

Article

# A Reliable Auto-Robust Analysis of Blood Smear Images for Classification of Microcytic Hypochromic Anemia Using Gray Level Matrices and Gabor Feature Bank

Bakht Azam <sup>1</sup>, Sami Ur Rahman <sup>1,\*</sup>, Muhammad Irfan <sup>2</sup>, Muhammad Awais <sup>3</sup>, Osama Mohammed Alshehri <sup>4</sup>, Ahmed Saif <sup>4</sup>, Mohammed Hassan Nahari <sup>4</sup> and Mater H. Mahnashi <sup>5</sup>

<sup>1</sup> Department of Computer Science and IT, University of Malakand, Chakdara 18801, Pakistan; bazam@uom.edu.pk

<sup>2</sup> College of Engineering, Electrical Engineering Department, Najran University, Najran 61441, Saudi Arabia; irfan16.uetian@gmail.com

<sup>3</sup> School of Computing and Communications, Lancaster University, Bailrigg, Lancaster LA1 4YW, UK; m.awais11@lancaster.ac.uk

<sup>4</sup> Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Najran University, Najran 61441, Saudi Arabia; Omalshehri@nu.edu.sa (O.M.A.); amsaif8080@gmail.com (A.S.); Mhnahari@nu.edu.sa (M.H.N.)

<sup>5</sup> Department of Medicinal Chemistry, Pharmacy School, Najran University, Najran 61441, Saudi Arabia; matermaha@gmail.com

\* Correspondence: srahman@uom.edu.pk

Received: 15 July 2020; Accepted: 15 September 2020; Published: 17 September 2020



**Abstract:** Accurate blood smear quantification with various blood cell samples is of great clinical importance. The conventional manual process of blood smear quantification is quite time consuming and is prone to errors. Therefore, this paper presents automatic detection of the most frequently occurring condition in human blood—microcytic hyperchromic anemia—which is the cause of various life-threatening diseases. This task has been done with segmentation of blood contents, i.e., Red Blood Cells (RBCs), White Blood Cells (WBCs), and platelets, in the first step. Then, the most influential features like geometric shape descriptors, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Gabor features (mean squared energy and mean amplitude) are extracted from each of the RBCs. To discriminate the cells as hypochromic microcytes among other RBC classes, scanning is done at angles ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ ). To achieve high-level accuracy, Adaptive Synthetic (AdaSyn) sampling for imbalance learning is used to balance the datasets and locality sensitive discriminant analysis (LSDA) technique is used for feature reduction. Finally, upon using these features, classification of blood cells is done using the multilayer perceptual model and random forest learning algorithms. Performance in terms of accuracy was 96%, which is better than the performance of existing techniques. The final outcome of this work may be useful in the efforts to produce a cost-effective screening scheme that could make inexpensive screening for blood smear analysis available globally, thus providing early detection of these diseases.

**Keywords:** erythrocytes; RBCs; segmentation; classification; anemia; reliable

## 1. Introduction

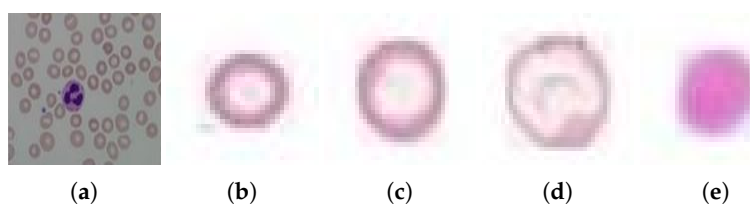
Anemia is an abnormal condition in human blood in which the amount of red blood cells (and therefore their oxygen-carrying capacity) is inadequate to fulfill the physiologic needs of the body.

The World Health Organization (WHO) estimates that 42% of children less than 5 years of age and 40% of pregnant women worldwide are anaemic [1,2]. According to the translated prevalence by [3], 496 million are nonpregnant women, 32 million are pregnant women, and 273 million are children with anemia in 2011. When a peripheral blood smear slide is viewed under a microscope, it is found to be composed of three main cell types, i.e., Red Blood Cells (RBCs), White Blood Cells (WBCs), and platelets. Normal RBCs are round and reddish in color, having a central pallor area in the center indicating hemoglobin level; see Figure 1. The diameter of normal RBCs ranges from about 6.2 to 8.2 micrometer ( $\mu\text{m}$ ), and its thickness ranges from about 2 to 2.5 micrometers ( $\mu\text{m}$ ) [4]. The amount of hemoglobin indicates whether a human is healthy or has a potential disease that affects RBC production or function. This indication is through several parameters of whole RBC (size, shape, and number) and the amount of hemoglobin it contains. The size of RBCs indicates three main defects in human blood: normocytosis, microcytosis, and macrocytosis. Normocytosis is an anemia of normocytes, where the size of RBCs is normal. Microcytosis occurs when the size of the RBC is less than that of the normal reference range and is expressed by a blood index called Mean Corpuscular Volume (MCV) of less than 80 femtoliters, whereas macrocytosis is when the size of the RBC is greater than the normal reference range and is expressed by an MCV greater than the normal range [5].

Patients of anemia may be categorized into mild, moderate, and severe according to WHO classification, i.e., mild anemia (hemoglobin 9.0–10.9 g/dL) moderate anemia (hemoglobin 7.0–8.9 g/dL), and severe anemia (hemoglobin <7.0 g/dL) [1]. The amount of hemoglobin on the other hand specifies two other abnormalities: hypochromic, if the mean concentration of hemoglobin in a single cell is lower than the normal reference range, i.e., 27 to 32 picogram (pg) (approximately) and is stated by Mean Corpuscular Hemoglobin Concentration (MCHC), and hyperchromic, if the range is above the said reference range [6].

The above discussed terms are parameters used in distinguishing the primary types of anemic conditions. When both cell size and hemoglobin defects are combined, e.g., if microcytic is taken from the first category and combined with hypochromic from the second category, then it will define the features of microcytic hypochromic anemia, whereas if macrocytic and hyperchromic are combined, it will give macrocytic hyperchromic anemia, etc. [7,8]. The diseases which are most frequently diagnosed in South Asian countries especially in Pakistan are iron deficiency anemia, sideroblastic anemia, and thalassemia, which are categorized and defined under microcytic hypochromic anemia [9]. These parameters may be obtained from peripheral blood smear slides prepared by hematologists in clinical labs of hospitals proceeded by grabbing digital images, which is an area of interest.

This work is organized as follows: related work is explained in Section 2, the preparation and visualization of the blood smear is explained in Section 3. To better understand the nature of improvement, we carried out a thorough study that builds up a methodology in which blood smear after following the necessary procedures is processed with image processing techniques for segmentation. The primary steps of image processing are image acquisition, preprocessing (conversion into color spaces, splitting channels, and enhancement techniques), binarization, segmentation, etc. [10]. Then, the classification step explained in Section 4 is conducted with Adaptive Synthetic (AdaSyn) sampling to balance the datasets and with Locality Sensitive Discriminant Analysis (LSDA) for the reduction of unnecessary features of all the contents visible in a blood smear. The rest of paper presents the results (Section 5) and finally the conclusion and discussion (Section 6).



