International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Extensive Study of Different Types of Leukemia using Image Processing Techniques

M. Abirami[*1], Revathy S.[2], Swathika R.[3], K. C. Rajheshwari[4], T. A. Mohanaprakash[5]

Submitted: 22/10/2023 **Revised**: 18/12/2023 **Accepted**: 26/12/2023

Abstract: The use of computer-aided diagnosis (CAD) has been on the rise during the last few years. Different diseases, such as leukaemia, can be detected using a variety of machine learning methods. The bone marrow and/or blood are both affected by leukaemia, which is a Sickness caused by WBCs. Detection of leukaemia at an early stage is essential to the patient's recovery and prognosis. The two primary forms of leukaemia that have evolved as a result of scientific developments are acute and chronic leukaemia. Myeloid and lymphoid cells are subtypes of any given sort. As a result, four separate types of leukaemia exist. A variety of strategies can be used to identify leukaemia subtypes. In spite of this, there has yet to be a complete examination of these strategies. Such a review will be quite helpful to those who are just starting out in this field of study and want to learn more. To fill the hole, this publication provides an extensive overview of previous studies on blood smear image analysis and leukaemia detection. For the most part, the review concentrates on discussing the underlying procedures and their claimed results. It also lists the problems in the field that have been resolved and the outstanding ones that remain.

Keywords: *Leukemia, Machine Learning, White Blood Cells, Blood Smear Images, Bone Marrow, Chronic Leukemia*

1. Background

1.1. Leukemia.

It is a disorder that affects WBC, and platelets make up blood. Platelets aid in the creation of a blood clot and in preventing bleeding. The tissues are accomplished by RBCs known as erythrocytes. When it comes to fighting sickness and infection, WBC is in charge. There is a condition known as leukaemia in which a large number of immature WBC is produced. It is a form of cancer that affects the bone core and blood, damaging the human body's immune scheme. Acute leukaemia and chronic leukaemia are the two primary types of leukaemia based on how quickly they progress. Chronic leukaemia affects WBC that can behave regularly and increase more slowly than acute leukaemia, in which the infected WBC grow rapidly and do not function normally. However, this can be dangerous because it's difficult to tell apart from a normal WBC in some cases based on the size and form of the CML.

1.1.1. Acute Lymphocytic Leukemia.

WBC in the marrow is the primary cause of ALL, a WBC malignancy that is most frequent in youngsters. Detection of ALL can be challenging due to the symptoms, which include fatigue, weakness, and joint and bone pain, being similar to flu and other prevalent illnesses. ALL can be classified in three ways: L1, L2, and L3 [1].

1.1.2. Acute Myeloid Leukemia.

Most people are diagnosed with acute leukaemia (AML) when the bone core begins to produce blasts and immature WBC. It may also result in abnormal RBCs and platelets. Early symptoms of AML, such as the common cold, Signs and symptoms may vary depending on the type of blood cell affected. Frequent nosebleeds and gum bleeding are two of the symptoms of acute myeloid leukaemia (AML), which can be characterized by fever and bone pain. AML is distinct from other forms of leukaemia because it has eight distinct subtypes. [2].

1.1.3. Chronic Leukemia Lymphoblastic.

CLL is a blood disease that worsens over time. Adults are more probable to experience this, although children are much less likely. Weight loss, fevers, night sweats, and recurring infections are all signs of CLL.

1.1.4. Chronic Myeloid Leukemia.

Slow-growing leukaemia, chronic myelogenous leukaemia (CML), can transform into acute myeloid leukaemia (AML), which spreads rapidly and is difficult to cure.

[1] Assistant Professor, Department of Computer Science and Engineering, Panimalar Engineering College, Chennai – 600123, INDIA, abiprj@gmail.com, https://orcid.org/0000-0003-4935-8938
[2] Assistant Professor, Department of Computer Science and Engineering, St. Joseph's Institute of Technology, OMR, Chennai-119, INDIA, revathymurthy440@gmail.com
[3]Assistant Professor, Department of Information Technology, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam 603 110, INDIA, swathikar@ssn.edu.in
[4] Associate Professor, Department of Computer Science and Engineering, Sona College of Technology, Salem- 636005, INDIA, rajeswarikc@sonatech.ac.in
[5]Associate professor, Department of Computer Science and Engineering, Panimalar Engineering College, Chennai – 600123, INDIA, tamohanaprakash@gmail.com, https://orcid.org/0000-0002-6885-4710
** Corresponding Author Email: abiprj@gmail.com*

Chronic, rapid, and blast stages of &is can all be seen as stages of the disease. A patient's leukaemia is at its strongest and slowest growth in the chronic stage. The blood cells in the second stage, known as the extended stage, are still immature. The explosion stage is the third and last stage. Figure 1 depicts a visual representation of blood structure and kinds of leukaemia.

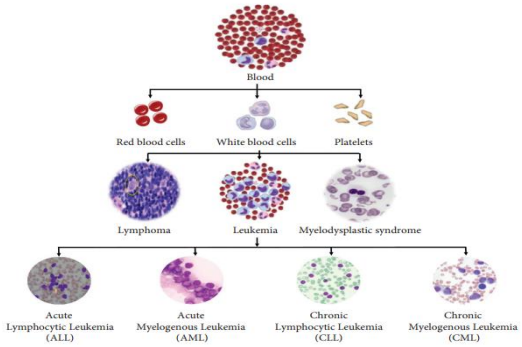

Fig1. Blood and leukemia types.

Hematologists must recognise the presence of leukaemia and its specific type to minimise medical hazards and to establish the suitable treatment for leukaemia. The detection of leukaemia through optical blood smear screening is a significant and time-consuming stage that requires the supervision of an expert. Many CAD strategies exist to address these issues. However, these methods have some shortcomings and need to be improved in terms of accuracy and overall competence.

2. Diagnosis of leukemia using Computer-aided:

These CAD tools can identify whether a blood smear contains leukaemia by looking at the morphology of the cells in it and then classifying them as either normal or positive. Rendering to leukaemia types, ALL, CLL, AML, and CML are further sub classified. In some systems, the count is between normal and aberrant. Figure 2 depicts the overall workflow employed by various CADx systems.

