Deep Learning Techniques in Bone Marrow Cell Classification: A Review

Rishabh Hanselia[1], Dilip Kumar Choubey*2[0000-0002-1233-7159], Neeraj Kumar Yadav[3]

[1,*2,3] Department of Computer Science & Engineering, Indian Institute of Information Technology Bhagalpur, Bihar, India
{[1]rishabhhanselia@gmail.com, [1]rishab.cse.2102020@iiitbh.ac.in
*[2]dilipchoubey_1988@yahoo.in, *[2]dkchoubey.cse@iiitbh.ac.in,
[3]neeraj.240201020@iiitbh.ac.in, [3]neerajyadavkhg@gmail.com}

Abstract. Evaluating bone marrow cells is essential for diagnosing hematological disorders, the current manual procedure has many inherent problems. One significant hurdle is the complexity and variability of cell morphology, making accurate classification difficult even for experienced experts. Moreover, subjectivity introduces inconsistencies among individuals, leading to unreliable results. Manual classification is also time-consuming, constrained by limited training resources, and prone to intra-observer variability. The development of automated AI powered systems, standardized protocols and allocation of more resources are all key components of efforts to increase the accuracy and speed of bone marrow cell classification. The authors have evaluated and thoroughly analyzed previous studies on bone marrow cell classification that have used a variety of deep learning and machine learning approaches. This study provides an oversight of the existing deep learning and machine learning being used to aid and even improve traditional cytology.

Keywords: Deep Learning, Bone Marrow Cell, Cytology, CNN, Accuracy, AUC

1 Introduction

Bone marrow cells are essential because they are responsible for both making and developing the different kinds of blood cells in the body. They encompass Red Blood Cells or RBC (erythrocytes), responsible for transporting oxygen, White Blood Cells or WBC (leukocytes), acting in immune system defenses (thrombocytes), necessary for blood coagulation process. Haematopoiesis is the process by which cells transform and develop into all the various types of blood cells takes place in the bone marrow. Under a microscope closely examining these bone marrow cells provides crucial hints for diagnosing a variety of blood related illnesses.

To classify bone marrow cells, it simply involves grouping cell types based on their unique features and roles. The important kinds are RBC, WBC, and platelets. The way RBCs are categorized depends on their stage of development and the amount of

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2362725

hemoglobin they contain. Neutrophils, lymphocytes, monocytes, eosinophils and basophils are among the many different subtypes that make up the WBC population. The two primary characteristics that determine a platelet's classification are its size and the presence of internal granules. The cells must be precisely categorized using instruments like flow cytometry, immunohistochemistry and microscopes in order to diagnose haematological disorders. This accurate classification helps doctors evaluate what the cells are made of, find cells that look abnormal and figure out the disease's stage and seriousness.

There are a number of significant difficulties with the manual approach to bone marrow cell classification. Because of the intricate shapes of the cells it is intrinsically laborious, subjective and prone to high variability. The inconsistent results are major problem technicians frequently disagree (intra- and inter-observer variability), which is made worse by inadequate training. These issues have led researchers to concentrate their efforts on developing artificial intelligence-based automated systems and tools. The main goal of these new systems is to make bone marrow cell classification more accurate, more consistent, and much faster. All things considered, these developments promise to significantly improve diagnosis in the study of blood disorders (hematopathology). This occurs because they eliminate human bias, provide reliable results and address the main shortcomings of manually classifying cells.

The remaining portions of the manuscript is organized as follows: Literature Review is described in section 2, Discussions are presented in section 3.

2 Literature Review

Song et al. [1] addressed cell analysis with a synchronized deep autoencoder network in their groundbreaking study. The system was specifically made to process images from bone marrow histology while simultaneously identifying and classifying cells. To create an automated system, Kimura et al. [2], combined two distinct algorithms, the Convolutional Neural Network (CNN) and XGBoost. This system's primary purpose was to distinguish between two dangerous conditions: aplastic anemia (AA) and myelodysplastic syndrome (MDS). Morphogo is an automated analysis system for bone marrow cell classification developed by Fu et al. [3]. The computational basis of this system is a deep, 27-layer Convolutional Neural Network (CNN). Wang et al. [4] sought to automatically differentiate between AML, MDS and AA. They used a CNN model that had been pre-trained on the sizable Image-net dataset to examine bone marrow smear images in order to complete this diagnostic task. An automated system to assess the shape (morphology) of bone marrow cells was developed by Matek et al. [5]. This was accomplished by using particular CNN classifiers, specifically a sequential network architecture and a ResNeXt. Choi et al. [6] employed a dual-stage CNN to develop a systematic technique for the WBC differential count. The VGGnet architecture served as the model for this network's structure. Tayebi et al. [7] created a thorough AI framework (an end-to-end architecture) in their study to enable AI-powered computational pathology analysis. Wu et al. [8] created BMSNet, a cutting-edge deep learning system that enables quicker evaluation of bone marrow smears, to help hematologists diagnose and track diseases. Ananthakrishnan et al. [9] used a Siamese network as their main

