AI for Healthy vs. Acute Lymphoblastic Leukemia Cell Classification

Rambaldi Matteo, Carraro Amedeo

Results

Proposed best model/models to be used

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. 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.

As the C-NCM 2019 dataset was unbalanced, the obtained accuracy may not reflect reality. F1-score was chosen as the metric to overcome this problem. The metric also allows the comparison with other studies, as the teams who participated in SBILab's challenge also used the F1-score.

The optimal hyper-parameters, i.e., the number of hidden layers, the number of neurons per layer, and the optimizer function, were chosen based on the performances of the ANN generated by combining the parameters presented below. The table also shows the optimal parameters.

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

Tables/figures – results of both models after given feedback

Method	F1-Score
ANN	84.66%
SVM	77.43%

Discussion

Main findings

The code was developed in Python 3.7 using machine learning libraries such as Tensorflow, Keras, and Mlxtend; the image processing libraries OpenCV, PIL, and Pyradiomics; and many other libraries responsible for data processing and manipulations, such as Numpy, Scipy, Pandas, CSV, Os, Multiprocessing, Queue, and Timeit. As the C-NCM 2019 dataset was unbalanced, the obtained accuracy may not reflect reality. F1-score was chosen as the metric to overcome this problem. The metric also allows the comparison with other studies, as the teams who participated in SBILab's challenge also used the F1-score.

The experiments were done on an Intel Core i7-12700H CPU @ 2.3 GHz × 14, with 16 GB of RAM and NVIDIA GeForce RTX 3050 Ti.

Notice that these approaches use less computational power than all convolutional neural network approaches submitted to this challenge.

Comparison of results with the literature

Yongsheng Pan, Mingxia Liu, Yong Xia and Dinggang Shen proposed the neighborhood-correction algorithm (NCA) to address this challenge, which consists of three major steps,

Method	F1-Score
ANN	84.66%
NCA	91.04%
VGG16	80.79%
RsN18	82.84%

including fine-tuning a pretrained residual network using training data and producing initial labels and featuremaps for test data, constructing a Fisher vector for each cell image based on its feature maps, and correcting the initial label of each test cell image via the weighted majority voting based on its most similar neighbors.

Atmika Honnalgere and Gaurav Nayak was to use transfer learning to adapt a VGG16 neural network with batch normalization, pretrained on the ImageNet dataset, to classify malignant and normal cells.

Christian Marzahl, Marc Aubreville, Jörn Voigt and Andreas Maier proposed a deep learning-based pipeline utilizing a dedicated preprocessing scheme with normalization and two augmentation methods. For training, we split the data into fivefold on subject level and use a ResNet18 network for classification and an additional regression head for bounding box prediction. During inference, the models are ensembled using a majority vote on the classification results.

Discuss/validate the results based on the literature related to the selected medical domain

BMA is the gold standard for leukemia diagnosis, it is an invasive procedure done under anesthesia. Exams done with peripheral blood are less invasive and may sometimes be preferred, even being less accurate than BMA. Recent studies reported very good results obtained with peripheral blood flow cytometry (PBFC). Lam et al. reported a sensitivity of 99.7% and a specificity of 98.5% obtained with PBFC. A disadvantage of flow cytometry is the requirement of marker reagents that may not be readily available in all laboratories, especially in third world countries, so blood smear image analysis may be an alternative. An F1-score of 93.70% is not accurate enough for disease diagnosis but can serve as a tool for assisting oncologists.

Discuss any limitations of this study

Refine our methodology by focusing on adding features similar to the 1387 features chosen for our interpretable model, especially textural and low-order statistical features. We may also test the inclusion of other classification models, such as decision trees, and linear and quadratic discrimination analysis. We may test our approach in a more complex dataset and try to solve a a multi-class classification problem. A possible way to improve our results is to use the scores returned by different classifiers before the decision, i.e., late soft fusion, and to integrate them in different ways, such as using their averages. It is also possible to use more sophisticated techniques, such as alpha-integration, and to optimize the weight of each primitive model in order to minimize the least mean squared error (LMSE) or the minimum probability of error (MPE).

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