Computers in Biology and Medicine 178 (2024) 108691

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A deep learning approach for automatic recognition of abnormalities in the cytoplasm of neutrophils☆

Kevin Barrera [a,*], José Rodellar [a], Santiago Alférez [a], Anna Merino [b]

[a] Technical University of Catalonia, Barcelona East Engineering School, Department of Mathematics, Barcelona, Spain
[b] Hospital Clínic of Barcelona-IDIBAPS, Biochemistry and Molecular Genetics Department, CORE Laboratory, Biomedical Diagnostic, Barcelona, Spain

ARTICLE INFO

Keywords:
Blood cell automatic recognition
Blood cell morphology
Convolutional Neural Networks
Cytoplasmic inclusions
Deep learning
Generative Adversarial Networks
Neutrophils

ABSTRACT

Background and objectives: This study aims to develop and evaluate NeuNN, a system based on convolutional neural networks (CNN) and generative adversarial networks (GAN) for the automatic identification of normal neutrophils and those containing several types of inclusions or showing hypogranulation. Methods: From peripheral blood smears, a set of 5605 digital images was obtained with neutrophils belonging to seven categories: Normal neutrophils (NEU), Hypogranulated (HYP) or containing cryoglobulins (CRY), Döhle bodies (DB), Howell–Jolly body-like inclusions (HJBLI), Green–blue inclusions of death (GBI) and phagocytosed bacteria (BAC). The dataset utilized in this study has been made publicly available. The class of GBI was augmented using synthetic images generated by GAN. The NeuNN classification model is based on an EfficientNet-B7 architecture trained from scratch. Results: NeuNN achieved an overall performance of 94.3% accuracy on the test data set. Performance metrics, including sensitivity, specificity, precision, F1-Score, Jaccard index, and Matthews correlation coefficient indicated overall values of 94%, 99.1%, 94.3%, 94.3%, 89.6%, and 93.6%, respectively. Conclusions: The proposed approach, combining data augmentation and classification techniques, allows for automated identification of morphological findings in neutrophils, such us inclusions or hypogranulation. The system can be used as a support tool for clinical pathologists to detect these specific abnormalities with clinical relevance.

1. Introduction

1.1. Neutrophils and cytoplasmic inclusions

Peripheral blood (PB) is an essential tissue that circulates throughout the human body, performing vital functions such as transporting oxygen and nutrients. Composed mainly of red blood cells (erythrocytes), white blood cells (leukocytes) and platelets, which are suspended in plasma. White blood cells are fundamental in the immune system, protecting the body against infections and diseases [1]. Within this category, neutrophils are the most abundant class, around 60% [2]. These cells contribute in the initial stages of the inflammatory response by eliminating microbes and strengthening the body's defenses, mainly through their phagocytic capacity [3]. This ability helps to the direct eradication of infectious agents. In addition to their phagocytic capacity, neutrophils contain granules that harbor enzymes and antimicrobial substances, these are released to modulate the inflammatory response and recruit other immune cells to the site of infection or tissue damage [4].

The presence and morphological characteristics of different types of inclusions within neutrophils, as illustrated in Fig. 1, can provide the clinical pathologist with diagnostic information on some clinical conditions that may be critical for the patient [5]. Döhle bodies are blue-gray or pale-blue cytoplasmic inclusions composed of endoplasmic reticulum together with glycogen granules and ribosomal proteins [6]. They are associated with pregnancy, infection, inflammation and myelodysplastic syndrome, among others. Cryoglobulins may be seen as single or multiple round, weakly basophilic inclusions that displace the nucleus [7]. The finding of cryoglobulins inside neutrophils has been associated with several infections, renal, hepatic, autoimmune, hematologic and neoplastic diseases [8]. Howell–Jolly body-like inclusions (HJBLI) within neutrophils correspond to small fragments of nuclear DNA [9]. Their presence in blood is not infrequently observed in an

☆ This document is a result of the research project "Computational System for the Diagnosis of Acute Leukemia and Lymphoma from Peripheral Blood Images: Proof of Concept and Roadmap for Technological Valorization". Funded by the Ministry of Science and Innovation of Spain, PDC2022-133514-I00.

* Corresponding author.
E-mail addresses: kevin.barrera@upc.edu (K. Barrera), jose.rodellar@upc.edu (J. Rodellar), santiago.alferez@upc.edu (S. Alférez), amerino@clinic.cat (A. Merino).

https://doi.org/10.1016/j.compbiomed.2024.108691
Received 3 April 2024; Received in revised form 12 May 2024; Accepted 1 June 2024
Available online 4 June 2024

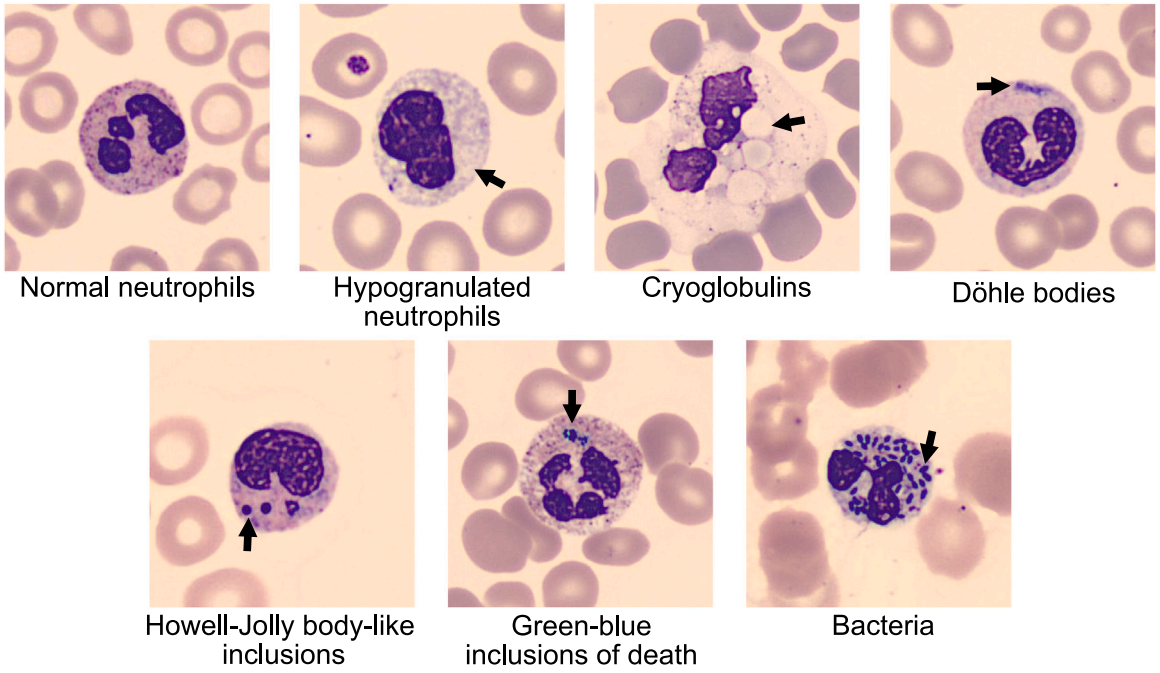

Fig. 1. Samples of normal neutrophils (NEU) and six different types of cytoplasmic abnormalities in neutrophils: Hypogranulated neutrophils (HYP), Cryoglobulins (CRY), Döhle bodies (DB), Howell–Jolly body-like inclusions (HJBLI), Green-blue inclusions of death (GBI)and Bacteria (BAC).

heterogeneous group of patients immunocompromised by a viral infection, chemotherapy, or immunosuppressive treatment in the context of a transplant procedure [10–12]. The presence of green-blue inclusions (GBI) within neutrophils is a critical morphological finding [13]. These inclusions correspond to the lipofuscin pigment coming from the destruction of liver parenchyma cells and phagocytosed by neutrophils. Their finding means that the patient has acute and severe liver failure which, without adequate support, will cause death [14]. Laboratories should be proactive in screening for these inclusions [15]. Detection of bacteria inside neutrophils by phagocytosis in the peripheral blood smear review indicates that the patient has a serious infection (sepsis) [16–18]. The reduction in the content of cytoplasmic granules in neutrophils is considered an abnormality. The identification of hypogranulated neutrophils through morphological examination plays a relevant role in the initial diagnosis of myelodysplastic syndromes [19].

