Spatial Maturity Regression for the Classification of Hematopoietic Cells

Philipp Gräbel
Institute of Imaging and Computer Vision
RWTH Aachen University
Aachen, Germany
graebel@lfb.rwth-aachen.de

Julian Thull
Institute of Imaging and Computer Vision
RWTH Aachen University
Aachen, Germany

Martina Crysandt
Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation
University Hospital RWTH Aachen University
Aachen, Germany

Barbara M. Klinkhammer
Institute of Pathology
University Hospital RWTH Aachen University
Aachen, Germany

Peter Boor
Institute of Pathology
University Hospital RWTH Aachen University
Aachen, Germany

Tim H. Brümmendorf
Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation
University Hospital RWTH Aachen University
Aachen, Germany

Dorit Merhof
Institute of Imaging and Computer Vision
RWTH Aachen University
Aachen, Germany

Abstract—In contrast to peripheral blood, cells in bone marrow microscopy images are not only characterized by the cell lineage but also a maturity stage within the lineage. As maturation is a continuous process, the differentiation between various stages falls into the category of (ordinal) regression. In this work, we propose Spatial Maturity Regression – a technique that regularizes the learning process to enforce a sensible positioning of maturity stages in the embedding space. To this end, we propose and evaluate several curve models, target definitions and loss function that incorporate this domain knowledge. We show that the classification F-scores improve up to 2.4 percentage points when enforcing regression targets along learnable curves in the embedding space. This technique further allows visualization of individual predictions by providing the projected position along the learnt curve.

Index Terms—representation learning, cell classification, embedding guides

I. INTRODUCTION

Hematopoietic diseases such as leukemia require a detailed analysis of the distribution of cells for a successful diagnosis. The cell count is typically performed with microscopy images of human bone marrow, as this enables insights into the maturation process, which is affected by many forms of leukemia. Due to the high workload for medical experts, it is desirable to support clinicians with an automated pipeline. An automated pipeline has fewer (time) restrictions on the number of cells evaluated for each cell count. This leads to a considerably higher number of cells, which is beneficial for the subsequent statistical analysis as a basis for diagnosis.

In the bone marrow, cells can form several different cell types, also denoted as cell lineages. Within each lineage, various maturity stages from immature to mature can be observed. Only the mature cells are capable of fulfilling their designated biological tasks. Hence, a deviation from typical ratios of immature to mature cells often indicates diseases

such as leukemia and is, therefore, of major importance for the diagnosis.

Several approaches have been proposed for the classification of hematopoietic cells [1]–[3]. However, while the various maturity stages are often differentiated into specific classes, the process itself is of continuous nature. This continuity as well as the predecessor-successor relationship between cell types of the same lineage should also be considered by the classification algorithm, for example by using (ordinal) regression methods.

Typically, regression is performed by replacing the classification head of a neural network with an estimator for a scalar value. As such, it is a well-established technique. However, the incorporation of multiple lineages is not straight-forward with this architecture. This use-case is better suitable for representation learning techniques, which offer ways to adapt the embedding space directly, for example with the Triplet Margin Loss (TML) [4].

A similar idea has been published by Kwitt et al. [5]. By utilizing known, continuous features and extracting them from trained representations, the authors obtain more suitable representations for classification. In contrast to this work, we regularize the entire learning process by incorporating ordinal information on the maturation process within cell lineages into training. To this end, we define curves in the embedding space that define the cell maturation through regression targets along the curves. The ordinal maturity information is thus transformed into high-dimensional target representations, ordered by their maturity progression along curves.

We propose two linear as well as a polynomial curve model based on Bézier curves. All curve models are partially learnt from the current representations produced by the neural network and partially defined by hand, reducing the available degrees of freedom. We further include an option for the regression targets that incorporates knowledge about the cell frequency in the dataset.

For the loss function, we propose an n-dimensional, projection-based as well as a parametric loss version that works in lower dimensions.

The proposed methods do not only aim at improving classification scores, they also yield better interpretable results. Utilizing projections onto the curve model, individual predictions can be explained in more detail – for example, for identification of samples that are in between two subsequent maturity stages.

