

Processing and evaluating graphical biomedical data



Author: Ing. Linda Blahová

Supervisor: doc. Ing. Jozef Kostolný, PhD.

UNIVERSITY OF ŽILINA
Faculty of Management Science
and Informatics

Motivation

In the current scientific landscape, data processing is crucial, particularly in disciplines such as biomedicine and biotechnology. Utilising image-based biomedical data has significantly enhanced healthcare, aiding in swift and accurate diagnoses and treatment decisions. As the healthcare landscape is shifting towards prevention and personalised medicine, necessitating the efficient processing and analysis of vast datasets. Photomicroscopic imaging of patient cells is at the forefront, bridging the gap between medicine and technology. Notable applications include the automated identification of cancerous cells and, as explored in this thesis, the identification and classification of leukocytes in a blood smear, which has the potential to aid in diagnosing various diseases, including blood cancers, infections, autoimmune disorders, and allergies.

Problem

Image data is significantly used in biomedicine to visualise anatomical and pathological processes in the human body. When done manually, image data processing is complex, time-consuming, and prone to error. This thesis tackles these issues by concentrating on the automation of processing microscopic blood sample data, which is critical given the enormous amount of data. The thesis examines existing processing possibilities and tools, looking for ways to improve and validate them, with the objective of eventually speeding the processing of such vital biological imaging data.

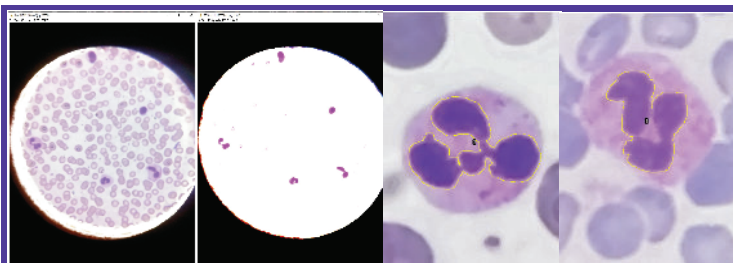


Figure 1 - image segmentation to identify leukocytes

Methodology

1. Analysis of the sample images from software (consistency and standards) and microscopic anatomy aspects
2. Analysis of the available techniques and tools that are generally used in biomedicine
3. Experimental search for a sequence of operations for both approaches and the whole process automation
4. **Approach 1 - leukocytes identification (Fig. 1)**
segmentation - histogram thresholding, color channels, region of interests analysis
Approach 2 - leukocyte type (single cell, Fig. 2)
classification - morphological feature extraction, k-best feature selection, active machine learning
5. Algorithms for verification and accuracy evaluation

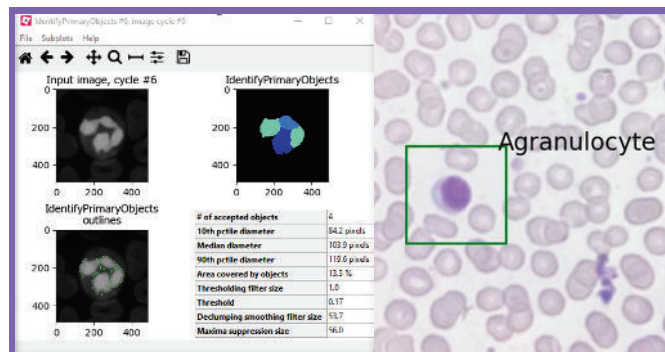


Figure 2 - feature extraction and machine learning classification to predict the leukocyte type

Experiments

To validate the relevance and robustness, both techniques were evaluated using two trials on distinct samples. One dataset album vs. all dataset albums were utilised for the segmentation. Classification was evaluated with the same configuration on two different datasets.

Results

The proposed solution was evaluated using balanced accuracy score and confusion matrices. The system proved to be accurate and configurable to any dataset via configuration files. The lower accuracy in Exp. 2 of the Approach 2 was due to lack of samples, which is demonstrated in Fig. 3.

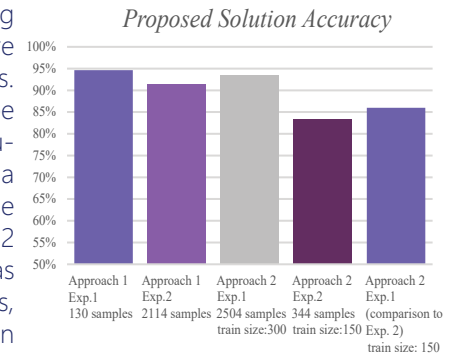


Figure 3 - experiments accuracy score

Contributions, Applications and References

- A comprehensive overview of various biomedical data types, with an emphasis on image data and their processing, that students can use as scripts.
- A novel solution that does not require manual selection of just the high-quality samples, that performs well even with large amounts of data and is adaptable to various datasets. The approach, together with the automation of all the necessary actions accelerates the identification and classification of leukocytes in laboratories.
- The solution was incorporated into the biobank being developed at our faculty.
- The implemented system and its outcomes were presented at an international conference IDAACS 2023 and published [1].

[1] L. Blahova, J. Kostolny, "Analysis of Photomicroscopy Images Using Active Learning Techniques", 2023 12th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 2023, pp. 798-803 (in press)