1.2. Computer-based recognition models involving neutrophils

Modern microscopy systems with integrated image processing and classification software allow digital morphological analysis of peripheral blood smears. They have been shown to be effective in automatically recognizing normal leukocytes, but have limitations in identifying the presence of abnormal blood cells associated to a wide variety of malignancies [20]. Consequently, pathologists in their practice in clinical laboratories use these systems mainly at a pre-examination stage. Driven by the availability of good quality digital images and the potential of deep learning methods such as convolutional neural networks (CNN), in recent years a significant research effort has been directed towards the development of models for the automatic morphological recognition of abnormal blood cells for diagnostic purposes. An example is the state of the art focused on the diagnosis of acute leukemia [21–23].

In this context, neutrophils have been treated primarily as part of the normal leukocytes. Different CNN-based models have been proposed for the classification of the five types of normal leukocytes [24,25]. Regarding morphological abnormalities in neutrophils, hypogranulation has been the only case in which computational models have been reported in the literature. The presence of less granulated dysplastic neutrophils in PB is difficult to detect by human eyes, which causes a considerable inter-observer variance in their manual classification. Two relevant CNN-based models were proposed to detect hypogranulated neutrophils in patients with myelodysplastic syndromes [26,27]. However, automated detection of other specific cytoplasmic inclusions, such as those described above, has not been sufficiently addressed in the literature and represents an emerging field of study with high clinical relevance yet to be explored in the present work.

1.3. Objectives of this work

In this paper, we develop and validate a new CNN-based system specifically designed for the automatic detection and classification of normal neutrophils and six different types of cytoplasmic abnormalities in neutrophils. The system will be trained to identify normal neutrophils and those that exhibit morphological alterations such as Döhle bodies, cryoglobulins, Howell–Jolly body-like inclusions, green-blue inclusions, phagocytosed bacteria, and hypogranulation. CNNs can learn complex patterns from images through a series of filters to extract representative features from the input images and obtain a prediction at their output. However, the effectiveness of these models depends largely on the amount and diversity of well annotated images available for training. This is particularly problematic for identifying specific cytoplasmic inclusions in neutrophils, which have low prevalence and representation.

The main contributions of this work are:

1. A new database of annotated images of neutrophils with cytoplasmic abnormalities is created, which will be released under appropriate license for use by the scientific community.
2. A methodology is proposed to artificially create morphologically correct images of specific cytoplasmic inclusions in neutrophils. It is applied to generate synthetic samples of the class of green-blue inclusions. The approach may be extensible to augment other rare types of morphological alterations.
3. A convolutional neural network called NeuNN is designed and trained to detect and classify the different types of cytoplasmic abnormalities in neutrophils. NeuNN leverages the complementary advantages of CNN and GAN architectures to achieve the

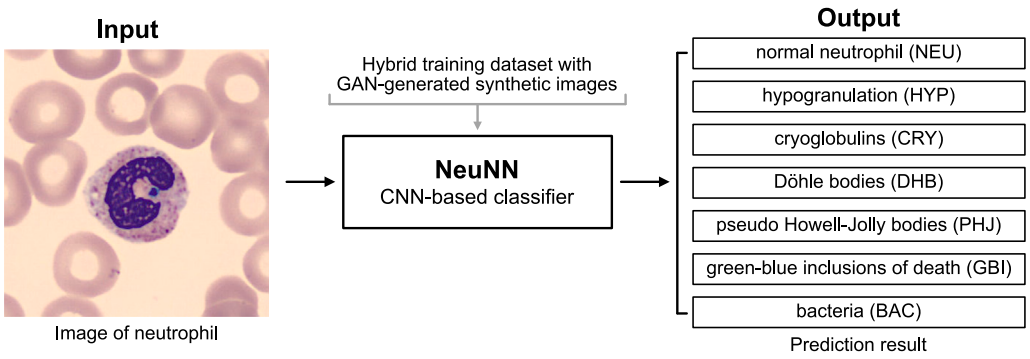

Fig. 2. General structure of the Neutrophil Neural Network (NeuNN) for the automatic identification of morphological alterations in neutrophils using deep learning models. NeuNN is based on a CNN classifier to identify seven types of neutrophils from input images. A GAN model is used to generate synthetic images as a data augmentation technique. The hybrid dataset, including original and synthetic images, is used to train NeuNN.

proposed objective with acceptable performance. The design and the training method are explained in detail.

4. The performance of NeuNN is assessed on a test set with images from patients and complex cases involving multiple simultaneous inclusions are analyzed together with the advice of expert pathologists.

To the authors' knowledge, there is no previous research addressing the automatic identification of all six types of cytoplasmic inclusions together.

2. Methodology

2.1. Overview

The ultimate goal of this work is to develop a computational model (NeuNN) to identify various morphological alterations in neutrophils, as illustrated in Fig. 2. The input is a set of digital images of neutrophils from a PB smear and the output is their classification into one of the following classes:

- Normal neutrophils (NEU)
- Hypogranulated neutrophils (HYP)
- Cryoglobulins (CRY)
- Döhle bodies (DB)
- Howell-Jolly body-like inclusions (HJBLI)
- Green-blue inclusions of death (GBI)
- Bacteria (BAC)

The approach uses convolutional neural networks (CNN) with special care in the problem posed by the fact that some of the classes under study have limited samples, which requires the generation of synthetic images modifying a GAN-based model described in [28].

In the subsequent Section 2.2, we present all the details involving the image acquisition, including the process to generate the artificial images. The structure and methodology behind the NeuNN model are detailed in Section 2.3, while its training is described in Section 2.4. The experimental evaluation of the automatic classification capability of NeuNN is presented in Section 3. Subsequently, Section 4 discusses our findings, explores the limitations of the current system and proposes possible improvements along with future lines of research. Finally, Section 5 summarizes the conclusions derived from this study.

2.2. Neutrophil image dataset

2.2.1. Original images

To train the NeuNN model, we started by collecting images at the Core laboratory of the Hospital Clinic of Barcelona. Blood samples where obtained from patients using EDTA-K3 as anticoagulant. Following this, smears were automatically prepared and stained using the SP10 system (Sysmex, Kobe, Japan) and the May Grünwald-Giemsa stain. The resultant smear was then processed through the CellaVision DM96 cell morphology analyzer, which supplied the individual images. Clinical pathologists validated the images through detailed visual morphological analysis, correctly categorizing and labeling them into one of the cell classes of interest for this study.

Table 1 shows the number of images used in this research for each class, with a total of 5605. They have a resolution of 360 × 363 pixels and are presented in the RGB (Red, Green, Blue) color space.

This dataset is available at the following link: https://data.mendeley.com/datasets/rh3jw43hjs/1. Interested researchers can access and download the images directly, using them in accordance with the Creative Commons Attribution 4.0 International license (CC BY 4.0). This license permits use, distribution and reproduction in any medium, provided the original source is properly cited.

2.2.2. Generation of synthetic images of neutrophil inclusions

The 5605 images in Table 1 exhibit a notable imbalance between the classes. Specifically, there is a predominance of normal Neutrophils (NEU) with 4595 images compared to the less represented class, the green-blue inclusions of death (GBI) with only 42 images. We performed preliminary studies using conventional augmentation techniques, such as rotation, translation, scaling, inversion and brightness adjustments [29], which were shown to be ineffective in compensating for the very limited number of GBI images to achieve effective training of CNN models. Given this situation, in this work we propose the generation of synthetic images using SyntheticCellGAN [28], an existing tool specifically designed to overcome image scarcity, which uses generative adversarial networks (GAN) [30] to create synthetic images of low-prevalence leukocytes and abnormal cells.

The original SyntheticCellGAN was composed of two models in series. First, GAN A used a WassersteinGAN architecture [31] to convert random data into basic low-resolution cell images. Subsequently, GAN B used a SuperResolution structure [32] to refine and obtain images of higher resolution and with more precise details in cellular morphology. In [28] a specific GAN B_n was designed to generate images of lymphocytes (B_1), monocytes (B_2), eosinophils (B_3), normal neutrophils (B_4), basophils (B_5), atypical promyelocytes (AP) (B_6) and hairy cells (HC) (B_7).