II. Image Data

Our hematopoietic cell dataset consists of images obtained from human bone marrow samples. The samples are stained using Pappenheim-staining [6] and digitized with a magnification of $63\times$ using automatic immersion oiling. The resulting whole slide images are analyzed by medical experts, who extract representative regions, similar to the typical diagnostic workflow.

Annotations are obtained in a semi-automatic process by first performing automatic detection using U-Net and Water-

shed as proposed by Gräbel et al. [7]. After manual validation and correction, the respective cell type (class) is assigned to each individual cell by medical experts. This constitutes the ground truth. In this work, we focus on two lineages that show maturity regression: blasts plus neutrophilic cells (promyelocyte, myelocyte, metamyelocyte, band granulocyte, segmented granulocyte) and cells of the erythropoiesis (proerythroblast, basophilic erythroblast, polychromatic erythroblast, orthochromatic erythroblast). Additionally, we consider basophilic and eosinophilic granulocytes, lymphocytes, promonocytes and monocytes as additional classes that are not optimized with respect to the proposed techniques as they do not show a sufficient maturity progression.

This results in $4\,560$ annotated cell patches of size $224 \times 224\,\mathrm{px}$. Figure 1 shows examples for both cell lineages.

III. Methods

We propose Spatial Maturity Regression (SMR), a regularization technique that performs regression of the maturity stages of individual cell lineages along curves in the embedding space. SMR aims at improving the representations learnt by a neural network by computing sensible cell type representations based on maturity progression information. This requires a definition of curve models, target layouts along the curves as well as appropriate loss functions which will be presented in the next sections.

A. Curve Models and Regression Targets

In this work, we utilize three methods of obtaining a curve model from our representations as shown in Figure 2. Per lineage, this results in one model that is required to define regression targets for maturity stages within this lineage. We describe the relationship between cell types of the same lineage either using linear or polynomial models. Whilst the linear curve models allow for better interpretability of learnt representations, the polynomial model allows for the most degrees of freedom – fitting curve models more closely to the data.

The first method requires pre-defined directions, which are used to position lines of fixed length centered on the mean of the embeddings of the lineage along this direction. This reduces many degrees of freedom, especially because the
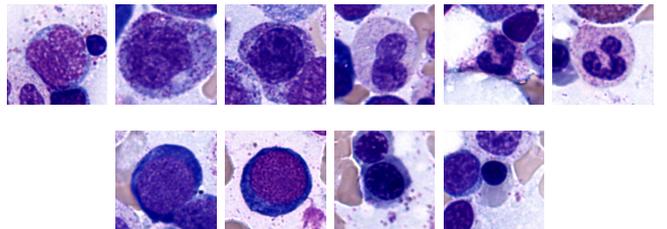
Fig. 1. The top row shows a blast and neutrophilic granulocytes (from left to right): promyelocyte, myelocyte, metamyelocyte, band granulocyte and segmented granulocyte. The second row shows cells of the erythropoiesis: proerythroblast as well as a basophilic, polychromatic and orthochromatic erythroblast.
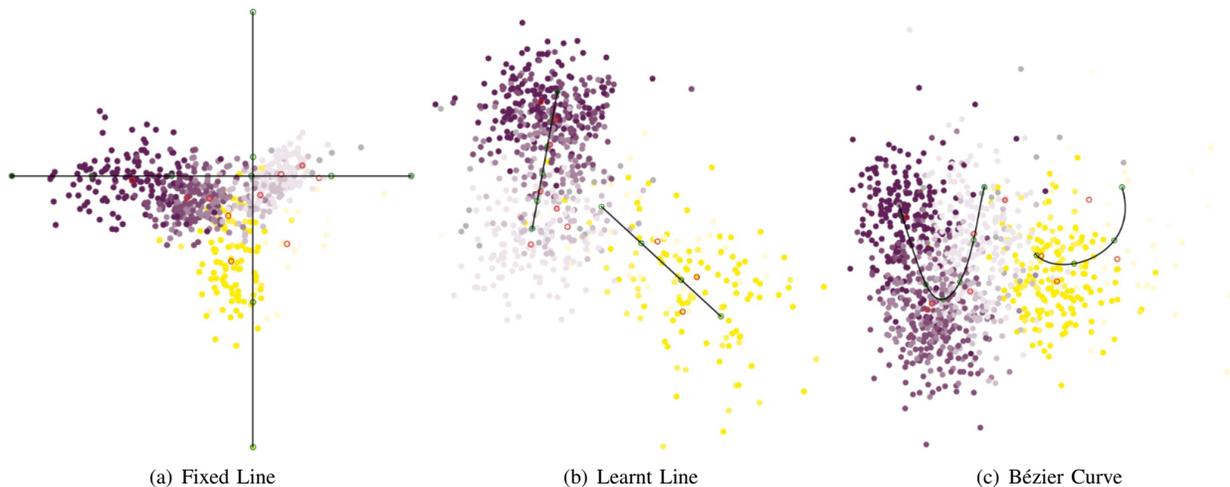
Fig. 2. Three proposed curve models. Points represent points of the embedding for neutrophilic granulocytes (purple) and cells of the erythropoiesis (yellow) ranging from immature (bright) to mature (dark) cells. Black curves represent the line models and the green circle represent the regression targets on the curve. Red circles represent cluster means of the embeddings for each class. Shown for the first two embedding dimensions for visualization.