method for automated bone marrow cell classification. By contrasting its performance with that of other hybrid machine learning models such as CNN combined with SVM and CNN combined with XGBoost they expanded the scope of their investigation. In order to create a novel diagnostic system based on cell morphology, Liu et al.'s [10] study used a Region-based CNN (R-CNN). This was developed to increase the precision of bone marrow cell detection. New datasets containing the distinct cell types present in bone marrow aspirate (BMA) were produced by Chandradevan et al. [11]. A machine learning algorithm that could identify and categorize the BMA cells was then created using these data. There are many review articles which give us idea to write this manuscript. Some of the articles are stated here. Hanselia and Choubey [12], have reviewed a thorough survey applying machine learning and data mining methods to forecast the mortality of patients receiving hematopoietic stem cell transplantation (HSCT). Srivastava & Choubey [13] have done a rigorous survey on heart disease. Sharma et al. [14] have done an extensive survey on breast cancer by using computational intelligence. Parthvi et al. [15] have done a rigorous study on leukemia by using machine learning and data mining approaches. Choubey et al. [16], [17] have done a rigorous analysis on diabetes classification on diabetes by using computational intelligence. Choubey et al. [18] used the Fuzzy Decision Tree (FDT) method for their classification. Further feature selection has been performed by using PSO_SVM followed by the same FDT classification method on diabetes. This step performs well compared to previous. Table 1 summarizes the existing research in the field of bone marrow cell analysis.

Table 1. A summary of the existing research on the identification of bone marrow cells

Ref. No.	Dataset Used	Techniques Used	Advantages	Limitations	Performance	Purpose
1	A collection of 52 images of bone marrow trephine biopsies from UHCW were used in the study; these images included samples from five ET and five PMF patient cases.	Set of three combined methodologies: the synchronized deep autoencoder network, the curve-support Gaussian model, and a neighborhood selection mechanism.	This approach greatly improves detection and classification accuracy thanks to a special, innovative neighborhood selection mechanism. By employing similar high-level feature representations to lower layer complexity, it also streamlines the network.	Convolutional layers can be added to the network to improve segmentation and classification capabilities. Better performance and quicker training will result from this enhancement.	Precision 0.9266, Recall 0.9674, and F-1 score 0.9466 (Other frameworks); Precision 0.8712, Recall 0.8879, and F-1 score 0.8795 (DL frameworks)	The goal is to design and develop a model that lessens the burden of managing the increase in computational complexity during digital histology training.
2	A collection of 3,261 peripheral blood (PB) smears, 1,165 were from patients with various	The system integrated the XGBoost (eXtreme Gradient Boosting)	Using peripheral blood (PB) smears, the researchers proposed the first fully automated CNN-driven system for the preliminary diagnosis of	Although the system is regarded as supplemental, it was able to diagnose MDS with over 90% accuracy. Its training patterns might not accurately reflect population variation because	Sensitivity 93.5%, specificity 96.0%, (DLS); sensitivity 96.2%,	The goal of experiment was to design deep convolutional neural networks capable of assisting in

