

WHITE BLOOD CELL CANCER CLASSIFICATION USING MODIFIED HONEY BEE OPTIMIZATION ALGORITHM (MHBO) BASED FEATURE SELECTION AND NEURAL NETWORK CLASSIFICATION MODEL

D. Hema Priya¹, S. S. Rajasekar², G. Ranjith³

¹ PG Scholar, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam.

² Associate Professor, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam.

³ Assistant Professor, Department of Artificial Intelligence and data Science, Bannari Amman Institute of Technology, Sathyamangalam.

Abstract: A blood malignancy known as leukemia is defined by the bone marrow's abnormal and uncontrollably high production of leukocytes, or white blood cells (WBCs). It is possible to identify and diagnose illnesses early by examining photographs of minute blood cells. lately, hematopathologists have been analyzing, detecting, and identifying leukemia kinds in patients utilizing image-processing techniques. Since no specialized equipment is required for lab testing, image detection is a quick and affordable way of detection. Several image processing tools are created for obtaining significant data from medical photos to enhance patient diagnosis. In this research, a feed forward-back propagation neural network framework is presented, and the kind of cancer found in the cells is ultimately predicted. First, the provided photos are subjected to the filtering procedure. The second method is the Fuzzy Inference System (FIS) technique, which is reliable in the face of changing illumination levels and considers the stability levels of color elements to identify edges among color bone marrow microscopic pictures. Third Active contour detection provides an accurate curve of the boundary edge of the cells. And then the background separation is done by Pinched Flow Fractionation (PFF). Then the feature extraction is carried out by the Modified Honey Bee optimization algorithm (MHBO). Finally, the Feed Forward Back Propagation Neural Network (FFBNN) framework-based classifier is proposed for predicting the white blood cells. Based on the findings, this suggested approach created an automated system that allows medical practitioners to accurately diagnose all forms and subtypes of the disease.

Keywords: Bone marrow microscopic images, Fuzzy Inference System (FIS), Pinched Flow Fractionation (PFF), Modified honey bee optimization algorithm (MHBO), Feed Forward Back Propagation Neural Network framework

1. Introduction

The three main cell types that compose blood are white blood cells, platelets, and red blood cells. Red blood cells are essential for removing carbon dioxide from the body and transferring oxygen from the heart to all tissues [1]. They constitute as much as half the total blood volume. Since WBCs are the body's first line of defense against illnesses, they are essential to the immunological system. Consequently, precise WBC categorization is essential and is frequently requested. Based on how the cytoplasm appears, WBCs can be divided into two categories. Granulocytes, which comprise neutrophils, basophils, and eosinophils, constitute the first kind [2]. Monocytes and lymphocytes constitute the second type, known as agranulocytes. The lymphatic system, which produces blood cells, is where it all begins. Firstly, it enters the body through the bone marrow and distributes to every blood cell. WBCs typically develop in accordance with the body's requirements, they are produced inadequately and lose their functionality. They are frequently identified by their dark purple hue, but because of their diversity in texture and form, identification and additional processing become extremely problematic [3]. A group of cells known as leukocytes are very different from one another. WBCs can be identified by their size and form, but one difficult feature is that they are encircled by other blood elements.

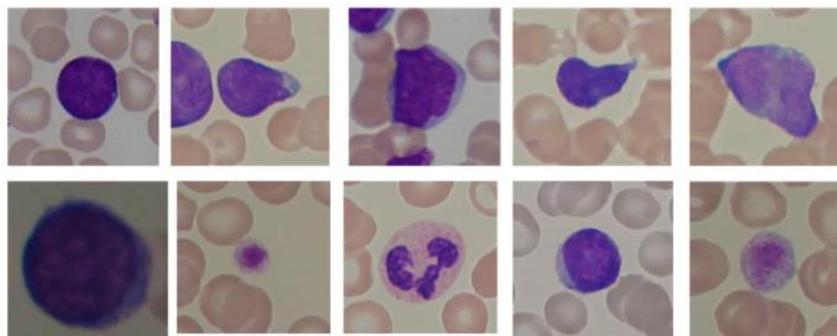


Figure 1. samples from the dataset showing benign (top) and malignant (bottom) lymphocytes.

As illustrated in Figure 1, lymphocytes possess a relatively regular shape, with smooth and regular edges around their nuclei. In contrast, lymphocytes diagnosed with Acute Lymphocytic Leukemia (ALL), also referred to as lymphoblasts, small cytoplasmic cavities identified as vacuoles, and their nuclei contain spherical components called nucleoli [4]. The disease's diagnosis worsens with the severity of the morphological abnormalities that have been reported. The ability to quickly and accurately identify the kind of leukemia is crucial in ensuring that the right treatment is given for that specific type. A complete blood count (CBC) is the first step in its discovery. A bone marrow biopsy is suggested for the patient if the count is abnormal. So studies are conducted to show the existence of leukemic cells utilizing morphological bone marrow and peripheral blood slide inspection [5]. To classify aberrant cells into their respective leukemia classes and subgroups, hematologists will employ light microscopy to analyze particular tissues while scanning for anomalies in the cytoplasm. This categorization is

applied for predicting the clinical symptoms of the disease, and the patient must be treated correctly. For unidentified reasons, the bone marrow generates an enormous quantity of aberrant WBC. Because pathology instruments are expensive, manual leukemia diagnosis is performed in pathology, requiring a lot of time [6]. For quick and accurate outcomes, an automated procedure is employed. This method involves processing a picture of a blood sample, segmenting the nucleus, and then categorizing the cells as either blast or normal.

