Literature Review: Detecting Haematological Diseases from Bone Marrow Cytology

Abstract

The manual classification of bone marrow (BM) cell morphology, a pivotal aspect of haematological diagnosis, is performed thousands of times daily due to the lack of comprehensive data sets and trained models. This process is often time-consuming and susceptible to human errors. The effectiveness of deep learning algorithms in biomedical applications is proven. The impact is undeniable and unquestionable as these techniques use extensive datasets encompassing diverse disease classes, meticulously annotated by medical professionals. This also eliminates any scope for error, shortens the diagnosis time and enhances the accuracy. Over time, deep learning techniques have improved further with new advancements. This paper aims to compile the existing deep-learning models and analyse the methodologies as laid out in several available papers. The review thoroughly explores recent advancements in deep learning techniques tailored to classify BM cells, aiming to enhance the detection of haematological diseases.

Introduction

Examination of bone marrow (BM) cell morphologies is a crucial step in diagnosing diseases affecting the hematopoietic system [1, 2]. Due to the complexity of automating this process, medical professionals are still responsible for performing the microscopic examination and morphological classification of BM cells. This process of physical evaluation of the specimens can be challenging and laborious, especially when obscure BM smears are involved [3,4]. The potential for errors in manual cytological evaluation of bone marrow samples arises from its sole dependence on the medical examiner's knowledge and expertise, thereby increasing the risk of inaccurate diagnoses [5]. Furthermore, pathology labs may lack competent personnel to execute this analysis [6].

Many earlier studies on automated cytomorphologic classification are on classifying physiological cell types or peripheral blood smears, which limits the relevance of their findings to the classification of leukocytes in the BM for the diagnosis of haematological malignancies [7, 8-13]. In recent years, the study of bone marrow cytology has involved the use of Convolutional Neural Networks (CNN), a subset of Deep Learning [5]. However, deep-learning approaches to BM cell classification have been constrained by low sample numbers, very few disease categories and the related data being kept as private [14-17]. Consequently, the effectiveness of employing CNNs for image classification hinges on the availability of a substantial volume of high-quality image data that is meticulously annotated. However, obtaining these annotations from medical professionals can be challenging due to the associated costs [18, 19].

For educational purposes and research on automated BM cell classification, Matek et al. [20] have presented a large data set of 171,374 expert-annotated single-cell images of BM smears. They trained two CNN classifiers on this dataset to classify single-cell images of BM leukocytes. Based on the dataset provided by Matek et al. [20], a Convolution and Attention based model architecture is proposed in [7] for the classification of BM cells. The study in [5] introduces a custom CNN model trained on a subset of six major labels taken from the dataset provided by Matek et al. [20].

This article aims to conduct a thorough analysis of existing literature to facilitate research focused on automating the classification of bone marrow cytology. This automation aims to assist clinicians in rapidly and accurately detecting haematological disorders.

Dataset

The dataset [21] furnished by Matek et al. [20] is an Expert-Annotated Dataset of Bone Marrow Cytology in Hematologic Malignancies. It has more than 170,000 annotated cells from 945 patients with different haematological conditions. It uses a special stain called May-Grünwald-Giemsa/Pappenheim to colour the cells. It uses a brightfield microscope with 40x magnification and oil immersion to capture the images. It is part of the TCIA (The Cancer Imaging Archive), a public resource for medical and bioinformatics research. It was created by a collaboration of three institutions: The Munich Leukaemia Laboratory (MLL), the Fraunhofer IIS, and the Helmholtz Munich [20-22]. **Figure 1** shows sample images of BM cells as present in the dataset. **Figure 2** shows the distribution of 171,374 single-cell BM smear images into 21 classes in the dataset.

The process of cytological analysis involves three primary steps [6]:

1. Detection of the region of interest (ROI), which is identified manually by skilled medical experts.
2. Precise localization of distinct bone marrow cells and non-cellular objects.
3. Classification of the localized cells into specific categories based on their cytological features.