	hematological disorders, was generated at Juntendo University Hospital in Tokyo, Japan, between the years 2017 and 2018.	algorithm for decision-making with a CNN-based Deep Learning Solution (DLS) for image recognition.	MDS. There is potential for this innovative tool to be incorporated into automated diagnostic systems for additional hematological disorders.	of the small sample size and single-center data. Importantly only one particular combination of DLS, CNN and XGB algorithms was tested in the study. .	specificity 100%; AUC 0.990 (MDS)	differentiating between myelodysplastic syndrome and aplastic anemia.
3	230 direct bone marrow aspirate smears were gathered from Army Medical University's Xinqiao Hospital's daily inspections and archives. The average age of the patients (154 females and 76 males) was \$44.02 \pm 21.22 years. Data gathering time – 26th nov 2018 to 30th jan 2019	Morphogo is an automated system that analyzes data using a 27-layer CNN.	The system can successfully detect and display blasts, dysplastic, and immature cells in bone marrow smears by using digital image acquisition and automated cell analysis. Hematology laboratories can benefit practically from this shift to a digital workflow, which greatly increases efficiency and offers a standardized examination process.	To improve the AI system, extensive training on a big, diverse dataset is essential. The incapacity of current instruments to precisely capture the morphological details of small erythroblasts is one of their limitations. Furthermore, one major diagnostic challenge that needs to be addressed is morphological recognition during burst analysis.	Accuracy 85.7%, sensitivity 69.4, and specificity 97.2%	A novel artificial intelligence (AI) system was created to automatically classify bone marrow cells and explore its potential clinical applications.
4	A cohort of 547 bone marrow smears specimen were used in the study with 115 samples sourced via the ASH Image Bank and 432 samples supplied by the clinic.	Two output layers for classification were demonstrated by the Imagenet pretrained CNN model, allowing for both two-class and three-class predictions.	Due to the lack of previous reports, the classification task is unique. Importantly, the model's robust performance on the external validation set indicates that it is still dependable and efficient despite current limitations.	Future research should concentrate on both disease progression and disease detection. Larger-scale research is required due to the study's small sample size. Differences between the augmented and original samples persisted despite the application of data augmentation, which might have had an impact on the outcomes. Analysis of MDS and AML subtypes was also hindered by the ASH Image Bank's limited data. Furthermore, the limited understanding of the features extracted by deep	AUC 0.985, Accuracy 91.4%, Sensitivity 0.992 (two classification), AUC 0.968, Accuracy 92.9%, Sensitivity 0.857 (three classification)	The goal of the study was to create a CNN model that could use bone marrow smears to automatically differentiate between acute myeloid leukemia (AML), myelodysplastic syndrome (MDS) and aplastic anemia (AA).

				learning made it impossible to fully interpret the model's misclassifications. The findings' generalizability is further limited by the absence of comprehensive patient data.		
5	Between 2011 and 2013, the MLL Munich Leukemia Laboratory collected expert-annotated single-cell images from 961 patients with hematological disorders.	Pretrained CNN classifiers: ResNeXt, sequential network architecture	The suggested model performs better than classifiers based on features.	To lessen the impact of label noise, semi-supervised or unsupervised methods should be applied. Their effectiveness has previously been demonstrated in fields such as cell cycle reconstruction and erythrocyte assessment.	0.95 precision (tolerant), 0.92 precision (strict), 0.85 recall (tolerant) and 0.71 recall (strict)	To enhance the automation of bone marrow (BM) cell morphology assessment through the use of contemporary image classification algorithms.
6	At Seoul National University Hospital's Department of Laboratory Medicine, 2,174 patch images were gathered.	The VGG-net model served as the model for the 16-layer Dual-stage CNN architecture that was used.	The proposed method exhibited exceptional classification performance, effectively categorizing raw images without the need for single cell segmentation or manual feature extraction. By employing Convolutional Neural Networks (CNN), the method demonstrated rotation and location invariance. These outcomes emphasize the potential of the proposed method as an automated system for white blood cell differential count, offering promising advancements in the field.	A comprehensive range of White Blood Cell (WBC) types, including lymphocytes and plasma-type cells, as well as different disease representations including leukemia stages, must be covered by the classification model. In the field of hematology, this comprehensive method guarantees accurate and efficient classification, allowing for better analysis and diagnosis. The network employed in this research was initially created for tasks involving generic picture classification. To meet the particular needs of the study, a distinctive architecture is required.	Accuracy 97.06%, Precision 97.13%, Recall 97.06%, and F-1 score 97.1%	The goal was to develop an automated WBC differential count system that would significantly lower the variation in results produced by human hematologists.
7	Digital WSI from 1,247 patients was collected over a one-year period. The diagnosis	Two models are used in the end-to-end AI architecture: a YOLO	For automated bone marrow cytology, a novel end-to-end AI architecture was developed. In a variety of clinical scenarios,	External validation is impeded by the scarcity of annotated public or academic digital pathology datasets, particularly for cytology and	Accuracy 0.97, ROC (AUC) 0.99, Precision 0.75, F1-Score 0.78	The goal is to facilitate the integration of AI into computational pathology, which may