In this paper, SyntheticCellGAN is modified to generate artificial images of cells with GBI. The modified system is shown in Fig. 3. Since GBI is a specific inclusion inside a neutrophil, our initial idea was to use the original GAN trained to generate normal neutrophils (B_4) to directly generate neutrophils with the desired inclusions. However, the first experiments showed that the generated GBI were randomly distributed throughout the image. This result was not satisfactory, since the true GBI are observed on the cytoplasm of neutrophils. To address this issue, we propose an alternative two-step strategy that allows control over the location of inclusions. The first step is to define a visual marker in

Table 1
Description of the data set with images of normal neutrophils and the different morphological alterations. The table specifies the techniques used in processing smears to obtain digital images.

Smear processing		Types of neutrophils	Number of samples
Preparation	SP10 system	Normal neutrophils (NEU)	4595
Staining	May Grünwald-Giemsa	Hypogranulated neutrophils (HYP)	494
Collection	EDTA	Cryoglobulins (CRY)	191
Microscope	CellaVision DM96	Döhle bodies (DB)	139
Image size	360 × 363 pixels	Howell-Jolly body-like inclusions (HJBLI)	71
Color space	RGB	Green-blue inclusions of death (GBI)	42
Laboratory	Hospital Clínic of Barcelona	Bacteria (BAC)	73

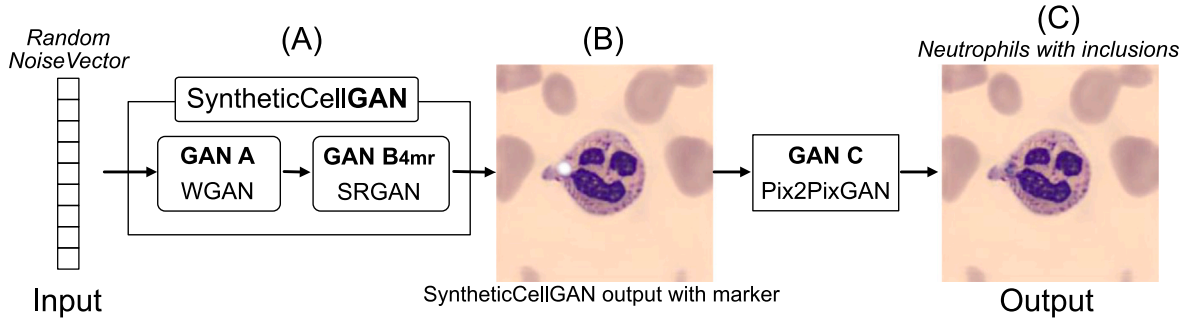

Fig. 3. Diagram of the workflow for generating artificial cytoplasmic inclusions using generative adversarial networks (GAN). In the initial stage (A), a random noise vector is fed into SyntheticCellGAN to generate an artificial neutrophil with a marker in the cytoplasm (B). Subsequently, this image is processed by a second GAN, resulting in the creation of synthetic neutrophils that present blue–green death inclusions (GBI) (C).

the form of a white circle within the cytoplasm to indicate the desired region for inclusion. The second step consists of filling this marker with a realistic cytoplasmic inclusion.

The first step is illustrated in Fig. 3.A, where the original GAN B_4 is modified to GAN B_{4mr} to generate images of neutrophils that have white markers in their cytoplasm. The data set used to train this model consist of pairs of images: normal neutrophils and the same neutrophils with a white circle manually inserted into the cytoplasm. This supervised learning technique allows the model to learn to replicate the location of the marker in the synthetic images. At the output, images of artificial neutrophils are obtained that include the marker in the cytoplasm, as shown in Fig. 3.B.

For the second step, a new GAN C, with a pix2pix architecture [33], is built for inserting a realistic GBI within the marked region, as shown in Fig. 3.C. Pix2pix performs an image transformation that allows to convert an image from one type to another, maintaining or modifying specific details [34]. In our study, this is used to transform white markers to realistic GBI inclusions. To train this transformation a pairwise training was performed by selecting 21 of the original GBI images in Table 1 and pairing them with images of neutrophils with white markers manually placed in the GBI zones. This pairing allowed GAN C to learn how to convert marked white regions into realistic images closely resembling the original cells with GBI inclusions, obtaining natural transitions when considering neighboring cytoplasm pixels.

The design and training methodology of GAN C was consistently similar to the methodology proposed for SyntheticCellGAN [28], using the same hyperparameters. For data augmentation, the 21 paired images were subjected to three transformations (90-degree rotation, a vertical flip, and a horizontal flip), leading to 84 paired images in total. The training was performed by 5 epochs of separate training and reached convergence in 8 epochs of joint training.

The final outcome was a number of synthetic images of neutrophils that exhibited GBI. By adding these images to those in Table 1, the representation of the GBI class can be similar to other less represented classes such as HJBLI and BAC.

2.3. Structure of the Neutrophil Neural Network

In the development of an automatic classification system it is necessary to find a balance between the neural network architecture, data partitioning and class balancing. In this work, multiple experiments were carried out evaluating different configurations of these three components. At the end of these tests, the combination that provided the best performance in the classification was selected. The following subsections indicate the neural network architecture, the data partitioning and the balancing strategy adopted from the experiments carried out.

2.3.1. Partition and class balance for NeuNN learning and testing datasets

Due to imbalance of the image database, we had to reconsider the image splitting strategy. Although the conventional approach based on the 80/20 rule (Pareto principle) usually allocates 80% of the data for learning and 20% for testing [35], when applied in our case, this division did not provide an adequate representation of the least represented classes. To solve this problem, we determined that an approximate 50/50 random split for the learning and test sets was the most beneficial, especially for underrepresented classes. The number of images for each class in this partition are shown in Table 2. The test set was separated to be used exclusively for the assessment of the final model. This ensures an objective test, without interference from the learning process.

In turn, the learning set was split into training and validation, allocating 80% and 20% of the images, respectively. The training set provides the patterns necessary to train the model, while the validation set acts as a feedback mechanism during the training phase. By evaluating the performance of the model on the validation set, it is possible to identify and correct problems, such as overfitting, allowing the model to be trained iteratively until optimal performance is achieved. Finally, our approach establishes three distinct sets: train, validation, and test as detailed in Table 2.

To compensate the very low number (21) of images of GBI class, we used the GAN-based system described in Section 2.2.2 to generate 17 images for training and eight for validation, resulting in a hybrid dataset with a total of 34 images for training and eighth for validation. The number of samples generated is not too large to avoid excessive dependence on the synthetic GBI, which could affect the variability of the real data and minimize under-representation biases [36,37]. This strategy achieves a reasonable balance of images between the three HJBLI, GBI and BAC classes, helping to minimize disparities, avoiding

Table 2
Distribution of images in the training, validation and test sets. In the GBI group (*), data augmentation was achieved with GAN for 21 artificial sample creation. The training set comprises a hybrid blend of 17 real and 17 synthetic images, while the validation set includes 4 real and 4 synthetic images.

Types of neutrophils	Dataset		
	Learning		Test
	Train	Validation	
NEU	1850	463	2282
HYP	197	51	246
CRY	76	19	96
DB	56	14	69
HJBLI	30	7	34
GBI*	17*	4*	21
BAC	32	8	33

overfitting and generalization problems when used to train and test classification models [38].

After balancing these three classes against each other, random subsampling was necessary to counteract the marked predominance of the NEU class to avoid biased predictions by the classification models. In addition, upsampling was needed to increase the images of the other classes until achieving a global balance.

To do upsampling, possible variations that could affect the performance of the classifier in real conditions were simulated. These variations include differences in illumination, changes in cell orientation or position, variations in size, and distortions caused during the imaging process. To optimize the performance of NeuNN under these variations, random image transformations were implemented during the training process. These transformations, fine-tuned through a series of iterative experiments, demonstrated improved accuracy in the final classification tests. Among the techniques applied, rotations in a range of 0 to 360 degrees and vertical flips were included, allowing the model to learn to recognize cells regardless of their orientation. In addition, a random zoom was applied with a variability from $0.5X$ to $2.0X$, with the aim of adapting the model to different scales related to the size and capture distance of the cell. The brightness of the images was adjusted in a range of ± 0.4 to simulate various lighting conditions.