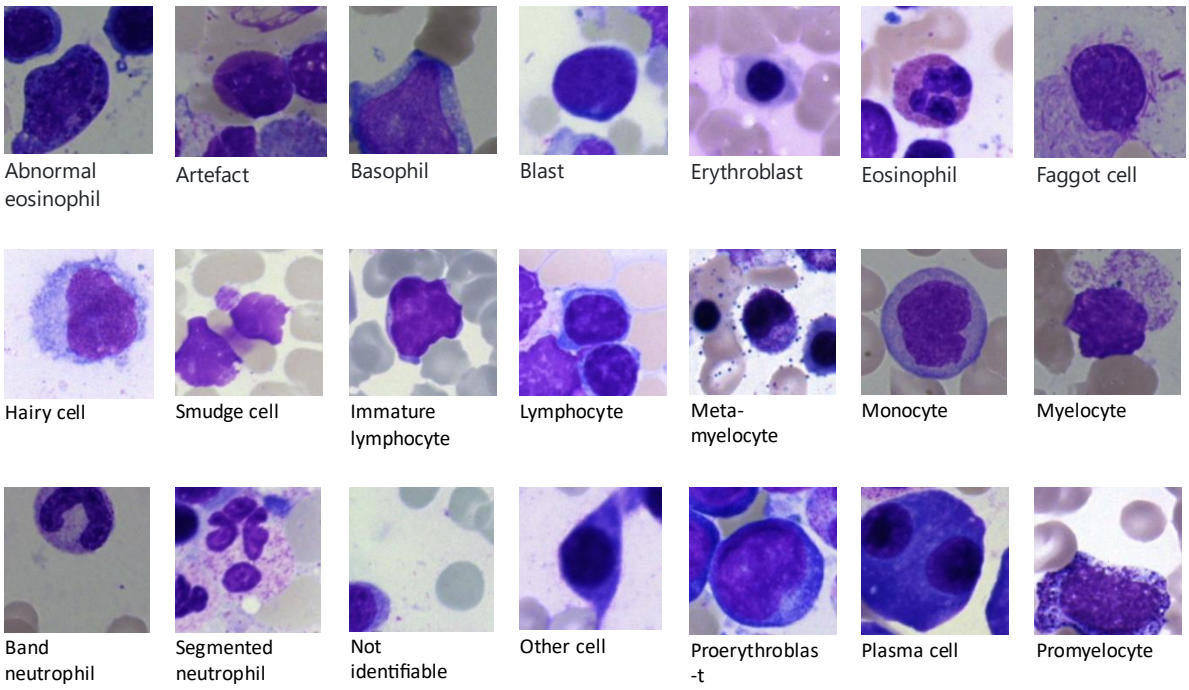

Figure 1: Samples of the BM cell images from the dataset provided by Matek et al [20].