	was the only basis for the annotation, and the images were processed without any identifying information.	model for detection and class assignment, and a DenseNet model for choosing Region of Interest (ROI) tiles.	the model demonstrated remarkable accuracy and precision in both ROI detection and object classification. This achievement demonstrates its potential to incorporate AI into computational pathology and increase the effectiveness and precision of hematology diagnosis.	hematopathology workflows. The models' propensity to overfit the local data is influenced by this scarcity.	average, 0.31 Log-average miss rate	eventually result in quicker and more accurate diagnoses for medical professionals.
8	Over the course of three years, from January 1, 2016, to December 31, 2018, 122 bone marrow smears were photographed.	The YOLO v3 architecture is used by BMSNet, a Convolutional Neural Network (CNN).	The findings of this study have the potential to significantly improve clinical practice by supporting the diagnosis and prognosis of numerous diseases.	Only six visiting employees were used to test the model's performance. Despite the fact that the initial images were taken by skilled staff, the study proposes that an automatic slide scanner is required for an ideal image acquisition process in order to eliminate the operator effect.	AUC 0.948 (5% of blasts), AUC 0.942 (20% of blasts)	The goal was to create the deep learning tool BMSNet to help hematologists interpret bone marrow smears in order to speed up diagnosis and track illness.
9	170,000 carefully annotated cell images made up the dataset. The bone marrow smear samples of 945 patients presenting various blood-related conditions were the source of the images.	Siamese network, integrated methods CNN + SVM and CNN + XGBoost.	Instead of focusing only on extracting features from individual images, the methodology emphasizes comparing and contrasting images, regardless of whether they have the same class label.	In order to determine the ROI in cell images and infer the model's reliance on convolutional units for encoding, future research will use YOLO and Grad-CAM. Researchers will also look into various image augmentation methods to address class imbalance.	Weighted Average Recall 92%, Validation Accuracy 84%, Accuracy 32% (CNN + SVM), Accuracy 28% (CNN + XGB), Accuracy 91%, and Weighted Average Recall for validation 91% (Siamese neural)	In order to achieve the best overall accuracy, a method to counteract the impact of imbalanced class on deep neural network training is sought.
10	The dataset used in this study included 4,451 images of real	Faster Region-Convolutional Neural	Compared to skilled diagnostic specialists, the recently created	The model may have been biased toward that particular population and imaging configuration because	Recall 0.710, precision 0.496,	The objective was to develop a novel morphological

	bone marrow cells.	Network (R-CNN)	system processed data faster.	it was trained on 4,451 fields from 70 patients at a single facility. This implies that the model might have trouble with uncommon samples.	F1-score 0.575	diagnosis system specifically for identifying bone marrow cells through a deep learning–based object detection approach.
11	A total of 17 patient samples of Wright-stained bone marrow aspirate (BMA) smears were obtained for routine patient care. These smears were deidentified and scanned for further analysis. Between June and August of 2015, the bone marrow laboratory at Emory University Hospital prepared the smears consistently.	Fast Region-Convolutional Neural Network (R-CNN)	The findings could significantly improve clinical practice by helping diagnose and prognosticate a variety of diseases.	Since cells from those cases were excluded from training, the study's classification accuracy for disease cases needs to be improved. Furthermore, bias toward well-preserved cytologic areas was introduced by the use of tiny Regions of Interest (ROIs). The advantages of using higher-resolution scanned images should be examined in future research, particularly as their clinical acquisition becomes more practical. Lastly, the study does not provide precise performance standards for the clinical validation of this new approach, which is still in its infancy.	AUC 0.959 (detection), AUC 0.982 (classification)	The objective was to create an automated system for bone marrow aspirate (BMA) sample differential cell counting (DCC), which is essential for accurately classifying hematologic disorders.

The performance of the models employed in the numerous studies that have been examined for this paper and referenced in Table 1 is graphically displayed in Fig. 1, 2, 3, 4, 5, 6, and 7. The following is a visual representation of the performance metrics: The accuracy is shown in Figure 1, the AUC in Figure 2, the precision in Figure 3, the recall in Figure 4, the F1 Score in Figure 5, the sensitivity in Figure 6 and the specificity in Figure 7. The paper reference number is shown in Figures 1 through 7 along with a number of performance metrics, including Accuracy, AUC, Precision, Recall, F1 Score, Sensitivity, and Specificity.

Fig. 1 represent the graphical comparison of accuracy for the mentioned techniques as used in Table 1.

