# Histopathological Transfer Learning for Acute Lymphoblastic Leukemia Detection

Angelo Genovese*, Mahdi S. Hosseini†, Vincenzo Piuri*, Konstantinos N. Plataniotis‡, Fabio Scotti*
* Department of Computer Science, Università degli Studi di Milano, Italy.
{*angelo.genovese, vincenzo.piuri, fabio.scotti*}*@unimi.it*
† Department of Electrical & Computer Engineering, University of New Brunswick, Fredericton, NB, Canada
*mahdi.hosseini@unb.ca*
‡ Department of Electrical & Computer Engineering, University of Toronto, ON, Canada
*kostas@ece.utoronto.ca*

***Abstract*—The detection of Acute Lymphoblastic (or Lymphocytic) Leukemia (ALL) is being increasingly performed with the help of Computer Aided Diagnosis (CAD) systems based on Deep Learning (DL), that support the pathologists in performing their decision by analyzing the blood samples to determine the presence of lymphoblasts. When using DL, the limited dimensionality of ALL databases favors the use of transfer learning techniques to increase the accuracy in the detection, by considering Convolutional Neural Networks (CNN) pretrained on the general purpose ImageNet database. However, no method in the literature has yet considered the use of CNNs pretrained on histopathology databases to perform transfer learning for ALL detection. In fact, the majority of histopathology databases in the literature has either a small number of samples or limited ground truth labeling possibilities (e.g., only two possible classes), which hinders the effectiveness of training CNNs from scratch. In this paper, we propose the first method based on histopathological transfer learning for ALL detection, that trains a CNN on a histopathology database to classify tissue types, then performs a fine tuning on the ALL database to detect the presence of lymphoblasts. As histopathology database, we consider a multi-label dataset with a significantly higher number of samples and classes with respect to the literature, which enables CNNs to learn general features for histopathology image processing and hence allow to perform a more effective transfer learning, with respect to CNNs pretrained on ImageNet. We evaluate the methodology on a publicly-available ALL database and considering multiple CNNs, with results confirming the validity of our approach.**

***Index Terms*—Deep Learning, CNN, ALL, Transfer Learning, Histopathology**

## I. Introduction

Acute Lymphoblastic (or Lymphocytic) Leukemia (ALL) refers to a disease affecting the blood cells, which spreads rapidly through the body and can result in fatal consequences if left undiagnosed and untreated. To ensure a detection of ALL in a timely manner and consequently the possibilities of recovery, one of the main steps consists in the inspection of peripheral blood samples to analyze white blood cells. The inspection and consequent diagnosis is usually performed by an expert pathologist in a manual way, by analyzing possible malformations in white blood cells. Such malformations are an indication of the presence of lymphoblasts, that occur normally in bone marrow. However, a greater number of lymphoblasts than normal in peripheral blood can be a sign of ALL [1], [2].

This work was supported in part by the EC within the H2020 Program under projects MOSAICrOWN and MARSAL and by the Italian Ministry of Research within the PRIN program under project HOPE. We thank the NVIDIA Corporation for the GPU donated.

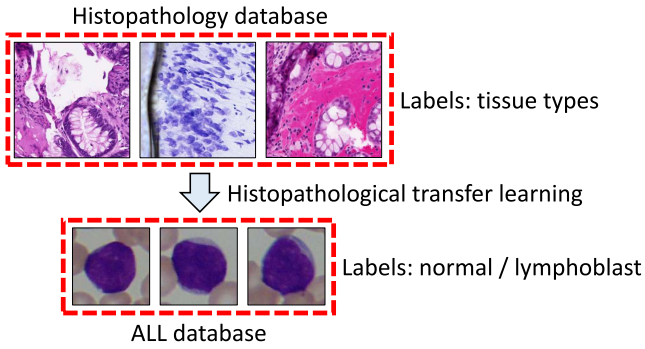


Fig. 1. Histopathological transfer learning: using a CNN trained on histopathological data and fine-tuned for Acute Lymphoblastic Leukemia (ALL) detection.

The manual inspection of blood samples is a time consuming process which leads to fatigue and hence affects the precision of the diagnosis. Therefore, there is a growing interest in developing Computer Aided Diagnosis (CAD) systems that can help pathologists in performing the diagnosis in a faster and more reliable way. In fact, CAD systems partially automate the process by detecting lymphoblasts using image processing and Machine Learning (ML) techniques [3]. In particular, ML methods based on Deep Learning (DL) and Convolutional Neural Networks (CNN) are being increasingly studied because of their high accuracy in several fields, making them the state of the art in numerous application scenarios [4], [5] as well as in several medical fields [6].