(a) Fixed Line   (b) Learnt Line   (c) Bézier Curve

relationship between different lineages is unalterable during training. However, this approach yields a partially interpretable embedding.

The second method also utilizes straight lines but removes the direction restriction. Instead, the direction is defined through the vector with the largest singular value after singular value decomposition (SVD). SVD is performed on the centered (mean removed) mean embeddings of the cell types for the given lineage.

The third method is based on polynomial curve fitting to further increase the degrees of freedom. As a polynomial curve model we choose Bézier curves. Based on Bernstein polynomials, the mean cell type embeddings can be interpreted as control points that yield smooth curves. Other curve definitions, such as splines or curves based on Lagrange polynomials, are theoretically possible as well, allowing for curve models of variable curvature.

For each of the three methods, we propose two methods to define the regression targets on the curve. These are defined such that the beginning of the curve is denoted by the value $t = 0$ (most immature) and the ending of the curve by $t = 1$ (most mature).

Firstly, we propose to place the regression targets for each of the C classes in a lineage equidistantly on the curve:

$$t_{\text{uniform}} = \left[\frac{c}{C - 1} \right]. \tag{1}$$

Equal distances assume a maturation process in which each maturity stage has the same relationship to the next stage. However, the continuous nature of the maturation process as well as the resulting uncertainty in the labeling process do in practice not allow for such an assumption. Due to this maturity discretization issue, some cell types account for much larger portions of the maturation process than others.

We try to overcome this problem by additionally proposing the usage of regression targets based on their logarithmic relative frequency in the dataset:

$$t_{\text{frequency}} = \left[\frac{\sum_{i=1}^{c} \frac{1}{2}(\log f_{i-1} + \log f_i)}{\sum_{i=1}^{C-1} \frac{1}{2}(\log f_{i-1} + \log f_i)} \right], \tag{2}$$

with $t_{\text{frequency}} = 0$ for $c = 0$ and f_c the relative sample frequency of the c-th class. This mapping ensures that classes that contain only few samples are assigned a smaller range of target values.

Based on these targets, the appropriate coordinates on the curve model l are defined, such that $l(t_c)$ is the desired target embedding of class c.

B. Loss Computation

The regression targets can be enforced by minimizing the distance between the predicted embedding and the target position for the respective cell type on the curve. In this case, we propose using the L_1 distance in combination with the triplet margin loss (TML). Let B be a batch of predicted embeddings, l the line model and $x_i \in B$ the embeddings with labels y_i and respective layout targets t_{y_i}.

$$\mathcal{L}_{\text{normal}} = \frac{\alpha}{|B|} \sum_{x_i \in B} L_1(x_i, l(t_{y_i})) + (1 - \alpha)\mathcal{L}_{\text{TML}}(B) \tag{3}$$

We furthermore propose to utilize the distance along the curve instead of the full L_1 distance. This way the maximum number of degrees of freedom is retained while still enforcing the ordering defined by the model. This requires a projection $p(x)$ of the predicted embeddings x onto the curve. A root-finding algorithm can be utilized to compute the projections. To simplify the loss-function, we then compute the L_1 distances between projected points and corresponding targets.