The current state of work on WBC categorization can be largely divided into two groups: (a) deep learning techniques and (b) classical techniques. The classical techniques include strategies that suggest effective preprocessing procedures to identify significant characteristics in WBC pictures and categorize them with baseline classifiers. Here are some noteworthy studies in this field that are addressed [7]. a technique that minimizes similarity while choosing the eigenvectors from color pictures of blood cells. Next, the eigen cells are categorized utilizing the Bayesian classifier according to density and color data. Leukocyte cytoplasm and nucleus are separated utilizing fuzzy C-means clustering. Support vector machines (SVMs) extract and classify different geometric, color, and statistical features. A gradient vector flow-based approach to boundary removal and mean-shift clustering is presented for picture segmentation. The divided nucleus and cytoplasm are utilized for obtaining an ensemble of characteristics, that are subsequently categorized utilizing a random forest method. Different approaches executed worse than the multinomial logistic regression technique. Utilizing reversion values like partial F values, a stepwise linear discriminant evaluation technique is employed to extract particular features from blood pattern images and categorize them [8]. WBC cancer detection approach utilizing a random forest classifier in conjunction with multiple clustering, morphological, and image pre-processing processes. The recommended approach classifies different forms of cancer utilizing a decision tree learning technique that leverages predictors at every node to improve decision-making. To diagnose the type of leukemia disease, several advanced medical tests or examinations can be conducted including flow cytometry, immunophenotyping, and molecular probing. But all those examinations are expensive, time-consuming, and rely on the operator's fatigue and capacity. As a result, microscopic analysis of blood slides is recommended for diagnosing leukemia as a quick, affordable, and accurate diagnostic approach is constantly required. Several image processing tools are created to extract significant data from medical photos to enhance patient diagnosis. In this research work presents a Feed Forward Back Propagation Neural Network framework and lastly forecasts the kind of cancer found in cells.

The remaining study is organized as follows: Section 2 describes several techniques for utilizing deep learning (DL) and machine learning (ML) to identify WBC.

2. Literature Review

This section reviews the recent methods for the detection of classifying of white blood cancer utilizing DL and ML classifiers.

Ding et al [9] introduced a CNN framework and employed stacking, an ensemble technique, to improve the final prediction accuracy. This enabled them to place eighth in the initial round and tenth in the final round. For the preliminary and final phases, the weighted F1 scores assessed utilizing the approach were 0.8674 and 0.8552, respectively. It's a rather realistic outcome for a work that requires this level of ability. Sahlol et al [10] presented an enhanced hybrid strategy for effective WBC leukemia categorization. utilize VGGNet, a potent CNN framework that has been pre-trained on ImageNet, to extract features from WBC photos. Following feature extraction, an SESSA is employed to filter the data. This bio-inspired optimization approach eliminates noisy and strongly related characteristics while choosing the pertinent features. Utilizing two publicly available WBC Leukemia reference datasets, implemented the suggested method and were able to accomplish both high accuracy and decreased computing cost. Merely 1K out of 25K features obtained with VGGNet were selected by the SESSA optimization, which also increased accuracy. The outcomes surpass a number of convolutional network approaches and score amongst the best obtained on these datasets. It is anticipated that a multitude of additional image categorization work may use from the integration of CNN feature extraction and SESSA feature optimization.

Joshi et al [11] suggested WBC segmentation utilizes an automatic Otsu's threshold blood cell segmentation technique in conjunction with picture improvement. To distinguish between blast cells and regular lymphocyte cells, the kNN classifier was applied. The method is applied on 108 photos that are accessible in public picture datasets. With this procedure, accuracy is 93%. For the purpose of categorizing all subtypes of ALL, Sallam et al. [12] presented a new technique employing the k-means clustering method. First, the suggested method's stage is image preparation. The second phase that was performed is feature extraction, which yields the descriptive features of the images. The EGWO approach was utilized in the third step to determine the key features which may characterize the blood cell histology. Utilizing the k-means clustering method, EGWO will begin to choose the search agents according to specific criteria as the optimal cluster center. A number of supervised classifiers were tested examined, including RF, KNN, SVM, and NB. A high degree of accuracy of 99.22%, sensitivity of 99% and precision of 99%, is achieved by the suggested approach.

Vincent et al [13] proposed a unique method for dividing WBCs into normal and aberrant categories to identify leukemic illnesses, namely, AML, CML, ALL, and CLL. One of the most important steps in the classification scheme that is feature selection using PCA, which will be forwarded to concatenated neural network classifier system. The result of feature selection shows that among 13 features, the most important feature is the correlation between Hausdorff Dimension and Gray Level Co-occurrence Matrix features. Ahmad et al [14] presented a combination strategy for effective WBC subtype identification. Utilizing transfer learning on deep

neural networks that had previously trained, the best deep features are first taken from improved and segmented WBC pictures. ECMPA is employed to filter the feature vector that was serially fused. This meta-heuristic optimization technique, chooses the prominent traits and eliminates the weaker ones. A number of baseline classifiers with dissimilar kernel values are employed to classify the reduced feature vector. A publicly available collection of 5000 synthetic pictures representing five distinct WBC subtypes is employed for verifying the suggested approach. The system reduces the feature vector's size by more than 95% while maintaining an average accuracy of 99.9% overall. Additionally, the feature selection method outperforms traditional meta-heuristic methods in terms of convergence efficiency.

Kumar et al [15] introduced a DCNN then determined the kind of cancer that exists within the cells. The simulator precisely 94 times out of 100 managed to obtain the samples and duplicate all of the results. With an overall accuracy of 97.2%, it outperformed traditional ML techniques. According to this research, when tested on the retrieved dataset, the DCNN algorithm's efficiency is comparable to that of the well-known CNN designs with a lot fewer variables and processing time. The framework works well as a tool to identify the kind of bone marrow cancer.

3. Proposed Methodology

In this research presents a feed-forward backpropagation neural network structure, the kind of cancer found in the cells is eventually predicted. Initially, the filtration process is applied to the given images. Second, The FIS is a reliable process for detecting edges in color bone marrow microscopic pictures. It considers the stability degrees of color elements and is resilient to changes in illumination levels. Third Active contour detection provides an accurate curve of the boundary edge of the cells. And then the background separation is done by Pinched Flow Fractionation (PFF). Then the feature extraction is carried out by the Modified Honey Bee optimization algorithm (MHBO). Finally, the Feed Forward Back Propagation Neural Network (FFBNN) framework-based classifier is proposed for predicting the white blood cells. Figure 2. displays the procedure for the suggested white cell cancer classification.

