Polarimetric imaging combining optical parameters for classification of mice non-melanoma skin cancer tissue using machine learning

Thi-Thu-Hien Pham [a,b,*], Thanh-Ngan Luu [a,b], Thao-Vi Nguyen [a,b], Ngoc-Trinh Huynh [c], Quoc-Hung Phan [d], Thanh-Hai Le [e,**]

[a] *School of Biomedical Engineering, International University (VNU-HCM), Ho Chi Minh City, Viet Nam*
[b] *Vietnam National University HCMC, Ho Chi Minh City, 700000, Viet Nam*
[c] *Department of Pharmacology, University of Medicine and Pharmacy at Ho Chi Minh City, HCMC, Viet Nam*
[d] *Mechanical Engineering Department, National United University, Miaoli 36063, Taiwan*
[e] *Department of Information Technology Specialization, FPT University, Ho Chi Minh City, 700000, Viet Nam*

A B S T R A C T

Polarimetric imaging systems combining machine learning is emerging as a promising tool for the support of diagnosis and intervention decision-making processes in cancer detection/staging. A present study proposes a novel method based on Mueller matrix imaging combining optical parameters and machine learning models for classifying the progression of skin cancer based on the identification of three different types of mice skin tissues: healthy, papilloma, and squamous cell carcinoma. Three different machine learning algorithms (K-Nearest Neighbors, Decision Tree, and Support Vector Machine (SVM)) are used to construct a classification model using a dataset consisting of Mueller matrix images and optical properties extracted from the tissue samples. The experimental results show that the SVM model is robust to discriminate among three classes in the training stage and achieves an accuracy of 94 % on the testing dataset. Overall, it is provided that polarimetric imaging systems and machine learning algorithms can dynamically combine for the reliable diagnosis of skin cancer.

1. Introduction

Skin cancer is one of the most common forms of cancer and has a rapidly growing incidence and mortality [1]. Skin cancer can be broadly categorized as melanoma skin cancer (MSC) or non-melanoma skin cancer (NMSC), where NMSC was the fourth most common cancer in 2020 with over 1 million new cases recorded [2,3]. Cancer detection and diagnosis is traditionally performed using pathologic biopsies. However, biopsies are not only invasive and expensive, but can also result in scarring of the afflicted area [4]. Consequently, several new techniques for performing robust skin cancer diagnosis have been proposed in recent decades, including dermatoscopy [5,6], optical coherence tomography (OCT) [7,8], and reflectance confocal microscopy (RCM) [9–12]. However, the effectiveness of dermatoscopy as a diagnosis tool depends heavily on the level of competence of the examiner and the complexity of the

---

assessed lesions. Meanwhile, detecting early-stage melanomas using OCT still poses a significant challenge due to the sheer number of different forms which NMSC lesions may take [13]. For RCM, the image resolution decreases as the depth of penetration increases, and hence RCM provides only a limited ability to detect deep tissue samples [14,15]. Thus, the requirement for developing simple and reliable technique for skin cancer detection is nesserary.

Owing to its advantages of non-invasiveness, polarized laser light is emerging as a new technique for the diagnosis and staging of various cancers, including skin cancers [16]. As the polarized light propagates through the light-scattering media such as skin tissue, the randomness of the tissue structure depolarizes the light and scatters it back. The change in polarization state between the initial illuminating light and backscattered light, respectively, then yields much valuable information regarding the tissue structure [17]. Notably, cancerous tissue responds differently to polarized light than normal tissue due to the structural alterations caused by the cancer to the extracellular matrix and cells [18,19]. Consequently, capturing the backscattered light with a CCD camera (or some other detectors) and also analyzing the optical parameters can provide rich information for the subsequent characterization and classification of the tissue of interest. A. Vahidnia et al. [20] employed a straightforward and efficient polarimetric technique to distinguish between benign and cancerous lesions based on three important polarimetric parameters (i.e. diattenuation, retardance, and depolarization). L. Jütte et al. [21] developed a semi-automatic feature-based registration approach to compare the effects of various registration techniques on polarimetric analysis.