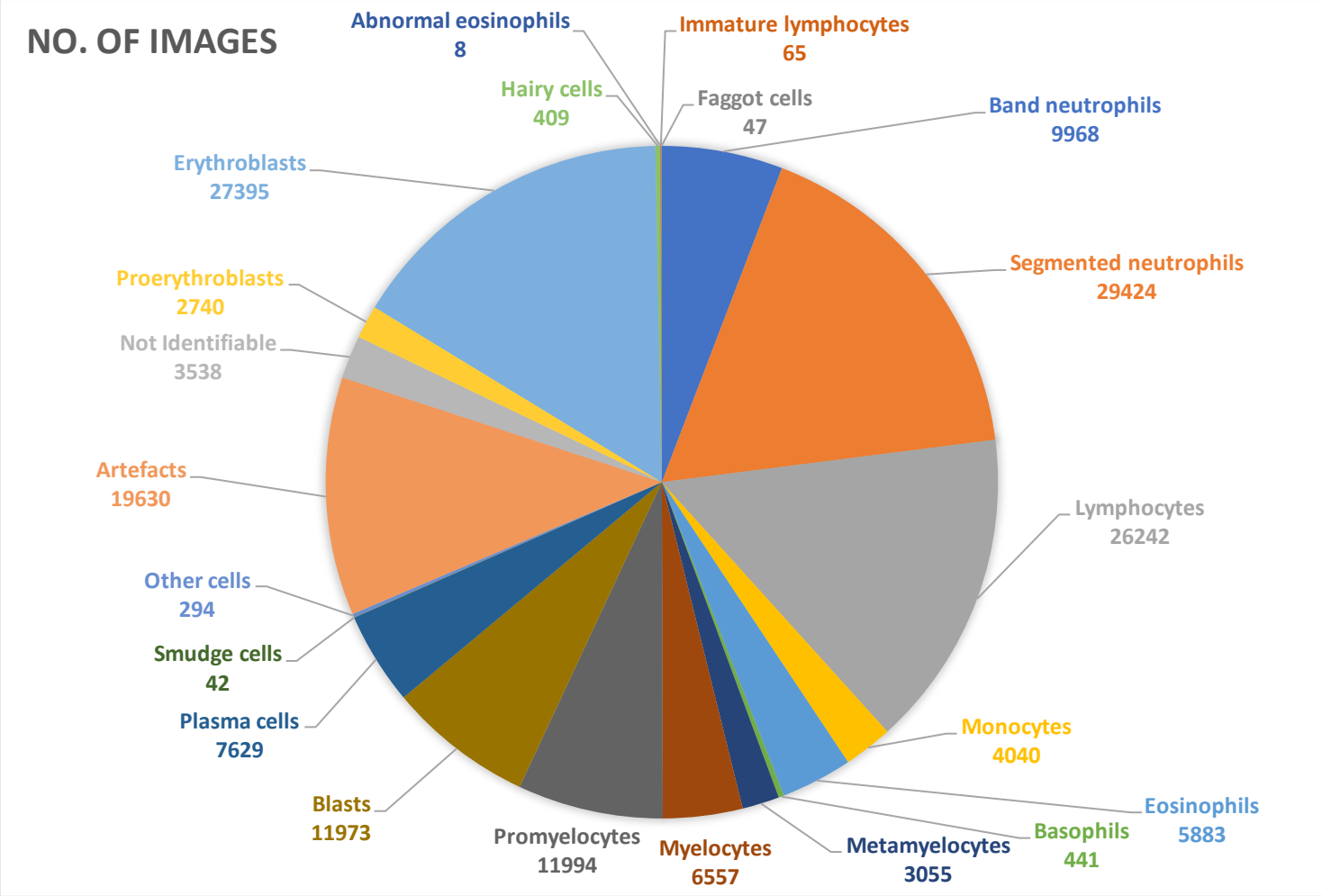

Figure 2: Distribution of 171,374 images in various cell classes in the dataset presented by Matek et al. [20]

Methodology

This paper provides a detailed analysis of three major deep learning models that have been trained on the dataset provided by Matek et al. [20]. They are ResNeXt-50 [20], CoAtNet [7] and BoMaCNet [5].

ResNeXt-50

The ResNeXt-50 architecture, developed by Xie et al. [23], achieved second place in the 2016 ImageNet Large Scale Visual Recognition Challenge [24] for image classification. This architecture had previously been applied in the classification of peripheral blood smears [25], making it a suitable choice for morphologically classifying bone marrow cells. A noteworthy characteristic of the ResNeXt architecture is its limited number of hyperparameters, specifically, the cardinality hyperparameter C, which is set at 32, in line with the original research [23]. The network was modified to accommodate 250 × 250-pixel images, and the number of output nodes was tailored to 22, with the fusion of two nodes to represent the 21 comprehensive morphological classes within the annotation scheme. Consequently, the final network consisted of 23,059,094 trainable parameters, and the prediction of the cell class was based on the output node with the highest activation [20].

The networks were trained using NVIDIA Tesla V100 graphics processing units, and the ResNeXt model required approximately 48 hours of computational time to complete its training. For each class, 80%

of available images were used for training, while the remaining 20% were allocated for network evaluation, employing a random stratified train-test split. Data augmentation was performed after this split. Fivefold cross-validation entailed a stratified partition of the dataset into five distinct, non-overlapping folds, each encompassing roughly 20% of the images within the corresponding cell class. Following this, five unique networks were individually trained for 13 epochs, with each network utilizing a different fold for testing while the other four folds were employed for training. Subsequently, the outcomes were averaged across the five distinct networks. To evaluate the consistency of the results with respect to network structure, a less complex sequential model, previously employed for training a leukocyte classifier in peripheral blood [25], was also trained. After modifying input and output channels to align with the ResNeXt model, the sequential model featured a total of 303,694 trainable parameters. The distribution of data into test and training sets for various folds matched the approach used in training the ResNeXt model [20].