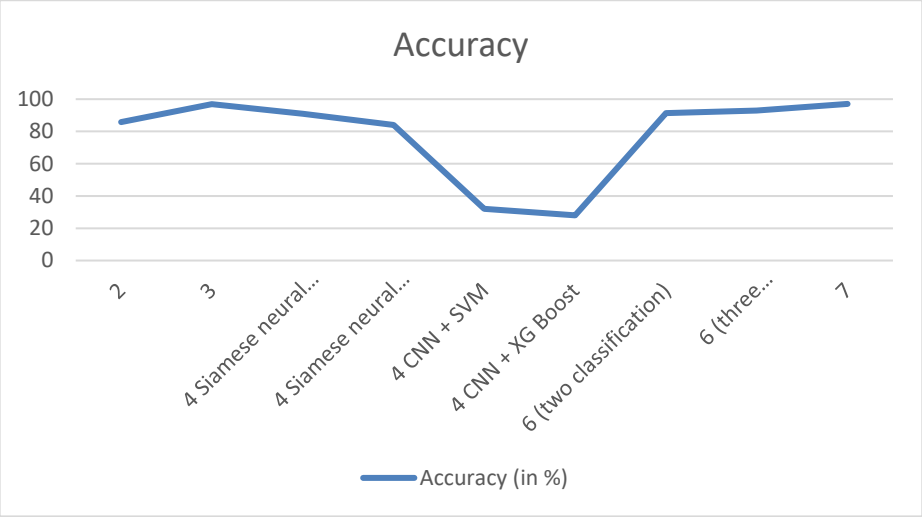


Fig. 1. Accuracy comparison for the listed techniques

The AUC (Area Under the Curve) results for the methods listed in Table 1 are graphically compared in Figure 2.

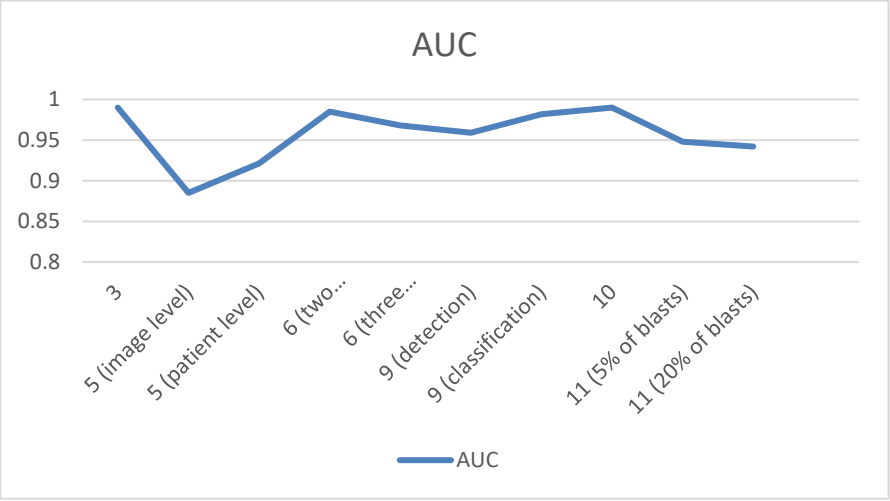


Fig. 2. Comparison of AUC for the mentioned techniques

Fig. 3 represent the graphical comparison of precision for the mentioned reference number as used in Table 1.

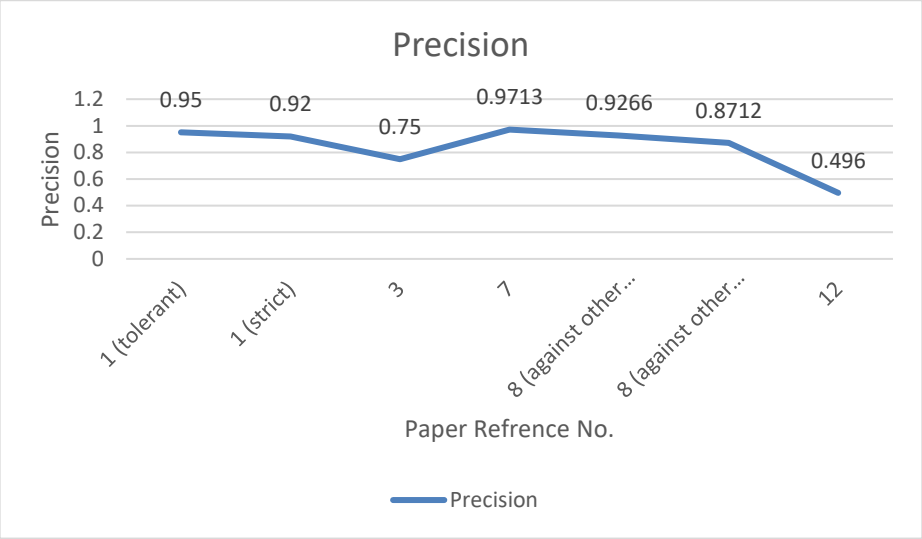


Fig. 3. Comparison of Precision for the mentioned paper Ref. No. as in Table 1

Fig. 4 represent the graphical comparison of recall for the mentioned reference number as used in Table 1.

