Haematological Cell Image Classification using Self-Supervised and Transfer Learning

Dissertation

zur Erlangung des akademischen Grades einer
Doktorin der Medizin (Dr. med.)

an der
Medizinischen Fakultät der Universität Hamburg

vorgelegt von

Laura Wenderoth

aus

Dortmund

2025

Betreuer:in / Gutachter:in der Dissertation: Prof. Dr. René Werner

Gutachter:in der Dissertation: Prof. Dr. Tobias Knopp

Vorsitz der Prüfungskommission: Prof. Dr. Tobias Knopp

Mitglied der Prüfungskommission: Prof. Dr. Frank Ückert

Mitglied der Prüfungskommission: PD Dr. Andreas Block

Datum der mündlichen Prüfung: 01.12.2025

Contents

1 Presentation of the Publication

1.1 Introduction

Haematological malignancies, including leukaemia, lymphoma, and myeloma, represent a broad spectrum of diseases with complex cellular characteristics and numerous subtypes. Accurate identification of the malignancies is crucial for precise diagnosis, treatment selection, and disease monitoring [1, 2]. In this regard, blood and bone marrow smears, along with cell counting techniques, play an important role in diagnostics, enabling the identification, subtyping, and monitoring of abnormal cell populations [3].

In clinical settings, manual counting of blood and bone marrow cells from smears is a fundamental procedure performed by skilled laboratory professionals [4]. The process involves visually examining the smear under a microscope and counting cells of various types manually to determine their frequency. This procedure is time-consuming [4] and affected by variations between examiners as it depends on their expertise [5, 6]. Automating the process - cell detection and classification - using Deep Learning (DL) techniques significantly improves efficiency [7, 8].

However, there are three key shortcomings in the existing DL approaches: (1) the need for *huge labour-intensive, manually labelled datasets*, (2) the requirement of *specialized and expensive hardware and training time*, (3) *limited transferability* between laboratories. First, large datasets requiring labour-intensive manual labelling are necessary to train supervised DL models. Second, training DL models for classification requires specialised hardware, such as a Graphics Processing Unit (GPU), and can take several days, consuming great amounts of energy. Third, trained blood cell classification models cannot be easily transferred to classify blood cell images acquired from different laboratories or scanners due to the lack of standardization in staining methods, optical magnifications, and colour representations [9, 10]. Variations in sample preparation, staining protocols, colour intensity, and imaging conditions can substantially affect classification performance and robustness [11, 12, 13].

In our previous study [14], we demonstrated that Self-Supervised Learning (SSL) can reduce the labelling requirement to approximately 250 labels per class while achieving performance comparable to Supervised Learning (SL) on Bone Marrow (BM) datasets. This approach holds promise for addressing the labelling bottleneck. However, the other outlined challenges remain unsolved. Each laboratory and digitization method still requires training a new model on specialized hardware, such as GPUs. For this reason, my research addresses all three challenges by developing an approach that:

- Enables transferability across different blood datasets without requiring extensive training.

- Eliminates the need for large, labour-intensive manually labelled datasets – approximately 50 labels per class are sufficient.

- Operates without specialised hardware – all computations can be performed on typical consumer-grade laptops using only Central Processing Unit (CPU), without the need for GPUs.

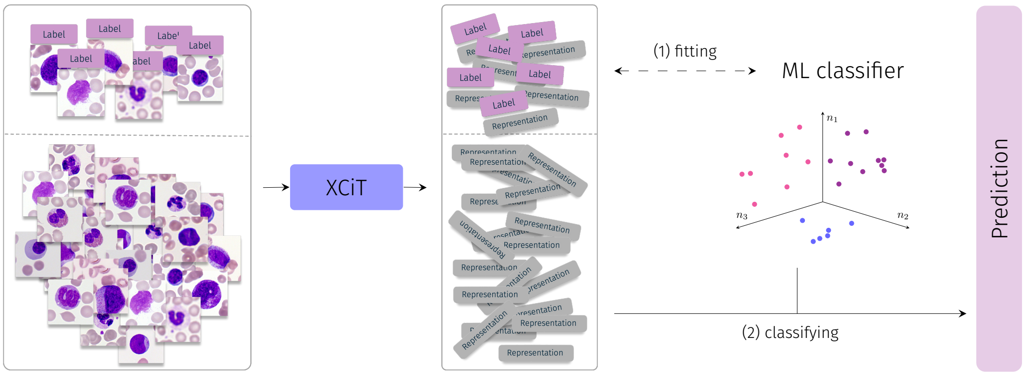


Figure 1.1: Illustration of the proposed blood cell image classification approach. Features are extracted for the entire dataset using a pretrained XCiT model on bone marrow cell images with SSL - no additional training or specialized hardware required. A lightweight classifier is trained on features from only 50 labelled cells per class. Once fitted, it accurately classifies the remaining unlabelled dataset, reducing labelling effort and computational demands.

As shown in Figure 1.1, the approach begins by extracting cell image features for the whole dataset using a feature extractor pretrained on bone marrow data. For a small labelled subset (50 cells per class), these extracted features and their labels are used to train a lightweight classifier. Both steps can be performed on a typical consumer-grade laptop using only CPUs. The trained classifier can then accurately classify unlabelled cell images, making the method adaptable to various datasets while minimizing labelling effort and hardware requirements.

1.2 Related Work

Traditionally, SL on large labelled datasets has been the dominant approach for blood cell classification, achieving high performance in controlled settings. However, this approach has notable limitations: it relies heavily on extensive manual labelling, often requires specialized hardware for training, and struggles to generalize across different datasets. While several studies have attempted to address one or two of these challenges, as summarized in Table 1.1, none have fully addressed all three.

For example, Acevedo et al. [15] aimed to reduce hardware dependency by applying transfer learning with a deep neural network pretrained on ImageNet. They utilised the pretrained network to extract image features, which were then used to train a traditional supervised ML classifier. This approach achieved an unbalanced accuracy of 90.5% on the Acevedo blood dataset [16], falling short of the 94.9% achieved by standard supervised learning. Although it reduced computational costs, it still required the entire labelled Acevedo blood dataset of over 17,000 images. This suggests that transfer learning from non-medical images, such as natural objects, is of limited use in highly specialised fields like haematology.

Long et al. [17] explored reducing the labelling burden by training on a subset of the Acevedo blood dataset [16]. They randomly selected 1,000 samples from a total of 17,092, using approximately 6% of the images and their labels. However, this resulted in a substantial accuracy drop from 99.3% to 82.2%, highlighting the sensitivity of supervised methods to labelled data availability. This example

Table 1.1: Overview of blood cell image classification studies addressing key limitations: scarcity of labelled data (1), specialized hardware requirements (2), and transferability (3) to unseen datasets.

Paper	Dataset	Year	Adressed Shortcoming	Description	Results
Acevedo et al. [15]	Acevedo Blood [16]	2019	Specialized Hardware (2)	Used pretrained ImageNet to extract features and trained an ML classifier on the features using all labels; Transfer Learning.	90.5% unbalanced accuracy - drop in accuracy compared to training using the cell images.
Long et al. [17]	Acevedo Blood [16]	2021	Labelled Data (1)	In one subexperiemt used only ∼ 6% of labels (1000 random cell images with labels of a 17,092 cell image dataset) with fully supervised training on the cell images.	82.2% accuracy - drop in accuracy compared to training using all labels and images.
Elhassan et al. [18]	Matek Blood [19], Acevedo Blood [16]	2022	Specialized Hardware (2)	Trained a CNN on Matek Blood to extract 128 features, then used these features to train an ML classifier; applied the trained CNN feature extractor to Acevedo Blood and trained a supervised ML classifier; Generalizability.	Achieved SoTA accuracy: 97.5% on Matek Blood, 96.4% on Acevedo Blood showcasing generalizability of their feature extractor.
Chen et al. [13]	Acevedo Blood [16], their own blood cell dataset (∼ 10% of the size of Acevedo)	2023	Transferability (3)	Employed a CNN trained through SSL to recognize various image transformations (raw, colour jitter, grayscale, etc.) for feature extraction without labelled data. Classifiers were trained on the extracted features using the labelled target dataset.	Achieved 88.9% accuracy on Acevedo and 71.8% on their own dataset when training CNN and classifier on the respective datasets. When using the CNN trained on the other dataset for feature extraction, accuracy dropped to 67.1% on their own dataset but increased to 91.3% on Acevedo. This suggests that pretraining on larger, similar datasets can improve performance, though not all datasets are suitable for pretraining.
Nielsen et al. [14]	Matek BM [20], two other medical datasets	2023	Labelled Data (1)	Employed the DINO SSL framework for feature extraction without labelled data, using a vision transformer (XCiT). Conventional ML classifiers (SVM, LR, KNN) were trained on the extracted features with as few as 250 labelled samples per class. Evaluated across three distinct medical imaging modalities.	Achieved state-of-the-art performance with only 1% to 10% of labelled data compared to supervised baselines. Demonstrated that SSL with limited labels can achieve high accuracy while reducing labelling requirements, but still requires specialized hardware for SSL pretraining.
Wenderoth et al. [21] (Ours)	Acevedo Blood [16], Raabin Blood [22], Matek Blood [19], Matek BM [20]	2025	Labelled Data (1), Specialized Hardware (2), Transferability (3)	Used a transformer-based encoder trained through SSL on Matek BM dataset without labelled data to extract features, followed by training lightweight ML classifiers (SVM, LR, KNN) with limited labelled samples (50 per class) for each target dataset. Assessed direct transfer and domain adaptation for three blood cell datasets.	Achieved superior transferability with balanced accuracy up to 91% training the classifier with minimal labelled data. Direct transfer (classifier trained on BM dataset) without adaptation yielded lower, but still competitive, accuracies. Demonstrated that bone marrow datasets are well-suited for SSL-based pretraining, effectively facilitating efficient cross-domain transfer to blood cell datasets.

illustrates the challenge of maintaining high performance with limited annotations, emphasizing the need for more data-efficient approaches.

To address these challenges, recent research has increasingly explored SSL as a means to extract useful feature representations from unlabelled data. For example, Chen et al. [13] used SSL to train a Convolutional Neural Network (CNN) to recognise various image transformations, achieving 88.9% accuracy on the Acevedo blood dataset [16] and 71.8% on their own, smaller dataset, which is approximately 10% the size of the Acevedo blood dataset. However, the transferability of the learned features was inconsistent, with performance depending heavily on the dataset used for pretraining. These findings indicate that while SSL has potential, effective transfer learning in haematology may require more carefully designed pretraining approaches.

In our previous research, we explored SSL in haematology to address challenges related to extensive labelling. In Nielsen et al. [14], we applied the DINO SSL method to three medical image datasets, including a public bone marrow dataset containing over 170,000 images [20]. The extracted features enabled the training of lightweight ML classifiers, such as Support Vector Machines (SVM), Logistic Regression (LR), and K-Nearest Neighbours (KNN), using minimal labelled data. For the classification of the BM dataset, we used only 250 samples per class.

The results show that SSL consistently outperforms SL when working with minimal datasets. For the BM dataset, a balanced accuracy of 73% was achieved using 250 images per class. Our findings confirm that SSL can reduce labelling demands while maintaining strong performance. However, our previous work did not explore transfer learning, where pretraining on one dataset improves performance on another while reducing hardware requirements. This gap motivated further research into whether bone marrow cell images offer a better pretraining source than natural image datasets, such as ImageNet for blood cell image classification.

In this work, we propose a novel approach combining SSL and transfer learning specifically for haematology. By pretraining on bone marrow cell images — which share morphological similarities with peripheral blood cells — we aim to create a feature extractor that can generalize across different blood cell datasets. Unlike previous studies, we minimize the labelling effort by requiring only 50 labelled cells per class. Furthermore, our method is designed to be computationally efficient, requiring no specialized hardware. This makes it accessible for routine laboratory use, addressing all three of the identified limitations: labelling effort, hardware dependency, and transferability.

1.3 Fundamentals

To understand why bone marrow is hypothesized to be well-suited for pretraining feature extraction in blood cell image classification, we must examine the physiological relationship between blood and bone marrow. In the following Section 1.3.1, this relationship will be explained in detail. Additionally, for feature extraction, a neural network is required - a structure capable of learning meaningful features from images. In our case, it is achieved through the Vision Transformer (ViT), which is briefly introduced in Section 1.3.2.

1.3.1 Haematological Maturation Process

Bone marrow and blood are intrinsically linked through the development and maturation of haematopoietic cells, as Figure 1.2 illustrates. All blood and bone marrow cells originate from a common stem cell, progressing through sequential differentiation stages that involve precursor and progenitor populations primarily located in the bone marrow. Immature precursor cells are abundant in the bone marrow, where most differentiation occurs, while mature cells predominantly circulate in peripheral blood [10]. Haematopoietic cells can be categorized into five main lineages, each giving rise to specific cell types. These lineages originate from a common haematopoietic stem cell and follow distinct differentiation pathways, highlighting the fundamental connection between bone marrow and peripheral blood.

Erythropoiesis is responsible for the production of red blood cells (erythrocytes), which are essential for oxygen transport. The process involves a series of stages, including the formation of proerythroblasts and erythroblasts, culminating in mature erythrocytes.

Granulopoiesis generates granulocytes, which include neutrophils, eosinophils, and basophils. The developmental sequence involves several stages: myeloblast, promyelocyte, myelocyte, metamyelocyte, band cell, and finally mature granulocytes. Neutrophils are the most abundant granulocytes and are critical for innate immunity, while eosinophils and basophils play key roles in allergic responses and parasitic defence.

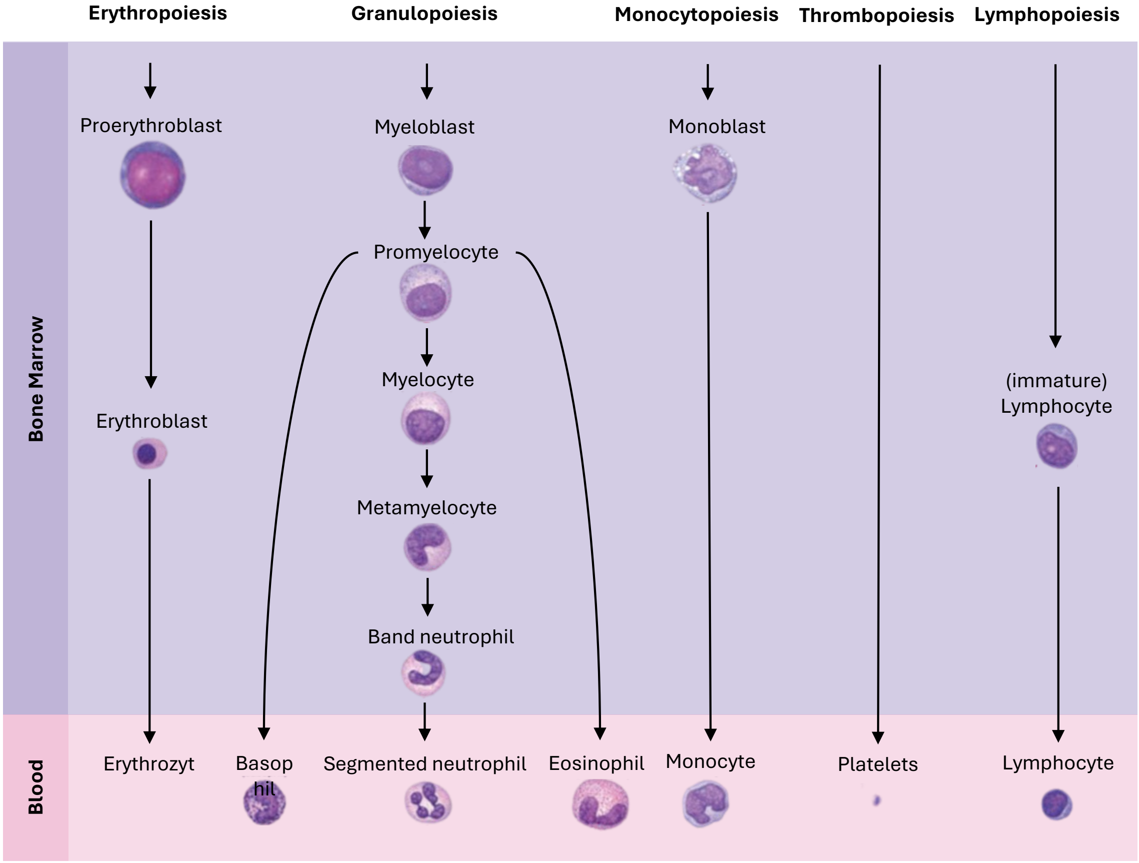


Figure 1.2: Maturational sequence of haematopoietic cells, illustrating the differentiation pathways of key lineages. Cells highlighted with a pink background (bottom) represent those predominantly found in peripheral blood under normal physiological conditions, while cells with a purple background (top) are primarily located in the bone marrow. Adapted from Ward, Cherian, and Linden [23].