Moreover, Krappe et al.'s feature-based approach [12] involved the use of minimum redundancy selection [26], which led to the selection of more than 6,000 features per cell for training a support vector machine. Furthermore, a slightly altered training strategy was adopted, designating 70% of the data for training and 30% for evaluation. To ensure the robustness of the results in light of this slight variation in the data split strategy, the ResNeXt-50 model was trained using a data split that maintained a 70% training and 30% test division. The outcomes displayed only marginal discrepancies compared to the results obtained through fivefold cross-validation [20].

CoAtNet

CoAtNet is an innovative hybrid model that amalgamates components from both Convolutional Neural Networks (CNNs) and Transformer models. Its design aims to leverage the advantages offered by both ConvNets and Transformers in a unified network, leading to superior performance across datasets of different sizes, all while upholding resource efficiency. CoAtNet is equipped with inductive biases that enable it to generalize effectively, similar to ConvNets. Additionally, it leverages the scalability and faster convergence properties of Transformers, enhancing overall efficiency. [7]

The study explores the fusion of convolution and attention in machine learning, focusing on two crucial aspects: generalization and model capacity. Convolutional layers excel at generalization, whereas attention layers excel at model capacity. By merging convolutional and attention layers, CoAtNet achieves superior generalization and capacity. This hybrid model is primarily designed for image classification and relies on two key features:

a) The combination of depth-wise convolution and self-attention using a straightforward relative attention mechanism. [7]

b) The strategic stacking of convolution layers and attention layers, leading to significant improvements in generalization, capacity, and efficiency. [7]

The model architecture encompasses both convolution and self-attention operations. The convolutional layer reduces input dimensionality, while the MBconv blocks, featuring an inverted structure, first expand the input by 4x and then capture spatial interactions before compression. Depth-wise convolution is applied independently to each channel. [7]

With regards to the feed-forward network (FNN) module and self-attention blocks, a comparable structure for expansion-compression that is employed in MBConv blocks is utilized. Self-attention offers the benefit of having a receptive field that spans all spatial locations and calculates weights based on re-normalized pairwise similarities [7].

CoAtNet offers architectural flexibility, allowing for the last three stages to be composed of either Convolution or Transformer blocks, enabling various model configurations. [7]

Input-Adaptive Weighting strengthens self-attention's capacity to discern relationships among diverse input elements, while the Global Receptive Field expands the receptive field employed in self-attention. The prime model architecture synergizes the Global Receptive Field and Input-Adaptive Weighting features of self-attention with the Translation Equivariance typical of CNNs. This combination enhances generalization, particularly for smaller datasets [7].

The core idea involves the summation of a global static convolution kernel with the adaptive attention matrix, either before or after softmax initialization, to achieve the desired characteristics in the model architecture. [7]

BoMaCNet

In the research study, the researchers introduced BoMaCNet, a custom CNN model tailored for the classification of bone marrow cell images from various individuals. BoMaCNet was trained on a subset of 96000 images taken from the dataset provided by Matek et al. [20]. 1600 images we taken from each of the six major classes chosen by the researchers. BoMaCNet employs a composition of 7 convolutional layers and 5 dense layers in its architecture. These convolutional layers utilize a 3 x 3 kernel size, with filter sizes ranging from 32 to 2048 for the individual layers. To enhance learning efficiency, Batch Normalization [17] is applied following each convolutional layer.

Following Batch Normalization, an average-pooling operation is performed using a 2 x 2 kernel. It was established that, within the model, average pooling exhibits superior performance compared to min-pooling and max-pooling architectures when it comes to image identification. For the activation function in all convolutional layers, Rectified Linear Unit (ReLU) [8] is employed, as it is known for its effectiveness in convolutional networks.

After the convolutional layers, a flatten layer is integrated to downsize the dimensions of the preceding layers, transforming them into a one-dimensional array for utilization as input in the subsequent layers. The model proceeds with the application of dense layers, comprising 256, 128, 64, 32, 16, and finally 6 units in the last layer, which correspond to the 6 label categories. In the ultimate layer, a softmax activation function [5] is employed to produce the final output.

