail.com

Vijay Bhaskar Reddy Dinnepu

AgWiQ Technologies Pvt. Ltd.,
Bengaluru, Karnataka, India
Email: vijaybrd@gmail.com

Abstract: Acute lymphoblastic leukaemia (ALL) detection through a complete blood count test is often flagged to an expert pathologist for confirmation which is time-consuming, observer-specific and involves intensive labour. The study proposes an efficient computer aided diagnosis (CAD) method based on image processing and machine learning models to assist doctors in analysing microscopic images. This study aimed to investigate the combined discriminative qualities of shape and texture features, as well as the best fit feature subset selection technique, to achieve high accuracy and a low false positive rate for classification of healthy and ALL infected leukocyte cell images. This approach outperformed existing models with an accuracy of 92.3%, a precision of 96%, and a false positive rate of 3.846%. As a result, the proposed methodology is capable of more precisely classifying the images into healthy and ALL affected cell images, assisting physicians in the detection and diagnosis processes.

Keywords: acute lymphoblastic leukaemia; ALL; computer aided diagnosis; CAD; image processing; leukocytes; machine learning; microscopic images; feature extraction; feature selection; SVM.

Reference to this paper should be made as follows: Tali, R.V., Borra, S. and Dinnepu, V.B.R. (2023) 'CLAHE enhanced hybrid feature descriptors for classification of acute lymphoblastic leukaemia in blood smear images', *Int. J. Biomedical Engineering and Technology*, Vol. 43, No. 4, pp.309–328.

Biographical notes: Renuka Veerappa Tali received her Master's degree from BMS College of Engineering, Bangalore, and is presently pursuing her PhD at Visvesvaraya Technological University, India. Her research interests are image processing, machine learning, and data analytics.

Surekha Borra received her PhD in 2015 from Jawaharlal Nehru Technological University, Hyderabad, India. She started her academic career as an Assistant Professor in 2004 and served in various engineering colleges for 17 years. Currently, she is a Professor in the Department of Electronics and Communication Engineering, K.S. Institute of Technology, Bengaluru, India. Her research interests include image and video analytics, information security and signal processing. She has received Woman Achiever's Award from The Institution of Engineers (India) for her prominent research and innovative contributions, and several research grants from the Government of Karnataka, India.

Vijay Bhaskar Reddy Dinnepu is a Founder and CEO of many agricultural startups in India. He is a passionate and a seasoned professional with core expertise in data networking, embedded software, artificial intelligence, and Ethernet switching. He is an alumnus of IIT Madras. He has been an achiever from his schooldays with scholarships given by state government for his outstanding academic brilliance. The same brilliance was carried forward from IIT campus to corporate corridors of IT majors in Wipro, Intel, and Cisco. His 23 years of corporate stint was galore with awards for his innovation and technology solutions before his entrepreneurial journey with Agwiq Technologies. He is the winner of AI for Social Good by NASSCOM, winner of Digitizing India Award by Cisco & CNBC TV18, winner of Sankalp Award for Best Agriculture Startup, India's most prestigious social enterprise award.

1 Introduction

Advancements in artificial intelligence (AI), data science, and computer vision have drastically improved the diagnosis accuracy and conventional perception of laboratory haematology process patterns and norms. Leukocyte image analysis has been the subject of much research over the past few decades. Leukocytes are responsible for building a strong immune system in the human body and fight against foreign invaders like viruses, bacteria, and fungi. Leukaemia is caused by the abnormal and uncontrolled growth of lymphocytic cells (Pan et al., 2012), with chronic lymphoblastic leukaemia (CLL) and acute lymphoblastic leukaemia (ALL) being the most common types. Lymphocytes are usually found in lymph nodes, the spleen, the thymus, the bone marrow and spread across the respiratory and digestive systems. They develop from lymphoblast cells into mature infection-fighting cells called lymphocytes: B cells and T cells. These early immature cells develop into ALL at various stages, and are most seen in 3–7 age group children, causing low survival rate if left untreated at the earliest. 6,000 new cases are diagnosed every year with symptoms like weakness, bone pain, fever, frequent infections, excessive sweating, sudden unexpected weight loss, etc. According to the American Cancer Society (2019), bone marrow aspiration and biopsy, blood tests, lumber puncture, lymph node biopsy, chromosome tests, flow cytometry, immunohistochemistry, cytogenetics, and imaging tests (X-ray, CT scan, MRI, ultrasound, etc.) should be used to diagnose leukaemia. Biopsy (Loeffen et al., 2020) is conducted to determine the definitive type of

cancer and its stage. A tiny hollow needle is inserted through the skin, and a sample of tissue or tumour is extracted and sent to a laboratory for analysis. This test, when conducted on children, leads to discomfort and certain complications like bleeding, injury, or infection. As a result, it should be avoided in children to the extent possible.

A complete blood count test, being the most economical and simple of all other tests, is often suggested by doctors to detect leukaemia. Pathologists use careful microscopic observation of blood smears to determine the number and type of leukocytes based on their usual colour, shape, size, and texture. The accuracy of this manual technique is low and partially dependent on the experience level of pathologists (Acharjee et al., 2016; Rajinikanth et al., 2020). Furthermore, pathologists who are tired may miss vital evidence while examining and take a long time to reveal the reports (Bhattacherjee et al., 2020). Leukocytes can also be divided into healthy and sick cells, or into the sorts of cells that make up each type, using a variety of automated methods employing laser, radio frequency, direct current, the impedance method, optimal temperatures, volumes, as well as various blood smear slide staining procedures. These tools offer a high degree of precision and accuracy in calculating and recognising normal blood cells. However, they lack the sensitivity needed to recognise premature, irregular, or blast cells. To solve this issue, analysers frequently flag samples based on their cell population, that demands trained personnel to perform peripheral smear tests to spot abnormal cells. This resulted in the use of computer vision and image processing techniques to detect and classify leukocytes in peripheral microscopic blood smear images. Leukocyte image analysis has received a great deal of attention in recent decades. Several computerised models for classifying leukocytes into healthy and diseased cells or constituent types have been created. Computer algorithms based on image processing (IP) and machine learning (ML) models promise to reduce laboratory workload and personnel costs while simultaneously improving patient care. Even if they have been shown to be more effective in the current situation, the automatic haematology analysers that are now in use have several issues that need to be properly addressed considering image quality, image acquisition standards, illumination settings, cell density, presence of overlapping cells, the effect of staining artefacts, similarity between class samples, imbalance, and inadequate datasets.