Currently, most of DL-based approaches for the detection of ALL are considering learning procedures that increase the classification accuracy [7]–[9], original network architectures suited to processing blood images [10]–[12], or DL-based preprocessing steps [1], with the purpose of detecting lymphoblasts with an increasingly higher precision [3], [13].

In most methods in the literature, the limited dimensionality

of ALL databases favors the use of transfer learning to increase the accuracy of DL-based methods [1], [7], as happens in several medical imaging fields due to the scarcity of medical records, when compared to the high cardinality of samples in general purpose databases (e.g., ImageNet) [14]. However, no method in the literature has considered the use of a histopathological transfer learning to increase the accuracy in detecting ALL. The histopathological transfer learning consists in pretraining the CNN model on a histopathology database (source domain) for tissue type classification and fine-tuning on an ALL database (target domain) for lymphoblast detection. The lack of any histopathological transfer learning in the literature is probably caused by the limitations of most histopathology databases, that include a small number of samples, usually with a limited classification possibility (e.g., only two classes, such as healthy or cancer). Such limitations hinder the effectiveness of training CNNs from scratch using histopathology databases, limiting the comprehensiveness of the learned feature space, and therefore the possibility of performing transfer learning.

Recently, histopathology databases with a significantly higher number of samples and classes and with multi-label classifications have been proposed [15]. Such databases can permit to train CNNs from scratch using histopathology images, learning general features for medical image processing, and enabling to perform a histopathological transfer learning, more effective for classifying blood samples with respect to using CNNs pretrained on ImageNet. By using the histopathological transfer learning, the source domain is more similar to the target domain, with respect to considering general purpose images (Fig. 1). In fact, it has been described in the literature that an increased similarity between source and target domains increases the effectiveness of the transfer learning [16]. Moreover, a CNN pretrained on a histopathology database can be used to detect cancer with limited fine-tuning [17].

In this paper, we propose the first histopathological transfer learning for ALL detection. The method is based on the HistoTNet[1], a CNN pretrained for tissue classification on a histopathology database [15] and then fine-tuned for ALL detection. We evaluate the proposed method on the Acute Lymphoblastic Leukemia Image Database (ALL-IDB)[2] [18], with results showing that our histopathological transfer learning approach increases the detection of accuracy of lymphoblasts over using CNNs pretrained on the ImageNet database.

The paper is structured as follows. Section II reviews the related works. Section III introduces the methodology. Section IV describes the experimental results. Finally, Section V concludes the work.

## II. Related Works

It is possible to divide the methods for ALL detection in three main categories: *i)* handcrafted feature extraction and shallow ML classifiers; *ii)* handcrafted feature extraction and Deep Learning (DL); and *iii)* pure DL.

The methods in the first category apply a handcrafted feature extraction step to process the images, then perform the classification using a shallow ML classifier (e.g., SVM) [19]–[24]. The methods in the second category also apply a handcrafted feature extraction step, but then consider a deep classifier (e.g., CNN) for the classification [25]–[28]. Lastly, the methods in the third category only consider DL techniques, without applying handcrafted feature extraction, therefore exploiting the capability of deep models to automatically learn feature representations [5]. In most cases, pure DL methods represent the most accurate approaches for ALL detection [1], [7]–[12].

It is possible to further divide the pure DL methods based on the approach used to achieve a more accurate classification of the blood sample images as normal or lymphoblast. In particular, it is possible to consider three categories: *i)* more efficient learning procedures; *ii)* original network architectures; *iii)* DL-based preprocessing.

Methods in the first category include the approaches proposed in [7], [8], based on pre-training a CNN on general purpose images and then fine tuning it on the ALL dataset. Similarly, the approach described in [9] uses a CNN pretrained on the ImageNet database, then adopts a feature selection procedure based on swarm optimization to better adapt to blood sample images.

Methods in the second category propose modifications of CNN architectures to better capture the details of blood sample images, such as the technique described in [10], which proposes a convolutional layer that can be added to existing CNNs (e.g., AlexNet [29]), with the purpose of preprocessing the blood sample images and performing a stain deconvolution operation. Similarly, the approach proposed in [12] introduces a variation of the ResNet architecture [30] able to capture both the global and local features of blood samples.

