DIJEMSS: Dinasti International Journal of Education Management and Social Science
E-ISSN: 2686-6331
P-ISSN: 2686-6358
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/dijemss.v7i1

Classification of Human Blood Cells Based on Subtypes Using Linear Discriminant Analysis on Microscope Images

Raihan Herlambang[1], Asril Adi Sunarto[2], Fathia Frazna Azzahra[3]
[1]Informatics Engineering Study Program, Faculty of Science and Technology, Muhammadiyah University of Sukabumi, Indonesia, raihanhrlambang@gmail.com
[2]Informatics Engineering Study Program, Faculty of Science and Technology, Muhammadiyah University of Sukabumi, Indonesia, asriladi@ummi.ac.id
[3]Informatics Engineering Study Program, Faculty of Science and Technology, Muhammadiyah University of Sukabumi, Indonesia, fathiafrazna@ummi.ac.id

Corresponding Author: raihanhrlambang@gmail.com[1]

Abstract: This study aims to develop an automated classification system for human blood cells using microscopic images to improve the accuracy of subtype identification. To overcome the limitations of manual classification, the research adopts an artificial intelligence approach using the Linear Discriminant Analysis (LDA) algorithm, chosen for its effectiveness in dimensionality reduction and data group separation. The study follows the Knowledge Discovery in Databases (KDD) methodology, involving data selection, preprocessing (normalization, enhancement, segmentation, and noise removal), feature extraction using the Gray Level Co-occurrence Matrix (GLCM), and classification into eight blood cell types. The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. The research aims to contribute to more efficient and accurate medical image classification systems and demonstrate the potential of LDA in AI-based medical applications.

Keywords: Blood Cell Classification, Linear Discriminant Analysis (LDA), Knowledge Discovery in Databases (KDD), Gray Level Co-occurrence Matrix (GLCM), Digital Image Processing

INTRODUCTION

Blood is a vital component of the human body that functions as a transport medium within the metabolic system. Through the circulatory system, essential substances such as nutrients and oxygen are delivered throughout the body. Generally, blood consists of three main types of cells: red blood cells (erythrocytes), white blood cells (leukocytes), and platelets (thrombocytes). Erythrocytes are responsible for transporting oxygen from the lungs to body tissues and carrying carbon dioxide from the tissues back to the lungs for exhalation. Leukocytes serve as the body's defense system against infections caused by bacteria or viruses. Platelets play a role in the blood clotting process. When the body is injured, platelets form blood clots to prevent excessive bleeding and aid in wound healing (Wonohadidjojo, 2021).

White blood cells, or leukocytes, consist of two subtypes: granulocytes and agranulocytes. Granulocytes include basophils, neutrophils, and eosinophils, while agranulocytes include lymphocytes and monocytes. These subtypes are distinguished based on the presence of cytoplasmic granules (Prasthio dkk., 2022).

Erythrocytes, commonly known as red blood cells, are the most abundant type of blood cells in the human body. These cells have membranes that contain hemoglobin, a protein crucial for transporting oxygen throughout the body and giving blood its distinctive red color. The lifespan of erythrocytes is relatively short, around 120 days. After this period, about 1% of erythrocytes break down and are naturally replaced through a regeneration process in the body. Red blood cells can be classified into two subtypes based on their shape and size: isomorphic erythrocytes and dysmorphic erythrocytes. Isomorphic erythrocytes have a normal and uniform shape and size, while dysmorphic erythrocytes exhibit abnormal shapes, which can indicate disorders or abnormalities in the body (Arviananta dkk., 2020).

Platelets, also known as thrombocytes, play a crucial role in maintaining bodily stability, especially in the blood clotting process. One of their main functions is to form clots at injury sites to prevent excessive bleeding. Structurally, platelets consist of four major zones, each with distinct functions. The outermost zone, called the peripheral zone, contains receptors and essential molecules that help activate and adhere platelets to injury sites. The membrane zone is part of the cell membrane and enables interaction with the surrounding environment. The cytoskeleton zone provides shape and structural integrity, while the organelle zone contains components directly involved in the clotting process. With this complex structure, platelets play a vital role in the body's wound response and maintaining circulatory system integrity (Boudreaux & Christopherson, 2020).

There are various subtypes of human blood cells, such as basophils, eosinophils, erythroblasts, immunoglobulins (Ig), lymphocytes, monocytes, neutrophils, and platelets. Currently, the identification process of these cells is typically performed manually through microscopic observation by experts. This method is time-consuming and prone to inconsistencies due to its reliance on subjective observation. To address these limitations, there is a need for an automated technology-based system capable of classifying blood cells more quickly, efficiently, and accurately. One approach is to utilize machine learning methods, which can recognize specific patterns in microscopic images and group blood cells according to their subtypes (Alkafrawi & Dakhell, 2022).

