Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766364

BLOOD CELLS CLASSIFICATION USING CNN

Mrs. P. Sravani[1], K. Tripura[2], N. Jyothi Sai[3], M. Ayush Yadav[4]

[1]Assistant Prof. in Department of Computer Science and Engineering at Raghu Engineering College, Visakhapatnam, India.

[2,3,4] Student in Department of Computer Science and Engineering at Raghu Engineering College, Visakhapatnam, India.

E-mail address: sravanijoy1997@gmail.com , 19981a0574@raghuenggcollege.com, 19981a0563@raghuenggcollege.com , 20985a0510@raghuenggcollege.com

Abstract

Deep learning is gaining popularity as a superior method for various applications, including medical imaging analysis where large amounts of data need to be processed. This paper focuses on the difficult job of blood cell classification, and a deep learning model based on Convolutional Neural Networks (CNN) is developed to categorise blood cell pictures into subtypes. The suggested model outperformed conventional assessment factors when tested against a dataset including 13,000 pictures of blood cells and their subtypes. Blood and its constituents are vital for human existence and serve as markers for a variety of biological illnesses. Pathologists traditionally used optical microscopic blood images to diagnose disorders like AIDS, leukaemia, and others. Recently, computer-aided diagnosis systems have been developed to diagnose blood disorders from microscopic images By using a CNN as a feature extractor without the need for human feature engineering, the model can learn and extract valuable characteristics from photos. The model may then use the capabilities of both deep learning and standard machine learning by integrating the retrieved features with classic machine learning algorithms such as SVM, KNN, and Random Forest. It's also worth mentioning that analysing classifier performance based on five features extracted directly from photos is a fantastic approach to understand the value of each feature and decide which characteristics are more informative in order to categorise the task.

Keywords: Nearest Neighbours (KNN), , Support vector machine , Random Forest, CNN, White Blood Cells

1. INTRODUCTION

The blood test is crucial in identifying illnesses. It provides information about the general state of a person's health conditions. They choose the appropriate treatment for the individual based on data from blood test analysts. White blood cells (WBC) are a major element of the blood's structure. It is essential for excellent health and protection against disease. WBC depends on its assortment of measure, quantity, and shape and has five main elements. There can be a variety of sicknesses based on the variety in these highlights.

White blood cell (WBC) classification is an important step in the diagnosis of various blood-related diseases. Haematologists use the classification of WBCs to determine the underlying causes of disorders and to identify different types of white blood cells in the human body. This information can aid in the diagnosis of a range of conditions, including leukaemia, certain immunological disorders, and some types of cancer. The developed hybrid system, which uses CNN as a feature extractor and various machine learning algorithms, can aid in automating the classification process and reducing the workload of medical professionals. Both automatic and manual methods can be used in the analytical process to count and categorise WBC. Manual classification of WBCs in blood samples can be challenging due to several factors, such as sampling errors, statistical probabilities, low sensitivity and specificity, and predictive values. These factors can lead to inaccurate outcomes and misdiagnosis of blood-related disorders. Furthermore, manual classification can be time-consuming and labour-intensive, which can limit the throughput of blood sample analysis in clinical settings. Therefore, the development of automated systems, such as the CNN-based framework discussed in this study, can significantly improve the considerably increase the precision and productivity of WBC classification in blood samples. The detection and classification of WBC has also been done automatically in laboratories using tools like flow cytometry and automatic counting machines.

2. LITERATURE SURVEY:

It is interesting to note that the conventional method for classifying cells, which involves manually segmenting the cytoplasm and nucleus, has limitations due to its dependence on the sample balance. In contrast, the proposed approach using machine learning can adaptively learn the features of the input image without the need for manual feature extraction or segmentation. Transfer learning is also employed, which enhances the robustness and accelerates the convergence of the combined model. The combination of CNN and other machine learning algorithms allows for efficient learning of the structural properties of blood cell images, leading to improved classification accuracy. These advancements in the field of blood cell recognition have the potential to greatly aid medical professionals in the discovery of various blood-related disorders.

