Computer Methods and Programs in Biomedicine 229 (2023) 107314

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Automatic generation of artificial images of leukocytes and leukemic cells using generative adversarial networks (syntheticcellgan)☆

Kevin Barrera[a], Anna Merino[b], Angel Molina[b], José Rodellar[a,*]

[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

Article history:
Received 29 July 2022
Revised 29 November 2022
Accepted 14 December 2022

Keywords:
Blood cell synthetic images
Blood cell automatic recognition
Deep learning
Convolutional neural networks
Generative adversarial networks
Image generation

ABSTRACT

Background and objectives: Visual analysis of cell morphology has an important role in the diagnosis of hematological diseases. Morphological cell recognition is a challenge that requires experience and in-depth review by clinical pathologists. Within the new trend of introducing computer-aided diagnostic tools in laboratory medicine, models based on deep learning are being developed for the automatic identification of different types of cells in peripheral blood. In general, well-annotated large image sets are needed to train the models to reach a desired classification performance. This is especially relevant when it comes to discerning between cell images in which morphological differences are subtle and when it comes to low prevalent diseases with the consequent difficulty in collecting cell images. The objective of this work is to develop, train and validate SyntheticCellGAN (SCG), a new system for the automatic generation of artificial images of white blood cells, maintaining morphological characteristics very close to real cells found in practice in clinical laboratories. Methods: SCG is designed with two sequential generative adversarial networks. First, a Wasserstein structure is used to transform random noise vectors into low resolution images of basic mononuclear cells. Second, the concept of image-to-image translation is used to build specific models that transform the basic images into high-resolution final images with the realistic morphology of each cell type target: 1) the five groups of normal leukocytes (lymphocytes, monocytes, eosinophils, neutrophils and basophils); 2) atypical promyelocytes and hairy cells, which are two relevant cell types of complex morphology with low abundance in blood smears. Results: The images of the SCG system are evaluated with four experimental tests. In the first test we evaluated the generated images with quantitative metrics for GANs. In the second test, morphological verification of the artificial images is performed by expert clinical pathologists with 100% accuracy. In the third test, two classifiers based on convolutional neural networks (CNN) previously trained with images of real cells are used. Two sets of artificial images of the SCG system are classified with an accuracy of 95.36% and 94%, respectively. In the fourth test, three CNN classifiers are trained with artificial images of the SCG system. Real cells are identified with an accuracy ranging from 87.7% to 100%. Conclusions: The SCG system has proven effective in creating images of all normal leukocytes and two low-prevalence cell classes associated with diseases such as acute promyelocyte leukemia and hairy cell leukemia. Once trained, the system requires low computational cost and can help augment high-quality image datasets to improve automatic recognition model training for clinical laboratory practice.

1. Introduction

Peripheral blood (PB) is the living tissue that circulates throughout the human body carrying nutrients and oxygen. It contains white blood cells or leukocytes, red blood cells and platelets suspended in plasma. Access to PB is minimally invasive, so its analysis is very useful in clinical diagnosis. Leukocytes play an important role in the human immune system, defending the body from infectious microorganisms or invasive foreign substances. Normal

☆ This document is a result of the research project "Computational hematopathology: deep learning solutions for the diagnosis of hematological diseases from peripheral blood cell images (CellsiMaticDeep)". Funded by the Ministry of Science and Innovation of Spain, PID2019-104087RB-I00.

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

https://doi.org/10.1016/j.cmpb.2022.107314

mature leukocytes are classified into five groups: neutrophils, lymphocytes, monocytes, eosinophils, and basophils. When leukocytes do not follow the normal maturation process, an invasive proliferation of abnormal cells without control (cancer) occurs. When this happens in bone marrow, these abnormal immature cells (blasts) go massively to the PB, and produce the diseases known as acute leukaemia. When uncontrolled proliferation occurs in the lymphatic system, the diseases are called lymphomas, and abnormal lymphocytes can be found in the PB.

Identification of normal and abnormal cell types is done by visual morphological inspection of the blood smear. This analysis is the basic principle for the assessment of more than 80% of hematological diseases [1]. Purely human visual assessment is prone to subjectivity and variability and requires expert personnel.

In modern clinical laboratories, the current trend is to perform morphological analysis with a multidisciplinary combination between the knowledge of the clinical pathologist and medical instruments that implement classification algorithms. The process begins with a blood smear, goes through a microscope and an analyzer that provides digital images of cells for the clinical pathologists to perform a visual inspection. They use qualitative descriptions of cell morphology to identify and diagnose different types of leukemia and lymphoma [1].

In this scenario, more and more artificial intelligence methods have been proposed to assist in hematological diagnosis, particularly through machine learning and, more recently, deep learning models for automatic recognition of morphological patterns associated with specific hematological diseases [2–4]. They complement the visual review by adding quantitative objectivity and consistency.

The works in [5,6] are examples of CNN models proposed to classify the different normal leukocyte groups. Automatic recognition of acute leukemia has been mainly approached in two different ways: differentiating lymphoblasts and leukocytes [7,8] and classifying various lymphoblast lineages [9,10]. A CNN model was developed to distinguish neoplastic (leukemia) from non-neoplastic diseases (infections), as well as to recognize the lineage of leukemia [11].

Ideally, large, well-annotated image sets are required to train CNN models. As in many other medical domains, collecting image data sets is a delicate task that needs specific procedures and technology. In the case of PB cell images, strict standardized procedures are followed to stain blood smears and capture single cell images using a microscope and acquisition system. These procedures are followed by the labeling of each cell by expert clinical pathologists. Quality assessment is critical to clearly highlight the morphological features that can help discern between the wide variety of cell types. This is especially relevant in cases where morphological differences are subtle and in cases of low prevalence diseases, where it is difficult to collect a large number of samples.

To solve the low abundance of CNN training images, data augmentation techniques are often used through simple manipulations of the images, such as rotation, flip, contrast variation, luminosity, focal length, among others. These manipulations help create new cell samples for learning. However, these techniques have a limit of variability, so the database does not significantly change [12].

New advanced techniques for synthesizing artificial images have recently been proposed for database augmentation, such as neural-style transfer [13], adversarial training [14] and the generative adversarial networks (GANs) [15]. GANs use two models that are trained in a contradictory scenario: one model creates the artificial images and the other discriminates the artificial images from the real ones, providing a loss function to the first model to be optimized.

The objective of this research is to exploit the potential of GAN models to produce synthetic images of cells circulating in peripheral blood with morphological characteristics very close to real cells.

We focus on the five groups of normal leukocytes plus atypical promyelocytes (AP) and hairy cells (HC). The former are included because they are the most common in PB and their most relevant morphological characteristics are quite different from each other. This makes clinical pathologists familiar with its visual recognition. Consequently, it is natural that the first step for a new artificial system is the ability to create samples of these normal leukocytes. AP are associated to acute promyelocytic leukemia (APL), the detection of which is of the utmost importance since patients can suffer severe bleeding and die if prompt treatment is not initiated. These abnormal promyelocytes have a much more complex morphology, since they present a nuclear bilobed profile and abundant immature granulation [16]. Also complex in morphology, the hairy cells are small mature lymphocytes with oval nuclei and abundant greyish-blue cytoplasm with villous projections. Its presence in PB is indicative of hairy cell leukemia (HC) [17].

Both AP and HC are considered low-prevalence diseases and their characteristic abnormal cells are not abundant in PB smears, making them relevant challenges for artificial image synthesis.

The contributions of this paper are twofold:

1. We design a new two-stage system. In the first stage we use a Wasserstein structure to create basic low-resolution images. In the second stage, we use the concept of image-to-image translation to build specific models that generate high-resolution final images with the realistic morphology of each of the target groups. The training relies on data sets collected from patients in the daily practice of a reference hospital assessed and annotated by expert clinical pathologists. The diagnosis of patients with AP and HC have been confirmed by other complementary tests in the hospital.
2. We evaluate the created images in four different scenarios. The first is a quantitative evaluation of the realism of the artificial images created by the system using metrics for GAN. The second occurs in a clinical environment, where a group of expert clinical pathologists classify the artificial cells according to their morphology type. The third is the automatic classification of the artificial images using CNN that were previously trained with real images. The fourth approach experimentally certifies the ability of the synthetic images in training CNN models specifically developed to recognize the low-prevalence groups. This fourth case is particularly new and relevant in clinical pathology since the generation system is an effective data augmentation approach to train classification algorithms that can help in patient diagnosis.

1.1. Related works

The creation of artificial images is a topic of great relevance in Deep Learning research, seeking to imitate the sense of human creation through algorithms. GANs [18] are generative models that create artificial digital images, learn patterns and characteristics of real sets until obtaining a synthetic result that mimics the original and deceives human perception.

These algorithms make it possible to create sample images with considerable resolution. Its operation is based on two networks that compete with each other. The first network known as Generator G, create images by learning patterns from a real data set. The second network known as Discriminator D serves to differentiate the synthetic image created by the Generator from a real image in the database. The training of both networks is perfected in this competition until the Generator creates a synthetic image that tricks the Discriminator into distinguishing it as real.

The GANs can be used for segmentation, detection, classification, creation, reconstruction, among other applications. GANs have different variants depending on their use, the most used models being the following. DCGAN [19] allows the use of deep convolutions to stabilize training. Conditional GAN [20] architecture allows to introduce labels in the Discriminator to help better classify the input images, and SRGAN [21] are used to improve the quality of images at high resolution. CycleGAN [22] performs image translation by transforming the domain from one image to another completely different. InfoGAN [23] learns specific features for better imaging and WGAN [24] improves the stability of training in using a Critic instead of a Discriminator when creating images.

Interest in the use of GANs in medical imaging has increased considerably due to their versatility. The work in [25] provides an overview of GANs in medicine. This approach is used in the transformation of brain magnetic resonance imaging (MRI) to computed tomography (CT) implementing a CycleGAN architecture, achieving a quality similar to that performed by a CT [26]. Image segmentation with GANs is primarily approached to improve cardiac MRI images by combining CycleGAN with encoders [27] and retinal vascular segmentation with a high quality [28].

Recently, the manipulation of pixels to reconstruct medical images has been considered, being a topic of interest due to the numerous applications focused on diagnosis powered by classification convolutional neural networks. The work in [29] proposes a comparative study between variants of GANs to reconstruct knee magnetic resonance images, managing to enhance the reconstruction time of an MR. Work in [30] provides the use of GANs to create artificial CT images in lesions of cystic hepatic metastases and hemangiomas.

Increasing the database from artificial images is applied in the generation of realistic nodules in CT lungs to develop automatic segmentation algorithms [31]. The methodology of using random noise to generate medical images is used in 2D MR imaging of the brain [32]. The work of [33] provides the creation of 3D images with a high resolution of 240x240 in MR imaging of the chest.

The use of adversarial networks to create microscopic images of red and white blood cells has been rare until now. The work in [34] generates individual red blood cell images (256 x 256 resolution) from a urine dataset using a StyleGAN2 architecture. The evaluation is made by the review of medical experts, without reporting any use to classification models.

The work at [35] proposes an approach to create synthetic field images of red blood smears. The approach starts with the segmented masks of the single cells of the original real images. A GAN generates masked synthetic image samples by paying attention to the shape and distribution of cells. From the masked images generated, a pix2pixHD architecture is used to create the final synthetic field blood cell images. Synthetic imaging is used as a training set augmentation technique for segmentation and object detection in field images. The evaluation experiments conclude that the use of synthetic images in combination with real images minimally improves the segmentation score and is relatively better in the detection task.