Results

In paper [20], ResNeXt accurately predicts most morphological classes of leukocytes. Precise differentiation of individual classes can be difficult, especially those closely related in the leukocyte class. Consequently, certain predictions may be considered acceptable, even if they deviate from the label marked by the human annotator.

ResNeXt and a sequential neural network outperformed a feature-based classifier in all classes except segmented neutrophils and lymphocytes. This may be due to the feature-based classifier's explicit inclusion of nuclear shape parameters, which ResNeXt and the sequential network do not rely on.

The CoAtNet [7] model outperformed the ResNeXt [5] model because of its attention network feature which increased the learning curve for the algorithm. BoMaCNet achieved a training accuracy of

95.71% and a validation accuracy of 93.06% in classifying the bone marrow smear images into six classes.

Most models performed less favourably at identifying classes with few training examples, such as abnormal eosinophils and faggot cells. Hence, more training data would be required to correctly classify images of these cell types.

ResNeXt [20] and CoAtNet [7] both employed two explainability methods, SmoothGrad and Grad-CAM, to analyze and identify the critical regions of the input images influencing the network's classification decisions. The results demonstrated that the model had learned to ignore background elements like erythrocytes, cell debris, or parts of other cells while concentrating on the relevant input of a single-cell patch, i.e., the major leukocyte that was visible in it.

The methods indicate that the model is learning to focus on the images' most vital features. This boosts confidence that the model's predictions are founded on plausible traits.

Table 1 shows a comparative study of the results demonstrated by CoAtNet [7], ResNeXt [20] and BoMaCNet [5] models:

Class Name	Precision			Recall		
	CoAtNet [7]	ResNext [20]	BoMaCNet [5]	CoAtNet [7]	ResNext [20]	BoMaCNet [5]
Band neutrophils	0.97	0.91	-	0.96	0.91	-
Segmented neutrophils	0.95	0.95	0.98	0.97	0.85	0.96
Lymphocytes	0.94	0.90	0.93	0.93	0.72	0.91
Monocytes	0.81	0.57	-	0.79	0.70	-
Eosinophils	0.91	0.85	-	0.88	0.91	-
Basophils	0.74	0.14	-	0.70	0.64	-
Metamyelocytes	0.91	0.68	-	0.89	0.87	-
Myelocytes	0.88	0.78	-	0.87	0.91	-
Promyelocytes	0.98	0.91	0.95	0.98	0.89	0.94
Blasts	0.94	0.79	0.88	0.98	0.69	0.91
Plasma cells	0.93	0.81	-	0.95	0.84	-
Smudge cells	-	0.28	-	-	0.90	-
Other cells	-	0.22	-	-	0.84	-
Artefacts	-	0.82	0.92	-	0.74	0.92
Not Identifiable	-	0.27	-	-	0.63	-
Proerythroblasts	0.89	0.69	-	0.85	0.85	-
Erythroblasts	0.99	0.90	0.96	0.98	0.83	0.96
Hairy cells	0.92	0.80	-	0.88	0.88	-
Abnormal eosinophils	0.42	0.02	-	0.39	0.20	-
Immature lymphocytes	0.65	0.35	-	0.66	0.57	-
Faggot cells	0.83	0.17	-	0.87	0.63	-

Table 1: A detailed comparison of the approximate values of Precision and Recall for each BM class achieved by the chosen models. Note: (A) Null values indicate that the specific class was excluded from the training dataset by the researchers. (B) The precision and recall values considered for the ResNeXt model are tolerant values i.e., they accept a certain level of misclassification.

Figure 3 and **Figure 4** show bar plots for the precision and recall values of CoAtNet [7], ResNext [20] and BoMaCNet [5] models respectively. It can be observed that amongst CoAtNet and ResNeXt, CoAtNet has better precision than ResNeXt. Most classes are more precisely predicted by CoAtNet than ResNeXt. Even in case of classes that have very few images in the dataset provided my Matek et al. [20], like Faggot cells, Immature lymphocytes and Abnormal eosinophils, CoAtNet has a higher precision than ResNeXt.

Unlike precision, recall values fluctuate a little between CoAtNet and ResNeXt. ResNeXt has better recall for classes such as Eosinophils and Myelocytes while it has the same recall as CoAtNet for Proerythroblasts and Hairy cells. However, overall, CoAtNet has better recall for most cell classes as compared to ResNeXt.