$$\mathcal{L}_{\text{projected}} = \frac{\alpha}{|B|} \sum_{x_i \in B} L_1(p(x_i), t_{y_i}) + (1-\alpha)\mathcal{L}_{\text{TML}}(B) \quad (4)$$

In addition, we propose to perform the regression loss computation (as well as the curve model l' and target definition t') in a lower dimensional parametric space. The required dimensionality reduction r can be performed using the parametric version of t-SNE. This allows retaining even more degrees of freedom, as loss computation on parametric t-SNE representations effectively accounts for manipulations in the neighborhood relations only – not supplying strict positions for the embeddings.

$$\mathcal{L}_{\text{dim-red}} = \frac{\alpha}{|B|} \sum_{x_i \in B} L_1(r(x_i), l'(t'_{y_i})) + (1-\alpha)\mathcal{L}_{\text{TML}}(B) \quad (5)$$

Of course, a combination of dimensionality reduced and projected loss are possible.

In a preliminary hyper-parameter analysis on a subset of the data, we found that $\alpha = 0.2$ is a suitable weighting for losses that are not performed on dimensionality reduced embeddings. When using the t-SNE version, $\alpha = 0.6$ is used instead.

C. Experimental Setup

The basic setup is based on a previous analysis of network architectures and hyper-parameters for the given dataset [8]. We utilize DenseNet-121 [9] (pre-trained on ImageNet) as a backbone network to extract embeddings of length 256.

In training, the entire network is fine-tuned based on the proposed losses. The curve models and target definitions are updated at the end of each epoch based on the newly predicted embeddings.

Based on the embeddings, we obtain classification predictions using a hyper-parameter optimized RBF SVM [10]. The training of the SVM is performed after each epoch using a 5-fold cross-validation split of the training data only. Furthermore, we perform dropout ($p = 0.05$), as well as random crop and rotation as data augmentations. Early stop is applied after 50 epochs without validation F-score improvement. The results are obtained by performing a six-fold cross-validation.

IV. Results

Figure 3 shows the resulting macro F-scores for each proposed method. Furthermore, we show training with the Triplet Margin Loss (TML) as a baseline.

Performance increases compared to this baseline can be observed for several models, most notably fixed and linear curves with frequency-based targets for the combination of projected and dimensionality reduced loss, linear curves in general for the normal loss, and linear curves (both targets) and fixed curve (frequency-based) projected loss.

The largest improvement is obtained by utilizing uniform targets on a linear curve with the normal loss, which shows an F-score increase from 0.689 to 0.713.

V. Discussion

The results show that for the dimensionality-reduced version, curves with higher complexity perform better. This highlights that a certain degree of freedom is required. However, even the most successful Bézier curve does not lead to an improvement compared to classical TML.

For the projected losses as well as for the normal loss, improvements can be observed. In these cases, Bézier curves usually yield lower results, which indicates too many or unsuitable degrees of freedom. Instead, fixed and linear curves achieve better results.

Generally, a performance increase for the classification of hematopoietic cells when using spatial maturity regression can be observed. These results were obtained despite the additional difficulty introduced by optimizing two lineages of ten classes overall and five non-regressive classes simultaneously.

Projections of cell representations onto spatial target models can be observed in Figure 4. We project the baseline TML as well as representations produced using SMR with linear uniform targets and normal training onto learnt linear models for both lineages. We additionally plot the cell images whose representations lie closest to the cell type positions on the model for the SMR training case.

Figure 4 shows that the representations produced using SMR are better sorted and show lower deviations along the spatial model dimension for many cell types. The segmented granulocytes for example overlap nearly completely with the band granulocytes when trained with the TML. However, when using SMR many segmented granulocytes are positioned further to the right – resulting in a better sorting. For the metamyelocytes as well as the myelocytes it can be noticed that the representations experience a smaller spread along the model for SMR training. In case of the erythropoiesis very similar observations can be made, as for example the basophilic erythroblasts encounter less overlap with the polychromatic erythroblasts using SMR.