The work in [36] proposes a combination of autoencoders with StyleGAN to create four types of leukocytes (eosinophils, lymphocytes, monocytes, and neutrophils) with a resolution of 128 x 128 pixels. This approach requires a large number (12,500) of images. This may be the reason why basophils are not included in the target, since they are the least abundant leukocytes (0.5% of the total).

Extending the application of GANs to other types of cells is a complex task due to the specific morphological characteristics of each cell class. The work of [37] focuses on bone marrow aspirate smears. It uses a Wasserstein GAN with gradient penalty for the creation of various types of cells with a low resolution of 64 x 64 pixels. The size and quality of the image in clinical pathology is of utmost importance to adequately determine the morphological differences of the cells and give a good diagnosis. In this paper we focus on peripheral blood cells and propose a system for generating images of reasonable size in high resolution, evaluated not only through its use in training models but also confirmed by visual review by expert clinical pathologists.

2. System development

2.1. Overview

The basic units in this work are digital images of cells circulating in peripheral blood. They are structured as bitmaps, matrices of size $m \times n \times p$, where m, n are the number of pixels in height and width of the images and p is the color depth. Each color channel is represented by a value in a range from 0 to 255.

Our team has been working for several years on the manual and automated morphological recognition of different types of blood cells in the CORE laboratory of the Hospital Clínic of Barcelona for diagnostic purposes. In the laboratory workflow, $360 \times 363 \times 3$ images are obtained using the Cellavision DM96 [38], which is basically a system with a motorized microscope and a software that selects and supplies images of pre-classified cells. It is widely used in hematology laboratories.

The objective of this work is to develop, train and validate a system, named SyntheticCellGAN, to generate artificial images similar to those obtained by the CellaVision DM96. The general scheme is shown in Fig. 1, where the input is a random noise vector and the output is a high resolution artificial image belonging to one of the following cell classes:

- Lymphocytes
- Monocytes
- Eosinophils
- Neutrophils
- Basophils
- Atypical promyelocytes (AP)
- Hairy cells (HC)

The first five groups are normal leukocytes, which have a morphology that is easily recognized by clinical pathologists. AP and HC are more challenging as they are pathological cells with more complex morphology and much more difficult to find circulating in the blood because AP and HC are hematological diseases in which the number of leukocytes is usually low.

Artificial images are mainly intended to be used in the development of automatic classifiers. CNNs are trained on sets of normalized images. On most CNN architectures, the normalized image size is 224×224 pixels. We propose to create images with a resolution higher than this standard. For the development of the system we use a server with a 12 GB Nvidia Titan XP graphics processing unit (GPU). Creating larger images increases the computational cost. In this paper we have set a target size of 256×256 pixels as a reasonable balance between resolution and cost.

The SyntheticCellGAN (SCG) system uses two GANs to generate an artificial cell image, as illustrated in Fig. 1. First, GAN A transforms a vector of random noise numbers into an low resolution image of a mononuclear cell with basic general features such as a nucleus and cytoplasm, surrounded by several red blood cells as a background. The reason for having this type of target image is the similarity with the images obtained and used in clinical laboratories for hematology examinations. GAN B transforms the basic general images into specific images for each cell type under study, adapting the specific cell morphology and increasing the quality and resolution of the image. Our strategy is to design a dedicated GAN B for each separate class. GAN B_1 for lymphocytes, B_2 for

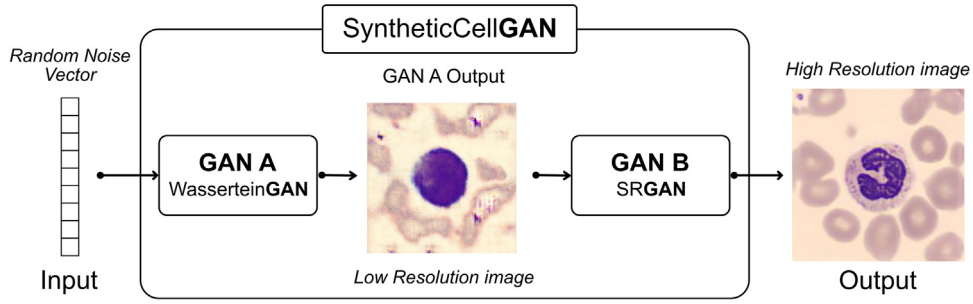

Fig. 1. General structure of the system to automatically create images of cells in peripheral blood using GANs as the main tool for morphological creation from a random noise matrix. *GAN: Generative Adversarial Network.

monocytes, B_3 for eosinophils, B_4 for neutrophils, B_5 for basophils, B_6 for atypical promyelocytes (AP) and B_7 for hairy cells (HC). They are trained independently of each other and, when the system is in operation mode, the user will choose which model is used depending on the type of cell to be generated.

After training the system, the last step is divided into four tests for the validation of the artificial images. The first test consists of a quantitative evaluation of the artificial images using metrics for GANs. The second test involves classifying the artificial images with visual inspection by several expert clinical pathologists. They determine if the generated images have a real cell morphology and can be considered valid. In the third test, the artificial images are analyzed by automatic classifiers that were previously trained with real images from our research group [6,11], verifying the effectiveness of the morphological creation. In the fourth test, a classifier is trained using artificial images and is used to classify images of real-life cells. The test results are presented in Section 4.

2.2. Structure of GAN a

Deep learning algorithms learn from an enriched database. GANs use two groups of images: one is the source group Z and the other is the destination group Y. The generator G learns to imitate the data distribution of Y to transform the first group into the second, resulting in a new group of artificial images X. Having a low number of samples Y and Z can cause overfitting and a poor image. The addition of random noise can be seen as a form of data augmentation [39], and typically in the Z dataset a small amount of noise is added to help blur the data.

As shown in Fig. 2A, the GAN A creates an artificial matrix image x_A with values in the range of 0 to 255 for the three color channels. The size of the image can be modified depending on the architecture of the G. In our case, these images have a size of $128 \times 128 \times 3$ to optimize the available GPU processing power. The input source group Z_A is made up of random noise vectors z_A. For the destination group Y_A, we use real cell images y_A whose details are given in Table 1 in Section 3.1.

Considering X_A and Y_A as two probability distributions, the goal of GAN is to approximate the data of X_A to the data of Y_A by decreasing the distance between theses distributions. The distance between points of two distributions is quantified with measures of divergence. Conventional GANs use the Jensen-Shannon divergence metric [40,41] to quantify the difference between continuous distributions. However, when using random numbers in the origin group Z_A, the result of the generator G_A will be random matrices x_A, and the calculation of the divergence between both distributions will not be continuous, affecting the learning stability of the GAN.

We initially explored models that use random numbers in the source group Z_A as ConditionalGAN (CGAN), VanillaGAN and Deep Convolutional GAN (DCGAN). In CGAN, the source group Z_A must be formed by a combination between a random vector and Y_A, so the generation of images would depend on a real set. VanillaGAN and DCGAN use random numbers in the source group Z_A for their learning, however in our case the result was not as expected, the instability of learning affected the resulting matrix x_A, making a more stable model necessary.

Table 1
Total number of peripheral blood cell images used in this work. The images are grouped by class for each data set. The training set in GAN A is used to create low resolution images and the training set in GAN B improves the image based on each specific morphology.

Image set Y	Cells class	Number of images
A	Mononuclear cells	40,953
B_1	Lymphocytes	4928
B_2	Monocytes	653*
B_3	Eosinophils	218*
B_4	Neutrophils	236*
B_5	Basophils	391*
B_6	Atypical promyelocytes	285*
B_7	Hairy cells	108*

* Cell groups that increased the number of images to 1000 with data augmentation techniques.

For this reason, in this work we use the Earth Mover Distance (EMD) or Wasserstein metric [24]. It represents the minimum average distance necessary to move the data from the distribution X_A to Y_A, calculating the distance of a point x_A with respect to several points of Y_A, thus avoiding discontinuity. Implementing this metric in the GANs, the architecture is known as the Wasserstein GAN or WGAN [24].

2.2.1. Generator GAN a

The generator G_A is a neural network that is responsible for creating the artificial images x_A. The architecture of G_A is shown in Fig. 2B. The input is a random noise vector z_A of size $1 \times 1 \times 100$ with values between 0 and 1, which is a usual standard [42]. The vector is reshaped as a concatenation into an array of size $4 \times 4 \times 100$. This random array goes through transposed convolutions [43], which are used to increase the size of the feature map. Initially, each value of the input array of size 4×4 is multiplied by the kernel of size 4×4 with a stride of 2, obtaining an array of size 8×8 in an upsampling process in the first transposed convolution. In total, five transposed convolutions are performed, which result in the final images x_A of size $128 \times 128 \times 3$.

2.2.2. Critic GAN a

In adversarial networks, the images created by the generator are evaluated by another module with respect to the images in the training set. In the WGAN structure, this module is called critic. The architecture of the Critic C_A is shown in Fig. 2B. The $128 \times 128 \times 3$ artificial image goes through five convolutions to

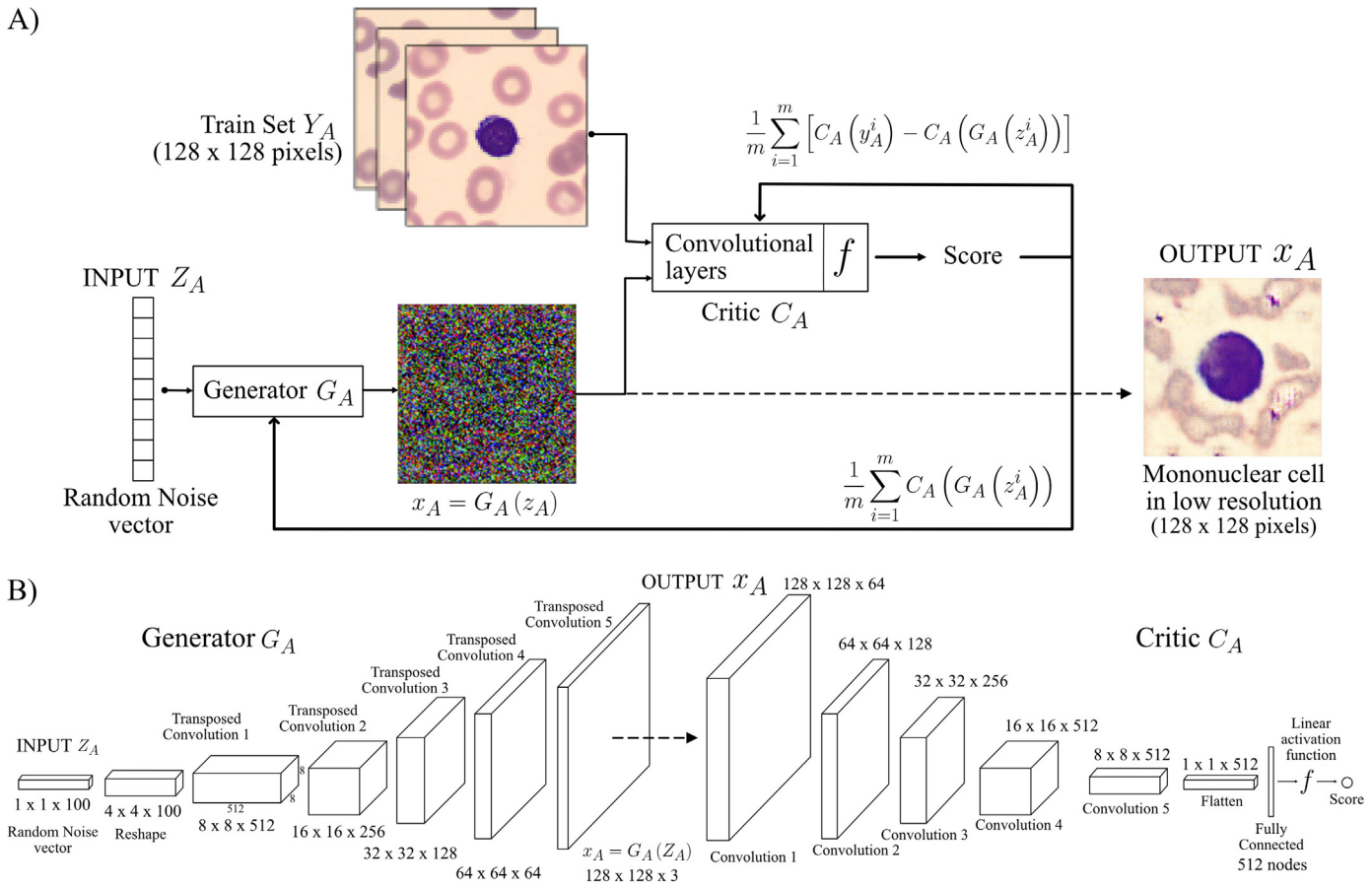

Fig. 2. GAN A architecture. WassersteinGAN is used to generate low-resolution image sets from random noise. A random noise matrix is the input to the generator. From this a synthetic image is created and the Critic rates the image against the real database. As a result, low-resolution images of mononuclear cells are obtained.

learn the relevant features [44]. In this process the image is reduced in size with a kernel of size 4×4 and a stride of 2. Finally, it is flattened and linked to a fully connected layer.