Methods in the third category use DL techniques to preprocess the images. In particular, the approach described in [1] uses CNNs trained using an unsupervised procedure to perform an intelligent tuning of the preprocessing parameters, with the purpose of realizing an adaptive unsharpening of the blood samples.

To the best of our knowledge, no method in the literature has considered the use of a histopathological transfer learning to increase the accuracy in detecting ALL.

## III. Methodology

This section describes our original methodology for histopathological transfer learning. The method is based on training a CNN on a histopathology database for tissue type classification, substituting the last fully-connected layer with a layer configured for detecting lymphoblasts, and fine-tuning the resulting CNN on ALL detection. In the remainder of the paper, we refer to the resulting CNN as $HistoTNet$.

Our method executes the following steps: *A)* histopathology training; *C)* creation of $HistoTNet$; *C)* deep ALL classification. Fig. 2 shows the outline of the methodology.

[1] http://iebil.di.unimi.it/HistoTNet/index.htm
[2] https://homes.di.unimi.it/scotti/all

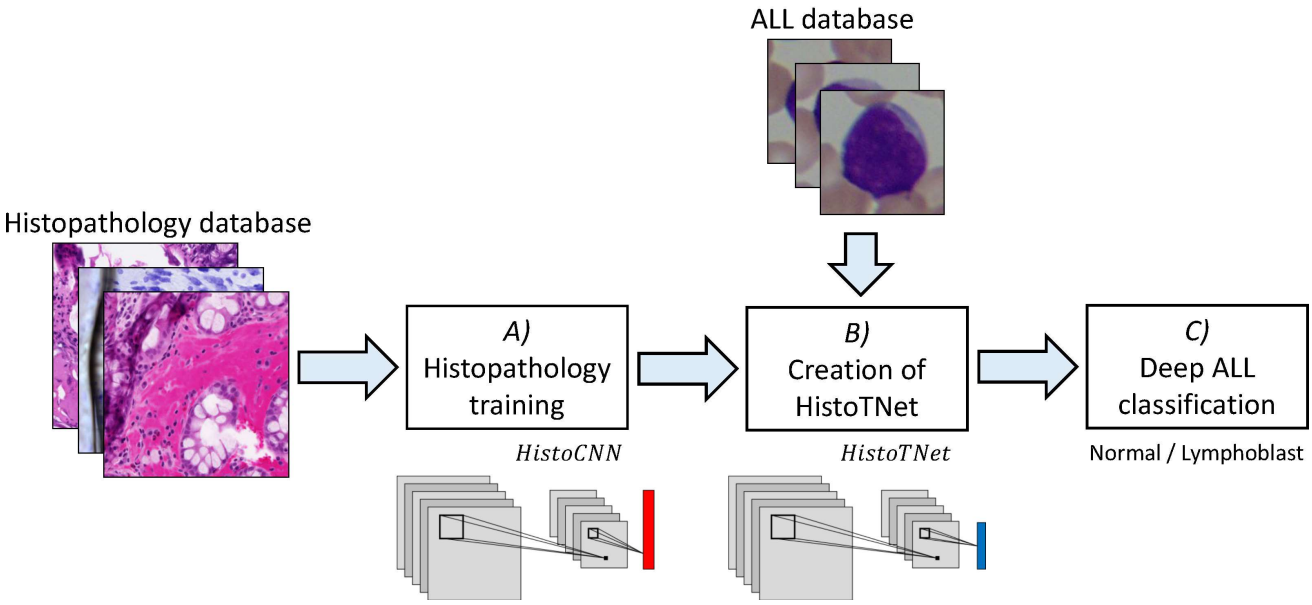


Fig. 2. Outline of the proposed methodology. The method executes the following steps: *A)* histopathology training, in which we train the $HistoCNN$, a CNN for histopathology tissue type classification; *B)* creation of $HistoTNet$, in which we adapt the $HistoCNN$ for binary classification (0: *normal*; 1: *lymphoblast*); *C)* deep ALL classification, in which we apply the $HistoTNet$ on the ALL images to predict the presence of a lymphoblast.