BoMaCNet has high and mostly consistent precision and recall values as compared to both, CoAtNet and ResNeXt.

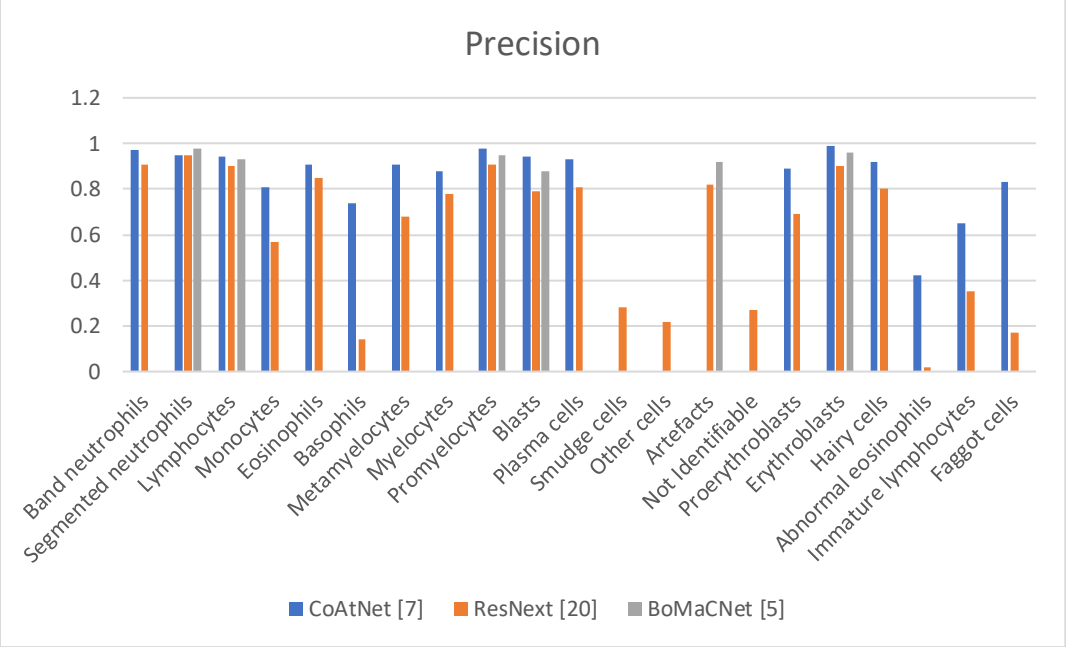

Figure 3: A bar plot depicting the precision values of CoAtNet [7], ResNeXt [20] and BoMaCNet [5] models.