The end result was twofold: a balanced train set with 200 images for each class and a balanced validation set with 40 images for each class. This division guarantees a more uniform distribution of classes, allowing a complete learning process of the classifier [39].

2.3.2. Model selection

After addressing the strategies for the datasets configuration, we will now focus on the structure and operation of NeuNN. To determine the most effective neural network architecture, a comparative analysis was performed using the end balanced data set described in Section 2.3.1. Models evaluated include ConvNeXt [40], ResNet34, ResNet50, ResNet152 [41], ResNeXt [42], VGG16, VGG19 [43], ViT-G [44], EfficientNet-L2 [45] and EfficientNet-B7 [46]. The performance metric used was the accuracy obtained for the test set, and a 24 GB Nvidia Xeon Graphics Processing Unit (GPU) was used to train these models.

The results, summarized in Table 3, indicate the accuracy achieved by each architecture and the training time per epoch. All architectures exceeded 80% accuracy. However, it was EfficientNet-B7 that stood out, achieving an accuracy of 94%. After this analysis, EfficientNet-B7 was selected for the model in the NeuNN system. This architecture is a member of the EfficientNet family, known for its scaling method that consistently adjusts the width, depth and resolution of the network using predefined coefficients. These models are generally more compact compared to other convolutional networks, which positions them as a good choice for embedded systems. EfficientNet-B7 is one of the deepest models in this family as it consists of 66 million parameters, which gives a good balance between accuracy and complexity of the model.

Table 3
Comparison of neural network architectures: accuracy and training time per epoch (in seconds).

	Accuracy (%)	Training time per epoch (s)
ConvNeXt	91	71
ResNet34	81	38
ResNet50	83	42
ResNet152	87	65
ResNeXt	89	68
VGG16	92	47
VGG19	92	49
ViT-G	85	85
EfficientNet-L2	92	74
EfficientNet-B7	94	52

Table 4
Summary of NeuNN architecture based on EfficientNet-B7.

Input: 224 × 224 × 3		
Layers		
Conv2D	32 filters 3 × 3	112 × 112 × 32
	7 MBConv blocks	
MBConv 1	16 filters, 3 × 3 DConv, SE	112 × 112 × 16
MBConv 2	24 filters, 3 × 3 DConv, SE	56 × 56 × 24
MBConv 3	40 filters, 3 × 3 DConv, SE	28 × 28 × 40
MBConv 4	80 filters, 3 × 3 DConv, SE	28 × 28 × 80
MBConv 5	112 filters, 3 × 3 DConv, SE	14 × 14 × 112
MBConv 6	192 filters, 3 × 3 DConv, SE	7 × 7 × 192
MBConv 7	320 filters, 3 × 3 DConv, SE	7 × 7 × 320
Conv2D	1280 filters 7 × 7	7 × 7 × 1280
Dense	1280 nodes	
Dense	1000 nodes	
Softmax	7 outputs	

2.3.3. NeuNN architecture

The EfficientNet-B7 structure of the NeuNN classification model is shown in Fig. 4. The architecture is composed of multiple convolutional layers that progressively extract features from images of neutrophils with morphological alterations, ended with the fully connected classifier. Table 4 presents the summary of the NeuNN structure.

The architecture begins with an input layer that receives images of size 224 × 224 pixels with 3 color channels (RGB). The input is then processed by a 2D-type convolutional layer [47]. A 2D convolutional layer (Conv2D) is a classic convolution in CNN, filters are passed that sweep the input image to detect features such as edges, textures and patterns. In this case, 32 filters of size 3x3 are applied. A stride is applied during convolution to reduce the spatial dimensions of the image. At the output of the layer, 32 feature maps of 112 × 112 pixels are obtained.

Next come seven MBConv (Mobile Inverted Bottleneck Conv) blocks [48]. These blocks are characteristic of the EfficientNet architecture. The MBConv block is made up of a sequence of convolutional and recalculation layers to extract features, as illustrated in Fig. 4. This block is the main contributor to optimizing the use of computational resources, memory and feature extraction of EfficientNet. The process for the first MBConv block is explained below and it is the same for all seven MBConvs. The MBConv block starts with a convolution 1 × 1 to reduce the number of channels from 32 to 16 in order to simplify the representation of feature maps.

The result is then processed by a Depthwise Convolution (DConv) [49]. This convolution treats each channel independently, applying a specific filter for each channel instead of combining information between channels. This separate approach reduces the amount of calculations required by avoiding processing and combining all channels with each filter at once, which would be computationally expensive.

After the Depthwise Convolution, the process continues with the Squeeze-and-Excitation (SE) module [50]. The SE module dynamically adapts the weighting of each channel, ensuring that most important

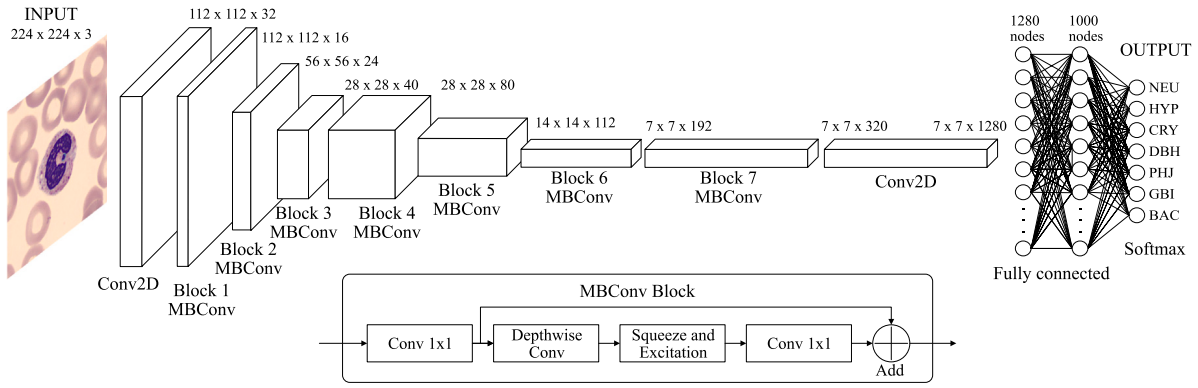

Fig. 4. Structure of the EfficientNet-B7 architecture used in NeuNN. The diagram represents the architecture in its entirety: starting with the input of a digital image of a neutrophil with morphological alterations, followed by successive processing stages in layers and convolutional blocks, including traditional convolutions and MBConv blocks, characteristic of EfficientNet, to extract the relevant features.
At the end of the process, the features pass to a fully connected layer that, through a softmax function, ends in the prediction between the seven possible output classes.

channels have a greater impact on the final representations. The process is carried out through two steps. The Squeeze step reduces the spatial information of each channel, generating a single value per channel that represents its relative importance. The Excitation step then recalculates and adjusts these values. This recalibration prioritizes channels that are most informative and reduces the relevance of less meaningful channels.

A convolution 1×1 is applied to the output of the SE to obtain 16 channels at the output. Finally, a residual connection Add is made that takes the output of the initial convolution 1×1 and adds it with the output of the last convolution 1×1, as shown in Fig. 4. This allows the original information to flow through the block and be combined with the processed features, giving the network a way to retain the initial useful information combined with the extracted features between the DConv and SE.

The seven sequentially applied MBConv blocks extract progressively more complex features by operating on smaller regions of the image. The Block1 MBConv transforms the feature into $112 \times 112 \times 16$. MBConv Block2 drops its resolution to $56 \times 56 \times 24$. Block3 MBConv then changes the resolution to $28 \times 28 \times 40$, and Block4 MBConv keeps the resolution but scales the channels to $28 \times 28 \times 80$. This progression continues through MBConv Block5, with a resolution of $14 \times 14 \times 112$, MBConv Block6 with $7 \times 7 \times 192$, and finally MBConv Block7 with a resolution and depth of $7 \times 7 \times 312$.

