Training a Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia

Dimitar Georgiev

196004

Abstract - Acute Lymphoblastic Leukemia is a rare type of blood cancer common in both children and adults, mostly occurring in children between the ages of two and five. It is the most common cause of death from cancer among children. Due to its aggressive nature, it progresses rapidly and results in death within weeks if left untreated. This is why detection and diagnosis of Acute Lymphoblastic Leukemia in an early stage is key for treating and curing it. The process of diagnosis includes taking blood (or bone marrow) samples from a patient, taking microscopic images of white blood cells and performing analysis. Since healthy and malignant cells are very similar, manual examination of the cells may not always produce accurate results and is time-consuming and expensive in general. Using deep learning, this process can be significantly sped up while also producing more accurate results, at lesser cost. In this paper, I present and discuss my attempt to create and train a deep learning classifier for identifying both healthy and malignant cells which indicate Acute Lymphoblastic Leukemia.

I. INTRODUCTION

Computer assisted diagnostic tools play a crucial role in modern medicine. They provide valuable insight and help professionals make conclusive decisions about a patient’s diagnosis. Such tools assist doctors in making an accurate diagnosis on time by using interdisciplinary technologies, usually combining pathological or radiological images with artificial intelligence and computer vision. In this paper, I discuss my attempt to make a computer assisted diagnostic tool using deep learning to detect and classify Leukemia.

Acute Lymphoblastic Leukemia (ALL) is a type of blood and bone marrow cancer characterized by continuous replication of abnormal white blood cells – immature lymphocytes, also called lymphoblasts. While in a healthy organism, the rate of lymphocyte replication is regulated and normal lymphoblasts develop into normal lymphocytes (mature white blood cells that fight infections), ALL disrupts both these processes and causes the lymphoblasts to replicate with a very high rate. The excessive amount of lymphoblasts interferes with the production of new blood cells and their function in general, causing serious harm to different organs and the overall immune system.

Although the general cause is unknown, the underlying mechanism of ALL involves multiple genetic mutations that result in rapid cell division. Some of these mutated genes can be inherited but they can also occur during a person’s lifetime. It is important to note that environmental risk factors are also needed to encourage ALL in addition to the mutations.

Since ALL is a very aggressive disease, it is very important to detect it in an early stage in order for it to be cured. As it progresses rapidly, the treatment is more effective the earlier it starts. After a sample of the blood or bone marrow is taken from the patient it is examined under a microscope. Doctors make microscopic images of the cells in order to further analyze them and identify whether the cells are healthy or malignant. The problem is that morphologically, healthy and malignant cells are very similar and often indistinctive to the human eye. For these reasons manual examinations are unreliable since they are time-consuming, often inaccurate and expensive. The ideal solution for making a fast, cheap and reliable diagnosis is a properly trained classification model which will serve as an automated diagnostic tool. The data and possible solutions will be discussed further in this paper.

II. RELATED WORK

Some of the articles which helped me the most to understand ALL in general, are Acute Lymphoblastic Leukemia on Leukemia & Lymphoma Society [1] and the article with the same name on Wikipedia [2]. I think understanding the problem is very important in order to solve it with deep learning approaches.

Ms. Minal D. Joshi et al. discuss white blood cells segmentation and classification in their paper [3] and propose a Leukemia detection pipeline which involves a k-Nearest Neighbours (kNN) classifier. This method achieves a surprisingly good accuracy of 93%, but is tested on only 108 images. The paper presents an insightful solution to this problem, nonetheless.

Amjad Rehman et al. in their paper [4] describe the process of making a classifier for identification of Leukemia that achieves an even greater accuracy of 97.78%. This paper compares the performances of multiple classifiers such as kNN, Support Vector Machine (SVM) and Naïve Bayes classifiers and a deep learning classifier utilizing convolutional neural networks, that in fact achieves the above-mentioned accuracy.

Niranjana Sampathila et al. also discuss their attempt to train a custom classifier for the same purpose [5]. By also using a deep learning classifier with convolutional layers, they achieved an accuracy of 95.54%. This paper contains useful information about data processing and augmentation techniques for image data.

The starting point for this project was an open competition I found on CodaLab [6] that describes the problem in detail and provides a dataset of a very good quality. The dataset used for this competition can be found through the Cancer Imaging Archive [7]. In this project, I use this exact dataset because of multiple reasons. The most important reason is that the microscopic images of white blood cells are already segmented (the cells are detected and extracted as separate images). The data originates from 118 patients and contains around 15000 microscopic images of good quality. The cells are manually labelled by an expert oncologist and can be counted reliable. The dataset and its preprocessing will be discussed further in more detail.