**Figure 1.** (a) A normal blood smear image and (b) microcytic hypochromic, (c) normocytic hypochromic, (d) macrocytic hypochromic, and (e) microcytic hyperchromic anemia.

## 2. Related Work

Various state-of-the-art image processing-based techniques for computer-aided disease detection are present excessively, some of which are the most relevant to our work presented here. Moallem et al. [11] used a three-step algorithm for segmentation of overlapped cells in blood smear images by extracting a binary mask in the first step, then an adaptive mean shift algorithm was used to centrally localize a cell, and finally the gradient vector flow algorithm was used to draw boundaries for separation of cells. In the given article, there is no shape analysis done, which fails to classify the significant classes of RBCs. Tomari et al. [12] used some geometric features for the classification of RBCs through Artificial Neural Network (ANN). Alomari et al. [13] worked on the automatic quantification of WBCs and RBCs by using an iterative structured circle detection algorithm, but in the given technique, it was found that it is restricted to circular cells. The technique suggested by Aggarwal et al. [14] is intensity-based Otsu thresholding where segmentation of RBCs infected with a parasite was done, but the segmentation by intensity may suffer extremely due to luminance variations and other photographic conditions. Tek et al. [15] suggested a technique for recognition and classification of malarial RBCs' parasites and species in peripheral blood smears. Shape, color, and local granulometry features were extracted from the area of interest, and k Nearest Neighbor (kNN) classifier was applied to classify them from extracted features. Chen et al. [16] evaluated blood smear slides having hemolytic anemia by determining the chain codes for finding the edges of cells, separated the cells with the help of concavity measurement, and classified cells with a bank of classifiers. A work done by Xu et al. [17] for the detection of sickle cells is segmentation of red cells in the first step, then separation of overlapped cells using random walk algorithm in the next step, and classification through Deep Convolutional Network in the last step. Sharma et al. [18] used a median filter for image smoothing and watershed for overlapped cells separation. Then, they implemented a three-feature vector, circularity metric, aspect ratio, and radial signature, trained with KNN classifier for the recognition of three types of RBCs called sickle cells, elliptocytes, and dacrocytes. A work for the detection and classification of parasites in RBCs was done by Ahirwar et al. [19]. They generated geometric, color attributes and gray level texture feature sets and used artificial neural network for their classification. A fuzzy logic technique was used by Bhagavathi et al. [20] for segmentation of RBCs and WBCs. Morphological operations and hough transform method for circle detection were used by Chandrasiri et al. [21] for the detection and analysis of red blood cells, but the proposed system was unable to determine and analyze the extensive number of clumped or overlapped regions. The above presented approaches can do better in a situation where the population of clumped cells is low. In highly populated clumped cells, accuracy suffers. In our proposed approach, the issue is fixed to much extent by leveraging the concavities and texture-based features in each cell as explained in the following Materials and Methods section.

## 3. Materials and Methods

Our proposed plan of work for the detection of hypochromic microcytic cells in sample blood smear images consisted of the following series of steps shown in Figure 2.

### 3.1. Blood Smear Slide Preparation

A consistent blood distribution and proper lucidity are required for reliable blood smear analysis. It can be done by starting with a drop of sample blood at one end of glass slide, which is smeared quickly and gently with a wedge technique to form a thin edge, where all cells are able to be analyzed separately, especially RBCs [22]. This whole process was done by an expert laboratory technician in the local hospital.

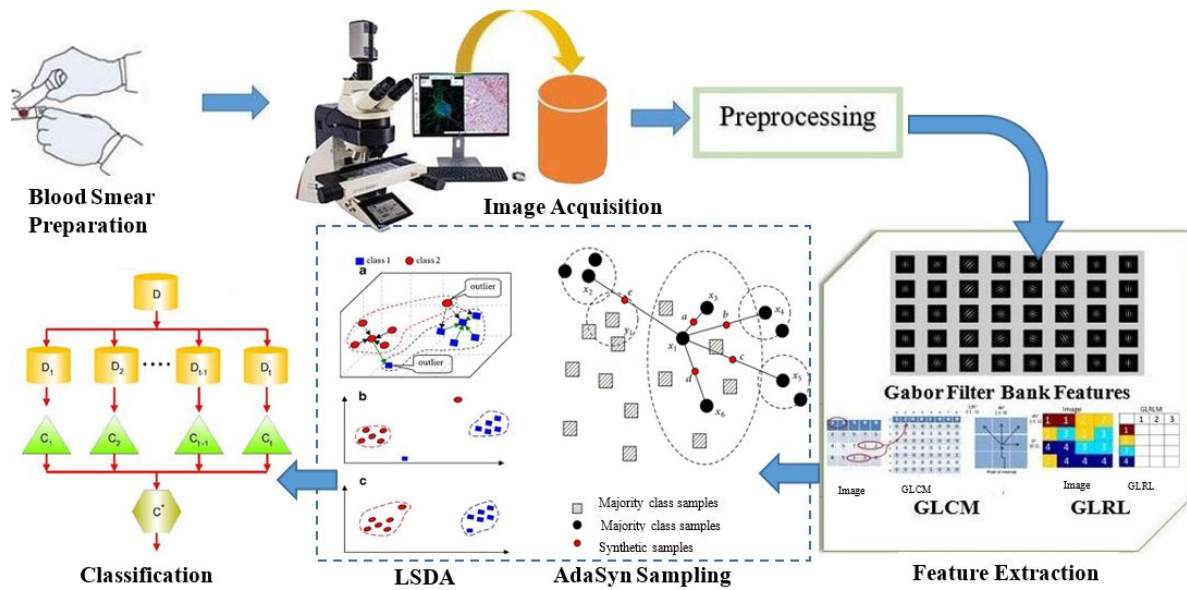


Figure 2. A step-by-step proposed plan of work.

### 3.2. Image Acquisition

After the staining process was completed, the image acquisition step started by using a 400× field of a microscope with oil immersion, keeping the horizontal and vertical resolution at 180 dpi and the image dimension at 2592 × 1944 pixels. The images were captured and labelled properly.

### 3.3. Preprocessing

A green channel of RGB blood smear was selected, enhanced, and smoothed using Balance Contrast Equalization Technique (BCET) and median filter. After getting a fine gray-scale image shown in Figure 3, it was quantized with a scale factor  $F_g$ , calculated from mean intensity value of  $g(x, y)$  calculated in Equations (1) and (2), where  $Q(x, y)$  is the resultant quantized image.

