Global Healthcare Transformation in the Era of Artificial Intelligence and Informatics
J. Mantas et al. (Eds.)

doi:10.3233/SHTI250704

Optimal Set of Features for Leukaemia Images with Extracted Areas of Interest

Marinela BRANESCU[1], Stephen SWIFT, Allan TUCKER and Steve COUNSELL
The Department of Computer Science, Brunel University, West London, United Kingdom

Abstract: Feature extraction was found effective in image classification across various studies. From a dataset containing leukemia images of four main categories 17 features were extracted including Haralick texture features, the size and number of white blood cells, and average colours. The process of feature extraction produced high accuracy when evaluated with some machine learning classifiers. Nonetheless, selective application of features was implemented on the modified dataset by extracting regions of interest (ROI). This approach resulted in 131,071 different combinations of features; some configurations achieved superior accuracy. This research aimed to identify the optimal combination of features within an expanded leukemia dataset through ROI extractions. Notably, cell size and count emerged as significant factors contributing to enhanced accuracy.

Keywords: Haralick Texture Features, RGB colors, Regions of Interest

1. Introduction

The four main types of leukaemia need different treatments and early identification [2], the morphological characteristics are transformed into extractable features in smear test image evaluation [1]. Extracting regions of interest is found efficient in improving the image classification [3]. Features evaluation respective Haralick features, average colours, size and number were effective in improving the classification accuracy [1,2].

This study is evaluating the best combination of features by using selective feature application where the collected features from a set of leukaemia images with extracted regions of interest belonging to the four main categories of the disease are evaluated.

2. Methods

The initial experiment involves dividing 312 images belonging to the four main categories of leukaemia (ALL, AML, CLL, CML) on regions of interest resulting an expanded dataset that includes either a white blood cell or clusters of cells in each image. From the processed images (Figure 1), 12 Haralick features, three average colours, size and count of the cells are extracted as numerical values (Table 1).

This data is then analysed, with a machine learning classifier respective Function.MultilayerPercepton. Evaluating the extracted features, a median accuracy was

[1] Corresponding Author: MARINELA BRANESCU; E-mail: branescu.iuliana@gmail.com.

achieved, however to increase it, selective feature application was applied. From the resulting in 131,071 combinations of features with varying levels of accuracy the best combinations are selected (Table 1).

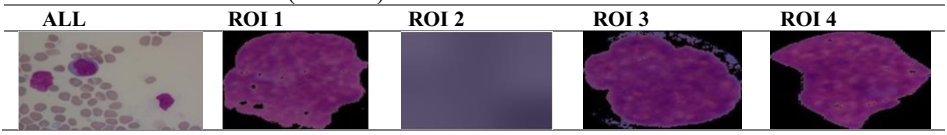

Figure1. Dividing the dataset per regions of interest resulting in an expanded dataset used in feature extraction.

3. Results

Reviewing the optimal feature combinations (Table 1) it is observed that integrating both cell size and count yields the highest accuracy for this configuration. Conversely, color features appear to be less effective compared to other options within the selection of 17 features.

Table 1. Best 5 performing combination of features for the images where ROI were extracted.

Combinations	110000000E	0100000000E	1100000100E	1110000000E	1100001000E
Accuracy	0.975791	0.974941	0.960285	0.959862	0.959014

Note: In the Table 1 a value of 0 indicates that a feature is excluded from combination, while a value of 1 signifies it's inclusion. The first feature listed is size, which is succeeded by the number of cells, followed by 12 Haralick features and three average colours metrics: average blue, average red, and average green. E summaries 7 of 0 respective omitted features and is used to reduce the space in the presented table..

4. Discussions and Conclusions

Selective feature application indicates the combination of features yielding the greatest accuracy observed in another research [2]. Here the number of the white blood cells and size are optimal in increasing the accuracy of classification as seen in Table 1. This investigation continues research on leukaemia image classification contrasting traditional feature extraction and Convolutional Neural Network. The 17 used features were optimal in classifying the original images before extracting regions of interest.

It is evident that the significance of features varies in this setup, where regions of interest were identified and subsequently analyzed for classification. Various configurations, leukemia subtypes, or classification models may yield differing results in terms features performance in evaluation. Enhancements to the research could be achieved by incorporating alternative feature sets and assessing their effectiveness.

References:

[1] Löfstedt T, Brynolfsson P, Asklund T, Nyholm T, Garpebring A. Gray-level invariant Haralick texture features. PLOS ONE. 2019 Feb; 14(2). doi:10.1371/journal.pone.0212110.

[2] Yadav DP, Kumar D, Jalal AS, Kumar A, Singh KU, Shah MA. Morphological diagnosis of hematologic malignancy using feature fusion-based deep convolutional neural network. Scientific Reports. 2023 Oct; 13(1). doi:10.1038/s41598-023-44210-7.

[3] S R, BM C, Ramani DM, Harsur MS. CNN based multi-view classification and roi segmentation: A survey. Global Transitions Proceedings. 2022 June; 3(1), pp. 86–90. doi:10.1016/j.gltp.2022.04.019.