Monocytopoiesis produces monocytes, which are precursors to macrophages and dendritic cells. The differentiation process starts with monoblasts, progresses to monocytes, and leads to the formation of mature macrophages in tissues. Monocytes are crucial for phagocytosis and immune regulation.

Megakaryopoiesis produces megakaryocytes, which are responsible for platelet formation. Megakaryocytes undergo a process of endomitosis and cytoplasmic fragmentation to release platelets into circulation, essential for haemostasis.

Lymphopoiesis generates lymphocytes, including T cells, B cells, and natural killer cells, which are fundamental to adaptive and innate immunity. Lymphocytes develop in specialised environments such as the bone marrow and the thymus.

In summary, the strong physiological relationship between blood and bone marrow is evident, as all five lineages first develop in the bone marrow before releasing mature cells into the bloodstream. The transferability of a model trained on bone marrow cells to blood datasets holds significant promise due to the biological relationship between the two. Immature precursor cells in the bone marrow differentiate into mature cells found in peripheral blood, meaning that features learned from bone marrow data are inherently relevant for blood cell classification tasks. The shared developmental origin provides a strong foundation for effective model transfer.

1.3.2 Vision Transformer

Our classification method employs a modified Vision Transformer (ViT) [24] known as the Cross-Covariance Image Transformer (XCiT) [25]. The following section introduces the original ViT and outlines the advantages of XCiT over ViT.

ViT In recent years, deep learning has made great progress in the field of image processing, with one being the ViT. This model has demonstrated high accuracy in object recognition tasks [24]. To fully comprehend the functioning of ViTs, it is essential to first understand the core principles that underpin these models.

Unlike humans, computers do not perceive images in the same way. Instead, they represent images as grids of numbers, where each number corresponds to the intensity or colour value of an individual pixel. To interpret an image, the ViT does not process the entire image at once. Rather, it divides the image into smaller square regions known as patches. Each patch is treated as a distinct unit of information.

Once the image is divided into patches as illustrated in Figure 1.3, each patch is flattened into a one-dimensional vector of pixel values. These vectors are then linearly projected through a trainable weight matrix, producing an embedding that encapsulates significant features, such as textures and edges and ensures a consistent embedding size.

The patch embeddings are then concatenated to form a sequence. To maintain spatial information, a learned positional encoding is added to each embedding, as ViT do not inherently capture the position of elements within a sequence. These positional encodings represent the original spatial location of each patch, allowing the model to recognise and preserve spatial relationships. The resulting sequence of patch embeddings, now enriched with positional information, is subsequently input into the ViT encoder.

To analyse the relationships between different image patches, ViTs utilise self-attention, a mechanism that compares each patch with every other patch to comprehend the overall structure of the image [26]. To enable this comparison between patches, the model uses three matrices: query (Q), key (K), and value (V). The query (Q) defines what a patch is seeking from other patches, much like a word in a sentence searching for related words. The key (K) acts as a descriptor of the patch's content, allowing other patches to assess its relevance. Finally, the value (V) contains the actual information of the patch, which is passed along if the query finds a strong match with the key. In summary, the self-attention mechanism determines the level of attention each patch should give to every other patch by calculating attention scores using the following formula:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \times V. \tag{1.1}$$

The term QK^T calculates the similarity between each patch's query and all other patches' keys, effectively measuring their relevance to one another. The denominator, $\sqrt{d_k}$, serves as a scaling factor, preventing excessively large values in the dot product, which could destabilise training by causing extremely small gradients. The softmax function then normalises these similarity scores into a probability distribution, ensuring that the attention weights assigned to all patches sum to one. Finally, these weights are applied to the value matrix V, allowing the model to aggregate information from the most relevant patches while downplaying less significant ones.

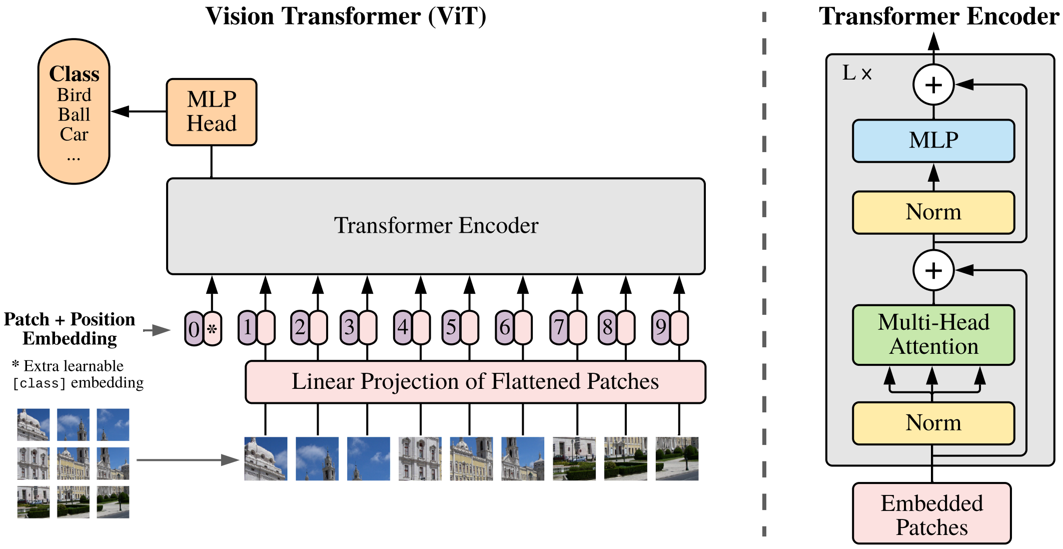


Figure 1.3: Visualisation of Vision Transformer architecture. The image is first divided into non-overlapping patches, which are then flattened and linearly embedded into a fixed-size vector for each patch. These patch embeddings are augmented with positional encodings to retain spatial information. The resulting sequence of embeddings is passed through a standard Transformer architecture, consisting of multiple layers of self-attention and feed-forward networks, to model global dependencies across the image. The output is passed through a classification head to produce the final prediction. Copied from Dosovitskiy et al. [24].

By comparing the query and key vectors across patches, the model determines which patches should influence each other, guiding its attention to the most significant image regions.

Rather than computing attention once for each patch, ViTs use multiple attention heads in parallel, a mechanism known as multi-head self-attention. Each attention head learns different relationships within the image, allowing the model to capture diverse aspects of the data.

The attention heads can be visualised by mapping the attention weights between image patches, revealing how the model assesses the relevance of different regions. These visualisations highlight the learned relationships and potentially show which parts of the image focus on key features, such as edges or textures, illustrating the model's understanding of spatial relationships and dependencies.

After passing through multiple ViT layers, the model aggregates the information from all image patches. The final representation, which captures the relevant features of the image, is then passed through a Multi-Layer Perceptron (MLP). The MLP processes the aggregated information to produce the final classification output, such as identifying the object or category in the image.

Vision Transformers have demonstrated significant success due to their ability to capture global dependencies across the entire image, enabling them to learn complex relationships and contextual information.

XCiT While the ViT achieves strong performance, its self-attention mechanism is computationally expensive, with quadratic time complexity $O(N^2)$, where N is the number of patches, making it impractical for high-resolution images. The Cross-Covariance Image Transformer (XCiT) [25] addresses this by introducing Cross-Covariance Attention (XCA), which shifts from patch-wise to channel-wise attention by switching and transposing the query and key matrices. Additionally, XCiT normalises the query and key matrices so that their values lie within the range of $[-1, 1]$, a step that improves the stability of training. These normalised matrices are referred to as $\hat{K}$ and $\hat{Q}$. As a result, the XCA formula is defined as:

$$\text{Cross-Covariance Attention}(Q, K, V) = V \times \text{Softmax}\left(\frac{\hat{K}^T\hat{Q}}{\tau}\right). \tag{1.2}$$

Here, τ is a temperature scaling factor, and the softmax function normalises the cross-covariance matrix $\hat{K}^T\hat{Q}$, producing the attention weights. These weights are then applied to the value matrix V, determining how much each feature channel contributes to the final output. This approach reduces computational complexity to linear with respect to the number of patches, enabling more efficient processing of high-resolution images, while maintaining strong performance and scalability.

1.4 Material and Methods

This section provides a detailed overview of the datasets used in the study, highlighting the significant differences in their digitisation and illustrating the challenges in transferring our approach across datasets. We then introduce the SSL method DINO and explain its application. Following this, the classification pipeline is presented, along with the experiments conducted. Visualisations of these can be found in Figures 2 and 3 in the paper in Section 2.

1.4.1 Datasets

The chosen datasets consist of digitised, stained blood and bone marrow smears. This study utilises one bone marrow dataset and three blood datasets. In the original paper in Section 2, the different cell classes and their distributions within the datasets are detailed in Tables 1 and 2. Therefore, this section focuses on a thorough explanation of dataset creation to highlight the impact of digitisation methods on potential deep learning classification algorithms.

Matek Bone Marrow Dataset [20]: The bone marrow dataset utilized in this study comprises a collection of 171,374 single-cell images obtained from bone marrow smears of 961 patients diagnosed with a wide range of haematological diseases. These diseases encompass myeloid and lymphoblastic malignancies, lymphomas, and non-malignant and reactive changes. The dataset was sourced from the Munich Leukaemia Laboratory MLL, spanning between 2011 and 2013. The image acquisition process involved applying May-Grünwald-Giemsa/Pappenheim staining and capturing using a CCD camera mounted on a brightfield microscope (Zeiss Axio Imager Z2). A 40× oil immersion objective was employed, yielding images with a size of 250 × 250 pixels.

The dataset encompasses a total of 21 distinct classes, each representing a specific cell type or morphological category. These classes include band neutrophils, segmented neutrophils, lymphocytes, monocytes, eosinophils, basophils, metamyelocytes, myelocytes, promyelocytes, blasts, plasma cells, proerythroblasts, erythroblasts, hairy cells, abnormal eosinophils, immature lymphocytes, smudge cells and other cells, artefacts, and unidentifiable cells.

It is worth noting that the class "artefacts" includes cells that are deemed unidentifiable, while the class "other cells" encompasses morphological classes not represented by the other specific cell types. Importantly, the dataset shows a highly imbalanced distribution across the cell classes. Detailed information can be found in Table 2 of the paper in Section 2.

Matek Blood Dataset [19]: The first blood cell dataset employed in this study comprises a total of 18,365 single-cell images, each possessing a size of 400 × 400 pixels. These images were extracted from peripheral blood smears obtained from 100 patients diagnosed with different subtypes of Acute Myeloid Leukaemia (AML) at the University Hospital Munich between 2014 and 2017.

The blood smears were scanned at ×100 optical magnification using an M8 digital microscope/scanner equipped with oil immersion. The cells were classified into 15 distinct categories by trained specialists. These categories include segmented neutrophils, banded neutrophils, typical lymphocytes, atypical lymphocytes, monocytes, eosinophils, basophils, myeloblasts, promyelocytes, bilobed promyelocytes, myelocytes, metamyelocytes, monoblasts, erythroblasts, and smudge cells.

Acevedo Blood Dataset [16]: The second blood dataset utilized in this study encompasses a publicly available compilation of 17,092 blood cell images acquired at the Hospital Clinic of Barcelona over a span of four years, from 2015 to 2019. These images were subjected to staining using the May-Grünwald-Giemsa technique and possess a size of 360 × 363 pixels.

Two domain experts participated in the annotation and classification of these images, resulting in the assignment of labels to eight distinct groups. These groups comprise neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (including promyelocytes, myelocytes, and metamyelocytes), erythroblasts, as well as platelets or thrombocytes

Notably, the data collection for this dataset exclusively involved healthy patients without any infections, haematological disorders, or oncological conditions.

Raabin Blood Dataset [22]: The third blood dataset used in this study consists of 14,514 blood cell images from 56 peripheral blood smears and was obtained in 2021. Two domain experts classified the images into five categories: lymphocytes, monocytes, neutrophils, eosinophils, and basophils. The blood smears were obtained from healthy individuals, except for one case of Chronic Myeloid Leukaemia (CML), which was specifically used for basophil extraction.

The smears were stained using the Giemsa technique and imaged at 100× magnification using Olympus CX18 and Zeiss microscopes. Additionally, smartphone cameras, including Samsung Galaxy S5 and LG G3, were used for image acquisition. Due to the recording method, multiple images of the same cell may be present in the dataset. However, a predefined split into training and test sets ensures a systematic separation for evaluation.

1.4.2 Self-Supervised Learning - DINO

In Machine Learning (ML), supervised and unsupervised learning are fundamental paradigms. Supervised learning trains models on large labelled datasets, adjusting parameters to minimize prediction errors and achieve high accuracy. While effective, it is resource-intensive due to the need for extensive human-annotated labels. In contrast, unsupervised learning discovers patterns and structures in data without relying on labelled examples, making it useful for clustering and anomaly detection. Bridging these paradigms, SSL has emerged as a subset of supervised learning. Rather than relying on external labels, SSL methods create auxiliary tasks, such as rotation prediction, jigsaw puzzles (rearranging shuffled image patches), and predicting context or missing information. Through these self-generated tasks, the model builds representations that capture the underlying structure of the data, effectively learning without direct human intervention. [27].

An example of SSL is Distillation with No Labels (DINO) [27], developed by Facebook AI Research. DINO employs a teacher-student architecture, which involves two networks with identical structures: the student network and the teacher network. The core idea is that the student network learns to replicate the teacher's output, providing supervision without labelled data. Training minimises their representation differences using temperature-weighted cross-entropy loss. The teacher network receives a higher-information version of the input image, such as higher resolution or larger patches, compared to the student network. To introduce varying levels of information, DINO uses a multi-crop strategy, where each image is divided into multiple views at different resolutions. Both networks process the high-resolution views, capturing global features, while only the student processes the lower-resolution views, capturing local features. This approach helps the student learn patterns across different image scales, making the model more robust to variations in size and location.

The teacher network can act as a supervisor despite not being explicitly trained on the dataset, as it is updated continuously through a momentum-based mechanism called Exponential Moving Average (EMA), allowing the teacher to evolve based on the student's learning. EMA smooths the teacher's updates by weighting the most recent student network parameters along with a decaying average of previous teacher network values. This gradual update ensures that the teacher network remains stable and reliable, incorporating the student's learning without overreacting to short-term changes.

A key issue in machine learning is data collapse (or representational collapse), where the model's outputs become excessively similar across different inputs. While this may satisfy the learning objective, it leads to ineffective learning and poor generalisation. Specifically, it means that both the teacher and student produce identical outputs, regardless of variations in the input cell images.

DINO employs two key strategies to avoid data collapse: centring and output sharpening. The centring mechanism stabilizes training by adding a bias term to the teacher network's output. This bias, updated via an EMA of the teacher's previous outputs, prevents representational collapse by maintaining variability in the teacher's outputs.

The second strategy, output sharpening, involves adjusting the teacher network's output distribution to make it more distinct. By lowering the temperature of the teacher's softmax function, the outputs become sharper, reducing the risk of overly smooth or similar responses. This sharpening process encourages the model to focus on finer distinctions in the data, making it more robust and able to differentiate between various inputs. Together, centring and output sharpening allow DINO to maintain stability during training, preventing representational collapse and ensuring the model learns useful and diverse feature representations.

DINO extends knowledge distillation to self-supervised learning through a combination of multi-crop augmentations, momentum-based teacher updates, and techniques such as centring and sharpening to prevent representational collapse. These components enable DINO to learn high-quality, scalable representations without requiring labelled data, demonstrating strong performance across various downstream tasks. Additionally, its design enhances flexibility across different neural network architectures, making it a versatile approach to pretraining.