The CAD tools based on neural networks (NN) and deep learning (DL) models (Khadse et al., 2020; Renuka and Surekha, 2021) could be an alternative, and many researchers have been working for decades to build models capable of classifying blood cells as malignant (unhealthy or abnormal) or benign (healthy or normal) (Gedik, 2022) using these algorithms. Rodrigues et al. (2016) proposed a simple technique to automate the classification of lymphocytes into normal and abnormal in blood smear images, using morphological operators (Vincent and Chandra, 2022) to preprocess ALLIDB images. Decision trees, naïve Bayes, SVM (Kshirsagar et al., 2021; Kalaiselvi et al., 2022) and K-nearest neighbour (KNN) classifiers, along with cross validation, were implemented for comparison and reported an accuracy of 85%. An ant colony optimisation-based hybrid feature selection with cosine similarity and SVM classifier was proposed (Sweetlin et al., 2017). The run-based recruitment strategy was applied to select the best features. This study extracted 22 grey level co-occurrence matrix (GLCM) features and 18 geometric features with an accuracy of 81.66%. An elephant herd optimisation algorithm that minimises the misclassification rate as an objective function is proposed (Sahlol et al., 2017). A feed forward neural network (NN) was used to classify the ALLIDB images, which resulted in 91.8% accuracy. Mishra et al. (2017) proposed a

technique based on the discrete cosine transform (DCT) for feature extraction. The Wiener filter was used to preprocess the ALLIDB images, and SVM was used as a classifier (Li et al., 2018). Ninety features were extracted using DCT, and 89.76% accuracy was reported. Shree and Kumar (2018) also proposed a similar approach for classification using the GLCM model's extracted features and an NN classifier.

Moshavash et al. (2018) presented a robust decision support system for the classification of healthy and malignant leukocytes with ensemble classifiers. The ensemble classifier is comprised of supervised models like SVM, KNN, naïve Bayes, and decision tree. The class labels from these models were approximated using weights, and majority vote labels were taken as the final decision. Similar approaches were carried out with variants of SVM kernel functions like multi-layer perceptron (MLP), Gaussian radial basis function, polynomial, quadratic, and linear. To classify ALL cells, a feature extraction method based on a grey-level run length matrix was presented (Mishra et al., 2018) in which 11 features were extracted and fed to the SVM classifier. De Faria et al. (2018) proposed an approach based on local descriptors like SURF and SIFT to extract features, and a bag of visual words (BOVW) was used to represent them. SVM and MLP supervised classifiers were used on three different datasets. Histopathology image classification using local descriptors was proposed by Öztürk and Akdemir (2018), where many supervised classifiers like boosted trees, KNN, and SVM are compared for performance. Among many feature extraction methods like GLCM, local binary pattern (LBP), and grey level run length matrix, the segmentation-based fractal texture analysis method picked up more prominent features than any other descriptor. Ahmed et al. (2019) presented a convolutional neural network (CNN)-based classification (Ahuja et al., 2019), which is computationally complex but incorporates feature extraction within itself. The approach was used to classify leukaemia subtypes from microscopic images. Five-fold cross validation with SGD and ADAM optimisers was used to tune the classifier model for 100 epochs. This model resulted in 88.25% accuracy. A multi-features-based segmentation and classification approach of leukocyte types was proposed (Benomar et al., 2021) where 155 leukocyte cells were segmented from 87 colour images using the watershed algorithm. Various morphological, colour, and texture features were extracted from both segmented nucleus and cytoplasm images. Overall accuracy of 95.86% was reported in classifying five types of leukocytes using a random forest classifier.

In the existing methods, raw images are processed directly. Some literature reports few methods of preprocessing but does not validate them with any metric to define quality improvement. No information is reported on false positive and false negative rates, which are crucial metrics to detect and diagnose any disease. Also, different datasets are used, some of which are benchmark datasets while others are handcrafted datasets that are hard to obtain. Some techniques were experimented on with their own private datasets and cannot be validated due to their non-availability to the research community. Further, images are subjected to a variety of clinical conditions, including image acquisition via various scanners, staining modalities, non-uniform illumination conditions, and microscopic magnification. Hence, irrespective of any application, images need custom adjustment of range and brightness values to correct image artefacts so that visual understanding and interpretation are possible. The difference between healthy and unhealthy cells can be recognised efficiently only when classifier models are fed with the most discriminative features. Hence, the challenge of developing a robust and highly differentiable feature extractor model remains unsolved. The removal of

redundant features is also a major task that can improve the training speed and complexity of a classifier model. Classifier model parameter optimisation must be carried out in a more systematic and controlled way. Further, the deployed model should be evaluated with relevant metrics to validate the system. The current classification models depend on the type of image dataset used, segmentation models, kinds of features extracted, and types of classifier models (Kshirsagar et al., 2018; Benomar et al., 2021). Hence, huge scope awaits many researchers to experiment and explore various techniques with a variety of combinations that can best suit multiple datasets.

This study proposes a new approach to classify leukocytes in microscopic blood smear images into healthy and ALL infected unhealthy, using the freely available benchmark dataset ALL Image Database (ALLIDB) to achieve a low false positive rate.

The highlights are as follows:

- images are preprocessed using the contrast limited adaptive histogram equalisation (CLAHE) enhancement model and evaluated with a no reference image quality score NIQE
- various texture feature extraction techniques, such as LBP, GLCM, HOG, and SURF are analysed
- SFS and PCA feature selection models are applied and compared to select the best feature vectors
- the optimised SVM classifier model is trained using selected features and tested
- the classifier model is evaluated with standard metrics.

2 Preliminaries

2.1 Contrast limited adaptive histogram equalisation

Histogram equalisation transforms all image pixels using a transformation derived from the histogram of the image. While this works well with uniformly distributed pixel intensities, it does not work well with variable pixel intensities. Adaptive histogram equalisation (AHE) addresses this issue by transforming every pixel with a function adapted by its neighbourhood region. The drawback of AHE is over amplification of the noise in near-constant regions. CLAHE is an improved variant of AHE that limits noise amplification (Tarandeep et al., 2017). CLAHE is basically characterised by two parameters: the number of tiles and the clip limit to control the contrast of the image (Fan et al., 2020). Tile is a two-element vector [M N] with N columns and M rows. The original image is divided into regions of contextual rectangles specified by tile. A contrast function is computed on each tile, thereby enhancing its contrast such that the histogram of each tile matches approximately the desired histogram shape, called the distribution value. The falsely induced boundaries are removed by combining neighbourhood tiles using bilinear interpolation. To avoid saturation in areas of uniform intensity, a contrast factor known as ‘clip limit’ is used. The clip limit ranges from 0 to 1, where a higher value leads to high contrast.

2.2 Local binary pattern

Humans can identify objects based on colour, shape, and size (Maiti et al., 2021), whereas machines follow specific patterns and textures to discriminate among objects (Dabeer et al., 2019). The process of identifying and describing a unique, distinguishable set of characteristics in the form of a numerical vector is called feature extraction. LBP is a popular feature descriptor known for computational ease and promisingly good performance. LBP (Khalil et al., 2018; Alhindi et al., 2018) computes local texture representation of an image. The neighbouring pixel values are compared with a central point under consideration to check if value is greater or less than the central point. A 2D array equal to the input image size with entries as LBP values is constructed, and a histogram is computed over it. If 3×3 neighbourhoods are taken, then 256 patterns from 0 to 255 will construct 256 bins of LBP codes. These feature bins for every image can be stored as an array in an excel sheet or CSV file for further classification.