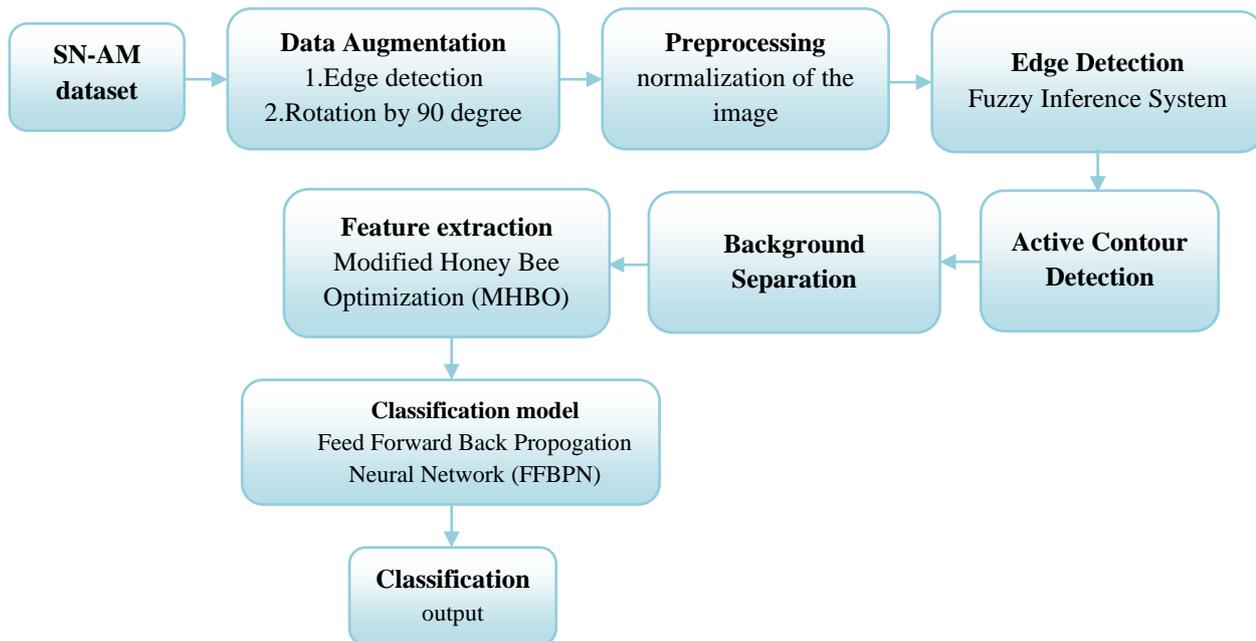


Figure 2. The overall process of the proposed methodology

3.1. Dataset Description

The SN-AM dataset applied in the suggested investigation was obtained from two distinct dataset collections [16]. The first section of the dataset comprises ninety photos of patients with B-ALL, or B-Lineage Acute Lymphoblastic Leukemia. The matching ALL image's nucleus mask and background mask. However, the second portion of the collection is comprised pictures of people who were diagnosed with multiple myeloma (MM), or 100 photographs total. The dataset contains images in BMP format with dimensions of 2560 x 1920 pixels. The integrated structure is utilized to train the suggested CNN approach, which distinguishes between MM and ALL cancer cell types.

3.2. Data Augmentation

The image is rotated and edges are extracted to start augmenting the SN-AM dataset. There are training and testing sets of scrambled photos. For the simulation to perform successfully during the evaluation phase, there ought to be an important quantity of data accessible because the object of interest needs to be present in a variety of sizes, positions, and lighting situations. Images can be altered using a variety of methods to create new data from preexisting ones. The first method is to rotate the photos by ninety degrees [17]. The photographs are all 90 degrees rotated since the model must be able to identify the object in any direction. The image is filtered in the following way, which yields an image that only includes the original image's edges and boundaries. Overfitting occurs in the absence of data augmentation, making it more challenging for the algorithm to be applied to new instances that weren't included in the training set.

3.3. Pre-Processing of Images

Handling null values, normalization, one-hot encoding, multi-collinearity, splitting, scaling the data, rearranging the data, and other operations are all included in pre-processing. The altered data acquired prior to feature selection in the planned research is first normalized, and after that, it is shuffled and split into training and testing sets. 25% of the dataset is utilized to test the framework, while the remaining 75% is utilized for training.

3.4. Edge Detection Using Fuzzy Inference System

This work describes a unique fuzzy logic reasoning strategy-based fuzzy image segmentation technique (FIS) for edge recognition in digital pictures without threshold value determination or training procedure requirements [18]. The suggested method starts by employing a floating 3x3 binary matrix to split the images into different sections. To find the edge, a direct fuzzy inference algorithm assigned a range of values that were different from one another in the floating matrix.

1) Fuzzy Image Processing

All methods that comprehend, indicate, and handle images, their segments, and their features as fuzzy sets are collectively denoted as fuzzy image processing. The issue needs to be solved and the fuzzy method utilized will determine what is represented and processed. The three primary phases of fuzzy image processing are image fuzzification, membership value modification, and, if required, image defuzzification. The reason for the fuzzification and defuzzification stages is that don't have fuzzy hardware. Thus, the processes that enable the processing of images utilizing fuzzy methods are the coding of image data and the decoding of the findings. The middle phase of fuzzy image processing is where its major strength is located.

The membership values are altered by suitable fuzzy procedures once the picture data have been fuzzy-analyzed and converted from the gray-level plane to the membership plane. This can take various forms, such as fuzzy integration, fuzzy clustering, fuzzy rule-based, and more.