The combination of polarimetric and deep learning has been developed for biosensor application [22–25]. Ivanov et al. [26] applied three machine learning algorithms namely Logistic regression, Random Forest and Support vector machine, to classify healthy and tumor tissue from the polarimetric dataset with a precision up to 0.93. Sindhoora et al. [27] proposed Stokes-Mueller matrix imaging and Support vector machine model for carcinoma tissue classification with an accuracy of 91 %. In previous studies, the present group performed a combined method based on Mueller matrix decomposition and artificial intelligence (AI) classification [28], and a hybrid method consisting of Mueller matrix element values and AI processing [29]. Drawing on these studies, the present research proposes a further method for performing skin cancer detection by examining the change in optical properties of skin tissue samples caused by cancerous development utilizing a polarimetric imaging system based on a CCD camera and a machine learning model. It is noted that our previous studies [28,29] only used the results of the polarization properties and the Mueller matrix element values (1D) as input in AI classification models. In contrast, in this proposed study, the images of the Mueller matrix (2D) combined the polarization properties were used as the input of the classification model. From the previous and present results, it is possible to make a comparison table to choose the data of polarization properties, Mueller matrix values, Mueller matrix images or combination as the best input factor for skin cancer classification model. Therefore, in the proposed approach, the backscattered light from the skin tissue is captured by the CCD camera and processed by the Stokes-Mueller method to determine the Mueller matrix image and corresponding optical properties of the sample. The elements of the Mueller matrix and optical properties of the sample are then further processed by a machine learning model to classify the skin tissue into one of three different classes: normal, papilloma (i.e. precancerous), and squamous cell carcinoma (SCC).
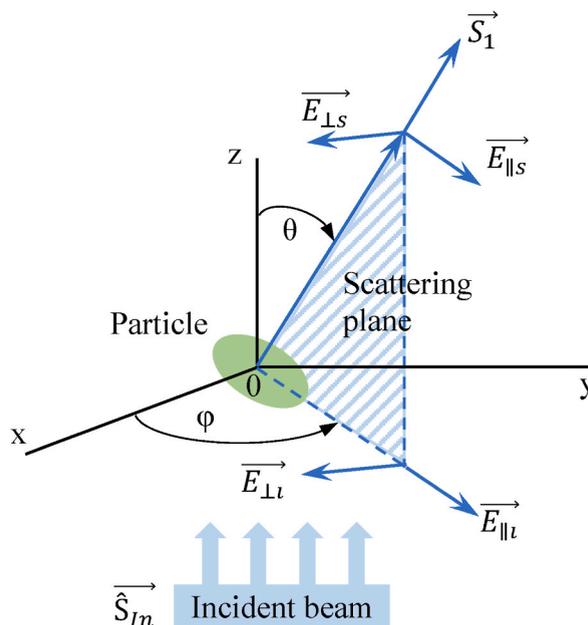

Fig. 1. Mathematical modeling of scattered polarization light [16].

2. Stokes-Mueller matrix formalism

Mathematically, an incidence polarized light $\overrightarrow{S_0}$ (oriented parallel to the z-axis) can be represented by two orthogonal linearly-polarized waves, $\overrightarrow{E_{\parallel l}}$ and $\overrightarrow{E_{\perp l}}$, where $\overrightarrow{E_{\parallel l}}$ and $\overrightarrow{E_{\perp l}}$ are parallel and perpendicular to the scattering plane, respectively. Consider the case shown in Fig. 1, where the incidence light meets a particle at the origin O and the scattering planes, P and $P_\perp$, have angles of θ and φ relative to the direction of the light, respectively. The change in polarization state (amplitude and phase) of the scattered light, $\overrightarrow{S_1}$, can thus be represented by the vectors $\overrightarrow{E_{\parallel s}}$ and $\overrightarrow{E_{\perp l}}$.