$$F_g = \frac{1}{N} \sum_1^N g(x, y) \tag{1}$$

$$Q(x, y) = F_g * \text{round}\left(\frac{g(x, y)}{F_g}\right) \tag{2}$$

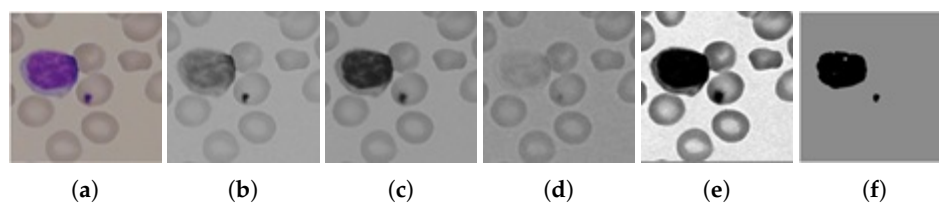
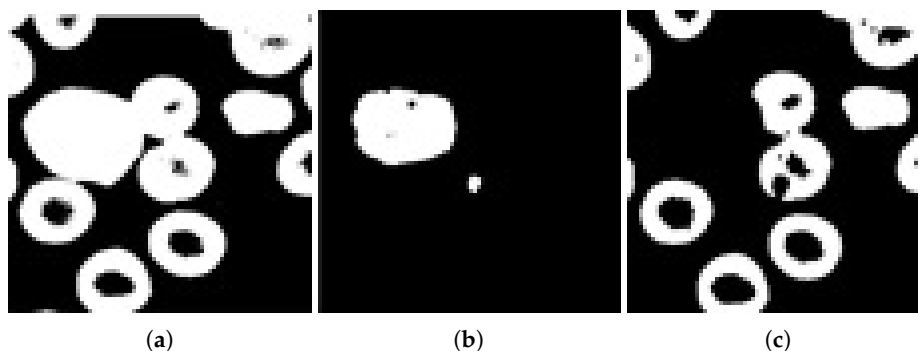


Figure 3. Output images after the preprocessing step: (a) original RGB image, (b) red channel, (c) green channel, (d) blue channel, (e) enhanced green channel, and (f) quantized image leaving behind RBCs.

### 3.4. Segmentation

A global or automatic thresholding technique was used in this work for the segmentation of only RBCs (leaving behind WBCs and platelets). A two-step binarization technique was followed for getting noise-free binary images. In the first step, the quantized image was binarized containing white blood cells; in the second step, the whole original image was binarized. Finally, we used the XOR gate

(Exclusive OR) operation to remove white blood cells. The output along with other processed images are shown in Figure 4.



**Figure 4.** (a) Binarized original image of a sample blood smear, (b) binarized image of a sample blood smear quantized, and (c) exclusive OR of (a,b).

### 3.5. Feature Extraction

A three-way feature extraction plan was implemented. The first was geometric morphological feature extraction, the second was intensity features and texture features using (GLCM and GLRLM), and the third was texture features using the Gabor Filter Bank.

#### 3.5.1. Geometric Morphology of Microcytic Hypochromic RBCs

Keeping in view the visual size and shape of hypochromic microcytes, the useful features were calculated as follows:

- **Area:** Area is an important geometrical feature for the detection of microcytes, being small in size compared to other blood cells.
- **Circularity:** A size-invariant shape descriptor given in Equation (3) which describes a shape to be circular, if the value is closer to 1 and noncircular if the value is closer to 0, where  $A$  is Area and  $P$  is parameter of a cell.

$$Circularity = \frac{4\pi A}{P^2} \quad (3)$$

- **Rectangularity:** It determines the degree of elongation with respect to a rectangle. Equation (4) shows its calculation, where  $A_s$  is area of a shape and  $A_r$  is the area of minimum bounding rectangle.

$$Rectangularity = \frac{A_s}{A_r} \quad (4)$$

- **Concavity:** This property is used to determine how much an object is concave; we applied it on the shapes for identification of the amount of central pallor area occupied in an RBC, given by Equation (5)

$$Concavity = \frac{A_c}{A_H} \quad (5)$$

- **Convexity:** A cell convexity can be determined by Equation (6), which identifies a shape through its boundary convexity.

$$Convexity = \frac{P_C H}{P_C} \quad (6)$$

#### 3.5.2. RBC Texture Feature Calculation

Features other than geometry include texture information on RBCs. In this research work, we extracted RGB intensity-based features (mean and variances), Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Gabor feature bank. This set of features

associated with the dispersal of chromatin matter in the RBCs is helpful in the classification of hypochromic, hyperchromic, and normochromic cells given in Figure 1 above. The measure will indicate that either the RBC has a deep red color compared to the others (cells having more haemoglobin) or are less red, with more central pallor area and having a smaller haemoglobin ratio. The texture features calculated are presented below:

- RGB mean and variance of hypochromic microcytic RBCs:

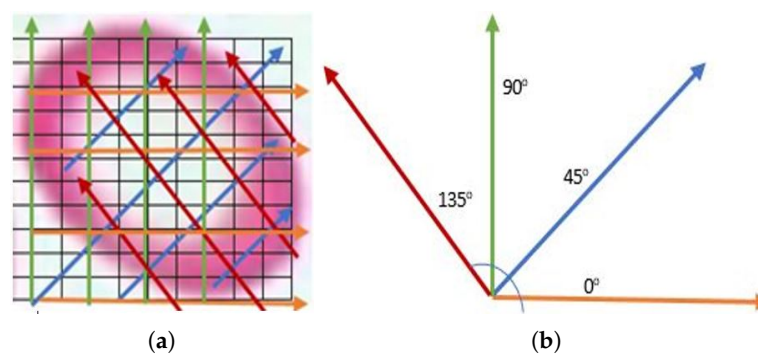
The mean values  $\mu_r$ ,  $\mu_g$ , and  $\mu_b$  of pixels of each RBC in R, G, and B, respectively, were calculated in Equation (7).

$$\begin{aligned}\mu_r &= \frac{1}{N} \sum_1^N r(x, y) \\ \mu_g &= \frac{1}{N} \sum_1^N g(x, y) \\ \mu_b &= \frac{1}{N} \sum_1^N b(x, y)\end{aligned}\quad (7)$$

The variances  $(\sigma_r)^2$ ,  $(\sigma_g)^2$ , and  $(\sigma_b)^2$  in the channels R, G, and B, respectively, are calculated in Equation (8).

$$\begin{aligned}(\sigma_r)^2 &= \frac{\sum(X - \mu_r)^2}{N} \\ (\sigma_g)^2 &= \frac{\sum(X - \mu_g)^2}{N} \\ (\sigma_b)^2 &= \frac{\sum(X - \mu_b)^2}{N}\end{aligned}\quad (8)$$

- GLCM features of Hypochromic Microcytic RBCs: GLCM is the distribution of cooccurring pixel values defined over an  $N \times N$  image  $P$  at a specific offset, or every  $P$ 's element determines the occurrences of a pixel with value of gray level,  $i$ , lifted by a certain distance to a pixel with value  $j$ . Our next six textural features are GLCM features. The mean of 6 GLCM features were determined for offset values conforming to  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  consuming 8 gray levels (see Figure 5).



**Figure 5.** (a) Orientation of angles overlaid on a sample red blood cell (RBC) and (b) scanning through  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ .

**Maximum Probability:** It measures the strongest response of the cooccurrence matrix. The range of values is  $[0, 1]$  as given in Equation (9), where  $P_{ij}$  is the pixels of gray image.

$$P = \max_{ij}(P_{ij}) \quad (9)$$

**Correlation:** The degree of correlation of a pixel to its neighbor is determined by the correlation factor of the cooccurrence matrix, ranging from 1 to  $-1$  given by Equation (10). This measure cannot be defined if any of the standard deviation  $\sigma$  is 0 for the two existing correlations, perfect positive and perfect negative correlation.

$$\text{correlation} = \sum_{i,j=1}^K \frac{(i - m_r)(j - m_c)p_{(ij)}}{\sigma_r \sigma_c} \quad (10)$$

**Pixels intensity contrast:** It is a measure of intensity contrast between a pixel and its neighbor over the entire image (calculated in Equation (11)).

$$\text{contrast} = \sum_{i,j=1}^k (i - j)^2 p_{ij} \quad (11)$$

**Energy:** It is the measurement of uniformity in the intensities of an image (as given in Equation (12)). Its value is 1, if an image is constant and 0 if the intensities are variable.

$$\text{Energy} = \sum_{i,j=1}^k (P_{(i,j)})^2 \quad (12)$$

**Homogeneity:** It measures the spatial closeness of the distribution of elements in the cooccurrence matrix to the diagonal given by (13). The values range is  $[0, 1]$ , and the maximum value is attained when the matrix is a diagonal.

$$\text{homogeneity} = \sum_{i,j=1}^k \frac{P_{ij}}{1 + |i - j|} \quad (13)$$

**Entropy:** It measure the degree of variability of the elements of the cooccurrence matrix. Its value is 0 if all intensities of  $P_{ij}$  are 0 and is maximum when all  $P_{ij}$  are equal. It may be calculated by (14).

$$\text{entropy} = - \sum_{i,j=1}^k P_{ij} \log_2 P_{ij} \quad (14)$$

- **Run length matrix features of each RBC:** The other textural features are created on the gray-level run length matrix (calculated in Equations (15)–(24)). The  $l * K$  matrix  $p$ , where  $l$  is the number of gray levels and  $k$  is the maximum run length, is defined for a certain image as the total runs with pixels of gray level  $i$  and run length  $j$ . Likewise, as in the GLCM, the run length matrices were calculated using 8 gray-levels for  $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ , and  $135^\circ$ .

**Short Run Emphasis (SRE):**

$$\text{SRE} = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^k \frac{P_{ij}}{j^2} \quad (15)$$

**Long Run Emphasis (LRE):**

$$\text{LRE} = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^k (p_{ij}) j^2 \quad (16)$$

**Gray-Level Nonuniformity (GLN):**

$$\text{GLN} = \frac{1}{R} \sum_{i=1}^k \left( \sum_{j=1}^l (p_{ij}) \right)^2 \quad (17)$$

Run Length Nonuniformity (RLN):

$$RLN = \frac{1}{R} \sum_{j=1}^l \left( \sum_{i=1}^k p_{ij} \right)^2 \quad (18)$$

Low Gray-level Run Emphasis (LGRE):

$$LGRE = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^l \frac{p_{ij}}{i^2} \quad (19)$$

High Gray-level Run Emphasis (HGRE):

$$HGRE = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^l (p_{ij}) i^2 \quad (20)$$

Short Run Low Gray-level Emphasis (SRLGE):

$$SRLGE = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^l \frac{p_{ij}}{i^2 j^2} \quad (21)$$

Short Run High Gray-level Emphasis (SRHGE):

$$SRHGE = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^l \frac{p_{ij} i^2}{j^2} \quad (22)$$

Long Run Low Gray-level Emphasis (LRLGE):

$$LRLGE = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^l \frac{p_{ij} j^2}{i^2} \quad (23)$$

Long Run High Gray-level Emphasis (LRHGE):

$$LRHGE = \frac{1}{R} \sum_{i=1}^k \sum_{j=1}^l (p_{ij}) i^2 j^2 \quad (24)$$

- Gabor Feature Extraction: like a human visual processing system, the Gabor filter extracts features at different amplitudes and orientation.

$$\begin{aligned} \psi(x, y) &= \frac{f^2}{\pi \gamma \eta} e^{-\left(\frac{f^2}{\gamma^2} x'^2\right) - \left(\frac{f^2}{\eta^2} y'^2\right)} e^{j2\pi f x'} \\ x' &= x \cos \theta + y \sin \theta \\ y' &= x \sin \theta + y \cos \theta \end{aligned} \quad (25)$$

The Gabor filter is the product of a 2D Fourier basis function and origin-centred Gaussian given in Equation (25), where  $f$  is the central frequency of the filter,  $\gamma$  and  $\eta$  are the sharpness or bandwidth measure along the minor and major axes of Gaussian respectively,  $\theta$  is the angle of rotation, and  $(\eta/\gamma)$  is the aspect ratio. The analytical form of this function in frequency domain is given in Equation (26) as follow:



$$\psi(u, v) = e^{-\frac{\pi^2}{f^2}(\gamma^2(u'-f)^2 + \eta^2 v'^2)} \quad (26)$$

$$u' = u \cos \theta + v \sin \theta$$

$$v' = u \sin \theta + v \cos \theta$$

In the frequency domain given by Equation (27), the function is a single real-valued Gaussian centered at  $f$ . A simplified version of a general 2D Gabor filter function in Equations (25) and (26) was formulated by [23], which implements a set of self-similar filters, i.e., Gabor wavelets (rotated and scaled forms of each other, irrespective of the frequency  $f$  and orientation  $\theta$ ).

Gabor bank or Gabor features were created from responses of Gabor filters in Equations (25) and (26) by using multiple filters on several frequencies  $f_m$  and orientations  $\theta_n$ . Frequency in this case corresponds to scale information and is thus drawn from [23]

$$f_m = k^{-m} f_{max}, m = \{0, \dots, M - 1\} \quad (27)$$

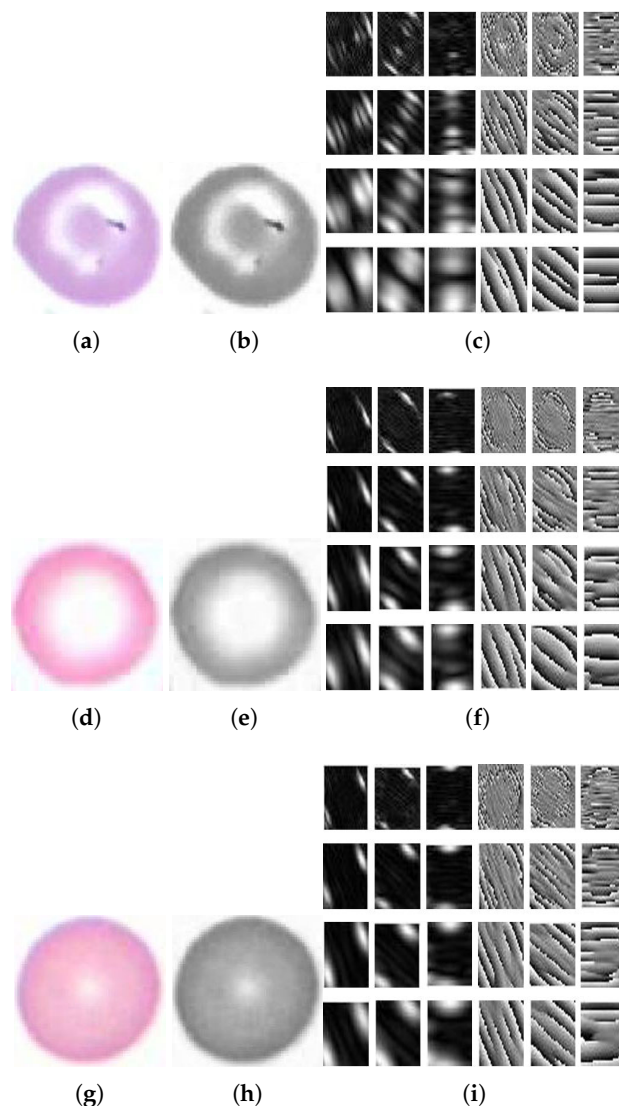
where  $f_m$  is the  $m$ th frequency,  $f_{max}$  is the highest frequency desired, and  $k > 1$  is the frequency scaling factor. The filter orientations are drawn from [24]. Gabor features were calculated at 4 wavelengths (3, 6, 9, and 12) and 3 orientations  $\theta$  (30°, 60°, and 90°); see Figure 6a–c. Then, each filter was convolved with the real image, and the response image of the same image was produced; here, each image gave us a feature vector. Each feature vector consisted of mean amplitude and mean squared energy. Finally, two matrices were obtained, that were of  $[1 \times 12]$  each. The matrices were appended to each other, and a  $[1 \times 24]$  matrix was produced for one image having a  $[n \times 24]$  vector for  $n$  images for supplementary training purpose in the preceding step of classification (as shown in Figure 6).

