Cite this article: Sheeba, C. M., & Sarojini, K. (2023, May). Optimal feature selection and detection of sickle cell anemia detection using enhanced whale optimization with clustering based boosted C5.0 algorithm for tribes of Nilgiris. *Journal of Current Science and Technology, 13*(2), 205-220.
https://doi.org/10.59796/jcst.V13N2.2023.1737

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Optimal feature selection and detection of sickle cell anemia detection using enhanced whale optimization with clustering based boosted C5.0 algorithm for tribes of Nilgiris

C. Maria Sheeba and K. Sarojini*

Department of Computer Science, L.R.G Government Arts College for Women, Tripur,
Tamil Nadu 641604, India

*Corresponding author; E-mail: ksarojini1467@gmail.com

Received 23 May 2022; Revised 6 October 2022; Accepted 5 November 2022;
Published online 15 July 2023

Abstract

Degradation of red blood cells (RBC) causes many diseases, like sickle cell anemia. Diagnosing this disease takes more time because peripheral blood samples must be examined under the microscope. Since the isolated RBC observation is subjective and high error rate leads to reduce the accuracy, the technology is needed to perform this approach. To fulfil these requirements, optimal feature selection and detection of sickle cell anemia using enhanced whale optimization with clustering based boosted C5.0 algorithm in tribes of Nilgiris is proposed in this manuscript. The input dataset is taken from real data set via non-government organization named NAWA (situated in Kotagiri). The reason behind of Nilgiri region is considered in this work is a collection of sickle cell anemia test results of the tribe people who lives in different areas of Nilgiri region. These images are pre-processed using wavelet packet transform cochlear filter bank method to eradicate the noises and to improve the superiority of image. After that, the features are extracted using force-invariant improved feature extraction method. To select the optimal features, enhanced whale optimization (EWO) algorithm is used. These optimal features are classified utilizing clustering based boosted C5.0 algorithm. The proposed method is activated on PYTHON. The proposed method shows 52.32%, 43.78%, 32.78% and 45.90% higher accuracy compared with the existing methods, such as RGSA-MLP-SCA, CRFA-SVM-SCA, AO-LSTM-SCA and BOA-CNN-SCA.

Keywords: *classification; clustering based boosted C5.0 algorithm; enhanced whale optimization; features selection; sickle cell anemia; sickle cell disease.*

1. Introduction

Sickle cell anemia is a genetic disorder, it is categorized by a change in the shape of red blood cells (RBCs) from soft donut shape or half-moon shape to the crescent (Carden, Fasano, & Meier, 2020; Rajesh, Shajin, Rajani, & Sharma, 2022). Degenerated cells lack plasticity, block small blood vessels, impairing blood flow, which reduces the survival of red blood cells, finally it causes anemia (Wonkam et al., 2020; Shajin, Rajesh, & Thilaha, 2020). The World Health Organization has designated this condition as a health priority. Without proper care, the lower oxygen levels of blood and blockages of blood vessel can cause persistent pain, severe anaemia, bacterial and recurrent infections, necrosis, clinical signs of thrombosis in internal organs, central nervous system disease, episodes of neurophysiological disease even patient's mortality

(Reeves et al., 2022; Rajesh, Shajin, Chandra, & Kommula, 2021).

A few details concerning sickle cell disease is given below (Delgado-Font et al., 2020):

– Around 5% of people worldwide have trait genes for haemoglobin diseases, primarily thalassemia and sickle cell disease.

– Some locations have up to 25% of the population carrying these genes.

– In Africa, sickle cell disease is prevalent.

– Each year, more than 300,000 infants are born with severe haemoglobin abnormalities.

Therefore, it is crucial to monitor these patients. It involves observing peripheral blood samples utilizing microscope, a time-consume process (Onalo et al., 2021; Shajin, Rajesh, & Raja, 2022). Additionally, a professional must do this method, and its error rate is considerable due to the subjective character of the observation. These problems are exacerbated when the count is high. A technique to assess a patient's clinical condition is to list the various RBC kinds according to its morphology (Olupot-Olupot et al., 2022). They are normal (discocyte), distorted elongated (sickle cells). This quantitative assessment has a number of issues, from expert disagreements to the challenge of considering stable estimation (Domingos et al., 2020). The process of extracting the features is critical in morphological examination of digital image configurations (Hindawi et al., 2020). To examine the degradation in erythrocytes present in peripheral blood samples, various models are designed, each has own properties and reported performance results.

People who live with sickle cell disease (SCD) suffer from a genetic hemoglobin structural abnormality for which there is no known cure. Cord blood or allogeneic bone marrow transplantations have been successful in curing some infants with sickle cell anemia (SCA) (Fields et al., 2022). SCA has 3 components: (i) pain syndrome, (ii) anemia and its repercussions, and (iii) organs failure. At least one sickle cell trait is passed down from one parent in the SCD family of hemoglobinopathies, all of which have a clinical impact (Soltani et al., 2020). SCA or sickle cell anemia occurs when the sickle gene is passed down from one parent to other.

SCA can also be caused by the sickle gene being passed down in combination with the Hemoglobin C (Hb C) gene, resulting in Hemoglobin sickle cell (Hb SC) disease, or β-thalassemia gene (β^0 or β^+), resulting in sickle -β^+-thalassemia or sickle - β^0-thalassaemia, accordingly some β-globin structural variants causing different combinations like HbSO Arab, and HbSD disease (Diop, & Pirenne, 2021). Homozygous HbSS along compound heterozygous disorders HbSβ^0-thalassemia, HbSβ^+- thalassemia, HbSC diseases are the most common kinds of SCD in the population. HbSS and HbSβ^0-thalassemia are termed as SS or SCA because of their clinical resemblance, and these genotypes are linked to the most severe symptoms. SCD problems that fall into one of several categories may be more common than others (Verlhac et al., 2021). As a result, SCA is more likely than other types of SCD to cause repeated painful conditions, severe anemia requiring blood transfusions, and acute chest syndrome (ACS).

Patients with SCD have defective hemoglobin in their red blood cells, known as sickle hemoglobin or hemoglobin S. The oxygen-carrying protein hemoglobin is discovered in RBCs (Gour, Dogra, Bhatt, & Nandi, 2020). People with sickle cell disease are born with two defective hemoglobin genes: one from each parent. Infections that can cause sickle cell disease include:

- Hemoglobin Sβ^0 thalassemia
- Hemoglobin Sβ^+ thalassemia
- Hemoglobin SC
- Hemoglobin SD
- Hemoglobin SE
- Hemoglobin SS

As with all forms of SCD, a person's body produces hemoglobin S due to at least one faulty gene. Anemia known as SCD occurs when a person has two copies of the hemoglobin S gene. The most prevalent and most severe kind of SCD is this one. Both hemoglobin SC and hemoglobin S thalassemia are frequent forms of SCD.

Red blood cells are a crucial component of the blood molecule, which is made up of many cell types. On the other hand, Anemia is a disorder in which there was a deficiency of healthy RBCs, resulting in exhaustion. Acute and chronic anemia is the two most common types of anemia. In acute anemia, the symptoms appear quickly, whereas in chronic anemia, the symptoms slowly appear. Pregnancy-related water weight gain, iron deficiency, decreased production of erythropoietin, and blood loss during menstruation is all factors that contribute to the development of anemia (Abdullahi et al., 2020). Drepanocytos or SCA is a condition caused by the presence of sickle-shaped RBCs in the bloodstream. The hemoglobin gene undergoes mutation, resulting in the generation of these cells. SCA is a hereditary disorder because it is caused by parents who have two defective genes that they pass on to their children.