After these MBConv blocks, there is a 2D convolutional layer (Conv2D). It serves as a high-level extractor, that is, it allows even more complex features to be extracted from the abstract representations of the last MBConv, resulting in a feature map of size $7 \times 7 \times 1280$. The result is flattened into a one-dimensional vector and passed through fully connected layers. The first fully connected layer consists of 1280 nodes, followed by a second layer of 1000 nodes, as indicated in the Table 4. Although originally this 1000-node structure was designed to fit the ImageNet database [51], which has 1000 classes, in our context it serves as an intermediate layer before the final decision layer. The final Softmax layer with seven outputs assigns the probabilities to each of the classes of morphological alterations in neutrophils. The class with the highest probability is considered the final prediction of the model (output).

2.4. Training the Neutrophil Neural Network

The training of the classifier was carried out using the balanced set with 200 images for each class. The Adam optimization algorithm was used with an initial learning rate of 0.001 [52]. The multiclass cross entropy loss function presented for deep learning model in [53], was chosen due to its wide effective implementation in the literature in multiclass classification problems:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} \cdot \log(\hat{y}_{ic}) \quad (1)$$

In this equation C represents the number of classes (7); y_{ic} is 1 if the actual class of the sample i is c, and 0 otherwise; and $\hat{y}_{ic}$ indicates the predicted probability that the sample i belongs to the class c. This function calculates the average loss per sample in a batch, quantifying the discrepancy between the model predictions and the true labels.

Having a training set with 1400 images, it was necessary to divide this set into batches of 20 images to manage the computational load. Each batch was processed through the CNN in what is called an iteration. With 70 batches per training set, each complete cycle through the 70 batches is known as a learning epoch. After completing each epoch, the performance of the CNN was evaluated using the balanced validation set with 40 images for each class.

To measure the effectiveness of the model, metrics such as training and validation loss were calculated, using the cross entropy in Eq. (1), as well as the overall accuracy of the validation set. This evaluation process was repeated at the end of each epoch.

In this study, the determination of the number of training epochs was based on the analysis of the validation set metrics. Initially the number of training epochs was pre-established, monitoring validation loss and accuracy as each learning epoch passed. However, an excessive number of epochs can lead to model overfitting, while an insufficient number can result in poor performance. To avoid overfitting and achieve adequate performance, the early stopping technique [54] was implemented. This technique involves monitoring validation metrics at each epoch and stopping training when no improvements in these metrics are observed after a predefined number of epochs. To determine if there is an improvement, the validation loss of the current epoch is compared to that of the previous epoch. The model is considered to improve only if the reduction in validation loss exceeds a preset threshold, which, through testing, was set to 0.05. If this improvement is not achieved for three consecutive epochs, training stops automatically. At this point, the model state corresponding to the last significant improvement observed in validation loss is preserved. The model training process was carried out over 52 epochs as shown in Fig. 5, where the training and validation losses and the accuracy of the validation set are presented.

3. Quantitative evaluation of NeuNN

In this section, specific metrics are presented to evaluate the performance of the NeuNN model, focusing on its classification ability. For an

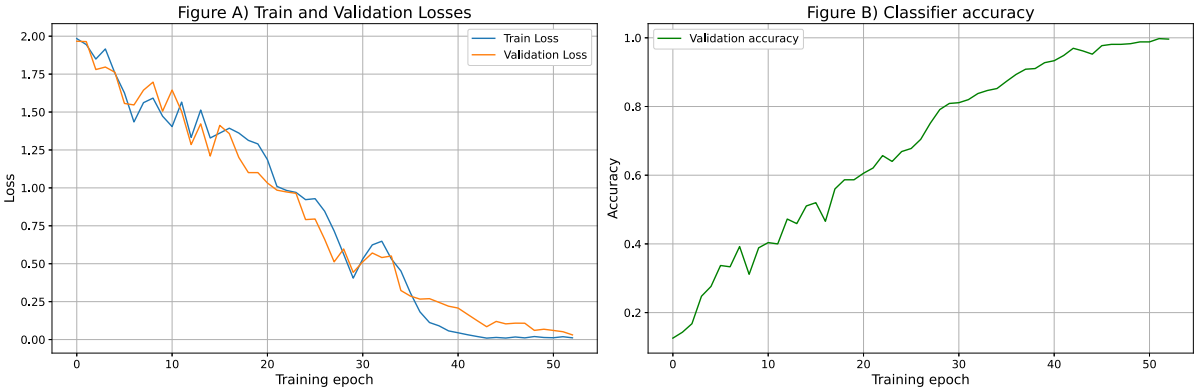

Fig. 5. Evolution of classifier training over 52 epochs: (A) Progression of training and validation losses, training loss measures how well the model fits the training data while validation loss measures how well the model fits on new data. (B) Progression of image classification accuracy from the validation set, indicating the percentage of images correctly classified by the model over time.

Table 5
Confusion Matrix for the classification performed by NeuNN on the test set. The absolute results and the number of images correctly identified per class are shown. The classification for the classes Normal neutrophils, Hypogranulation, Cryoglobulins, Döhle Bodies (DB), Pseudo Howell–Jolly Bodies, Green-blue inclusions of death and Bacteria had accuracy percentages per class respectively of 100%, 97%, 99% , 80%, 94%, 95% and 95%. The model has an overall accuracy of 94.3%.

		Predicted class						
		NEU	HYP	CRY	DB	HJBLI	GBI	BAC
True class	Normal neutrophils (NEU)	**1.00** (2282)	0	0	0	0	0	0
	Hypogranulation (HYP)	0	**0.97** (239)	0.01 (1)	0.02 (6)	0	0	0
	Cryoglobulins (CRY)	0	0	**0.99** (95)	0	0	0.01 (1)	0
	Döhle bodies (DB)	0	0.20 (14)	0	**0.80** (55)	0	0	0
	Howell-Jolly body-like inclusions (HJBLI)	0	0.06 (2)	0	0	**0.94** (32)	0	0
	Green-blue inclusions of death (GBI)	0.05 (1)	0	0	0	0	**0.95** (20)	0
	Bacteria (BAC)	0	0	0.05 (2)	0	0	0	**0.95** (31)

Table 6
Sensitivity, specificity, precision, F1 score, Jaccard Index and Matthews correlation coefficient (MCC) values of the classification results of the test set.

Classes	Sensitivity	Specificity	Precision	F1 score	Jaccard Index	MCC
Normal neutrophils (NEU)	1.00	0.992	0.952	0.976	0.952	0.972
Hypogranulation (HYP)	0.97	0.957	0.789	0.870	0.770	0.852
Cryoglobulins (CRY)	0.99	0.990	0.943	0.966	0.934	0.960
Döhle bodies (DB)	0.80	0.997	0.976	0.879	0.784	0.867
Howell-Jolly body-like inclusions (HJBLI)	0.94	1.000	1.000	0.969	0.940	0.965
Green-blue inclusions of death (GBI)	0.95	0.998	0.969	0.969	0.940	0.965
Bacteria (BAC)	0.95	1,000	0.974	0.974	0.950	0.971
Average values	0.94	0.991	0.943	0.943	0.896	0.936

objective quantitative evaluation, we used the test set, composed of cell images that did not participate in the training or validation phases of the model. This set includes a total of 2781 images, the details of which can be found in the Table 1. Their classification using the NeuNN model is summarized in the confusion matrix in Table 5, where the predicted classes (in columns) are compared with the true classes verified by clinical diagnosis (in rows).