The critic in WGAN has a different conceptual view on the evaluation of artificial images compared to the usual discriminators. These are trained to distinguish between artificial and real images using a probability with a sigmoid activation function at the end. However, learning to distinguish random noise matrices from real images in a discriminator is very fast and unstable [24], so it does not provide enough information to the generator to minimize the distance between the X_A and Y_A distributions and, consequently, create better images. In WGAN, the sigmoid function is changed to a linear activation function to make learning more stable. At the output of the linear function, a value is obtained that qualifies the realism or falsity of the image with a scalar score.

2.2.3. Loss functions in GAN a

As a general principle, the GAN A uses a traditional WGAN loss function. Consider a batch of m initial noise images z_A^i and another batch of m images of the train set y_A^i, $i = 1, \cdots, m$. The generator G_A creates artificial images $x_A = G_A(z_A^i)$, while the critic gives a score $C_A(x_A^i)$ for the artificial images and $C_A(y_A^i)$ for the training images.This score feeds back to the generator and critic, so they can improve their respective learning. The score values should be higher as they are closer to the actual training samples. Therefore, the generator G_A must be trained to maximize the following loss function:

$$L_{G_A} = \frac{1}{m}\sum_{i=1}^{m} C_A\left(G_A\left(z_A^i\right)\right) \tag{1}$$

This measures the effectiveness of the generator G_A in tricking the Critic C_A. On the other hand, the Critic C_A must be trained to maximize the following loss function:

$$L_{C_A} = \frac{1}{m}\sum_{i=1}^{m}\left[C_A\left(y_A^i\right) - C_A\left(G_A\left(z_A^i\right)\right)\right] \tag{2}$$

This means training the critic to distinguish between real and artificial instances.

Both G_A and C_A have a simultaneous learning, as will be described in Section 3.

2.3. Structure of GAN b

The cell images x_A generated by GAN A have basic morphological features and a low resolution with a size of $128 \times 128 \times 3$. The GAN B has to modify the pixels of these images until an image with an acceptable cell morphology and the prescribed size is obtained.

For the design and learning of GAN B we must consider the morphology of each type of cell. If we develop a single GAN B that creates all the cell types proposed in this study, the learning is generalized and its result in our tests is an artificial image with a mixture of features, morphologically incorrect. Therefore, we propose to design a GAN B_i specific for the morphological characteristics of each class $i = 1, \ldots, n$ where n is the number of cell types in our study. This section presents the complete design of a generic GAN B, since the model is the same for each type of cell, varying the set of learning images indicated in Table 1. The structure of GAN B is shown in Fig. 3A.

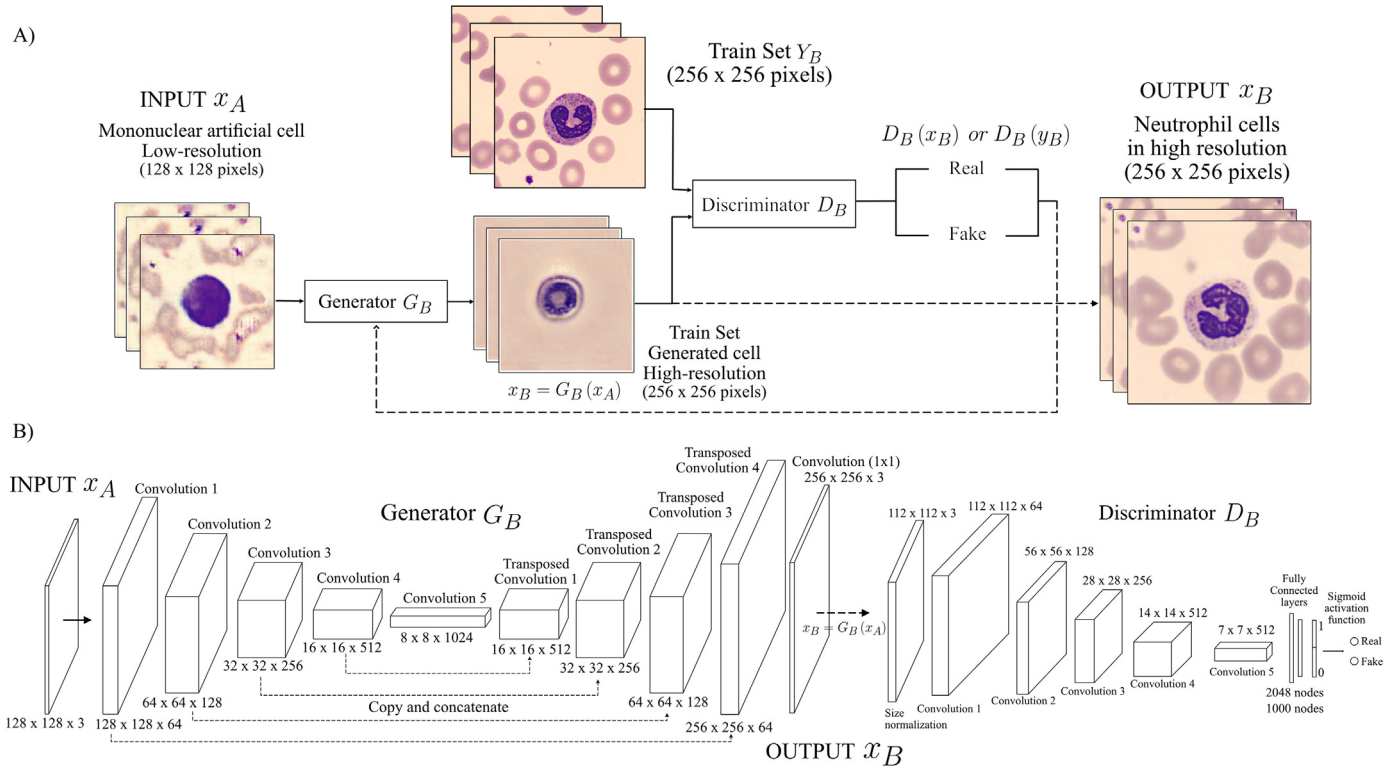

Fig. 3. GAN B architecture. The generator G_B improves the resolution and increases the size of the input image x_A, transforming the pixels according to the training set. The discriminator D_B distinguishes the real images from the synthetic ones until the generator G_B creates a high-resolution images that fool the discriminator. As a result, images of cells are obtained that are morphologically acceptable with a high resolution, typically 256 x 256 pixels. A GAN B is specifically designed for each cell class under consideration. The figure shows the specific case of neutrophil cells.

We adopt the working principle of super resolution images [21], using progressive transmission to improve quality and transposed convolutions to increase size. Simultaneously, we use the working principle of image-to-image translation [45] to transform the cell morphology of GAN A into each cell class in our study. We combine these concepts in the Generator G_B.

The artificial images x_B supplied by the generator G_B and the real images of the training set Y_B are classified by the Discriminator D_B, which is trained to discern images between real or fake.

2.3.1. Generator GAN b

The images x_A must be transformed to images of real cells y_B. To recognize significant regions of interest in a cell (groups of pixels), we use semantic segmentation. Deep learning algorithms based on semantic segmentation typically use sizeable data sets. However, in our case the number of HC and AP samples is limited. An efficient segmentation method for biomedical images was proposed by [46], using a compression and expansion network (U-NET) to perform segmentation with few training samples. This structure is adopted for the generator G_B as shown in Fig. 3B.

In the compression stage, the input image x_A of size $128 \times 128 \times 3$ goes through five descending convolutions until obtaining a feature map of $8 \times 8 \times 1024$. In expansion, this matrix is enlarged using the principle of super resolution with transposed convolutions. In the fourth transposed convolution we set the output size to $256 \times 256 \times 64$.

To perform the image-to-image translation ($x_A - to - Y_A$), it is necessary to extrapolate the pixels. So we connect the compression and expansion layers to each other, as indicated by rows in Fig. 3B. We copy the compression feature map and concatenate it into the corresponding expansion layer. In this process, the resulting image is influenced by the real images of the training set Y_B. Finally, a 1×1 convolution is performed on each pixel, obtaining the image x_B with the three color components.

It is necessary to mention that the fourth transposed convolution can be modified depending on the desired output size. In our case, the generator is limited depending on the capacity of the available GPU, with 12 GB being the standard in most of this work. To explore this aspect, we occasionally used a higher capacity GPU (24 GB) for the HC cell class. This made possible to obtain artificial images this class with a size of $360 \times 360 \times 3$.

2.3.2. Discriminator GAN b

The discriminator D_B is a binary classifier, which distinguishes high-resolution artificial cell images x_B from real cell images y_B. As we mentioned in the previous section, the G_B generator can create images of size $256 \times 256 \times 3$ and $360 \times 360 \times 3$. We normalize the input size of the discriminator so that the classification works with different possible sizes of input images. Specifically, we set the input size to 112×3. As shown in Fig. 3B, these images go through five convolutional layers and two fully connected layers. The last layer has a sigmoid activation function to obtain a binary classification with a probabilistic score in the range of 0 and 1.

2.3.3. Loss functions in GAN b

Contrary to GAN A, GAN B has a discriminator D_B that classifies the images with a probability. The loss function of D_B is a penalty given to the discriminator when it wrongly classifies a real image as artificial and vice versa, which is defined as follows:

$$L_{D_B} = \frac{1}{m}\sum_{i=1}^{m} \log D_B\left(y_B^i\right) + \log\left(1 - D_B\left(x_B^i\right)\right) \tag{3}$$

where, for batches of size m, y_B^i are the real images of the training set, $x_B^i = G_B(x_A^i)$ are the images created by the generator, and D_B is the activation function of the discriminator.

This function calculates the probability of realism of the images. $D_B(y_B^i)$ is the probability that a real image is discriminated as real, while $1 - D_B(x_B^i)$ is the probability that an artificially created image will be discriminated as false. Therefore, the objective of the discriminator training is to maximize the loss function (3).