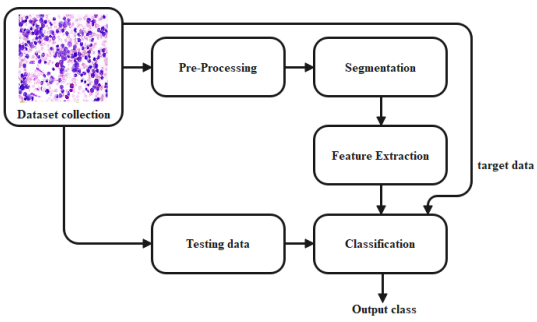

Fig2. Analysing leukocyte morphology in blood smear pictures using a DIP system workflow

2.1. Datasets

Use these data sets for research on the identification and diagnosis of leukaemia by looking at the blood smear images.

(1) **Acute Lymphoblastic Leukemia Record for Image Processing:** ALL blast cells will be segmented and classified in this dataset [3]. A Canon Power Shot G5 camera was used to capture the photographs. They can be downloaded in.jpg file with a resolution of 2592 x 1944 pixels and a 24-bit colour depth in the.png format. All-IDB1 and All-IDB2 versions of this dataset are available. 108 blood slide photos are included in the ALL-IDB1 dataset. Each.xyc ile provides the x,y coordinates of the centre of the leukemic cells. There is a second version of the dataset, called ALL-IDB2, which comprises from the images of ALL-IDB1 (i.e., blast cells).

(2) **American Society of Hematology (ASH):** Blood slide photos from a variety of disorders can be found in this online database [4]. More than 4,879 photos of ALL, lymphoma, myeloid problem, anaemia, and other diseases are included.

(3) **C-NMC Leukemia Dataset:** It is made accessible as part of the ALL test in ISBI-20192, and it contains dataset [5]. An estimated 15,000 450x450-pixel photos in BMP format are included in the collection, along with their associated labels in a CSV file.

(4) **The Munich Acute Myeloid Leukemia Dataset:** Around 18,000 pictures of single AML cells are included in this dataset [6]. Under the direction of professionals, they have been separated from the blood slides. Each image has a resolution of 400 x 400 pixels and is in the.tif file format. Morphological classes are discussed and summarized in.txt format, while their labels are supplied in.dat format.

(5) **SN-AM Dataset:** A bone marrow slides are included in this collection [7]. (MM). they were taken with a Nikon Eclipse-200 camera at 2560 x 1920 resolutions and are available in the.bmp format. Images of B-ALL and MM make up the dataset.

(6) **Clinical Proteomic Tumor Analysis Dataset:** This dataset [8] is employed to ascertain the molecular foundation of cancer's aetiology. To study the phenotype of cancer, radiologic images and pathology data are gathered. To view the tissue slide photos, you must use the.svs format, whereas the clinical data is stored in the JSON file format.

(7) **Hematology Atlas (HA):** Images of leukaemia, anaemia, parasites, and fungal infections are included in this dataset [9]. All, AML and hairy cell leukaemia are all included in this collection of 88 photos.

2.2 Pre-processing

Workflow pipeline stage 2 is when the following operations are conducted to the blood smear picture in order to make it suitable for the next stages.

(1) Noise filtering:

A fixed-pattern noise (FPN) can affect blood smear pictures. Filters like Wiener are used to reduce random noise. Despite the fact that these filters reduce noise, they also soften the image's crisp characteristics, such as cell borders. There are two methods for getting rid of FPN after random noise has been removed: re-calibration and background subtraction.

(2) Contrast enhancement:

Digital images of properly slides show the nucleus and cytoplasm of cells, as well as the blood cells and plasma. It is possible to use the contrast difference to segment leukocytes in a number of ways. Slide preparation errors, such as incorrect ink solution concentration, unstable staining agent, or under- or over-exposure, the nucleus and cytoplasm of the cells must be enhanced first. Contrast enhancement is necessary to properly segment WBCs, nuclei, and cytoplasm using contrast. Histogram equalization, linear contrast stretching, or both techniques in combination can be used to improve the overall contrast of a blood smear image. A non-linear transformation function is used for histogram equalization, which is both simple and fast. If you're concerned about losing contrast in small areas and in the background, this may not be the best option for you. Enhancement with fewer severe effects circumvents this histogram equalization limitation. Intensity information from the entire image is used in global contrast enhancement techniques to improve overall contrast.

(3) Colour space conversion:

The default colour model is easy to create and manipulate colours with the RGB model. If you're looking to process blood smear images, you may want to look elsewhere. The RGB model produces a non-linear and discontinuous colour space. Changes in hue are difficult to track due to the colour space's discontinuities. In addition, the RGB model's colour hue is affected by changes in illumination, making tracking and analysis difficult [10]. As a result, numerous studies have altered the input image's colour model during the preprocessing stage. The perceptually uniform CIE colour space with the CIELab colour model have been used to implement the HSV, HSL and HSI colour models in [11-12]. Moshavash et al. [13] used the CIELab and CMYK colour models. As an alternative, Mohapatra et al.[14] used the CIELab colour model and the HSV colour model to segment leukocytes and nuclei.

***(4)* Data augmentation:**

If you're going to use deep learning, you need a lot of training data to get it right. Because of the difficulty of obtaining and labelling medical data, this requirement of deep learning is not met in this area. It's one of the biggest issues that prevents. To solve this problem, data augmentation is used to significantly increase the sum of training samples without actually collecting new data It has features like random cropping, padding, rotation, and horizontal and vertical lip. When Kassani et al. [15] performed data augmentation on leukemic datasets, they used horizontal lip, vertical lip, and contrast adjustment to nearly eight-fold increase the dataset's size Image rotation and mirroring were used by Shaique [16] and Ghosh et al. [17] to perform augmentation. In order to improve the performance of their LeukoNet classifier, Mourya et al. [18] used amine.

2.3. Segmentation

Smear image following the pre-processing stage. Because blood cells come in a variety of forms and sizes, this is a difficult task. It is usually done in a series of steps, due to its complexity. Segmentation methods.