The rapid pace of technological development demands continuous adaptation and innovation to tackle various challenges with effective solutions. One such advancement is the use of artificial intelligence (AI) in computer systems, especially for classification tasks. In the context of machine learning, the ability to classify microscopic images with high accuracy is essential. One algorithm that can be used for this process is Linear Discriminant Analysis (LDA). This algorithm excels at maximizing the separation between classes while minimizing variation within each class. Although LDA is an established algorithm, recent studies show that it remains effective and relevant in medical image classification as long as the features used align with its characteristics (Meenakshi dkk., 2022).

The main issue in this study lies in the image classification process of microscopic images. The case study focuses on blood cell classification, which presents its own challenges due to variations in cell size, shape, and color. These variations are typically caused by differences in lighting during medical image acquisition. Therefore, the data processing stage becomes a crucial part of ensuring that the features extracted from the images are of high quality and effectively support the classification process. Several key steps—such as normalization, segmentation, and feature extraction—are commonly applied in medical imaging research, as they significantly improve the performance of classification methods. In this study, the classification approach uses the Linear Discriminant Analysis

(LDA) algorithm, which requires relevant and well-structured features to produce optimal results (Gatc & Maspiyanti, 2022).

The LDA approach is known for its simplicity in processing data, including in the context of medical imaging. One of its advantages is the ability to produce classification results that are clearly interpretable, making the model's output easier to understand.

In this research, the data used were obtained from public sources on the Kaggle platform, widely used in various studies due to its completeness and diversity. The dataset contains thousands of human blood cell images from various subtypes, requiring a classification method capable of handling such complexity thoroughly and accurately. LDA was selected as the classification algorithm for its ability to process complex data while maintaining an understandable structure (Kouzehkanan dkk., 2022).

Linear Discriminant Analysis (LDA) is a method not only capable of reducing data dimensionality but also of highlighting the differences between groups or classes within a dataset. In statistics and machine learning, LDA is used to find the most effective feature combinations to distinguish two or more classes. One of its main advantages is its ability to retain important information from extracted features—especially those related to texture patterns and tissue structures—in the form of coefficients that represent data characteristics. However, in practice, the resulting features can often have high dimensionality, which may become an obstacle in data processing. LDA offers a solution by reducing dimensions without losing essential information, and it can enhance classification performance by projecting data into an optimal space where class differences become clearer and easier to identify (Zhu dkk., 2022).

Human blood consists of various types, such as red blood cells (erythrocytes), white blood cells (leukocytes), and platelets (thrombocytes). These three types of cells play essential roles in maintaining the body's metabolic functions. For example, white blood cells include components such as lymphocytes, monocytes, and neutrophils, which are crucial to the immune system. The presence and characteristics of blood cells are critical indicators for certain diseases, such as infections, anemia, or leukemia. In medical practice, identifying blood cell types is usually done through manual microscopic analysis by experts. However, this approach has limitations, especially in terms of accuracy and result consistency. Therefore, to improve the accuracy and efficiency of the blood cell image classification process, a technology-based system is needed to assist experts in processing and recognizing cell images more accurately and quickly. Such an approach is expected to minimize errors and support more accurate medical diagnoses (Lin dkk., 2023).

Based on this background, manual identification and classification of human blood cells still has limitations in terms of efficiency, dependence on individual expertise, and the potential for bias. Thus, an automated approach based on digital image processing and artificial intelligence is necessary to improve accuracy and efficiency. Linear Discriminant Analysis (LDA) is chosen due to its ability to reduce dimensions and separate classes effectively while remaining competitive in medical image classification when relevant features are used. Considering challenges such as variations in shape, size, and color of human blood cells, this study aims to implement LDA for human blood cell classification using microscopic images. Hence, the title selected is: "Classification of Human Blood Cells by Subtype Using Linear Discriminant Analysis (LDA) on Microscopic Images."

METHOD

Research Method

This research utilizes the Knowledge Discovery in Databases (KDD) approach, which includes the stages of Selection, Preprocessing, Transformation, Data Mining, Evaluation, and Knowledge. KDD serves as a structured method for extracting valuable information from

large and complex data sets. Each stage in this process plays a critical role in ensuring that the resulting information is truly accurate and reliable.