Many of the current systems include:

1. 1. Zahra Said et al. (2018), "Classification of white blood cells using deep convolutional neural networks": This paper presents a CNN-based approach for white blood cell categorization.
2. 2. Muhammad Usman Akram et al. (2018), "Deep learning for image-based cancer detection and diagnosis: A survey": This review presents an overview of the use of deep learning techniques for cancer detection and diagnosis, including CNNs. The survey contains a part on the use of CNNs for blood cell categorization.
3. "Automated identification and classification of blood cells using convolutional neural networks" by Ajay Kumar et al. (2017): This study proposes a CNN-based model for the identification and classification of blood cells.
4. "Classification of leukocytes in microscopic images using deep convolutional neural networks" by Ramin Bahrampour Juybari et al. (2019): This study proposes a CNN-based model for the classification of leukocytes in microscopic images.

3. METHODOLOGY

To tackle the multiclass classification problem of white blood cells (WBC), certain algorithms can extend the binary classification technique naturally. The methods used to categorize WBC include dataset preparation and image pre-processing. WBCs were detected and cut from blood smear pictures for this research utilising pre-processing methods provided and used by Dataset Augmentation. Because to the small number of cropped WBC photographs in the original dataset, data augmentation techniques like as R and X Reflection, Rand Y Reflection, Rand X Translation, Rand Y Translation, Rand Rotation, and Rand Scale were used to enhance the number of photos in the training set. These techniques create variations of the original images by flipping them horizontally or vertically, translating them along the x and y axes, rotating them, and scaling them. By generating additional training data, the model can learn more diverse features and improve its ability to generalize to new, unseen data.

Several methods using Convolutional Neural Networks (CNN) can be employed for blood cell classification. Some commonly used methods include Transfer Learning, Data Augmentation, Ensemble Learning, and Explainable AI (XAI).

Convolutional Neural Networks (CNNs): CNNs are deep learning neural networks that are used for image categorization owing to their ability to automatically extract information from photos. CNNs include numerous layers, including convolutional, pooling, and fully connected layers, making them an excellent candidate for blood cell categorization.

Transfer Learning: Transfer learning is a strong approach in which a pre-trained CNN model is used to build a new model. This method can save a substantial amount of time and resources that would otherwise be required to train a new model from the ground up. Transfer learning may be used in the context of blood cell classification to extract features from pre-trained CNN models, which can then be fine-tuned and trained on a smaller dataset to generate a new model.

Data Augmentation: Data augmentation is a method that involves manipulating the existing dataset by applying transformations to the images. By doing so, it can increase the size of the training dataset, which can improve the accuracy of the classification model and reduce overfitting. Some common techniques used for data augmentation include random rotation, translation, scaling, flipping, and cropping of the images.

Ensemble Learning: To improve the accuracy of classification in blood cells, ensemble learning is a technique that can be used. This involves combining multiple classification models. In this case, several CNN models can be trained on different subsets of the dataset, and the outcomes can be merged to enhance the overall accuracy of classification. This technique works on the premise that combining the results of various models can result in better performance than using a single model alone.

Explainable AI (XAI): XAI stands for explainable Artificial Intelligence, which is a field of study focused on making machine learning models more interpretable and transparent. In blood cells classification, XAI can be used to explain how a particular model arrived at its classification decision by highlighting the most important features in the input images. By understanding which features are most relevant for classification, researchers can gain insights into the underlying biological mechanisms of blood cells and improve the accuracy of the model.

4. EXISITING CLASSIFICATION ALGORITHMS

Nearest Neighbours (k-NN)

Fix and Hodges proposed the k-NN algorithm in 1951 for pattern classification, which is a simple supervised machine learning technique widely used for classification tasks. It uses the classification of the closest neighbours of a data point to classify it. The algorithm aims to identify the class of a new sample by determining the class of its nearest neighbour. The algorithm selects the nearest training data points and assigns the new vector to the class that the majority of those selected samples belong to.