1.4.3 Classification Pipeline

The cell image classification pipeline consists of two main stages: self-supervised feature extraction and cell classification. In the first stage, the XCiT (see Section 1.3.2) is trained using the self-supervised approach DINO [27] and the sparsam implementation [14] on the BM Matek dataset. We selected XCiT for its improved computational efficiency over the original ViT and its demonstrated effectiveness in self-supervised learning, particularly with DINO [25]. The SSL pertaining on BM allows the model to extract meaningful features from cell images without requiring annotations. Hyperparameters follow the settings defined in Caron et al. [27]. Once trained, XCiT converts each image into a 384-dimensional feature vector.

In the second stage, the extracted features are used to train lightweight supervised machine learning classifiers. We used three different classifiers: Support Vector Machine (SVM), Logistic Regression (LR), and K-Nearest Neighbors (KNN). Training with scikit-learn's default hyperparameters ensures consistency and ease of implementation. These classifiers often outperform deep learning models on small labelled datasets, offering greater robustness and adaptability.

1.4.4 Experiments

We conducted two experiments to evaluate our classification pipeline. The first, direct transfer, assesses how well the approach transfers from bone marrow to blood without any fitting or fine-tuning on blood datasets. The second, domain adaptation, tests the generalization of SSL-extracted features across datasets by fitting ML classifiers on a small number of labelled samples from target blood datasets. The same SSL feature extractor, trained on the BM dataset, was used in both experiments. Performance was measured using balanced accuracy and class-specific sensitivity.

The SSL feature extractor was initially trained on the Matek BM dataset for 48 hours. For the direct transfer experiment, a small labelled subset (250 samples per class) from the Matek BM dataset was used to fit the classifiers based on the extracted feature representations. To assess robustness, the classifiers were trained 100 times with randomly selected labelled samples. Their performance was then evaluated on three blood test datasets, considering all shared classes between the blood datasets and the Matek BM dataset while accounting for potential class discrepancies. For comparison, a supervised benchmark model using XCiT for end-to-end deep learning classification was trained three times on the same image data for 24h.

For domain adaptation, the classifiers were trained on varying amounts of labelled samples from the target blood cell dataset (ranging from 5 to 2000 per class) using the extracted feature representations. For comparison, supervised deep learning benchmark models from the literature, trained on over 10,000 labelled samples, were used as a reference.

We also explored the transfer from blood to bone marrow by training the feature extractor on the Matek Blood dataset and then transferring it to the Matek Bone Marrow dataset. The process was evaluated using the same experimental setup, encompassing both direct transfer and domain adaptation.

The experiments were performed on an Nvidia A40 GPU with 48GB of Random-Access Memory (RAM) for model training and feature extraction. The AMD EPYC 7543 processor, equipped with 32 cores and 1TB of RAM, was used for the computational tasks involved in model fitting and evaluation.

For data splitting and evaluation, each dataset was divided into a training set (70%) and a test set (30%), while preserving class distributions. The SSL feature extractor was trained on the entire training set, while the classifiers were trained using randomly selected stratified subsets of the training data. Evaluation was performed on the test set. To assess performance consistency, the supervised deep learning model was trained three times with random initialisations on the same training set.

1.5 Results

The following section summarises the results for both direct transfer, where no adaptation to the target dataset is applied, and domain adaptation, where the classifier is fine-tuned using varying sample sizes per class from the target dataset. A distinction is made between the transfer from bone marrow to blood and from blood to bone marrow. A detailed presentation of the results, including visualisations, is provided in Figure 4 and Tables 3 and 4 in the original paper in Section 2.

1.5.1 Transfer from Bone Marrow to Blood

Direct Transfer The results of direct transfer show that our approach, when applied from the Matek bone marrow dataset to three different blood cell datasets, achieved substantial classification accuracy. For the Matek blood dataset, our approach reached a balanced accuracy of around 64% across 11 cell types, compared to the SL model with a balanced accuracy of around 41%. The Acevedo dataset showed a slightly lower accuracy of 53% for 7 cell types, compared to SL model with around 46%, while the Raabin dataset achieved 63% accuracy for 5 cell types, compared to SL model 46%. This demonstrates the superiority of our approach over traditional SL methods in terms of transferability and generalizability.

Domain adaption The results of domain adaptation, where the ML classifier was trained using varying sample sizes from the target blood dataset, demonstrated performance comparable to or even surpassing state-of-the-art results, with only 50 labelled samples per class.

For the Matek blood dataset, using 50 labelled samples per class with the SVM classifier, a balanced accuracy of 76% was achieved, compared to 66% with the baseline supervised learning method using the whole dataset [19]. The superior accuracy compared to the state-of-the-art supervised baseline demonstrates that our approach can match or exceed the performance in terms of accuracy. Our approach outperformed supervised deep learning models, especially for smaller classes like erythroblasts, monoblasts, and bilobed promyelocytes, with accuracy exceeding 90%. Logistic Regression (LR) achieved the highest accuracy of 78% with 100 samples per class. However, accuracy decreased for larger sample sizes across all classifiers due to class imbalance, with larger classes (>500 samples) benefiting from more data, while smaller classes (<100 samples) were under-represented.

For the Acevedo blood dataset, both LR and SVM achieved similar performance, reaching 91% accuracy with 50 samples per class and 97% accuracy with 2000 samples, surpassing the literature benchmark of 96% [15] with 500 samples per class. The accuracy did not decrease with larger sample sizes due to the more balanced class distribution. For 50 labelled samples per class, high accuracy (>90%) was achieved for basophils, eosinophils, lymphocytes, neutrophils, and platelets, while immature granulocytes showed 79% sensitivity.

For the Raabin blood dataset, LR and SVM achieved 95% accuracy with 50 samples per class and 97% with 2000 samples, not surpassing the literature benchmark of 98% [22]. Using our approach with 50 labelled samples, high accuracy (>90%) was achieved for all classes. While the SL baseline was not exceeded, the model's performance trained using our approach could improve with more labelled dataset-specific images.

Training Time and Efficiency In terms of computational performance, the training of the ML classifiers takes approximately 1 second, with SSL pretraining requiring a one-time 48-hour training for all datasets. In comparison, the benchmark approach from the literature requires up to four days training per dataset [19].

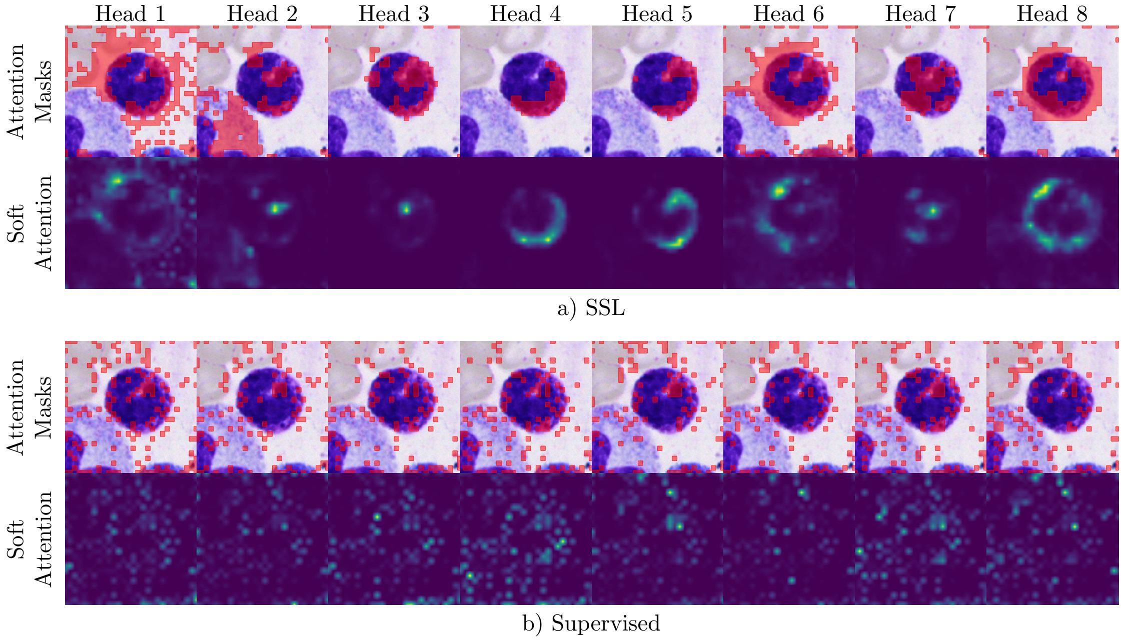
1.5.2 Transfer from Blood to Bone Marrow

Applying our approach to transfer the model trained on the Matek blood dataset to bone marrow cell classification resulted in a low accuracy of 43%. This result represents an improvement over the supervised baseline model, which achieved only 20%. Using the blood feature extractor and fine-tuning the classifier on 50 labelled bone marrow samples per class resulted in a slight increase in accuracy to 48%, although it still fell short of the BM dataset benchmark of 51% [20]. SSL pretraining on blood images did not provide an advantage for bone marrow classification.

1.6 Discussion

This thesis addressed three key shortcomings in existing blood cell classification methods: the need for large amounts of manually labelled data, high computational demands, and poor model transferability across different staining and digitalisation conditions. By combining SSL with transfer learning, we overcame these limitations, demonstrating a more efficient, resource-friendly, and adaptable approach to haematological cell image classification.

Our approach achieved state-of-the-art results on the Matek blood dataset using a classifier trained only on Matek bone marrow cells, surpassing supervised models trained on Matek blood with considerably shorter training time. While colour



a) SSL

b) Supervised

Figure 1.4: Visualization of attention from eight attention heads in the image transformer model for a basophilic granulocyte. The upper panel shows attention in the SSL approach, while the lower panel represents the supervised DL baseline. In each panel, high-attention areas (red; $\geq$ 80% of the maximum value) are overlaid on the original image. The bottom row displays continuous attention values per head (brighter colour = higher attention). The SSL model focuses on the cell, whereas the supervised approach does not, suggesting SSL's superior generalization capability.

similarities between the Matek datasets may have played a role, the strong direct transfer performance underscores the effectiveness of our method. For the Acevedo and Raabin blood datasets, direct transfer without using any fine-tuning on the blood cell dataset did not exceed the performance of dataset-specific deep learning models, but still outperformed supervised models trained on Matek bone marrow. These findings demonstrate that our approach enhances robustness and generalisation by learning meaningful and transferable cell features for classification. As shown in Figure 5 of the original paper in Section 2 and Figure 1.4, attention head visualisations reveal that SSL-trained models focus on relevant cellular features, whereas supervised models tend to overfit. This suggests that SSL does not simply memorise dataset-specific patterns but captures broader cellular characteristics, improving generalisability.

While the accuracy of direct transfer alone was not optimal for medical applications, domain adaptation significantly improved performance. By fitting lightweight ML classifiers with as few as 50 labelled samples per class from the target blood dataset, accuracy reached state-of-the-art levels, particularly for smaller but clinically relevant cell types. These results highlight the effectiveness of our approach, showing that pretraining on bone marrow alone provides a strong foundation while reducing computational demands, as the SSL-trained model only needs to be trained once for all blood datasets.

We also explored pretraining with blood cell images and transferring to bone marrow, although this approach did not achieve results meeting medical standards. From a biological perspective, it is plausible that models trained on mature blood cells may not generalise effectively to precursor bone marrow cells. From a computational viewpoint, one potential explanation is the considerable size disparity between the datasets, with the Matek blood dataset being notably smaller than the bone marrow dataset. This reduced size may have limited the model's ability to capture the full spectrum of features in peripheral blood images. Furthermore, differences in image backgrounds could contribute to the observed difficulties. Blood smears are predominantly composed of erythrocytes, whereas bone marrow smears contain a higher density of nucleated cells, which may provide more informative features for extraction. The greater heterogeneity of cell types and subtypes in bone marrow, along with the inherent variability in staining, preparation methods, and biological factors, likely contributes to the increased complexity and diversity of the bone marrow images. While these factors strongly suggest the reasons for the less successful transfer from blood to bone marrow, further studies are needed to conclusively confirm these observations.

In summary, this thesis demonstrated that SSL combined with transfer learning effectively addresses challenges in blood cell image classification, such as manual labelling, high computational demands, and poor transferability. By reducing reliance on large labelled datasets, optimising computational resources, and improving model transferability, our approach offers a more accessible, sustainable, and adaptable alternative to traditional supervised DL approaches. However, areas for improvement remain. Future work should focus on enhancing domain adaptation, particularly under varying imaging conditions, and extending this approach to whole-slide images to further minimise manual intervention.

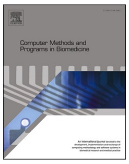
2 Publication

Computer Methods and Programs in Biomedicine 260 (2025) 108560

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine



Transferable automatic hematological cell classification: Overcoming data limitations with self-supervised learning

Laura Wenderoth [a,b,c], Anne-Marie Asemissen [d], Franziska Modemann [d], Maximilian Nielsen [a,b,c,1], René Werner [a,b,c,1,*]

[a] Institute for Applied Medical Informatics, University Medical Center Hamburg-Eppendorf, Christoph-Probst-Weg 1, 20251 Hamburg, Germany
[b] Institute of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
[c] Center for Biomedical Artificial Intelligence (bAIome), University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
[d] II. Department of Medicine, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

ARTICLE INFO

Keywords:
Self-supervised learning
Domain adaption
Domain transfer
Leukemia
Cell classification
Label efficiency

ABSTRACT

Background and Objective: Classification of peripheral blood and bone marrow cells is critical in the diagnosis and monitoring of hematological disorders. The development of robust and reliable automatic classification systems is hampered by data scarcity and limited model generalizability across laboratories. The present study proposes the integration of self-supervised learning (SSL) into cell classification pipelines to address these challenges.
Methods: The experiments are based on four public hematological single cell image datasets: one bone marrow and three peripheral blood datasets. The cell classification pipeline consists of two parts: (1) SSL-based image feature extraction without the use of image annotations, and (2) a lightweight machine learning classifier applied to the SSL features and trained on only a small number of annotated images.
Results: Direct transfer of SSL models trained on bone marrow data to peripheral blood data resulted in higher balanced classification accuracy than the transfer of supervised deep learning counterparts for all blood datasets. After adaptation of the lightweight machine learning classifier with 50 labeled samples per class of the new dataset, the SSL pipeline surpasses supervised deep learning classification performance for one dataset and classes with rare or atypical cell types and performs similarly on the other datasets.
Conclusions: The results demonstrate that SSL enables (1) extraction of meaningful cell image features without the use of cell class information; (2) efficient transfer of knowledge between bone marrow and peripheral blood cell domains; and (3) efficient model adaptation to new datasets using only a few labeled data samples.

1. Introduction

Cytomorphological analysis of peripheral blood and bone marrow is essential for diagnosing and classifying hematological disorders [1,2] as well as subtyping and monitoring abnormal cell populations [3]. Central analysis steps are currently performed by laboratory professionals [4]. The process involves visually examining the smear under a microscope, identifying malignant or atypical cells, and manually counting cells of various types to determine their frequency. This procedure is time-consuming [4] and affected by variations between examiners [5,6]. Automating aspects of the process would significantly improve efficiency and reliability. With the advances in artificial intelligence and deep learning (DL)-based image analysis, especially automation of cell detection and classification appear promising [7,8]. However, hematological cells, especially those in bone marrow, comprise >30 distinct subcategories, each playing different roles in various physiological and pathological processes. The complex morphological variations among these cells make it challenging to automate their differentiation and reliably identify rare or atypical cells.