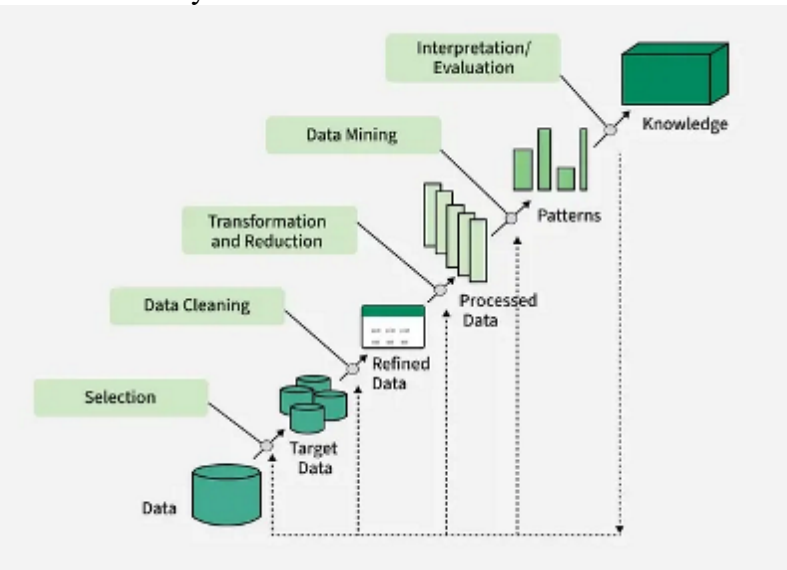


Figure 1. Knowledge Discovery Methodology in Databases
Source: (KDD Process/Overview, n.d.)

Data Collection Techniques

In this study, data collection focuses on obtaining microscopic images of human blood cells for classification using the Linear Discriminant Analysis (LDA) method. This stage is crucial to ensure the data is high-quality, relevant, and representative of the research objectives. A secondary data collection method is used, sourcing labeled and verified datasets from trusted platforms such as public databases, hospitals, research institutions, and platforms like Kaggle.

The dataset includes various types of blood cells—red blood cells, white blood cells, and platelets—along with their respective subtypes required for accurate classification. Key considerations in data collection include high image quality for clear visual features, diverse samples representing different clinical conditions, and consistent lighting. All images are annotated by medical professionals to ensure labeling accuracy.

After data collection, a verification and validation process is conducted to ensure the data is usable and aligned with research goals. This involves checking image formats, completeness, and consistency of visual characteristics. Irrelevant or inconsistent data is filtered out to maintain classification model accuracy.

Using secondary data offers advantages such as reduced time and cost, along with access to large and diverse image sets. However, challenges include the need to align existing data with research requirements and potential variability in image quality and formats that could affect analysis results.

Research Location and Object

This research used a dataset of microscopic images of human blood cells obtained from validated secondary data sources, such as public repositories that provide accurately annotated blood cell images. Therefore, the research location focused on data processing and analysis using the provided computer equipment.

This research was conducted in a laboratory or research environment that supports digital image processing and machine learning. If the research is in collaboration with an institution or laboratory, the research location may include a healthcare institution.

RESULT AND DISCUSSION

Selection

The initial step in the KDD method begins with selecting the data to be used in this study. The data source comes from the Kaggle platform, consisting of microscopic images of human blood cells grouped by subtype, such as red blood cells (erythrocytes), white blood cells (leukocytes), and platelets. The total number of images reached 17,092, obtained from a hospital laboratory in Barcelona. This data covers eight cell categories: neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (such as promyelocytes, myelocytes, and metamyelocytes), erythroblasts, and platelets. Each image has a resolution of 360 x 360 pixels, is in JPG format, and has been labeled by a medical professional. All images were taken from individuals without infections or a history of hematological or oncological diseases.

This dataset has been equipped with labels to facilitate the training and testing of the model. These labels are useful for helping the system recognize and differentiate between various types of blood cells in normal peripheral conditions.