Regarding the retrieved cell images, it can be observed that the cell lineage differentiation can be visually traced along the models for both lineages – supplying us with a mean to interpret the produced representations in the context of cell maturation.

Compared to simple classification algorithms, spatial maturity regression allows a more detailed prediction: by utilizing the curve models in the prediction step, we can obtain a regression estimate. This allows, for example, the identification of samples that lie at the border of two subsequent classes in the continuous maturation process.

The overall increase of 2.4 percentage points in macro F-score is a considerable improvement of the classification method. Even though other publications [2], [3] report larger absolute results, these often result from dataset simplification and are, therefore, not comparable. In contrast to other state-of-the-art datasets [1]–[3], our image data is more challenging due to several aspects: Firstly, we utilize more data than [1] including more challenging images with respect to variability
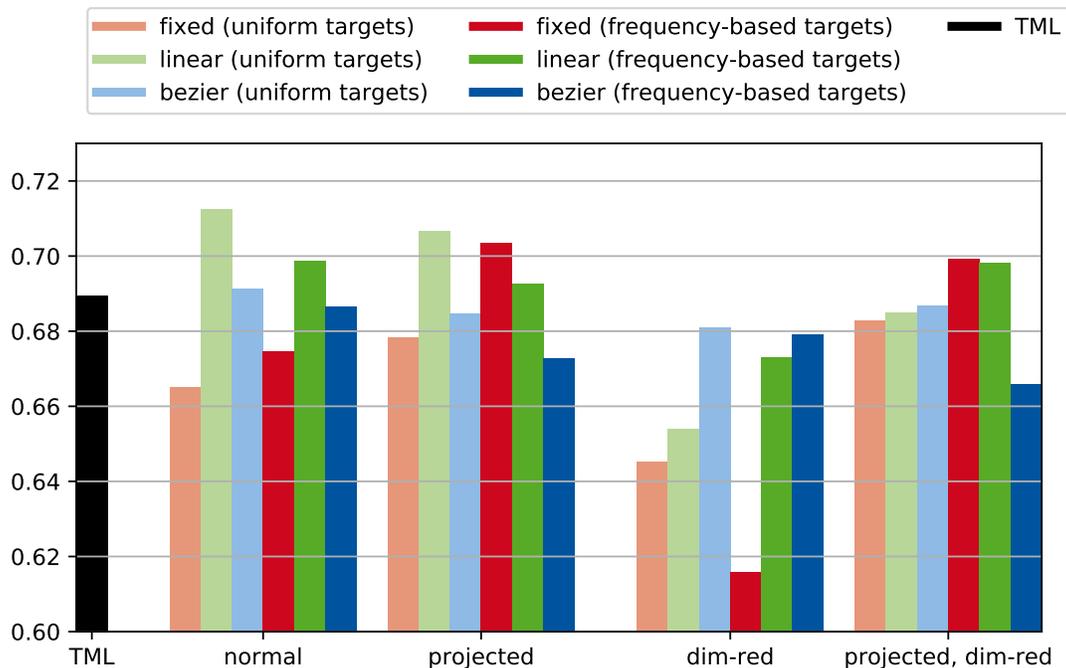
Fig. 3. Resulting macro F-scores for the various proposed methods.
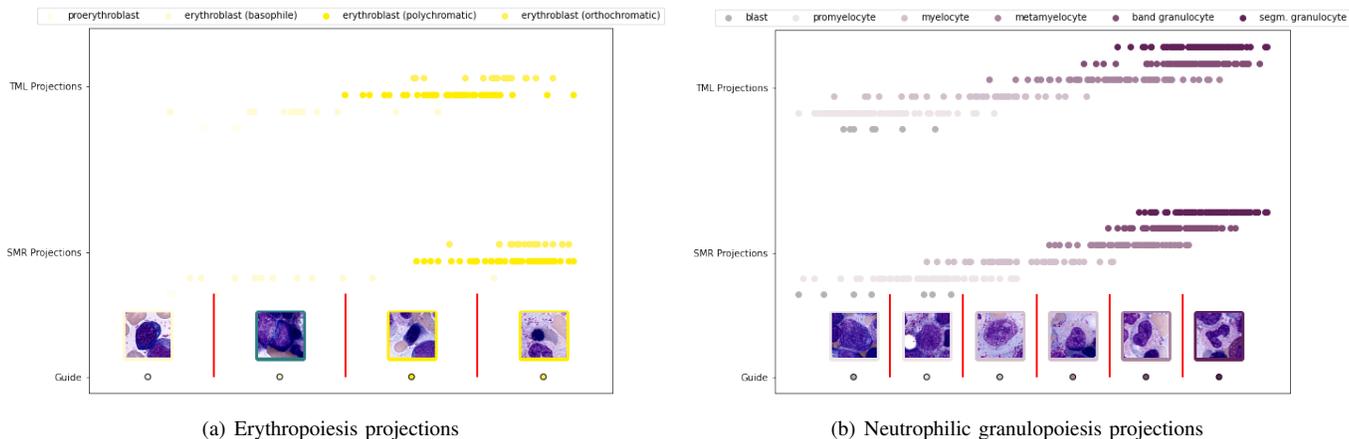
(a) Erythropoiesis projections