From the confusion matrix, several metrics were calculated to evaluate the performance of the NeuNN model proposed in [55]. These metrics are the sensitivity or true positive rate (TPR), which measures the model's ability to correctly identify positive cases. The specificity, or true negative rate (TNR), evaluates how well the model correctly identifies negative cases. Precision or positive predictive value (PPV), which determines the proportion of positive classifications that are correct. The F1 Score, which is the average of the precision and sensitivity. The non-uniform distribution of the seven classes in our data set introduces a class imbalance, which can lead to a bias in conventional evaluation metrics towards the majority class. To address this issue, additional evaluation metrics have been incorporated. These include the Jaccard Index and the Matthews Correlation Coefficient (MCC) proposed in [56–58]. The Jaccard Index measures the proportion of true positives to the union of predicted and actual cases. This metric minimizes the impact of a high number of true negatives, which is common in imbalanced data sets. On the other hand, the MCC provides an effective evaluation of the quality of binary classifications, even in situations with highly disproportionate classes. For the evaluation of the metrics, each class is considered in a binary way, treating the class of interest as one category and grouping all others into the opposite category. The expressions to calculate these metrics are as follows:

$$\text{TPR} = \frac{TP}{TP+FN} \qquad \text{TNR} = \frac{TN}{TN+FP}$$

$$\text{PPV} = \frac{TP}{TP+FP} \qquad \text{F1 Score} = 2 \cdot \frac{\text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}}$$

$$\text{Jaccard index} = \frac{TP}{TP+FP+FN}$$

$$\text{MCC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}}$$

where TP are the true positives, FN the false negatives, FP the false positives, and TN the true negatives. These values are obtained from the confusion matrix.

The results of the performance metrics of each class are given in Table 6. Analysis of the results from this table indicates a high overall performance of the model, with a global accuracy of 94.3%. Classes such as Normal Neutrophils (NEU) and Howell–Jolly body-like inclusions (HJBLI) stand out for their high sensitivity (1.00 and 0.94

respectively) and specificity (0.992 and 1.000 respectively). However, a lower performance is observed for the Döhle Bodies (DB) with sensitivity 0.80 and Jaccard index (0.784), and for Hypogranulation (HYP) with some lower scores, which require additional attention. These cases will be discussed in more detail in the next section.

4. Discussion

The identification of morphological alterations in circulating neutrophils in peripheral blood is of clinical interest. The presence of hypogranulation or different types of cytoplasmic inclusions in neutrophils can be indicative of a wide range of pathological conditions, including myelodysplastic disorders, bacterial infections, immunosuppression, acute liver failure, among others. Although the automated image analysis systems currently used in clinical laboratories allow the preclassification of the normal leukocytes, they are not able to detect any type of cytoplasmic abnormalities and inclusions inside neutrophils. Our study proposes a CNN-based system with EfficientNet B7 architecture for the automatic classification of hypogranulated neutrophils and neutrophils with five types of cytoplasmic inclusions. To the authors' knowledge, there is no previous research in the literature addressing this objective.

The remaining content of this section is dedicated to discussing the following points: model development and training process, interpretation of the results, and demands on computational resources.

4.1. Considerations about the development and training process

The development of NeuNN required the implementation of multiple strategies to overcome the challenges associated with the training data set. One of the main issues was the imbalance between classes, with a predominance of normal neutrophils (4595 images) compared to less represented categories such as HJBLI, BAC and GBI, the latter being the one that presented the smallest number of available images. To address underrepresentation, classical data balancing techniques were first applied, which were effective for HJBLI and BAC classes, but insufficient for GBI.

For the GBI case, different data augmentation techniques were tested, such as geometric transformation and style transfer. However, the strategy that showed a significant impact on improving NeuNN training was the incorporation of synthetic images for GBI, generated by a GAN system, creating a hybrid dataset. During several iterations of NeuNN training, we noticed that overloading artificial images was counterproductive to the correct classification of real images. The model, during its learning stage, demonstrated high accuracy in identifying the synthetic images within the validation set, however, it incorrectly classified the real images belonging to the GBI class from the same set, suggesting an artificial overrepresentation. Consequently, we adjusted the proportion of synthetic images until we reached a balance that minimized biases in learning the GBI class. This adjustment resulted in the three minority classes reaching a similar number of samples, balancing the data set without needing to resort to artificial augmentation for the HJBLI and BAC classes.

Another challenge in the learning process was determining the optimal number of epochs to achieve adequate convergence, that is balanced between the risk of overfitting with too many epochs and the danger of an undertrained model with insufficient epochs. The early stopping technique was implemented, which consists of monitoring the loss metric in the validation set after each training epoch. The decision to apply early stopping was preferred over other anti-overfitting strategies, such as regularization or dropout. This choice was based on a comparison that indicated early stopping as the method with the highest classification performance in the validation set. We set a minimum improvement for early stopping technique threshold of 0.05 in validation set loss reduction as a criterion to determine that the model was still learning. This value was selected after an experimentation process, in which the accuracy was measured and compared by epoch in both the training and validation sets, paying attention to the difference in metrics between the two. We observed that performance tended to be similar on both sets when the improvement threshold of 0.05 was applied. The model was trained with early stopping until it was automatically stopped after no improvement was recorded with the set threshold for three consecutive epochs, which occurred after 52 epochs.

Additionally, we explore the effects of varying the depth of the EfficientNet architecture on the performance of NeuNN. By experimenting with different configurations, from EfficientNet-B0 to EfficientNet-B7, we sought to identify the optimal balance between model complexity and computational efficiency. Initial tests with shallower versions, such as B0, B1, and B2, showed faster training times, but were only effective in classifying normal neutrophils (NEU). Configurations such as B3, B4, B5 and B6 were computationally more demanding, improving classification performance for the NEU, HYP and CRY classes, but indicating poor performance in the classes with few samples. The deepest configurations, EfficientNet-B7 and EfficientNet-L2, exhibited improvements in all classes, achieving high accuracy, as shown in Table 3. EfficientNet-L2, being significantly larger and more complex, could be considered for applications with high computational resources. On the other hand, EfficientNet-B7, with a large depth and a high number of filters per layer, in our study showed a notable reduction in training times compared to L2 and achieved higher accuracy. The final selection of EfficientNet-B7 was based on these parameters. The NeuNN system was designed with future implementation in mind, so this architecture offers feasibility for implementations on computers with moderate computational characteristics.

4.2. Interpreting performance metrics

The performance of the NeuNN on the test set (Table 2) was first evaluated using conventional metrics such as sensitivity, specificity, precision and F1 score. Table 6 shows that HJBLI, BAC, GBI and DB exhibited comparatively lower sensitivity while having very high specificity. This is particularly evident in the Döhle bodies (DB) with the lowest sensitivity (0.80) and a specificity of 0.997. It is also notable that the hyproganulation class (HYP) has the lowest precision (0.789).

To counteract the influence of the non-uniform distribution of samples in the test case, and carry out a more equitable evaluation, it was decided to use alternative metrics such as the Jaccard index and the Matthews correlation coefficient (MCC). These metrics are useful in evaluating machine learning systems where it is common to find underrepresented or overrepresented categories [56]. The Jaccard index measures the similarity between the sets of true positives and the union of the predicted and real cases, minimizing the effect of the excess of true negatives typically found in unbalanced data. The values obtained for this metric indicated an acceptable overall performance of NeuNN, with an average of 0.896 of the seven classes evaluated. On the other hand, the MCC offers a balanced evaluation for both unbalanced data and overlapping classes, indicating considerable classification ability with an average value of 0.936 on the test set.

When analyzing the value of the Jaccard and MCC metrics for each class, we observe that their lowest values occur in the DB and HYP classes, with lowest values between 0.770 and 0.867, while in the other classes these values exceed 0.94. The particular relation between DB and HYP is also observed in the confusion matrix in Table 6, where 14 (20%) of the images originally labeled as DB are predicted as HYP by the model. The next section is devoted to analyze these images in more detail.

4.3. Interpretation of DB and HYP misclassifications

To understand the classification of the 14 images under analysis, Fig. 6 (A) displays the probabilities given by the softmax function at the end of the NeuNN model (see Fig. 4) to each of the seven possible classes, taking into account that the class with the highest probability is the final prediction. Four images labeled as DB (4,8,9 and 10) were predicted as HYP with high probabilities (above 0.95). In contrast, the remaining 10 images showed a similar probability for both classes DB and HYP. A complementary view is given in Fig. 6 (B) using a bubble plot for these 10 images. The center of each bubble represents the average softmax probabilities per class, and the size is proportional to the corresponding standard deviation. We observe that the bubbles of DB and HYP classes are very close. Since the HYP bubble center is slightly higher, it is understandable that the images were predicted by the model as belonging to class HYP, but the proximity of the DB probability indicates uncertainty in the classification, suggesting feature overlap between these two classes.