In the case of the generator G_B, the following loss function is used:

$$L_{G_B} = \frac{1}{m}\sum_{i=1}^{m} log\left(1 - D_B\left(G_B\left(x_A^i\right)\right)\right) \quad (4)$$

where x_A^i denotes an image that enter the generator and $G_B(x_A^i)$ is the image created by G_B. If realistic enough, the probability of $D_B(G_B(x_A^i))$ approaches 1 and the loss function L_{G_B} will tend minus infinity. Therefore, the training goal for the generator is to minimize the loss function (4).

3. System training

3.1. Image datasets

In this work we use images obtained from peripheral blood smear samples in the CORE laboratory of the Hospital Clḍnic de Barcelona. Smears are automatically stained on the system Syxmex SP10i [47] with May Grõnwald - Giemsa [48].

Cell-by-cell images are obtained from the field images using the CellaVision DM96 system, with a resolution of 360×363 pixels. The images are identified, stored and annotated in our database by clinical pathologists. Diagnosis of patients is made according to clinical data, morphology, immunophenotype, cytogenetics and molecular biology.

Table 1 shows the images of the cells grouped by class and the different data sets arranged to train the system. The first set is Y_A, it consists of 40,953 peripheral blood mononuclear cells images for GAN A to learn basic morphological features and create cells with a single round nucleus, cytoplasm and red blood cells in the background of the artificial image. The size of these images is reduced to 128×128 for the GAN A training.

In the case of the GANs B, we group the images for each cell type. GAN B_1 has 4928 lymphocytes, GAN B_2 653 monocytes, GAN B_3 218 eosinophils, GAN B_4 236 neutrophils, and GAN B_5 391 basophils. To create low-prevalence cells, we collect 285 images of atypical promyelocytes for GAN B_6 and 108 images of hairy cells for GAN B_7. We use classical data augmentation techniques such as rotation and flipping to obtain more morphological information on the GAN sets B_{2-7} up to 1000 samples of each class. We reduce the size set of the GAN B_{1-6} to 256×256 for training on the Nvidia Titan XP 12 GB GPU and keep the original size set of the GAN B_7 for the training on 24 GB Nvidia Xeon GPU.

3.2. Training of GAN a

The training is an iterative process that uses the set of real images Y_A and the set of images x_A created by the generator G_A. The period in which all the images pass through the network is known as the training epoch. To improve the efficiency in the learning process the sets are iteratively processed in small batches of 20 images. The generator and critic training are performed simultaneously, that is, the weights G_A and C_A are constantly updated. This procedure has two steps and is performed at each training epoch.

In the first step we freeze the weights of the Critic C_A. Then a batch of random vectors enter the Generator G_A, which are converted into artificial images x_A. The critic C_A evaluates these images with a score. Generator losses are calculated based on its ability to fool the critical using the loss function (1) and the generator weights G_A are updated.

In the second step we unfreeze the weights of the Critic C_A and freeze the weights of G_A. The batch of real images of Y_A is passed through the Critic C_A and the scores of these images are stored. Then the batch of artificial images x_A goes through the Critic C_A and the scores are also saved. With both scores, the critic loss function is calculated using (2). This process leads to updating the weights of C_A.

Using WGAN we can perform a large number of learning epochs. In this case we use 600. At the end of the epochs, it is interesting to visualize images x_A to get an idea of the learning progress. Fig. 4 (a) shows several samples at different epochs. Interestingly, the creation of low-resolution mononuclear cells improves progressively up to epoch 600. This agrees with Fig. 4 (c), where the G_A loss decreases and the C_A loss increases until both converge. The peaks observed in the losses vary oppositely due to the simultaneous training of G_A and C_A. Part (c) in Fig. 4 illustrates that the creation of low-resolution mononuclear cells fails due to overfitting when the number of epochs exceed 600.

3.3. Training of GANs b

All GANs B are trained following the same procedure. In this section, we describe in detail the case of GAN B_6 to synthesize atypical promyelocytes as a representative.

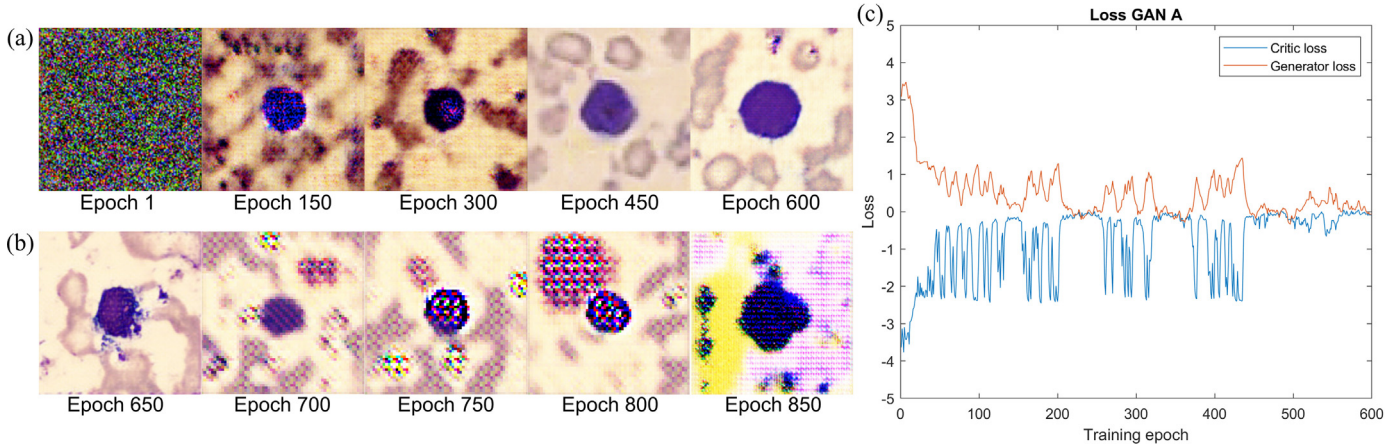

Fig. 4. GAN A training progress. a) Examples of artificial images x_A created at different training epochs with increasing quality. b) Examples of artificial images created with poor quality due to overfitting by exceeding the optimal number of training epochs. c) Evolution of training losses.

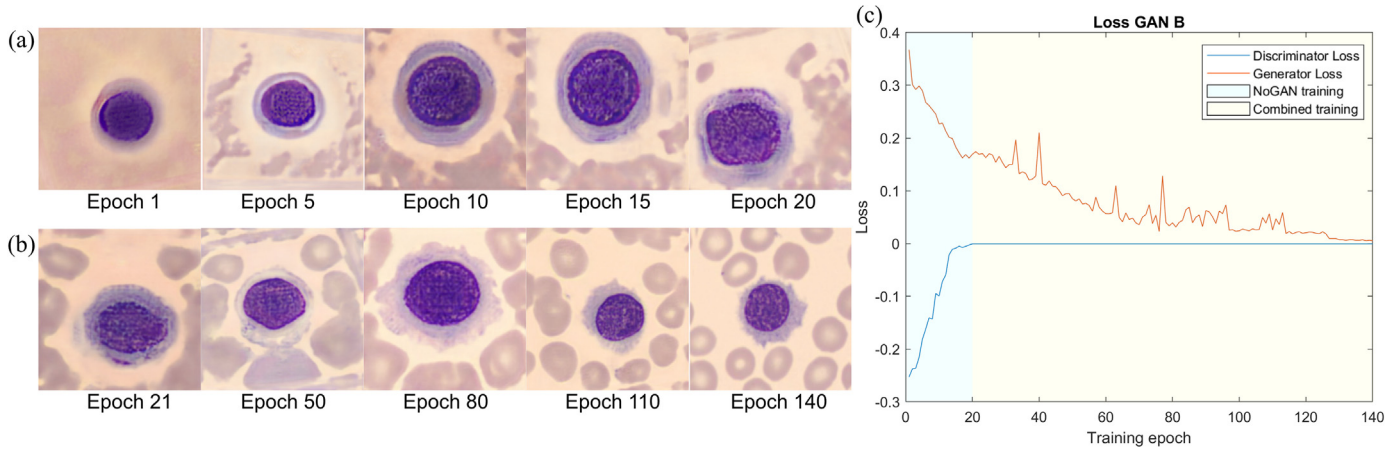

Fig. 5. GAN B training progress. a) Examples of artificial images x_B created in noGAN training. b) Example of artificial images x_B created in combined training. c) Evolution of training losses.

Simultaneous generator and critical training used in GAN A limits its GPU memory. This procedure is not feasible for GAN B, since the algorithm needs a higher capacity to modify and change the resolution of the image. Alternatively, we adopt the NoGAN method [49], which was developed by the DeOldify project. This method trains the discriminator and the generator separately. Its main advantages are a shorter learning time, greater generation control and loss stability.

We initially used GAN A to create a set of 1000 images of artificial mononuclear cells x_A in low resolution (128×128 pixels). We also have the training set Y_B with 1000 real images in high resolution (256×256 pixels). In the initial individual training of the generator G_B with a U-Net architecture, we use default weights and batches of eight images x_A are passed. We perform 20 individual training epochs for the generator G_B, as seen in Fig. 5 (c). A set of 1000 artificial images x_B was obtained. On the other hand, visual inspection of the generated images x_B in Fig. 5 (a) shows that the cell morphology does not yet resemble atypical promyelocytes.

For the individual training of the D_B discriminator, the set of real images Y_B and the set of 1000 artificial images are used. The binary classifier has an input normalization of 112×112 to minimize processing on the GPU. The D_B is trained for 20 epochs until the discriminator losses approach 0.

After the separate training, we carry out a combined training of 120 epochs as seen in Fig. 5 (c). In this case, this training serves to adjust the weights of the generator based on the prediction of the discriminator. We freeze the discriminator weights, since the generator loss already has its maximum value (0). Every epoch the generator G_B creates artificial images x_B, these images enter the discriminator D_B and it predicts whether the image is real or fake. In each workout the weights of the generator are updated.

After the separate training, we carry out a combined training to continue adjust the generator weights based on the prediction of the discriminator. As illustrated in Fig. 5 (c), 120 training epochs are performed keeping the discriminator weights frozen, since the generator loss already has its maximum value (0). Every epoch the generator G_B creates artificial images x_B, these images enter the discriminator D_B and it predicts if the image is real or fake. In each training the weights of the generator are updated. Visual inspection of the generated images x_B in Fig. 5 (b) shows that the cell morphology progressively tends to resemble atypical promyelocytes, which is the ultimate goal.

The training is repeated for each of the seven GAN B changing the real image sets Y_B as described in Table 1 and the number of epochs for each type of cell. Fig. 6 shows examples of the artificial images in high resolution for each case.

4. Experimental assessment

The purpose of this section is to evaluate the effectiveness of the trained SCG system to create morphologically correct artificial images. Four types of tests are presented in the following subsections.

4.1. Quantitative evaluations

We consider several quantitative measures to evaluate the quality of the images generated by the system. The most commonly used metrics to assess GANs are Frechet Inception Distance (FID) [50], Inception Score (IS) [51], and Learned Perceptual Image Patch Similarity (LPIPS) [52].

FID is used to calculate the distance between feature vectors extracted from two sets of generated and real images. In this evaluation we calculate the FID between sets of 250 artificial images of each GAN B_i and 250 real images of each class. We use a pre-trained InceptionV3 network with its original weights for feature extraction and resize all images to 299 x 299 (inception's original training data size). The FID calculation is done using the 2048 features obtained from the last pooling layer of InceptionV3 to represent the image. The lower the distance, the better the image quality. IS metric is used to rate the quality and variability of the generated images with a score, which uses the classification of the generated and the real images using a pre-trained InceptionV3 model. The higher, the better. The LPIPS metric is used to evaluate the perceptual diversity in a set of images in a way close to the human perception. In this evaluation this metric is calculated for the sets of 250 images generated from each GAN B_i. The higher, the better. Table 2 presents the results of each metric. A quick glance shows that FID has the best (lowest) values for neutrophils (GAN B_4) and basophils (GAN B_5), while the worst values correspond to lymphocytes (GANs B_1). The IS score is similar for all the GANs B_i, so that each class of the artificial images created by SCG have a similar quality and variability. The LPIPS score is similar in all classes, so there is a degree of diversity within each class created by SCG.