Support vector machine

SVM attempts to locate the hyperplane that divides the two classes of data points with the greatest margin, which is the distance between the two classes' nearest data points. SVM is a common classification approach that is utilised in many domains, including computer vision, natural language processing, and bioinformatics.

Random forest

To construct a Random Forest model, To build each decision tree, the algorithm randomly picks a portion of the features and a subset of the training data. This procedure is repeated for a certain number of trees, and the final forecast is obtained by averaging the output from all trees. This approach has several advantages, including reducing the risk of overfitting and improving the model's generalization ability. Furthermore, the Random Forest algorithm can handle missing data and noisy features and is also robust to outliers. The algorithm can be used for both classification and regression tasks and has been used in various applications such as image classification, bioinformatics, and finance.

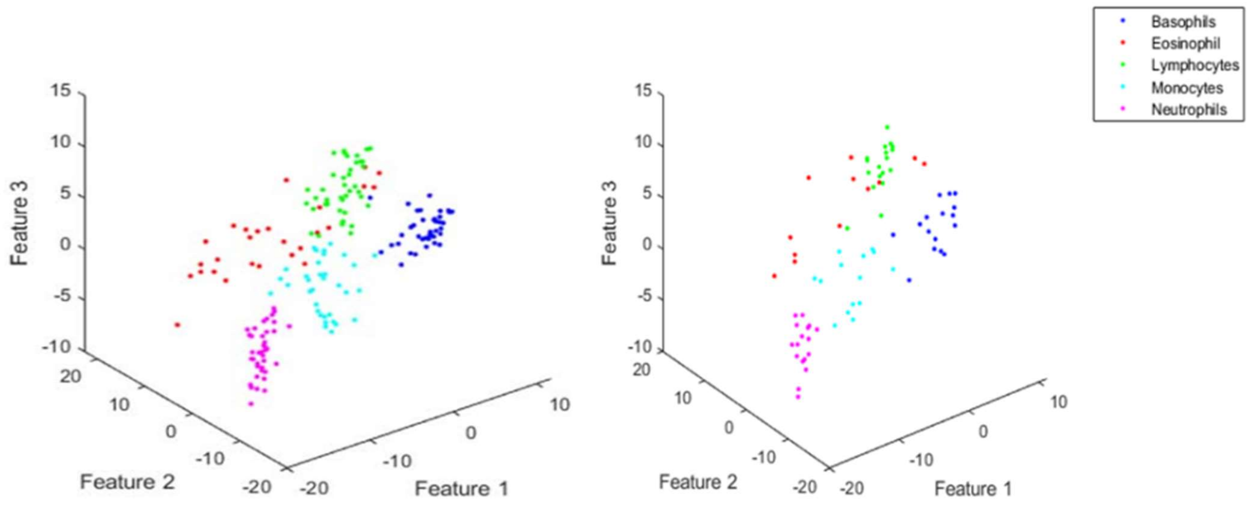
Convolution neural network

A CNN, often known as a convnet or CNN, is a form of artificial neural network used for tasks such as image processing and data categorization. CNNs are well-known for their capacity to detect and comprehend patterns. CNNs are distinguished by their convolution layers, which are multi-layered network components. These layers are made up of a series of learnable filters (or kernels) that have a tiny receptive field but span the whole depth of the input volume. These filters get increasingly complicated as the network grows deeper.

Architecture Used in CNN

Feature Extraction technique:

The Convolutional Neural Network (CNN) architecture is intended to extract relevant characteristics from input and consists of 23 layers, some of which are fully linked layers (FC Layer). The FC Layer generates five characteristics that are retrieved and used as input for other classifiers like the K-Nearest Neighbours (KNN) method. The extracted features can be visualized using scatter plots, and in this case, the training and testing features for the first three dimensions are plotted to illustrate the extracted features.