### 3.6. ADASYN Sampling

A significant aspect in classification and learning is to show a reasonable dataset to guarantee that no inclination is presented by an imbalanced information distribution. A method that has been used in previous works is Adaptive Synthetic Sampling (ADASYN) to enhance the classification accuracy by balancing the datasets, thus decreasing bias factors [25]. Table 1 shows that the original dataset is partially imbalanced; therefore, we applied ADASYN to overcome this problem and to balance the dataset. After applying ADASYN sampling, the database then consisted of 354 microcytic, 327 normocytic, 312 macrocytic, 340 hypochromic, and 380 normochromic images of blood smears.

**Table 1.** Balanced datasets after applying Adaptive Synthetic Sampling (ADASYN).

Cell Type	Original	Synthetic	Total
Microcyte	157	197	354
Normocytes	270	57	327
Macrocytes	101	211	312
Hypochromic	157	150	340
Normochromic	380	0	380



**Figure 6.** (a) Original image of hyperchromic macrocytic RBC, (b) gray level image of hyperchromic macrocytic RBC, (c) Gabor filter bank of hyperchromic macrocytic RBC, (d) original image of hypochromic microcytic RBC, (e) gray level image of hypochromic microcytic RBC, (f) Gabor filter bank features of hypochromic microcytic RBC, (g) original image of hyperchromic microcytic RBC, (h) gray level image of hyperchromic microcytic RBC, and (i) a hyperchromic microcytic RBC with its Gabor filter bank features.

### 3.7. Features Reduction

To maintain variation among interclass data samples, it is necessary to reduce the dimensionality of an original dataset. We used Gray Level Cooccurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Gabor filter bank, which collectively produced 52 features for a single cell image in a blood smear image. Therefore, to reduce the dimensionality, a Locality Sensitive Discriminant Analysis (LSDA) [26] approach was applied separately to the features extracted from each cell (shown in Figure 7). We use LSDA because it is significant in the case where there are no sufficient training samples. LSDA uses local structures, and it is generally more important than global structure for discriminant analysis. LSDA determines a projection using the local manifold structure, which results in the maximization of the margin among data points from different classes at every local area. Various experiments on the existing datasets showed an improvement over the Linear Discriminant Analysis (LDA).

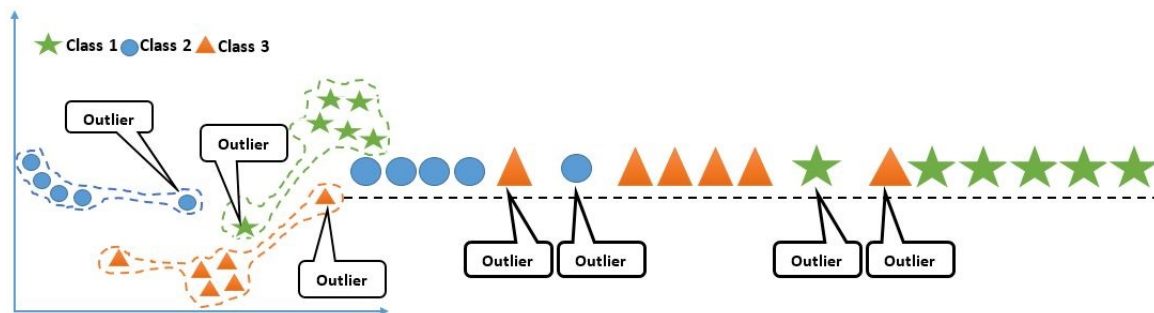


Figure 7. Feature reduction using Locality Sensitive Discriminant Analysis (LSDA).

#### 4. Classification

In different situations, varied instances show a tendency to a specific classification tool. Therefore, iterative experiments have been performed for the selection and determination of an optimal tool. The tools determined during the process were random forest and multilayer perceptual modal. These are ideal classification tools in this situation. For classification, the training data set was prepared with the extracted features mentioned above. The images were labelled with ground truth from the existing dataset images, containing Iron Deficiency Anaemia (IDA) images, mostly. The instances of dataset comprised of geometric morphological feature, GLCM, GLRL, and Gabor texture features are given in the features below. We used the classical machine learning methods as the overfitting and underfitting anomalies of pretrained deep learning algorithms suffer in accuracy of results due to less sufficient data in datasets. The performance of our classification technique is given in the Results Section 5.

#### 5. Results

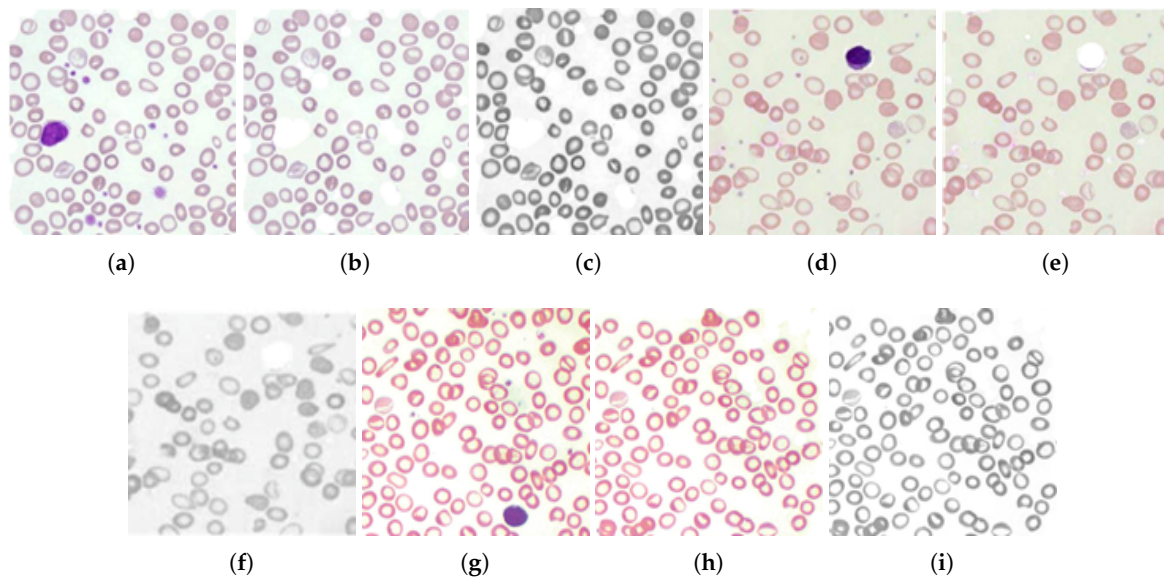
The results of the proposed work are as follows.