SCD can be diagnosed at an earlier stage, allowing doctors and patients to take advantage of treatment options such antibiotics, blood transfusions or painkillers. Counting millions of red blood cells in a single spell necessitates a time-consuming and error-prone manual assessment, diagnosis, and cell count. SCD may be accurately detected in the human body using data mining techniques as a classifier algorithm (Wang et al., 2021; Risoluti et al., 2020). The intention of this work is to classify and predict SCA using a data mining technique with feature selection to better understand the disease. The proposed method comprises data preparation, data analysis, training and testing, feature selection, and implementation of the classification algorithm. Figure 1 shows the sickle cell anemia.

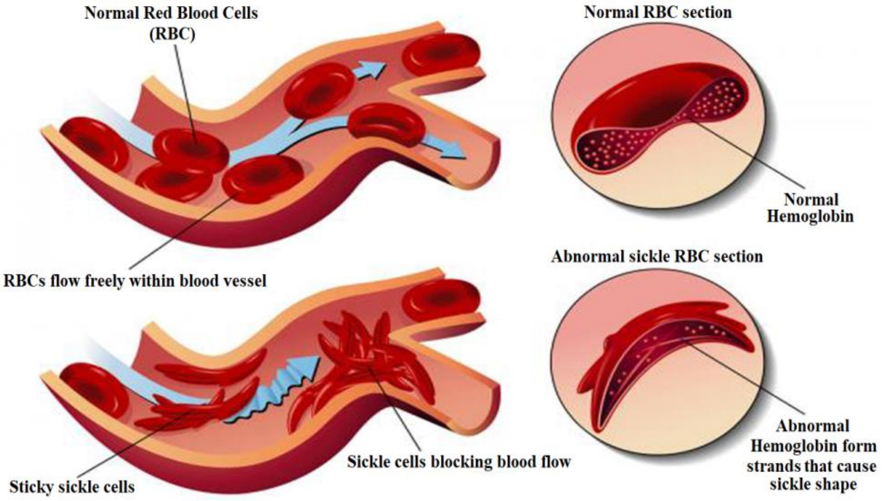

Normal Red Blood Cells (RBC)
Normal RBC section
Normal Hemoglobin
RBCs flow freely within blood vessel
Abnormal sickle RBC section
Abnormal Hemoglobin form strands that cause sickle shape
Sickle cells blocking blood flow
Sticky sickle cells

Figure 1 Sickle cell anemia

Degradation of RBC causes many diseases, like sickle cell anemia, it leads to pain and severe anemia. Since peripheral blood samples must be examined under a microscope, a lot of time to monitor patients with these disorders. The technology is needed to perform this approach due to isolated RBC observation is subjective and high error rate, which reduces the accuracy. To overcome these issues, an Optimal Feature Selection and detection of Sickle Cell Anemia Detection using Enhanced Whale Optimization with Clustering based Boosted C5.0 Algorithm in Tribes of Nilgiris is proposed in this manuscript. The novelty of this manuscript is "to optimally identify the SCA with high accuracy by lessening error rate and difficulty".

1.1 Related works

Most of the SCA prediction related to medical data or medical image classification. The medical data-based classification is considered for the prediction of SCS. Mostly, data mining-based techniques and machine learning based techniques implemented based on medical data classification are analyzed here.

Petrović, Moyà-Alcover, Jaume-i-Capó, & González-Hidalgo (2020) have presented a sickle-cell disease detection assist appropriate machine learning method selection. The pre-processed together with segmented microscopic imageries were assumed to assure higher quality of feature. Then employed the models for extracting features through blood cells, also machine learning categorized its morphology. The best parameters were searched from the resulting data at feature extraction. It provides higher accuracy with lower F-score.

Singh, Ojha, Bhoi, Bissoyi, and Patra (2021) have presented a Prediction of hydroxy urea effect on sickle cell anemia patients utilizing machine learning method. The chemotherapeutic drug Hydroxy urea was considered by concerned patient to re-mediate medical issues. The purpose of machine learning approach was to predict the effects of the drug Hydroxy urea treatment in different sickle cell anemic patients. Support vector machine classified different data division methods. It attains better precision with lesser sensitivity.

Delgado-Font et al. (2020) have suggested sickle cell anemia based disease detection by classifying RBC shape in peripheral blood imageries. By utilizing Chan-Vese active contour mode, the objects of interest were segmented. The analysis was acted to categorize the RBCs named called erythrocytes with the help of circular and elliptical shape factor. The images of patient blood samples were gathered from hospital at Santiago de Cuba. It provides lower computational time with higher error rate.

Balamanikandan, and Bharathi (2022) have presented mathematical modelling for identifying SCA utilizing Quantum graph theory with Aquila optimization classifier. The aim was to extract elasticity attributes and separate as normal or sickle cell anemia in RBC. The elasticity attributes were extracted through Quantum graph theory, where the spanning tree formation was done by graph structure along Hemoglobin quantization. The extracted features were enhanced by Aquila optimization and categorized by cascaded Long Short-Term memory to reach classified result of sickle cell, normal cell. It provides greater accuracy with lower specificity.

Dong and McGann (2021) have presented modifying the medical paradigm of hydroxyurea treatment for SCA through precision medicine. The laboratory with sparse sampling required 10 μL blood collected 15, 60, 180 minutes after test dosing. Bayesian adaptive control assess hydroxyurea exposure resulting HbF responses >30–40%, levels beyond what was attained with typical weight-base and error dosing escalation. The hydroxyurea dosing has changed the treatment paradigm of hydroxyurea therapy. It provides higher sensitivity with lower F-score.

Zhang, Li, Xu, and Li (2020) have presented an Automatic semantic segmentation of RBCs for sickle cell disease. U-Net architecture offer exact localization for image semantic segmentation. The deformable convolution acts free-form feature learning process deformation, thus making the network proficient for certain cell morphologies along image settings. Then, dU-Net was tested under microscopic RBC imageries from patients with sickle cell disorder. It provides higher complexity with lower accuracy.

da Silva, Rodrigues, & Mari (2020) have presented an Optimize data augmentation policies for convolutional neural networks depending upon sickle cells categorization. The data augmentation was dependent upon hyperparameter optimization for scaling better policies via Bayesian optimization. Convolutional Neural Networks (CNN) architecture was considered to categorize sickle cell images. It provides lower error rate with high computational time.

Medical detection of SCA is a major issue in the biomedical field. There are several methods are explained in Literature survey (Petrović, Moyà-Alcover, Jaume-i-Capó, & González-Hidalgo et al., 2020; Singh, Ojha, Bhoi, Bissoyi, & Patra, 2021; Delgado-Font et al., 2020; Balamanikandan, & Bharathi, 2022; Dong, & McGann, 2021; Zhang, Li, Xu, & Li, 2020; da Silva, Rodrigues, & Mari, 2020) to upgrade the classification models performance for breast cancer detection. The existing Random Forest, K-Nearest Neighbors (kNN), Naïve Bayes does not give adequate accuracy and the error rate was maximized. However, there are still numerous issues that must be correctly resolved, like biomedical datasets, which result from the unequal distribution of anemia disease. To tackle this problem, EWOA-CBC5.OA-SCA method is proposed. EWOA-CBC5.OA-SCA has some benefits, like it produces newly samples with identical distribution to original samples, then the produced samples can expand the dataset, thus dealing the detection complexity in huge samples. The ability of data augmentation strategy to adapt the distribution of training has been significantly improving classifier performance on a count of datasets. The proposed EWOA-CBC5.OA-SCA method achieves high accuracy for predicting the SCA.