A particular issue of automated cell classification systems is their lack of generalizability and transferability to new laboratories [9,10]. Variations in sample preparation, staining protocols, and imaging conditions significantly affect the classification performance and robustness [11–13]. This is at least partly because current systems rely on

[*] Correspondence author at: Institute for Applied Medical Informatics, University Medical Center Hamburg-Eppendorf, Christoph-Probst-Weg 1, 20251 Hamburg, Germany.
E-mail address: r.werner@uke.de (R. Werner).
[1] M.N. and R.W. contributed equally to this study.

https://doi.org/10.1016/j.cmpb.2024.108560
Received 25 September 2024; Received in revised form 1 December 2024; Accepted 7 December 2024
Available online 9 December 2024

supervised learning approaches using deep neural networks [14–18]. These algorithms learn from labeled datasets to classify cell images, iteratively adjusting the network parameters to minimize the difference between the network prediction and cell types assigned by the human expert. Due to the large number of network parameters, thousands of cells per class need to be manually labeled by experts, and each combination of different staining standards and changed scanning processes requires a labor-intensive time-consuming generation of a new specific training dataset [9]. In addition, supervised training of cell classification models suffers from the scarcity of annotated datasets that represent rare or atypical cell populations [19–22], although these are often of high clinical relevance.

To address issues presented by inadequate training data, domain adaptation (DA) strategies, also referred to as transfer learning [23], are applied [24]. The idea is to exploit established knowledge to solve new problems [9]. Acevedo et al. [25] used image representations generated by a deep neural network pretrained on ImageNet, a large-scale natural image database, for subsequent blood cell classification by a classical machine learning (ML) classifier. They achieved an average accuracy of 90.5 % across eight classes, but the ML classifier training required the entire dataset of 17,092 labeled blood cell images. For the same dataset, Long et al. [26] reported a classification accuracy of 99.3 % when trained on the entire dataset. When they used only 1000 training data points, the accuracy dropped to 82.2 %. This highlights the significant impact of the number of labeled images to achieve high accuracy in classifying hematological cells using current state-of-the-art DA and DL approaches and underscores the importance of exploring alternative approaches to alleviate the data labeling burden.

One approach that is currently attracting interest in the field of artificial intelligence is the combination of self-supervised learning (SSL) and open datasets [27,28]. SSL focuses on leveraging unlabeled data to learn meaningful image representations. By formulating surrogate tasks that do not require expert annotations, SSL algorithms aim to exploit an intrinsic structure within the unlabeled data. The learned representations can then be input into lightweight ML algorithms to solve the actual task [29]. Successful application of SSL for hematological cell classification in bone marrow datasets has been demonstrated by Nielsen et al. [29]. Trained on a large public dataset of approximately 170,000 cell images from bone marrow smears, image representations were extracted and subsequently used to train a supervised ML classifier with a small set of 100 labeled images per class. Despite the scarcity of training samples, the authors reported promising performance, showcasing the potential of SSL in enhancing cell classification performance with minimal labeled data. Chen et al. [13] further described SSL-based classification of blood cells using a larger (about 18, 000 images) and a smaller (about 1000 images) public blood cell dataset. When they trained their SSL feature extractor on the smaller dataset and fine-tuned the ML classifier using the features computed for a part of the larger dataset, they achieved an average classification accuracy of 91.3 % (8 classes; classifier fitting still based on several thousand labeled images). In contrast, unsupervised feature extraction on the larger dataset and classifier training on the smaller dataset resulted in an average accuracy of only 67.2 % (6 classes).

Given the currently unclear and preliminary results, the present study investigates the ability of SSL for blood cell classification, with a specific focus on the feasibility of transferring knowledge and learned representations of cells in bone marrow smear images to cell classification in peripheral blood smears. To simplify the notation, in the following, the classification of cell images from bone marrow smears is referred to as the classification of bone marrow cell images and the classification of cell images from peripheral blood smears as the classification of blood cell images. The key contributions of the present study are:

- **Cross-domain model transferability**. It will be shown that classifiers trained on bone marrow images using SSL can be directly applied to peripheral blood cell image classification, regardless of staining or digitalization methods used in different laboratories and with superior transferability compared to current state-of-the-art supervised DL.
- **Domain adaption with limited samples per class**. It will be demonstrated that SSL allows for efficient domain adaptation for blood cell classification with only a minimal amount of labeled data.
- **Reproducibility and transparency**. The study emphasizes transparency and reproducibility by using only public datasets and providing the source code and the trained models without limitations to promote rigorous scientific debate and support the continued progress of the field.

2. Methods

2.1. Cell datasets

The study is based on four established, public hematological single cell image datasets: one bone marrow dataset [30] and three peripheral blood datasets [19–21]. Dataset details are listed in Table 1. All datasets provide cropped single cell images and do not require application of cell detection or segmentation algorithms. In the following, the datasets are named after the first authors of the corresponding articles or according to the dataset name given in the publications: Matek bone marrow (BM) dataset; and Matek, Acevedo, and Raabin blood datasets. The Matek BM and blood datasets can be assumed to be produced in the same laboratory. A summary of the cell classes and class distributions for the different datasets is given in Table 2. To illustrate the varying staining conditions, representative sample images for the five cell classes that exist in the datasets are shown in Fig. 1.

2.2. Cell image classification pipeline

The cell image classification pipeline consisted of two independent parts. First, a DL-based image encoder was trained through SSL, only driven by the image data without any annotations, to extract relevant cell features. Second, the obtained low-dimensional image representations were used to train a lightweight ML classifier, using a limited labeled image subset for the classification task at hand. The source code for the experiments and the trained models can be accessed at github.com/IPMI-ICNS-UKE/cell-classification.

2.2.1. Self-supervised extraction of cell image features

The cell image features for subsequent classification were extracted

Table 1
General characteristics of the four hematologic image datasets used in this study.

Domain	Bone marrow	Peripheral blood		
Dataset	Matek [30]	Matek [19]	Acevedo [21]	Raabin [20]
# cell images	171,374	18,365	17,092	17,965
# classes	21	15	8	5
# patients	945	100	N/A	N/A
# smears	N/A	200	N/A	72
Patients with disease included?	Yes	Yes	No	Yes
Stain type	May-Grünwald-Giemsa / Pappenheim	N/A	May-Grünwald-Giemsa	Giemsa
Image size (pixels)	250 × 250	400 × 400	360 × 363	512 × 512
Digitalization	Zeiss Axio Imager Z2 (40 × oil immersion)	M8 digital microscope/ scanner (100 × oil immersion)	CellaVision DM96	Olympus CX18 and Zeiss (100 ×) + smartphone

Table 2
Overview of cell classes and sample numbers involved in haematopoiesis, including myelopoiesis, lymphopoiesis, thrombopoiesis and of other cell types for the used four datasets.

	Bone marrow	Peripheral blood		
	Matek [30]	Matek [19]	Acevedo [21]	Raabin [20]
Erythropoiesis				
Proerythroblast	2740			
Erythroblast	27,395	78	1551	
Myelopoiesis				
Blast	11,973			
Myeloblast		3268		
Fagott cell	47			
Immature granulocyte	21,606	145	2895	
Promyelocyte	11,994	70		
Promyelocyte (bilobed)		18		
Myelocyte	6557	42		
Metamyelocyte	3055	15		
Neutrophil granulocyte	39,392	8593	3329	10,862
Segmented neutrophil	29,424	8484		
Band neutrophil	9968	109		
Basophil	441	79	1218	301
Eosinophil	5883	424	3117	1066
Abnormal eosinophil	8			
Monoblast		26		
Monocyte	4040	1789	1420	795
Lymphopoiesis				
Lymphocyte	26,242	3937	1214	3609
Lymphocyte (atypical)		11		
Lymphocyte (immature)	65			
Plasma cell	7629			
Hairy cell	409			
Thrombopoiesis				
Platelets			2348	
Other cell types				
Smudge cell	42	15		
Artefact	19,630			
Not identifiable	3538			
Other cells	294			
Total	171,374	18,365	17,092	17,965

through a DL image encoder, utilizing the *DINO* algorithm by Caron et al. [31] and the *sparsam* implementation by Nielsen et al. [29]. The feature extraction utilized a student-teacher setup. Both the teacher and student network had the same deep learning architecture, in our case a cross-covariance image transformer (XCiT) architecture. The teacher model weights were calculated as the exponential moving average of the student model weights. During training, different heavily augmented crops of the same cell image were passed through both networks, and the student model was forced to produce a similar image representation to the teacher using temperature-weighted cross-entropy. A peculiarity of the DINO algorithm is the usage of multiple global and local crops (smaller than 25 % of the original image) for the student, while the teacher receives only global crops, see Fig. 2(a). This approach has been shown to generate meaningful image representation for downstream tasks like image classification [31]. The student model parameters were updated with backpropagation using the Adam optimizer. Hyperparameters were chosen as proposed in the original work by Caron et al. [31] and can be found in our github repository.

As a transformer-based encoder architecture, XCiT relies on an intrinsic mechanism known as attention, which dynamically assigns relative importance to different image areas with respect to other regions and the model output. In turn, the attention weights of the last network layer can be used to gain insight into the model by visualizing the impact of an image region on the output, providing some explainability for the otherwise opaque decision-making processes of the model. In the present study, a multi-head attention with eight heads was used, with each head offering an attention map that can be used for model behavior visualization behavior. For technical details, please refer to Dosovitskiy et al. [32].

2.2.2. *Image feature-based cell classification*

After learning to extract the presumably meaningful cell image representations by SSL, the task of cell classification remains. In the present study, three classifiers were fitted: support vector machine (SVM), logistic regression (LR), and k-nearest neighbors (KNN). The default values of the Python package scikit-learn were used as the classifier hyperparameters. Compared to deep learning classifiers, all three approaches usually offer better performance for a small number of samples, while maintaining fast and robust adaptability to new data.

2.3. *Experiments*

Two sets of experiments were conducted: First, the capabilities of SSL cell image representations and subsequent ML classifiers for direct model transfer between datasets and domains were evaluated. The second step involved testing the adaptability and performance of the ML classifiers when trained with only a small number of labeled samples from a new dataset. An overview of the performed experiments is given in Fig. 3. Classification performance was evaluated by balanced accuracy and class-specific sensitivity. All experiments were performed on an NvidiaA40 GPU with 48GB of RAM and an AMD EPYC 7543 with 32 cores and 1 TB of RAM.

2.3.1. *Direct model transfer*

The initial experiments involved training an SSL feature extractor on the Matek BM dataset (Fig. 3, left; only trained once for all experiments, and training took approximately 48 h). Then, the ML classifiers were adapted using cell image representations obtained from the feature extractors using the BM dataset (Fig. 3, middle; approximately 1 s). The number of labeled samples was limited to 250 per class, motivated by the Nielsen et al. [29]. The composed model, an SSL feature extractor and an ML classifier, was applied to all three blood test datasets to assess the SSL generalization capabilities (Fig. 3, right). The same procedure was repeated with the blood Matek dataset as base dataset.

The BM dataset and the three blood datasets have a different number of classes, with some classes missing in some datasets and other classes combined into a superclass. While the classifiers were trained using the classes of the Matek BM dataset (the source domain), during the final evaluation only the output probabilities for the classes of the target dataset were considered. For superclasses like neutrophil granulocytes, which are differentiated into banded and segmented neutrophils in the Matek BM dataset, a blood cell was labeled as a neutrophil granulocyte if the BM classifier assigned the cell to either a banded or a segmented neutrophil. The same applied to immature granulocytes. If a cell was classified as a metamyelocyte, myelocyte or promyelocyte, it was labeled as an immature granulocyte.

For comparison purposes and to provide a supervised benchmark, the same model architecture (XCiT) was trained as an end-to-end DL cell image classification system (loss function: weighted cross entropy; Fig. 2, lower panel; training time approximately 24 h). Model training was based on the same image data that were used for training the SSL

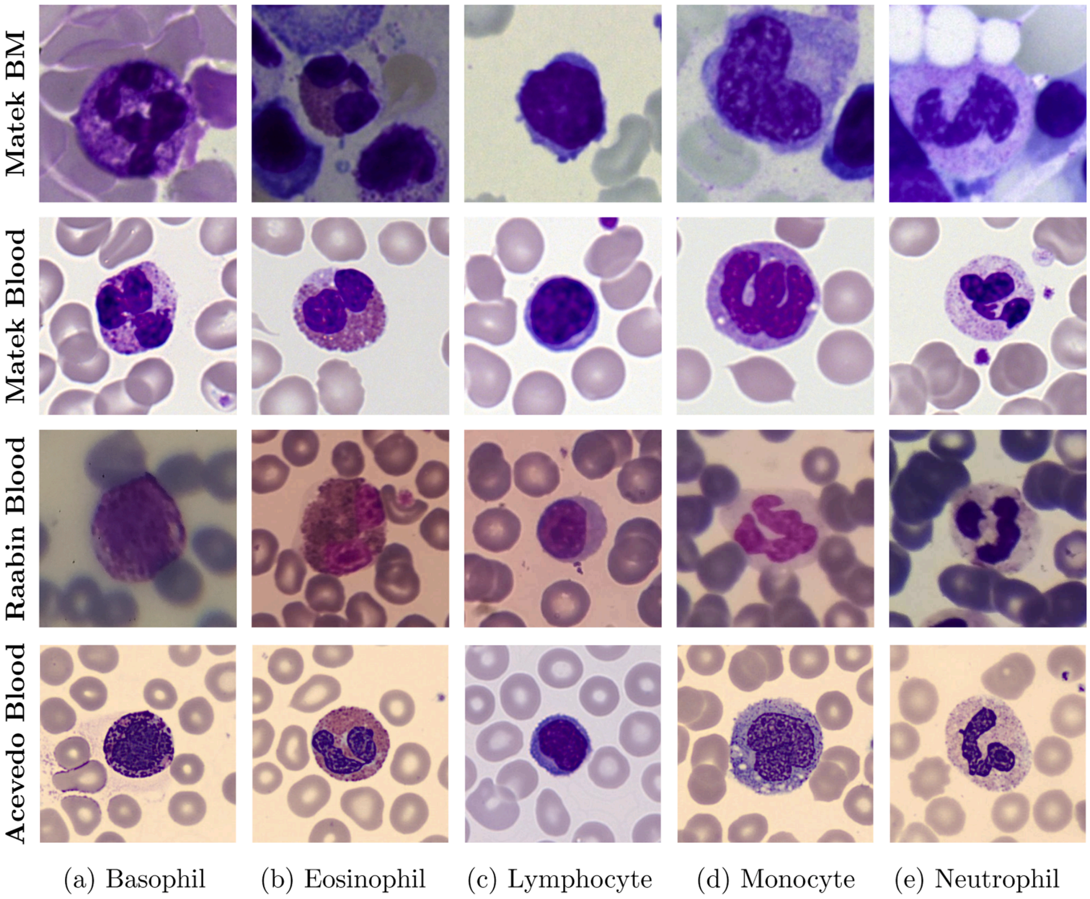


Fig. 1. Representative images of five different cell classes from the four datasets used in this study. From top to bottom: Matek bone marrow (BM), Matek blood, Raabin blood, and Acevedo blood. The images show color variations due to the different staining conditions and magnification, resulting in a different appearance of the cells. In the BM images, various additional cells are often visible around the central cell, whereas in the blood datasets, the central cell is typically surrounded only by erythrocytes.

models. Hyperparameters and data augmentation strategies were consistent with the SSL setting.

To investigate the robustness of the different pipelines, ML classifier fitting was repeated 100 times with random labeled samples of the training dataset (relative class frequencies maintained during sampling). Due to long training times, the training of the supervised baselines was only repeated three times.

2.3.2. *Domain adaption*

The second set of experiments investigated the adaptability of the SSL pipeline to new datasets under the constraint of having only a small number of labelled samples available. Cell image representations were again extracted using the SSL feature extractor trained on the Matek BM dataset. Different from the direct model transfer experiments, the classifier was now fitted with a small labeled subset of [5, 25, 50, 100, 250, 500, 1000, 2000] samples per class from the specific target blood dataset. The classification performance was evaluated as a function of the number of labeled samples. The classifier fitting was repeated 100 times to investigate the robustness of the classification performances.

For the domain adaptation experiments, the benchmark results for supervised DL-based classification were directly extracted from the publications of the datasets and associated articles [20,25,33]. All benchmark models were pretrained on the ImageNet dataset and subsequently trained on a part of the specific cell image dataset. In each case, the training datasets consisted of >10,000 labeled samples.

2.3.3. *Data split and evaluation protocol*

All datasets were randomly split into a training set (70 %) and a test set (30 %) while preserving relative class frequencies. The SSL encoder was trained exclusively on the training set to prevent data leakage. For evaluation of the 100 repeated runs of the SSL experiments, the ML classifier was trained using random stratified subsets of the training set and evaluated on the test set. The supervised DL benchmark model for direct transfer was trained three times on the same training set but with random initial states. Due to the computational cost, the DL experiment was repeated only three times.