The relationship between the scattered polarization state, S_O, and initial polarization state, $\widehat{S}_{In}$, can be described by the linear equation $S_O = M\widehat{S}_{In}$, where M is the Mueller matrix of the scattering sample. In general, the polarization state of any arbitrary light can be described by four Stokes parameters, I, Q, U and V, where $I^2 = Q^2 + U^2 + V^2$, as shown in Eq. (1).

$$\widehat{S}_{In} = \begin{bmatrix} I_1 \\ Q_1 \\ U_1 \\ V_1 \end{bmatrix} \tag{1}$$

where I is the light intensity, and Q, U and V are the horizontal linear, 45-degree linear, and circular polarization light components, respectively [30]. The optical characteristics of optical samples can be fully determined using four input lights with different states of linear polarization ($\widehat{S}_{0^0} = [1,1,0,0]^T$, $\widehat{S}_{45^0} = [1,0,1,0]^T$, $\widehat{S}_{90^0} = [1,-1,0,0]^T$, and $\widehat{S}_{135^0} = [1,0,-1,0]^T$) and two input lights with different states of circular polarization ($\widehat{S}_{RHC} = [1,0,0,1]^T$ and $\widehat{S}_{LHC} = [1,0,0,-1]^T$). For each input light, a 4×4 Mueller matrix, M, describes the polarization properties of the sample which has form

$$\begin{bmatrix} I_O \\ Q_O \\ U_O \\ V_O \end{bmatrix} = [M] \begin{bmatrix} I_1 \\ Q_1 \\ U_1 \\ V_1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} I_1 \\ Q_1 \\ U_1 \\ V_1 \end{bmatrix} \tag{2}$$

in the Mueller matrix imaging technique, the input and output Stoke vectors are replaced by corresponding images of the 36 states of polarization [31]. The Mueller matrix in Eq. (2) is then computed as shown in Table 1. Where H, P, V, and M, are denoted for the linear polarized at $0°$, $45°$, $90°$, and $135°$. While L and R are denoted for the left-handed circular polarization light and right-handed circular polarization light, respectively. Then, the Mueller matrix calculation formula based on 36 polarization images is shown in Eq. (3).

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix}$$
$$= \begin{bmatrix} HH+HV+VH+VV & HH+HV-VH-VV & PH+PV-MH-MV & RH+RV-LH-LV \\ HH-HV+VH-VV & HH-HV-VH+VV & PH-PV-MH+MV & RH-RV-LH+LV \\ HP-HM+VP-VM & HP-HM-VP+VM & PP-PM-MP+MM & RP-RM-LP+LM \\ HR-HL+VR-VL & HR-HL-VR+VL & PR-PL-MR+ML & RR-RL-LR+LL \end{bmatrix} \tag{3}$$

33. Non-melanoma skin cancer model in mice

The experiment implemented on 100 healthy male Swiss albino mice purchased from Nha Trang Institute of Vaccine and Medical Biology (IVAC), Vietnam. It is noted that this study was conducted on mice rather than humans because (1) when conducting experiments on mice, the stages of skin cancer are created and controlled better, and being proactive in the quantity and quality of samples; (2) mouse has many similarities to humans in terms of genome, anatomy, physiology and genetics; and (3) in the next research steps, experiments can be conducted in vivo when capturing the images of polarization states directly on anesthetized mice. During the experiment, mice were provided with water and pellets that bought from the IVAC. All mice were housed in standard conditions at $75°F$ and 45% humidity. Each mouse was housed separately in $12 \times 12 \times 15$ cm^3 cell, with 6 cells in 1 cage. All

Table 1
Symbols and definitions of optical parameters.

Name	Symbol	Definition
Orientation angle of linear birefringence	α	
Linear birefringence	β	$2\pi(n_s - n_f)l/\lambda_0$
Orientation angle of linear dichroism	θ_d	
Linear dichroism	D	$2\pi(\mu_s - \mu_f)l/\lambda_0$
Circular birefringence	γ	$2\pi(n^- - n^+)l/\lambda_0$
Circular dichroism	R	$2\pi(\mu^- - \mu^+)l/\lambda_0$
Depolarization index	Δ	