Various breast cancer categorization models specified in literature survey, but those models have certain limitations, such as lower identification rate SCA samples examined with higher fraction of Normal samples, during classification procedure the SCA accuracy is categorized and decrease the Normal accuracy, some methods not classify the Normal and thalassemia exactly, but gives the accuracy of the normal samples as well as increased the error rate, some methods displays accuracy of the image, but computational complex is raised. The proposed method handles all these complexities and attains more accuracy.

2. Objectives

The main contributions of this manuscript are summarized below

- Optimal feature selection and detection of sickle cell anemia using enhanced whale optimization with clustering based boosted C5.0 algorithm in tribes of Nilgiris is proposed.
- The input dataset is taken from real data set via a nongovernment Organization named NAWA (situated in Kotagiri). Nilgiris is considered in this work is a collection of sickle cell anemia test results of the tribe peoples who lives in the different areas of Nilgiris region.
- Then these images are pre-processed using wavelet packet transform cochlear filter bank method (Hamsa, Shahin, Iraqi, & Werghi, 2020) to eradicate the noises and enhance the superiority of the image.
- Afterward, the features are extracted utilizing force-invariant improved feature extraction method (Islam et al., 2021). It extracts the robust features effectually from input dataset.
- Then to select the optimal features of enhanced whale optimization (EWO) algorithm (Valayapalayam Kittusamy, Elhoseny, & Kathiresan, 2022) is used.
- These optimal features are classified by clustering based boosted C5.0 algorithm (Zhang et al., 2021).
- The proposed model is implemented on PYTHON tool.
- The performance metrics, such as accuracy, sensitivity, PPV, specificity, NPV, MCC, error rate, computational time, complexity is examined.
- The obtained results are compared with existing methods, such as Sickle-cell disease detection assist selecting the proper machine learning model (RGSA-MLP-SCA) (Petrović, Moyà-Alcover, Jaume-i-Capó, & González-Hidalgo, 2020), Prediction of hydroxyurea effect on sickle cell anemia patients utilizing machine learning model (CRFA-SVM-SCA) (Singh, Ojha, Bhoi, Bissoyi, & Patra, 2021), Mathematical modelling to identify sickle cell anemia utilizing Quantum graph theory with Aquila optimization classifier (AO-LSTM-SCA) (Balamanikandan, & Bharathi, 2022) and Optimizing data augmentation policies for convolutional neural networks depending on sickle cells categorizations (BOA-CNN-SCA) (da Silva, Rodrigues, & Mari, 2020) methods.

The leftover manuscript is structured as: Segment 2 deliberated the related works. Segment 3 explains the proposed method. Segment 4 exemplifies the results with discussion. The conclusion is presented in Segment 5.

3. Methodology

3.1 Proposed model for detecting SCA

In the proposed model, the sickle cell anemia disease is to classify based on the real-time data. This real-time data was collected from tribe peoples of different regions around Nilgiris district. The proposed model is a hybrid data mining approach, which includes a feature selection and classification techniques for disease prediction. The dataset contains 14 number of features related to the SCA. For features selection, the Enhanced Whale Optimization algorithm is used. For classifying the data with selected features, the clustering based boosted C5.0 algorithm is used. Finally, based on the results, the performance was compared with the existing models for validation. Figure 2 shows the proposed framework for SCA detection.

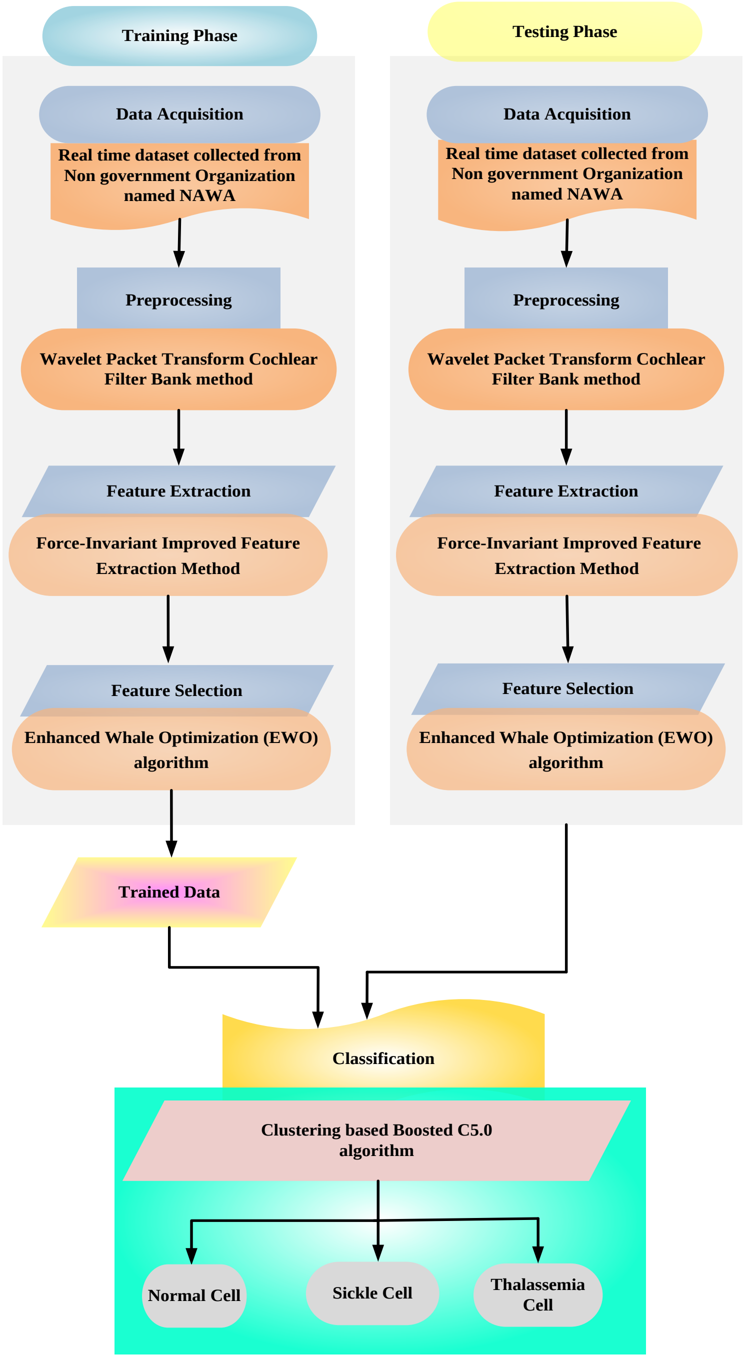

Training Phase
Testing Phase
Data Acquisition
Real time dataset collected from Non government Organization named NAWA
Data Acquisition
Real time dataset collected from Non government Organization named NAWA
Preprocessing
Wavelet Packet Transform Cochlear Filter Bank method
Preprocessing
Wavelet Packet Transform Cochlear Filter Bank method
Feature Extraction
Force-Invariant Improved Feature Extraction Method
Feature Extraction
Force-Invariant Improved Feature Extraction Method
Feature Selection
Enhanced Whale Optimization (EWO) algorithm
Feature Selection
Enhanced Whale Optimization (EWO) algorithm
Trained Data
Classification
Clustering based Boosted C5.0 algorithm
Normal Cell
Sickle Cell
Thalassemia Cell