(b) Neutrophilic granulopoiesis projections

Fig. 4. Embedding projections onto spatial target models. SMR embeddings were produced using linear uniform targets trained with normal loss without the usage of class-weights. Red lines indicate SMR decision boundaries on the model that separate the cell classes. For each cell type position on the model the cells with the nearest neighboring SMR representations are visualized.

in staining and visual appearance. This also makes the dataset more challenging compared to [2], which is based on selected image regions chosen specifically for their simple visual interpretability. In contrast to [3], we focus on a larger number of classes.

VI. CONCLUSION

In this work, we propose spatial maturity regression – a technique to regularize the learning process of neural networks by performing regression for ordinal maturity information along learnt curves in the embedding space. We investigate three curve models, two regression target mappings and different loss functions. Several combinations of those improve macro F-scores by up to 2.4 percentage points for hematopoietic cell classification. In addition to higher classification scores, the obtained results are better interpretable by visualizing samples as points on the curve.

REFERENCES

[1] P. Gräbel, M. Crysandt, R. Herwartz, M. Hoffmann, B. M. Klinkhammer, P. Boor, T. H. Brümmendorf, and D. Merhof, "Evaluating out-of-the-box methods for the classification of hematopoietic cells in images of stained bone marrow," in *Computational Pathology and Ophthalmic Medical Image Analysis*. Springer, 2018, pp. 78–85.

[2] R. Chandradevan, A. A. Aljudi, B. R. Drumheller, N. Kunananthaseelan, M. Amgad, D. A. Gutman, L. A. D. Cooper, and D. L. Jaye, "Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells," *Laboratory Investigation*, vol. 100, no. 1, pp. 98–109, Sep. 2019. [Online]. Available: https://doi.org/10.1038/s41374-019-0325-7

[3] J. W. Choi, Y. Ku, B. W. Yoo, J.-A. Kim, D. S. Lee, Y. J. Chai, H.-J. Kong, and H. C. Kim, "White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks," *PLOS ONE*, vol. 12, no. 12, p. e0189259, Dec. 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0189259

[4] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[5] R. Kwitt, S. Hegenbart, and M. Niethammer, "One-shot learning of scene locations via feature trajectory transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 78–86.

[6] T. Binder, H. Diem, R. Fuchs, K. Gutensohn, and T. Nebe, "Pappenheim stain: Description of a hematological standard stain – history, chemistry, procedure, artifacts and problem solutions," *Journal of Laboratory Medicine*, vol. 36, no. 5, pp. 293–309, 2012.

[7] P. Gräbel, Özcan Özkan, M. Crysandt, R. Herwartz, M. Bauman, B. M. Klinkhammer, P. Boor, T. H. Brümmendorf, and D. Merhof, "State of the art cell detection in bone marrow slide images," *Journal of Pathology Informatics*, 2021.

[8] P. Gräbel, G. Nickel, M. Crysandt, R. Herwartz, M. Hoffmann, B. M. Klinkhammer, P. Boor, T. H. Brümmendorf, and D. Merhof, "Systematic analysis and automated search of hyper-parameters for cell classifier training," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2020.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.