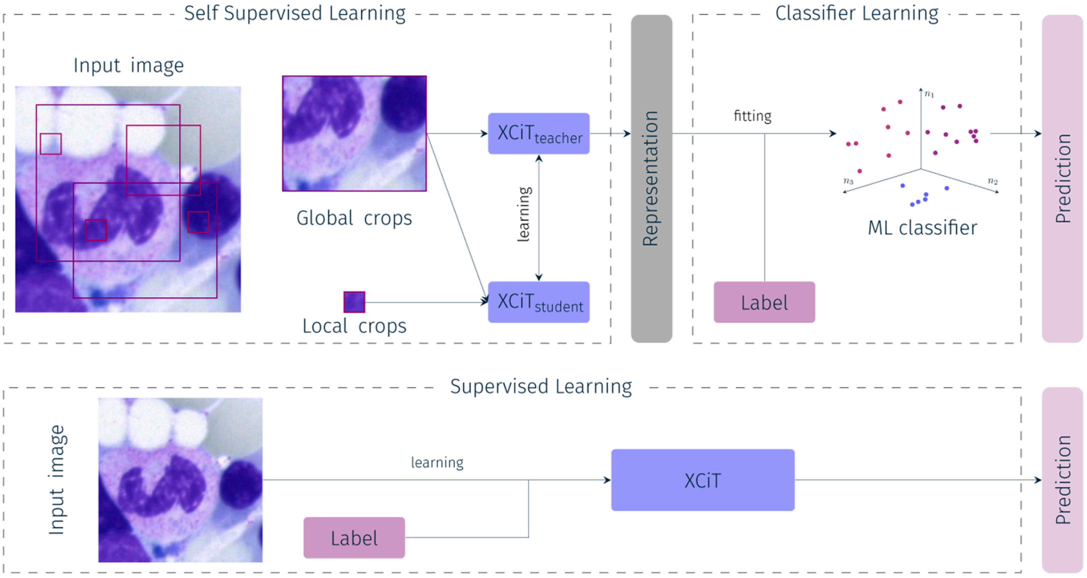


Fig. 2. Comparison of self-supervised learning (SSL, upper part) and supervised deep learning (DL) training processes (lower part). During SSL, random crops of an input image are generated, resulting in global (large) and local (small) crops. Two global crops are then passed through the teacher network (XCiT: cross-covariance image transformer, network architecture used in this study). At the same time, the student network receives the same global and additional five local crops. The learning objective is that student and teacher models extract similar representations for the different crops of the image. During the training process, only the student network is updated by backpropagation, while the teacher network weights are maintained as an exponential moving average of the student weights. The trained teacher network is finally used to extract 384-dimensional feature representations of the images. These representations serve as the basis for the subsequent classification task, which is solved by fitting a lightweight machine learning (ML) classifier based on a limited amount of labeled training data. During SSL, no labels / image annotations are used. In contrast, the supervised DL training process involves the use of a label for each input image. The learning objective is to directly solve the classification task in an end-to-end manner, resulting in a task- and dataset-specific image representation.

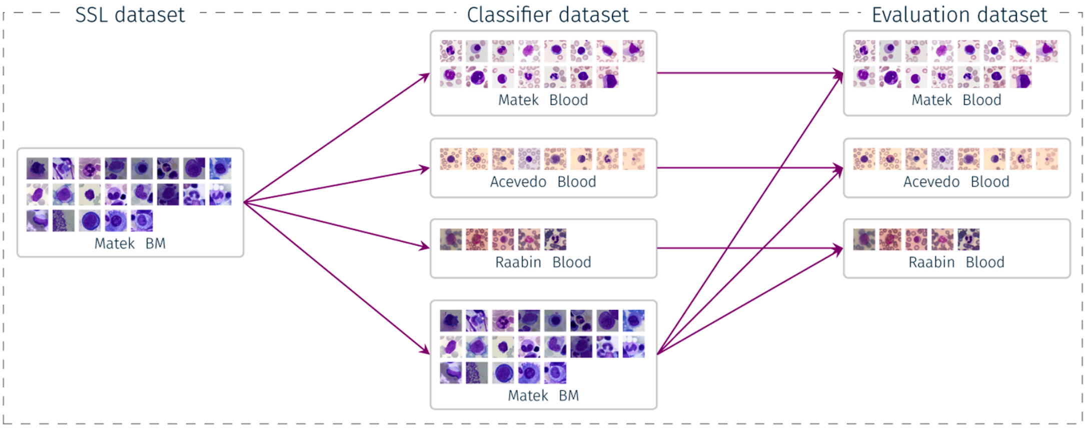


Fig. 3. Overview of the experiments. The SSL feature extractor was trained on the Matek BM dataset. Based on the extracted features, lightweight ML classifiers were fitted. Two types of experiments were conducted: direct model transfer experiments, which were based on the BM dataset, and domain adaptation experiments, which used a limited amount of labeled data from the target dataset. The classification performance was then assessed using the evaluation datasets, specifically the test data of the target dataset.

3. Results

3.1. Direct model transfer

Before the model transfer experiments, the performance of the ML classification approaches (SVM, LR, KNN) was evaluated for the test datasets of the Matek BM and blood datasets. Best classification performance was achieved by the SVM, and subsequent results refer to this configuration.

3.1.1. Transfer of BM models to peripheral blood smear datasets

The results for the transfer of the SSL feature extractor and SVM

Table 3
Cell classification results comparing the SSL feature extractor and the SVM classifier (fitted with 250 labeled samples per class) with a supervised approach, both trained on the Matek bone marrow datasets, for classification of the cell images of the three blood datasets (Matek, Acevedo, and Raabin). The values represent the sensitivity (mean and standard deviation) in percentages for each cell class. If no number is given for a specific class, this class was not included in the Matek bone marrow dataset or the specific peripheral blood dataset.

Dataset	Matek [19]		Acevedo [21]		Raabin [20]	
Learning approach	Ours (SSL)	supervised	Ours (SSL)	supervised	Ours (SSL)	supervised
Erythroblast	93.1 ± 2.2	94.2 ± 2.5	58.4 ± 3.6	73.9 ± 5.1	–	–
Immature granulocyte	80.1 ± 5.1*	74.4 ± 8.9*	92.0 ± 3.0	16.9 ± 7.6	–	–
Promyelocyte	60.0 ± 11.5	11.1 ± 5.9	–	–	–	–
Myelocyte	79.8 ± 11.8	64.1 ± 14.5	–	–	–	–
Metamyelocyte	61.0 ± 15.7	13.3 ± 9.4	–	–	–	–
Neutrophil granulocyte	99.0 ± 0.4**	99.9 ± 0.0**	95.4 ± 1.6	99.9 ± 0.1	99.9 ± 0.1	99.9 ± 0.2
Segmented neutrophil	98.4 ± 0.5	99.7 ± 0.1	–	–	–	–
Band neutrophil	20.0 ± 5.7	10.1 ± 3.8	–	–	–	–
Basophil	34.8 ± 6.0	0.0 ± 0.0	23.4 ± 5.0	1.6 ± 0.2	86.1 ± 4.6	21.0 ± 14.5
Eosinophil	76.7 ± 4.4	38.1 ± 13.6	13.8 ± 5.6	51.9 ± 26.8	7.1 ± 2.8	19.5 ± 4.8
Monocyte	83.5 ± 3.3	30.6 ± 3.3	33.1 ± 6.9	0.0 ± 0.0	34.4 ± 10.5	0.1 ± 0.2
Lymphocyte	78.1 ± 2.7	87.9 ± 2.4	52.6 ± 10.8	77.6 ± 6.7	87.6 ± 3.5	90.2 ± 1.3
Smudge cell	21.0 ± 4.4	0.0 ± 0.0	–	–	–	–
Total	**64.3 ± 2.0**	**40.8 ± 2.9**	**52.7 ± 2.0**	**46.0 ± 4.6**	**63.0 ± 2.7**	**46.1 ± 3.4**

[*] Cells are considered correctly assigned to the superclass *immature granulocytes* if they are assigned to one of the classes: *promyelocyte, myelocyte, metamyelocyte*. The resulting sensitivity for the superclass is not taken into account when computing the balanced accuracy (i.e., the total values).

[**] Cells are considered correctly assigned to the superclass *neutrophil granulocyte* if they are assigned to one of the classes: *segmented neutrophil, band neutrophil*. The resulting sensitivity for the superclass is not taken into account when computing the balanced accuracy.

classifier trained on the Matek BM dataset to the three peripheral blood smear datasets are summarized in Table 3. The table also shows the classification sensitivity for the supervised DL baseline trained on the Matek BM dataset. Applying the SSL BM smear cell image classification pipeline to the Matek blood dataset resulted in a balanced accuracy of 64 % for the 11 classes that were present in both the Matek BM and the blood dataset. The Acevedo blood dataset yielded a balanced accuracy of 53 % (7 classes), and for the Raabin blood dataset, an accuracy of 63 % (5 classes) was achieved. Furthermore, the transfer of the SSL BM models led to a considerably higher balanced accuracy for all three blood datasets than direct transfer of the DL counterpart that was trained in an end-to-end supervised manner. The confusion matrices corresponding to Table 3 and detailed information on the distributions of the sensitivity values of the repeated experiment runs are given in the Appendix, Figs. A1 and A2. This indicates a higher degree of transferability and generalizability of SSL models compared to standard supervised deep learning approaches.

3.1.2. Transfer of peripheral blood smear models to bone marrow images

A transfer of the SSL feature extractor and ML classifiers trained on the Matek blood dataset to the task of bone marrow cell image classification resulted in a low balanced accuracy of around 42 %. In comparison, the supervised baseline model trained on the same Matek blood dataset achieved a balanced accuracy of 20 %. These findings underscore the superior generalizability of the proposed SSL approach. In addition, we employed the blood feature extractor and trained the ML classifier using 50 labeled samples per class from the BM dataset. The accuracy improved to 48 %. The corresponding benchmark article published in conjunction with the BM dataset [34] reports a balanced accuracy of approximately 69 %. The per-class sensitivity values can be

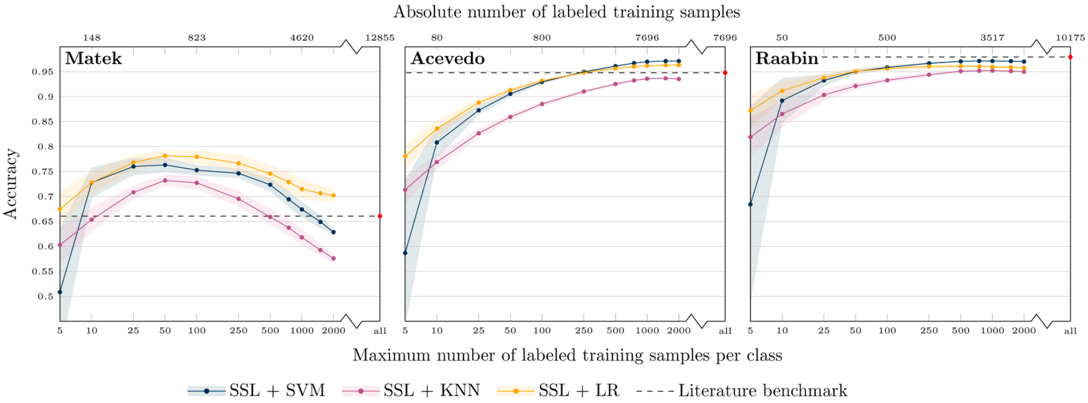


Fig. 4. Balanced accuracy for the domain adaptation experiments as a function of the number of samples per class of the specific blood datasets that were used for the classifier fitting. The data is shown for the three blood datasets and the different ML classifiers (support vector machine, SVM; logistic regression, LR; k-nearest neighbours, KNN). The SSL cell image feature extractor was always the same, trained on the Matek bone marrow dataset. The literature benchmark line refers to the accuracy given in the original dataset-specific publications (supervised end-to-end deep learning cell classification).

found in the Appendix, Table A1. Thus, no advantage was observed from SSL pretraining on blood cell images for bone marrow cell classification, and subsequent experiments focus the transfer of bone marrow smear image representations to the blood datasets.

3.2. Domain adaption

Table 4
Domain adaption: cell classification sensitivity using an SSL feature extractor trained on the Matek bone marrow dataset and a SVM classifier trained with only 50 labeled cell images of the corresponding blood dataset (format: mean and standard deviation in percentages). The results are compared to the sensitivity values reported in the original dataset publications or directly related articles (Table 1 of Matek et al. [33]; Fig. 12 of Acevedo et al. [25], values for the Vgg16-based model; Table 8, values of the ResNext50 run, of Kouzehkanan et al. [20]; the Vgg16-based and the ResNext50 models showed the best balanced accuracy in the respective publications), corresponding to supervised deep learning systems trained with >10,000 labeled samples of the same dataset.

	Matek		Acevedo		Raabin	
	Ours	Ref. [33]	Ours	Ref. [25]	Ours	Ref. [20]
Erythroblast	99.5 ± 1.4	87 ± 9	93.3 ± 2.1	91.9	–	–
Myeloblast	85.8 ± 1.9	94 ± 2	–	–	–	–
Immature granulocytes	*85.4 ± 1.4* *	–	79.2 ± 3.0	92.2	–	–
Promyelocyte	65.6 ± 2.2	54 ± 20	–	–	–	–
Promyelocyte bilobed	100 ± 0.0	41 ± 37	–	–	–	–
Myelocyte	84.9 ± 1.5	43 ± 7	–	–	–	–
Metamyelocyte	15.6 ± 8.3	13 ± 27	–	–	–	–
Neutrophil granulocyte	*94.4 ± 1.1***	–	93.9 ± 1.3	99.6	92.4 ± 2.2	99.6
Segmented neutrophil	85.6 ± 1.7	96 ± 1	–	–	–	–
Band neutrophil	85.8 ± 2.8	59 ± 16	–	–	–	–
Basophil	73.9 ± 4.9	82 ± 7	90.0 ± 2.8	94.3	99.9 ± 0.4	100
Eosinophil	85.1 ± 2.7	95 ± 1	91.4 ± 2.0	99.6	96.1 ± 1.5	98.8
Monoblast	99.1 ± 3.2	58 ± 26	–	–	–	–
Monocyte	82.4 ± 2.0	90 ± 5	86.7 ± 3.3	95.3	90.7 ± 2.4	91.5
Lymphocyte	90.0 ± 2.1	95 ± 2	93.5 ± 2.5	96.8	95.9 ± 1.3	100
Lymphocyte atypical	14.0 ± 20.2	7 ± 13	–	–	–	–
Platelet	–	–	97.0 ± 0.9	99.6	–	–
Smudge cell	80.0 ± 0.0	77 ± 20	–	–	–	–
Total	**76.5 ± 1.6**	**66.1**	**90.6 ± 0.6**	**96.2**	**95.0 ± 0.6**	**98.0**

[*] Cells are considered correctly assigned to the superclass *immature granulocytes* if they are assigned to one of the classes: *promyelocyte, promyelocyte bilobed, myelocyte, metamyelocyte*. The resulting sensitivity for the superclass is not taken into account when computing the balanced accuracy.

[**] Cells are considered correctly assigned to the superclass *neutrophil granulocyte* if they are assigned to one of the classes: *segmented neutrophil, band neutrophil*. The resulting sensitivity for the superclass is not taken into account when computing the balanced accuracy.

The results of training the ML classifier on a limited number of labeled samples from the target domain are given in Fig. 4. The figure shows the balanced accuracy for different sample sizes and classifiers (SVM, LR, KNN) for the three blood datasets. For the Matek blood dataset, LR achieved the highest accuracy and surpassed the performance of the supervised deep learning baseline of the original paper by Matek et al. [33] with only ten samples per class. The accuracy increased up to 78 % with samples per class. With more samples per class, the trend was reversed and the balanced accuracy dropped, which was due to the strong class imbalance within the dataset. Larger classes (>500 samples) benefited from more samples during the classifier fitting, while smaller classes (<100 samples, therefore underrepresented in a larger training set) were assigned less accurately. Similar trends were observed in the Acevedo blood dataset, where LR and SVM performed comparably well. With only 50 samples per class, a balanced accuracy of 91 % was achieved, which increased up to 97 % with 2000 samples per class, surpassing the accuracy of the literature baseline. The classification accuracy did not decrease with even larger sample sizes, as this dataset had roughly balanced class sizes. The Raabin blood dataset followed a similar pattern, with both LR and SVM showing similar performance. With 50 labeled samples per class, a classification accuracy of 95 % was achieved, which increased up to 97 % with 2000 samples per class.