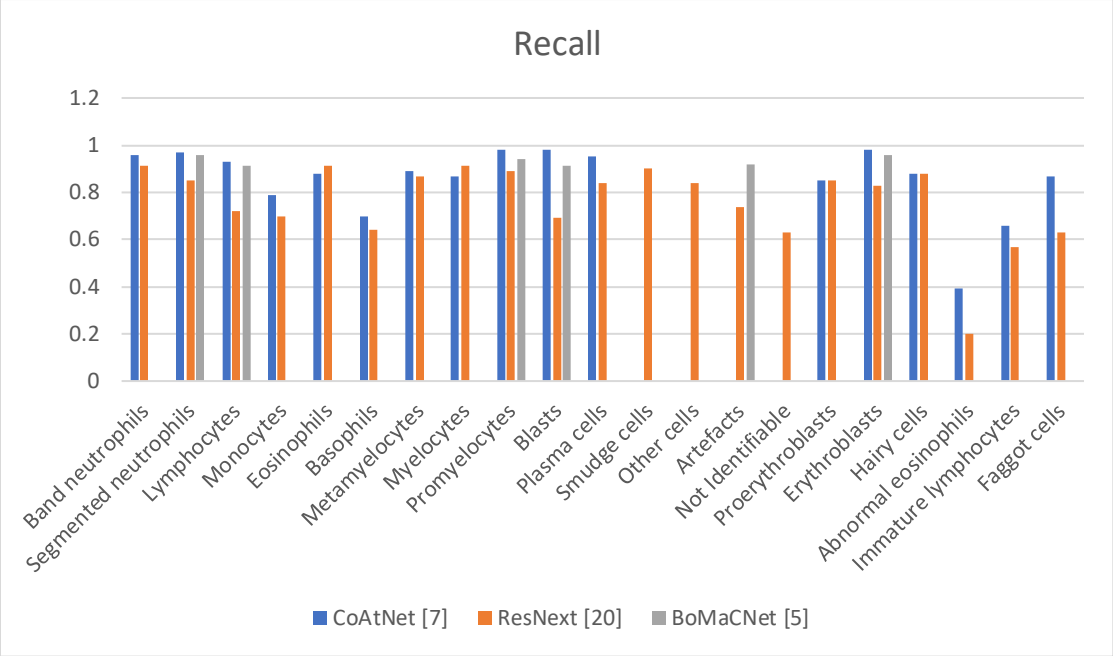

Figure 4: A bar plot depicting the recall values of CoAtNet [7], ResNeXt [20] and BoMaCNet [5] models.

Conclusion

In recent times, significant efforts have been made to automate the analysis and classification of bone marrow cytology. While traditional feature-based classifiers rely on manually crafted single-cell features from digital images, convolutional neural networks learn from the images and exhibit greater accuracy in predicting bone marrow cell classes. Additionally, extensive datasets with diverse disease classes and expert annotations are being made available publicly for educational and research purposes. In the near future, medical experts will be able to leverage sophisticated deep learning techniques to accelerate and improve the precision of haematological disease diagnosis.

References

[1] Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H et al. The 2016 revision of the WHO classification of lymphoid neoplasms. Blood.. 2016 May 19;127(20):2375-90. doi: 10.1182/blood-2016-01-643569. Epub 2016 Mar 15. PMID: 26980727; PMCID: PMC4874220.

[2] Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR et al. Diagnosis and management of acute myeloid leukemia in adults: European LeukemiaNet recommendations from an international expert panel. Blood. 2017 Jan 26;129(4):424-447. doi: 10.1182/blood-2016-08-733196. Epub 2016 Nov 28. PMID: 27895058; PMCID: PMC5291965.

[3] Briggs C, Longair I, Slavik M, Thwaite K, Mills R et al. Can automated blood film analysis replace manual differentiation? An evaluation of the CellaVision DM96 automated image analysis system. Int J Lab Hematol. 2009 Feb;31(1):48-60. doi: 10.1111/j.1751-553X.2007.01002.x. Epub 2007 Dec 20. PMID: 18177438.

[4] Angulo J, Flandrin G.. Automated detection of working area of peripheral blood smears using mathematical morphology. Anal Cell Pathol. 2003;25(1):37-49.. doi: 10.1155/2003/642562. PMID: 12590176; PMCID: PMC4618829.

[5] A. S. Abeed, A. Atiq, A. A. Anjum et al., "BoMaCNet: A CNN Model to Detect Bone Marrow Cell Cytology," 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2022, pp. 686-691, doi: 10.1109/ICCIT57492.2022.10054976.

[6] Tayebi, R.M., Mu, Y. et al. Automated bone marrow cytology using deep learning to generate a histogram of cell types. Commun Med 2, 45 (2022). https://doi.org/10.1038/s43856-022-00107-6

[7] Satvik Tripathi, Alisha Isabelle Augustin, Rithvik Sukumaran et al., HematoNet: Expert level classification of bone marrow cytology morphology in hematological malignancy with deep learning, Artificial Intelligence in the Life Sciences, Volume 2, 2022, 100043, ISSN 2667-3185, https://doi.org/10.1016/j.ailsci.2022.100043.

[8] Krappe, S., Benz, M., Wittenberg, T. et al., "Automated classification of bone marrow cells in microscopic images for diagnosis of leukemia: a comparison of two classification schemes with respect to the segmentation quality", in Medical Imaging 2015: Computer-Aided Diagnosis, 2015, vol. 9414. doi:10.1117/12.2081946.

[9] Reta, Carolina & Altamirano Robles, Leopoldo & Gonzalez et al. (2015). Correction: Segmentation and Classification of Bone Marrow Cells Images Using Contextual Information for Medical Diagnosis of Acute Leukemias.. PLoS ONE. 10. 10.1371/journal.pone.0134066.