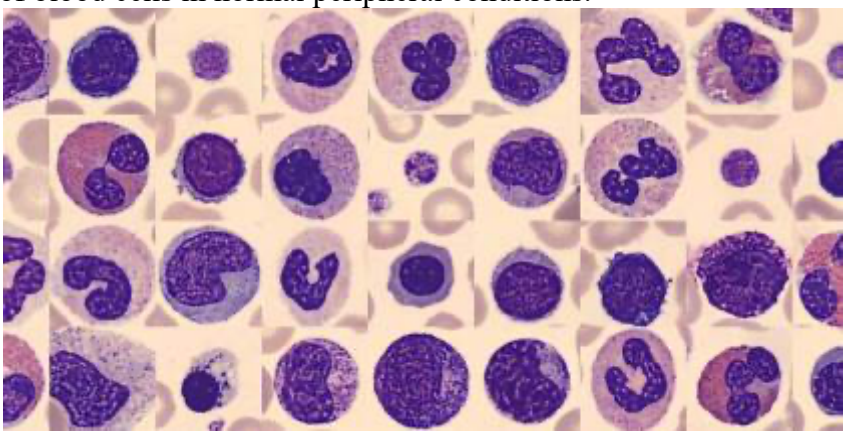
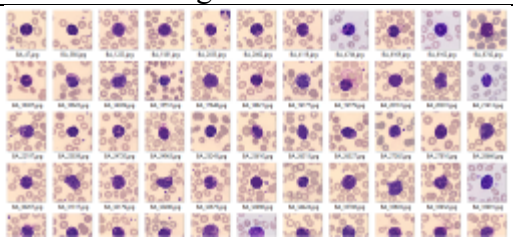
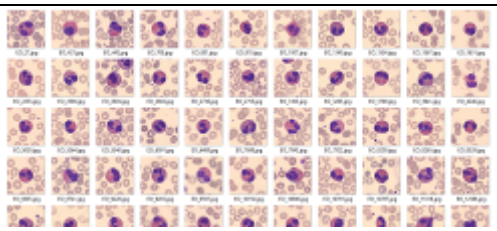
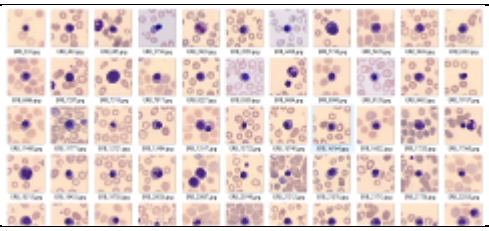
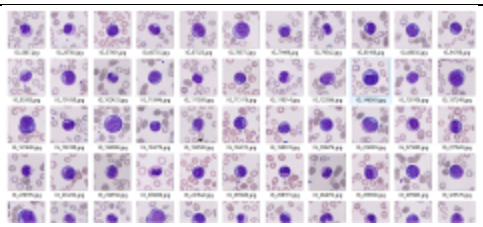
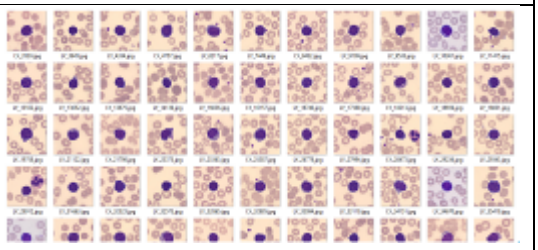
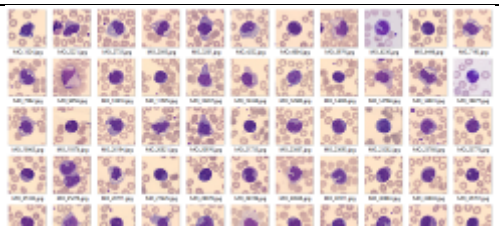
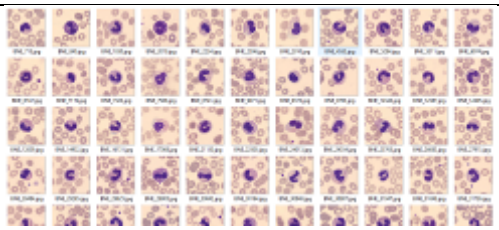
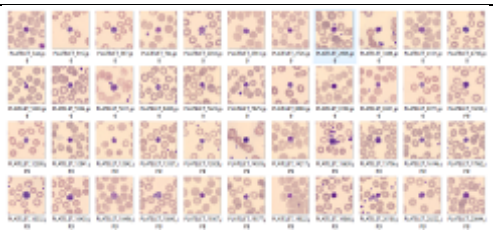


Figure 1. Selection

Table 1. Dataset Image

Name	Image Collection
Basophil	
Eosinophil	
Erythroblast	

IG	
lymphocyte	
Monocyte	
Neutrophil	
Platelet	

Source: Kaggle (Blood Cells Image Dataset, n.d.)

Table 2. Number of Datasets per Subtype

Basophil	Image 1218
Eosinophil	Image 3117
Erythroblast	Image 1551
IG	Image 2895
Lymphocyte	Image 1214
Monocyte	Image 1420
Neutrophil	Image 3329
Platelet	Image 2348

The collected data will then be prepared for preprocessing so that it can be used for further analysis.

Preprocessing

At this stage, the collected image data undergoes a process of adjustment and quality improvement to make it suitable for further analysis. This process is carried out using TensorFlow, starting with the data augmentation stage. One of the main steps is normalizing pixels to a value range between 0 and 1 using a rescaling technique. Additionally, several transformations are applied to enrich the data variation, such as flipping the image horizontally using RandomFlip, randomly rotating the image using RandomRotation, zooming in on objects in the image using RandomZoom, and adjusting the contrast level using RandomContrast. All of these processes aim to increase the diversity of the training data while maintaining image quality so that the model can learn optimally.

```
augment = tf.keras.Sequential([
  tf.keras.layers.Rescaling(1./255),
  tf.keras.layers.RandomFlip("horizontal"),
  tf.keras.layers.RandomRotation(0.1),
  tf.keras.layers.RandomZoom(0.1),
  tf.keras.layers.RandomContrast(0.1),
])

tf_model = tf.keras.Sequential([
    augment,
    tf.keras.layers.Conv2D(32, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(32, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(32, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(NUM_CLASSES)
])
```

Figure 2. Preprocessing

Preprocessing Results:

Model: "sequential_1"

Layer (type)	Output Shape	Param #
sequential (Sequential)	(None, 64, 64, 3)	0
conv2d (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 32)	9,248
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 32)	0
conv2d_2 (Conv2D)	(None, 12, 12, 32)	9,248
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 32)	0
flatten (Flatten)	(None, 1152)	0
dense (Dense)	(None, 128)	147,584
dense_1 (Dense)	(None, 8)	1,032