(1) Rule-based methods:

Leukocytes, their nucleus, and their cytoplasm are analysed using heuristic criteria drawn from past understanding of their structure. WBCs can be segmented using rules based on intensity, colour apply several rules sequentially to achieve the desired result. These techniques have a noteworthy advantage in that the constituent rules can be applied in a variety of ways. Further classification of rule-based approaches:

(a) Intensity-thresholding based approaches:

Leukocytes can be segmented from the blood smear image using intensity thresholding, which is one of the most straightforward methods. Leukocyte nuclei were segmented by using a global threshold, we use our past knowledge of the blood smear images and fine-tune our estimation by increasing the sum of edge points with large gradients. Morphological techniques are employed to remove single points and lines created by noise from the segmented regions. For leukocyte segmentation, Abd Halim et al. [19] employed a fixed global threshold of = 100 on the S component of HSI colour space transformed ALL and AML pictures. Methods that use adaptive thresholding such as Otsu thresholding, Zack algorithm are superior to those that use local or global thresholding. Leukocyte nuclei have been segmented using Otsu's approach and several post-processing techniques in [20]. For the segmentation of microscopic cells, the Zack method is commonly employed. For leukocyte segmentation, Toh et al. used color-based thresholding instead of grayscale thresholding. They used the R, G, and B value ratio to partition the blood smear image.

Segmenting the nucleus and cytoplasm was accomplished using G to B colour value ratios of 0.85 and 0.75, respectively. After that, the image is subjected to the median filter in order to eliminate any remaining noise.

(b) **Edge detection-based methods:**

These methods of discontinuities in the image to designate the cellular borders. The nucleus border in several research including [21] and [22], respectively. Safuan et al. [23] employed CHT to identify and count WBCs in blood smear pictures. A mixture of S channel from HSV color space and C channel from CMYK color space is used as input to the algorithm. They observed that a circle radius of 10−12 pixels is most suitable to detect WBCs even in the cluster region.

(c) **Region-based methods:**

The coherence qualities of picture pixels are exploited to organize them into homogeneous regions using region-based approaches. When it comes to seed-based region-growing (SBRG) techniques, an initial seed is required and then propagates outwards from that seed. Recursively, they look at the features of the pixels around the seed and group pixels with similar characteristics into distinct regions. Seeded region growth area extraction (SRGAE) was used by Harun et al. [25] to separate leukocytes using an SBRG technique modification. Because the approach concentrates on each pixel rather than the entire region, it is less susceptible to errors resulting from poor illumination or noisy limits. For the segmentation of leukocytes, another family of region-based approaches uses the watershed method. When using the watershed method, the nucleus-cytoplasm boundary is smoothed down using morphological operations following the segmentation step. Some approaches of region-based segmentation use clustering to group together neighbouring pixels with comparable features in each region. The nucleus and cytoplasm of WBCs have been segmented by a variety of methods, including K-means clustering (KMC), K-medoids clustering. For acute leukaemia blood cells, the author from [25] examined the performance of three distinct forms of clustering, classical K-means (CKM), FCM. Each algorithm is evaluated for both its qualitative and quantitative performance. All and AML blast cells are better segmented by the MKM-clustering algorithm than by the other two techniques.

(2) Deformable model-based methods:

In order to segment an object, the deformable replicas use flexible 2D or 3D curves. When an object is initialized, its curves are shaped by a combination of internal and external forces, as well as constraints specified by the user. During the evolution process, while the external forces push the model toward the object's border. Even in the face of severe image quality variability, these models are extremely effective at segmenting a wide range of different-shaped objects. The deformable model-based approaches are categorized according to how the model is defined in the shape domain:

(a) ***Parametric model-based approaches:***

Using a small number of factors, these models create a representation of an object's geometrical shape and appearance. The model's parameters are iteratively modified in response to the image's internal and exterior forces. The segmented output can be described as equilibrium when both forces are equal. Convergence and concise representation are two advantages of these models. ACMs, sometimes known as snakes, to separate the nuclei of leukocytes from blood smear pictures. The nucleus of a cell is difficult to segment because of its amorphous shape. It is easy for snakes to modify their shape, which allows them to successfully segment the nucleus of a cell. Nucleus segmentation can be followed by the use of a thresholding-based approach to distinguish cytoplasm from background. GVF snake is another active contour model that is used to segment cells in the blood smear image. Due to the broad capture area, the gradient-based ACM is superior to other gradient-based models. Furthermore, the GVF snake converges in one-ninth of the time it takes a regular snake model to do the same.

(b) ***Geometric model-based methods:***

Using three-variable function level sets, two-dimensional curves can be depicted. When it comes to topology control, numerical computation stability, and avoiding self-intersections, these types of models outperform their parametric counterparts. The nucleus segmentation of WBCs has been examined utilising the level set method. It has been used in conjunction with other algorithms by a small number. Al-Dulaimi [26] used a convolution of active smooth function to modify the level-set technique. As a result of the change, blast cells were properly segmented. They employed a level set technique shadowed by a clever edge detector for the segmentation. Using this method, you can only benefit from basic borders; more complex ones aren't.

(3) Machine learning methods:

Deep technologies have been widely used in the medical field to segment numerous anatomical components such as the lungs, heart, and liver using machine learning-based methodologies. Each pixel in the image is classified as either belonging to the anatomical component or the background using one of these approaches, which create a binary or multi-class classifier. The term "pixel classification based approaches" refers to the fact that these algorithms classify each individual pixel.

The PC-based approaches typically use encoder-decoder

architecture. Convolution and pooling are used in the encoder network to learn feature maps at different scales. All photos are upsampled such that they are the same size after decoding. RBCs and WBCs were segmented from peripheral blood smear pictures using SegNet. VGG16 weights are used to initialise the SegNet network's weights. This architecture is better at defining boundaries [27].