##### 5.1. Dataset

The images we used in our work were collected from a local hospital. At the selected hospital, a high frequency of microcytic hypochromic patients were observed. The available datasets were also searched for such images, and a set of 150 images (comprising of about 80 hypochromic microcytes per image) were created based on ground truth, labeled by expert hematologists.

##### 5.2. Qualitative Results

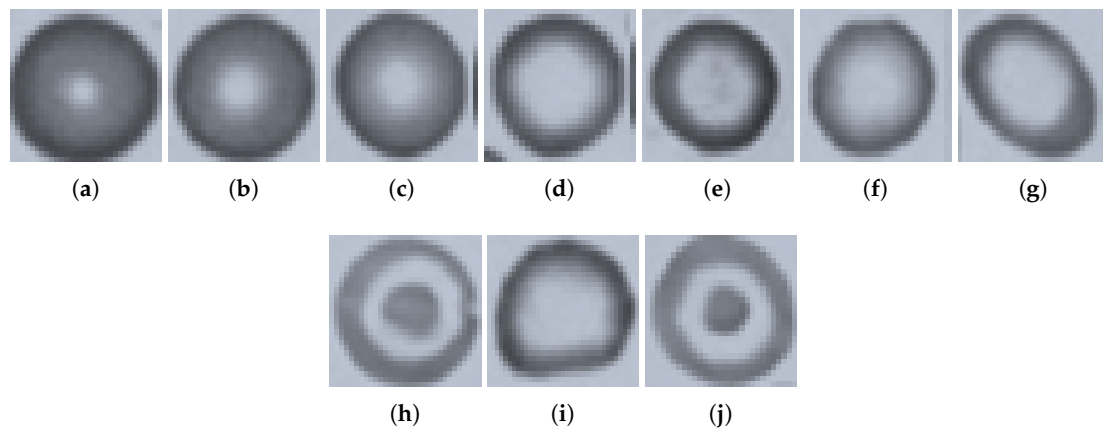
Hypochromic microcytes are the cells we are interested in for segmentation, the cell with a more pallor area and smaller than normal mean sizes of RBCs (Figure 8). These types of cells have less pixel area with more hole area proportionality. A hole containing cells was assumed to be a pallor cell and was less chromatic than the cell having no hole. The chromaticity factor is inversely proportional to the proportionality of central pallor: the bigger the central pallor, the less chromatic the cell will be and the cell will have a big hole in it after carrying out binarization operations, and hence, the concavity of the cell will also be more. The qualitative results of the GLCM, GLRL, Gabor mean amplitude, and Gabor mean square energy are given in Tables 2–5. The qualitative results are also demonstrated in Figures 8 and 9. The sample cells for which the features are extracted and demonstrated in Tables 2, 3, 4 and 5 are shown in Figure 10.



**Figure 8.** (a,d,g) Original RGB images of a sample blood smear, (b,e,h) removal of white blood cells (WBCs) from the blood smear images, and (c,f,i) enhanced gray-scale images of RBCs in blood smears.

### 5.3. Quantitative Results

This work has two main purposes: one is to segment the red blood cells accurately and then to detect the microcytic hypochromic cell. To evaluate the detection algorithm, the precision, recall, F1-score and accuracy metrics were calculated (using Equations (28)–(31)), where TP, TN, FP, FN are true positive, true negative, false positive, and false negative, respectively. To evaluate the detection algorithm, the precision, accuracy, recall, and F1-score matrices were calculated. The performance-wise results of different algorithms are given in Table 6.



**Figure 9.** Classes of cells among RBCs (a,b) microcytic hypochromic cells, (c) microcytic normochromic cells, (d) macrocytic hyperchromic cells (e,f) microcytic hyperchromic cells (g) macrocytic Hyperchromic cells, and (h,j) codocytes or target cells.

**Table 2.** Gray Level Cooccurrence Matrix (GLCM) features extracted from each cell in a sample image.

Cell No.	Contrast	Correlation	Energy	Homogeneity
Cell1	85.40	406.62	123.52	105.8
Cell2	80.02758	−112.132	136.3636	108.52
Cell3	116.9	394.59	100	93.68
Cell4	96.18	94.716	117.85	100.0
Cell5	121.1	−79.49	80	92.77
Cell6	106.2	−148.7	80	98.03
Cell7	86.6	−201.2	120	104.07
Cell8	88.12	88.12	120	104
Cell9	106.1	97.47	100	98
Cell10	100	100	100	100

**Table 3.** Gray Level Run Length (GLRL) features.

Cell No.	SRE	LRE	GLN	RLN	RP	LGRE	HGRE	SRLGE	SRHGE	LRLGE	LRHGE
Cell1	0.56	44.24	116.35	113.27	8.74	2.09	50.11	0.22	12.46	24.79	209.57
Cell2	0.7	40.25	126.78	108.01	8.69	2.26	39.37	0.29	14.12	25.91	127.59
Cell3	0.43	44.88	135.02	139.61	8.47	2.02	51.45	0.18	10.59	24.89	268.0
Cell4	0.5	65.18	137.38	115.35	12.93	2.62	48.2	0.23	7.83	45.05	134.07
Cell5	88.24	106.62	94.77	106.14	102.68	93.53	110.29	80.23	86.07	96.51	118.61
Cell6	115.4	100.85	94.51	98.77	103.26	92.73	99.82	120.77	73.02	88.5	120.18
Cell7	0.39	61.57	132.04	111.73	11.22	2.6	43.64	0.19	7.9	42.7	130.44
Cell8	1.25	89.76	103.12	103.36	29.92	7.01	81.33	0.59	19.3	70.39	130.52
Cell9	0.33	47.64	309.2	216.5	11.1	2.44	28.86	0.15	6.02	32.81	254.73
Cell10	0.37	45.41	355.65	206.29	10.46	2.56	28.17	0.18	5.99	31.69	325.68