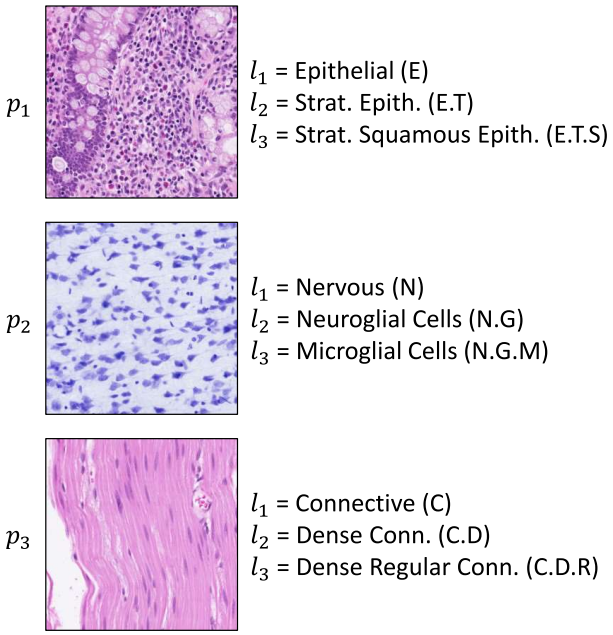


Fig. 3. Examples of patch samples $p$ in the histopathology database, with associated set of labels. For each patch sample, the label $l_{i+1}$ represents a more precise classification than $l_i$.

### A. Histopathology Training

We consider a histopathology database containing Whole Slide Imaging (WSI) samples, collected from diverse histological tissue organs such as epithelial, skeletal, nervous, and adipose. Each WSI sample is split into several patches, each with its own label, to facilitate the processing by the CNN. The patches in the database are multi-label, with each patch that can have multiple labels (e.g., if the patch represents multiple tissue types).

The labels for each patch sample in the database are organized in a hierarchical way, structured in $n$ levels, with each increasing level considering labels with a more precise definition. Each patch sample $p$ has therefore the set of labels $L(p) = \{l_i\}_{i=1}^{n}$ associated, with the label $l_1$ corresponding to the most coarse classification, and the label $l_n$ corresponding to the most precise classification. For example, a patch sample $p$ can have the following labels [15]:

$$\begin{array}{lcl} l_1 & = & \text{Epithelial (E)}; \\ l_2 & = & \text{Stratified Epitelial (E.T)}; \\ l_3 & = & \text{Stratified Squamous Epithelial (E.T.S)} \end{array} , \quad (1)$$

with $l_{i+1}$ having a more precise classification than $l_i$. Fig. 3 shows some example of patch samples, each with the associated set of labels.

To perform the histopathology training, we train a CNN for tissue type classification, for each of the $n$ levels of labels. To train the CNN, we consider the CNN architectures and training procedures described in [15]. As a result, we obtain $n$ different $HistoCNNs$, trained to classify tissue types in the histopathology database, for each level of labels: $\{HistoCNN_i\}_{i=1}^{n}$. The architecture of $HistoCNN$ is build according to models in the literature (e.g., ResNet [30] and VGG16 [31]). These models include a set of convolutional layers, followed by a set of fully-connected layers. The sizing of the last fully-connected layer is performed according to the cardinality of the classes in the label set (e.g., the models pre-trained on the ImageNet database have a last fully-connected layer with 1000 neurons in output, corresponding to the 1000 possible classes in the labels).

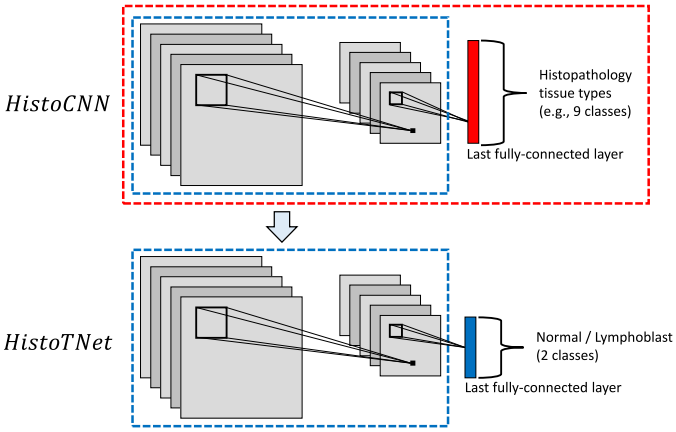


Fig. 4. Example of $HistoTNet$, obtained by substituting the last layer of the HistoCNN with a fully-connected layer with 2 neurons as output.