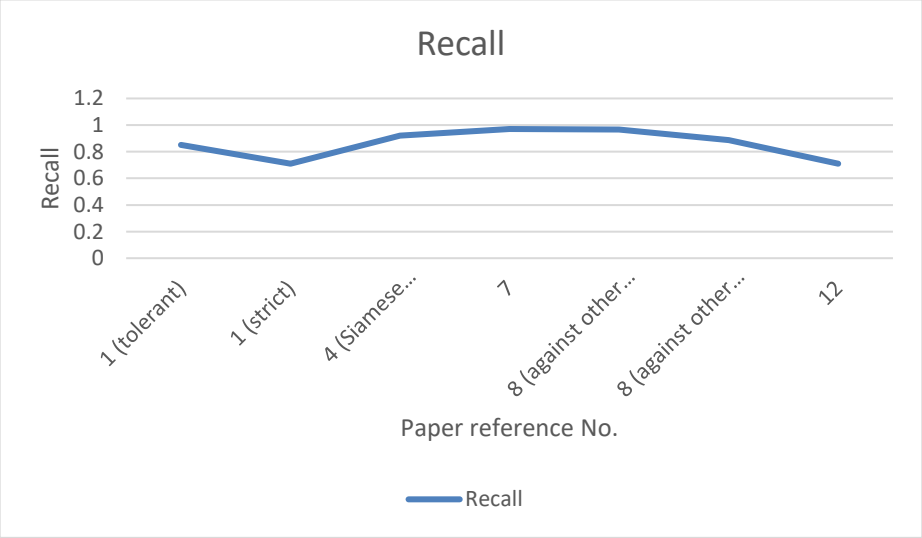


Fig. 4. Comparison of Recall for the mentioned paper Ref. No. as in Table 1

Fig. 5 represent the graphical comparison of F1 Score for the mentioned reference number as used in Table 1.

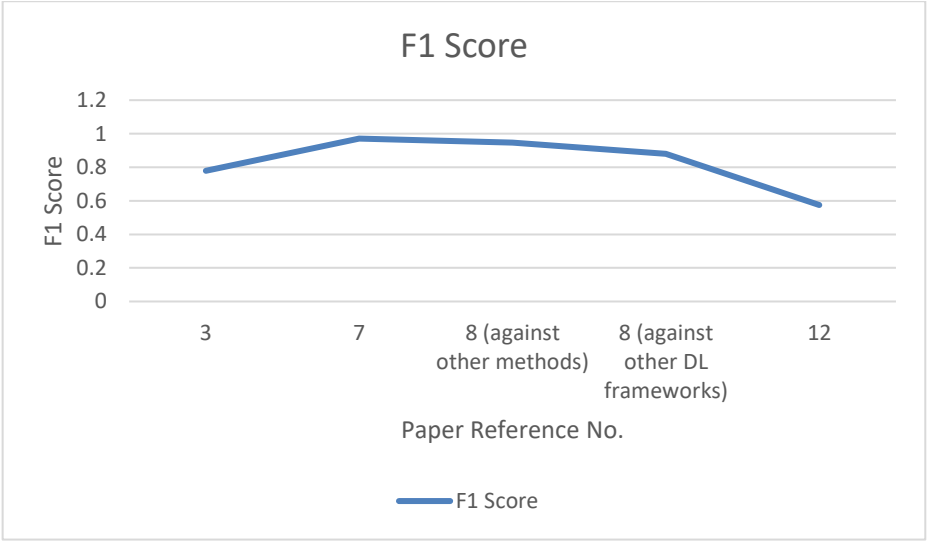


Fig. 5. Comparison of F1 Score for the mentioned paper Ref. No. as in Table 1

Fig. 6 represent the graphical comparison of Sensitivity for the mentioned reference number as used in Table 1.