**Table 4.** Gabor's Mean Squared Energy (MSE).

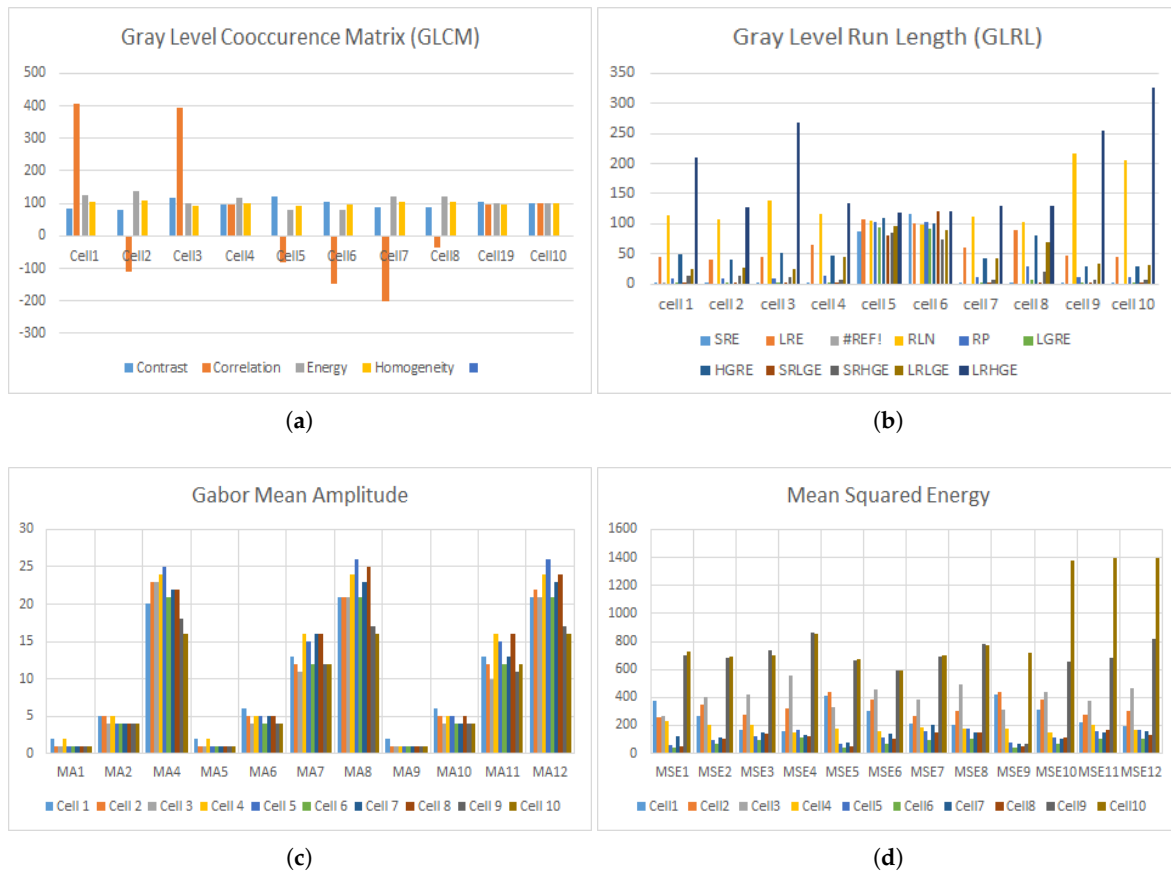
Cell No.	MSE1	MSE2	MSE3	MSE4	MSE5	MSE6	MSE7	MSE8	MSE9	MSE10	MSE11	MSE12
Cell1	376	263	165.7	155.2	414.9	301.7	216.3	204.5	424.4	313.3	226.3	197
Cell2	258	350	274.6	316.8	436.3	380.9	264.9	306.7	439.7	382.9	273.3	299
Cell3	269	403	419.8	554.9	328.8	457.5	386.6	493.4	314.5	440.3	377.1	462
Cell4	234	205	207.9	153.3	177.8	157.8	186.5	174.7	173.5	153	200.3	172
Cell5	62	96	121.3	164.9	72.1	110.4	157.8	173.6	73.3	109.7	155.8	167
Cell6	45	70	91.5	113.5	40.8	70.5	93.3	104.4	41.5	72.4	102	105
Cell7	120	115	149.8	136	79.1	140	201.2	149.7	70.6	102.6	147.2	156
Cell8	50	101	138.6	120	51.8	101	149.7	146	54.2	113.4	163.6	131
Cell9	701	682	737.9	862.7	669	592	692	784	70	653.5	685.5	820
Cell10	730	695	700	850	670	590	701	770	715	1373.9	1396.1	1397

**Table 5.** Gabor's Mean Amplitude (MA).

Cell No.	MA1	MA2	MA3	MA4	MA5	MA6	MA7	MA8	MA9	MA10	MA11	MA12
Cell1	2	5	11	20	2	6	13	21	2	6	13	21
Cell2	1	5	12	23	1	5	12	21	1	5	12	22
Cell3	1	4	11	23	1	4	11	21	1	4	10	21
Cell4	2	5	16	24	2	5	16	24	1	5	16	24
Cell5	1	4	13	25	1	5	15	26	1	5	15	26
Cell6	1	4	11	21	1	4	12	21	1	4	12	21
Cell7	1	4	13	22	1	5	16	23	1	4	13	23
Cell8	1	4	14	22	1	5	16	25	1	5	16	24
Cell9	1	4	12	18	1	4	12	17	1	4	11	17
Cell10	1	4	12	16	1	4	12	16	1	4	12	16

**Table 6.** Performance evaluation of each classification algorithm with statistical measurement tools.

Statistical	K-means Clustering (%)	Logistic Regression (%)	Naive Bayes (%)	Proposed Classifier (%)
Precision	81.1	86.4	87.1	92.3
Accuracy	80.7	86.2	88.3	93.2
Recall	80.3	83.2	84.3	95.4
F1-Score	79.8	82.1	84.1	94.1



**Figure 10.** Graphs showing the results of four types of texture features, (a) GLRLM (b) GLCM (c) GMSE (d) GMA.

$$Precision = \frac{FP}{TP + FP} \tag{28}$$

$$Recall = \frac{TP}{TP + FN} \tag{29}$$

$$F1Measure = 2 * \frac{precision * Recall}{precision + Recall} \tag{30}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{31}$$

### 6. Conclusions

The goal of this study is to develop and improve a robust algorithm for the analysis of blood smear images for the classification of microcytic hypochromic anaemia. Many studies are present in the literature that focus on the classification of blood cells, but there are very rare studies found on the classification of normal and abnormal blood slides as a specific anemic disease. Moreover, the existing

state-of-the-art techniques are very expensive and their operation is very difficult. Our proposed system has a reasonable accuracy rate and processing time. The proposed system is capable of detecting various chromatic status of blood and accurately estimates the boundary pixels of RBCs at diverse photographic conditions. The results showed that our algorithm is a better automatic segmentation methods for blood smear images. The geometric and three potential texture features of RBCs, i.e., GLCM, GLRL, and Gabor texture features at 4 different degrees of scan (3, 6, 9, and 12) and 3 orientations  $\theta$  ( $30^\circ$ ,  $60^\circ$ , and  $90^\circ$ ) were extracted. After feature extraction and the feature reduction technique, the features were then put in two powerful machine learning algorithms (random forest and multilayer perceptron) using ensemble learning technique. The overall classification accuracy was 96%. Which was compared to the existing techniques and were found better for classification. The present work may be extended to a 3D volume estimation of blood cells, which is necessary in finding the accurate blood indices like Mean Corpuscular Volume (MCV), Mean Corpuscular Haemoglobin (MCH), and Mean Corpuscular Haemoglobin Concentration (MCHC). The proposed system may facilitate pathologists by getting quicker results with high True Positive (TP) and True Negative (TN) rates in the initial stage of diagnosis. It is user-friendly and easily operable with less expenses in terms of cost. Furthermore, the system may be made available on the web without any association of special equipment and user knowledge. Therefore, it can be considered an added value to the existing automated analysis blood smear tools.