Figure 2 Proposed framework for SCA detection

3.1.1 Pre-processing using Wavelet Packet Transform Cochlear Filter Bank

The pre-processing is done to eradicate the noise as well as assure the images superiority to detect the sickle cell anemia diseases from the tribal's of the Nilgiris. To remove unwanted noises from the microscopic cell images, Wavelet Packet Transform Cochlear Filter Bank is used. It has the array of wavelet filters to distort images into sub-bands over dissimilar regions of frequency spectrum, without losing the time domain characterization as enabled by Fourier transform, which is helpful for image processing. In this, the input microscopic cell images are decomposed in the low pass filter bank and the high pass filter bank using every level. After that, the samples are decomposed into the Wavelet Packet Transform (WPT) Cochlear Filter Bank, input microscopic cell images denotes $y(l)$, decomposed sample represents y_l and it consists of modulating samples m_l and the carrier samples c_l for determining the transforms. Then the pre-processing for input samples is given in equation (1)

$$SCA_{\text{Pre}-processing} = F\left[\underset{l}{WPT}[y(l)]\right] \tag{1}$$

where F is represented as the envelop detector of the WPT for removing impurities from the blood to detect the SCA. Then the WPT is combined with the Cochlear Filter Bank and the discrete modulation with input samples $y(l)$ is given in equation (2),

$$SCA_{\text{Pre}-processing} = DFT\left[F\left[\underset{l}{WPT}[y(l)]\right]\right] = \sum_{m=0}^{L-1} N(m,l)e^{\frac{(-2\pi mj)}{L}}, j = 0,1,....L-1 \tag{2}$$

where L is represented as the length of the Fourier transform, DFT is represented as the discrete Fourier transform. By this process, image noises are removed and the quality of image is improved.

3.1.2 Feature Extraction using Force-Invariant Improved Feature Extraction Method

The pre-processed images are fed to feature extraction stage to extract the image features to accurately classify the sickle cell anemia disease. *Force-Invariant* Improved Feature Extraction Method is a computer vision algorithm to detect, describe, and match local *features* in images. Here, the image features are extracted based on the shape or geometry, color, and texture features using *Force-Invariant* Improved Feature Extraction Method. The input image is represented as $y(jT), j = 0,1,2,3,...M-1$ and T denotes count of samples $T = \frac{1}{fs}$ and the feature extraction is given in equation (3).

$$\sum_{j=0}^{M-1} y(j)^2 = \frac{1}{M}\sum_{l=0}^{M1} Y[l]Y^*[l] = \sum_{l=0}^{M1} L[l] \tag{3}$$

where $Y^*[l]$ is represented as the conjugate of $Y[l]$ and $L[l]$ with the input image features. Using this equation 3, features are extracted, viz shape or geometry, color, texture features. The shape gives cell measurement depends on its geometrical aspects and it is defined as the Ratio of squared perimeter C to object area O. If the value is huge, the convexity shape shift away, also it contains holes and it is given in equation (4)

$$\frac{|C|^2}{|O|} \tag{4}$$

Color feature is defined as the Statistical measures with respect to color spaces for cells feature extraction. It tries to separate different cells by color. In which, RGB, HSV, CIEL*a*b* color spaces are used for classification tasks. Consider I implies color space channel, N implies number of pixels in the region of interest O (every identified cell) at channel I, $I(P)$ implies pixel P intensity value, channel color mean μ with its σ standard deviation is computed using equation (5),

$$\mu_{I,O} = \frac{1}{N}\sum_{P\in O} I(P), \sigma_{I,O} = \sqrt{\frac{1}{N-1}\sum_{P\in O}(I(P)-\mu)^2} \tag{5}$$

Texture feature is used to measure the degree of asymmetry and it is specified in equation (6),

$$Texture = \frac{1}{N}\sum_{P\in O} I(P), \sigma_{I,O} = \frac{1}{N}\sum_{j\in 1}\left(\frac{x_j-\mu}{\sigma}\right)^3 \tag{6}$$

From the total features, some relevant features are selected to detect the SCA with high accuracy. Here μ is selected as the optimal parameter for selecting important features from the input dataset to detect the SCA.

3.1.3 Feature selection using Enhanced Whale Optimization (EWO) algorithm

Enhanced Whale Optimization (EWO) algorithm is utilized for selecting optimal features from the input dataset (given in Table 1) to detect the sickle cell anemia disease in early stage and to reduce the death rate of tribal’s in Nilgiris. EWO algorithm is used for selecting optimal features from the input dataset, the original dataset contains many features that increases the overfitting problem and the time complexity, to reduce this issues EWO algorithm. EWOA algorithm is a metaheuristic approach that selects the optimal parameters using exploration and exploitation phases. The EWO algorithm is based on bubble-net hunting approach using humpback whales. To hunt a group of fishes, they prefer to stay near to the surface. As a result, humpback whales move around their prey in the shrinking circles while still following the spiral structured paths, resulting in characteristic bubble patterns along the way or “9” shaped path. Whales' positions can be updated by either the shrinking encircling approach or the spiral model, which has 50% chance of being selected during optimization. In the following manner, their formulations are defined:

Figure 3 portrays the flow chart of whale optimization algorithm.

Step 1: Initialization

Initialize the initial parameters of the Whales for selecting the optimal features from the input dataset and it is represented as F attributes and K number of instances and the primary number of attributes are represented as FET. The initial population is initialized based on whale h_j, where $j = 1,2,....n$ is given in equation (7)

$$h = Y.h_j(x) + F.(h_z(x) - h_j(x)) \qquad (7)$$

Let Y specifies positioning of whale for selecting the optimal features, $h_z(x)$ and $h_j(x)$ indicates search area’s lesser and upper limit on j^{th} dimension, K indicates uniform random count at [0, 1] range.

Step 2: Random generation

After initialization, randomly generate the exploration and exploitation phases of the EWO for attaining best solution.

Step 3: Fitness Function

Fitness function is calculated for optimally select the important features from the input dataset for detecting the sickle cell anemia in early stage and to reduce the death rate using the parameter μ and it is given in equation (8),

$$Fitness\ Function = Max(\mu), \mu = \ for\ selectingoptimal\ features \qquad (8)$$

Step 4: Update the exploration and exploitation phases of whale for selecting optimal features.

The exploration and exploitation phases are used to optimal features from the input dataset. Feature selection is given in equation (9),

$$(x_{j,l}(z+1) = z_{j.l}(s) + \beta \oplus Levy(\mu) \qquad (9)$$

Let $y_{j,l}(s)$ as initial whale location, β specifies optimum power parameter. Then the random walk of levy distribution specifies μ . To get new solution randomly chosen the j^{th} iteration,

Humpback whales find out the prey positioning as well as updating the better search agent in each iteration, then the presentation is exhibited in eq^n (10), (11),

$$F = |K.Y(t) - X(t)|, \qquad (10)$$

$$K(t+1) = Y(t) - Fet_{selection}.F \qquad (11)$$

Where $K.Y$ implies coefficient vector, t specifies maximum number of iteration, EWO performs exploitation and exploration phase, if $K < 1$ acts exploitation the exploration is done. This exploitation phase upgrades optimal selection of features from the input dataset.

Step 5: Termination

The EWO selects the optimal features for identifying SCA in starting stage and to reduce the death rate of tribal’s in Nilgiris. The algorithm iterates step 4 to step 3 until fulfil the halting criteria time iteration $j = j + 1$.