2) Fuzzy Sets and Fuzzy Membership Functions

The system was implemented with the understanding that the output and input image that was acquired after defuzzification are 8-bit quantized, meaning that their gray levels are consistently within the range of 0 and 255. The intensities of every parameter were represented by fuzzy sets, which were linked to the linguistic parameters "Black," "Edge," and triangles for the input and output, accordingly The functions utilized for executing the "or" and "and" operations were the maximum and minimum functions. The fuzzy sets that were produced by utilizing each inference criterion to the input data were linked utilizing the add function, and the lom of the resulting membership function was applied to determine the system's output. This defuzzification process was

carried out utilizing the Mamdani approach. The output's three membership functions' values are intended to divide the image's black, white, and edge values.

3) Inference Rules Definitions

The inference criteria are based on whether the weights of the eight gray-level neighbors' pixels are more or less black or white. One of these rules' most potent features is its ability to immediately extract all of the edges from the processed image. In this investigation, every pixel in the processed image is being assayed by examining its neighbors' circumstances. Every pixel's status is determined by scanning all grays with a floating 3x3 mask. Here are some explanations of the intended regulations. The verified pixel is an edge when all of the grays in a line are represented by black and the balance of grays by white. The vertical and horizontal direction lines and gray level values around the verified pixel of the mask are the subject of the above four parameters. The center pixel denotes the edge if the weights of the next four pixels are levels of black and the weights of the subsequent four neighbors are degrees of white. According to the values of the gray level weights, the remaining four rules handle the eight neighbors in a similar manner. The edges, the white, and the black pixels are detected by the newly provided rules as well as another set of criteria. The black and white sections as well as the contours are contributed by the final photographs. The input grays range from 0-255 in terms of gray intensity from the fuzzy construction side. The required rules dictate when the gray level is transformed into membership function values. Following defuzzification, the FIS output is once more displayed with values between 0 and 255, at this point, the edges, black, and white are identified. Based on the evaluated photos in this research, it was determined that the best results were obtained in the black range, which is defined as zero to 80 gray values, and 80 to 255, which indicates that the weight is white.

3.5. Active Contour detection

"Active contouring" is the process of producing dynamic frameworks with forces and limitations in an image for the objective of segmentation. In order to create a contour, contour frameworks explain the edges of objects or other features of the image. Utilizing a variety of contour methods and imposed external and internal forces, the simulation's curvature is calculated. There is always a connection between the energy functional and the curve shown in the figure. The deformable variations are controlled by a combination of internal energy and forces generated by the image, which are meant to regulate the contour's placement on the image. The limitations for contour segmentation of a particular image are determined by the criteria. The necessary form is obtained by finding the minimum of the energy functional. A contour's deformation is described by a collection of points that locate the contour. The contour that fits the required image contour reduces the energy functional.

The gradient vector flow (GVF) [19] method is a variation of the snake or active contour approaches. The conventional snake approach has two disadvantages: Concave border contour convergence is not well achieved by

the snake curve flow whenever it is initiated far from the optimum. Moreover, the GVF approach utilizes the GVF field as an energy constraint to determine the contour flow.

The following procedures are employed to identify the GVF field. The first step is to determine the edge mapping function $f(x, y)$ from the image $I(x, y)$. The edge mapping function for binary images is given by Equation (1). However, $G_\sigma(x, y)$ is a 2D Gaussian function that has the standard deviation σ as a statistical variable.

$$f(x, y) = -G_\sigma(x, y) * I(x, y) \quad (1)$$

Equation (2) provides the edge map function for grayscale pictures, whereas the gradient operator is ∇

$$f(x, y) = -|\nabla[G_\sigma(x, y) * I(x, y)]|^2 \quad (2)$$

The equilibrium solution that lowers the functional energy is a gradient vector flow field. The variable μ determines the existence of two distinct terms in the functional energy: the data term and the smoothing term. The image's noise level determines the parameter value; a high noise level necessitates raising the parameter. The smoothing term that creates the contour's edge rounding is the primary source of issue or restriction with gradient vector flow. Consequently, an increase in μ values lessens edge rounding but somewhat compromises the contour's smoothing condition. The energy functional Eq. (3) defines the GVF.

$$\varepsilon = \iint \mu(u_x^2 + v_x^2 + v_y^2) + |\nabla f|^2 |g - \nabla f|^2 dx dy \quad (3)$$

The gradient vector flow obtained through the Euler equations is defined by the variable g in this equation. The Laplacian operator specified by these equations is complemented by two other equations, both (4) and (5).

$$\mu \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0 \quad (4)$$

$$\mu \nabla^2 v - (v - f_y)(f_x^2 + f_y^2) = 0 \quad (5)$$

Common gradient operators, including Prewitt, Sobel, or isotropic operators, are utilized to compute the equation's f_x and f_y values. The GVF field is characterized by these variables. The conventional snake algorithm's energy limitations are replaced with the GVF field $g(x, y)$ once it has been determined. Under these limitations, the iterative process of continuously computing the curve flow for the structural characterizing of the contour happens.

3.6. Background Separation by Pinched Flow Fractionation (PFF)

White blood cells (WBCs) and cancer cells are separated utilizing the pinch flow fractionation (PFF) background separation approach [20]. There are two inlets and three outlets on the PFF devices. The first outlet is for particles less than or equal to d_c , the second is for particles larger than or equal to d_c , and the third drains the

buffer solution. The channel geometry is utilized for calculating the d_c of a PFF device with three outlets. Because of mass conservation, the flow rate through the pinched segment has to exceed the total flow rate utilizing the outlet channels.

$$Q_{pinched} = Q_{small} + Q_{large} + Q_{drain} \quad (6)$$

The expression is made simpler by expressing the flow rates via the big particle exit and the drain on the basis of the tiny particle outlet.