2.4. Feature Engineering

To identify, quantify, or describe the qualities of objects in an image, we employ features, which are collections of non-redundant and useful numerical values. To be successful, feature engineering calls for a high level of technical competence as well as a thorough understanding of the subject matter. The features of shallow ML-based methods are intuitive and manually created. Manual subtasks such as extracting features and normalising and selecting features are part of this procedure. In order to conduct segmentation or classification jobs effectively, it is difficult to extract only the elements that are absolutely necessary and sufficient. Due of the automatic feature extraction, deep learning-based algorithms are used. Feature extraction, normalisation, and selection are common subtasks utilised in shallow machine learning algorithms described in this area.

(1) ***Feature extraction:*** Blast cells and non-blast cells can be distinguished primarily by three types of characteristics.

(a) ***Geometrical features:*** One of the most essential criteria in detecting whether a cell is a blast or not is the geometrical shape of its nucleus. Shape characteristics have been described using both region and border descriptors. These features were employed by Huang et al. [28] in their research. In Putzu et al. [29] calculated the area under each leukocyte's convex hull. An object's density is determined by its solidity, which may be determined using this method. Blast cells are present when the solidity is less than 1, approximately 0.90.

The roundness of the segmented items was calculated to identify them as leukocytes. An object is classed as leukocyte if its solidity is greater than 0.93 and non-leukocyte if its roundness is greater than 0.8. The roundness and solidity (value=0.98) features were also utilised by Sahol et al. [30] to achieve successful classification outcomes.

The roughness of the nucleus boundary of the blast cell has been measured using this dimensionalless metric. As a result, geometric features are frequently combined with additional features like colour and texture to improve segmentation accuracy.

(b) ***Textural features:*** Subsegments and their interrelationships in a digital image are known as texture because of the way they are arranged. Variations in intensity or frequency are common manifestations of these kinds of arrangements. The following are a few methods for describing textural features.

(i) ***Fourier descriptors:***

The Fourier transform function is used to encode an image's texture with the help of Fourier descriptors. Descriptors for textural information often use the 2-D discrete Fourier transform (DFT) or the fast Fourier transform. For cell images, DFT or FFT are used to calculate frequency domain features like kurtosis, and skewness. Wold's decomposition model was utilised by Reta et al. [31] to analyse the image's textural properties. The harmonic field in cells is solved using DFT in this model, which identifies the most useful harmonic peaks. This model has the advantage of resembling the human vision system (HVS) and assessing images for both periodic and random textures. Aside from the fact that it can withstand translation, rotation, and scaling on a given image without losing any of its properties,

(ii) ***Wavelet texture features:***

Discrete wavelet transformations (DWT), Gabor wavelet transforms (GWT), Sym4 and Db4 wavelet transforms are used to extract these features. When it comes to temporal resolution, the wavelet transform beats out the Fourier transform handily. When looking for leukaemia in blood cells, Neoh, et al. [32] used wavelet analysis to examine the nucleus structure using Haar.

(iii) ***Haralick texture features:***

A surrounding grey levels are counted. Checking for co-occurring values is done along a specific axis and at a specific distance D. ofset value ((-1,0), (0-1), (1-1), (0-1)) is used to extract statistical metrics such as contrast, correlation, homogeneity and energy from GLCM. Four GLCMs were used by Roopa et al. [33] to extract textural features. There are four different orientations with unit distances for each of them. A total of 54 textual features were used in the GLCM to derive descriptors for each object. Additionally, textural descriptors like kurtosis and skewness are incorporated. AML and ALL detection have relied heavily on the use of textural cues.

(iv) Laws ***texture method:***

This technique uses a variety of kernels to identify the texture of a photograph. There were five one-dimensional masks originally employed, notably level, spot, wave, edge, and ripples. Matrix values for these one-dimensional masks are as follows: [+1 4 6 4 1 +1], [-1 1 2 1 1 -1], "[-1 +2 1 2 1 +1]" and "[+1 2 1 2 +1]" for the first and second masks. Hedge et al. [33] utilised laws textural properties to classify WBC in peripheral blood smear pictures.

(c) ***Chromatic features*****:**

A variety of classification and segmentation tasks can benefit greatly from the use of chromatic data like intensity and hue. As contrast to textural features, these traits are highly discriminative and computationally affordable to extract. The hue of blood cells has been utilised extensively (i.e., nucleus and cytoplasm). The RGB channels of the peripheral blood smear image were used to generate colour features. As an alternative to directly calculating the intensity values, Nasir et al. used the mean of the colour intensity.

Blood cells can be identified in an ink stain image using additional features such as mean and standard deviation computed from the RGB channels to variations in illumination, making it a poor choice for colour representation. Color attributes based on different colour spaces such as CIE L*a*b, HSL, and HSI were used by Fatichan et al. [34].

(2) ***Feature normalization:***

Extracted characteristics have a wide range of values. With data normalization, the varying value ranges can be converted to a single scale without affecting the original data. The model's performance and training stability are both enhanced as a result of this.

(4) Feature selection:

Many features are extracted during the feature extraction procedure. It's important to note that not all of the aspects are equally important in making a decision. Machine learning algorithms may suffer if they have to deal with a large number of irrelevant features. In order to limit the sum of features in a feature space, feature selection, or dimensionality reduction, is an essential activity. The following are the broad classifications of feature selection techniques:

(a) ***Filter methods*****:** The correlation input factors that have a high correlation with the desired outcome are then chosen. Features can be selected without relying on the machine learning algorithm and at a low cost in terms of processing resources. Linear Discriminant Analysis (LDA), It is closely similar to PCA, which also identifies the axes with the greatest variation, but does not take target classes into account.

Immature lymphoblast cells technique, which has been refined throughout time. Classification of lymphoblasts has been carried out using LDA. LDA has a difficult time deciding which features to include. As a result, PCA has been employed prior to LDA in order to make feature selection a breeze. ANOVA is a statistical technique that uses a p-value to determine the influence of individual features. There are other features that have lower p-values (usually less than 0.05) that fail to distinguish between different groups.