**Author Contributions:** The research conceptualization and methodology were done by B.A., S.U.R., and M.I. The technical and theoretical framework was prepared by M.A., O.M.A., and A.S. The technical review and improvement were performed by M.H.N. and M.A. The overall technical support, guidance, and project administration were done by M.H.M. and O.M.A. Finally, responses to the queries of the reviewers were done by B.A. and M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors acknowledge the Deanship of Scientific Research, Najran University, Kingdom of Saudi Arabia for their support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. World Health Organization. *Haemoglobin Concentrations for the Diagnosis of Anaemia and Assessment of Severity*; World Health Organization: Geneva, Switzerland, 2011.
2. Farooqi, N.; Naqvi, M.; Jaffery, S.; Anemia, H.F. Anemia; moderate to severe during pregnancy. *Prof. Med. J.* **2014**, *21*, 247–252.
3. Stevens, G.A.; Finucane, M.M.; De-Regil, L.M.; Paciorek, C.J.; Flaxman, S.R.; Branca, F.; Peña-Rosas, J.P.; Bhutta, Z.A.; Ezzati, M. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant women for 1995–2011: A systematic analysis of population-representative data. *Lancet Glob. Health* **2013**, *1*, e16–e25. [[CrossRef](#)]
4. Jandl, J.H. Blood: Textbook of hematology. *Blood* **1996**, *4*, 3.
5. Brecher, G.; Jakobek, E.F.; Schneiderman, M.A.; Williams, G.Z.; Schmidt, P.J. Size distribution of erythrocytes. *Ann. N. Y. Acad. Sci.* **1962**, *99*, 242–261. [[CrossRef](#)] [[PubMed](#)]
6. Vreugdenhil, G.; Baltus, C.; Eijk, H.V.; Swakk, A. Anaemia of chronic disease: Diagnostic significance of erythrocyte and serological parameters in iron deficient rhuematioid arthritis pateints. *Rheumatology* **1990**, *29*, 105–110. [[CrossRef](#)] [[PubMed](#)]
7. Aulakh, R.; Sohi, I.; Singh, T.; Kakkar, N. Red cell distribution width (rdw) in the diagnosis of iron deficiency with microcytic hypochromic anemia. *Indian J. Pediatr.* **2009**, *76*, 265–268. [[CrossRef](#)] [[PubMed](#)]
8. Bachman, A.L. Macrocytic hyperchromic anemia in early infancy: Report of a case and review of the literature. *Am. J. Dis. Child.* **1936**, *52*, 633–647. [[CrossRef](#)]
9. Habib, M.A.; Black, K.; Soofi, S.B.; Hussain, I.; Bhatti, Z.; Bhutta, Z.A.; Raynes-Greenow, C. Prevalence and predictors of iron deficiency anemia in children under five years of age in pakistan, a secondary analysis of national nutrition survey data 2011–2012. *PLoS ONE* **2016**, *11*, e0155051. [[CrossRef](#)] [[PubMed](#)]
10. Gonzalez, R.C.; Woods, R.E.; Eddins, S.L. *Digital Image Processing Using MATLAB*; Pearson-Prentice-Hall: Upper Saddle River, NJ, USA, 2004; Volume 624.

11. Moallem, G.; Sari-Sarraf, H.; Poostchi, M.; Maude, R.J.; Silamut, K.; Hossain, M.A.; Antani, S.; Jaeger, S.; Thoma, G. Detecting and segmenting overlapping red blood cells in microscopic images of thin blood smears. In Proceedings of the Medical Imaging 2018: Digital Pathology, Houston, TX, USA, 10–15 February 2018.
12. Tomari, R.; Zakaria, W.N.W.; Jamil, M.M.A.; Nor, F.M.; Fuad, N.F.N. Computer aided system for red blood cell classification in blood smear image. *Procedia Comput. Sci.* **2014**, *42*, 206–213. [[CrossRef](#)]
13. Alomari, Y.M.; Abdullah, S.; Huda, S.N.; Azma, R.Z.; Omar, K. Automatic detection and quantification of wbcs and rbcs using iterative structured circle detection algorithm. *Comput. Math. Methods Med.* **2014**, *2014*, 979302. [[CrossRef](#)] [[PubMed](#)]
14. Aggarwal, P.; Khatter, A.; Vyas, G. An intensity threshold based image segmentation of malaria infected cells. In Proceedings of the 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 15–16 February 2018; pp. 549–553.
15. Tek, F.B.; Dempster, A.G.; Kale, Z. Parasite detection and identification for automated thin blood film malaria diagnosis. *Comput. Vis. Image Underst.* **2010**, *114*, 21–32. [[CrossRef](#)]
16. Chen, H.-M.; Tsao, Y.-T.; Tsai, S.-N. Automatic image segmentation and classification based on direction texton technique for hemolytic anemia in thin blood smears. *Mach. Vis. Appl.* **2014**, *25*, 501–510. [[CrossRef](#)]
17. Xu, M.; Papageorgiou, D.P.; Abidi, S.Z.; Dao, M.; Zhao, H.; Karniadakis, G.E. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLoS Comput. Biol.* **2017**, *13*, e1005746. [[CrossRef](#)] [[PubMed](#)]
18. Sharma, V.; Rathore, A.; Vyas, G. Detection of sickle cell anaemia and thalassaemia causing abnormalities in thin smear of human blood sample using image processing. In Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016; pp. 1–5.
19. Ahirwar, N.; Pattnaik, S.; Acharya, B. Advanced image analysis based system for automatic detection and classification of malarial parasite in blood images. *Int. J. Inf. Technol. Knowl. Manag.* **2012**, *5*, 59–64.
20. Bhagavathi, S.; Niba, S.T. An automatic system for detecting and counting rbc and wbc using fuzzy logic. *ARPJ. Eng. Appl. Sci.* **2016**, *11*, 6891–6894.
21. Chandrasiri, S.; Samarasinghe, P. Morphology based automatic disease analysis through evaluation of red blood cells. In Proceedings of the 2014 5th International Conference on Intelligent Systems, Modelling and Simulation, Langkawi, Malaysia, 27–29 January 2014; pp. 318–323.
22. Adewoyin, A. Peripheral blood film—a review. *Ann. Ib. Postgrad. Med.* **2014**, *12*, 71–79. [[PubMed](#)]
23. Daugman, J.G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A* **1985**, *2*, 1160–1169. [[CrossRef](#)] [[PubMed](#)]
24. Ilonen, J.; Kamarainen, J.-K.; Paalanen, P.; Hamouz, M.; Kittler, J.; Kalviainen, H. Image feature localization by multiple hypothesis testing of gabor features. *IEEE Trans. Image Process.* **2008**, *17*, 311–325. [[CrossRef](#)] [[PubMed](#)]
25. Yung, M.; Brown, E.T.; Rasin, A.; Furst, J.D.; Raicu, D.S. Synthetic sampling for multi-class malignancy prediction. *arXiv* **2018**, arXiv:1807.02608.
26. Gao, Y.-L.; Luo, S.-Z.; Wang, Z.-H.; Chen, C.-C.; Pan, J.-Y. Locality sensitive discriminative unsupervised dimensionality reduction. *Symmetry* **2019**, *11*, 1036. [[CrossRef](#)]