Total params: 168,008 (656.28 KB)
Trainable params: 168,008 (656.28 KB)
Non-trainable params: 0 (0.00 B)

Gambar 3. Hasil *Preprocessing*

```
Epoch 1/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 39s 71ms/step - accuracy: 0.4370 - loss: 1.4523 - val_accuracy: 0.7384 - val_loss: 0.7894
Epoch 2/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 3s 7ms/step - accuracy: 0.7506 - loss: 0.6627 - val_accuracy: 0.8057 - val_loss: 0.5610
Epoch 3/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 6s 9ms/step - accuracy: 0.8147 - loss: 0.5064 - val_accuracy: 0.7972 - val_loss: 0.5878
Epoch 4/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 5s 7ms/step - accuracy: 0.8459 - loss: 0.4274 - val_accuracy: 0.8464 - val_loss: 0.4778
Epoch 5/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 3s 7ms/step - accuracy: 0.8738 - loss: 0.3593 - val_accuracy: 0.8575 - val_loss: 0.4216
Epoch 6/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 3s 8ms/step - accuracy: 0.8821 - loss: 0.3230 - val_accuracy: 0.8209 - val_loss: 0.5259
Epoch 7/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 5s 8ms/step - accuracy: 0.8990 - loss: 0.2789 - val_accuracy: 0.7794 - val_loss: 0.7031
Epoch 8/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 4s 10ms/step - accuracy: 0.9063 - loss: 0.2676 - val_accuracy: 0.7744 - val_loss: 0.7554
Epoch 9/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 4s 8ms/step - accuracy: 0.9149 - loss: 0.2430 - val_accuracy: 0.7993 - val_loss: 0.6908
Epoch 10/10
428/428 ━━━━━━━━━━━━━━━━━━━━ 5s 7ms/step - accuracy: 0.9204 - loss: 0.2179 - val_accuracy: 0.8382 - val_loss: 0.4869
```

Figure 4. Preprocessing Epoch Results

In the data augmentation process, the Convolutional Neural Network (CNN) method was used to build the model architecture. The model consists of three main Conv2D layers, each with 32 filters of size 3x3, equipped with ReLU activation functions to introduce non-linearity to the detected features. Each Conv2D layer is followed by a MaxPooling2D layer, which functions to reduce image dimensions while preserving important information, thereby speeding up the training process. After feature extraction, the data is flattened using a Flatten layer to feed into the Dense layer. In this final stage, the Dense layer contains 128 neurons with ReLU activation to optimize the model's learning before moving on to classification.

The model is designed using the Adam optimizer, which is known for its efficiency in speeding up training and achieving optimal results. To evaluate the classification performance, the Sparse Categorical Crossentropy loss function is used, which is suitable for multi-class classification problems with numeric labels. In addition, the model is evaluated using a confusion matrix or other evaluation metrics to determine how accurately the model makes predictions. Before training, the model is compiled by setting the input dimensions

according to the image size. A summary of the model architecture is then displayed, including the number of parameters used in each network layer.

The preprocessing process utilizes the prepared dataset, namely train_ds_tuned as the training data and val_ds_tuned as the validation data. The model is then trained for a predetermined number of epochs. During each epoch, the system displays the loss and accuracy values for both training and validation data. The training results show that the model effectively learns the patterns in the data, as evidenced by decreasing loss values and increasing accuracy for both datasets, indicating strong model performance during training.

Transformation

The transformation stage aims to extract important features from the preprocessed blood cell images. These features include shape, texture, and color, which are key elements in distinguishing between blood cell subtypes. This information is later used in the classification process to accurately identify the type of blood cell. The transformation components include:

1. Conv2D (Convolutional Layer) – Detects local features in the image using filters.
2. MaxPooling2D – Reduces data dimensions through downsampling while retaining the main features.
3. Flatten – Converts the 2D or 3D convolution results into a 1D array for use in other classification models.

```
# Alternative feature extractor for LDA
inputs = tf.keras.Input(shape=(img_height, img_width, 3))
x = augment(inputs)
x = tf_model.layers[1](x)  # First Conv2D layer
x = tf_model.layers[2](x)  # MaxPooling2D layer
x = tf_model.layers[3](x)  # Second Conv2D layer
x = tf_model.layers[4](x)  # MaxPooling2D layer
x = tf_model.layers[5](x)  # Third Conv2D layer
x = tf_model.layers[6](x)  # MaxPooling2D layer
x = tf_model.layers[7](x)  # Flatten layer
features_extractor = tf.keras.Model(inputs, x)

train_features = []
train_labels = []
for images, labels in train_ds:
    features = features_extractor(images).numpy()
    train_features.extend(features)
    train_labels.extend(labels.numpy())
```

Figure 5. Transformation

Transformation Results:

```
monocyte: 1420 images
ig: 2895 images
neutrophil: 3329 images
basophil: 1218 images
lymphocyte: 1214 images
erythroblast: 1551 images
eosinophil: 3117 images
platelet: 2348 images
```

Figure 6. Number of Images Per Subtype

Data Mining

After feature extraction, the Linear Discriminant Analysis (LDA) algorithm is applied as the primary classification technique. LDA simplifies the data by reducing the number of dimensions while simultaneously finding the best combination of features that can optimally separate each type of blood cell. The implementation process is as follows:

1. Build a discriminant function to differentiate subtypes.
2. Analyze the distribution between classes to determine the optimal line.
3. Train the model with training data to recognize patterns in the subtypes.
4. Test the model with test data to determine whether it performs well and can classify.