In this paper, we perform the sizing of the last fully-connected layer for each $HistoCNN$, according to the cardinality of the classes in each level of the labels considered. For example, in the considered histopathology database, the cardinality of the classes in the first level $|\{l_1\}| = 9$, while the cardinality of the classes in the second level, since they express a more precise labeling, is $|\{l_2\}| = 23$.

### B. Creation of HistoTNet

We consider an ALL database composed of images of blood samples, describing either healthy white blood cells or lymphoblasts. Each image in the database has a binary label (0: *normal*; 1: *lymphoblast*) and can therefore be classified in one of 2 classes.

To create the $HistoTNet$, we first adapt each $HistoCNN$ by substituting the last layer with a fully-connected layer configured for binary classification, with 2 neurons as output (Fig. 4). As a result, we obtain $n$ different HistoTNets: $\{HistoTNet_i\}_{i=1}^{n}$.

### C. Deep ALL classification

To train each $HistoTNet$, we perform a fine-tuning on the training subset of the ALL database. To compensate for the limited dimensionality of the samples in the ALL database, we apply data augmentation techniques, by randomly flipping or rotating each image in the training phase.

Lastly, to perform the classification, we apply each $HistoTNet$ on the images in the testing subset of the ALL database. The output of $HistoTNet$ for each image is a binary value indicating the predicted presence of a lymphoblast (0: *normal*; 1: *lymphoblast*).

## IV. Experimental Results

In this section, we describe the used databases, the CNN training, the evaluation procedures and error measures, and the accuracy results of the classification.

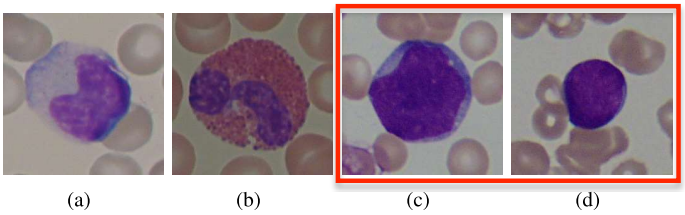


Fig. 5. Examples of blood sample images in the ALL-IDB2 database [18], describing white cells: (a,b) *normal* white cells; (c,d) *lymphoblasts*. The images show only the area around white cells, having been cropped from digital acquisitions of peripheral blood samples.

TABLE I
Overview of the different $HistoCNNs$ obtained by training a CNN on the ADP histopathology database.

| | | CNN Architecture | |
|---|---|---|---|
| | | **ResNet18** | **VGG16** |
| **Level *** | **1** | $HistoCNN_{ResNet18,1}$ | $HistoCNN_{VGG16,1}$ |
| | **2** | $HistoCNN_{ResNet18,2}$ | $HistoCNN_{VGG16,2}$ |
| | **3** | $HistoCNN_{ResNet18,3}$ | $HistoCNN_{VGG16,3}$ |

Notes. * = Level of histopathology labels (see Section III-A).

### A. Used Databases

As the histopathology database, we consider the Atlas of Digital Pathology (ADP) [15][3]. The ADP database contains $17,668$ patches, each labeled according to three levels of labeling (Fig. 3). Each patch can describe a portion of WSI with multiple tissues, hence the labels are not mutually exclusive and each patch can have multiple labels. For example, in the case of the first level of labels $l = 1$, each patch is associated to a label vector of 9 elements, where 9 is the cardinality of the classes in the first level of labels $|\{l_1\}| = 9$.

As the ALL database, we consider the ALL-IDB2 database [18], containing 260 images of peripheral blood samples describing white cells. The images are cropped to show only the region of interest around the cell (Fig. 5). Each image is associated with a binary label (0: *normal*; 1: *lymphoblast*). The images were captured using an optical microscope and a Canon PowerShot G5 camera, with different magnifications of the microscope, ranging from $300x$ to $500x$. All the images have size $H \times W = 256 \times 256$ pixels, with 24 bit color depth.

### B. CNN Training

As the backbone of the $HistoCNNs$, we consider two CNN architectures described in the literature, ResNet18 [30] and VGG16 [31], and for each architecture we train the corresponding $HistoCNN$ for each of the 3 levels of labels. In total, we obtain 6 $HistoCNNs$, shown in Table I. Moreover, to address the possibility of having patches with multiple labels, we append a sigmoid layer after the last fully-connected layer in each CNN, rather than using a softmax layer.