$$Q_{pinched} = (1 + \alpha + \beta)Q_{small} \quad (7)$$

where α and β are the hydraulic resistance ratios within the corresponding other exit and the small particle outlet (R_s/R_l and R_s/R_d). It has been noted that this notion only applies to big channel aspect ratios. Microfluidic channels are commonly thought to have a constant flow profile, $H/W \gg 1$ and $W/H \gg 1$, and not for small aspect ratio cross sections like the pinched segment. Consequently, the velocity profile, $v_x(y, z)$, has to be considered. The pinched segment's velocity profile is integrated throughout the channel's height (z -direction) and width (y -direction) to determine the flow rates.

$$\int_0^H dz \int_{-w_p/2}^{w_p/2} dy v_x(y, z) = (1 + \alpha + \beta) \int_0^H dz \int_{-w_p/2}^{w_p/2} dy v_x(y, z) \quad (8)$$

Observe that the equation contains the critical diameter, d_c . Through numerical solution of the Navier-Stokes equation with no-slip boundary conditions at the wall, the velocity of rectangular channels was determined. Next, by substituting the velocity expression into equation (8) and finding α , the ideal hydraulic resistance ratio (α) was determined.

The required critical diameter is obtained from a variety of sets of dimensions, some more useful than others. Since the largest cell aggregates are anticipated to be about 20 μm in size, a 30 μm channel height was selected to prevent clogging. By allowing the outlet channels to continue straight from the pinched portion to the outlets, the lengths were determined. The channel lengths must be in the centimeter range because the injection molded chip has a 5 cm diameter. Consequently, equation (8) was able to be employed to optimize the channel widths. two devices were used: The first device, known as the non-adjusted PFF, contains a d_c that is used for separation when pressure is applied solely to the buffer inlet and sample. One way to modify the d_c of the second device is to exert pressure on the drain. It's known as adjustable PFF.

3.7. Feature Extraction using Modified Honey Bee Optimization (MHBO)

- **The basic principle of HBO**

The HBO method mimics the hive's queen mating routine. To start the mating flight, the queen takes off and dances away from the nest, after which the drones follow her and mate with her in the air. That indicates that some drone sperm will enter the queen's spermatheca and contribute to the colony's genetic pool [21]. The queen fertilizes eggs with the sperm that she has saved. A queen randomly selects a blend of the sperm kept in the spermatheca every time she lays fertilized eggs. Once the drones have mated with the queen once, they will perish. The queen mates repeatedly. One way to think of the mating flight is as a series of state-space transformations. The queen takes off with a certain amount of energy and speed at the beginning of her flight, and she returns to her nest when her spermatheca fills up or her energy drops below a certain threshold. The queen travels at a certain speed among the many states in space, mating probabilistically with the drones that follow her in each state. By employing an annealing function, a drone mates with the queen in a probabilistic manner.

$$Prob(Q, D) = e^{\frac{\Delta(f)}{Speed(t)}} \quad (9)$$

wherein $\Delta(f)$ is the absolute variance among D and Q's fitness, and $Prob(Q, D)$ is the likelihood of including drone D's sperm to queen Q's spermatheca. The queen's speed at time t is expressed as $speed(t)$. It is clear that there is a high chance of mating when the drone reaches the same level of fitness as the queen, or while the queen is just starting off on her mating flight. The drone's sperm is kept in the queen's spermatheca if the mating is effective. The following formulas are used to decrease the queen's energy and speed upon each change in space.

$$Speed(t + 1) = \alpha \times speed(t) \quad (10)$$

$$Energy(t + 1) = \alpha \times energy(t) \quad (11)$$

whereas $\alpha \in (0,1)$ implies that with each stride and transition, the queen's speed and energy will decrease. The queen utilizes the sperm kept in the spermatheca to start new broods through the crossover operator following the mating flight. The quantity of the new broods has greater research capacities since the genotype in the drones' sperm varies among the broods. Because their only responsibility in the colony is to care for the broods, the workers are hired to enhance the broods that the crossover operator produces. Each worker in the HBMO method is a heuristic that aids in the improvement of a group of broods. There are two portions to the quantity of workers ($w = w1 + w2$), There are $w1$ quantity of single local search techniques and $w2$ quantity of variations on these methods. Every newly generated brood will randomly select one worker to enhance them. The new brood, who are more fit than the queen, will take her place. If not, the brood joins the drones on the subsequent trip if its fitness surpasses that of a drone.

Given the foregoing description of the fundamental characteristics of honey bee mating, it is evident that the HBMO method is composed of three primary parts: 1. The queen's mating flight: in this part, the drones follow the queen as she begins her mating flight and mates with them. 2. Brood generation: Utilizing the sperm in its spermatheca, a queen produces several fresh broods in this step. 3. Brood enhancement: In this part, newly formed broods randomly select a worker to become better versions of themselves. This element has the potential to improve the method's exploring capabilities. In combinatorial optimization situations, the fundamental HBMO method works well and is generally better than the genetic and particle swarm optimization algorithms. It is limited to solving single-objective optimization issues. The study focuses on enhancing the fundamental HBO method in order to address the issue that it is challenging to maximize multiple objectives simultaneously with the fundamental HBO method.

The differentiation structure of the broods in the single-objective optimization issue is achieved by merely evaluating the fitness value. When solving multi-objective optimization issues, this approach is not appropriate. With the goal of resolving the multi-objective optimization issue in this study, a rapid non-dominated sorting approach is employed to build a broods differentiation technique.

Workers related to the local search process according to the aforementioned technique. The following are the specific steps:

Step 1: In the job shop scheduling structure, a maximum amount of worker iterations to enhance broods is set to LS_{max} , set $t = 1$.

Step 2: Choose a worker at arbitrarily from the potential worker bee w_i , $i \in \{1,2\}$. w_i is used to enhance brood τ , and new brood τ' will be acquired following the enhancement.

Step 3: If brood τ' dominates brood τ , τ is substituted by τ' .

Step 4: Set $t = t + 1$, if $t < LS_{max}$, go to step 2. If not, worker cultivation operations are stopped.