(b) ***Wrapper methods*****:**

In order to select an ideal subset of features, these approaches incorporate feature selection with the learning process. Using stacked cross-validation, the selection procedure is iterative and, as a result, computationally expensive. For feature selection, the wrapper techniques employ either forward or backward selection. In terms of forward selection algorithms, the Tabu search is one of the best when it comes to solving the problem of feature selection optimally. Chain Tabu search was employed by Parvaresh et al. [35] to iteratively explore optical characteristics without getting stuck in local minima. Particle swarm optimization (PSO), probabilistic incremental programme evolution (PIPE), and social spider optimization algorithm (SSOA) are some of the other wrapping methods used to automatically classify leukaemia from blood smear images.

3.5. Classification

A classifier assigns a class label, such as "ALL positive," "ALL negative," at the end of the workflow pipeline. In some cases, a cell count is also generated by a regression module in addition to the class labels. Classifier and repressor learning from the training phase is used to complete these assignments. Classes other than instance-based use a labelled training dataset and alter their model parameters until they start correctly classifying a big enough number of input photos in their training phase. A trained classifier is then used to classify photos that were not previously part of the training set. This section provides a list of blood smear image classifiers that have been utilised in previous investigations:

(1) Instance-based Classifiers:

Classification is carried out by these classifiers by comparing example with the previously trained instances. Memory-based classifiers are a subcategory of these classifiers. A classifier's performance is greatly affected by the value of k, even if it is straightforward to build. Classifying blasts in blood samples from patients with acute leukaemia has been done by Supardi et al. [36]. They found that k = 4 was the most accurate classification method in their experiments. For the provided dataset, Purwanti et al. [37] tested all odd values of k ranging from 1 to 15, and found that k = 7 produced the best results. Many researches have utilised the k-NN classifier for blood smear image processing, but the non-parametric nature of the classifier has numerous limitations. Parametric classifiers such as SVM and ANN.

(2) Support Vector Machine (SVM):

The SVM is a top-notch classifier in machine learning. Even if you have data with non-linear relationships, our algorithm is capable of learning from it. Through the use of

hyperplane fitting, it seeks to categorise the input instances. For the hyper-plane to have the greatest distance from nearby data points, it must be fitted precisely. Depending on the data, it's possible that the given data points are linearly interconnected. Kernel functions are used in SVM to solve this problem by transforming linearly inseparable data into linearly separable data in higher-dimensional space. It's easy to apply linear and polynomial kernels, however RBF kernels are known for producing reliable results. The blood smear pictures were classified as non-leukemic or leukemic using an SVM classifier. Using multi-class SVM, the pictures are further sub-classified into L1-L3 sub-types. For each type of kernel, the parameter values are optimized for maximum accuracy. The accuracy achieved was employed in all of these investigations using the SVM classifier. There is a need to examine the presentation of SVM and other kernels on larger datasets, as well.

(3) Artiicial Neural Networks (ANN):

Because of its generalisation and parallel processing capabilities, ANNs, who are based on biological neural networks, have grown in popularity recently. It is through this training process that the ANNs learn, updating the weights of each neuron. ANNs can be trained to produce the required result at any point during the process. Multi-category classification is much easier to implement with ANNs than with SVMs.

To detect leukaemia in blood smear images, researchers utilised a feedforward multi-layer perceptron (MLP). There have been studies using MLP to identify and classify acute leukaemia from blood smear pictures with a 91% success rate. For all categorization, Vincent et al. [39] employed a LM optimization algorithm is used to train both subnetworks. Madhukar et al. [40] employed the identical pre-classification steps that they did. NNs, rather than SVMs, were utilised in the classification step, which raised classification accuracy by 4% as a result of the difference in strategy. Discriminative characteristics must be of a high quality in order for ANNs with shallow depth (i.e., fewer layers) to perform well. It is a challenging task that demands substantial subject knowledge and specialised training to extract high-quality features. As a solution to the problem of shallower ANNs, researchers investigated increasing the depth of the ANN. However, tweaking a large number of trainable parameters is a difficult operation when trying to train deep ANNs. It takes a lot of computing power and a lot of training data to achieve this level of tweaking. In the previous decade, we've figured out how to fix these issues. Because of the availability of GPUs and big training datasets, deep neural networks (ANNs) have become a lot easier to train. Today, deep CNN, play a critical role when it comes to medical imaging tasks such as brain tumour segmentation and mitotic cell detection in medical imaging. In order to improve the categorization of melanoma and basal skin cancer, this research presents a unique strategy that makes use of ResNet models and uncertainty quantification approaches. By enhancing the precision and consistency of skin cancer detection, it makes important contributions to the area of medicine [48].

CNN's performance was also compared to that of a feed-forward network. According to the researchers, deep CNN has a classification accuracy of higher than that of the feed-forward system. Li [41] found that using CNN for image classification of ALL outperforms other neural networks. They found that CNN was the best, k-NN was the worst, and the other NNs performed somewhere in the middle. The concept of transfer learning has also helped to raise the profile of convolutional neural networks. An image dataset like ImageNet from Large Scale Visual Recognition Challenge (LSVRC) can be used to train a CNN model if customised training data are scarce. Natural image datasets provide the CNN with general low-level features like edges and blobs, while the customised dataset provides more specific high-level features. Transfer learning was used to classify all images from peripheral blood smears. Classification has been achieved through transfer learning. This study unveils a novel machine learning ensemble strategy for improving hair loss forecasting performance. This strategy provides a complete and rigorous framework for early diagnosis and intervention in the area of hair loss prevention by integrating the capabilities of many algorithms [47]. The paper presents a novel approach for multi-class skin cancer classification, combining a Hybrid Dynamic Salp Swarm Algorithm with Weighted Extreme Learning Machines and incorporating transfer learning techniques. This research offers a promising method to improve the accuracy and effectiveness of skin cancer diagnosis, making it relevant for the healthcare and machine learning communities [49].