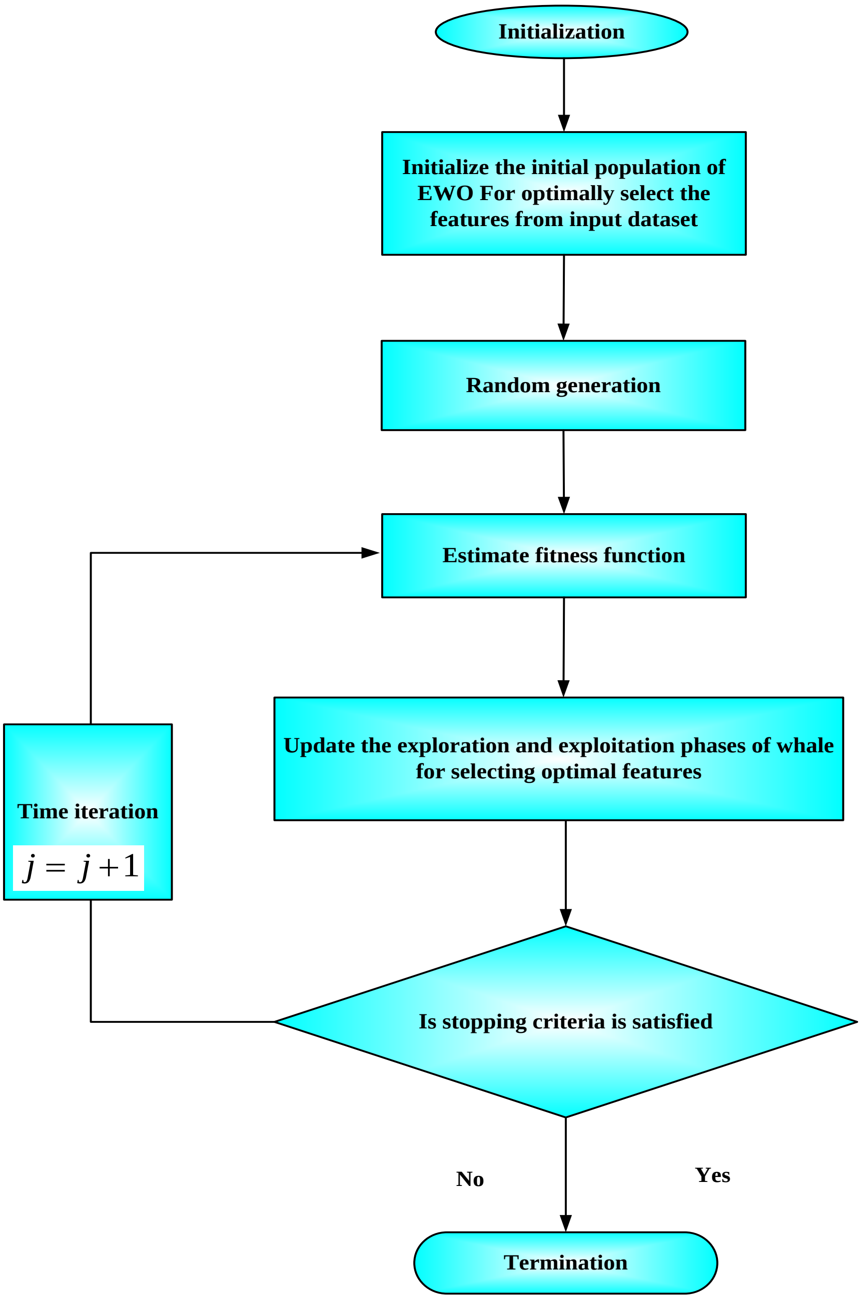

Initialization
Initialize the initial population of EWO For optimally select the features from input dataset
Random generation
Estimate fitness function
Update the exploration and exploitation phases of whale for selecting optimal features
Time iteration
$j = j + 1$
Is stopping criteria is satisfied
No
Yes
Termination

Figure 3 Flow chart of EWO algorithm for optimally selects the features from input dataset

3.1.4 Classification using Clustering based *Boosted C5.0 approach*

The Clustering based Boosted C5.0 is used to classify the SCA with maximal accuracy. The feature selected data is fed to Clustering base Boosted C5.0 to classify the SCA. The cluster-base under sampling mode first utilizes affinity propagation clustering for categorizing the majority class as well as the count of K-means clusters is predetermined, so every data sample owns to particular clusters in majority class. Every data sample is assigned to distinct cluster in 2nd clustering using K-means method with a preset count of clusters that is deemed as 1st stage. Then, the distance amid the samples is scaled by standardized Euclidean distance metrics. Finally, the sample including far distance is sampled; thus, the balancing trained data set acquired. Cluster-base Boosted C5.0 method is given in below steps:

Step 1: The original dataset is represented as A, therefore $A = \{a_j\}, j = 1,2,...m$, where a_j is represented as n instance along n-tuple of attribute values A, v_l implies center point, $f_{average}$ represented as average distance from the instance, v_l as cluster center μ_k, A_j^i denotes j^{th} input component of member i, A_j^{ki} denotes j^{th} input component of member μ_k, M as total count of data points, k as predefined cluster.

Step 2: The input feature selected data is represented as the majority class A is clustered through affinity propagation clustering approach for determining the count of clusters.

Step 3: Select k points randomly represent k initial clusters centroid from A, k is predefined in Step 2.

Step 4: The classified equation based cluster-based Boosted C5.0 approach is represented in equation (12).

$$J(G_k) = \sum_{a_j \in G_k} \|a_j - \mu_k\| \tag{12}$$

Where $J(G_k)$ is represented as a_j and the point μ_k is given in equation (12)

Step 5: The SCA disease is successfully detected or otherwise repeats the step 4-2. Then the final result is given as A'.

The proposed model uses the cluster-base under sampling strategy to exclude them from majority class, because border samples have a greater impact on categorization than internal samples. In cluster-base under sampling model, the size of training set is lessened and balanced, as a result, decreases the runtime including space complex in categorization process. This is a merit of classification performance. Boosted C5.0, an innovative top-down decision tree approach considers the property with the greatest information gain ratio to be the training attribute, and then uses the recursion process to generate a tree. C5.0 has some advantages, like considerable lower error rate, lesser memory, and high efficiency on the basis of rule quest research. So, it can make simpler decision tree. Boosting is a vital technology of C5.0. Every sample has equal at the process of iteration; then tuned every sample weight. These samples contain maximal weight. Boosting is an ensemble mode to combine numerous weak classifiers to get strong hypotheses. The fundamental concept of boosting is the instances allocate weights that reflect its significance. Besides, every misclassified minority events assigns maximal value of cost, i.e., boosting allocates maximal weight to samples resulting difficult to realize. Based on C5.0, C5.0 algorithm, the boosting technique is successfully implemented. Finally, cluster-based Boosted C5.0 algorithm accurately classifies the SCA, normal, Thalassemia.

4. Results and discussion

The Optimal Feature Selection and detection of Sickle Cell Anemia Detection using Enhanced Whale Optimization with Clustering based Boosted C5.0 Algorithm in Tribes of Nilgiris (EWOA-CBC5.OA-SCA) is discussed in this segment. The proposed model is implemented and experimented on PYTHON tool 3.7.12 version. The experiments are done in PC along configuration of Intel Core i7-10700 CPU running in 2.9, 4.8 GHz, 8 GB of RAM, 64-bit Windows 10-OS. The performance of EWOA-CBC5.OA-SCA method is analyzed with performance metrics. The derived outcomes are estimated to the existing RGSA-MLP-SCA, CRFA-SVM-SCA, AO-LSTM-SCA methods.