3.8. Feed Forward Back Propagation Neural Network (FFBPN)

The FFBPN method was employed to train a two-way iteration structure. First, input weights are computed in a forward step; second, updates to weights and errors are computed in a backward step [22]. The normalization of the training data was done inside the [0–1] range. Thirty percent of the data was split equally between testing and validation, with the remaining seventy percent being utilized to train the framework. Second, until the requirements were satisfied, the framework was trained to utilize Equation (12).

$$x_k = \sum_i^n w_{ki} x_i \quad (12)$$

where x_k is the variable's initial value, x_i is its updated value. and w_{ki} is the weight connection value of the neuron. Equation (13) shows that the activation function across the input and the hidden layer was "logsig."

$$f(x) = 1/(1 + e^{-x}) \quad (13)$$

Equation (13) connecting the hidden layer and the output layer utilized the purelin function:

$$f(x) = x \quad (14)$$

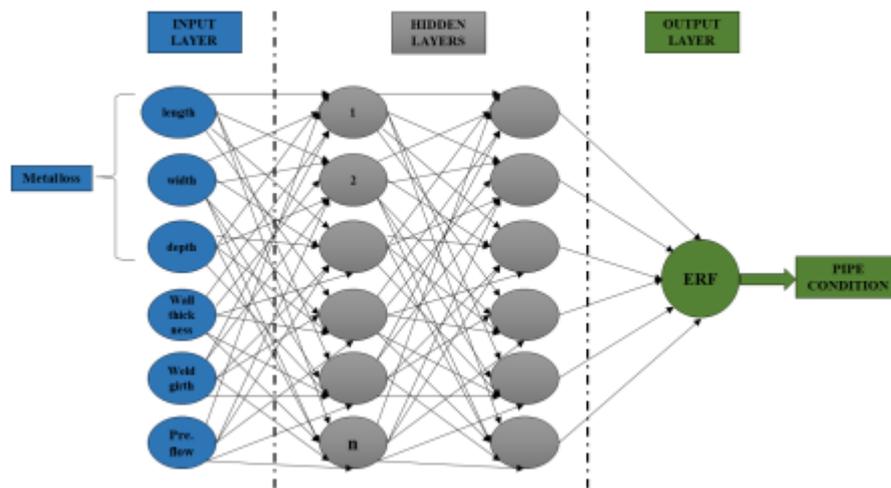


Figure.3. Structure of FFBPN

The FFBPN-based system is trained with historical data that is accessible in the first training phase. In the event that the simulation does not perform up to the expected level, the Levenberg-Marquardt (LM) backpropagation method will allow the procedure to be repeated until it finds the optimal requirement [20]. 15% of the data was utilized for validating the framework that was trained at the validation step after the algorithm training was finished. The framework was tested utilizing the remaining 15% of the data sets after accuracy was attained, and the findings showed that the algorithm was correct, with an R2 value that was almost 1.0. As seen in Figure 3, the network is made up of three layers: the input, hidden, and output layers. The linear function Purelin was utilized in the output layer of the suggested network, and the sigmoid function Tansig was employed in the hidden layer. After the input layer gets the signals from the other source, the hidden layer's job is to convert them into a form that the output signal can utilize. The suggested neural network structure features two functions: "tansig" and "purelin." Another name for this neural structure is a multi-layer perceptron. Any quantifiable function can be approximated utilizing standard multilayer perceptrons to the required accuracy rate.

For network training, the backward propagation technique was employed. A significant benefit of the multilayer perceptron is the ease with which the coefficients are modified through the application of the backpropagation method, a technique that has proven to be highly effective in real-world applications. This is

referred to as a supervised learning approach in the language employed for explaining neural networks because, in a learning phase, the network's output is compared with a known intended signal to provide an indicator of its performance. the network is performing. A type of steepest-descent technique is the backpropagation technique, where the error signal, which is the variation from the neural network's present output, and the intended output signal, is employed to modify the output layer's weights before being applied to the hidden layers' weights, continually returning the network's direction approach the inputs. Therefore, even though the neural network utilizes only feedforward to process the input signals and produce an output, throughout learning, the mistake is transmitted back from the output to the network's input, allowing the weights to be adjusted. There, the backpropagation technique composition will be demonstrated with a simplified network consisting of a single *linear* output neuron as the source of the single output signal and one hidden layer. The gradient descent technique is employed by the back-propagation neural network, a supervised learning technique, to reduce the error between the target output and the projected output. When the mean square error function achieves the end condition, repeat the computation above and make any necessary adjustments to the weights and thresholds. The adaptive approach to choosing the initial hidden layer node number depends on the mean square error value. Take note that there is more input layer nodes (m) here than output layer nodes (Q). Therefore, define the quantity of the first hidden layer nodes' numbers as $[(m + Q)/2, (2m + 1)]$ to sustain the neural network's convergence.

4. Results and Discussion

A comparison study was conducted in this part to demonstrate that ML performs effectively as data scales. The suggested ML technique is analyzed utilizing a variety of picture classification methods that ML provides. The suggested FFBN approach was evaluated with benchmark datasets to gauge its effectiveness. Additionally, these datasets are separated into training, test, and new sets, where various classifiers are learned and assessed. The suggested FFBN technique's classification accuracy as well as that of other classifiers for white cell cancer classification. The suggested FFBN-based learning approach and all current classifiers are tested in a static setting where the entire dataset is considered at once, and the accuracy is measured using the 10-fold cross-validation method. In addition to classification accuracy, the classifier is evaluated using the average findings for each classifier and the statistical metrics provided in equations (15)–(18). The ratio of accurately discovered positive results to all expected positive findings is termed as precision.

$$\text{Precision} = \text{TP}/\text{TP}+\text{FP} \quad (15)$$

The ratio of accurately detected positive results to all data is termed as sensitivity.

$$\text{Recall} = \text{TP}/\text{TP}+\text{FN} \quad (16)$$

The precision and recall weighted average is termed as F1 score. It requires false positives and false negatives as an outcome.