(4) Ensemble classifiers: To put it simply, the Latin word "insimul" means "all together" in English. Consequently, an ensemble classifier combines numerous classifiers that learn multiple hypotheses and work together to tackle a specific problem. It is possible to combine the results of individual classifiers in a variety of ways, such as by majority vote or averaging. Recent years have seen the rise of ensemble classifiers. KNN, MLP, and SVM were used to detect leukaemia by Mohapatra and colleagues [42]. Using MobileNetV2 and its variants as the base classifiers, Verma et al. [43]. Escalante et al. [44] used the particle swarm model selection (PSMS) method. A set of n-independent binary classifiers are used to develop an n-class classifier. The n-classifiers are used. The research paper employs an ensemble learning technique that integrates Logistic Regression, Gradient Boosting, and Random Forest algorithms to enhance the prediction of

Chronic Kidney Disease. By combining the predictive power of these diverse algorithms, the approach significantly improves accuracy, as demonstrated by lower RMSE and MSE values, making it a valuable tool for early CKD perception and potential application in various predictive tasks [46]. The paper introduces an innovative approach for skin cancer prediction, utilizing an Enhanced Genetic Algorithm in conjunction with Extreme Learning Machines. This research contributes to the field of computer science and smart technology by offering a promising method to enhance the accuracy and efficacy of skin cancer diagnosis [51].

3. Research Gaps:

The following research holes were discovered after a thorough evaluation of the relevant literature. Many studies do not take into account cells that overlap at the segmentation stage. During the staining technique, the cells overlap in a large number of real-world situations. Different bio-inspired algorithms may be used for segmentation, which may show to be effective. Classifier performance can still be better-quality using a variety of optimization strategies. In the later stages of leukaemia, the disease is extremely hazardous. ALL, Acute AML, Chronic Lymphoblastic Leukemia (CLL) [45] are some of the many forms. Pathologists face a significant barrier in identifying these pathogens because only their morphology can speak to them.

Subtypes of ALL include L1, L2, and L3, which are all subtypes of ALL. Blast cells and white blood cells (WBC) in the blood are examined for morphological characteristics before being assigned to one of three groups. There are numerous sub-types of ALL, each of which indicates a distinct infection and a separate treatment plan for the patients who suffer from it. Researchers in this field tend to overlook the existence of these sub-types. Similarly, there are many subtypes of AML, ranging from M0 to M7. The treatment instructions for each of these sub-types are also distinct. Furthermore, the distinctness of types T1, T2, and T3 is still a problem for scholars. In most situations, the diagnosis of these various subtypes is overlooked. In most circumstances, performance can only be measured in terms of precision. Errors in accuracy can be corrected. An extremely small sample size is used to test the accuracy of several phases of blood cell analysis. This research introduces a new method for medical picture categorization by bringing together Gated Deep Reinforcement Learning and Red Deer Optimization. This study makes a significant contribution to the study of medical images and artificial intelligence by proposing a way to improve the reliability and productivity of medical picture categorization [51]. This research focuses on improving picture quality using Digital Object Tracking that is platform-specific. It makes a contribution to cognitive computing and information processing by presenting a strategy to enhance picture quality across many file types [52].

4. Conclusion:

Performing blood smear images is a hard endeavor. Slide preparation issues, such as stain fluctuations, stain debris, and image acquisition issues, such as noise and light dispersion, can affect digital blood smear photographs. There are several pre-processing processes that make the input image suitable for subsequent processing. These include noise filtering, contrast improvement, and colour space conversion after the photos have been processed, blood cells are extracted via segmentation. Even while current segmentation algorithms do a good job of segmenting sparsely scattered cells, it is still an outstanding research question how to accurately segment cells that overlap in densely packed locations. Based on cell size and shape, overlapping cells can be separated using the watershed algorithm controlled by a marker or concavity analysis. In recent years, there have been few machine learning-based techniques algorithms for overlapping cell segmentation. Many difficulties confront the classifiers, including the need for a large amount of training data, classification accuracy that is sufficient, and results that are generalizable, reproducible, and understandable. Solutions like population-based augmentation and AutoAugument

Acknowledgements

No funding sources.

Author contributions

All authors are equally contributed in preparing, experimenting and reviewing the article.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] S. Shafique and S. Tehsin, “Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks,” Technology in Cancer Research & Treatment, vol. 17, 2018.

[2] https://www.cancercenter.com/cancer-types/leukemia/types/acute-myeloid-leukemia.

[3] Ruggero Donida Labati, Vincenzo Piuri, and Fabio Scotti. 2011. ALL-IDB : THE ACUTE LYMPHOBLASTIC LEUKEMIA IMAGE DATABASE FOR IMAGE PROCESSING. 2011 18th IEEE International Conference on Image Processing (2011), 2045ś2048. DOI: http://dx.doi.org/10.1109/ICIP.2011.6115881

[4] ASH: https://imagebank.hematology.org/. ([n. d.]).

[5] Sbilab. https://competitions.codalab.org/competitions/20395#learn_the_details-data-description. ([n. d.]).

[6] Christian Matek, Simone Schwarz, Carsten Marr, and Karsten Spiekermann. 2019. A Single-cell Morphological Dataset of Leukocytes from AML Patients and Non-malignant Controls. (2019). DOI:http://dx.doi.org/10.7937/TCIA.2019.36F5O9LD

[7] Anubha Gupta and Rita Gupta. 2019. SN-CanData: White Blood Cancer Dataset of B-ALL and MM for Stain Normalization. (2019). DOI:http://dx.doi.org/10.7937/TCIA.2019.OF2W8LXR

[8] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). 2019. Imaging Data from the Clinical Proteomic Tumor Analysis Consortium Acute Myeloid Leukemia [CPTAC-AML] collection. (2019). DOI:http://dx.doi.org/10.7937/TCIA.2019.

[9] N. Medeiros. http://hematologyatlas.com/principalpage.htm. ([n. d.]).

[10] Edgar Chavolla, Daniel Zaldivar, Erik Cuevas, and Marco A Perez. 2018. Color Spaces Advantages and Disadvantages in Image Color Clustering Segmentation. In Advances in Soft Computing and Machine Learning in Image Processing. Springer, 3ś22.

[11] Muhammad Sajjad, Siraj Khan, Zahoor Jan, Khan Muhammad, Hyeonjoon Moon, Jin Tae Kwak, Seungmin Rho, Sung Wook Baik, and Irfan Mehmood. 2016. Leukocytes classiication and segmentation in microscopic blood smear: a resource-aware healthcare service in smart cities. IEEE Access 5 (2016), 3475ś3489.

