AI for Healthy vs. Acute Lymphoblastic Leukemia Cell Classification

Rambaldi Matteo, Carraro Amedeo

Data Description - Changes in Delivable 2

The dataset used in this study was provided by the research team of SBILab. The C-NMC 2019 dataset consists of 15,114 images of lymphocytes collected from 118 subjects. These images were split into a training, preliminary and test sets. Each image set contains single-cell images of healthy or malignant lymphocytes previously labeled by a team of oncologists.
The cells were dyed using the Jenner–Giemsa stain technique, that is a histological staining method commonly used in cytology and histology laboratories for staining blood smears, bone marrow aspirates, and other cellular samples. The first step involves preparing thin smears of the specimen to be examined. The smears are then fixed to the slide using a fixative solution, commonly methanol, to preserve the cellular structures and prevent degradation. The fixed smears are then immersed in a solution of Giemsa stain. Giemsa stain is a mixture of eosin (a red dye) and methylene blue (a blue dye), along with other components. The slides are incubated in the Giemsa stain solution for a specific period of time, typically around 20-30 minutes. During this time, the stain penetrates the cells and binds to various cellular components, revealing specific morphological details. After incubation, the slides are rinsed with buffered water or a similar solution to remove excess stain. Once rinsed, the slides are air-dried or gently blotted to remove excess moisture. They are then mounted with a coverslip using a mounting medium, such as Canada balsam or a synthetic resin. The stained slides are examined under a light microscope at various magnifications. The Giemsa stain reveals different cellular structures and components based on their affinity for the eosin and methylene blue components of the stain.
The SBILab team preprocessed these images using segmentation, image enhancement, and normalization techniques. Individual lymphocytes were segmented from blood smear images and placed in the center of them; each picture has 450×450 pixels and a black background.

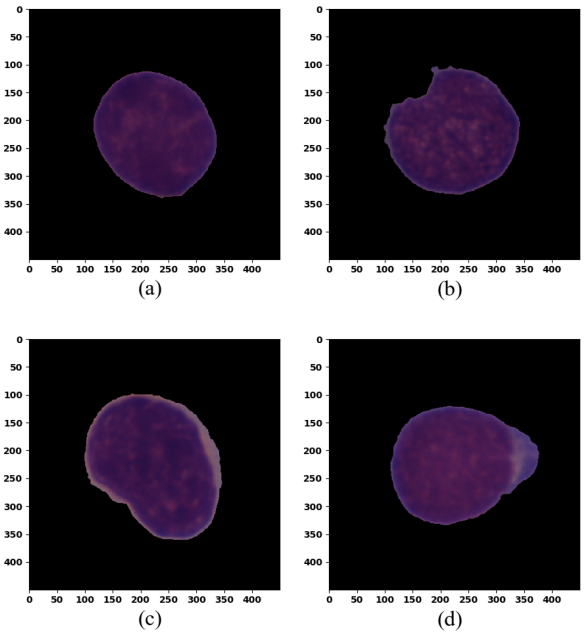

Figure 1. C-NMC 2019 dataset samples. The images (a,c) are malignant lymphocytes, and (b,d) are healthy lymphocytes.

Features

For the features extraction, from each image contained in the dataset, we extracted an array of 1387 features. We used low-order statistical, textural, morphological, contour, and DCT features extracted from each lymphocyte image.
We obtained the low-order statistics from each channel of the images in both RGB and HSV formats. The textural fea-

Feature Type	Number
Low-order statistical	108
Textural	75
Morphological	20
Contour	160
DCT	1024
Total	1387

tures were calculated using the coefficients of co-occurrence matrices. These coefficients represent the different gray level combinations that occur in the image and can be used in image classification tasks. We used features obtained from the gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level dependence matrix (GLDM), gray level size zone matrix (GLSZM), and neighboring gray-tone difference matrix (GLDM). The morphological features used, indicate the general shape of a lymphocyte. We obtained the contour features from the discrete Fourier transform of the centroid distance function (CDF) of the lymphocyte. The CDF represents the distance between the lympocyte centroid and each pixel of its contour. In the end, we calculated the DCT from the lymphocyte image converted to grayscale, producing a matrix with 202.500 DCT coefficients. The size of of this matrix was the same as the number of pixels in each image (450×450). We mapped the coefficients to a 1D array using a zigzag scan and used only the first 1024 lowest frequency coefficients. Finally, we combined all the features into a unique vector for each sample image for the training phase and normalized all values by subtracting each value from the column's mean and dividing it by the column's standard deviation.