4.1. Dataset acquisition

The real-time data was collected from different regions around Nilgiris district. Totally, 300 samples were collected, which includes 187 female samples and 113 male samples. The age of the patients varies from 3 to 72. The dataset contains 13 features which are described in the following Table 1.

Table 1 Details of the dataset

Features	Description	Ratio
Weight	Patient's weight.	Depends on growth and age.
Hemoglobin (Hb)	The patient's haemoglobin level. The standard unit of measurement is grams (gm) per decilitre (dL).	Normal children: 11.5-16.5 (g/dl) S-thalassemia: 9.6 ± 0.8 (g/dl) sickle cell: 10.8 + 1.8 (g/dl)
Mean Corpuscular Volume (MCV)	RBCs, also known as erythrocytes, are measured by an MCV, which is the average size of an RBC.	Normal children: 11.5-16.5 Cu/urine S-C disorder: 75.4 ± 6.0 Cu/urine Sickle-thalassemia:70.4 ± 7.6 Cu/urine
Platelets (PLTS)	When we are injured, our bodies rely on platelets to help them form clots and stop the flow of blood.	Platelet count, 109/ L: 346 ± 530
Neutrophils(NEUT)	Microorganisms are killed by a type of immune cell that helps against infection. It is a part of the immune system's WBC type.	Neutrophils, 109/ L: 5.4 ± 11
Count of Reticulocyte (RETIC A)	Measures how quick RBCs moves oxygen from lungs to every cell.	Normal children: 0% - 2 % S-thalassemia: 9.7% ± 3.7% Sickle cell: 5.1% ± 2.2%
Alanine aminotransferase (ALT)	Checks for damage of liver	ALT: 24 ± 2.0 U/L
Reticulocyte Count (RETIC %)	Measures how quick RBCs moves oxygen from lungs to every cell.	Normal children: 0% - 2 % S-thalassemia: 9.7% ± 3.7% Sickle cell: 5.1% ± 2.2%
Hb F	Primary oxygen carry protein in human fetus.	7.9 % ± 13.7 %
Bilirubin (BILI)	Estimates the bilirubin level in blood. It assists doctor to find the health states such as anemia and liver infection.	BILI: 61.56 ± 10.26 μM
Body Bio Blood (BIO)	Features of dietary that are generated from result of blood.	Depends on weight of the body.
Aspartate Aminotransferase	Checks for damage of liver	Aspartate Aminotransferase: 49 ± 23.0 U/L
Lactate dehydrogenase	Measures the level of LDH in blood. The target was to detect the location and severity of tissue damages in the kidney and liver.	Lactate dehydrogenase: 487 ± 58 U/L

4.2 Performance metrics

The performance metrics, such as accuracy, PPV, NPV, specificity, MCC, error rate, sensitivity, computational time, complexity are evaluated to assess the performance of the proposed method.

- True Positive (T_P): Normal exactly categorized as Normal
- True Negative (T_N): Abnormal exactly categorised as Abnormal
- False Positive (F_P): Abnormal inexactly categorized as Normal
- False Negative (F_N): Normal inexactly categorized as Abnormal

4.2.1 Accuracy

It is the ratio of exact prediction to overall number of proceedings in the dataset. This is computed by eq^n (13),

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{13}$$

4.2.2 Sensitivity

It is the quantity of real positives that predicted accurately and is computed by eq^n (14),

$$Sensitivity = \frac{T_p}{F_N + T_p} \tag{14}$$

4.2.3 Specificity

This is the amount of true outcomes divided by the amount outcomes and is scaled by eq^n (15),

$$specificity\ value=\frac{T_p}{T_p+F_N} \tag{15}$$

4.2.4 Specificity

It consists of true negative rate, this is scaled by eq^n (16),

$$Specificity=\frac{F_p}{T_p+F_p} \tag{16}$$

4.2.5 Computation Time

The time required for final output image quality segmentation using eq^n (17),

$$C_{Time}=J*CPI/P \tag{17}$$

Here J implies image quantity, CPI implies cycle per instruction, P signifies segmentation period.

4.2.6 Positive Predictive Value

This is a metric for determining the accuracy of a positive prediction. It is expressed in equation (18),

$$PPV=\frac{TP}{TP+FP} \tag{18}$$

4.2.7 Matthews Correlation Coefficient

This is a model evaluation measure and is more accurate statistical value that generated a higher rate only if the predictions get better results in all four confusion matrix (TP, FP, TN, FN), proportional to the quantity of positive and negative factors in the data set, this is mathematically formulated in eq^n (19),

$$MCC=\frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{19}$$

4.2.8 Negative Predictive Value (NPV)

This is the proportion of patients with negative findings who are actually disease free and it is computed using eq^n (20),

$$NPV=\frac{TN}{TN+FN} \tag{20}$$

4.3 Simulation result

Figure 4 displays the output of Sickle cell anemia disease from the blood.

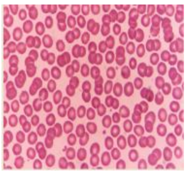
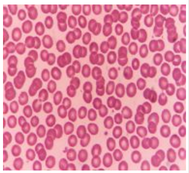
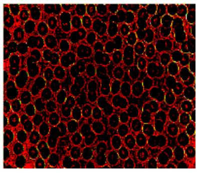
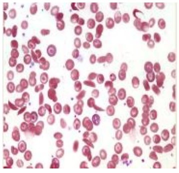
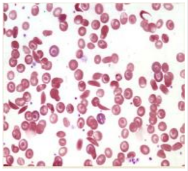
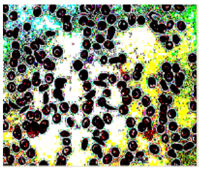
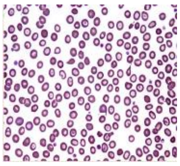
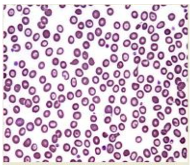
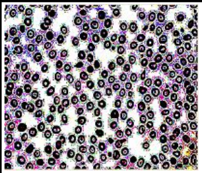
Input Image	**Pre-Processed image**	**Feature extracted and feature selected image**	**Classification**
			**Normal**
			**Sickle cell anemia**
			**Thalassemia Cell**

Figure 4 Output result of the Sickle cell anemia disease from the blood

The simulation result of proposed EWOA-CBC5.OA-SCA method is depicted in Table 2. In this, the proficiency of proposed approach is compared with existing RGSA-MLP-SCA, CRFA-SVM-SCA, AO-LSTM-SCA and BOA-CNN-SCA methods.