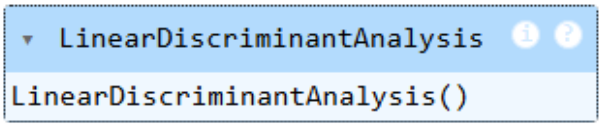


Figure 7. LDA

Evaluation

After the model is implemented, its performance is evaluated using the following evaluation metrics:

1. Accuracy, which measures the number of correct predictions.
2. Precision and recall, which measure how well the model can recognize each blood cell.
3. F1-score, which combines precision and recall to provide a comprehensive picture.

	precision	recall	f1-score	support
monocyte	0.79	0.84	0.81	232
ig	0.99	0.97	0.98	652
neutrophil	0.91	0.86	0.89	312
basophil	0.80	0.86	0.83	592
lymphocyte	0.88	0.89	0.88	240
erythroblast	0.85	0.81	0.83	274
eosinophil	0.97	0.95	0.96	647
platelet	0.99	0.98	0.98	469
accuracy			0.91	3418
macro avg	0.90	0.89	0.90	3418
weighted avg	0.91	0.91	0.91	3418

Figure 8. LDA Model Evaluation

Confusion Matrix Results:

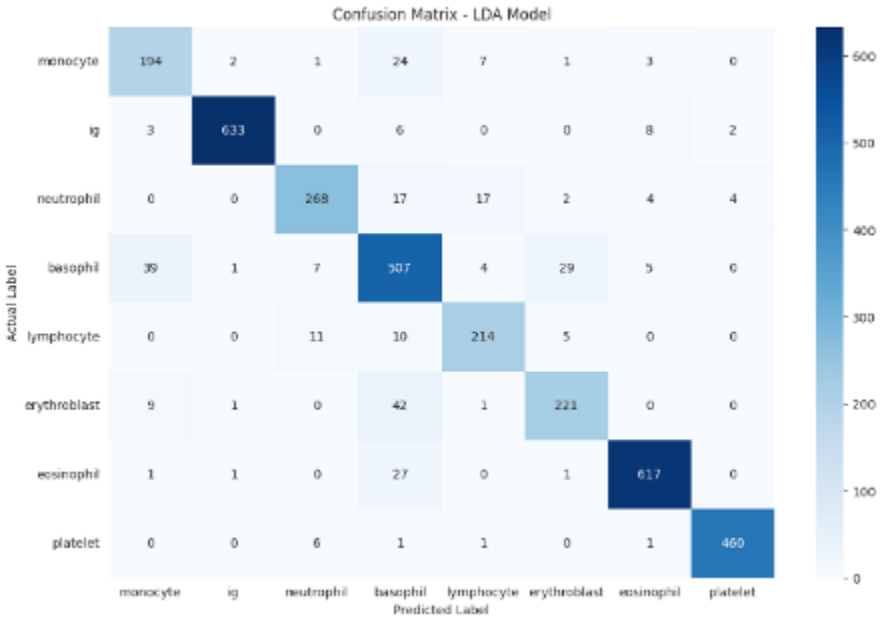


Figure 9. Evaluation

Knowledge

By classifying human blood cells using the Linear Discriminant Analysis method, useful information was obtained to more precisely identify each blood cell subtype. The model was evaluated by measuring accuracy, precision, recall, and f1-score, which together provide an indication of how well the model differentiates blood cell types. From this analysis, cell morphological patterns can be more clearly recognized, allowing characteristics such as the size, shape, and texture of each cell to be mapped. Furthermore, the distribution of data between classes also provides an understanding of the feature separation capability, which is useful for early detection of potential blood disorders.

