Machine-learning-based classification of Stokes-Mueller polarization images for tissue characterization

K M Sindhoora 1, K U Spandana 1, D Ivanov 2, E Borisova 2, U Raghavendra 3, S Rai 4, S P Kabekkodu 5, K K Mahato 1 and N Mazumder 1, *

1 Department of Biophysics, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, India -576104
2 Institute of Electronics, Bulgarian Academy of Sciences, 72 Tsarigradso Chaussee Blvd., 1784, Sofia, Bulgaria
3 Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India - 576104
4 Department of Pathology, Kasturba Medical College, Mangalore, Karnataka, India – 575001
5 Department of Cell and Molecular Biology, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, India -576104

*E-mail: nirmal.mazumder@manipal.edu

Abstract. The microstructural analysis of tissues plays a crucial role in the early detection of abnormal tissue morphology. Polarization microscopy, an optical tool for studying the anisotropic properties of biomolecules, can distinguish normal and malignant tissue features even in the absence of exogenous labelling. To facilitate the quantitative analysis, we developed a polarization-sensitive label-free imaging system based on the Stokes-Mueller calculus. Polarization images of ductal carcinoma tissue samples were obtained using various input polarization states and Stokes-Mueller images were reconstructed using Matlab software. Further, polarization properties, such as degree of linear and circular polarization and anisotropy, were reconstructed from the Stokes images. The Mueller matrix obtained was decomposed using the Lu-Chipman decomposition method to acquire the individual polarization properties of the sample, such as depolarization, diattenuation and retardance. By using the statistical parameters obtained from the polarization images, a support vector machine (SVM) algorithm was trained to facilitate the tissue classification associated with its pathological condition.

1. Introduction

Novel microscopy techniques are needed to produce high-resolution, contrast-and-brightness enhanced images with minimum sample destruction. Although the present optical microscopic techniques, such as wide-field microscopy, fluorescence microscopy, scanning transmission electron microscopy etc. enabled one to probe into microscopic levels, they are either limited by the quality of the output image or by the instrumentation cost. The fluorescence polarization microscopy, being a contrast-enhancing technique, is used to acquire high-contrast images; however, labelling the sample can cause photobleaching and sample destruction. To overcome these disadvantages, the Stokes-Mueller
polarization technique can be used as a powerful method of analyzing a sample’s properties in detail [1]. The manual image analysis of tissue samples under a microscope is a very tedious and time-consuming process due to the complex nature of biological entities, which in turn demands an expert pathologist to get the accurate output. To overcome these limitations, an automatic fast and robust image processing technique is desirable [2].

Machine learning (ML) is one such field of image analysis, a subcategory of artificial intelligence (AI), which is a part of computer science that deals with the study of computers exhibiting human-like intelligence. The main goal of AI includes enhancing the ability of a computer to learn a problem and solve the same to provide a better solution by its own; it has applications in the fields of robotics, natural language processing (NLP), gaming etc. ML includes statistical techniques (algorithms) that enable machines to improve at the task with experience [3]. This is the main difference between ML and conventional programming, in which the computer is programmed with inputs and set of rules to get a preferred output, whereas in ML the computer is trained rather than programmed (figure 1).

![Figure 1. Traditional programming vs Machine learning.](image)

ML algorithms can be of the following types: supervised learning (SL), unsupervised learning (UL) and reinforcement learning (RL). It can be used to solve problems like classification, clustering and regression [3]. Hence, it is used in identifying spam emails, malware filtering, search engine result refining and related post suggestions in social media etc. [4]. ML is also used in sample characterization in Raman spectroscopy [5], spectral analysis and classification of pathological samples [6], quantitative phase imaging of red blood cells to diagnose haematological disorders [7], classification of a microscopic image of oral squamous [8] etc. Support vector machine (SVM) is one of the ML techniques used to solve a classification problem. The trained SVM model separates the datasets using a hyperplane [3].

2. Classification using SVM

The ML-algorithms-based classification reveals the important features of the sample under study; also, it performs well with smaller datasets. In the present study, tissue images of both normal and tumor regions of ductal carcinoma tissue are obtained by Stokes-Mueller polarization microscope with various input polarization states. Further, the polarization parameters, such as degree of polarization (DOP), degree of linear polarization (DOLP), degree of circular polarization (DOCP) and anisotropy, are reconstructed using the Matlab platform. From each of these polarization parameters, a gray level co-occurrence matrix (GLCM) is constructed. GLCM is a [8×8] matrix providing information regarding the co-occurrence of pixel values in the given image. A total of 19 GLCM parameters, such as contrast, correlation, energy, entropy etc., of normal and tumor regions of the tissue are calculated to find a significant feature to perform machine learning classification. The overall methodology of the current work is presented in figure 2.
3. Results and discussion

In this study, five normal and five tumor regions from ductal carcinoma tissue are selected and imaged with different input polarization states. Further, the four polarization parameters are reconstructed from 0 and 90-degree input polarization to train two different SVM classifiers – SVM1 and SVM2 with 0 and 90-degree input polarization images, respectively. Both classifiers are built with the same hyperparameters in order to maintain uniformity; 80% of the dataset is used for training and the remaining 20% is used for testing. A total of 19 GLCM features from each of image are used for training the classifier. The results of SVM1 training is presented in figure 3. The area under the receiver operating characteristic (ROC) curve shows that the accuracy of the model is 100%.

![Figure 3. ROC curve and confusion matrix for SVM1.](image-url)
Similarly, SVM2 is trained; the area under the ROC curve shows the accuracy of the model to be 91%, as shown in figure 4.

![SVM 2](image)

**Figure 4.** ROC curve and confusion matrix for SVM2.

SVM1 has a better classification accuracy in comparison with SVM2, which indicates that the polarization parameters at 0 degree input polarization reveal more distinct features required for ML classification.

### 4. Conclusion

The present study provides insights regarding the sample imaging and feature extraction for ML-based classification. However, further development is needed to improve the model’s performance by finding significant features for classification. Further, the ML-based image classification model is not fully automated and requires manual feature extraction. This step yields a deeper vision regarding the significant features of the sample under study. With the acquired knowledge on the sample features, it is easier to extend the classification problems from machine learning to deep learning, which is a widely used image analysis technique.

### Acknowledgements

We acknowledge Prof. K. Satyamoorthy, Director, Manipal School of Life Sciences, MAHE for his encouragement and Manipal Academy of Higher Education, Manipal, India for providing the infrastructure facilities. We also thank MAHE for Dr TMA Pai Ph. D. Scholarship. This work is supported by Department of Science and Technology (DST) - Science and Engineering Research Board (SERB) (Project No. ECR/2016/001944) and DST (Project No. DST/INT//BLG/P-03/2019), Government of India.

### References

[1] Mazumder N, Qiu J, Kao FJ and Diaspro A 2017 *Journal of Optics* 19 025301
[2] Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H and Ozcan A 2017 *Optica* 4 1437
[3] Chollet F 2017 *Deep learning with python* (Manning Publications Co.)
[4] LeCun Y, Bengio Y and Hinton G 2015 *Nature* 521 436–444
[5] Ye Z 2005 *IEEE Transactions on Automation Science and Engineering* 2 67–73
[6] Nayak G S *et al.* 2006 *Biopolymers: Original Research on Biomolecules* 82 152–166
[7] Kim G, Jo Y, Cho H, Min H S and Park Y 2019 *Biosensors and Bioelectronics* 123 69–76
[8] Rahman T, Mahanta L, Chakraborty C, Das A and Sarma J 2018 *Journal of Microscopy* 269 85–93