Table 2 Performance metrics of proposed method

Performance Metrics		**Methods**				
		RGSA-MLP-SCA	**CRFA-SVM-SCA**	**AO-LSTM-SCA**	**BOA-CNN-SCA**	**EWOA-CBC5.OA-SCA (Proposed)**
Accuracy	Normal cell	62.34	78.52	67.28	82.45	99.63
	Sickle Cell	85.63	79.65	84.27	65.27	98.52
	Thalassemia Cell	68.26	87.24	71.27	74.28	96.34
Sensitivity	Normal cell	74.52	65.28	89.32	56.48	97.64
	Sickle Cell	88.52	79.63	66.49	89.32	96.32
	Thalassemia Cell	78.56	87.32	85.46	65.32	99.48
Specificity	Normal cell	63.15	74.26	85.32	75.48	96.18
	Sickle Cell	85.64	79.63	59.78	67.85	98.36
	Thalassemia Cell	78.45	80.27	89.64	72.49	97.28
Positive Predictive Value (PPV)	Normal cell	63.14	79.64	84.27	62.49	95.26
	Sickle Cell	71.29	79.63	69.38	84.27	96.48
	Thalassemia Cell	82.35	78.23	82.14	88.27	98.26
Matthews Correlation Coefficient (MCC)	Normal cell	85.36	79.65	84.27	65.24	97.48
	Sickle Cell	76.32	84.27	62.34	79.65	96.32
	Thalassemia Cell	59.36	78.26	81.27	56.38	94.27
Negative Predictive Value (NPV)	Normal cell	63.21	57.29	84.36	74.28	95.32
	Sickle Cell	74.25	65.32	75.28	84.27	97.28
	Thalassemia Cell	52.36	64.21	78.39	69.24	94.18
Error Rate		37.66	21.48	32.72	17.55	0.37
Computation Time		0.23	0.45	0.52	0.24	0.02
Time Complexity		0.41	0.52	0.32	0.65	0.05

Table 2 tabulates the performance metrics of proposed method. The performance of EWOA-CBC5.OA-SCA method provides 52.32%, 43.78%, 32.78% and 45.90% higher accuracy for normal cell, 34.78%, 34.78%, 25.78% and 37.09% higher accuracy for sickle cell, 29.07%, 19.06%, 29.67% and 29.05% higher accuracy for thalassemia cell; 34.78%, 24.67%, 32.89% and 43.67% higher sensitivity for normal cell, 21.05%, 23.10%, 32.10% and 42.15% higher sensitivity for sickle cell, 32.10%, 22.30%, 44.21% and 30.21% higher sensitivity for thalassemia cell; 23.10%, 45.32%, 41.20% and 28.64% higher specificity for normal cell, 28.65%, 29.65%, 33.26% and 35.24% higher specificity for sickle cell, 32.10%, 21.30%, 32.10% and 23.65% higher specificity for thalassemia cell; 23.10%, 41.20%, 33.10% and 21.04% higher PPV for normal cell, 23.10%, 32.10%, 44.32% and 32.10% higher PPV for sickle cell, 29.35%, 23.10%, 19.65% and 22.31% higher PPV for thalassemia cell; 32.18%, 21.05%, 31.07% and 31.05% higher MCC for normal cell, 32.15%, 33.21%, 26.35% and 33.21% higher MCC for sickle cell, 31.08%, 22.10%, 39.64% and 32.10% higher MCC for thalassemia cell; 21.45%, 32.10%, 36.52% and 24.30% higher NPV for normal cell, 32.10%, 39.67%, 50.27% and 39.67% higher NPV for sickle cell, 29.07%, 19.06%, 29.67% and 29.05% higher NPV for thalassemia cell; 34.90%, 29.67%, 30.67% and 36.90% lower error rate; 23.89%, 17.90%, 18.56% and 29.56% lower computational time; 23.63%, 19.63%, 32.18% and 52.89% lower time complexity than the existing methods, such as RGSA-MLP-SCA, CRFA-SVM-SCA, AO-LSTM-SCA and BOA-CNN-SCA.

5. Conclusion

In this manuscript, The Optimal Feature Selection and detection of Sickle Cell Anemia Detection using Enhanced Whale Optimization with Clustering based Boosted C5.0 Algorithm in Tribes of Nilgiris is proposed. The input dataset is taken from real data set via nongovernment Organization named NAWA (situated in Kotagiri). The reason behind Nilgiris is used in this work is a collection of sickle cell anemia test results of the tribe peoples who lives in the different areas of Nilgiris region. Then these images are pre-processed using Wavelet Packet Transform Cochlear Filter Bank method to eradicate noises and enhance the superiority of the images. After that, the features are extracted utilizing. *Force-Invariant* Improved Feature Extraction Method. To select the optimal features, Enhanced Whale Optimization (EWO) algorithm is used. These optimal features are classified utilizing Clustering based Boosted C5.0 algorithm. The proposed model is implemented on PYTHON tool. The efficiency of the proposed method shows 23.65%, 22.67%, 32.76% and 33.75% higher precision compared with existing RGSA-MLP-SCA, CRFA-SVM-SCA, AO-LSTM-SCA, BOA-CNN-SCA models.

This method provides high accuracy, but it is limited, because this method is not applicable in large dataset, so that deep learning base optimization algorithm is used to detect the SCA in early stage.

This EWOA-CBC 5.0A-SCA is appropriate in real time applications. EWOA-CBC 5.0A-SCA method can characterize the patients depend on their SCA by gathering real time data from tribal's of Nilgiris. Not only that, this method is used to classify the SCA efficiently, also compare the accuracy of different ML strategies including feature selection optimization algorithms that are utilized to identify the SCA in starting stage. The proposed approach exhibits greater prediction with classification tasks estimated with other ML classifier models. These tasks are employed in SCA care for disease probability prediction, screening, detection, treatment guidance, complication management. This application represents a successful tool to assist the detection of SCA patients.

6. References

Abdullahi, S. U., Wudil, B. J., Bello-Manga, H., Musa, A. B., Gambo, S., Galadanci, N. A., ... & DeBaun, M. R. (2020). Primary prevention of stroke in children with sickle cell anemia in sub-Saharan Africa: rationale and design of phase III randomized clinical trial. *Pediatric Hematology and Oncology*, *38*(1), 49-64. https://doi.org/10.1080/08880018.2020.1810183

Balamanikandan, P., & Bharathi, S. J. (2022). A mathematical modelling to detect sickle cell anemia using Quantum graph theory and Aquila optimization classifier. *Mathematical Biosciences and Engineering: MBE*, *19*(10), 10060-10077. https://doi.org/10.3934/mbe.2022470

Carden, M. A., Fasano, R. M., & Meier, E. R. (2020). Not all red cells sickle the same: contributions of the reticulocyte to disease pathology in sickle cell anemia. *Blood Reviews*, *40*, Article 100637. https://doi.org/10.1016/j.blre.2019.100637

da Silva, M., Rodrigues, L., & Mari, J. F. (2020). Optimizing data augmentation policies for convolutional neural networks based on classification of sickle cells. In *Anais do XVI Workshop de Visão Computacional* (pp. 46-51). SBC. https://doi.org/10.5753/wvc.2020.13479

Delgado-Font, W., Escobedo-Nicot, M., González-Hidalgo, M., Herold-Garcia, S., Jaume-i-Capó, A., & Mir, A. (2020). Diagnosis support of sickle cell anemia by classifying red blood cell shape in peripheral blood images. *Medical & Biological Engineering & Computing, 58*(6), 1265-1284. https://doi.org/10.1007/s11517-019-02085-9

Diop, S., & Pirenne, F. (2021). Transfusion and sickle cell anemia in Africa. *Transfusion Clinique etBiologique*, *28*(2), 143-145.

https://doi.org/10.1016/j.tracli.2021.01.013

Domingos, I. F., Pereira-Martins, D. A., Sobreira, M. J., Oliveira, R. T., Alagbe, A. E., Lanaro, C., ...& Santos, M. N. (2020). High levels of proinflammatory cytokines IL-6 and IL-8 are associated with a poor clinical outcome in sickle cell anemia. *Annals of Hematology*, *99*(5), 947-953. https://doi.org/10.1007/s00277-020-03978-8