To train the $HistoCNNs$, we split the ADP database using $90\%$ of the samples as training and $10\%$ as validation. We train the $HistoCNNs$ for 100 epochs, using the procedure and

[3] https://www.dsp.utoronto.ca/projects/ADP

TABLE II
OVERVIEW OF THE DIFFERENT $HistoTNets$ OBTAINED BY APPLYING THE PROPOSED METHOD FOR HISTOPATHOLOGICAL TRANSFER LEARNING.

| | | CNN Architecture | |
|---|---|---|---|
| | | **ResNet18** | **VGG16** |
| **Level *** | **1** | $HistoTNet_{ResNet18,1}$ | $HistoTNet_{VGG16,1}$ |
| | **2** | $HistoTNet_{ResNet18,2}$ | $HistoTNet_{VGG16,2}$ |
| | **3** | $HistoTNet_{ResNet18,3}$ | $HistoTNet_{VGG16,3}$ |

Notes. * = Level of histopathology labels (see Section III-A).

learning parameters described in [15]. For each $HistoCNNs$, we select the values of the weights for which we obtain the highest classification accuracy over the validation subset of the ADP database.

To create the $HistoTNet$, we apply the proposed methodology for histopathological transfer learning on each $HistoCNN$, obtaining the corresponding $HistoTNet$, shown in Table II. Then, we split the ALL database as $40\%$ training, $10\%$ validation, and $50\%$ testing. We train each $HistoTNet$ on the training subset of the ALL-IDB2 database using a stochastic gradient descent for $100$ epochs, with batch size $8$, a learning rate $lr = 0.02$, and momentum $m = 0.9$. Every $20$ epochs, the learning rate is halved $lr' = lr/2$. We fine-tune the $HistoTNet$ by using a deep tuning approach, enabling gradient update on all weights of the CNN and not only the last fully-connected layer. For each $HistoTNet$, we consider the values of the weights for which we obtain the highest classification accuracy on the training subset, then apply the $HistoTNet$ on the testing subset of the ALL-IDB2 database to compute the error measures. The training procedure is the same for all the different $HistoTNets$ shown in Table II.

### C. Evaluation Procedures and Error Measures

To compute the error measures, we consider an evaluation procedure based on a $n$-fold cross-validation, with $n = 2$, repeated $10$ times. At each repetition, the training, validation, and testing subsets are extracted randomly from the ALL-IDB2 database.

For each repetition, we apply the proposed methodology for histopathological transfer learning, in particular we perform the steps described in Section III-B to create and train the $HistoTNet$, and the steps described in Section III-C to perform the classification of ALL samples. Lastly, the results are averaged over the $10$ repetitions.

As error measures, we consider the metrics described in [18], which include the mean and standard deviation of the classification accuracy, described as the percentage of samples correctly classified over the total number of samples in the testing subset. Moreover, we consider the confusion matrix, describing the distribution of true positives, true negatives, false positives, and false negatives.

### D. Accuracy Results

Table III shows the accuracy results of the $HistoTNet$ for ALL classification, obtained using our methodology based on histopathological transfer learning. As a comparison, the Table shows also the corresponding results obtained using by performing a fine-tuning on the corresponding CNN pretrained on the ImageNet database, which currently represents the standard procedure in medical imaging [14]. From the Table, it is possible to observe that the proposed method for histopathological transfer learning permits to significantly increase the classification accuracy of ALL samples, with respect to using CNNs pretrained on the ImageNet database. In particular, the $HistoTNet_{ResNet18,2}$ achieves the best accuracy among the considered methods. Table IV shows the average confusion matrix obtained using $HistoTNet_{ResNet18,2}$ on the ALL-IDB2 database.

Using our approach, the increased accuracy over CNNs pretrained on the ImageNet database are, with most probability, due to the fact that the ALL database is significantly more similar to a histopathology database than a general purpose object database such as the ImageNet. In fact, it has been demonstrated that an improved similarity between the source and target domains improves the effectiveness of approaches based on transfer learning [16].