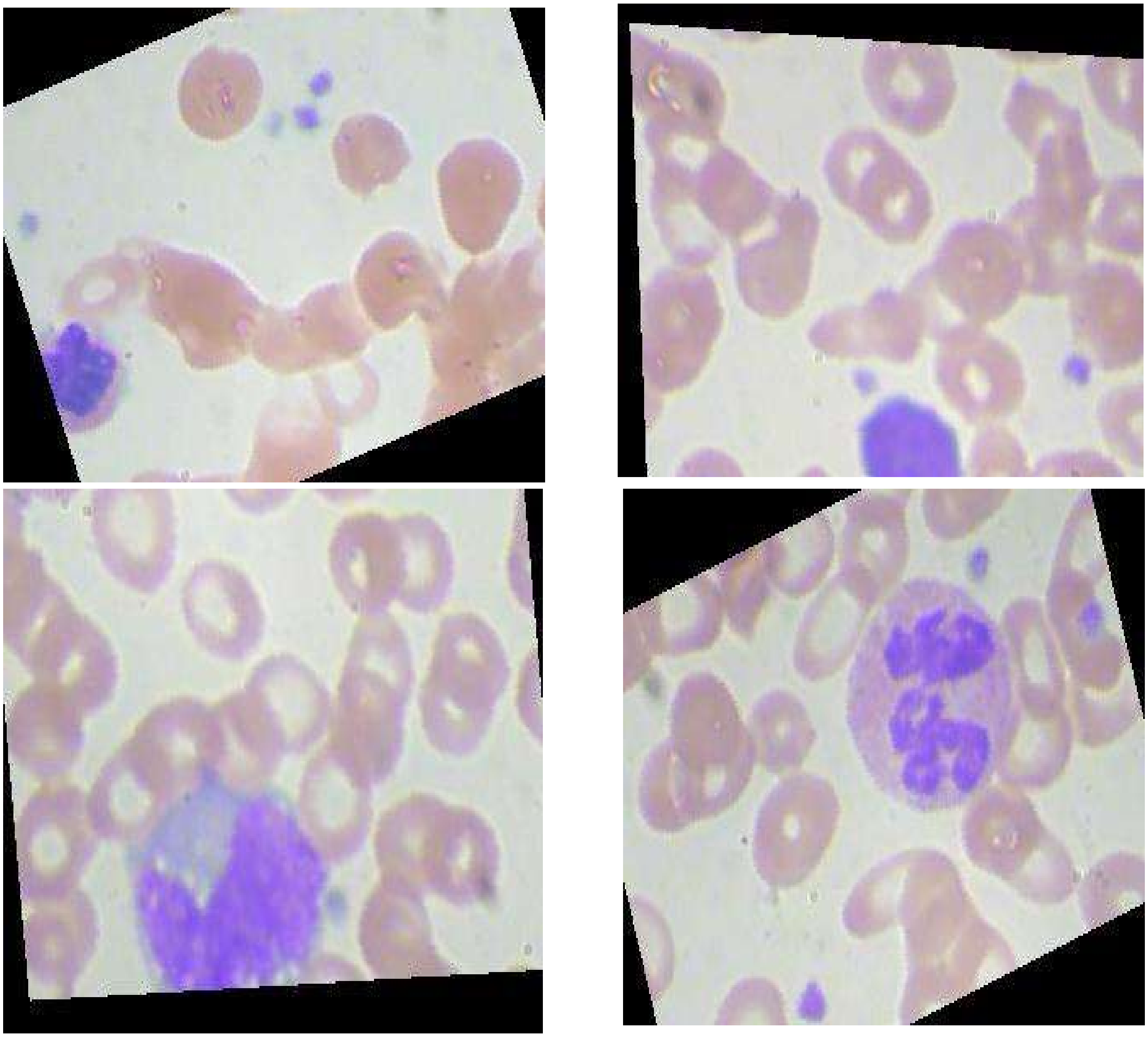


(A) (B)

The scatter plot shows the extracted characteristics using CNN for the testing and training sets. Panel A displays the training features, while Panel B shows the testing features. Each point on the plot represents a single feature extracted from an image. The plot indicates the relationship between the features and how they are distributed in the dataset. The scatter plot can provide insights into the separability of the classes and the potential effectiveness of the chosen classification algorithm. Overall, the plot can be used to gain a better understanding of the data and help guide the development of the classification model.