Dong, M., & McGann, P. T. (2021). Changing the clinical paradigm of hydroxyurea treatment for sickle cell anemia through precision medicine. *Clinical Pharmacology & Therapeutics*, *109*(1), 73-81. https://doi.org/10.1002/cpt.2028

Fields, M. E., Mirro, A. E., Binkley, M. M., Guilliams, K. P., Lewis, J. B., Fellah, S., ... & Lee, J. M. (2022). Cerebral oxygen metabolic stress is increased in children with sickle cell anemia compared to anemic controls. *American Journal of Hematology*, *97*(6), 682-690. https://doi.org/10.1002/ajh.26485

Gour, A., Dogra, A., Bhatt, S., & Nandi, U. (2020). Effect of natural products on improvement of blood pathophysiology for management of sickle cell anemia. In *Botanical Leads for Drug Discovery* (pp. 51-65). Singapore: Springer, https://doi.org/10.1007/978-981-15-5917-4_3

Hamsa, S., Shahin, I., Iraqi, Y., & Werghi, N. (2020). Emotion recognition from speech using wavelet packet transform cochlear filter bank and random forest classifier. *IEEE Access*, *8*, 96994-97006. https://doi.org 10.1109/ACCESS.2020.2991811

Hindawi, S., Badawi, M., Elfayoumi, R., Elgemmezi, T., Al Hassani, A., Raml, M., ...& Gholam, K. (2020). The value of transfusion of phenotyped blood units for thalassemia and sickle cell anemia patients at an academic center. *Transfusion*, *60*, S15-S21. https://doi.org/10.1111/trf.15682

Islam, M. J., Ahmad, S., Haque, F., Reaz, M. B. I., Bhuiyan, M. A. S., & Islam, M. R. (2021). Force-invariant improved feature extraction method for upper-limb prostheses of transradial amputees. *Diagnostics*, *11*(5), 843. https://doi.org/10.3390/diagnostics11050843

Olupot-Olupot, P., Connon, R., Kiguli, S., Opoka, R. O., Alaroker, F., Uyoga, S., ...& Williams, T. N. (2022). A predictive algorithm for identifying children with sickle cell anemia among children admitted to hospital with severe anemia in Africa. *American Journal of Hematology*, *97*(5), 527-536. https://doi.org/10.1002/ajh.26492

Onalo, R., Cooper, P., Cilliers, A., Vorster, B. C., Uche, N. A., Oluseyi, O. O., ... & Morris, C. R. (2021). Randomized control trial of oral arginine therapy for children with sickle cell anemia hospitalized for pain in Nigeria. *American Journal of Hematology*, *96*(1), 89-97. https://doi.org/10.1002/ajh.26028

Petrović, N., Moyà-Alcover, G., Jaume-i-Capó, A., & González-Hidalgo, M. (2020). Sickle-cell disease diagnosis support selecting the most appropriate machine learning method: Towards a general and interpretable approach for cell morphology analysis from microscopy images. *Computers in Biology and Medicine*, *126*, Article 104027. https://doi.org/10.1016/j.compbiomed.2020.104027

Rajesh, P., Shajin, F. H., Mouli Chandra, B., & Kommula, B. N. (2021). Diminishing Energy Consumption Cost and Optimal Energy Management of Photovoltaic Aided Electric Vehicle (PV-EV) By GFO-VITG Approach. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 1-19. https://doi.org/10.1080/15567036.2021.1986606

Rajesh, P., Shajin, F. H., Rajani, B., & Sharma, D. (2022). An optimal hybrid control scheme to achieve power quality enhancement in micro grid connected system. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, *35*, Article e3019. https://doi.org/10.1002/jnm.3019

Reeves, S. L., Freed, G. L., Madden, B., Wu, M., Miller, L., Cogan, L., & Dombkowski, K.

J. (2022). Trends in quality of care among children with sickle cell anemia. *Pediatric Blood & Cancer*, *69*(2), Article e29446. https://doi.org/10.1002/pbc.29446

Risoluti, R., Caprari, P., Gullifa, G., Massimi, S., Sorrentino, F., Maffei, L., & Materazzi, S. (2020). Innovative screening test for the early detection of sickle cell anemia. *Talanta*, *219*, Article 121243. https://doi.org/10.1016/j.talanta.2020.121243

Shajin, F. H., Rajesh, P., & Raja, M. R. (2022). An efficient VLSI architecture for fast motion estimation exploiting zero motion prejudgment technique and a new quadrant-based search algorithm in HEVC. *Circuits, Systems, and Signal Processing*, *41*(3), 1751-1774. https://doi.org/10.1007/s00034-021-01850-2

Shajin, F. H., Rajesh, P., & Thilaha, S. (2020). Bald eagle search optimization algorithm for cluster head selection with prolong lifetime in wireless sensor network. *Journal of Soft Computing and Engineering Applications*, *1*(1), 1-7.

Singh, B. K., Ojha, A., Bhoi, K. K., Bissoyi, A., & Patra, P. K. (2021). Prediction of hydroxyurea effect on sickle cell anemia patients using machine learning method. In *Advances in Biomedical Engineering and Technology* (pp. 447-457). Singapore: Springer, https://doi.org/10.1007/978-981-15-6329-4_37

Soltani, S., Zakeri, A., Tabibzadeh, A., Zandi, M., Ershadi, E., AkhavanRezayat, S., & Farahani, A. (2020). A literature review on the parvovirus B19 infection in sickle cell anemia and β-thalassemia patients. *Tropical Medicine and Health*, *48*(1), 1-8. https://doi.org/10.1186/s41182-020-00284-x

Valayapalayam Kittusamy, S. R., Elhoseny, M., & Kathiresan, S. (2022). An enhanced whale optimization algorithm for vehicular communication networks. *International Journal of Communication Systems*, *35*(12), Article e3953. https://doi.org/10.1002/dac.3953

Verlhac, S., Gabor, F., Paillard, C., Chateil, J. F., Jubert, C., Petras, M., ...& Drepagreffe Investigators. (2021). Improved stenosis outcome in stroke-free sickle cell anemia children after transplantation compared to chronic transfusion. *British Journal of Haematology*, *193*(1), 188-193. https://doi.org/10.1111/bjh.17178

Wang, W. C., Zou, P., Hwang, S. N., Kang, G., Ding, J., Heitzer, A. M., ...& Hankins, J. S. (2021). Effects of hydroxyurea on brain function in children with sickle cell anemia. *Pediatric Blood & Cancer*, *68*(10), Article e29254. https://doi.org/10.1002/pbc.29254

Wonkam, A., Chimusa, E. R., Mnika, K., Pule, G. D., Ngo Bitoungui, V. J., Mulder, N., ... & Adeyemo, A. (2020). Genetic modifiers of long-term survival in sickle cell anemia. *Clinical and Translational medicine*, *10*(4), Article e152. https://doi.org/10.1002/ctm2.152

Zhang, J., Chen, L., Tian, J. X., Abid, F., Yang, W., & Tang, X. F. (2021). Breast cancer diagnosis using cluster-based undersampling and boosted C5. 0 algorithm. *International Journal of Control, Automation and Systems*, *19*(5), 1998-2008. https://doi.org/10.1007/s12555-019-1061-x

Zhang, M., Li, X., Xu, M., & Li, Q. (2020). Automated semantic segmentation of red blood cells for sickle cell disease. *IEEE Journal of Biomedical and Health Informatics*, *24*(11), 3095-3102. https://doi.org/10.1109/JBHI.2020.3000484