TABLE III
ACCURACY RESULTS ON THE ALL-IDB2 DATABASE USING THE PROPOSED METHODOLOGY BASED ON HISTOPATHOLOGICAL TRANSFER LEARNING.

| **Ref.** | **Deep CNN** | **Classification Accuracy** (%) ($\text{Mean}_{\text{Std}}$) |
|---|---|---|
| [30] | ResNet18 (pretrained on ImageNet) | $88.69_{2.67}$ |
| [31] | VGG16 (pretrained on ImageNet) | $87.54_{3.15}$ |
| - | $HistoTNet_{ResNet18,1}$ | $95.38_{3.41}$ |
| | $\mathbf{HistoTNet_{ResNet18,2}}$ | $\mathbf{97.92_{1.62}}$ |
| | $HistoTNet_{ResNet18,3}$ | $97.38_{1.04}$ |
| | $HistoTNet_{VGG16,1}$ | $97.62_{1.90}$ |
| | $HistoTNet_{VGG16,2}$ | $97.62_{1.31}$ |
| | $HistoTNet_{VGG16,3}$ | $97.62_{1.31}$ |

TABLE IV
AVERAGE CONFUSION MATRIX OF THE $HistoTNet_{ResNet18,2}$ ON THE ALL-IDB2 DATABASE USING THE PROPOSED METHODOLOGY BASED ON HISTOPATHOLOGICAL TRANSFER LEARNING.

| | | **Predicted** | |
|---|---|---|---|
| | | 0 (*normal*) | 1 (*lymphoblast*) |
| **True** | 0 (*normal*) | TN = 49.23% | FP = 0.77% |
| | 1 (*lymphoblast*) | FN = 1.31% | TP = 48.69% |

Notes. TN = True Negatives; TP = True Positives; FN = False Negatives; FP = False Positives.

## V. Conclusion

In this paper we proposed the first approach in the literature based on histopathological transfer learning for Acute Lymphoblastic Leukemia (ALL) detection. The method is based on using transfer learing to improve the cancer detection accuracy in the case of databases with limited dimensionality, and on considering a source domain containing histopathology images, more similar to the target ALL domain with respect to the general purpose ImageNet database.

Our approach is based on pretraining a CNN on a histopathology database, then adapting the CNN for cancer detection and performing a fine-tuning step on the ALL database. The results on a publicly available database designed for ALL detection prove that our method is able to significantly increase the accuracy in detecting lymphoblasts, when compared to CNN pretrained on the ImageNet database.

Future works will consider the use of different databases as the source domain and the application on different databases for cancer detection.

## References


[1] A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, "Acute Lymphoblastic Leukemia detection based on adaptive unsharpening and Deep Learning," in *Proc. of ICASSP*, 2021.

[2] M. M. Amin, S. Kermani, A. Talebi, and M. G. Oghli, "Recognition of acute lymphoblastic leukemia cells in microscopic images using k-means clustering and support vector machine classifier," *J. Medical Signals Sens.*, vol. 5, no. 1, Jan. 2015.

[3] H. T. Salah, I. N. Muhsen, M. E. Salama, T. Owaidah, and S. K. Hashmi, "Machine learning applications in the diagnosis of leukemia: Current trends and future directions," *Int. J. Lab. Hematol.*, vol. 41, no. 6, Dec. 2019.

[4] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, Sep. 2018.

[5] M. Abukmeil, S. Ferrari, A. Genovese, V. Piuri, and F. Scotti, "A survey of unsupervised generative models for exploratory data analysis and representation learning," *ACM Computing Surveys*, 2021.

[6] S. Kulkarni, N. Seneviratne, M. S. Baig, and A. H. A. Khan, "Artificial intelligence in medicine: Where are we now?" *Acad. Radiol.*, vol. 27, no. 1, Jan. 2020, special Issue: Artificial Intelligence.

[7] S. Shafique and S. Tehsin, "Acute Lymphoblastic Leukemia detection and classification of its subtypes using pretrained Deep Convolutional Neural Networks," *Technol. Cancer Res. T.*, vol. 17, Jan. 2018.

[8] A. Rehman, N. Abbas, T. Saba, S. I. u. Rahman, Z. Mehmood, and H. Kolivand, "Classification of Acute Lymphoblastic Leukemia using Deep Learning," *Microsc. Res. and Tech.*, vol. 81, no. 11, Nov. 2018.

[9] A. Talaat, P. Kollmannsberger, and A. Ewees, "Efficient classification of white blood cell leukemia with improved swarm optimization of deep features," *Scientific Reports*, vol. 10, no. 2536, Feb. 2020.