[12] Sos Agaian, Monica Madhukar, and Anthony T Chronopoulos. 2014. Automated screening system for acute myelogenous leukemia detection in blood microscopic images. IEEE Systems journal 8, 3 (2014), 995ś1004.

[13] Zeinab Moshavash, Habibollah Danyali, and Mohammad Sadegh Helfroush. 2018. An automatic and robust decision support system for accurate acute leukemia diagnosis from blood microscopic images. Journal of digital imaging 31, 5 (2018), 702ś717.

[14] Subrajeet Mohapatra, Sushanta Shekhar Samanta, Dipti Patra, and Sanghamitra Satpathi. 2011. Fuzzy based blood image segmentation for automated leukemia detection. In 2011 International Conference on Devices and Communications (ICDeCom). IEEE, 1ś5.

[15] Sara Hosseinzadeh Kassani, Michal J Wesolowski, Kevin A Schneider, Ralph Deters, et al. 2019. A Hybrid Deep Learning Architecture for Leukemic B-lymphoblast Classiication. arXiv preprint arXiv:1909.11866 (2019).

[16] Sarmad Shaique and Samabia Tehsin. 2018. Acute lymphoblastic leukemia detection and classiication of its subtypes using pretrained deep convolutional neural networks. Technology in cancer research & treatment 17 (2018), 1533033818802789.

[17] Arna Ghosh, Satyarth Singh, and Debdoot Sheet. 2018. Simultaneous localization and classiication of acute lymphoblastic leukemic cells in peripheral blood smears using a deep convolutional network with average pooling layer. 2017 IEEE International Conference on Industrial and Information Systems, ICIIS 2017 - Proceedings 2018-January (2018), 1ś6. DOI:http://dx.doi.org/10.1109/ICIINFS.2017.

[18] Simmi Mourya, Sonaal Kant, Pulkit Kumar, Anubha Gupta, and Ritu Gupta. 2018. LeukoNet: DCT-based CNN architecture for the classiication of normal versus Leukemic blasts in B-ALL Cancer. arXiv preprint arXiv:1810.07961 (2018).

[19] NH Abd Halim, MY Mashor, AS Abdul Nasir, NR Mokhtar, and H Rosline. 2011. Nucleus segmentation technique for acute leukemia. In 2011 IEEE 7th international colloquium on signal processing and its applications. IEEE, 192ś197.

[20] Emad A Mohammed, Mostaja MA Mohamed, Christopher Naugler, and Behrouz H Far. 2013. Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding. In 2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, 1ś5.

[21] Subhash Rajpurohit, Sanket Patil, Nitu Choudhary, Shreya Gavasane, and Pranali Kosamkar. 2018. Identiication of Acute Lymphoblastic Leukemia in Microscopic Blood Image Using Image Processing and Machine Learning Algorithms. 2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018 Cll (2018), 2359ś2363.

[22] Hatungimana Gervais. 2016. Computer-Aided Screening for Acute Leukemia blood infection using gray-level intensity. 2016 International Conference on Information & Communication Technology and Systems (ICTS) (2016), 69ś74.

DOI:http://dx.doi.org/10.1109/ICTS.2016.7910275

[23] Syadia Nabilah Mohd Safuan, Mohd Razali Md Tomari, and Wan Nurshazwani Wan Zakaria. 2018. White blood cell (WBC) counting analysis in blood smear images using various color segmentation methods. Measurement 116 (2018), 543ś555.

[24] K Sudha and P Geetha. 2020. A novel approach for segmentation and counting of overlapped leukocytes in microscopic blood images. Biocybernetics and Biomedical Engineering 40, 2 (2020), 639ś648.

[25] Nor Hazlyna Harun, AS Abdul Nasir, Mohd Yusof Mashor, and Rosline Hassan. 2015. Unsupervised segmentation technique for acute leukemia cells using clustering algorithms. World Academy of Science, Engineering and Technology International Journal of Computer, Control, Quantum and Information Engineering 9 (2015), 253ś59.

[26] Khamael Al-Dulaimi, Inmaculada Tomeo-Reyes, Jasmine Banks, and Vinod Chandran. 2016. White blood cell nuclei segmentation using level set methods and geometric active contours. In 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, 1ś7.

[27] Thanh Tran, Oh-Heum Kwon, Ki-Ryong Kwon, Suk-Hwan Lee, and Kyung-Won Kang. 2018. Blood cell images segmentation using deep learning semantic segmentation. In 2018 IEEE International Conference on Electronics and Communication Engineering (ICECE). IEEE, 13ś16.

[28] Der-Chen Huang, Kun-Ding Hung, and Yung-Kuan Chan. 2012. A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images. Journal of Systems and Software 85, 9 (2012), 2104ś2118.

[29] Lorenzo Putzu, Giovanni Caocci, and Cecilia Di Ruberto. 2014. Leucocyte classiication for leukaemia detection using image processing techniques. Artiicial intelligence in medicine 62, 3 (2014), 179ś191.

[30] Ahmed T Sahlol, Ahmed M Abdeldaim, and Aboul Ella Hassanien. 2019. Automatic acute lymphoblastic leukemia classiication model using social spider optimization algorithm. Soft Computing 23, 15 (2019), 6345ś6360.

[31] Carolina Reta, Leopoldo Altamirano, Jesus A. Gonzalez, Raquel Diaz-Hernandez, Hayde Peregrina, Ivan Olmos, Jose E. Alonso, and Ruben Lobato. 2015. Segmentation and classiication of bone marrow cells images using contextual information for medical diagnosis of acute leukemias. PLoS ONE 10, 6 (2015), 1ś18. DOI:http://dx.doi.org/10.1371/journal.pone.0130805

[32] Siew Chin Neoh, Worawut Srisukkham, Li Zhang, Stephen Todryk, Brigit Greystoke, Chee Peng Lim, Mohammed Alamgir Hossain, and Nauman Aslam. 2015. An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images. Scientiic Reports 5 (2015), 1ś14. DOI:http://dx.doi.org/10.1038/srep14938.