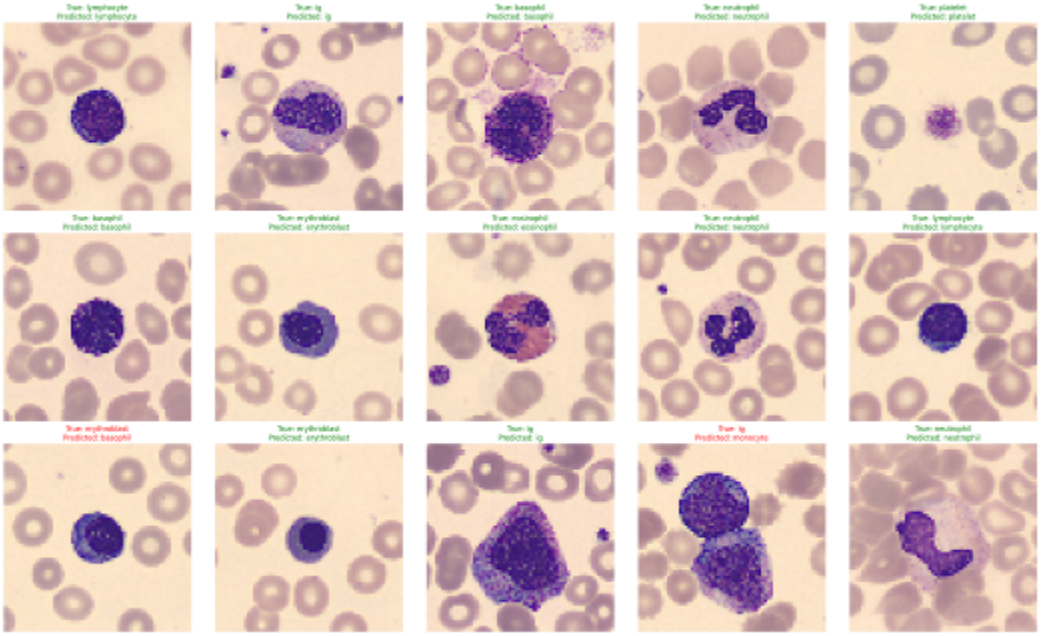


Figure 10. Knowledge

Web view:

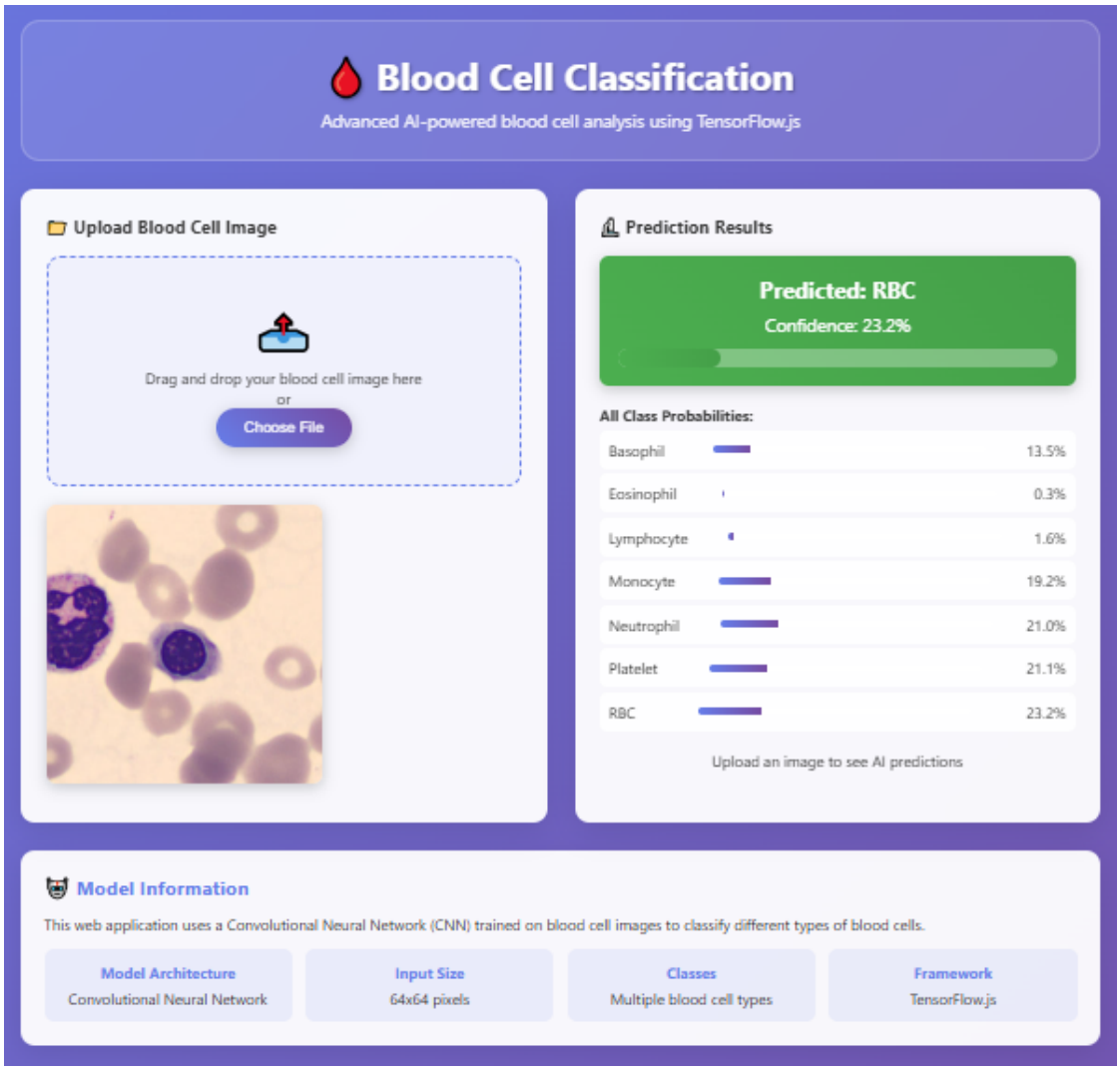


Figure 11. Program Results

CONCLUSION

Based on the research conducted on "Classification of Human Blood Cells by Subtype Using Linear Discriminant Analysis (LDA) on Microscopic Images," it can be concluded that the LDA method is capable of delivering good performance in classifying human blood cell images. The research process began with the collection of a microscopic blood cell image dataset, followed by a preprocessing stage to enhance image quality—such as normalization, contrast enhancement, and object segmentation. Subsequently, feature extraction was carried out, focusing on morphological features including shape, size, and texture of each blood cell.

Next, the Linear Discriminant Analysis (LDA) method was applied to reduce feature dimensionality while maximizing inter-class separation. The results showed that LDA was able to clearly distinguish between blood cell subtypes, such as red blood cells, white blood cells, and platelets. The accuracy level achieved was quite satisfactory, with consistent average accuracy across various tests, indicating that LDA is an effective method for microscopic image classification of human blood cells.

Moreover, the use of LDA in this study proved to offer a relatively fast and simple classification process, as LDA is computationally less complex while still maintaining good accuracy. Hence, this method is suitable for implementation in automated classification systems that require fast identification processes, such as in clinical laboratory settings or preliminary diagnostic tools.

However, there are some limitations in this study, such as the limited dataset size and the relatively low variation in image conditions. This may cause the resulting model to be less robust when applied to larger datasets or images captured under highly variable lighting and resolution conditions.

REFERENCES

Alkafrawi, I. M. I., & Dakhell, Z. A. (2022). Blood Cells Classification Using Deep Learning Technique. *Proceedings - 2022 International Conference on Engineering and MIS, ICEMIS 2022*. https://doi.org/10.1109/ICEMIS56295.2022.9914281

Arviananta, R., Syuhada, S., & Aditya, A. (2020). Perbedaan Jumlah Eritrosit Antara Darah Segar dan Darah Simpan. *Jurnal Ilmiah Kesehatan Sandi Husada*, *12*(2), 686–694. https://doi.org/10.35816/jiskh.v12i2.388

Boudreaux, M. K., & Christopherson, P. W. (2020). Platelet Structure. *Schalm's Veterinary Hematology, Seventh Edition*, 658–666. https://doi.org/10.1002/9781119500537.CH76;SUBPAGE:STRING:ABSTRACT;WEBSITE:WEBSITE:PERICLES;CTYPE:STRING:BOOK

Gatc, J., & Maspiyanti, F. (2022). Prediksi Parasit Plasmodium pada Citra Mikroskopis Sel Darah Merah dengan Convolutional Neural Networks. *Jurnal Buana Informatika*, *13*(1), 31–41. https://doi.org/10.24002/jbi.v13i1.5007

Kouzehkanan, Z. M., Saghari, S., Tavakoli, S., Rostami, P., Abaszadeh, M., Mirzadeh, F., Satlsar, E. S., Gheidishahran, M., Gorgi, F., Mohammadi, S., & Hosseini, R. (2022). A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-021-04426-x

Lin, E., Fuda, F., Luu, H. S., Cox, A. M., Fang, F., Feng, J., & Chen, M. (2023). Digital pathology and artificial intelligence as the next chapter in diagnostic hematopathology. *Seminars in Diagnostic Pathology*, *40*(2), 88–94. https://doi.org/10.1053/J.SEMDP.2023.02.001

Meenakshi, A., Ruth, J. A., Kanagavalli, V. R., & Uma, R. (2022). Automatic classification of white blood cells using deep features based convolutional neural network. *Multimedia Tools and Applications*, *81*(21), 30121–30142. https://doi.org/10.1007/s11042-022-12539-2

Prasthio, R., Yohannes, Y., & Devella, S. (2022). Penggunaan Fitur HOG Dan HSV Untuk Klasifikasi Citra Sel Darah Putih. *Jurnal Algoritme*, *2*(2), 120–132. https://doi.org/10.35957/ALGORITME.V2I2.2362

Wonohadidjojo, D. M. (2021). Perbandingan Convolutional Neural Network pada Transfer Learning Method untuk Mengklasifikasikan Sel Darah Putih. *Ultimatics : Jurnal Teknik Informatika*, *13*(1), 51–57. https://doi.org/10.31937/TI.V13I1.2040

Zhu, F., Gao, J., Yang, J., & Ye, N. (2022). Neighborhood linear discriminant analysis. *Pattern Recognition*, *123*, 108422. https://doi.org/10.1016/J.PATCOG.2021.108422