4.2. Visual inspection of synthetic cells

Five clinical pathologists participate in this assessment. They are experienced in visual assessment of peripheral blood smears using digital cell images acquired from the CellaVision DM96 system. They have contributed to the annotation of data bases of images for training and assessment of automatic classification systems [53].

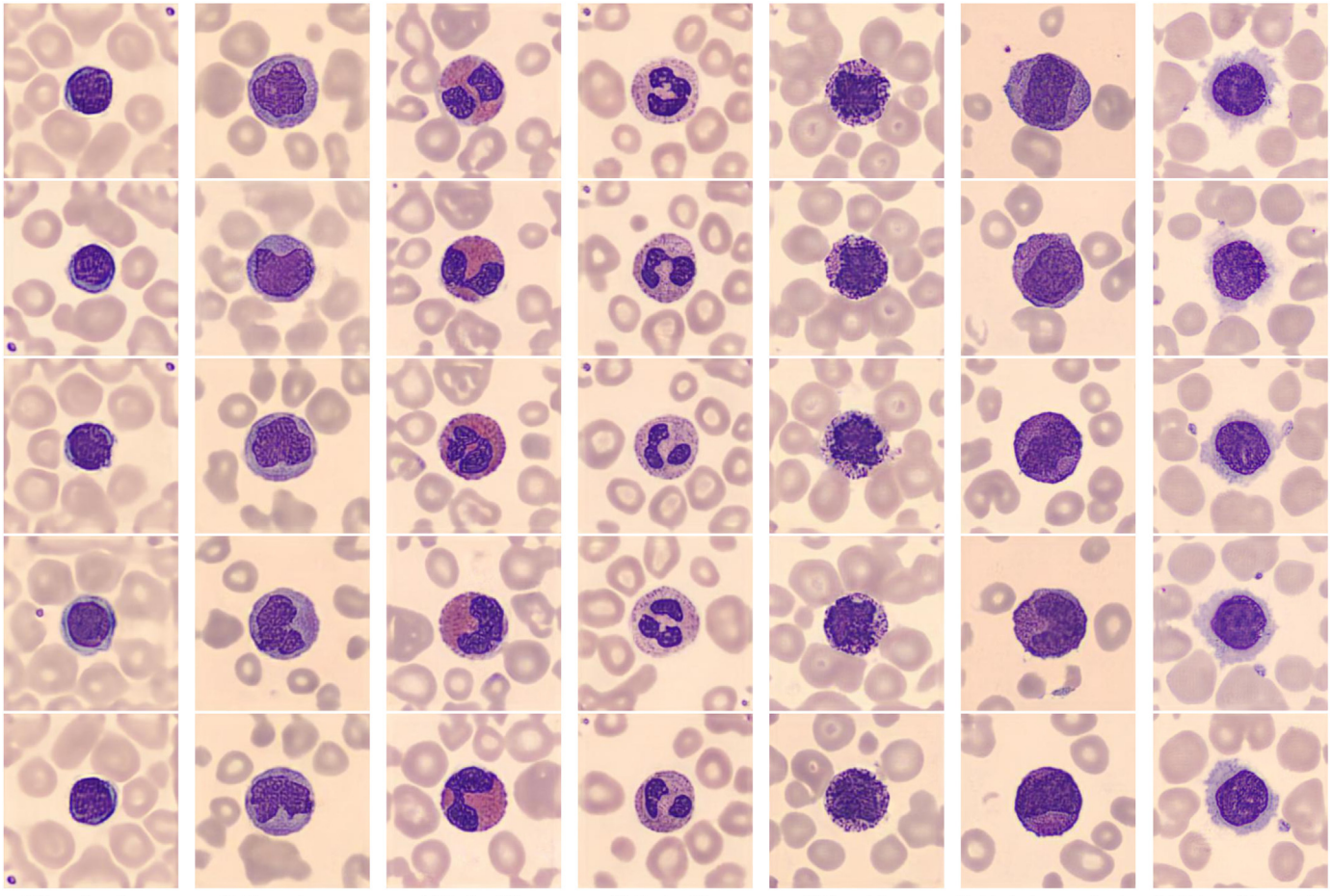

Fig. 6. Examples of synthetic images generated with the developed system. (a) Lymphocytes. (b) Monocytes. (c) Eosinophils. (d) Neutrophils. (e) Basophils. (f) Atypical promyelocytes. (g) Hairy cells.

Table 2
Quantitative evaluation metrics for SCG synthetic images. GAN B_1 generates lymphocytes, GAN B_2 monocytes, GAN B_3 eosinophils, GAN B_4 neutrophils, GAN B_5 basophils, GAN B_6 atypical promyelocytes and GAN B_7 hairy cells.

Metrics	SyntheticCellGAN models						
	GAN B_1	GAN B_2	GAN B_3	GAN B_4	GAN B_5	GAN B_6	GAN B_7
FID	75.766	47.341	36.249	13.921	12.918	29.463	31.279
IS	9.156 Ø 0.127	9.234 Ø 0.071	10.111 Ø 0.063	10.206 Ø 0.162	10.636 Ø 0.073	10.175 Ø 0.104	10.148 Ø 0.084
LPIPS	0.492	0.508	0.515	0.534	0.537	0.527	0.519

FID: the lower, the better. IS: the higher, the better. LPIPS: the higher, the better.

Clinical pathologists must blindly classify four pools of artificial images that are selected by increasing the complexity of cell morphology. In the first case, a set of 500 artificial leukocyte images are created, 100 of each of the following types: neutrophils, lymphocytes, eosinophils, basophils and monocytes. Table 3A shows the average classification result in a confusion matrix. The global accuracy is 99.92%.

In the second experiment, two new sets of cells are created that include smear cases with abnormal cells (AP and HC). The first set consists of 150 artificial cell images, 25 of each type of leukocyte along with 25 of AP. The second set is composed of 175 artificial cell images, 25 of each type of normal leukocyte, 25 AP and 25 HC. As in the previous experiment, the images are mixed and classified by the clinical pathologists. Table 3B and 3C show the confusion matrices for both cases with the experts' average classification results. Remarkable is the 100% accuracy, particularly in artificial AP and HC.

In the third visual inspection experiment, a new set of mononuclear cells is synthesized. These images have in common the characteristic of a single round nucleus. The set consists of 100 images, including 33 lymphocytes, 33 monocytes and 34 HC. The images are mixed and classified by the experts. The challenge is the differentiation between normal cells (lymphocytes and monocytes) and abnormal lymphoid cells (HC). The confusion matrix in Table 3D with the average classification of the clinical pathologists shows an accuracy of 100%.

4.3. Automatic classification of synthetic cells

In this section we use two CNN models, which were previously trained on images of real blood cells obtained from the CellaVision DM96, to classify sets of artificial images as another way to test their morphological realism. The first model was designed to classify among eight cell groups (neutrophils,

Table 3
Confusion matrix of classification results (in %) by expert clinical pathologists. A) Classification of a set of 500 artificial leukocytes (100 from each group). B) Classification of a set of 150 artificial cells including atypical promyelocytes (25 from each group). C) Classification of a set of 175 artificial cells including atypical promyelocytes and hairy cells (25 from each group). D) Classification of a set of 100 artificial mononuclear cells (33 lymphocytes, 33 monocytes and 34 hairy cells). The overall classification accuracy of each experiment is 99.92%, 100%, 100% and 100%, respectively.

A.

		Classification				
		Neutrophils	Lymphocytes	Eosinophils	Basophils	Monocytes
Ground Truth	Neutrophils	**100**	0	0	0	0
	Lymphocytes	0	**99.8**	0	0.2	0
	Eosinophils	0	0	**100**	0	0
	Basophils	0	0.2	0	**99.8**	0
	Monocytes	0	0	0	0	**100**

B.

		Classification					
		Neutrophils	Lymphocytes	Eosinophils	Basophils	Monocytes	Atypical promyelocytes
Ground Truth	Neutrophils	**100**	0	0	0	0	0
	Lymphocytes	0	**100**	0	0	0	0
	Eosinophils	0	0	**100**	0	0	0
	Basophils	0	0	0	**100**	0	0
	Monocytes	0	0	0	0	**100**	0
	Atypical promyelocytes	0	0	0	0	0	**100**

C.

		Classification						
		Neutrophils	Lymphocytes	Eosinophils	Basophils	Monocytes	Atypical promyelocytes	Hairy cells
Ground Truth	Neutrophils	**100**	0	0	0	0	0	0
	Lymphocytes	0	**100**	0	0	0	0	0
	Eosinophils	0	0	**100**	0	0	0	0
	Basophils	0	0	0	**100**	0	0	0
	Monocytes	0	0	0	0	**100**	0	0
	Atypical promyelocytes	0	0	0	0	0	**100**	0
	Hairy cells	0	0	0	0	0	0	**100**

D.

		Classification		
		Lymphocytes	Monocytes	Hairy cells
Ground Truth	Lymphocytes	**100**	0	0
	Monocytes	0	**100**	0
	Hairy cells	0	0	**100**

Table 4
Confusion matrix results (in %) when the synthetic images (2500, 500 per group) are classified by a CNN model that was trained on real blood cell images. The global classification accuracy is 95.36%.

		Predicted Class						
		Neutrophils	Monocytes	Eosinophils	Lymphocytes	Basophils	Erythroblasts	Immature granulocytes
Input synthetic cells	Neutrophils	**87.6**	1	5.4	0	0.6	0	5.4
	Monocytes	0	**100**	0	0	0	0	0
	Eosinophils	0	0	**98.8**	0	0.6	0	0.6
	Lymphocytes	0	0	0	**93.8**	0	6.2	0
	Basophils	0	0.8	0	2.4	**96.6**	0	0.2

eosinophils, basophils, lymphocytes, monocytes, immature granulocytes, erythroblasts, and platelets) [6]. The evaluation experiments exhibited a general accuracy of 96%. The second model (ALNet) was developed to help predict weather a patient is healthy, has acute leukemia, or has a viral infection [11]. ALNet classifies five groups of cells (monocytes, lymphocytes, atypical promyelocytes, blasts and reactive lymphocytes).

We create a first set of 2500 artificial images of leukocytes with the SCG system: neutrophils, monocytes, eosinophils, lymphocytes and basophils (500 from each class). We pass this set through the classifier [6]. Table 4 displays the classification results as a confusion matrix. The true positive rates for artificial neutrophils, monocytes, eosinophils, lymphocytes and basophils are 87.6%, 100%, 98.8%, 93.8% and 96.6% respectively.

We generate a second set of 1500 artificial images of cells with the SCG system: monocytes, lymphocytes and atypical promyelocytes (500 from each group), and pass them through the ALNet classifier [11]. The classification results are displayed in the confusion matrix in Table 5. The true positive rates for monocytes, lymphocytes and atypical promyelocytes are 100%, 91% and 91% respectively. The overall accuracy is 94%.

As a summary of this section, the models trained with real images classify synthetic images with performances that are not very far from those obtained when classifying real images.