[33] Roopa B. Hegde, Keerthana Prasad, Harishchandra Hebbar, and Brij Mohan Kumar Singh. 2019. Comparison of traditional image processing and deep learning approaches for classiication of white blood cells in peripheral blood smear images. Biocybernetics and Biomedical Engineering 39, 2 (2019), 382ś392. DOI:http://dx.doi.org/10.1016/j.bbe.2019.01.005

[34] Chastine Fatichah, Martin L Tangel, Fei Yan, Janet P Betancourt, M Rahmat Widyanto, Fangyan Dong, and Kaoru Hirota. 2015. Fuzzy feature representation for white blood cell diferential counting in acute leukemia diagnosis. International Journal of Control, Automation and Systems 13, 3 (2015), 742ś752.

[35] Hamed Parvaresh, Hedieh Sajedi, and Seyed Amirhosein Rahimi. 2018. Leukemia Diagnosis using Image Processing and Computational Intelligence. In 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES). IEEE, 000305ś000310.

[36] NZ Supardi, MY Mashor, NH Harun, FA Bakri, and R Hassan. 2012. Classiication of blasts in acute leukemia blood samples using k-nearest neighbour. In 2012 IEEE 8th International Colloquium on Signal Processing and its Applications. IEEE, 461ś465.

[37] Endah Purwanti and Evelyn Calista. 2017. Detection of acute lymphocyte leukemia using k-nearest neighbor algorithm based on shape and histogram features. In Journal of Physics: Conference Series, Vol. 853. IOP Publishing, 012011.

[38] Morteza MoradiAmin, Ahmad Memari, Nasser Samadzadehaghdam, Saeed Kermani, and Ardeshir Talebi. 2016. Computer aided detection and classiication of acute lymphoblastic leukemia cell subtypes based on microscopic image analysis. Microscopy research and technique 79, 10 (2016), 908ś916.

[39] Ivan Vincent, Ki-Ryong Kwon, Suk-Hwan Lee, and Kwang-Seok Moon. 2015. Acute lymphoid leukemia classiication using two-step neural network classiier. In 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV). IEEE, 1ś4.

[40] Monica Madhukar, Sos Agaian, and Anthony T

Chronopoulos. 2012. New decision support tool for acute lymphoblastic leukemia classification. In Image Processing: Algorithms and Systems X; and Parallel Processing for Imaging Applications II, Vol. 8295. International Society for Optics and Photonics, 829518.

[41] Richard Sipes and Dan Li. 2019. Using convolutional neural networks for automated ine grained image classiication of acute lymphoblastic leukemia. Proceedings - 3rd International Conference on Computational Intelligence and Applications, ICCIA 2018 (2019), 157ś161.

[42] Subrajeet Mohapatra, Dipti Patra, and Sanghamitra Satpathy. 2014. An ensemble classiier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Computing and Applications 24, 7-8 (2014), 1887ś1904.

[43] Ekansh Verma and Vijendra Singh. 2019. ISBI Challenge 2019: Convolution Neural Networks for B-ALL Cell Classiication. In ISBI 2019 C-NMC Challenge: Classiication in Cancer Cell Imaging. Springer, 131ś139.

[44] Hugo Jair Escalante, Manuel Montes-y Gómez, JesúS A GonzáLez, Pilar GóMez-Gil, Leopoldo Altamirano, Carlos A Reyes, Carolina Reta, and Alejandro Rosales. 2012. Acute leukemia classiication by ensemble particle swarm model selection. Artiicial intelligence in medicine 55, 3 (2012), 163ś175.

[45] Al-Tahhan, F.E., Sakr, A.A., Aladle, D.A., Fares, M.E.: Improved image segmentation algorithms for detecting types of acute lymphatic leukaemia. IET Image Process. 13(13), 2595–2603(2019).

[46] Dhanwanth, B. ., Vivek, B. ., Abirami, M. ., Waseem, S. M. ., & Manikantaa, C. (2023). Forecasting Chronic Kidney Disease Using Ensemble Machine Learning Technique. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 336–344. https://doi.org/10.17762/ijritcc.v11i5s.7035.

[47] Sai, C. N. V. ., Archana, E. ., Vivek, B. ., Dhanwanth, B. ., & K. S., V. . (2023). Enhancing Hairfall Prediction: A Comparative Analysis of Individual Algorithms and An Ensemble Method. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(6s), 499–508. https://doi.org/10.17762/ijritcc.v11i6s.6958.

[48] Deepa, D., V. Muthukumaran, V. Vinodhini, S. Selvaraj, M. Sandeep Kumar, and J. Prabhu. "Uncertainty Quantification to Improve the Classification of Melanoma and Basal Skin Cancer Using ResNet Model." Journal of Uncertain Systems 16, no. 01 (2023): 2242010.

[49] Panneerselvam, Ramya, and Sathiyabhama Balasubramaniam. "Multi-Class Skin Cancer Classification Using a Hybrid Dynamic Salp Swarm Algorithm and Weighted Extreme Learning Machines with Transfer Learning." Acta Informatica Pragensia 12, no. 1 (2023): 141-159

[50] Ramya, P., and B. Sathiyabhama. "Skin Cancer Prediction using Enhanced Genetic Algorithm with Extreme Learning Machine." Journal of Trends in Computer Science and Smart Technology 5, no. 1 (2023): 1-13.

[51] N. Ganesh, S. Jayalakshmi, R. C. Narayanan, M. Mahdal, H. M. Zawbaa and A. W. Mohamed, "Gated Deep Reinforcement Learning With Red Deer Optimization for Medical Image Classification," in IEEE Access, vol. 11, pp. 58982-58993, 2023, doi: 10.1109/ACCESS.2023.3281546

[52] G. Maheswari, K. Remya, R. Reenadevi, J. Surendiran, S. Prasad and K. Sathish, "Image Quality Enhancement by format-specific Digital Object Tracking," 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP), Bengaluru, India, 2022, pp. 1-6, doi: 10.1109/CCIP57447.2022.1005869