After this observation, the set of 14 images was evaluated by an expert clinical pathologist. Interestingly, the clinical pathologist confirmed that only the four mentioned images (4,8,9 and 10) were erroneously classified, while the remaining 10 images were neutrophils showing the presence of both cytoplasmic abnormalities, DB inclusions and hypogranulation. The explanation is that in some myelodysplastic syndromes, Döhle body inclusions and simultaneous neutrophil hypogranulation can be observed. Fig. 7 shows one of the 10 images that simultaneously presents a DB inclusion and a consistent hypogranularity in its cytoplasm. Therefore, it is explainable that the classification model assigns similar probabilities to both DB and HYP morphological alterations since they coexist in the cell image. Considering that in practice it is possible to find more than one morphological alteration in the same image, alternative techniques such as multi-label classification [59] and fuzzy logic [60,61] could be explored to determine the degrees of membership to each morphological alteration.

4.4. Limitations in the implementation and generalization of the system

One of the main limitations of the NeuNN system is related to the size of the data set available for training. Although data augmentation techniques were applied to compensate for underrepresented classes, the total number of images of such classes remains limited. The sample size is restrictive due to the low frequency of cytoplasmic inclusions in patient samples. Our data set was obtained over several years of clinical analyses, so expanding it with new cases can be time-consuming and requires continued efforts by pathologists.

Regarding the proposed dataset, evaluation metrics such as accuracy and sensitivity may be biased by the imbalance between classes. While addressed using alternative metrics such as the Jaccard index and Matthews correlation coefficient, additional research is needed on new fair evaluation approaches for unbalanced sets.

Another limitation is found in the high computational demand that the training of the proposed methodology represents, challenging its implementation in environments with limited resources. The model required a 24 GB Nvidia Xeon GPU for its robust processing. However, these requirements may restrict accessibility for institutions with more modest hardware. Furthermore, the choice of the EfficientNet B7 network with its 66 million parameters requires high-performance computing equipment to train it from scratch.

Finally, running the system in clinical laboratories from different hospitals introduces difficulties for the generalization of the model beyond the specific training conditions. NeuNN was trained on samples from a single center with standardized protocols, using samples prepared with May Grünwald-Giemsa staining and analyzed using the CellaVision DM96 system. Although we sought to increase tolerance to variations using image transformations in the data augmentation process, the diversity of conditions in clinical practice represents a challenge for generalization. Future work will address this limitation using deep learning methodologies to adapt the NeuNN system to the diversity of equipment and staining protocols used in different hospital environments.

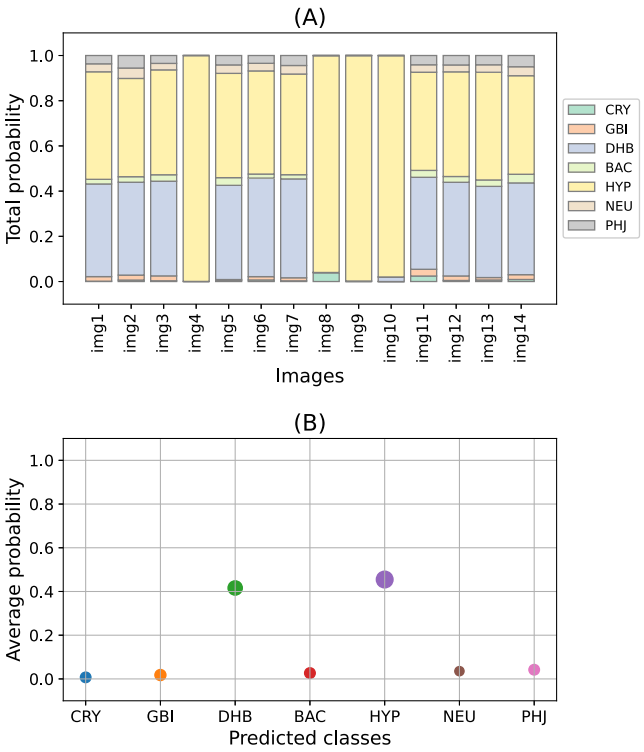

Fig. 6. Interpretation of Döhle bodies (DB) and hypogranulation (HYP) classification by the NeuNN model for 14 particular images. (A) For each image, a composite bar shows the probability assigned by the model to each neutrophil class under study. (B) Bubble plot for the 10 images that exhibit similar probabilities for DB and HYP, where centers indicate the average probability for each predicted class, and the bubble size reflects the standard deviation.

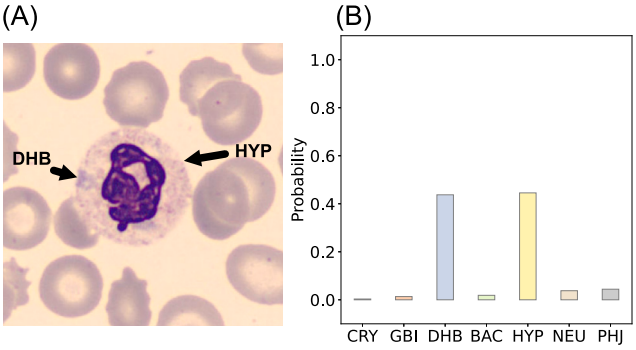

Fig. 7. (A) Image of a neutrophil showing hypogranulation and presence of a Döhle body inclusion. (B) Bar graph indicating the seven probabilities calculated by the model for this neutrophil. Both HYP and DB classes present comparable probabilities.

5. Conclusion

This work presented NeuNN, a system for the automatic detection and classification of cytoplasmic inclusions in peripheral blood neutrophils. The model showed good performance identifying normal neutrophils and those with hypogranulation, cryoglobulins, Döhle bodies, Howell–Jolly body-like inclusions, blue–green inclusions of death and bacteria. Additionally, this research contributed with the creation of a database of annotated images of these neutrophil abnormalities and

a GAN-based model for generating artificial images for classes with low prevalence.

As digital image analyzers and computational methods continue to advance rapidly, the approach proposed in this work serves as a basis in the automatic recognition of morphological alterations in neutrophils, which can be improved and expanded in future works and be useful as support tool for clinical pathologists.

CRediT authorship contribution statement

Kevin Barrera: Writing – original draft, Validation, Software, Methodology, Investigation. **José Rodellar:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Santiago Alférez:** Writing – review & editing. **Anna Merino:** Writing – review & editing, Visualization, Validation, Supervision, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgment

This work is part of a research project funded by the Ministry of Science and Innovation of Spain, with reference PDC2022-133514-I00.

K. Barrera appreciates the financial support of the Secretary of Higher Education, Science, Technology and Innovation of Ecuador as a personal sponsor entity.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.108691.