2.3 Grey level co-occurrence matrix

The most popular texture-based feature extraction method is GLCM (Alhindi et al., 2018; Dwaich and Abdulbaqi, 2021), which is a matrix that represents the spatial relationship between pixels. It explores how often pixel pairs occur in an image with values of intensity and position. Several statistics are derived from this matrix, such as contrast, correlation, energy, and homogeneity. Equations (1) to (5) specify GLCM features.

$$Energy = \sum_{x,y=0}^{M-1} \left(P_{xy}\right)^2 \tag{1}$$

$$Contrast = \sum_{x,y=0}^{M-1} P_{xy}(i-j)^2 \tag{2}$$

$$Correlation = \sum_{ix,y=0}^{M-1} P_{ij} \frac{(x-\mu)(y-\mu)}{\sigma^2} \tag{3}$$

$$\sigma^2 = \sum_{x,y=0}^{M-1} P_{xy}(i-\mu)^2 \tag{4}$$

$$Homogeneity = \sum_{x,y=0}^{M-1} \frac{P_{xy}}{1+(x-y)^2} \tag{4}$$

where

P_{xy} GLCM element

M total levels

μ GLCM mean

σ^2 variance of intensities of all reference pixels that contributed to GLCM.

Contrast is a metric of variance and inertia among image pixels, defines how a pixel is correlated to its neighbouring pixel over the entire image. The uniformity property is measured by energy, and homogeneity measures how closely GLCM elements are distributed to diagonal elements. This method provides four feature descriptors for each image.

2.4 Bag of visual words

The bag of words method is an approach to extracting powerful features from text data that are used to classify documents and perform language modelling. Unique patterns or image patches are identified, and their features are extracted into a bag. The steps involved in this approach are feature extraction and code book construction. The vocabulary generated in the code book is used to classify images. Features are comprised of key points and their descriptors (Mittal and Saraswat, 2019). Key points are unique image points that do not change even if images are shrunk, rotated, or even expanded. These key points can be represented as descriptors. Identifying stand-out key points and extracting descriptors can be done using algorithms like scale-invariant feature transform (SIFT), speeded up robust features (SURF), etc. Number of clusters of descriptors is formed and the centre of each cluster is used as vocabularies. The histograms created from vocabularies and frequency of vocabularies, form the bag of visual words in the proposed method. This method was proposed for text data classification, it can be applied on medical images too. It focuses on frequency of occurrence of pixels in an image which is suitable to the proposed image dataset.

2.5 Histogram of oriented gradients

HOG is a powerful shape descriptor that can be used in object detection applications. The HOG (Raghavendra et al., 2019) descriptor is based on the calculation of intensity gradients along x and y directions, which uniquely represent local shape and appearance within an image patch. The magnitude of the gradient is large at corners and edges due to the sharp change in intensity. The entire image is subdivided into small regions of size [M, N] for which oriented gradients are calculated. HOG generates a histogram for each of these regions separately. Later, the histogram of each region is normalised to reduce lightning variation. Finally, all the image's features are obtained by combining normalised histogram bins of smaller regions.

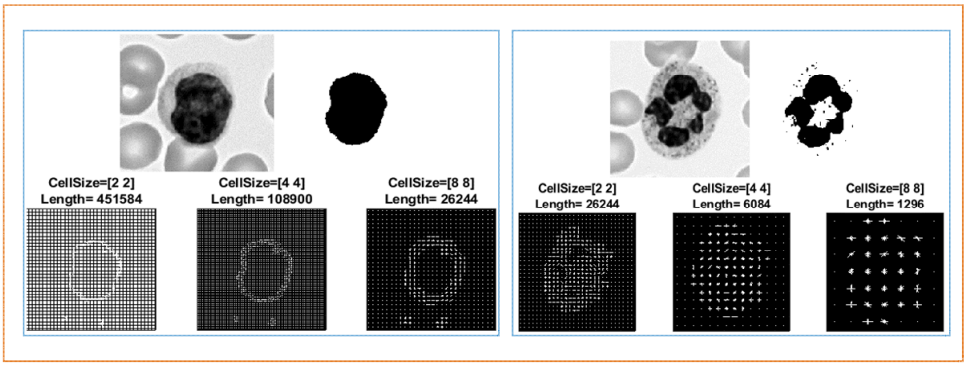
$$Mag\ Gradient = \sqrt{g_x{}^2 + g_y{}^2} \tag{6}$$

$$Direction\ of\ \ gradient\ \theta = \arctan\frac{g_y}{g_x} \tag{7}$$

where g_x and g_y are resultant gradients in x and y direction for the pixel.

Figure 1 depicts typical visualisations of HOG feature descriptors of various cell sizes. The dataset in RGB colour is converted to greyscale and later binarised to extract descriptors. As cell size decreases, the length of descriptors increases with increased memory space. The magnitude and direction of gradients can be computed using equations (6) and (7).

Figure 1 Visualisation of HOG feature descriptors for different cell size (see online version for colours)



2.6 Sequential forward selection

Feature selection reduces a high dimensional feature vector to a low-dimensional space that leads to better generalisation capabilities, reduced complexity, and reduced run-time (Kumar et al., 2021; Khaire and Dhanalakshmi, 2022; Pahuja et al., 2022). Feature selection algorithms require a search method and an objective function to evaluate selected subsets or candidates. The objective function estimates candidate subsets, returning a metric of their goodness as feedback to the search method to select new subsets. SFS follows heuristic search for features (Acharya et al., 2019) and evaluates using wrapper methods. It can add or delete features sequentially but tends to get trapped at local minima (Ashok and Aruna, 2016). It provides excellent computational efficiency by filtering out redundant and irrelevant features. It eventually starts with an empty set and sequentially adds the feature a+ that, when combined with the earlier features Xk, yields the highest objective function O(Xk+a+) (Dey et al., 2020). The SFS algorithm attempts to select the best features for a given ML algorithm through resampling or cross validation. The SFS method is more accurate since it gets tuned to meet classification constraints (Jain and Zongker, 1997). It has good generalisation ability and a mechanism to prevent overfitting. The only disadvantage of SFS is that once added, the feature cannot be discarded. Since it trains a classifier model for every subset of features, execution is very slow.

2.7 Principal component analysis

Classification is the process of categorising detected objects of interest into a set of predefined classes. Supervised classification learns from labelled features of training data and classifies unlabeled test data into predefined classes. Unsupervised classification corresponds to grouping unlabelled data into classes based on analysis made by software. A few supervised techniques include SVM, decision trees, naïve Bayes, and KNN. PCA is a statistical technique for obtaining low-dimensional data from high-dimensional data by selecting relevant features that encapsulate the most information from the dataset. The optimal search strategy and filter method of evaluation make PCA independent of predictive classification accuracy. The intrinsic properties of features, such as the distance between interclass points, are used to select them. The principle involved in

PCA (Song et al., 2010) is to find the largest eigenvalue of the covariance of input data points and choose the first principal component based on the largest variance. The next highest being next principal component, and so on. The principal components have no correlation with one another. It can compute non-iteratively on a dataset, which is much faster than a training session for a classifier.