Table 5
Confusion matrix results (in %) when the synthetic images (1500, 500 per group) are classified by a CNN model that was trained on real blood cell images. The overall classification accuracy is 94%. AP: Atypical promyelocytes.

		Predicted Class				
		Monocytes	Lymphocytes	AP	Blasts	Reactive lymphocyte
Input synthetic cells	Monocytes	**100**	0	0	0	0
	Lymphocytes	9	**91**	0	0	0
	AP	9	0	**91**	0	0

Table 6
Confusion matrix of the classifier (in %) to differentiate 6728 leukocytes (274 monocytes, 674 neutrophils, 249 basophils, 603 eosinophils and 4928 lymphocytes). The classification accuracies of each class are 97.8%, 98.5%, 99.2%, 99.5% and 99.4%, respectively. The global accuracy is 98.89%.

		Predicted Class				
		Monocytes	Neutrophils	Basophils	Eosinophils	Lymphocyte
Input real cells	Monocytes	**97.8**	0.4	1.5	0	0.4
	Neutrophils	0.9	**98.5**	0.6	0	0
	Basophils	0.4	0	**99.2**	0	0.4
	Eosinophils	0	0.3	0.2	**99.5**	0
	Lymphocyte	0.4	0.2	0	0	**99.4**

Table 7
Results of binary classifiers to differentiate lymphocytes and hairy cells. In cases A.1 - A.5 the classifier is trained using different ratios of real images of 3955 lymphocytes and 417 hairy cells. In case B the classifier is trained with a set of 7000 artificial images of each class only. In the cases C.1 - C.5, the classifier is trained with a hybrid database of 7000 artificial images of each class plus the different ratios of real images used in the cases A.1 - A.5.

Classifier	Type of data set	Number of images to train		Accuracy classifying real images of test set (in %)		
				Per class		Global Accuracy
		Lymphocytes	Hairy cells	Lymphocytes	Hairy cells	
A.1	100% real images	3955	417	99.9	80.2	**90**
A.2	75% real images	2966	313	99.9	79.3	**89.6**
A.3	50% real images	1978	209	99.7	76.9	**88.3**
A.4	25% real images	989	104	99.1	74.4	**86.7**
A.5	5% real images	198	21	96.4	64.5	**80.4**
B	artificial	7000*	7000*	81.2	94.2	**87.7**
C.1	artificial + 100% real images	10,955	7417	99.9	89.3	**94.5**
C.2	artificial + 75% real images	9966	7313	98.5	87.6	**93**
C.3	artificial + 50% real images	8978	7209	99.4	85.1	**92.3**
C.4	artificial + 25% real images	7989	7104	99.9	84.3	**92.1**
C.5	artificial + 5% real images	7198	7021	97.5	71.9	**84.7**

* Artificial images set created by the SCG system.

4.4. Automatic classification of real cells

The purpose of this section is to investigate if the artificial images generated by the SCG system can be used as a data augmentation tool for training classifiers.

Three classification scenarios are considered. In the first scenario, a CNN classifier is trained and tested to discriminate between five classes of leukocytes (monocytes, neutrophils, basophils, eosinophils and lymphocytes). In the second scenario, a CNN classifier is trained and tested to discriminate between lymphocytes and hairy cells. In the third scenario, a CNN model is trained and tested to distinguish between neutrophils and atypical promyelocytes. The classifiers are designed and trained as described in [54] with all images normalized to 224 x 224 pixels. No data augmentation techniques are used. The SeNet154 architecture is used and the models are trained with 10 epochs. In scenario 1 we use only artificial images for training. In the following two scenarios, the training is done with different combinations of real and artificial images, while the tests are done with the same set of real images to compare the classification accuracy of each trained model.

4.4.1. Scenario 1

The SCG creates 7000 artificial images of each type of leukocyte, which are used to train the classification network. After the training, we use the network to classify a set of real leukocytes with 4928 lymphocytes, 674 neutrophils, 603 eosinophils, 274 monocytes and 249 basophils images. The classification results of this scenario are displayed in the confusion matrix in Table 6. The classification accuracies of each class are 97.8%, 98.5%, 99.2%, 99.5% and 99.4%. The global accuracy is 98.89%.

4.4.2. Scenario 2

All the relevant information on the tests in this scenario are shown in Table 7. We initially have 4928 real images of lymphocytes and 538 hairy cells, which are divided randomly: 80% for training and 20% for testing. This gives 3955 lymphocytes and 417 hairy cells for training and 973 lymphocytes and 121 hairy cells to be classified at the end of each training as a test set.

We consider five cases (A.1-A.5) in which the classifier is trained only with real images, but using smaller percentages of the total available cells from 100% to only 5%. After training, the images of the full test set are classified and the individual accuracies and the global (average) accuracy are given. For additional tests, we use the SCG system to create an artificial set of 7000 lymphocytes and 7000 hairy cells. Case B in Table 7 represents the classifier trained with only the full set of artificial images. Cases C.1 - C.5 correspond to the training of the classifier with a hybrid set

Table 8
Results of binary classifiers to differentiate neutrophils and atypical promyelocytes. In the cases D.1 - D.5, we train the classifier with different ratios of real images of 3680 neutrophils and 2939 atypical promyelocytes. In case E we train with a set of 7000 artificial images of each class only. In the cases F.1 - F.5, we train with a hybrid database of 7000 artificial images of each class plus the different ratios of real images used in the cases D.1 - D.5.

Classifier	Type of data set	Number of images to train		Accuracy classifying real images of test set (in %)		
				Per class		Global Accuracy
		Neutrophils	Atypical promyelocytes	Neutrophils	Atypical promyelocytes	
D.1	100% real images	3680	2939	100	100	**100**
D.2	75% real images	2760	2204	99.8	100	**99.89**
D.3	50% real images	1840	1470	100	100	**99.86**
D.4	25% real images	920	735	99.8	99.7	**99.82**
D.5	5% real images	184	147	87.5	99.9	**93.70**
E	artificial	7000*	7000*	97.7	100	**98.85**
F.1	artificial + 100% real images	10,680	9939	100	100	**100**
F.2	artificial + 75% real images	9760	9204	99.8	100	**99.89**
F.3	artificial + 50% real images	8840	8470	99.9	100	**99.94**
F.4	artificial + 25% real images	7920	7735	100	100	**100**
F.5	artificial + 5% real images	7184	7147	99.8	100	**99.89**

* Artificial images set created by the SCG system.

that combines all the created artificial images with different proportions of the available real images.

4.4.3. Scenario 3

In this case we use a set of real images from a large number of blood samples: 4595 real images of neutrophils and 3679 of atypical promyelocytes. After a random split of 80% - 20%, 3680 images of neutrophils and 2939 atypical promyelocytes are available for training, and 915 neutrophils and 740 atypical promyelocytes are used as a test set. The relevant information on the tests are shown in Table 8.

As in Scenario 3, we consider five cases (D.1-D.5) where the classifier is trained only on real images using smaller proportions of the total cells from 100% to 5%. After training, the entire test set is classified. Table 8 shows accuracies ranging from 100% to 93.7%. For the further tests, the developed SCG system is used to create a set of artificial images of 7000 neutrophils and 7000 atypical promyelocytes. Case E in Table 8 corresponds to the classifier trained using only this full set. Cases F.1 - F.5 in Table 8 correspond to training the classifier with combinations of all the artificial images and different percentages of the available real images.

5. Discussion

This work is framed in the context of automatic classification of cells circulating in peripheral blood as a tool to assist clinical pathologists in rapid initial diagnosis of hematologic diseases. The substrate material is digital cell images obtained under the microscope from blood smears, while the essential methodologies are deep learning models (specifically CNN) for image classification. In general, these models rely on well-labelled large image sets to learn to extract morphological features. Collecting blood cell images needs standardized procedures and equipment followed by cell annotation by expert clinical pathologists. Quality and resolution are essential for learning morphological features allowing the experts and the computerized classification models to discern among a wide spectrum of cell types observed in blood.

Our work has been inspired by the success of generative adversarial networks (GAN) to create artificial images in different medical domains, mostly computed tomography and magnetic resonance imaging. The tangible result of the work has been a new two-stage system, SyntheticCellGAN (SCG), which creates images of seven types of blood cells. Five of them are the family of normal leukocytes: lymphocytes, monocytes, eosinophils, neutrophils, and basophils. The other abnormal two groups are atypical promyelocytes (AP) and hairy cells (HC).

Very few works have so far been reported in the literature on the generation of microscopic cell images using adversarial networks. As far as the authors know, the closest to our work are [36] and [37]. Four classes of blood leukocytes are synthesized in [36], with no basophils. They use a combination of automatic encoders with StyleGAN to generate images with a resolution of 128 x 128 pixels. A large set of 12,500 images is used, so basophils may not be included as they are the rarest leukocytes (0.5% of total).

The work in [37] synthesizes cell images obtained from bone marrow aspirate smears. The target is the five normal leukocyte classes plus erythroblasts, immature granulocytes, monoblasts, myeloblasts, normal promyelocytes and platelets. The work uses a Wasserstein GAN with gradient penalty to obtain cell images with a resolution of 64 x 64 pixels. In our opinion, this resolution is low. Extending the application of GANs to a variety of cell types is a complex task due to the specific and subtle morphological characteristics of each class. The quality and size of the image in clinical pathology is of utmost relevance to determine the morphological cell differences and provide an accurate diagnosis.

Our objective has been the synthesis of nucleated cells from peripheral blood and the proposed system has been designed to generate images with a high resolution of 256 x 256 pixels, which are close to the type of images provided by analyzers currently used in clinical laboratories. We have emphasized this target to be closer to clinical practice. So, besides the normal leukocytes, we have synthesized images of atypical promyelocytes since these cells are indicators of acute promyelocytic leukemia, whose patients can have severe bleeding and die if timely treatment is not established. In addition, we have synthesized hairy cells, which are indicative of hairy cell leukemia. We have found it worthwhile to focus on these two pathological cell types with complex morphological features and representative of two serious hematological diseases of low prevalence and be more ambitious to achieve a higher resolution.

The rest of this section deals with the following aspects: 1) analysis of the experimental results; 2) the main conceptual and practical points about the development and training of the model; and 3) computational resources for the proposed method.

5.1. Analysis of the results

The first test of the system was the evaluation of the quality and diversity of the artificial cells. The results of the quantitative evaluation in Table 2 are comparable to the quantitative tests in the papers [36] and [37].

The FID between the real images and artificial images generated by SCG is in the range of 75.766 to 12.918, the lower the better. The

value of this metric in [37] is between 76.3 and 67.2 when comparing various approaches for generation of microscopic cell images. It is particularly interesting to observe (see Table 2) the difference between the best FID scores (13.921 and 12.918 for the synthetic images of neutrophils and basophils, respectively) versus the worst highest score (75.766 and 47.34 for synthetic images of lymphocytes and monocytes, respectively). This may be interpretable in terms of some morphological features of these cell types. Lymphocytes encompasses a class of cells that exhibit significant intraclass morphological variability. Some lymphocytes may present granulation (T-lymphocytes) and others not (B-lymphocytes). Some have a greatly reduced cytoplasm, but some show it to be slightly larger. There is also variability in color, showing a more or less basophilic cytoplasm. This variability is naturally present when choosing a set of real images and is reflected when creating a set of artificial images. Relative to monocytes, these cells also exhibit some variability in size and nuclear outline. In fact, lymphocytes and monocytes are known to have a quite heterogeneous morphology between cells of the same class. On the other side, real neutrophils and basophils show much more intraclass homogeneity, which is also reflected in the sets of artificially synthesized images. Consequently, the FID distance between real and synthetic image sets of lymphocytes or monocytes is understandably greater than the distance between two sets with real and synthetic images of neutrophils or basophils.