[10] Chandradevan R, Aljudi AA, Drumheller BR et al.. Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells. Lab Invest. 2020 Jan;100(1):98-109. doi: 10.1038/s41374-019-0325-7. Epub 2019 Sep 30. PMID: 31570774; PMCID: PMC6920560.

[11] T. -H. Song, V. Sanchez, H. EI Daly and N. M. Rajpoot, "Simultaneous Cell Detection and Classification in Bone Marrow Histology Images," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1469-1476, Jul 2019, doi: 10.1109/JBHI.2018.2878945.

[12] Krappe ,S., Wittenberg ,T., Haferlach ,T., Münzenmayer ,C., "Automated morphological analysis of bone marrow cells in microscopic images for diagnosis of leukaemia: nucleus--plasma separation and cell classification using a hierarchical tree model of haematopoesis," Proc. SPIE 9785, Medical Imaging 2016: Computer-Aided Diagnosis, 97853C (24 March 2016); https://doi.org/10.1117/12.2216037

[13] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," CIMSA. (2005) IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, 2005., Messian, Italy, 2005, pp. 96-101, doi: 10.1109/CIMSA.2005.1522835.

[14] Mori, J., Kaji, S., Kawai, H. et al. Assessment of dysplasia in bone marrow smear with CNN. Sci Rep 10, 14734 (2020). https://doi.org/10.1038/s41598-020-71752-x

[15] K.K. Anilkumar, V.J. Manoj, T.M. Sagi, A survey on image segmentation of blood and bone marrow smear images with emphasis to automated detection of Leukemia, Bio-cybernetics and Bio-medical Engineering, Volume 40, 2020, ISSN 0208-5216, https://doi.org/10.1016/j.bbe.2020.08.010.

[16] Jin H., Fu X., Cao X. et al. Developing and Validating an Automatic Cell Classification System for Bone Marrow Smears: a Pilot Study. J Med Syst. 2020 Sep 7;44(10):184. doi: 10.1007/s10916-020-01654-y. PMID: 32894360; PMCID: PMC7476995.

[17] Choi J, Ku Y, Yoo B, et al. (2017) White blood cell differential count in maturation stages in bone marrow smear using dual-stage CNNs. PLoS ONE 12(12): e0189259. https://doi.org/10.1371/journal.pone.0189259

[18] H. Greenspan, B. van Ginneken and R. M. Summers, "Guest Editorial DL in Medical Imaging: Overview and Future Promise of an Exciting New Technique," in IEEE Transactions on Medical Imaging, vol. 35, pp. 1153-1159, 2016, doi: 10.1109/TMI.2016.2553401.

[19] Shen D, Wu G, Suk H. DL in Medical Image Analysis. Annual Rev Bio-med Engg. 2017;19:221-248. doi:10.1146/annurev-bioeng-071516-044442

[20] Christian Matek, Sebastian Krappe, Christian Münzenmayer, Torsten Haferlach, Carsten Marr; Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. Blood 2021; 138 (20): 1917–1927. doi: https://doi.org/10.1182/blood.2020010568

[21] Matek, C., Krappe, S., Münzenmayer, C., Haferlach, T., & Marr, C. (2021). An Expert-Annotated Dataset of BM Cytology in Hematologic Malignancies [Data set].. TCIA. https://doi.org/10.7937/TCIA.AXH3-T579

[22] Clark K, Vendt B, Smith K, Freymann J, et al. The Cancer Imaging Archive: Maintaining and Operating a Public Data Repository, Journal of Digital Imaging, Volume 26, December, 2013, pp 1045-1057. DOI: 10.1007/s10278-013-9622-7

[23]. Xie S, Girshick R, Dollar P, et al. Aggregated residual transformations for DNNs. 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017;5987-5995.

[24] Russakovsky O, Deng J, Su H, et al. ImageNet visual recognition challenge. Intl. J Comp. Visi. 2015;115(3):211-252.

[25] Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in AML with CNNs. Nat Mach Intell. 2019;1:538-544

[26] Schowe B, Morik K. Fast-ensembles of minimum redundancy feature Selection. In: Okun O, Valentini G, Re M, eds. Ensembles in MachineLearning Applications. Studies in Computational Intelligence, vol 373. Springer,Berlin, Heidelberg; 2011:75–95