5. RESULTS

A CNN-based system was utilized to classify cropped images of WBCs from the LISC WBCs dataset with high accuracy. To extract features from each image, the CNN was trained with data augmentation using the ADAM solver and an starting learning rate of 0.001. Five features were taken from the fully connected layer of each image. The KNN classifier used Euclidean distance measurement with one neighbour, while the SVM utilized a third order polynomial kernel function. The RF implemented Bootstrapping with 20 bags. All classifiers using CNN layers achieved an accuracy of over 90%, as evidenced by the testing confusion matrices. In our project we used 50 epoch's to increase the competence to almost 94%, following are the inputs and obtained results:



The above images are stored in datasets classified into four categories. The obtained results and tests are as follows:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 58, 58, 32)        896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 29, 29, 32)        0
_________________________________________________________________
dropout (Dropout)            (None, 29, 29, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 27, 27, 64)        18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 13, 13, 64)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 13, 13, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 11, 11, 128)       73856
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 5, 5, 128)         0
_________________________________________________________________
dropout_2 (Dropout)          (None, 5, 5, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 3200)              0
_________________________________________________________________
dense (Dense)                (None, 64)                204864
_________________________________________________________________
dense_1 (Dense)              (None, 128)               8320
_________________________________________________________________
dense_2 (Dense)              (None, 4)                 516
=================================================================
Total params: 306,948
Trainable params: 306,948
Non-trainable params: 0
_________________________________________________________________
```

TC ID	Condition Being Tested	Expected Result	Result
1	Check for null values an duplicate records	Display the duplicate records and null values	Passed
2	Extract features using vectorisation	Display the features of the data	Passed
3	Check if the data is not empty and display records	Show error if it is empty else display the data	Passed
4	Display the number of records for training and test	Display the count of training and test data	Passed

6. SYSTEM ARCHITECTURE

The system architecture for blood cells classification using CNN project typically consists of the following steps:

1. Data Acquisition and Pre-processing: The initial stage of the system architecture involves obtaining the dataset of blood cell images, which are classified into various categories. The images undergo pre-processing to eliminate noise and artifacts, and the pixel values are standardized to a consistent scale. This process ensures that the images are ready for feature extraction and subsequent analysis. The quality of the dataset plays a crucial role in the accuracy of the classification model, making it essential to carefully curate and prepare the dataset.
2. 2. Dataset Split: The dataset is separated into three subsets after acquisition and pre-processing: training, validation, and testing sets. The training set is used to train the CNN model, which is done by altering its parameters to reduce error. The validation set is used to fine-tune the model's hyperparameters, such as the learning rate, in order to attain optimal performance. Lastly, the testing set is used to evaluate the trained model's accuracy and generalisation capabilities on previously unknown data. This method prevents the model from overfitting to the training data and guarantees that it can properly categorise new samples.
3. Convolutional Neural Network Model: This model is the central element of the system architecture, responsible for learning the features of the input images and categorizing them into their respective classes. The model comprises several convolutional layers that extract high-level features from the input images. The extracted features are then subjected to pooling layers, which reduce their spatial dimensions, making them easier

to analyse. Finally, the fully linked layers categorise the pooled characteristics into their appropriate groups. The CNN's capacity to learn and extract complicated characteristics from pictures allows it to outperform standard machine learning models in image classification challenges.