3 Methodology

The proposed leukocyte classification methodology shown in Figure 2 incorporates preprocessing of the images, dataset preparation for training and testing, feature extraction and selection, classifier model training, testing, and evaluation with performance metrics. The steps involved are summarised as follows:

1 preprocess the images using the CLAHE enhancement algorithm

2 extract feature descriptors using LBP, GLCM, BOVW and HOG

3 select minimum features using SFS and PCA.

4 build a classifier with a training dataset using SVM algorithm with optimised hyper parameters.

5 evaluate the model's performance

6 compute performance metrics: accuracy, precision, F1-score, FNR, FPR, etc.

The images in the dataset are usually affected by many artefacts due to microscopic magnification, non-uniform illumination, staining modalities, and many other clinical conditions during image acquisition. To address these issues, image preprocessing is performed firstly, so that most of the meaningful information in the image can be obtained and used to interpret further processing.

The CLAHE enhancement of image is performed as follows:

1 read the original image input

2 convert an RGB colour space image to Lab colour space

3 scale values to a range of 0 to 1

4 perform CLAHE using the MATLAB function adapthisteq on L channel of the image by selecting a tile size of [M N] and clip limit within 0 to 1

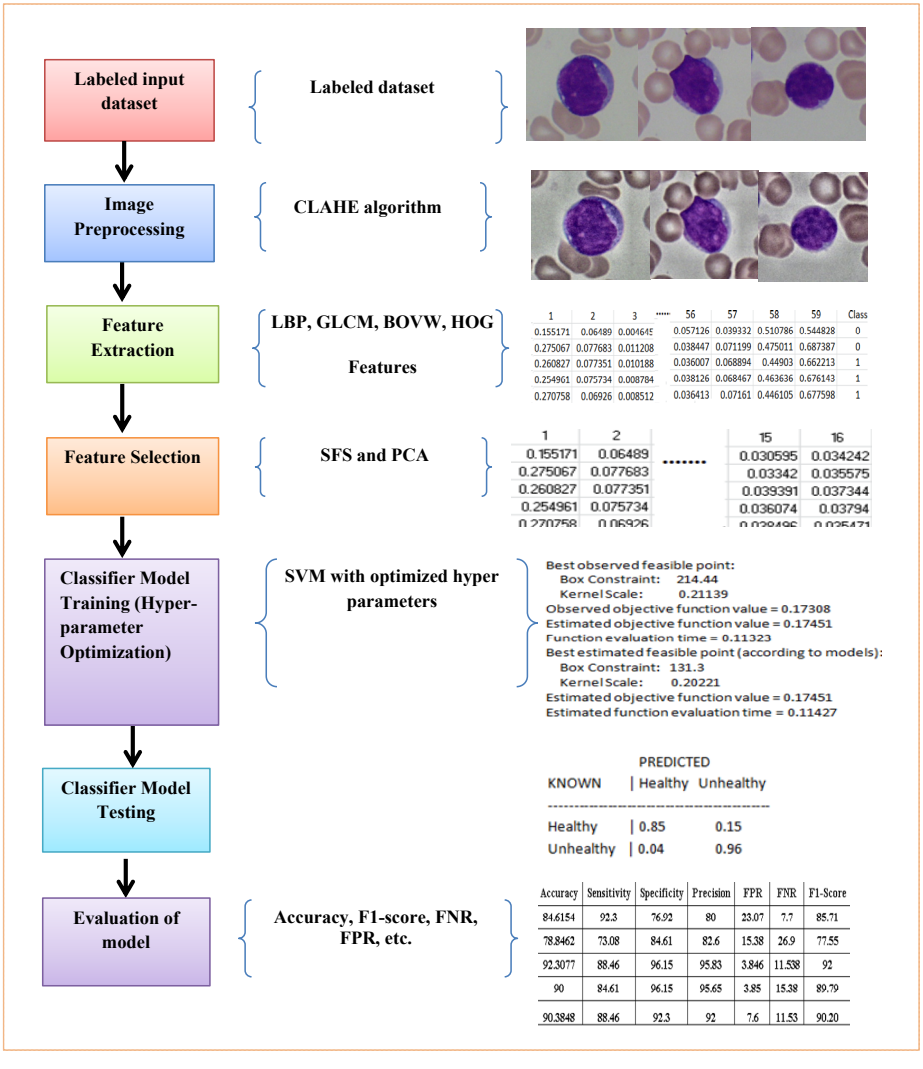
5 scale back L-channel values to normal.

6 convert the enhanced image back to RGB colour space.

7 compute the no reference quality measure using the natural image quality evaluator (NIQE) metric on original RGB input image and enhanced RGB image.

8 compare the NIQE values to check for enhancement.

Figure 2 Flow diagram of proposed methodology (see online version for colours)



Colour, shape, and texture features are powerful descriptors that can represent objects of interest in an image. The texture being the most eminent feature with high discriminative ability, this work explored the GLCM and LBP models to extract features from CLAHE enhanced image. Features are relevant to a specific problem and can be extracted either manually or automatically based on domain knowledge. Specialised algorithms or networks like deep neural networks capable of extracting features automatically without human intervention have a drawback of complex computations and expensive resources. In the proposed method, LBP, GLCM, BOVW, HOG feature extraction algorithms are implemented to obtain a set of features. The prominent features extracted from the previous step contain discriminating signatures along with some redundant records that need to be identified and removed. There are two types of features: features that are highly correlated with each other and the features that are highly correlated with the

target class. Features that have a high correlation with one another may have the same effect on the target variable, and hence can be discarded as a redundant feature. Features with a high correlation with the target class contribute more to the classifier model's quality and must be retained. Hence, in the proposed method, for the removal of redundant features and to reduce the training and testing complexity, SFS and PCA methods are employed.

Later, SVM is used for classification as it is computationally efficient on small and large balanced datasets, requires less memory space, and offers greater classification performance (Tsang et al., 2005; Menon, 2009; Li and Yu, 2014; Okwuashi and Ndehedehe, 2020; Althnian et al., 2021). Because, the SVM hyperplane completely relies on support vectors, the size of a dataset has no effect on its performance if all its data is in support vectors.