Corresponding quantitative results for the different classes and a minimal training set of 50 labeled cell images per class, that is, a minimum effort setting in terms of data labeling, and the SVM classifier are given in Table 4 and Fig. A2. For the Matek blood dataset, high accuracy for the most common physiological cell types like segmented neutrophils, typical lymphocytes, monocytes, eosinophils, and myeloblasts (crucial for diagnosing acute myeloid leukemia) was achieved by the SSL pipeline (accuracy above 80 %). This is in good agreement with the results of supervised deep learning approaches, which achieve a sensitivity of over 90 % for classes with many training images [33]. Different from the supervised DL approaches, also images of smaller classes were well classified using the SSL pipeline. For instance, the classification accuracy for erythroblasts, monoblasts, and bilobed promyelocytes was above 90 % for the proposed SSL approach, while it was between 41 % and 87 % for the supervised DL model. Some classes, like atypical lymphocytes or metamyelocytes, remained challenging for both approaches, but a noticeable improvement in the accuracy was observed for the SSL pipeline.

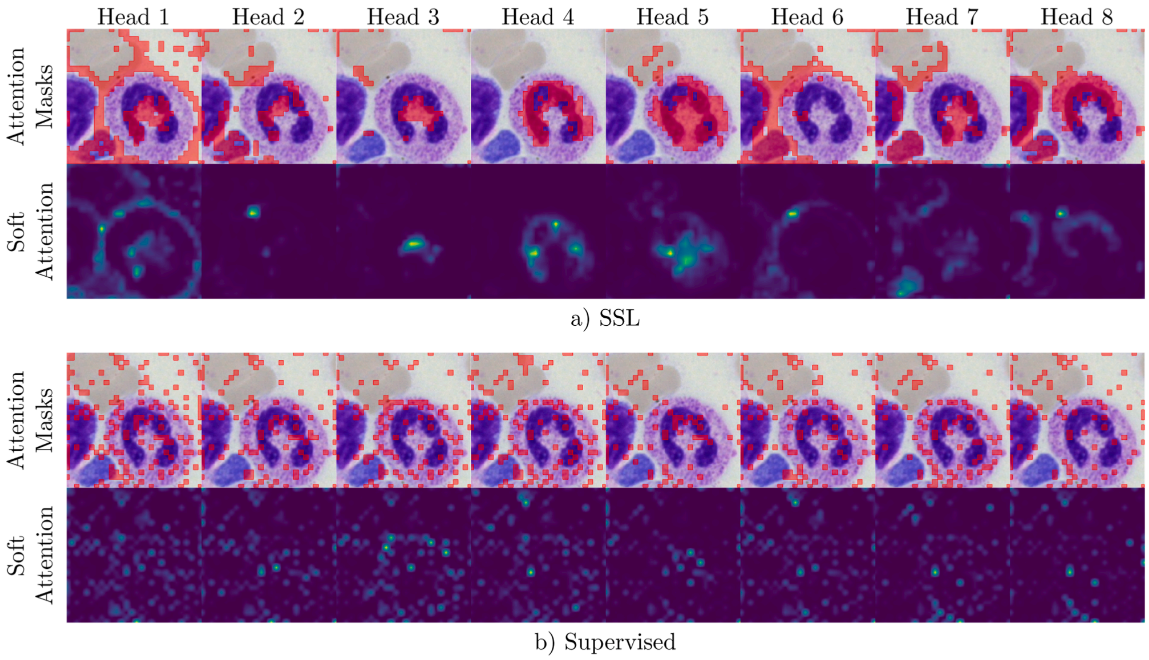
For the Acevedo blood dataset, a similar trend was observed, although the class imbalance was not as pronounced. Basophils, eosinophils, lymphocytes, neutrophil granulocytes, and platelets were all identified with an accuracy of over 90 %. Immature granulocytes proved to be more challenging for the SSL approach with a classification accuracy of 79 %.

For the Raabin dataset, again, high classification accuracy was obtained using the SSL bone marrow image 240 representation and the minimal labeled dataset (accuracy of >90 % for all classes). Although the balanced accuracy of the baseline supervised DL model was not surpassed, Fig. 4 illustrates that additional labeled dataset-specific cell images could improve the SSL pipeline classification accuracy.

4. Discussion

The present study demonstrated the effectiveness of SSL for hematological cell classification and cross-domain adaption with only a minimal labeled dataset. To the best of the authors' knowledge, the work was the first to show that SSL feature extractors trained on bone marrow smear cell images without the use of any class labels provide meaningful image representations to fit lightweight ML models for accurate peripheral blood cell image classification.

For one of the public blood datasets that were used as target domain data, the Matek blood dataset, it was even sufficient to directly transfer the SSL BM classifier to blood image cell classification to achieve state-of-the-art classification results of DL systems trained on the blood dataset-specific image data. Thus, a successful direct model transfer



a) SSL

b) Supervised

Fig. 5. Visualization of the attention information of the eight attention heads of the image transformer model for a representative cell image from the bone marrow dataset. Upper panel: attention for the SSL approach; lower panel: attention for the supervised DL baseline. For each panel, the areas with the highest attention (red; threshold at 80 % of maximum value) are superimposed on the original image. The lower row visualizes the continuous attention values for each head (high values: brighter color). The attention patterns differ notably for the two approaches: The SSL model shows a focused attention on the cell, whereas the supervised approach does not. This observation suggests a potential explanation for superior generalization capabilities of SSL compared to the standard supervised DL.

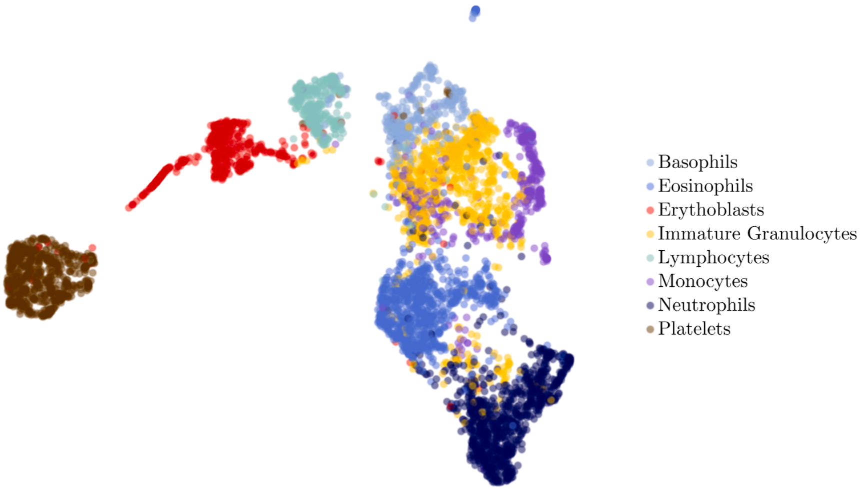


Fig. 6. Two-dimensional embedding of the features of the Acevedo blood cell images. The embedding is computed by a UMAP transform. The features are extracted using the SSL model that was trained on the Matek bone marrow dataset, without information about cell class labels. The visualization shows a clear clustering of cells according to the eight classes of the Acevedo blood dataset, supporting the hypothesis that the SSL approach enables the learning of meaningful image and cell representations.

between the bone marrow smear and the blood smear image domain is possible using SSL.

For the other two peripheral blood smear datasets, direct transfer of the BM SSL models did not achieve classification accuracy close to standard DL systems trained on the images of the specific blood datasets. Nonetheless, a direct transfer of the SSL models outperformed the direct transfer of BM-trained supervised DL counterparts. Thus, the experiments demonstrated the advantage of SSL to improve the generalizability of learned image representations. Fig. 5 supports this hypothesis by displaying the attention head information of the SSL trained (a) and supervised-trained (b) bone marrow transformer models of a representative cell image. The attention head information indicates the image areas on which the models focus. The SSL approach focuses on the cell, whereas the supervised approach does not. This suggests that the SSL approach extracts meaningful cell features, while the supervised approach appears to overfit to the specific training dataset. This may explain the poor transferability of standard supervised DL model representations to new domains.

Fitting the ML classifier with only a small number of dataset-specific labeled samples substantially increased the SSL classification accuracy. In particular, for smaller classes which are often clinically relevant, a higher accuracy was observed than for the literature baseline DL models. These findings further indicate that the SSL-extracted bone marrow smear image representations capture relevant cellular patterns. This is supported by Fig. 6, which shows a two-dimensional embedding of the SSL cell image features of the Acevedo blood test dataset. A clustering of the samples by cell class is apparent, although the representations were obtained by a feature extractor trained on bone marrow cell images without prior knowledge of any classes (neither blood nor bone marrow cell classes) or blood cell images.

The model transfer from the blood to the bone marrow smear image domain did not lead to the same favorable results. This could be due to various factors. First, the Matek blood dataset is considerably smaller, only about one-tenth of the size of the bone marrow dataset. This may have limited the ability of the SSL model to effectively capture representative cell features and patterns from the peripheral blood images. Moreover, the background surrounding the central cell in the images varies between bone marrow and peripheral blood cell smears. Blood images mainly feature erythrocytes, whereas bone marrow images show a high density of nucleated cells. In addition, bone marrow cells exhibit a wider variety of subcategories and greater variability in representation due to differences in staining, preparation methods, and biological characteristics of the cells. This disparity seems problematic for the proposed SSL-DINO approach; the detailed background and variability in appearance in the BM images seem advantageous to learn and extract meaningful features of the cell of interest. However, the balanced classification accuracy reported in the corresponding dataset article [34] is also relatively low, and the exact reason for this remains to be clarified in follow-up studies.

From a clinical perspective, a direct model transfer between laboratories and domains without loss of classification accuracy would be the ideal setting. If not feasible, the time-consuming process of expert-based data annotation should be minimized. The present work showed that domain adaptation using SSL cell image representations achieves high classification accuracy already with a small number of approximately 50 samples per class. This reduces the data labeling efforts by a large amount compared to standard supervised deep learning systems that are trained on >10,000 labeled cell images and up to 2000 labeled images per class. The proposed approach is therefore suitable for transferring it to other datasets. To foster this transfer, all described models and the corresponding source code are provided publicly available, including the trained SSL cell image feature extractors. This means that no specialized hardware is necessary to adapt the proposed approach to the data of interested readers. Using the trained SSL encoder, fitting of conventional ML classifiers by a limited amount of their image data and the corresponding image representations can be efficiently done using a standard computer. In our study, new fitting of the SVM classifier took approximately 1s. This is in contrast to supervised DL models, which typically demand extensive training time on server-grade hardware (in our case approximately 24 h, which is comparable to other publications [19]), rigorous experimental validation to mitigate overfitting, and substantial expertise to design and adapt training pipelines. These requirements make the DL training process less accessible, especially in terms of computational resources and expert knowledge.

In conclusion, the present study demonstrated the capabilities of SSL of cell image representation to improve the classification of hematological cell images. The reuse of learned, meaningful bone marrow image representations enables domain adaptation with only a few labeled cell images of the target dataset, reducing the effort and time of clinical experts to label their image data. This in turn could accelerate the automation of cell classification and thus the diagnosis and monitoring of hematological diseases.

However, the present study built on already cropped single cell images. For clinical application, the pipeline has to be extended to be able to handle whole-slide images, for example by integration of an instance segmentation step, and the resulting pipeline has to be benchmarked against corresponding end-to-end approaches [35]. Additionally, with advancements in tissue sectioning and staining techniques, it will be interesting to explore how these approaches translate to liquid hematological smears, potentially further reducing the need for manual intervention.

CRediT authorship contribution statement

Laura Wenderoth: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing. **Anne-Marie Asemissen:** Conceptualization, Supervision, Validation. **Franziska Modemann:** Validation, Writing – review & editing. **Maximilian Nielsen:** Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization, Visualization, Writing – review & editing. **René Werner:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Acknowledgements

We acknowledge financial support from the Open Access Publication Fund of UKE - Universitätsklinikum Hamburg-Eppendorf. We would also like to thank Thilo Sentker for his great tikz skills and his support in creating the graphical abstract.

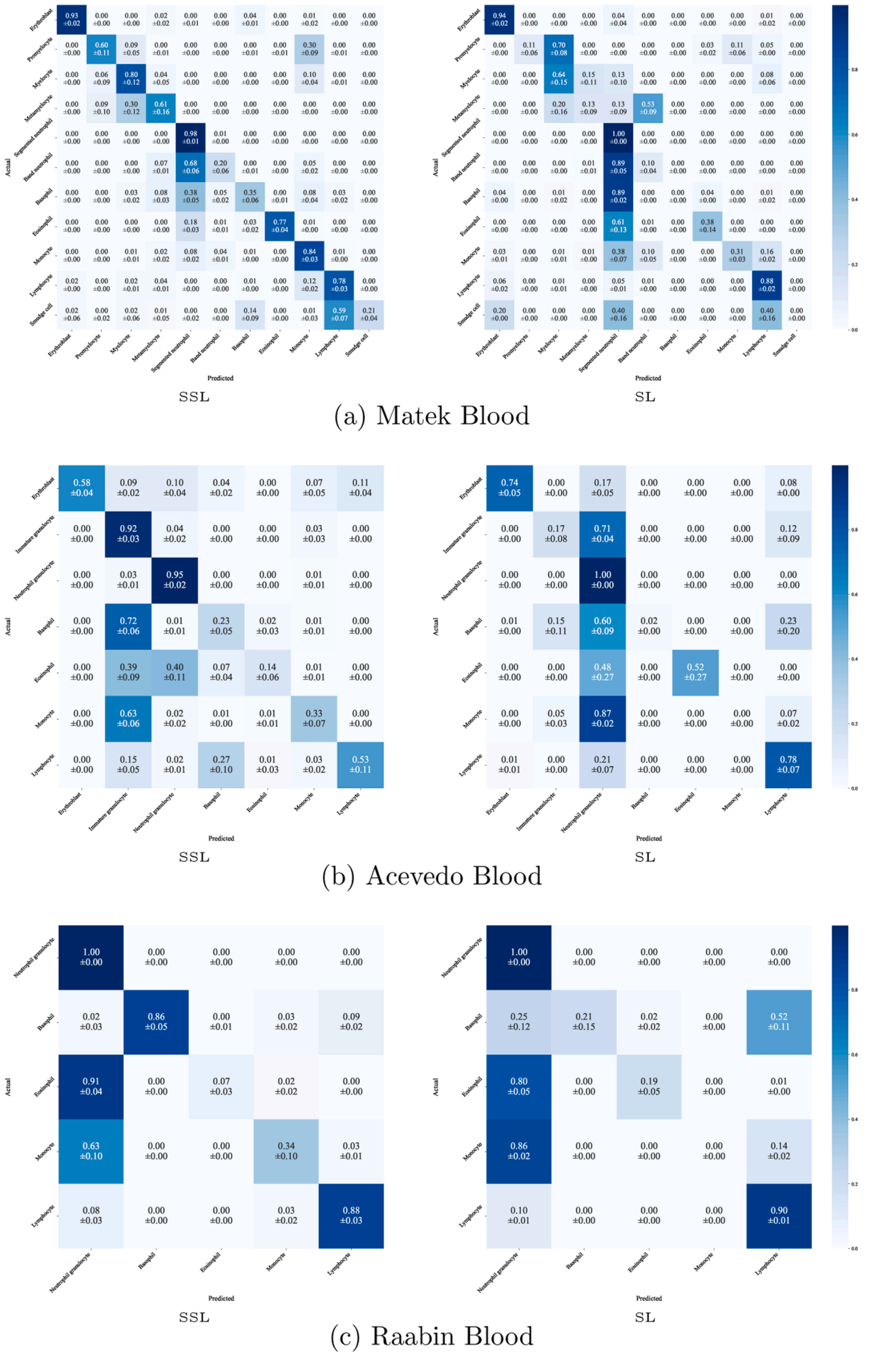
All authors declare that there are no competing financial interests in relation to the work described.