III. DATA

The dataset consists of 15000 microscopic segmented cell images from 118 patients, each image manually labeled by an oncologist, as mentioned before. The dataset itself is imbalanced, containing more images from lymphoblastic cells than healthy ones. Since imbalanced datasets make the learning curve of deep learning models unsteady and the model is prone to developing a bias, I decided to manually rearrange the dataset and create balanced training and validation sets, at the cost of cutting out surplus images. For a better effect, oversampling techniques are proven to work best in such cases, but since training a deep learning model on a dataset this big takes a very long time, I used the other method. The size of the training set is 4794 cell images (half of which lymphoblastic) and the size of the validation set is 300 images (half lymphoblastic).

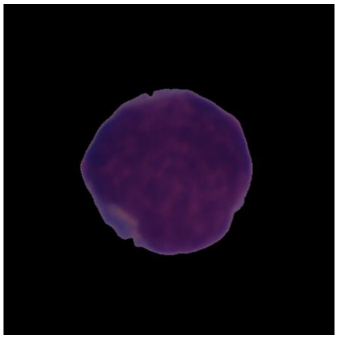


Figure 1: A lymphoblastic cell image

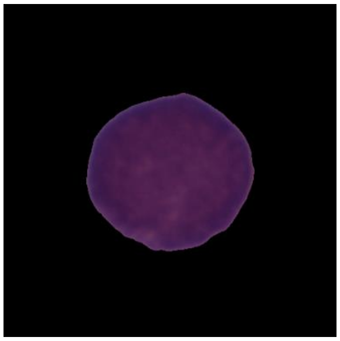


Figure 2: A healthy cell image

Every image in the dataset has a size of 450x450 pixels and contains one cell on a black background (as shown on Figures 1 and 2). Although models can directly be trained on these images, the learning process can usually be improved by preprocessing the dataset and using augmentation during training. Many pretrained computer vision related deep neural models such as EfficientNet or VGG require a certain color normalization and image size (based on the architecture and the dataset they were trained on). Since I use VGG and EfficientNet for feature extraction in this project, the dataset needs to go through transformers that rezise the images to 224x224 pixels and normalize the colors accordingly (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]). To make the learning process faster, I experimented with cropping the pictures to size 300x300 before resizing and normalizing, in order to eliminate the black pixels around the cells (black pixels have a value of 0 for every channel so they deactivate certain neurons in the neural networks). However, this method may crop some parts of bigger cells at the cost of a faster learning process. In terms of data augmentation, I experimented with random rotation applied to images during training, in the interval between -45 to +45 degrees. I do not believe other techniques such as random crops and color augmentation would significantly improve the learning process since every image comes in the same color palette and random cropping may only confuse the model.

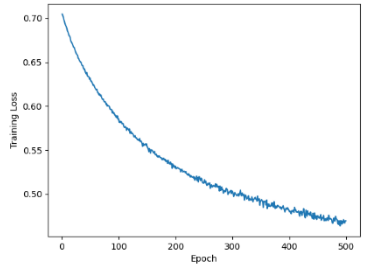
IV. METHODS

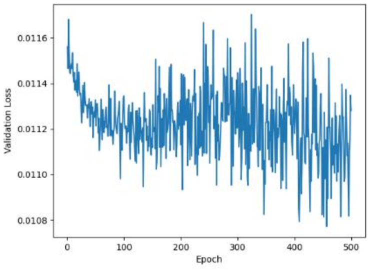
As in many other computer vision tasks, it is important to make an efficient feature extraction from the images before attempting classification. Image feature extraction refers to extracting a vector (tensor) from an image, that meaningfully represents the image itself. Since an image in the form of a three-dimensional matrix has little to no meaning by itself, the feature vector has to capture the meaning and essence of the image. Because training a computer vision feature extractor model is a difficult and time-consuming process, transfer learning is commonly used for this task. In this project I used a pretrained version of EfficientNet (trained on ImageNet, a popular image dataset). I also experimented with using VGG for feature extraction, but sticked to EfficientNet in further attempts. In order to use this pretrained architecture for feature extraction (because the model itself is meant for classification), I removed the last fully connected layer and softmax activation function and instead replaced it with one additional fully connected layer (of size 320) and one final node meant for binary classification. This way, the pretrained weights of EfficientNet are preserved for feature extraction, while the fully connected layers in the end of the models will be trained to classify Acute Lymphoblastic Leukemia. Before training, the weights of the convolutional layers intended for feature extraction need to be frozen, in order not to be altered during backpropagation. I experimented with training this model both with and without augmentation and with different parameters. The results will be further discussed in the next section.