The trained SVM classifier is then evaluated using various performance metrics such as false negative rate (FNR), false positive rate (FPR), precision, specificity, sensitivity, accuracy, and f1-score (Tali et al., 2021). These metrics are calculated by extracting data from the confusion matrix. A confusion matrix (Mimura, 2023) is a [M, N] matrix that summarises model performance, with M being the number of true classes and N being the predicted classes. The matrix compares predicted classes by the model to actual ground truth classes. For a binary classification problem, the confusion matrix is a 2×2 matrix, as shown in Table 1. True positive (TP) and true negative (TN) values define the predicted value that exactly matches the target class's true value. The errors: false positive (FP) and false negative (FN) define predicted classes that do not match the true class. False positives are a major source of concern in the medical field because ALL infected cells are misidentified as healthy cells, leading to expensive and harmful medication for patients who do not have the disease. At the same time, patients are missing a chance to be treated for their true illness (Allen, 2020). Misdiagnosis is a threat to humans that eventually leads to loss of life (Gárate-Escamila et al., 2020). Equations (8) to (15) depict the performance metrics of classification algorithm derived from a confusion matrix.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Sensitivity\ (Recall) = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$False\ positive\ rate\ (FPR) = \frac{FP}{FP + TN} \tag{12}$$

$$False\ negative\ rate\ (FNR) = \frac{FN}{FN + TP} \tag{13}$$

$$F1\text{-}Score = \frac{2Recall * Precision}{Recall + Precision}. \tag{14}$$

Table 1 Confusion matrix

True class	*Predicted Class*	
	True positive (TP)	False negative (FN)
	False positive (FP)	True negative (TN)

In this study, positive class is healthy cell and negative class is ALL infected cell. Accuracy in equation (8) is a metric used to evaluate binary classifiers. It measures how many predictions made by classifier model are correct. Sensitivity (recall) in equation (9) measures the capability to correctly predict the positive class. Specificity in equation (10) measures capability to correctly predict the negative class. Precision in equation (11) measures the positive prediction rate. F1-score is a measure of the accuracy of a classifier model in terms of recall and precision.

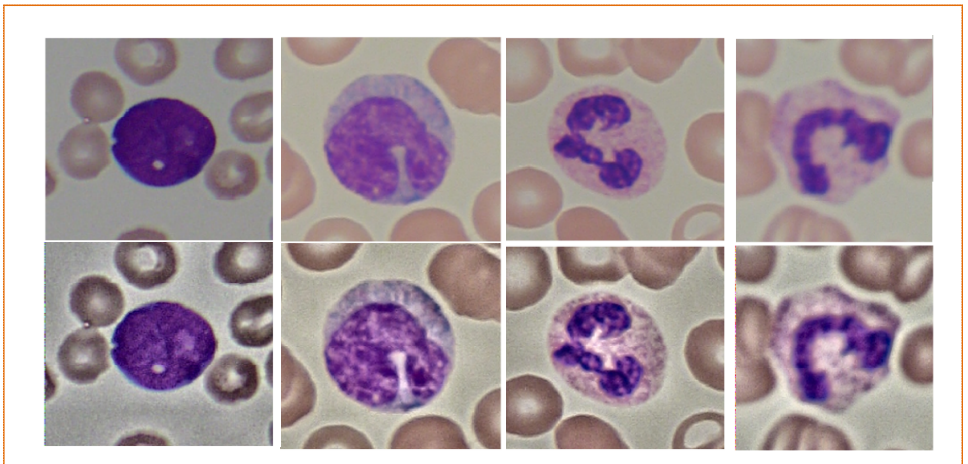
4 Results

The experimentation is carried out on a PC with a Core i3-2350M CPU at 2.3GHz, 8GB RAM and the Windows 10 operating system. The simulation of the proposed methodology is performed using MATLAB R2019a. The proposed methodology is implemented using a standard and freely available benchmark dataset called ALLIDB (Labati et al., 2011), which has two distinct datasets: ALLIDB1 and ALLIDB2. ALLIDB1 is used for cell segmentation, and ALLIDB2 is used for cell classification. ALLIDB1 contains 108 images for segmentation, and ALLIDB2 contains 260 cropped normal and blast cell images from ALLIDB1 for classification. The 260 images in the dataset are divided into two classes: 130 images are of healthy lymphocyte cells, and 130 images belong to unhealthy lymphocyte cells that are infected by ALL. The healthy cell images are considered positive class, and unhealthy cell images are negative class. The dataset is made up of 24 JPEG images with a colour depth of 24 captured by a Canon PowerShot G5 camera. A resolution of 257×257 was used to obtain 130 candidate lymphoblasts and healthy lymphocytes each, observed under a laboratory optical microscope with magnifications from 300 to 500.

ALLIDB images are dull, with broken edges and no discernible difference between nucleus and cytoplasm. The images have washed-out texture and low contrast, out of which distinguishable features cannot be extracted. Hence, the images were first enhanced using the CLAHE model with a tile size of [8×8] and clip limit of 0.01. Figure 3 shows CLAHE enhanced images (second row) in comparison to original RGB images (first row). The quality of images before and after enhancement is measured using the NIQE image quality metric, which is a reference less or blind image quality score.

Table 2 gives a comparison of NIQE metrics for some dataset image files before and after enhancement operations. Dataset images are named as Imxxx_y_output, where x represents the image number and y represents an unhealthy (1) or healthy (0) leukocyte cell. The NIQE values in the table indicate that the quality of the image is improved after the preprocessing step.

Figure 3 CLAHE enhanced Images (first row – original images, second row – enhanced images) (see online version for colours)

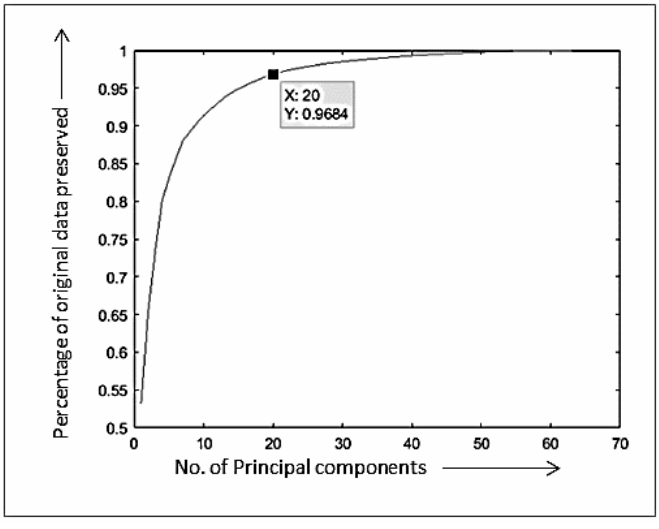


In the feature extraction process, 59 features are extracted using the LBP technique, and in parallel, GLCM extracted contrast, energy, correlation, and homogeneity features for each of the 260 enhanced images. The extracted features are split 80:20 between the training and test sets. K-fold cross validation of 10 is applied on to the validation set prepared from the training set. Later, the feature selection methods were applied to select the optimum features for classification using two methods: SFS and PCA. The plots of the number of principal components on the x-axis and percentage information retained along the y-axis indicate the optimal number of features to be selected as given in Figure 4.