[10] R. Duggal, A. Gupta, R. Gupta, and P. Mallick, "SD-Layer: Stain deconvolutional layer for CNNs in medical microscopic imaging," in *Proc. of MICCAI*, 2017.

[11] J. L. Wang, A. Y. Li, M. Huang, A. K. Ibrahim, H. Zhuang, and A. M. Ali, "Classification of white blood cells with PatternNet-fused ensemble of Convolutional Neural Networks (PECNN)," in *Proc. of ISSPIT*, 2018.

[12] P. Mathur, M. Piplani, R. Sawhney, A. Jindal, and R. R. Shah, "Mixup multi-attention multi-tasking model for early-stage leukemia identification," in *Proc. of ICASSP*, 2020.

[13] M. Alsalem, A. Zaidan, B. Zaidan, M. Hashim, H. Madhloom, N. Azeez, and S. Alsyisuf, "A review of the automated detection and classification of acute leukaemia: Coherent taxonomy, datasets, validation and performance measurements, motivation, open challenges and recommendations," *Computer Methods and Programs in Biomedicine*, vol. 158, May 2018.

[14] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, 2016.

[15] M. S. Hosseini, L. Chan, G. Tse, M. Tang, J. Deng, S. Norouzi, C. Rowsell, K. N. Plataniotis, and S. Damaskinos, "Atlas of Digital Pathology: A generalized hierarchical histological tissue type-annotated database for Deep Learning," in *Proc. of CVPR*, 2019.

[16] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010-2020)," *IEEE Trans. on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2020.

[17] M. S. Hosseini, L. Chan, W. Huang, Y. Wang, D. Hasan, C. Rowsell, S. Damaskinos, and K. N. Plataniotis, "On transferability of histological tissue labels in computational pathology," in *Proc. of ECCV*, 2020.

[18] R. Donida Labati, V. Piuri, and F. Scotti, "ALL-IDB: The Acute Lymphoblastic Leukemia Image Database for image processing," in *Proc. of ICIP*, 2011.

[19] S. Mishra, B. Majhi, P. K. Sa, and L. Sharma, "Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection," *Biomed. Signal Proces.*, vol. 33, Mar. 2017.

[20] J. Rawat, A. Singh, H. S. Bhadauria, J. Virmani, and J. S. Devgun, "Classification of acute lymphoblastic leukaemia using hybrid hierarchical classifiers," *Multimed. Tools Appl.*, vol. 76, no. 18, Sep. 2017.

[21] W. Srisukkham, L. Zhang, S. C. Neoh, S. Todryk, and C. P. Lim, "Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization," *Appl. Soft Comput.*, vol. 56, Jul. 2017.

[22] M. MoradiAmin, A. Memari, N. Samadzadehaghdam, S. Kermani, and A. Talebi, "Computer aided detection and classification of acute lymphoblastic leukemia cell subtypes based on microscopic image analysis," *Microsc. Res. Tech.*, vol. 79, no. 10, Oct. 2016.

[23] S. C. Neoh, W. Srisukkham, L. Zhang, S. Todryk, B. Greystoke, C. P. Lim, M. A. Hossain, and N. Aslam, "An intelligent decision support system for leukaemia diagnosis using microscopic blood images," *Sci. Rep.*, vol. 5, no. 14938, Oct. 2015.

[24] F. Scotti, "Automatic morphological analysis for Acute Leukemia Identification in peripheral blood microscope images," in *Proc. of CIMSA*, 2005.

[25] K. K. Jha and H. S. Dutta, "Mutual information based hybrid model and Deep Learning for Acute Lymphocytic Leukemia detection in single cell blood smear images," *Comput. Methods Programs Biomed.*, vol. 179, Oct. 2019.

[26] J. Laosai and K. Chamnongthai, "Deep-Learning-based Acute Leukemia classification using imaging flow cytometry and morphology," in *Proc. of ISPACS*, 2018.

[27] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," *Med. Biol. Eng. Comput.*, vol. 55, no. 8, Aug. 2017.

[28] M. I. Razzak and S. Naz, "Microscopic blood smear segmentation and classification using deep contour aware CNN and Extreme Machine Learning," in *Proc. of CVPRW*, 2017.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. of NIPS*, 2012.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of CVPR*, 2016.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. of ICLR*, 2015.