Methodology

Neural Network Training and Fine-Tuning

We use the extracted features vector to fed it into an ANN that discriminates the lymphocytes as either malignant or healthy, define using Keras, a high-level neural networks API, developed with a focus on enabling fast experimentation and prototyping of deep learning models. Neural networks in Keras are built as sequences of layers, each layer being a modular building block. This allows for easy construction of complex network architectures. To find the best architecture to solve our problem, we did an extensive search in the hyper-parameter space of our network using grid search.

Parameter	Values	Chosen Value
Hidden layers	1, 2, 3, 4	1
Batch Size	250, 750, 1000, 1500	250
Dropout	0.1, 0.25, 0.3, 0.5	0.1
Neurons Number	1024, 1536, 2048, 2560	2560
Activation	Prelu, Relu, Sigmoid, Softmax	Relu
Optimizer	Adamax, Adam, SGD	Adam
Kernel Initializer	Random Uniforme, Normal	Normal

After finding the best architecture among all ANN possibilities, we implemented a fine-tuning step to obtain, among other values, the best number of epochs. Since the best optimization method was the Adam function, it was essential to choose the best values for the learning rate, β_1, and β_2. These coefficients are responsible for controlling the exponential decay rates of the moving averages. The values tested were 0.01, 0.001, 0.005, 0.0001, 0.0005 for the learning rate. The values tested for β_1 and β_2 were 0.99, 0.98, and 0.97. The best value found for β_1 and β_2 was 0.97, and for the learning rate the best was 0.001.

Support Vector Machine

Lightweight classifier trained from the features vector; the central idea of this algorithm is to obtain hyper planes that separate the samples used for training into their respective classes. The points closest to the discrimination hyperplane are called support vector points, and the distances between these points and a hyperplane are called margins. The support vector machine technique searches for a separation hyperplane that maximizes the margins. We used a canonical SVM classifier with default parameters and linear discrimination.

Model Evaluation

F1-score

In statistical analysis of binary classification, the F1-score is a measure of predictive performance. It is calculated from the precision and recall of the test, where the precision is the number of true positive results divided by the number of all samples predicted to be positive, including those not identified correctly, and the recall is the number of true positive results divided by the number of all samples that should have been identified as positive. The metric also allows the comparison with other studies, as the teams who participated in SBILab's challenge also used the F1-score.

Accuracy

Accuracy is also used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition. That is, the accuracy is the proportion of correct predictions (both true positives and true negatives) among the total number of cases examined.

Recall

Recall, also known as sensitivity or true positive rate, is a metric used in binary classification models to evaluate how well the model identifies positive instances out of all actual positive instances. It measures the proportion of

actual positive instances that were correctly identified by the model.

Precision

Precision, also known as positive predictive value, is a metric used in binary classification models to evaluate the accuracy of the positive predictions made by the model. It measures the proportion of true positive predictions out of all instances that the model predicted as positive.

Initial results

Regarding the generation of the two models, so far we have completed only one trial on the preliminary test set. We will proceed to refine the computation of the two machine learning models and to define the evaluation metrics not only on the preliminary test set but also on the final one.

Method	F1-Score on Preliminary Test Set
ANN	84.66%
SVM	77.43%

References

- **GCTI-SN: Geometry-Inspired Chemical and Tissue Invariant Stain Normalization of Microscopic Medical Images**

  Anubha Gupta, Rahul Duggal, Ritu Gupta, Lalit Kumar, Nisarg Thakkar, and Devprakash Satpathy

- **Stain Color Normalization and Segmentation of Plasma Cells in Microscopic Images as a Prelude to Development of Computer Assisted Automated Disease Diagnostic Tool in Multiple Myeloma**

  Ritu Gupta, Pramit Mallick, Rahul Duggal, Anubha Gupta, and Ojaswa Sharma

- **Neighborhood Correction Algorithm for Classification of Normal vs . Malignant Cells**

  by Yongsheng Pan, Mingxia Liu, Yong Xia, and Dinggang Shen

- **Overlapping Cell Nuclei Segmentation in Microscopic Images UsingDeep Belief Networks**

  Rahul Duggal, Anubha Gupta, Ritu Gupta, Manya Wadhwa, and Chirag Ahuja

- **Segmentation of overlapping/touching white blood cell nuclei using artificial neural networks**

  Rahul Duggal, Anubha Gupta, and Ritu Gupta

- **SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging**

  Rahul Duggal, Anubha Gupta, Ritu Gupta, and Pramit Mallick