Ethics statement

The experiments described in this article were based on publicly available data published and provided by other working groups and institutions.

Appendix A

To complement the data summarized in Tables 3 and 4, the appendix contains the corresponding confusion matrices (Figs. A1 and A3) and a comparison of the distribution of the sensitivity values for the repeated runs of the experiments belonging to Table 3. In addition, Table A1 summarizes the results of the experiments on the direct transfer of an SSL feature extractor and an SVM classifier trained and fitted on the Matek blood dataset for single cell classification in the Matek BM dataset. The corresponding domain adaptation experiments results are also listed in the table and compared with literature benchmark results.



(a) Matek Blood

(b) Acevedo Blood

(c) Raabin Blood

Fig. A1. Confusion matrices for the results summarized in Table 3. Confusions matrices corresponding to the cell classification results displayed in Table 3, comparing the SSL-based classification feature extractor and the SVM classifier (fitted with 250 labeled samples per class) with a supervised approach (SL), both trained on the Matek bone marrow datasets, for classification of the cell images of the three blood datasets (Matek, Acevedo, and Raabin). The values represent mean relative frequencies, normalized by class size, i.e., the numbers of each row sum up to 1. Please zoom in to read the details.

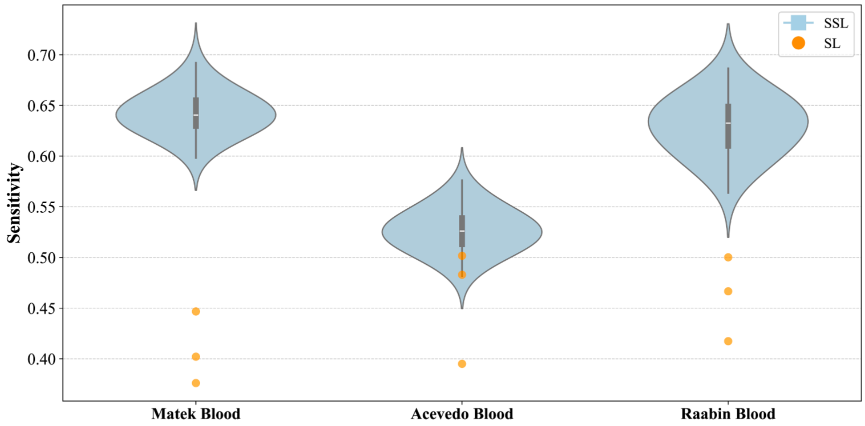
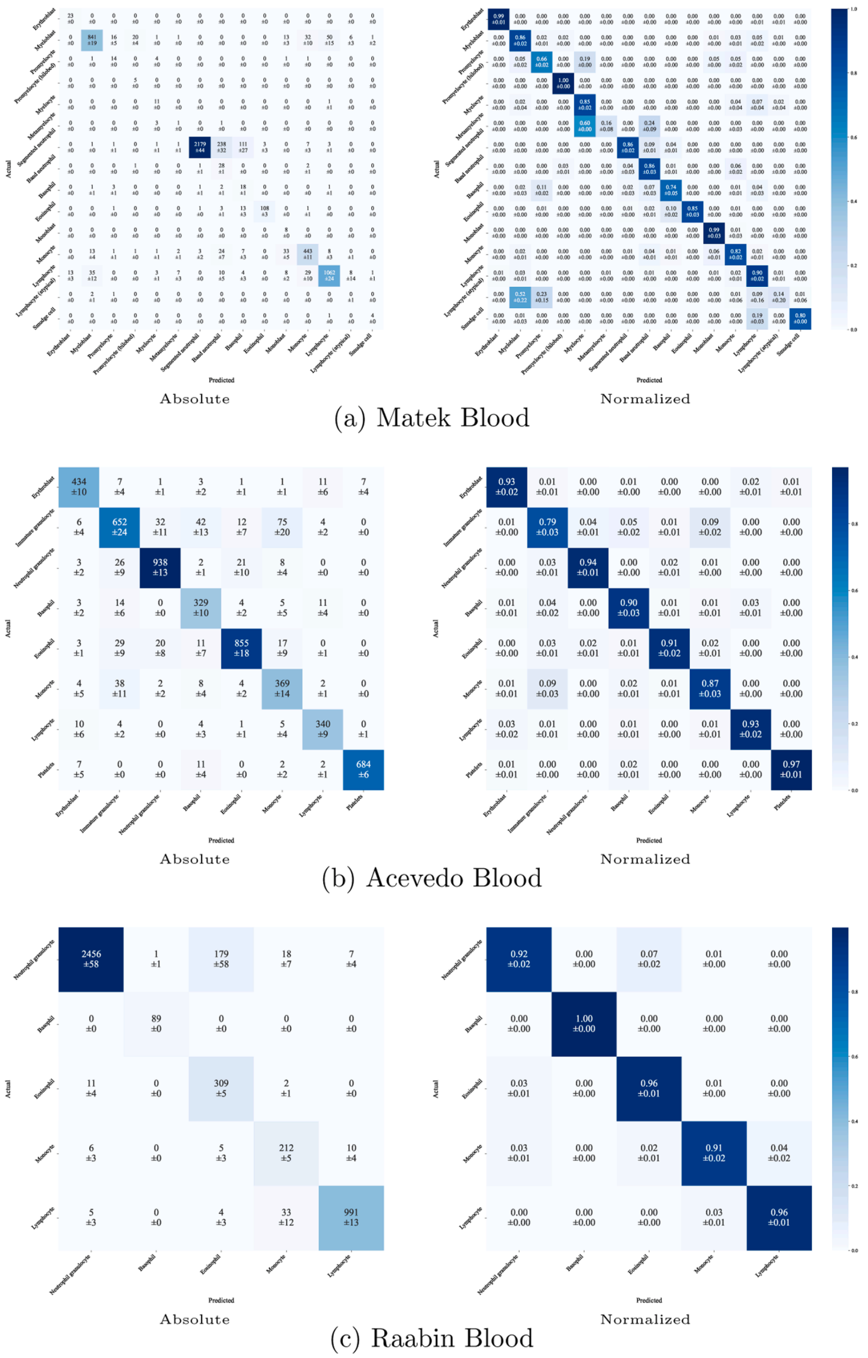


Fig. A2. Distribution of the sensitivity values of the repeated runs belonging to Table 3. The figure shows the distribution of the repeated experiment runs for each of the three blood datasets. The SSL distribution refers to the 100 repeated runs of each experiment performed using the self-supervised learning (SSL) pipeline. The three orange points refer to the three corresponding runs for the supervised deep learning (SL) counterparts.

Table A1
Cell classification results comparing the performance of an SSL feature extractor trained on the Matek blood dataset for classification of single cells of the Matek bone marrow dataset. For direct model transfer, an SVM classifier was fitted using 250 labeled samples per class of the blood dataset and applied for classification of single cell images of the bone marrow dataset. For the domain adaptation experiments, the SVM was fitted with 50 labeled samples of the bone marrow dataset. The literature benchmark values are taken from column Recall$_{\text{strict}}$ of Table 1 of the corresponding article [34]. The values represent the sensitivity (mean and standard deviation) in percentages for each cell class. If no number is given for a specific class, the class was not part of the Matek blood dataset.

Dataset	Matek Bone Marrow [30]		
	Direct transfer	Domain adaption	Ref. [34]
Proerythroblast	–	62.0 ± 3.2	63 ± 13
Erythroblast	48.9 ± 3.6	67.0 ± 2.5	82 ± 1
Blast	–	35.5 ± 2.8	65 ± 3
Fagott cell	–	26.2 ± 5.4	63 ± 27
Promyelocyte	35.1 ± 2.4	46.4 ± 2.8	72 ± 8
Myelocyte	15.6 ± 1.4	38.6 ± 4.5	59 ± 6
Metamyelocyte	0.5 ± 0.1	28.7 ± 3.1	64 ± 8
Segmented neutrophil	76.9 ± 3.5	62.7 ± 2.9	71 ± 5
Band neutrophil	21.2 ± 4.0	50.8 ± 3.5	65 ± 4
Basophil	18.6 ± 1.5	38.8 ± 3.5	64 ± 7
Eosinophil	38.3 ± 2.5	46.5 ± 2.7	91 ± 3
Abnormal eosinophil	-	0 ± 0	20 ± 40
Monocyte	51.1 ± 5.3	42.6 ± 2.6	70 ± 3
Lymphocyte	83.0 ± 3.4	46.5 ± 3.5	70 ± 3
Lymphocyte (immature)	–	62.5 ± 6.2	53 ± 15
Plasma cell	–	42.4 ± 2.5	84 ± 4
Hairy cell	–	64.9 ± 3.0	80 ± 6
Other	–	54.1 ± 3.8	84 ± 6
Smudge cell	76.3 ± 2.4	88.9 ± 4.1	90 ± 10
Artefact	–	57.4 ± 2.1	74 ± 6
Not identifiable	–	49.6 ± 2.9	63 ± 4
Total	**42.3 ± 0.6**	**48.2 ± 0.6**	**68.9**



Absolute Normalized

(a) Matek Blood

Absolute Normalized

(b) Acevedo Blood

Absolute Normalized

(c) Raabin Blood

Fig. A3. Confusion matrices for the results summarized in Table 4. Table 4 contains the results of the domain adaptation experiments for the SSL feature extractor trained on the Matek bone marrow dataset for single cell classification in the three blood datasets. The left column shows the actual cell numbers, illustrating the class imbalance inherent to the datasets. The right column contains the corresponding relative numbers, normalized by the class size (format: mean and standard deviation in percentages). Please zoom in to read the details.

References

[1] J.D. Khoury, E. Solary, O. Abla, et al., The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms, Leukemia 36 (7) (2022) 1703–1719.

[2] R. Alaggio, C. Amador, I. Anagnostopoulos, et al., The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms, Leukemia 36 (7) (2022) 1720–1748.

[3] T.I. George, Malignant or benign leukocytosis, Hematology 2012 (2012) 475–484.

[4] B.J. Bain, Diagnosis from the blood smear, N. Engl. J. Med. 353 (5) (2005) 498–507.

[5] X. Fuentes-Arderiu, D Dot-Bach, Measurement uncertainty in manual differential leukocyte counting, Clin. Chem. Labor. Med. 47 (1) (2009) 112–115.
[6] K. Sasada, N. Yamamoto, H. Masuda, et al., Inter-observer variance and the need for standardization in the morphological classification of myelodysplastic syndrome, Leuk. Res. 69 (2018) 54–59.
[7] N.M. Deshpande, S. Gite, R. Aluvalu, A review of microscopic analysis of blood cells for disease detection with AI perspective, PeerJ. Comp. Sci. 7 (2021) e460.
[8] J. Rodellar, S. Alférez, A. Acevedo, et al., Image processing and machine learning in the morphological analysis of blood cells, Int. J. Lab. Hematol 40 (Suppl 1) (2018) 46–53.
[9] Y.Y. Baydilli, U. Atila, A. Elen, Learn from one data set to classify all – a multi-target domain adaptation approach for white blood cell classification, Comput. Methods Programs Biomed 196 (2020) 105645.
[10] R. Lüllmann-Rauch, E. Asan, Taschenlehrbuch Histologie, 6th ed., Georg Thieme Verlag, Stuttgart, 2019.
[11] S.J. Wagner, C. Matek, S.S. Boushehri, et al., Make deep learning algorithms in computational pathology more reproducible and reusable, Nat. Med. 28 (9) (2022) 1744–1746.
[12] S. Tavakoli, A. Ghaffari, Z.M. Kouzehkanan, R. Hosseini, New segmentation and feature extraction algorithm for classification of white blood cells in peripheral smear images, Sci. Rep 11 (1) (2021) 19428.
[13] X. Chen, G. Zheng, L. Zhou, Z. Li, H. Fan, Deep self-supervised transformation learning for leukocyte classification, J. Biophoton 16 (3) (2023) e202200244.
[14] L. Boldu, 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.
[15] H. Chen, J. Liu, C. Hua, J. Feng, B. Pang, D. Cao, et al., Accurate classification of white blood cells by coupling pre-trained ResNet and DenseNet with SCAM mechanism, BMC Bioinform. 23 (1) (2022) 282.
[16] Q. Wang, S. Bi, M. Sun, Y. Wang, D. Wang, S. Yang, Deep learning approach to peripheral leukocyte recognition, PLoS One 14 (6) (2019) e0218808.
[17] F. Qin, N. Gao, Y. Peng, Z. Wu, S. Shen, A. Grudtsin, Fine-grained leukocyte classification with deep residual learning for microscopic images, Comput. Methods Programs Biomed 162 (2018) 243–252.
[18] M.M. Alam, M.T. Islam, Machine learning approach of automatic identification and counting of blood cells, Healthc. Technol. Lett 6 (4) (2019) 103–108.
[19] Matek C., Schwarz S., Marr C., Spiekermann K. A single-cell morphological dataset of leukocytes from AML patients and non-malignant controls. https://www.cancerimagingarchive.net/collection/aml-cytomorphology_lmu. Accessed 22 July 2024.
[20] Z.M. Kouzehkanan, S. Saghari, S. Tavakoli, P. Rostami, M. Abaszadeh, F. Mirzadeh, et al., A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm, Sci. Rep 12 (1) (2022) 1123.
[21] A. Acevedo, A. Merino, S. Alferez, A. Molina, L. Boldu, J. Rodellar, A dataset of microscopic peripheral blood cell images for development of automatic recognition systems, Data Brief 30 (2020) 105474.
[22] T.A. Elhassan, M.S. Mohd Rahim, M.H. Siti Zaiton, T.T. Swee, T.A. Alhaj, A. Ali, et al., Classification of atypical white blood cells in acute myeloid leukemia using a two stage hybrid model based on deep convolutional autoencoder and deep convolutional neural network, Diagnostics 13 (2) (2023) 196.
[23] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (2016) 9.
[24] H.E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M.E. Maros, T. Ganslandt, Transfer learning for medical image classification: a literature review, BMC Med. Imag. 22 (1) (2022) 69.
[25] A. Acevedo, S. Alferez, A. Merino, L. Puigvi, J. Rodellar, Recognition of peripheral blood cell images using convolutional neural networks, Comput. Methods Programs Biomed 180 (2019) 105020.
[26] F. Long, J.J. Peng, W. Song, X. Xia, Sang J. BloodCaps, A capsule network based model for the multiclassification of human peripheral blood cells, Comput. Methods Programs Biomed 202 (2021) 105972.
[27] R. Krishnan, P. Rajpurkar, E.J. Topol, Self-supervised learning in medicine and healthcare, Nat. Biomed. Eng 6 (12) (2022) 1346–1352.
[28] E. Tiu, E. Talius, P. Patel, C.P. Langlotz, A.Y. Ng, P. Rajpurkar, Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning, Nat. Biomed. Eng. 6 (12) (2022) 1399–1406.
[29] M. Nielsen, L. Wenderoth, T. Sentker, R. Werner, Self-supervision for medical image classification: state-of-the-art performance with 100 labeled training samples per class, Bioengineering 10 (8) (2023) 895.
[30] Matek C., Krappe S., Münzenmayer C., Haferlach T., Marr C. An expert-annotated dataset of bone marrow cytology in hematologic malignancies. https://www.cancerimagingarchive.net/collection/bone-marrow-cytomorphology_mll_helmholtz_fraunhofer. Accessed 22 July 2024.
[31] M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski, et al., Emerging properties in self-supervised vision transformers, in: *IEEE/CVF International Conference on Computer Vision* (ICCV), Montreal, QC, Canada, 2021, pp. 9630–9640.
[32] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An iamge is worth 16x16 words: transformers for image recognition at scale. 2021;arXiv: 2010.11929v2.
[33] C. Matek, S. Schwarz, K. Spiekermann, C. Marr, Human-level recognition of blast cells in acute myeloid leukemia with convolutional neural networks, Nat. Mach. Intell. 1 (2019) 538–544.
[34] C. Matek, S. Krappe, C. Münzenmayer, T. Haferlach, C. Marr, Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set, Blood 138 (20) (2021) 1917–1927.
[35] S.A. Tarimo, M.-A. Jang, E.E. Ngasa, H.B. Shin, H. Shin, Woo J.WBC YOLO-ViT, 2 Way - 2 stage white blood cell detection and classification with a combination of YOLOv5 and vision transformer, Comput. Biol. Med. 169 (2024) 107875.