Table 2 Comparison of original and CLAHE enhanced images using NIQE metric

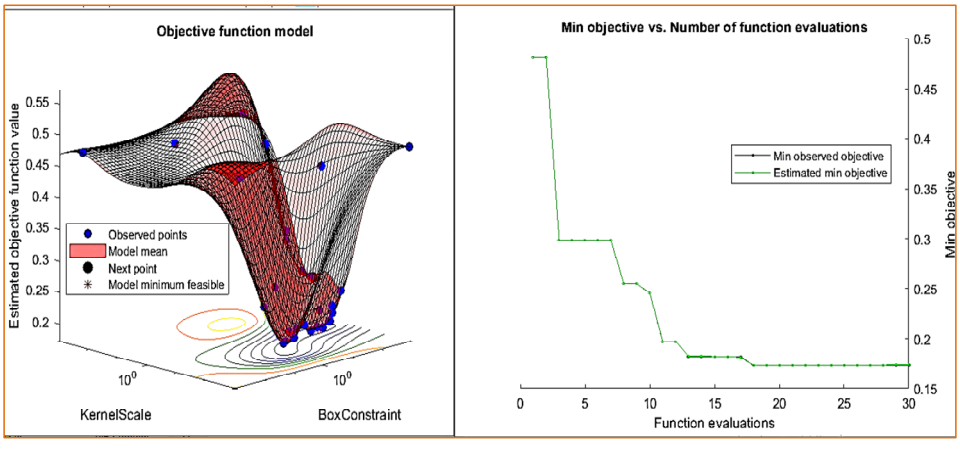
Images	*Images files*	*Original RGB image quality*	*CLAHE enhanced image quality*
Unhealthy	Im070_1_output	4.3077	5.1025
	Im044_1_output	4.5948	5.2775
	Im003_1_output	5.2841	5.8231
	Im065_1_output	5.1619	5.6302
	Im123_1_output	4.1421	4.6885
Healthy	Im131_0_output	4.3232	4.9369
	Im135_0_output	3.7128	4.2270
	Im142_0_output	4.4942	4.6160
	Im148_0_output	3.4664	3.7394
	Im175_0_output	3.6686	3.8779

Figure 4 Principal components verses percentage of original data preserved



While the SFS method validated optimal features using a Bayesian optimisation algorithm, the best hyperparameters were obtained after 30 iterations. The optimise hyper-parameter function is used to generate optimal expected and observed values of two parameters: box constraint (C) and kernel scale (γ). The model created with the best hyperparameters is validated using test data. Iteration plots of optimising hyperparameters are shown in Figure 5.

Figure 5 SVM hyper-parameter optimisation iteration (see online version for colours)



The iteration history and optimum values for k-fold cross validation of 10 features and the final feature columns selected are shown below:

K-fold cross validation = 10

Final columns included: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 62

Best observed feasible point:

Box constraint: 214.44

Kernel scale: 0.21139

Function evaluation time = 0.11323

Estimated objective function value = 0.17451

Observed objective function value = 0.17308

Best estimated feasible point (according to models):

Box constraint: 131.3

Kernel scale: 0.20221

Estimated function evaluation time = 0.11427

Estimated objective function value = 0.17451

The bag of visual words algorithm is applied to the database to select feasible feature point locations using the grid method. 652,288 SURF features from 208 training set images are extracted with a block width of [32 64 96 128] and a grid step of [8 8]. 500 visual word vocabularies are created using k-means clustering algorithm from 80% of the strongest features. Clustering converged after 16 iterations at 5.64 seconds per iteration. With 521,830 features, 96% accuracy for training data and 90% accuracy for the test dataset were obtained.

Table 3 shows a comparative analysis of performance (without and with CLAHE) for variously implemented feature extraction and feature selection models for the given 260 image dataset. Table 4 depicts the confusion matrix parameters for various feature extraction models implemented in this study. Results indicate that the combined LBP and GLCM model have a low false positive rate, and a bag of visual words indicates that the incorrect prediction rate is very low. Table 5 shows the performance comparison for each feature extraction model used.

Table 3 Comparison of performance for various Feature extraction and feature selection models

Feature extraction models	*Feature size*	*Feature selection model*	*Subset feature size*	*Accuracy (without CLAHE)*	*Accuracy (with CLAHE)*
LBP	59	PCA	16	57.69	82.6923
		SFS	16	78.8	84.6154
GLCM	4	PCA	4	57.7	67.3077
		SFS	4	67.3	78.8462
LBP+GLCM	*63*	PCA	20	60	70.1234
		SFS	*20*	*78.9*	*92.3077*
Bag of visual words	652,288	K-means clustering (in-built)	521,830	80.77	90
HOG	Cell size [4 4] – 108,900	NIL	108,900	76.9231	90.3846

Table 4 Confusion matrix parameters for various feature class

Feature extraction models	*True positive (TP)*	*False negative (FN)*	*False positive (FP)*	*True negative (TN)*
LBP	24	2	6	20
GLCM	19	7	4	22
LBP+GLCM	*27*	*3*	*1*	*21*
Bag of visual words	22	4	1	25
HOG	23	3	2	24

Table 5 Comparison of performance metrics for various feature extraction techniques

Features	*Accuracy*	*Sensitivity*	*Specificity*	*Precision*	*FPR*	*FNR*	*F1-score*
LBP	84.6154	92.3	76.92	80	23.07	7.7	85.71
GLCM	78.8462	73.08	84.61	82.6	15.38	26.9	77.55
LBP+GLCM	*92.3077*	*88.46*	*96.15*	*95.83*	*3.846*	*11.538*	*92*
Bag of visual words	90	84.61	96.15	95.65	3.85	15.38	89.79
HOG	90.3848	88.46	92.3	92	7.6	11.53	90.20

Table 6 shows the comparison of the proposed method with the existing techniques with respect to accuracy. Results show that LBP features are more efficient in discriminating positive classes, improving classifier sensitivity, which measures the probability of correctly classifying a positive class. Combining LBP and GLCM with SURF features better predicted negative classes. The highest precision of 95.833 and the highest F1-score of 92% are seen for combined LBP and GLCM features indicating that the combined effect of texture features (Humeau-Heurtier, 2019) is best for the ALLIDB dataset. The FPR and FNR are lowest for combined LBP and GLCM features, indicating better performance for the proposed methodology. Overall, the method demonstrated great promise in distinguishing healthy and ALL infected cells in microscopic blood smear images.