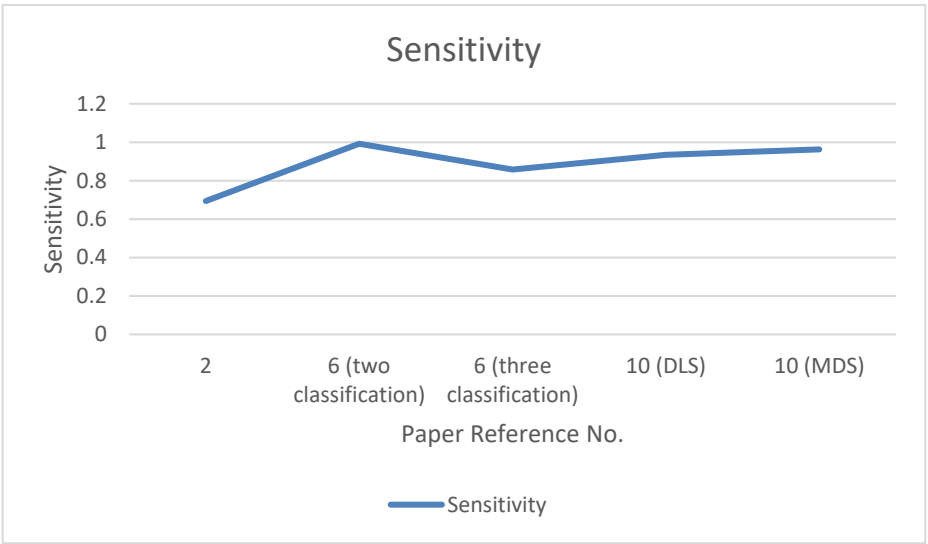


Fig. 6. Comparison of Sensitivity for the mentioned paper Ref. No. as in Table 1

Fig. 7 represent the graphical comparison of Specificity for the mentioned reference number as used in Table 1.

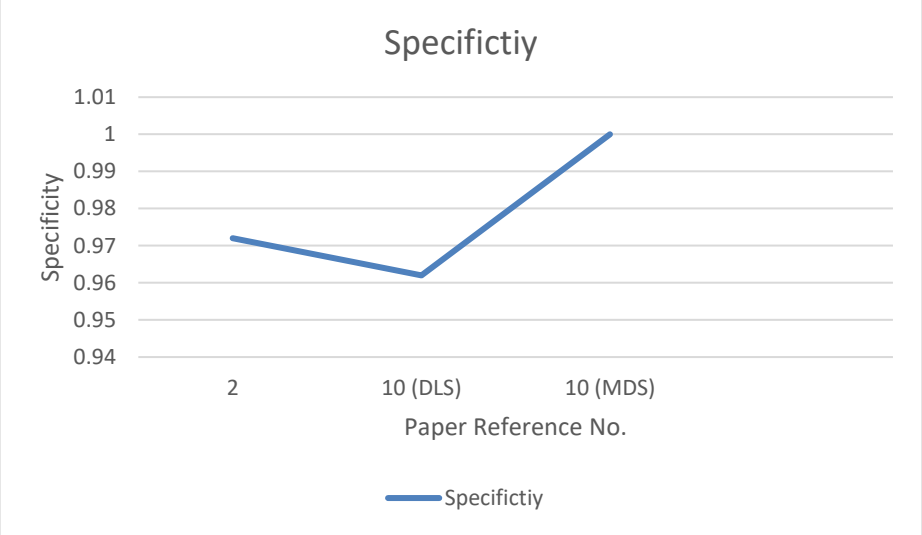


Fig. 7. Comparison of Specificity for the mentioned paper Ref. No. as in Table 1

3 Discussions

The authors of this work conducted a comprehensive review of all previous studies on the classification of bone marrow cells. They examined and analyzed various techniques employed in previous studies, comparing their performance, motivations, and contributions to the field of cytology. Through their analysis, the authors provided a comprehensive overview of the existing literature, highlighting the strengths and limitations of different approaches. Their study aimed to consolidate knowledge in the field and contribute to the advancement of bone marrow cell categorization techniques in cytology.

Traditionally, bone marrow cell classification relied on manual expert examination, which was a laborious, slow and subjective process. However, the emergence of artificial intelligence has transformed this field. Annotated bone marrow images have been used to develop and train Convolutional Neural Networks and other AI algorithms. These systems are used to classify cells based on their morphological traits. AI-based systems can analyze digital images quickly and accurately, providing standardized results and minimizing inter-observer variability. They can also detect rare cell types and abnormalities, improving diagnostic accuracy and supporting clinical decision-making. The integration of AI in bone marrow cell classification has transformed the process, offered faster and more precise results, and has the potential to advance research and clinical practice in hematopathology.