3 Summary

3.1 English

Accurate classification of peripheral blood and bone marrow cells is crucial for diagnosing haematological disorders. Traditional supervised Artificial Intelligence (AI) methods for blood cell image classification, which are trained on large labelled datasets, have dominated the field, achieving high performance in controlled environments. However, this approach has significant limitations: it depends heavily on extensive manual labelling, often requires specialized hardware for training, and struggles to generalise across different datasets. Although several studies have attempted to address one or two of these challenges, none has fully overcome all three.

To address these challenges, transfer learning presents a promising solution by transferring knowledge from a model trained on a large (annotated) dataset to a smaller target dataset. This method makes efficient use of existing information, enhancing performance with minimal additional annotation. Another effective approach is self-supervised learning (SSL), where algorithms can extract useful information from data without the need for human-made annotations. In this study, I combine SSL with transfer learning to improve blood cell classification, effectively overcoming the three limitations identified earlier.

To further illustrate the approach, SSL-based feature extraction is combined with a lightweight classifier trained on a small number of labelled samples. This strategy allows for effective representation learning with minimal reliance on large labelled datasets. Four datasets are used: one bone marrow and three peripheral blood cell image datasets. The feature extractor is trained using SSL on the bone marrow images. Two experiments are conducted: direct transfer, where classifiers are trained on bone marrow images, and domain adaptation, where classifiers are trained using a limited number of blood cell images. The performance of this pipeline is then compared to traditional SL methods, which require extensive labelled datasets for training.

The results demonstrate that this approach enhances the transferability of blood cell image classification. In direct transfer, the SSL pipeline achieved an accuracy between 53% to 64%, outperforming the supervised models, which achieved between 41% to 46%. In domain adaptation, the ML classifier, trained with approximately 50 labelled images per class, outperformed the supervised models, particularly in classifying rare or atypical cell types. These results highlight the value of combining transfer learning with SSL for knowledge transfer between bone marrow and peripheral blood. This study also tested transfer learning from blood to bone marrow, but the results were not favourable, likely due to differences in dataset size, image background, and domain variability.

In conclusion, transfer learning combined with SSL offers a promising alternative to traditional methods and provides a more efficient and scalable solution for automated blood cell image classification. Future work could focus on extending this approach to whole-slide images to further improve automation in cell image classification and diagnosis.

3.2 Deutsch

Die präzise Klassifikation von peripheren Blut- und Knochenmarkzellen ist von zentraler Bedeutung für die Diagnostik hämatologischer Erkrankungen. Traditionell dominieren überwachte KI-Methoden zur Blutzellklassifikation, die auf großen, annotierten Datensätzen basieren und in kontrollierten Umgebungen eine gute Performanz erzielen. Diese Ansätze weisen jedoch signifikante Einschränkungen auf: Sie sind stark auf die manuelle Annotation großer Datensätze angewiesen, erfordern oft spezialisierte Hardware für das Training und haben Schwierigkeiten, die Generalisierbarkeit auf unterschiedliche Datensätze sicherzustellen. Obwohl zahlreiche Studien versucht haben, einzelne dieser Herausforderungen zu adressieren, konnte bislang keine Methode alle drei Probleme vollständig lösen.

Um diese Herausforderungen zu bewältigen, stellt Transfer Learning eine vielversprechende Lösung dar, bei der Wissen von einem Modell, das auf einem großen (annotierten) Datensatz trainiert wurde, auf einen kleineren Ziel-Datensatz übertragen wird. Diese Methode nutzt somit bestehende Informationen effizient und verbessert die Leistung mit minimalen zusätzlichen Annotationen. Ein weiterer effektiver Ansatz ist das selbstüberwachte Lernen (self-supervised learning - SSL), bei dem Algorithmen nützliche Informationen aus den Daten extrahieren können, ohne dass menschliche Annotationen erforderlich sind. In dieser Thesis kombiniere ich SSL mit Transfer Learning, um die Blutzellklassifikation zu verbessern und so die drei oben identifizierten Einschränkungen effektiv zu überwinden.

Die SSL-basierte Merkmalsextraktion wird mit einem Klassifikator kombiniert, der auf einer kleinen Anzahl von annotierten Bildern trainiert wird. Diese Strategie ermöglicht ein effektives Repräsentationslernen mit minimaler Abhängigkeit von großen annotierten Datensätzen. Es werden vier Datensätze verwendet: ein Knochenmark-Datensatz und drei Datensätze für periphere Blutzellbild-Datensätze. Der Merkmalsextraktor wird mithilfe von SSL auf den Knochenmarkbildern trainiert. Zwei Experimente werden durchgeführt: direkter Transfer, bei dem Klassifikatoren auf Knochenmarkbildern trainiert werden, und Domänenadaptation, bei der Klassifikatoren unter Verwendung einer begrenzten Anzahl von Blutbildbildern trainiert werden. Die Performanz dieser Pipeline wird dann mit traditionellen überwachten Lernmethoden (supervised learning - SL) verglichen, die umfangreiche annotierte Datensätze für das Training erfordern.

Die Ergebnisse zeigen, dass mein Ansatz die Übertragbarkeit der Blutzellklassifikation verbessert. Beim direkten Transfer erzielte die SSL-Pipeline eine Genauigkeit von 53% bis 64%, was die SL Modelle übertraf, die eine Genauigkeit von 41% bis 46% erreichten. Bei der Domänenadaptation übertraf der ML-Klassifikator, der mit etwa 50 annotierten Bildern pro Klasse trainiert wurde, die überwachten Modelle, insbesondere bei der Klassifikation seltener oder atypischer Zelltypen. Diese Ergebnisse unterstreichen den Wert der Kombination von Transferlernen mit SSL für den Wissenstransfer zwischen den Domänen Knochenmark und peripherem Blut. Diese Studie testete auch das Transferlernen von Blut zu Knochenmark, aber die Ergebnisse waren nicht günstig, wahrscheinlich aufgrund von Unterschieden in der Datensatzgröße, dem Bildhintergrund und der Domänenvariabilität.

Zusammenfassend lässt sich sagen, dass Transferlernen in Kombination mit SSL eine vielversprechende Alternative zu traditionellen Methoden darstellt. Zukünftige Arbeiten können diesen Ansatz auf Whole-Slide-Bilder erweitern, um die Automatisierung der Zellklassifikation und Diagnose weiter zu verbessern.

4 Bibliography

1. Khoury JD et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. leukemia 2022; 36:1703–19
2. Alaggio R et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. Leukemia 2022; 36:1720–48
3. George TI. Malignant or benign leukocytosis. Hematology 2012 Dec; 2012. DOI: `10.1182/asheducation.V2012.1.475.3798515`
4. Bain BJ. Diagnosis from the Blood Smear. The New England Journal of Medicine 2005. DOI: `DOI:10.1056/NEJMra043442`
5. Fuentes-Arderiu X and Dot-Bach D. Measurement uncertainty in manual differential leukocyte counting. Clinical Chemistry and Laboratory Medicine 2009 Jan. DOI: `10.1515/CCLM.2009.014`. Available from: `https://www.degruyter.com/document/doi/10.1515/CCLM.2009.014/html`
6. Sasada K et al. Inter-observer variance and the need for standardization in the morphological classification of myelodysplastic syndrome. Leukemia Research 2018 Jun. DOI: `10.1016/j.leukres.2018.04.003`
7. Deshpande NM, Gite S, and Aluvalu R. A review of microscopic analysis of blood cells for disease detection with AI perspective. PeerJ Computer Science 2021 Apr. DOI: `10.7717/peerj-cs.460`
8. Image processing and machine learning in the morphological analysis of blood cells. International journal of laboratory hematology 2018 May. DOI: `10.1111/ijlh.12818`
9. Baydilli YY, Atila U, and Elen A. Learn from one data set to classify all – A multi-target domain adaptation approach for white blood cell classification. Computer Methods and Programs in Biomedicine 2020 Nov. DOI: `10.1016/j.cmpb.2020.105645`
10. Lüllmann-Rauch R and Asan E. Taschenlehrbuch Histologie. 6th ed. Stuttgart: Georg Thieme Verlag, 2019. DOI: `10.1055/b-006-163361`. Available from: `https://eref.thieme.de/10.1055/b-006-163361`
11. Wagner SJ et al. Make deep learning algorithms in computational pathology more reproducible and reusable. Nature Medicine 2022 Sep. DOI: `10.1038/s41591-022-01905-0`
12. Tavakoli S, Ghaffari A, Kouzehkanan ZM, and Hosseini R. New segmentation and feature extraction algorithm for classification of white blood cells in peripheral smear images. Scientific Reports 2021 Sep. DOI: `10.1038/s41598-021-98599-0`
13. Chen X, Zheng G, Zhou L, Li Z, and Fan H. Deep self-supervised transformation learning for leukocyte classification. Journal of Biophotonics 2023 Mar. DOI: `10.1002/jbio.202200244`. Available from: `https://onlinelibrary.wiley.com/doi/10.1002/jbio.202200244`

14. Nielsen M, Wenderoth L, Sentker T, and Werner R. Self-supervision for medical image classification: state-of-the-art performance with 100 labeled training samples per class. Bioengineering 2023 Apr. DOI: https://doi.org/10.3390/bioengineering10080895. Available from: http://arxiv.org/abs/2304.05163

15. Acevedo A, Alférez S, Merino A, Puigví L, and Rodellar J. Recognition of peripheral blood cell images using convolutional neural networks. Computer Methods and Programs in Biomedicine 2019 Oct. DOI: 10.1016/j.cmpb.2019.105020

16. Acevedo A, Merino A, Alférez S, Molina Á, Boldú L, and Rodellar J. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in Brief 2020 Jun. DOI: 10.1016/j.dib.2020.105474

17. Long F, Peng JJ, Song W, Xia X, and Sang J. BloodCaps: A capsule network based model for the multiclassification of human peripheral blood cells. Computer Methods and Programs in Biomedicine 2021 Apr. DOI: 10.1016/j.cmpb.2021.105972

18. Elhassan TAM, Rahim MSM, Swee TT, Hashim SZM, and Aljurf M. Feature Extraction of White Blood Cells Using CMYK-Moment Localization and Deep Learning in Acute Myeloid Leukemia Blood Smear Microscopic Images. IEEE Access 2022. DOI: 10.1109/ACCESS.2022.3149637

19. Matek C, Schwarz S, Marr C, and Spiekermann K. A Single-cell Morphological Dataset of Leukocytes from AML Patients and Non-malignant Controls. 2019. DOI: 10.7937/TCIA.2019.36F5O9LD. Available from: wiki.cancerimagingarchive.net/x/fgWkAw

20. Matek C, Krappe S, Münzenmayer C, Haferlach T, and Marr C. An Expert-Annotated Dataset of Bone Marrow Cytology in Hematologic Malignancies. 2021. DOI: 10.7937/TCIA.AXH3-T579

21. Wenderoth L, Asemissen AM, Modemann F, Nielsen M, and Werner R. Transferable automatic hematological cell classification: Overcoming data limitations with self-supervised learning. Computer Methods and Programs in Biomedicine 2025; 260:108560. DOI: 10.1016/j.cmpb.2024.108560. Available from: https://www.sciencedirect.com/science/article/pii/S0169260724005534

22. Kouzehkanan ZM et al. A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm. Scientific Reports 2022 Jan. DOI: 10.1038/s41598-021-04426-x

23. Ward JM, Cherian S, and Linden MA. Hematopoietic and Lymphoid Tissues. *Comparative Anatomy and Histology: A Mouse, Rat, and Human Atlas.* Second Edition. Elsevier, 2018. Chap. 19:365–401. DOI: 10.1016/B978-0-12-802900-8.00019-1. Available from: https://www.sciencedirect.com/science/article/pii/B9780128029008000191

24. Dosovitskiy A et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2021 Jun. arXiv:2010.11929 [cs]. DOI: 10.48550/arXiv.2010.11929. Available from: http://arxiv.org/abs/2010.11929

25. Ali A et al. XCiT: Cross-Covariance Image Transformers. *Advances in Neural Information Processing Systems.* Vol. 34. 2021 :20014–27

26. Vaswani A et al. Attention is All you Need. *Advances in Neural Information Processing Systems*. Ed. by Guyon I et al. Vol. 30. Curran Associates, Inc., 2017. Available from: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

27. Caron M et al. Emerging Properties in Self-Supervised Vision Transformers. Proceedings of the IEEE/CVF international conference on computer vision 2021 May. arXiv:2104.14294 [cs]. DOI: https://doi.org/10.48550/arXiv.2104.142944. Available from: http://arxiv.org/abs/2104.14294

5 List of Abbreviations

AI Artificial Intelligence.

AML Acute Myeloid Leukaemia.

BM Bone Marrow.

CML Chronic Myeloid Leukaemia.

CNN Convolutional Neural Network.

CPU Central Processing Unit.

DINO Distillation with No Labels.

DL Deep Learning.

EMA Exponential Moving Average.

GPU Graphics Processing Unit.

KNN K-Nearest Neighbours.

LR Logistic Regression.

ML Machine Learning.

MLP Multi-Layer Perceptron.

RAM Random-Access Memory.

SL Supervised Learning.

SSL Self-Supervised Learning.

SVM Support Vector Machines.

ViT Vision Transformer.

XCA Cross-Covariance Attention.

XCiT Cross-Covariance Image Transformer.

6 List of Figures

7 List of Tables

8 Declaration of Personal Contribution

Table 8.1: Summary of contributions to the individual tasks of the publication process

Data Collection and Experimental Design	
Identification of relevant datasets	Laura Wenderoth
Conceptualization of the experimental setup	Laura Wenderoth, Maximilian Nielsen, Rene Werner
Implementation	
Adaptation of the SSL DINO approach to the specific problem	Laura Wenderoth, considerably Maximilian Nielsen
Implementation of the experimental SSL setup	Laura Wenderoth
Implementation of the SL setup	Laura Wenderoth
Ensuring reproducibility and code refactoring	Laura Wenderoth
Data analysis and evaluation	Laura Wenderoth
Manuscript Preparation	
Drafting the initial manuscript version	Laura Wenderoth
Finalizing the manuscript	Rene Werner, Laura Wenderoth, Maximilian Nielsen
Proofreading	Maximilian Nielsen, Rene Werner, Laura Wenderoth, Franziska Modemann, Anne-Marie Asemissen
Figures and tables	Laura Wenderoth (assisted by Maximilian Nielsen)
Visual abstract creation	Thilo Senker, Maximilian Nielsen
Letter of rebuttal	Laura Wenderoth, Rene Werner, Maximilian Nielsen

9 Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe, insbesondere ohne entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten, verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Das gilt insbesondere auch für alle Informationen aus Internetquellen.

Soweit beim Verfassen der Dissertation KI-basierte Tools („Chatbots“) verwendet wurden, versichere ich ausdrücklich, den daraus generierten Anteil deutlich kenntlich gemacht zu haben. Die „Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG“ aus September 2023 wurde dabei beachtet.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich damit einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Datum Unterschrift

10 Acknowledgements

I would like to take this opportunity to express my heartfelt gratitude to those who have supported me throughout the course of this work.

First and foremost, I would like to thank Prof. Dr. René Werner for his valuable supervision, the provision of a workspace, and his support in integrating me into the Image Processing and Medical Informatics research group. His guidance and support were essential to the completion of this project, and I appreciate the trust he placed in me and the opportunity to undertake this research.

I am deeply grateful to Maximilian Nielsen for his steadfast support with technical challenges and his unwavering belief in this project, even during its most challenging phases. His encouragement and commitment were invaluable and played a key role in the success of this work.

I thank Dr. Anne-Marie Assemissen and Franziska Modemann for their clinical guidance and thoughtful advice throughout this work.

Finally, I wish to sincerely thank my closest friends, my partner and my parents for their steadfast support during moments of crisis and uncertainty. Your encouragement and belief in me made this achievement possible, and I am forever grateful for your presence and support during this journey.