Table 6 Binary classification comparison for proposed model with state of art methods

Dataset used	*Method*	*Classifier model (feature extraction method)*	*Accuracy*
ALLIDB	Rodrigues et al. (2016)	Image processing (morphological operators)	85%
	Ahmed et al. (2019)	CNN	88.25%
	Mishra et al. (2018)	SVM (grey level run length matrix features)	89.76%
	Moshavash et al. (2018)	SVM (LBP)	89.81%
	Sahlol et al. (2017)	NN (GLCM)	91.8%
	Proposed model	*SVM (LBP+GLCM)*	*92.3077%*

The experimentation shows that the results are completely dependent on the quality of the images, the features extracted, the classifier model used, and the software environment. The morphology of cells in the images shows that the difference between healthy and unhealthy ALL cells is not so prominent. Healthy lymphocytic cells are spherical in shape and contain a single large nucleus. The unhealthy lymphocytic cells are elongated, shape and their nucleus cover 90% of the cytoplasm. Small holes called vacuoles are seen within the nucleus of unhealthy cells. Hence, it is challenging for machine learning models to deal with such diverse and complex medical images. As compared to morphological operators used (Rodrigues et al., 2016; Li et al., 2018) for preprocessing the images, CLAHE results in better quality, as evident from Figure 3 and Table 2. The results obtained from feature extraction techniques like LBP and GLCM from CLAHE-enhanced images are good compared to features extracted from SURF and HOG. Since the structural similarity of cells in images is high, shape-based SURF and HOG feature extraction methods are not suitable, as evident from the results. The combined effect of texture features improved classification results significantly while using less time and memory. The SFS feature selection approach is highly robust and reliable as it selects features that meet classification constraints compared to hybrid feature selection. Because it is independent of data size, the SVM classifier converges faster and with less computation time and complexity than the feed-forward neural network and CNN. The proposed methodology is highly reliable and robust for such a balanced average dataset of medical images without the need for complex hardware. The proposed methodology, when employed as a CAD tool, is very economical and can be deployed in remote areas or villages for early prediction of disease and timely monitoring of treatment.

5 Conclusions

The classification of healthy and ALL infected leukocyte cell images is experimented on in this study using a combination of the CLAHE enhancement technique, LBP, GLCM feature extraction models, SFS feature selection model, and optimised SVM classifier model. The results show that the CLAHE model is quite effective at enhancing RGB dataset images with poor contrast and brightness. For the ALLIDB picture dataset, the combined LBP and GLCM features offer appropriate form and texture representative values. The combination of statistical and structural approaches, as well as their insensitivity to variations in illumination, resulted in the computation of good representative features. SFS feature selection outperformed PCA because it chooses a feature subset depending on the classifier prediction wrapping approach. The optimised linear SVM classifier works best with a binary class and a dataset of medium dimensionality. However, when images are rotated and local pixel variations are small, the texture descriptors LBP and GLCM perform poorly. Future study will involve experimenting with additional powerful descriptors on more complicated leukocyte cell picture collections.

References

Acharjee, S., Chakrabartty, S., Alam, M.I., Dey, N., Santhi, V. and Ashour, A.S. (2016) 'A semiautomated approach using GUI for the detection of red blood cells', *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, March, IEEE, pp.525–529.

Acharya, U.R. et al. (2019) 'Automated detection of Alzheimer's disease using brain MRI images – a study with various feature extraction techniques', *Journal of Medical Systems*, Vol. 43, No. 9, pp.1–14.

Ahmed, N. et al. (2019) 'Identification of leukemia subtypes from microscopic images using convolutional neural network', *Diagnostics*, Vol. 9, No. 3, pp.104–114.

Ahuja, R., Jain, D., Sachdeva, D., Garg, A. and Rajput, C. (2019) 'Convolutional neural network based American sign language static hand gesture recognition', *International Journal of Ambient Computing and Intelligence (IJACI)*, Vol. 10, No. 3, pp.60–73.

Alhindi, T.J. et al. (2018) 'Comparing LBP, HOG and deep features for classification of histopathology images', *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp.1–7.

Allen, J.A. (2020) 'The misdiagnosis of CIDP: a review', *Neurology and Therapy*, Vol. 9, No. 1, pp.43–54.

Althnian, A. et al. (2021) 'Impact of dataset size on classification performance: an empirical evaluation in the medical domain', *Applied Sciences*, Vol. 11, No. 2, p.796.

American Cancer Society (2019) *Cancer A-Z, Leukemia in Children* [online] https://www.cancer.org/content/dam/CRC/PDF/Public/8695.00.pdf.

Ashok, B. and Aruna, P. (2016) 'Comparison of feature selection methods for diagnosis of cervical cancer using SVM classifier', *Int. J. Eng. Res. Appl*., Vol. 6, No. 1, pp.94–99.

Benomar, M.L., Chikh, A., Descombes, X. and Benazzouz, M. (2021) 'Multi features-based approach for white blood cells segmentation and classification in peripheral blood and bone marrow images', *Int. J. Biomedical Engineering and Technology*, Vol. 35, No. 3, pp223–241.

Bhattacherjee, A., Roy, S., Paul, S. and Dey, N. (2020) 'Classification approach for breast cancer detection using back propagation neural network: a study', *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications*, pp.1410–1421, IGI Global.

Dabeer, S., Khan, M.M., and Islam, S. (2019) 'Cancer diagnosis in histopathological image: CNN based approach', *Informatics in Medicine Unlocked*, Vol. 16, pp.1–9, DOI: 100231.

de Faria, L.C., Rodrigues, L.F. and Mari, J.F. (2018) 'Cell classification using handcrafted features and bag of visual words', *Anais do XIV Workshop de Visão Computacional*, pp.68–73.

Dey, N. et al. (2020) 'Firefly algorithm and its variants in digital image processing: a comprehensive review', *Applications of Firefly Algorithm and Its Variants*, pp.1–28, Springer, Singapore.

Dwaich, H.A. and Abdulbaqi, H.A. (2021) 'Signature texture features extraction using GLCM approach in Android Studio', *Journal of Physics: Conference Series*, Vol. 1804, No. 1, p.012043, IOP Publishing.

Fan, R., Li, X., Lee, S., Li, T. and Zhang, H.L. (2020) 'Smart image enhancement using CLAHE based on an F-shift transformation during decompression', *Electronics*, Vol. 9, p.1374.

Gárate-Escamila, A.K., El Hassani, A.H. and Andrès, E. (2020) 'Classification models for heart disease prediction using feature selection and PCA', *Informatics in Medicine Unlocked*, Vol. 19, pp.1–11, DOI: 100330.

Gedik, N. (2022) 'A method for the classification of mammograms using a statistical-based feature extraction', *International Journal of Biomedical Engineering and Technology*, Vol. 38, No. 1, pp.99–108.

Humeau-Heurtier, A. (2019) 'Texture feature extraction methods: a survey', *IEEE Access*, Vol. 7, pp.8975–9000.

Jain, A. and Zongker, D. (1997) 'Feature selection: evaluation, application, and small sample performance', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 2, pp.153–158.

Kalaiselvi, T., Kumarashankar, P. and Sriramakrishnan, P. (2022) 'Machine learning approach for automatic brain tumour detection using patch-based feature extraction and classification', *International Journal of Biomedical Engineering and Technology*, Vol. 39, No. 4, pp.396–411.

Khadse, V.M., Mahalle, P.N. and Shinde, G.R. (2020) 'Statistical study of machine learning algorithms using parametric and non-parametric tests: a comparative analysis and recommendations', *International Journal of Ambient Computing and Intelligence (IJACI)*, Vol. 11, No. 3, pp.80–105.

Khaire, U.M. and Dhanalakshmi, R. (2022) 'Stability of feature selection algorithm: a review', *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, No. 4, pp.1060–1073.