References

[1] A. Merino, Manual de Citología de Sangre Periférica y Líquidos Biológicos, Editorial Panamericana, 2019.
[2] H.K. Lehman, B.H. Segal, The role of neutrophils in host defense and disease, J. Allergy Clin. Immunol. 145 (6) (2020) 1535–1544.
[3] F.R. DeLeo, L.-A.H. Allen, Phagocytosis and neutrophil extracellular traps, Facul. Rev. 9 (2020).
[4] A. Othman, M. Sekheri, J.G. Filep, Roles of neutrophil granule proteins in orchestrating inflammation and immunity, FEBS J. 289 (14) (2022) 3932–3953.
[5] M. Ong, D.M. Veillon, M.L. Nordberg, J. Cotelingham, Karyopyknotic cytoplasmic inclusions in neutrophils, J. La State Med. Soc. 162 (2010) 228–230.
[6] B. Bain, Blood Cells: A Practical Guide, Blackwell Science Ltd, London, 2003.
[7] F. Wang, G. Wang, X. Wang, Cryoglobulin crystals deposited in a peripheral blood film and phagocytosed by neutrophils, Br. J. Haematol. 195 (4) (2021) 484–484.
[8] A. Merino, C. Fernández de Larrea, B. J., Cytoplasmic inclusions in leukocytes associated to cryoglobulinemia and igg-kappa monoclonal gammapathy of undetermined significance, Eur. J. Haematol. 47 (86) (2011) 550.
[9] M. Akhtari, E. Waller, Howell-Jolly body-like inclusions in neutrophils, Blood 114 (4) (2009) 2860.
[10] H. Abdel-Monem, A. Prakasam, P. Thiagarajan, Body–like inclusions in neutrophils of a transplant recipient in association with ganciclovir therapy, Arch. Pathol. Lab. Med. 134 (2010) 809–810.
[11] B. Revol, A. Thiebaut-Bertrand, M. Carré, C.-E. Bulabois, P. Mossuz, J. Mondet, Quantification of Howell-Jolly body-like inclusions in ganciclovir toxicity using CellaVision DM 96 analyser, Br. J. Haematol. 174 (4) (2016) 637–639.
[12] M. Asad, N. Pathan, P. Fiedler, Neutrophilic Howell-Jolly body-like inclusions in concurrent hiv and sars Covid-19 virus infections, Am. J. Clin. Path. 160 (Supplement_1) (2023) S70–S71.
[13] R. Vicente-Steijn, A. Tomé, F. Maduell, M. Xipell, P. Castro, A. Molina, A. Merino, Green inclusions in neutrophils: A critical finding that must be reported, Int. J. Lab. Hematol. 42 (3) (2020) e101–e104.
[14] D. Welder, W. Chen, Green–blue inclusions in neutrophils and monocytes: A critical finding, Int. J. Hematol. 115 (2) (2022) 149–150.
[15] A. Merino, A. Molina, M. Rodríguez-García, R. Vicente-Steijn, A. Tomé, F. Maduell, M. Xipell, P. Castro, Detection and significance of green inclusions in peripheral blood neutrophils and monocytes, Int. J. Lab. Hematol. 43 (2) (2021) e92–e94.
[16] A. Fife, D. Hill, C. Barton, P. Burden, Gram negative septicaemia diagnosed in peripheral blood smear appearances, J. Clin. Pathol. 47 (1993) 82–84.
[17] L. Bigorra, A. Merino, The unusual presence of streptococcus gallolyticus within neutrophils in a patient with endocarditis and brain abscesses, Br. J. Haematol. 169 (2015) 308.
[18] E. Boero, I. Brinkman, T. Juliet, E. van Yperen, J.A. van Strijp, S.H. Rooijakkers, K.P. van Kessel, Use of flow cytometry to evaluate phagocytosis of staphylococcus aureus by human neutrophils, Front. Immunol. 12 (2021) 635825.
[19] S. Swerdlow, E. Campo, N. Harris, E. Jaffe, S. Pileri, H. Stein, et al., WHO Classification of Tumours of Haematopoietic and Lymphoid tissues, International Agency for Research on Cancer, 2017.
[20] A. Kratz, S.-h. Lee, G. Zini, J.A. Riedl, M. Hur, S. Machin, I.C. for Standardization in Haematology, Digital morphology analyzers in hematology: ICSH review and recommendations, Int. J. Lab. Hematol. 41 (4) (2019) 437–447.
[21] L. Vogado, R. Veras, F. Araujo, R. Silva, Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification, Eng. Appl. Artif. Intell. 72 (2018) 415–422.
[22] L. Boldú, A. Merino, A. Acevedo, A. Molina, J. Rodellar, A deep learning model (ALNet) for the diagnosis of acute Leukaemia lineage using peripheral blood cell images, Comput. Methods Programs Biomed. 202 (2021) 105999.
[23] S. Saleem, J. Amin, M. Sharif, G.A. Mallah, S. Kadry, A.H. Gandomi, Leukemia segmentation and classification: A comprehensive survey, Comput. Biol. Med. 150 (2022) 106028.
[24] Q. Wang, S. Bi, M. Sun, Y. Wang, D. Wang, S. Yang, Deep learning approach to peripheral Leukocyte recognition, PLoS ONE 14 (6) (2019) e0218808.
[25] A. Acevedo, S. Alférez, A. Merino, L. Puigví, J. Rodellar, Recognition of peripheral blood cell images using convolutional neural networks, Comput. Methods Programs Biomed. 180 (2019) 105020.
[26] K. Kimura, Y. Tabe, T. Ai, I. Takehara, H. Fukuda, H. Takahashi, et al., A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA, Sci. Rep. 9 (2019) 13385.
[27] A. Acevedo, A. Merino, L. Boldú, A. Molina, S. Alférez, J. Rodellar, A new convolutional neural network predictive model for the automatic recognition of hypogranulated neutrophils in myelodysplastic syndromes, Comput. Biol. Med. 134 (2021) 104479.
[28] K. Barrera, A. Merino, A. Molina, J. Rodellar, Automatic generation of artificial images of Leukocytes and Leukemic cells using generative adversarial networks (syntheticcellgan), Comput. Methods Programs Biomed. 229 (2023) 107314.
[29] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48, http://dx.doi.org/10.1186/S40537-019-0197-0/FIGURES/33, URL https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0.
[30] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A.A. Bharath, Generative adversarial networks: An overview, IEEE Signal Process. Mag. 35 (1) (2018) 53–65.
[31] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, 2017, arXiv:1701.07875.
[32] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690.
[33] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[34] K. Barrera, J. Rodellar, S. Alférez, A. Merino, Automatic normalized digital color staining in the recognition of abnormal blood cells using generative adversarial networks, Comput. Methods Programs Biomed. 240 (2023) 107629.
[35] D.M. McCarthy, R.S. Winer, The Pareto rule in marketing revisited: is it 80/20 or 70/20? Market. Lett. 30 (2019) 139–150.
[36] Y. Yang, Z. Xu, Rethinking the value of labels for improving class-imbalanced learning, Adv. Neural Inf. Process. Syst. 33 (2020) 19290–19301.
[37] A. Sahoo, A. Singh, R. Panda, R. Feris, A. Das, Mitigating dataset imbalance via joint generation and classification, in: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16, Springer, 2020, pp. 177–193.
[38] W. Li, J. Gou, Z. Fan, Session-based recommendation with temporal convolutional network to balance numerical gaps, Neurocomputing 493 (2022) 166–175.
[39] I. Muraina, Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts, in: 7th International Mardin Artuklu Scientific Research Conference, 2022, pp. 1–9.
[40] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.
[41] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[42] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1492–1500.
[43] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[44] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.
[45] Q. Xie, M.-T. Luong, E. Hovy, Q.V. Le, Self-training with noisy student improves imagenet classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10687–10698.
[46] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
[47] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[48] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.
[49] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
[50] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[51] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2009, pp. 248–255.
[52] Z. Zhang, Improved adam optimizer for deep neural networks, in: 2018 IEEE/ACM 26th International Symposium on Quality of Service, IWQoS, Ieee, 2018, pp. 1–2.
[53] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[54] L. Prechelt, Early stopping-but when? in: Neural Networks: Tricks of the Trade, Springer, 2002, pp. 55–69.
[55] D.M. Powers, Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation, 2020, arXiv preprint arXiv:2010.16061.
[56] A.W. Setiawan, Image segmentation metrics in skin lesion: accuracy, sensitivity, specificity, dice coefficient, jaccard index, and matthews correlation coefficient, in: 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia, CENIM, IEEE, 2020, pp. 97–102.
[57] D. Chicco, G. Jurman, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Gen. 21 (2020) 1–13.
[58] D. Chicco, Ten quick tips for machine learning in computational biology, BioData Min. 10 (1) (2017) 35.
[59] W. Qiu, J. Guo, X. Li, M. Xu, M. Zhang, N. Guo, Q. Li, Multi-label detection and classification of red blood cells in microscopic images, in: 2020 IEEE International Conference on Big Data (Big Data), IEEE, 2020, pp. 4257–4263.
[60] M. Versaci, G. Angiulli, P. Crucitti, D. De Carlo, F. Laganà, D. Pellicanò, A. Palumbo, A fuzzy similarity-based approach to classify numerically simulated and experimentally detected carbon fiber-reinforced polymer plate defects, Sensors 22 (11) (2022) 4232.
[61] S.K. Singh, V. Abolghasemi, M.H. Anisi, Fuzzy logic with deep learning for detection of skin cancer, Appl. Sci. 13 (15) (2023) 8927.