$$F1 \text{ Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (17)$$

Considering positive and negative values, accuracy is calculated:

$$\text{Accuracy} = (\text{TP} + \text{FP}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (18)$$

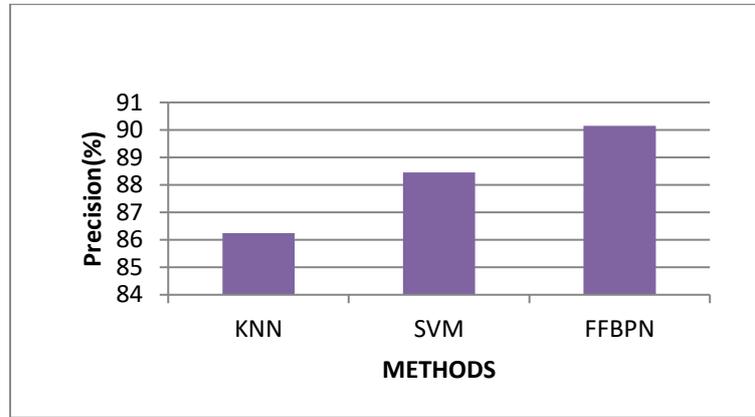


Fig.4. Precision comparison of the proposed FFBN

Figure 4 shows the results of a precision comparison among the suggested and current methods for identifying WBC data. When contrasted with the other ML methods, the FFBN simulation performed better overall on the datasets. The outcomes are in line with the error rate that was previously obtained and ascribed to the non-redundant rule sets that the suggested classification framework created. Considering the results, the FFBN strategy outperforms the current classification methods with respect to precision.

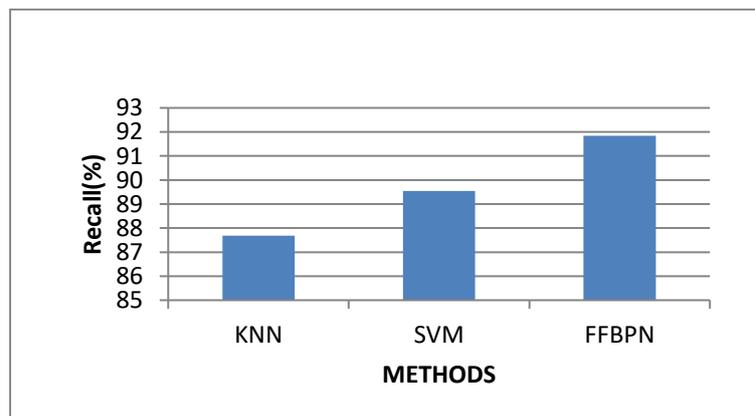


Fig.5. Recall comparison of the proposed FFBN

The memory comparison among the suggested and present approaches for classifying the WBC data is illustrated in Fig. 5. The primary focus of the data utilized is the diagnosis of people exhibiting WBC signs, taking into account a number of variables that typically impact the diagnosis outcome. As an outcome, it is thought of the prediction model as a categorization challenge that arises from having WBC or not. As an outcome, the suggested

supervised designs were used for the assigned task, and the outcomes were assessed and examined. To support the assessment results and the correctness of the simulations, the weak variables must be removed from the databases utilizing the feature selection method before utilizing these ML frameworks. For the purpose of diagnosing WBC, supervised classification frameworks and feature selection methods utilizing MHBO proved appropriate.

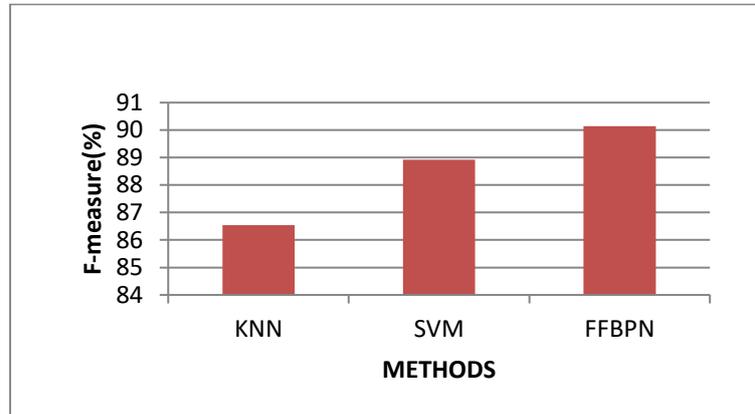


Fig.6. F-measure comparison of the proposed FFBN

The F-measure comparison among the suggested and present approaches for categorizing the WBC data is illustrated in Fig. 6. The parameter in the SN-AM database has the greatest correlation with the target class, as determined by the WBC test, according to both feature-selecting and categorization methods. When compared to other ML simulations, the one presented has the greatest accuracy rate measure in the database. Out of all the databases, the SN-AM database yields the best f-measure findings. From the graph, it is noted that the proposed FFBN model has high f-measure results than the existing methods.

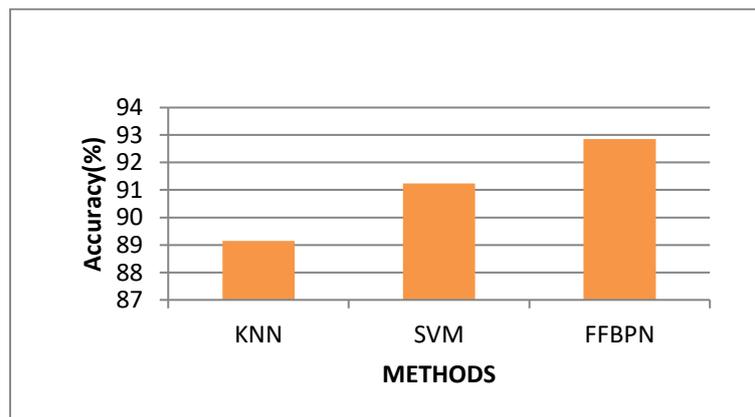


Fig.7. Accuracy comparison of the proposed FFBN

The accuracy comparison among the suggested and present approaches for classifying the WBC data is illustrated in Fig. 7. A supervised ML strategy is considered effective when it can accurately forecast the target and generalize predictions to new examples. Typically, accuracy, which has two subtypes: specificity and sensitivity

utilized to gauge the validity of an algorithm. The results indicate that the FFBN approach operates more accurately than the existing classification techniques.