Khalil, M., Ayad, H. and Adib, A. (2018) 'Performance evaluation of feature extraction techniques in MR-brain image classification system', *Procedia Computer Science*, Vol. 127, pp.218–225.

Kshirsagar, P.R. et al. (2021) 'Mental task classification using wavelet transform and support vector machine', *International Journal of Biomedical Engineering and Technology*, Vol. 37, No. 4, pp.368–381.

Kshirsagar, P.R., Akojwar, S.G. and Bajaj, N.D. (2018) 'A hybridised neural network and optimisation algorithms for prediction and classification of neurological disorders', *Int. J. Biomedical Engineering and Technology*, Vol. 28, No. 4, pp.307–321.

Kumar, T.A., Rajakumar, G. and Samuel, T.S.A. (2021) 'Analysis of breast cancer using grey level co-occurrence matrix and random forest classifier', *Int. J. Biomedical Engineering and Technology*, Vol. 37, No. 2, pp.176–184.

Labati, R.D., Piuri, V. and Scotti, F. (2011) 'All-IDB: the acute lymphoblastic leukemia image database for image processing', *18th IEEE International Conference on Image Processing*, IEEE, p.10.

Li, X. and Yu, W. (2014) 'Fast support vector machine classification for large data sets', *International Journal of Computational Intelligence Systems*, Vol. 7, No. 2, pp.197–212.

Li, Y., Xie, Y. and Zhang, Q. (2018) '3D gesture recognition based on handheld smart terminals', *International Journal of Ambient Computing and Intelligence (IJACI)*, Vol. 9, No. 4, pp.96–111.

Loeffen, E.A.H. et al. (2020) 'Reducing pain and distress related to needle procedures in children with cancer: a clinical practice guideline', *European Journal of Cancer*, Vol. 131, pp.53–67 [online] https://www.sciencedirect.com/science/article/pii/S0959804920301131.

Maiti, A., Chatterjee, B. and Santosh, K.C. (2021) 'Skin cancer classification through quantized color features and generative adversarial network', *International Journal of Ambient Computing and Intelligence (IJACI)*, Vol. 12, No. 3, pp.75–97.

Menon, A.K. (2009) *Large-Scale Support Vector Machines: Algorithms and Theory*, Research Exam, University of California, San Diego.

Mimura, M. (2023) 'Impact of benign sample size on binary classification accuracy', *Expert Systems with Applications*, Vol. 211, pp.1–10, DOI: 118630.

Mishra, S. et al. (2017) 'Microscopic image classification using DCT for the detection of acute lymphoblastic leukemia (ALL)', *Proceedings of International Conference on Computer Vision and Image Processing*, pp.171–180, Springer, Singapore.

Mishra, S., Majhi, B., and Sa, P.K. (2018) 'GLRLM-based feature extraction for acute lymphoblastic leukemia (ALL) detection', *Recent Findings in Intelligent Computing Techniques*, pp.399–407, Springer, Singapore.

Mittal, H. and Saraswat, M. (2019) 'Classification of histopathological images through bag-of-visual-words and gravitational search algorithm', *Soft Computing for Problem Solving*, pp.231–241, Springer, Singapore.

Moshavash, Z., Danyali, H. and Helfroush, M.S. (2018) 'An automatic and robust decision support system for accurate acute leukemia diagnosis from blood microscopic images', *Journal of Digital Imaging*, Vol. 31, No. 5, pp.702–717.

Okwuashi, O. and Ndehedehe, C.E. (2020) 'Deep support vector machine for hyperspectral image classification', *Pattern Recognition*, Vol. 103, pp.1–10, DOI: 107298.

Öztürk, Ş. and Akdemir, B. (2018) 'Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA', *Procedia Computer Science*, Vol. 132, pp.40–46.

Pahuja, G., Nagabhushan, T.N. and Prasad, B. (2022) 'A comparative study of feature projection and feature selection approaches for Parkinson's disease detection and classification using T1-weighted MRI scans', *International Journal of Biomedical Engineering and Technology*, Vol. 38, No. 1, pp.65–80.

Pan, C., Park, D.S., Yang, Y. and Yoo, H.M. (2012) 'Leukocyte image segmentation by visual attention and extreme learning machine', *Neural Comput & Applic.*, Vol. 21, pp.1217–1227, Springer-Verlag London Limited, DOI: 10.1007/s00521-011-0522-9.

Raghavendra, U. et al. (2019) 'Computer-aided diagnosis for the identification of breast cancer using thermogram images: a comprehensive review', *Infrared Physics & Technology*, Vol. 102, p.103041.

Rajinikanth, V., Dey, N., Kavallieratou, E. and Lin, H. (2020) 'Firefly algorithm-based Kapur's thresholding and Hough transform to extract leukocyte section from hematological images', *Applications of Firefly Algorithm and its Variants*, pp.221–235, Springer, Singapore.

Renuka, T.V. and Surekha, B. (2021) 'Acute-lymphoblastic leukemia detection through deep transfer learning approach of neural network', *Proceeding of First Doctoral Symposium on Natural Computing Research*, pp.163–170, Springer, Singapore.

Rodrigues, L.F. et al. (2016) 'Leukocytes classification in microscopy images for acute lymphoblastic leukemia identification', *XII Workshop de Visão Computacional, WVC*, Campo Grande, MS, Brazil, pp.68–73.

Sahlol, A.T. et al. (2017) 'Elephant herd optimization with neural networks: a case study on acute lymphoblastic leukemia diagnosis', *2017 12th International Conference on Computer Engineering and Systems (ICCES)*, IEEE, pp.657–662.

Shree, N.V. and Kumar, T.N.R. (2018) 'Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network', *Brain Informatics*, Vol. 5, No. 1, pp.23–30.

Song, F., Guo, Z. and Mei, D. (2010) 'Feature selection using principal component analysis', *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, IEEE, Vol. 1, pp.27–30.

Sweetlin, J.D., Nehemiah, H.K. and Kannan, A. (2017) 'Feature selection using ant colony optimization with tandem-run recruitment to diagnose bronchitis from CT scan images', *Computer Methods and Programs in Biomedicine*, Vol. 145, pp.115–125..

Tali, R.V., Borra, S. and Mahmud, M. (2021) 'Detection and classification of leukocytes in blood smear images: state of the art and challenges', *International Journal of Ambient Computing and Intelligence (IJACI)*, Vol. 12, No. 2, pp.111–139.

Tarandeep, T. et al. (2017) 'An updated CLAHE approach for image enhancement with advanced wavelets', *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 6, pp.383–387.

Tsang, I.W. et al. (2005) 'Core vector machines: fast SVM training on very large data sets', *Journal of Machine Learning Research*, Vol. 6, No. 4, pp.363–392.

Vincent, S. and Chandra, J. (2022) 'Automated segmentation and classification of nuclei in histopathological images', *International Journal of Biomedical Engineering and Technology*, Vol. 38, No. 3, pp.249–266.