The IS values of the SCG are between 9.156 and 10.636, the higher the better. The result of the IS metric confirms that each GAN B_i has a similar generation quality. In previous literature, the IS value ranges from 8.34 to 12.36. The LPIPS range between 0.492 and 0.537, the higher the better. The diversity is comparable in each generated class. In previous literature, LPIPS values are between 0.25 and 0.34.

Since this work was motivated by possible clinical interest, the second synthetic imaging examination was a direct visual assessment by expert clinical pathologists. The four experiments summarized in Table 3 were performed mixing different types of synthetic images in different combinations, and practically all of them were classified correctly. Experts commented positively on the high quality and resolution of the images. In addition, they highlighted several qualitative morphological features that are typically used to discern between different types of cells in the analysis of real images. As observed in the synthetic examples of Fig. 6, the regular profile of the nucleus with condensed chromatin and the scarce cytoplasm stands out for lymphocytes. For monocytes, the nucleus with a kidney-shaped profile and loose chromatin. For eosinophils, round and orange granulation. For neutrophils, segmentation of the nucleus. For basophils, round and basophilic granulation (almost black). For the atypical promyelocytes (AP), large size, irregular profile of the nucleus and high content of immature granulation in the cytoplasm. For hairy cells, irregular profile associated with villi in the cytoplasmic membrane. In summary, it was remarkable that clinical pathologists were able to recognize the cells without question by their morphological features realistically similar to those observed in real cells.

We found it interesting to carry out a test similar to the one carried out by humans, but this time with two available CNN models that were previously trained with real images. The first model [6] was trained to recognize neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes, erythroblasts, and platelets. In our case, only the first five types were synthesized. The five first columns of Table 4 show the classification percentages of these cell types. The overall accuracy (average of the diagonal true positive rates) is 95.36%, which is comparable to the overall accuracy of 96% reported in [6]. We observed a 6.2% of lymphocytes that were recognized by the model as erythroblasts. This is understandable since both have morphological similarities with round nucleus and condensed chromatin. We also observed 5.4% of neutrophils classified as immature granulocytes. Both are of the same cell lineage, with the main difference being that neutrophils are mature.

The second model was trained to classify five cell groups: monocytes, lymphocytes, atypical promyelocytes (AP), blasts and reactive lymphocytes [11]. In our experiment, only monocytes, lymphocytes and AP were synthesized. The classification results in Table 5 show that non of the synthetic images were identified as blasts or reactive lymphocytes. The results of the three synthetic classes were comparable to those reported in [11] with an overall accuracy of 94%.

Undoubtedly, a practical motivation for the generation of artificial images is the increase of data sets for model training, particularly when data is sparse or unbalanced. In this work we addressed two problems, related to hairy cells (HC) and atypical promyelocytes (AP), respectively.

For the first problem, a binary CNN classifier was designed to distinguish between HC and normal lymphocytes. It is a relevant problem since HC is a type of abnormal lymphocyte from which it is difficult to obtain peripheral blood images. In a first step, the model was trained only with real images, performing different experiments (A.1-A.5) using decreasing percentages of the total available images as shown in Table 7. In all cases, there is a clear HC imbalance with only around 10% of the total images. The tests with real images for all the experiments showed a progressive decrease of the accuracy with the reduction of the images used for training, particularly for HC. While the accuracy for lymphocytes ranged from 99.9% to 96.4%, it dropped from 80.2% to 64.5% for HC, when only 21 images were used for training.

The second step was to train the classifier using the synthetic images created by the SCG system in different proportions together with real images as described in Table 7. The first noteworthy comment is that using 7000 artificial images of each group, the accuracy of the HC classification (case B) drastically improves to 94.2%, while the accuracy for lymphocytes drops to 81.2%. This could be attributable to the fact that the artificial images of lymphocytes generated by the system are quite homogeneous, as shown in Fig. 6. This can favor that the training is not so rich and, consequently, the test with the real images gives a percentage of misclassifications. In cases C.1-C.5, using all the synthetic images but also including real images, the accuracy in lymphocyte classification recovers the high values of cases A.1-A.5. At the same time, the HC classification in C.1-C.5 maintains accuracy values always higher than those corresponding to cases A.1-A.5. In this sense, it is especially interesting to compare experiments A.5 and C.5 in which only 198 real lymphocytes and HCs are used. By adding the set of synthetic images, the accuracy increases from 96.4% to 97.5% for lymphocytes and from 64.5% to 71.9% for HC.

In the second problem, a CNN model was designed to distinguish between neutrophils and atypical promyelocytes (AP). Both are myeloid cells but neutrophils are normal mature granulocytes, while AP are immature abnormal cells circulating in blood in acute promyelocytic leukemia, which is an hematological disease whose annual incidence rate is 0.08 case per 100,000 population.

The model was first trained only with real images and different experiments (D.1-D.5) were performed using decreasing ratios of the total available images, see Table 8. As a difference from the previous problem, the number of training images is not very unbalanced (3680 neutrophils and 2939 AP). The classification accuracy of the test set is almost 100% for the two cell classes, except in case D.5 when only 5% of the real cells are used for training. In this case, the accuracy for neutrophils remains high at 87.5%. It is notable that for cases E to F.4, adding artificial images to the training data set does not change the high accuracies obtained previously with only real cells. And for the F.5 case with 184 neu-

trophils and 147 AP (only 5% of the total images), adding the set of synthetic images increases the accuracy of the neutrophils from 87.5% to 99.8%. This result is consistent with the widespread idea that the use of GAN-based data augmentation systems pays off for small enough data sets.

5.2. System development and training

The SCG system has been developed with a structure of two sequential GANs. Random vectors are the inputs Z, the real training set Y is formed by the cells of our database and the outputs of the system are the artificial images X.

Initially, we considered using a single GAN. The Z inputs were real cells from our database, which were modified by the GAN according to the characteristics of the training set. However, the number of images created was limited by the number of input images since it was a one-to-one correspondence. To mitigate this limitation, we changed the Z inputs to random vectors and restructured the GAN architecture. Since the distribution of random vectors is not continuous, we used the Wasserstein metric of the WGAN architecture for an efficient approximation to the images in our database. The output size of the artificial images obtained with this architecture was not satisfactory for our final objective. Moreover, if we wanted to create new cell classes that were not included in the training set, we would have to retrain the GAN.

Considering these factors, we decided to use two GANs. GAN A creates images of basic mononuclear cells in low resolution from random vectors, and GAN B modifies the morphology of these images, improving their quality. At the output, high resolution synthetic images are obtained. The first results showed that in this process the GAN B learned all the features of the training set without differentiating each class of cells. The high resolution artificial images obtained had a mixture of features that did not correspond to a particular real morphology. For this reason, we decided to design a specific GAN B_i for each cell class to be generated. These models were trained independently of each other.

Using this insight, we obtained final images that matched the desired specific training morphology. The database for GAN A training was composed of mononuclear cells obtained in the laboratory using CellaVision DM96 and reduced to a size of 128x128. The training set for each specific GAN B required high-quality images. In this sense, it was important to inspect irregularities and purge some images with the help of clinical pathologists.

The Wasserstein metric used in GAN A ensures stability in training. In our work, the generator and the critic were trained simultaneously for 600 epoch to reach the convergence of both loss functions. It is interesting to recall the information given by Fig. 4, which illustrates the progressive improving of the synthetic images until this convergence is reached.

In the case of the GAN B_i, we proposed an adverse learning technique using the NoGAN method, which implies a separate training for the generator and the discriminator. The generator approximates the data distribution of the basic mononuclear images to the specific morphology of each cell class. It is worth mentioning that in this process the computational resources such as memory, training and processing times are used efficiently because this resources are used only in the generators. We performed 20 epochs of training until the discriminator loss reached its target value (0) as seen in Fig. 5.

Since the generator loss was not yet stabilized, we performed additional training to adjust the generator weights based on the discriminator prediction with their optimal weights frozen. Visual inspection of the images generated at each epoch (see Fig. 5) showed the progressive evolution towards images resembling the real class together with the convergence of the generator losses. The entire training strategy adopted for each GAN B was crucial for the generation of high-resolution images for each target class.

5.3. Computational resources

It is relevant to discuss the computational resources and the time required to train the models. The resources are essentially reduced to the memory of the graphics processing unit (GPU) dedicated to perform the calculations, which relies on a suitable cooling device to dissipate the heat and a constant track of the training process.In this work a server unit with a 12 GB Nvidia Titan XP was used.

We implemented a batch training, where GAN takes small batches of images to learn their features until the full set of images is used in each training epoch. By increasing the batch size, the time of each training epoch will decrease but the computational effort will increase. We opted for a balance between effort and time using training batches of 8 images. With this configuration, the GPU occupied 80% of its capacity, the internal heat generated by the GPU was kept in the range of 65 degrees Celsius and the approximate time per training epoch was 20 minutes in each GAN B. In GAN A the time was much less with a duration of 5 minutes per epoch because the calculation was considerably reduced by generating images of smaller size. The training time of GAN A and B_i is a limitation to consider. At the end of each epoch, we save the GANs to avoid losing training progress due to some external event involving server disconnection. The approximate training time for the GAN B_i was 47 hours for each class and 50 hours for the GAN A.

For practical use, it is more important to check the runtime of the SCG system to synthesize new artificial images. We measure the runtime to create a batch of artificial images in four experiments: 1) generation of 100 low-resolution images of GAN A; 2) transformation from 100 low-resolution images to 100 high-resolution images in each GAN B_i; 3) transformation of the 100 low resolution images to 700 high resolution images of all the GANs B_i; and 4) the combined generation and transformation of 700 artificial images between GAN A and all GANs B_i. The average result of 10 measurement trials for the first experiment was 42.22 seconds, 2.35 seconds for the second experiment, 16.11 seconds for the third and 58.92 for the last experiment.In summary, the developed SCG system is capable of creating 100 synthetic samples of each class with a total of 700 artificial cell images in approximately 1 minute.

For the present work, we consider that the results were satisfactory for the available equipment. More powerful devices could be used to reduce the calculation effort and training time.

6. Conclusions and limitations

The final contribution of this work is a GAN-based method capable to generate synthetic images of leukocytes and leukemic cells morphologically similar to those circulating in peripheral blood. It has been implemented through a two-stage system. The first stage allows to generate a general class of basic blood cells, while the second is specialized in a specific class to generate high resolution images with realistic morphology. It is the user who chooses which model is used depending on the type of cell to generate. This realism has been validated by expert clinical pathologists and through artificial classifiers. The synthetic cells have proved to be useful for increasing image training sets and, once trained, the system requires a low computational cost, helping to create data sets in a short time. Therefore, we believe that it can help to train automatic recognition systems for the final clinical practice. Up to the authors' knowledge, this work is one of the few first contributions in artificial generation of blood cell images oriented to the clinical laboratory.

This work has limitations that motivate future works. The sample collection time for system formation is considerably long. The real images of GAN B_6 and B_7 were collected over several years, not all patients have hairy cells and atypical promyelocytes. For broader application, it is necessary to include malignant cell classes beyond those used in the present study. Our next step will be to add other types of acute leukemia and lymphoma cells. The sequential architecture should provide a flexible framework for extensions, as they can be achieved by designing specific GANs B trained for each new class.