5. Conclusion

Finding the most effective ML algorithm to diagnose people with certain symptoms of white blood cell cancer was the main objective. A number of procedures were required to choose the optimal ML strategy. Selecting the effective WBC questionnaire diagnosis technique and compiling a superior database for every age group was crucial. In this research, a feed forward-back propagation neural network framework is presented, and the kind of cancer found in the cells is ultimately predicted. Initially, the filtration process is applied to the given images. Second, Color bone marrow microscopic pictures can be edge detected utilizing the FIS technique, which is resistant to varying light levels and considers the stability levels of color elements. Third Active contour detection provides an accurate curve of the boundary edge of the cells. And then the background separation is done by Pinched Flow Fractionation (PFF). Then the feature extraction is carried out by a Modified Honey bee optimization algorithm (MHBO). And finally, the Feed Forward Back Propagation Neural Network (FFBNN) framework-based classifier is suggested for predicting the white blood cells. The highest performing classifier was found by analyzing AN-SM datasets in light of recall, F-measures, precision, and classification errors, as per the findings of the study. The findings show that, for the SN-AM dataset, the suggested FFBN approach yields an accuracy score of 99%. In the future, the deep learning model-based classification model will be incorporated into the WBC classification.

REFERENCES

1. Ullah, A., & Shoaib, M. (2022). Normal Versus Malignant Cell Classification in B-allwhite Blood Cancer Microscopic Images Using Deep Learning. *American Journal of Computer Science and Technology*, 5(3), 190-197.
2. Wang, Y. L., Ge, X. X., Wang, Y., Xu, M. D., Gong, F. R., Tao, M., ... & Li, W. (2018). The values of applying classification and counts of white blood cells to the prognostic evaluation of resectable gastric cancers. *BMC gastroenterology*, 18, 1-12.
3. Su, M. C., Cheng, C. Y., & Wang, P. C. (2014). A neural-network-based approach to white blood cell classification. *The scientific world journal*, 2014.
4. Anilkumar, K. K., Manoj, V. J., & Sagi, T. M. (2021). Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison. *Medical Engineering & Physics*, 98, 8-19.

5. Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2018). Classification of acute lymphoblastic leukemia using deep learning. *Microscopy Research and Technique*, 81(11), 1310-1317.
6. Kutlu, H., Avci, E., & Özyurt, F. (2020). White blood cells detection and classification based on regional convolutional neural networks. *Medical hypotheses*, 135, 109472.
7. Vogado, L., Veras, R., Aires, K., Araújo, F., Silva, R., Ponti, M., & Tavares, J. M. R. (2021). Diagnosis of leukaemia in blood slides based on a fine-tuned and highly generalisable deep learning model. *Sensors*, 21(9), 2989.
8. Jiang, M., Cheng, L., Qin, F., Du, L., & Zhang, M. (2018). White blood cells classification with deep convolutional neural networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(09), 1857006.
9. Ding, Y., Yang, Y., & Cui, Y. (2019). Deep learning for classifying of white blood cancer. In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging: Select Proceedings* (pp. 33-41). Singapore: Springer Singapore.
10. Sahlol, A. T., Kollmannsberger, P., & Ewees, A. A. (2020). Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. *Scientific reports*, 10(1), 2536.
11. Joshi, M. D., Karode, A. H., & Suralkar, S. R. (2013). White blood cells segmentation and classification to detect acute leukemia. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 2(3), 147-151.
12. Sallam, N. M., Saleh, A. I., Ali, H. A., & Abdelsalam, M. M. (2023). An efficient EGWO algorithm as feature selection for B-ALL diagnoses and its subtypes classification using peripheral blood smear images. *Alexandria Engineering Journal*, 68, 39-66.
13. Vincent, I., Shin, B. K., Kwon, S. G., Lee, S. H., & Kwon, K. R. (2014, July). Feature Selection using Principal Component Analysis for Leukemia Classification. In *Proceeding of the 10th International Conference on Multimedia Information Technology and Applications* (pp. 206-207).
14. Ahmad, R., Awais, M., Kausar, N., & Akram, T. (2023). White Blood Cells Classification Using Entropy-Controlled Deep Features Optimization. *Diagnostics*, 13(3), 352.
15. Kumar, D., Jain, N., Khurana, A., Mittal, S., Satapathy, S. C., Senkerik, R., & Hemanth, J. D. (2020). Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks. *IEEE Access*, 8, 142521-142531.
16. Anudba Gupta and Gupta Ritu. Sn-am dataset: White blood cancer dataset of b-all and mm for stain normalization [data set]. <https://doi.org/10.7937/tcia.2019.of2w8l1r>.
17. Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999, 2016.

18. Sabri, N., Aljunid, S. A., Salim, M. S., Badlishah, R. B., Kamaruddin, R., & Malek, M. A. (2013). Fuzzy inference system: Short review and design. *Int. Rev. Autom. Control*, 6(4), 441-449.
19. Xu, C., & Prince, J. L. (2020). Gradient vector flow. *Computer Vision: A Reference Guide*, 1-8.
20. Vig, A. L., & Kristensen, A. (2008). Separation enhancement in pinched flow fractionation. *Applied Physics Letters*, 93(20).
21. Yuce, B., Packianather, M. S., Mastrocinque, E., Pham, D. T., & Lambiase, A. (2013). Honey bees inspired optimization method: the bees algorithm. *Insects*, 4(4), 646-662.
22. Yu, X., Efe, M. O., & Kaynak, O. (2002). A general backpropagation algorithm for feedforward neural networks learning. *IEEE transactions on neural networks*, 13(1), 251-254.