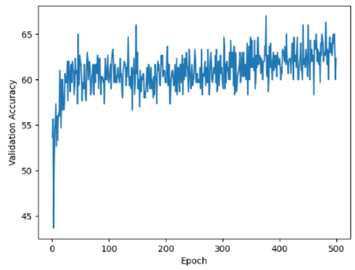
V. EXPERIMENTS & RESULTS

First, I made an attempt at training two different deep learning classifiers, one using VGG and the other using EfficientNet for feature extraction. I trained these two classifiers for 1000 epochs on a very small subsample of the dataset, in order to determine whether the models are capable to learn and compare their learning curves to choose a model to further work with. After this experiment, both models proved capable, with the EfficientNet version having a slightly smoother and more stable learning curve. Based on this, I decided to continue experimenting with this model, using the whole dataset (the balanced version discussed in the Data section). It would have been useful to experiment with more feature extraction models and continue testing more of these models further on, but training time was a major problem that would not allow me to do this in a reasonable time.

Next, I trained the EfficientNet based classifier on the whole balanced dataset for 500 epochs (which took around 3 days to train). In this attempt I did not use image cropping and augmentation. The images were simply preprocessed (resized and normalized) to fit the pretrained feature extraction model.



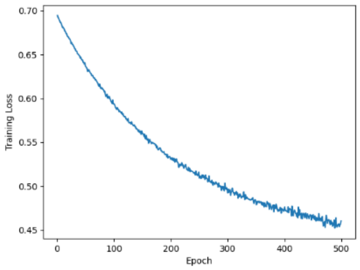


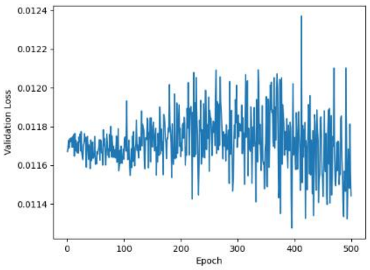


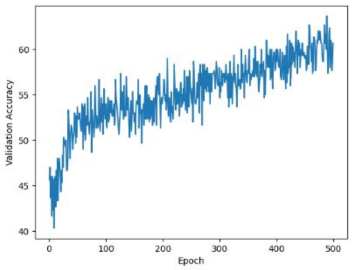
Figures 1, 2, 3: Training loss, validation loss and validation accuracy of the EfficientNet based classifier without using cropping and augmentation

The results of this attempt are visualized on figures 1, 2 and 3. By the end of the 500 epochs, the model is still learning, so it would benefit if the number of epochs is increased. Although the training loss decreases very smoothly over the epochs, the validation loss is a bit more noisy and unstable. However, it can be seen that the accuracy increases during the training process (although with a small rate of change). The maximum accuracy that was achieved with this attempt is 67%.

In my second attempt, I trained the same model by using both cropping and random rotation augmentation on the microscopic images. To best compare the performance of this model and the previous, I also trained this model for 500 epochs. The results are shown in figures 4, 5 and 6.

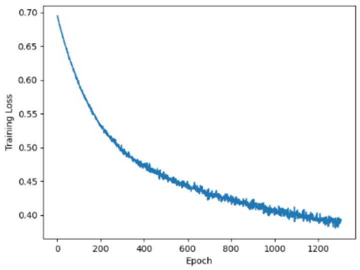


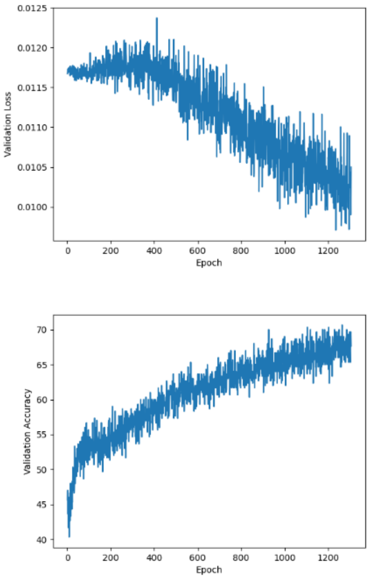




Figures 4, 5, 6: Training loss, validation loss and validation accuracy of the classifier using data augmentation and cropping

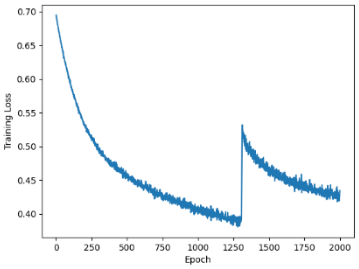
Compared to the previous attempt, the results are a little different but no significant improvement is achieved. However, the accuracy of this model increases with a higher rate than in the previous attempt and is more promising. Also, the validation loss seems more stable in this attempt. However, the point of using data augmentation is to make the model learn to generalize and achieve better results, with the cost of usually requiring more time to learn. Based on this, I decided to continue training this model for 800 more epochs.