In their comparative analysis, the authors examined how well Deep Learning models performed in comparison to more conventional statistical techniques for classifying bone marrow cells. While the Deep Learning models showcased superior performance, the study identified areas for further improvement. These potential enhancements were

thoroughly examined and discussed by the authors, providing valuable insights for future advancements in the field.

References

[1] T.-H. Song, V. Sanchez, H. E. I. Daly, and N. M. Rajpoot, "Simultaneous cell detection and classification in bone marrow histology images," *IEEE J. Biomed. Heal. informatics*, vol. 23, no. 4, pp. 1469–1476, 2018.

[2] K. Kimura *et al.*, "A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA," *Sci. Rep.*, vol. 9, no. 1, p. 13385, 2019.

[3] X. Fu *et al.*, "Morphogo: an automatic bone marrow cell classification system on digital images analyzed by artificial intelligence," *Acta Cytol.*, vol. 64, no. 6, pp. 588–596, 2020.

[4] M. Wang, C. Dong, Y. Gao, J. Li, M. Han, and L. Wang, "A deep learning model for the automatic recognition of aplastic anemia, myelodysplastic syndromes, and acute myeloid leukemia based on bone marrow smear," *Front. Oncol.*, vol. 12, p. 844978, 2022.

[5] C. Matek, S. Krappe, C. Münzenmayer, T. Haferlach, and C. Marr, "Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set," *Blood, J. Am. Soc. Hematol.*, vol. 138, no. 20, pp. 1917–1927, 2021.

[6] J. W. Choi *et al.*, "White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks," *PLoS One*, vol. 12, no. 12, p. e0189259, 2017.

[7] R. M. Tayebi *et al.*, "Automated bone marrow cytology using deep learning to generate a histogram of cell types," *Commun. Med.*, vol. 2, no. 1, p. 45, 2022.

[8] Y.-Y. Wu *et al.*, "A hematologist-level deep learning algorithm (BMSNet) for assessing the morphologies of single nuclear balls in bone marrow smears: algorithm development," *JMIR Med. informatics*, vol. 8, no. 4, p. e15963, 2020.

[9] B. Ananthakrishnan, A. Shaik, S. Akhouri, P. Garg, V. Gadag, and M. S. Kavitha, "Automated bone marrow cell classification for haematological disease diagnosis using siamese neural network," *Diagnostics*, vol. 13, no. 1, p. 112, 2022.

[10] J. Liu *et al.*, "A deep learning method and device for bone marrow imaging cell detection," *Ann. Transl. Med.*, vol. 10, no. 4, p. 208, 2022.

[11] R. Chandradevan *et al.*, "Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells," *Lab. Investig.*, vol. 100, no. 1, pp. 98–109, 2020.

[12] R. Hanselia and D. K. Choubey, "A Comparative Study for Prediction of Hematopoietic Stem Cell Transplantation-Related Mortality," in *International Conference On Innovative Computing And Communication*, 2023, pp. 641–652.

[13] K. Srivastava and D. K. Choubey, "Soft computing, data mining, and machine learning approaches in detection of heart disease: A review," in *International Conference on Hybrid Intelligent Systems*, 2019, pp. 165–175.

[14] D. Sharma, P. Jain, and D. K. Choubey, "A comparative study of computational

intelligence for identification of breast cancer," in *International conference on machine learning, image processing, network security and data sciences*, 2020, pp. 209–216.

[15] A. Parthvi, K. Rawal, and D. K. Choubey, "A comparative study using machine learning and data mining approach for leukemia," in *2020 International conference on communication and signal processing (ICCSP)*, 2020, pp. 672–677.

[16] D. K. Choubey, S. Paul, and J. Bhattacharjee, "Soft computing approaches for diabetes disease diagnosis: a survey," *Int. J. Appl. Eng. Res.*, vol. 9, no. 21, pp. 11715–11726, 2014.

[17] D. K. Choubey and S. Paul, "Classification techniques for diagnosis of diabetes: a review," *Int. J. Biomed. Eng. Technol.*, vol. 21, no. 1, pp. 15–39, 2016.

[18] D. K. Choubey, S. Paul, K. Bala, M. Kumar, and U. P. Singh, "Implementation of a hybrid classification method for diabetes," in *Intelligent innovations in multimedia data engineering and management*, IGI Global Scientific Publishing, 2019, pp. 201–240.