In this study, real images were collected in the CORE Laboratory of Hospital Clinic of Barcelona using May Grünwald-Giemsa staining and the CellaVision DM96 analyzer to obtain the digital cell images. The variability associated with various factors, such as different laboratories, staining protocols and analyzers, is a recurring limitation when it comes to learning-based approaches involving blood cell morphology. This variability can affect the morphology of synthetic images and, consequently, the performance of automatic classifiers. This is an issue not addressed in this paper and requires further standardization work which is now in progress.

Authors' contributions

Kevin Barrera developed and training the system, implemented the classifiers for the experiments, obtained the experimental results and contributed to the writing of the manuscript.

Anna Merino designed the experiments, leaded the morphological and clinical interpretations and supervised the manuscript writing.

Angel Molina contributed to image collection, morphological annotation and interpretation of the experiments.

José Rodellar supervised the overall project, advised on the deep learning methods, designed the manuscript contents and contributed to writing and editing.

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.

Acknowledgments

This work is part of a research project funded by the Ministry of Science and Innovation of Spain, with reference PID2019-104087RB-I00. K. Barrera appreciates the financial support of the Secretary of Higher Education, Science, Technology and Innovation of Ecuador as a personal sponsor entity. The collaboration of Javier Laguna, María Rodríguez and Laura Boldú, members of the Biochemistry and Molecular Genetics Department of the Hospital Clinic of Barcelona, is appreciated in the visual assessment of the synthetic cells.

References

[1] A. Merino, L. Puigví, L. Boldú, S. Alférez, J. Rodellar, Optimizing morphology through blood cell image analysis, Int. J. Lab. Hematol. 40 (2018) 54–61.
[2] J. Rodellar, S. Alférez, A. Acevedo A Molina, A. Merino, Image processing and machine learning in the morphological analysis of blood cells, Int. J. Lab. Hematol. 40 (2018) 46–53.
[3] H. El Archi, J. Khoury, Artificial intelligence and digital microscopy applications in diagnostic hematopathology, Cancers (Basel) 12 (2020) 797, doi:10.3390/cancers12040797.
[4] W. Walter, C. Haferlach, N. Nadarajah, I. Schmidts, C. Kuhn, W. Kern, T. Haferlach, How artificial intelligence might disrupt diagnostics in hematology in the near future, Oncogene 40 (2021) 4271–4280.
[5] A. Shahin, Y. Guo, K. Amin, A. Sharawi, White blood cells identification system based on convolutional deep neural learning networks, Comput. Methods Programs Biomed. 168 (2017) 69–80.
[6] 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.
[7] L. Vogado, R. Veras, F. Araujo, R. Silva, K. Aires, Leukemia diagnosis in blood slides using transfer learning in cnns and svm for classification, Eng. Appl. Artif. Intell. 72 (2018) 415–422.
[8] Y. Qiao, Y. Zhang, N. Liu, P. Chen, Y. Liu, An end-to-end pipeline for early diagnosis of acute promyelocytic leukemia based on a compact cnn model, Diagnostics 11 (7) (2021) 1237.
[9] S. Shafique, S. Tehsin, Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks, Technol. Cancer Res. Treat. 17 (2018). 1533033818802789
[10] S. Ramaneswaran, K. Srinivasan, R. Durai, P. Vincent, C.-Y. Chang, Hybrid inception v3 xgboost model for acute lymphoblastic leukemia classification, Comput. Math. Methods Med. 2029 (2021) 7519603.
[11] 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.
[12] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48, doi:10.1186/S40537-019-0197-0/FIGURES/33. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0
[13] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, M. Song, Neural style transfer: A Review, IEEE Trans. Vis. Comput. Graph 26 (11) (2020) 3365–3385, doi:10.1109/TVCG.2019.2921336.
[14] T. Bai, J. Luo, J. Zhao, B. Wen, Q. Wang, Recent advances in adversarial training for adversarial robustness, IJCAI Int. Joint Conf. Artif. Intell. (2021) 4312–4321, doi:10.48550/arxiv.2102.01356.
[15] L. Gonog, Y. Zhou, A review: generative adversarial networks, Proceedings of the 14th IEEE Conference on Industrial Electronics and Applications, ICIEA 2019 (2019) 505–510, doi:10.1109/ICIEA.2019.8833686.
[16] L. Boldú, A. Merino, S. Alférez, A. Molina, A. Acevedo, J. Rodellar, Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis, J. Clin. Pathol. 72 (11) (2019) 755–761.
[17] J. Paillassa, X. Troussard, Patients with relapsed/refractory hairy-cell leukemia, Cancer Rep. 5 (3) (2022) e1495.
[18] 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.
[19] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings (2015) 1511.06434v2, doi:10.48550/arxiv.1511.06434.
[20] M. Mirza, S. Osindero, Conditional Generative Adversarial Nets(2014). arXiv preprint: 1411.1784v11411.1784 10.48550/arxiv.1411.1784
[21] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, Others, 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.
[22] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
[23] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: interpretable representation learning by information maximizing generative adversarial nets, Adv. Neural Inf. Process. Syst. 29 (2016).
[24] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN (2017). arXiv preprint: 1701.07875v31701.07875
[25] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, A. Mukhopadhyay, GANs For medical image analysis, Artif. Intell. Med. 109 (July) (2020) 101938, doi:10.1016/j.artmed.2020.101938.
[26] J.M. Wolterink, A.M. Dinkla, M.H. Savenije, P.R. Seevinck, C.A. van den Berg, I. Išgum, Deep MR to CT synthesis using unpaired data, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10557 LNCS (2017) 14–23, doi:10.1007/978-3-319-68127-6_2.
[27] W. Yan, Y. Wang, S. Gu, L. Huang, F. Yan, L. Xia, Q. Tao, The domain shift problem of medical image segmentation and vendor-Adaptation by unet-GAN, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11765 LNCS (2019) 623–631, doi:10.1007/978-3-030-32245-8_69.
[28] J. Son, S.J. Park, K.-H. Jung, Retinal Vessel Segmentation in Fundoscopic Images with Generative Adversarial Networks (2017). arXiv preprint: 1706.09318 1706.09318
[29] G. Yang, J. Lv, Y. Chen, J. Huang, J. Zhu, Generative Adversarial Networks (GAN) Powered Fast Magnetic Resonance Imaging – Mini Review, Comparison and Perspectives(2021). arXiv preprint: 2105.01800 2105.01800
[30] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-Based synthetic medical image augmentation for increased CNN performance in liver lesion classification, Neurocomputing 321 (2018) 321–331, doi:10.1016/J.NEUCOM.2018.09.013.
[31] C. Han, Y. Kitamura, A. Kudo, A. Ichinose, L. Rundo, Y. Furukawa, K. Umemoto, Y. Li, H. Nakayama, Synthesizing diverse lung nodules wherever massively: 3D multi-Conditional GAN-Based CT image augmentation for object detection, Proceedings - 2019 International Conference on 3D Vision, 3DV 2019 (2019) 729–737, doi:10.1109/3DV.2019.00085.

[32] C. Han, L. Rundo, R. Araki, Y. Nagano, Y. Furukawa, G. Mauri, H. Nakayama, H. Hayashi, Combining noise-to-image and image-to-image GANs: brain MR image augmentation for tumor detection, IEEE Access 7 (2019) 156966–156977, doi:10.1109/ACCESS.2019.2947606.
[33] H. Uzunova, J. Ehrhardt, H. Handels, Memory-efficient GAN-based domain translation of high resolution 3D medical images, Comput. Med. Imaging Graph. 86 (2020) 101801, doi:10.1016/J.COMPMEDIMAG.2020.101801.
[34] X. Wu, Y. Wu, X. Li, X. Zhang, X. Zhou, Urine red blood cells generation using stylegan2 network, ACM Int. Conf. Proceed. Ser. (2020) 244–248, doi:10.1145/3433996.3434487.
[35] O. Bailo, D. Ham, Y.M. Shin, Red blood cell image generation for data augmentation using Conditional Generative Adversarial Networks, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops 2019-June(2019) 1039–1048. arXiv preprint: 1901.06219v21901.06219 10.48550/arxiv.1901.06219
[36] K. Liu, R. Shuai, L. Ma, ZeXu, Cells image generation method based on VAE-SGAN, Procedia Comput. Sci. 183 (2021) 589–595, doi:10.1016/J.PROCS.2021.02.101.
[37] D. Hazra, Y.C. Byun, W.J. Kim, C.U. Kang, Synthesis of microscopic cell images obtained from bone marrow aspirate smears through generative adversarial networks, Biology (Basel) 11 (2) (2022) 1–17, doi:10.3390/biology11020276.
[38] H. Eilertsen, C.E. Henriksson, T.-A. Hagve, The use of cellavision™ DM 96 in the verification of the presence of blasts in samples flagged by the sysmex XE-5000, Int. J. Lab. Hematol. 39 (4) (2017) 423–428.
[39] I. Goodfellow, Y. Bengio, A. Courville, Deep learning (adaptive computation and machine learning series), Cambridge Massachusetts (2017) 241–242.
[40] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural. Inf. Process Syst. 27 (2014).
[41] S. Kolouri, P.E. Pope, C.E. Martin, G.K. Rohde, Sliced Wasserstein auto-encoders, in: International Conference on Learning Representations, 2018.
[42] I. Marin, S. Gotovac, M. Russo, D. Božić-Štulić, The effect of latent space dimension on the quality of synthesized human face images, J. Commun. Softw. Syst. 17 (2) (2021) 124–133, doi:10.24138/jcomss-2021-0035.
[43] G. Liu, R. Taori, T.-C. Wang, Z. Yu, S. Liu, F.A. Reda, K. Sapra, A. Tao, B. Catanzaro, Transposer: universal texture synthesis using feature maps as transposed convolution filter, arXiv preprint arXiv:2007.07243 (2020).
[44] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–21, doi:10.1109/TNNLS.2021.3084827.
[45] 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.
[46] W. Weng, X. Zhu, U-Net: Convolutional Networks for Biomedical Image Segmentation, IEEE Access 9(2015) 16591–16603. arXiv preprint: 1505.04597v11505.04597 10.48550/arxiv.1505.04597
[47] S. Pipitone, L. Germagnoli, G. Da Rin, A. Di Fabio, A. Fanelli, F. Fiorini, S. Francione, A. Marini, A. Papa, A. Benegiamo, Others, Comparing the performance of three panels rules of blood smear review criteria on an italian multicenter evaluation, Int. J. Lab. Hematol. 39 (6) (2017) 645–652.
[48] E. Gemen, The sysmex SP1000i for automated bone marrow slide smear staining, LABMEDICINE 40 (2009) 23–27.
[49] F. Mameli, M. Bertini, L. Galteri, A. Del Bimbo, A NoGAN approach for image and video restoration and compression artifact removal, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 9326–9332.
[50] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, Adv. Neural Inf. Process. Syst. 30 (2017).
[51] S. Barratt, R. Sharma, A note on the inception score, arXiv preprint: 1801.01973(2018).
[52] H. Talebi, P. Milanfar, Learned perceptual image enhancement, in: 2018 IEEE international conference on computational photography (ICCP), IEEE, 2018, pp. 1–13.
[53] A. Acevedo, A. Merino, S. Alférez, Á. Molina, L. Boldú, J. Rodellar, A dataset of microscopic peripheral blood cell images for development of automatic recognition systems, Data in Brief, ISSN: 23523409, Vol. 30,(2020) (2020).
[54] J. Rodellar, K. Barrera, S. Alférez, L. Boldú, J. Laguna, A. Molina, A. Merino, A deep learning approach for the morphological recognition of reactive lymphocytes in patients with COVID-19 infection, Bioengineering 9 (5) (2022) 229, doi:10.3390/bioengineering9050229. https://www.mdpi.com/2306-5354/9/5/229/pdf