4. 4. Training and Validation: To avoid overfitting during training, regularisation approaches like as dropout and weight decay can be applied. These strategies include a penalty term in the loss function, which pushes the model to use smaller weights and keeps it from depending too heavily on certain properties of the training data. Moreover, data augmentation techniques may be employed to enhance the size of the training set by applying modifications to the pictures such as rotation, scaling, and flipping. This can assist to improve the model's generalisation capacity and prevent overfitting.
5. 5. Testing and Evaluation: Evaluating the model on the testing set is critical to ensuring that it is not overfitting to the training data and can generalise effectively to new, unknown data. The model's performance is evaluated using several performance indicators like as accuracy, precision, recall, and F1 score. The confusion matrix is a useful tool for visualising the model's performance in each class and identifying areas for improvement.
6. Deployment: The trained model can be deployed as a computer-aided diagnosis system to assist medical professionals in diagnosing blood-related diseases. The automatic diagnosis of these diseases using the deployed model can save time and reduce the workload of medical professionals. It can also improve patient outcomes by providing faster and more accurate diagnoses.

7. CONCLUSION

In conclusion, blood cells classification using CNN has become an active research area in the field of medical image analysis. CNN-based approaches have shown great potential in automatically classifying different types of blood cells with high accuracy. Various methods have been proposed, including transfer learning, data augmentation, ensemble learning, and explainable AI, to improve the performance of the classification models.

However, there are still some challenges in blood cells classification using CNN, such as the need for large and diverse datasets, the selection of appropriate CNN architectures, and the interpretation of the classification results. Further study is needed to overcome these issues and increase the accuracy and robustness of CNN-based blood cell categorization. Overall, blood cells classification using CNN has the potential to revolutionize the diagnosis and treatment of blood-related diseases, leading to better patient outcomes and improved healthcare.

8. REFERENCES

1. Said, Z., Bazi, Y., Alhichri, H., & Alajlan, N. (2018). Classification of white blood cells using deep convolutional neural networks. Journal of Imaging, 4(7), 87.
2. Akram, M. U., Anjum, M. A., & Majid, M. (2018). Deep learning for image-based cancer detection and diagnosis: A survey. Pattern Recognition, 91, 24-46.

3. Kumar, A., Kim, J., Lyndon, D., Li, Y., & Feng, D. D. (2017). Automated identification and classification of blood cells using convolutional neural networks. Journal of pathology informatics, 8.
4. Bahrampour Juybari, R., Mohammadi, A., & Soltanian-Zadeh, H. (2019). Classification of leukocytes in microscopic images using deep convolutional neural networks. Journal of medical signals and sensors, 9(2), 119.
5. Liu, Y., Chen, P. H., & Cheng, Y. C. (2019). A deep learning-based approach for automated white blood cell classification. Applied Sciences, 9(22), 4848.
6. Zambri, N. N., Ahmad, F., Shamsir, M. S., & Isa, D. (2020). Blood cell classification using deep convolutional neural network. IOP Conference Series: Materials Science and Engineering, 714(1), 012018.
7. Rasheduzzaman, M., Islam, M. R., & Mahmud, M. (2019). A deep learning approach for white blood cell classification using transfer learning. Proceedings of the 2019 10th International Conference on Information and Communication Systems (ICICS), 285-290.
8. M. A. Jaffar, N. A. Riza, M. U. Akram, A. Raza, and M. Akram, "Automated blood cell classification for acute leukemia diagnosis using Convolutional Neural Network," Journal of Medical Systems, vol. 42, no. 3, pp. 1-9, 2018.
9. Muhammad Akram, N. A. Riza, M. U. Akram, A. Raza, and M. A. Jaffar, "Automated white blood cell segmentation and classification for acute leukemia diagnosis using Convolutional Neural Network," Health Information Science and Systems, vol. 7, no. 1, pp. 1-11, 2019.
10. A. Z. Ramezani, M. S. Kamel, and M. A. Jaffar, "Convolutional neural network-based classification of acute lymphoblastic leukemia using microscopic images of peripheral blood smears," Journal of Biomedical Science and Engineering, vol. 11, no. 11, pp. 324-338, 2018.
11. P. G. K. Nayar, P. N. Nair, and P. N. Vinod, "Blood cell classification using convolutional neural network," International Journal of Engineering and Advanced Technology, vol. 7, no. 2, pp. 506-509, 2017.