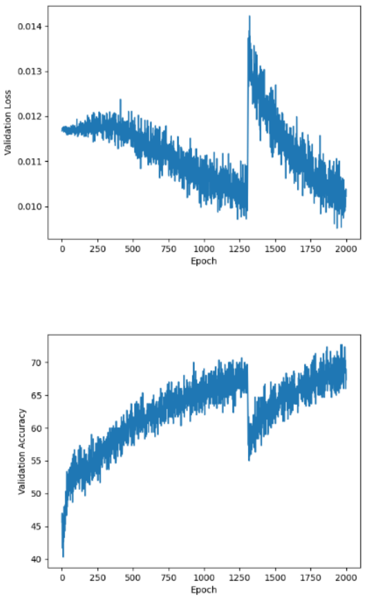




Figures 7, 8, 9: Training loss, validation loss and validation accuracy of the classifier trained for 1300 epochs using data augmentation and cropping

As we can see, the performance of this model significantly improves with increasing the number of epochs. The validation loss starts to stabilize and decrease after the 400th epoch, while the accuracy rises as much as 70%. Nonetheless, the model is still learning so further training is needed to improve its performance. This time, I decided to continue training the model for 700 more epochs, but this time by using image augmentation without cropping the images. I thought at this point, the model would easily be tuned to work with the whole images and maybe produce slightly better results.





Figures 10, 11, 12: Training loss, validation loss and validation accuracy of the model, tuned with augmented but not cropped images

The performance of the model is shown in figures 10, 11 and 12. In order to better visualize the tuning of the model with using whole images, the data is shown together with the previous training attempt (also the training is a continuation of it). At first, the performance significantly drops, but starts to improve with a very high rate. By doing this, the model even slightly surpasses its previous performance, this time reaching a maximum accuracy of 72%. The model is still learning by this point, so the training can be further continued to achieve better results. However these training attempts took an enormous amount of time to finish on my local machine (since online platforms such as Google Colab go to sleep after a certain amount of time and cancel the training process) and I could not increase the number of epochs much more and finish the training in a reasonable time.

VI. CONCLUSION

In this paper, I presented my approach for classifying Acute Lymphoblastic Leukemia based on microscopic cell images. Although the model reached only 72% accuracy, I am satisfied with the overall performance of the model, since it can be further trained over more epochs (it still has not reached its maximum learning capacity). Also the model proved capable of learning this task, which can be confirmed with the plots of the decreasing loss over the epochs. Further steps to achieve better performance would be to finish the training process (train the model for as much epochs as it continues learning). Another thing that would enable the model to better learn is to use the whole imbalanced dataset and use oversampling techniques instead of the approach I have taken in this project (This way the model will have more data to work with and generalize better). After these steps are taken, the model can be tested directly on the competition site [6] on the given (unlabeled) testing set in order to compare its performance to the other submitted models.

References

[1] Acute Lymphoblastic Leukemia, Leukemia & Lymphoma Society (https://www.lls.org/leukemia/acute-lymphoblastic-leukemia)

[2] Acute Lymphoblastic Leukemia, Wikipedia (https://en.wikipedia.org/wiki/Acute_lymphoblastic_leukemia)

[3] White Blood Cells Segmentation and Classification to Detect Acute Leukemia, Ms. Minal D. Joshi et al. IJETTCS 2013 (https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=91b3f21cd92fac05c7d8690034a182844b0f292f)

[4] Classification of Acute Lymphoblastic Leukemia Using Deep Learning, Amjad Rehman et al. 2018 (https://core.ac.uk/download/pdf/222832835.pdf)

[5] Customized Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia Using Blood Smear Images, Niranjana Sampathila et al. Healthcare 2022 (https://www.mdpi.com/2227-9032/10/10/1812)

[6] Classification of Normal vs Malignant Cells in B-ALL White Blood Cancer Microscopic Image: ISBI 2019, CodaLab (https://competitions.codalab.org/competitions/20395#learn_the_details)

[7] C_NMC_2019 Dataset: ALL Challenge dataset of ISBI 2019 (C-NMC 2019), Cancer Imaging Archive (https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52758223)