experimental protocols were approved by the scientific committee, specialty of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Viet Nam with licensing No. 03–2016/QĐ-SĐHD. All experiments were performed in accordance with relevant guidelines and regulations. The proposed study is reported in accordance with ARRIVE guidelines. Mice were stabilized for 1 week before the experiment, and the weight of each mouse at the beginning of the experiment was from 25 to 30 g. A group of 20 healthy mice was cared as a control group. Fig. 2 shows the non-melanoma tumors were established in mice by a two-stage chemical carcinogenesis protocol [32]. Briefly, the dorsal skin of healthy Swiss *albino* mice was shaved over an area of approximately 2×2 cm^2 one day before the experiment. Tumor initiation was then performed via a single application of the chemical mutagen 7,12-dimethylbenz[*a*]anthracene (DMBA, Sigma – Aldrich, Germany). Two weeks later, tumor promotion was conducted by the twice-weekly topical application of croton oil (Sigma – Aldrich, Germany) over a period of 20 weeks. The mice were then sacrificed by CO_2 dried-ice, and skin samples were isolated and fixed in 10 % formalin. It is noted that the skin samples were taken similarly location on dorsal regions of the mice for all groups. The tissues were embedded in paraffin wax, sectioned with a microtome, and then stained with Hematoxylin and Eosin (H&E). The stained slides were observed under a light microscope to classify and confirm the tumor progression from normal to benign papillomas and then to squamous cell carcinoma (SCC) by histologists. From the analysis results, all mice of tumor groups have abnormal skin, with 100 % of mice having papilloma, of which some become carcinoma.

In accordance with the histopathological analysis results from histologists, the tissue samples were classified into three types: healthy, papilloma, and SCC. Fig. 3 shows typical histopathological analysis results for the three classes. To ensure the accuracy of the classification results, each sample was measured at least three times.

4. Experimental setup

As shown in Fig. 4 (a) & (b), the polarization imaging system comprised two blocks: the generator block and the analyzer block. The generator block was used to produce incident lights with particular states of polarization, while the analyzer block was used to determine the polarization state of the beam scattered by the sample. The generator block consisted of a frequency-stable He–Ne laser (HNLS008R, Thorlabs Co.) with a central wavelength of 633 nm, a linear polarizer P1 (GTH5M, Thorlabs Co.) to generate four linear polarization states (i.e., 0° (denoted as *H*), 45° (denoted as *P*), 90° (denoted as *V*), and 135° (denoted as *M*)), a quarter-wave plate R1 (QWP0-63304-4-R10, CVI Co.) to produce left-handed circular polarization light (denoted as *L*) and right-handed circular polarization light (denoted as *R*), a convex lens (L1), and a concave lens (L2). Meanwhile, the analyzer block consisted of a linear polarizer (P2), a quarter-wave plate (R2), and a CCD camera (CCD, DCU224C, Thorlabs, Inc.) fitted with a zoom lens and connected to a computer. The P2 and R2 in the analyzer performed the same functions as P1 and R1 in the generator, respectively. It is noted that polarizers P1 & P2 and quarter-wave plates R1 & R2 were mounted on rotation motorized stages (SGSP-60YAW-0B, Sigma Koki Co.) to generate the 36 polarization states required to construct the Mueller matrix for each sample. When performing the experiments, the linear polarization states of generator block was produced by rotating the polarizer (P1) whereas the circular polarization states was generated by moving P1 out of the laser path with a slider and rotating the R1 to the right- and left-hand circular polarization states. The same procedure was made for the analyzer block. Total six polarization states of generator block associated with six polarization states of the analyzer block enable to obtained the Mueller matrix of samples with 36 images calculated from Eq. (3). For example, the combination (*HV*) the use of horizonal polarization light of generator block and vertical polarization light of analyzer block, respectively. Therefore, to obtain m$_{11}$, it is required four combination of measurements, namely (*HH*), (*VV*), (*HV*), and (*VH*), whereby m$_{11}$ = HH + HV + VH + VV.

4.1. Data processing

Machine learning models generally utilize a feature scaling normalization mechanism to scale the values of the various features to a consistent range (i.e., magnitude) in order to prevent one feature from dominating the others in the classification process [33]. Previous studies have shown that models operating on scaled data achieve a significantly better performance than those which use unscaled data [28,34]. In the present study, the features of each input image were normalized to scale the image intensities in range [0, 1] using the Min-Max method in accordance with Eq. (4).

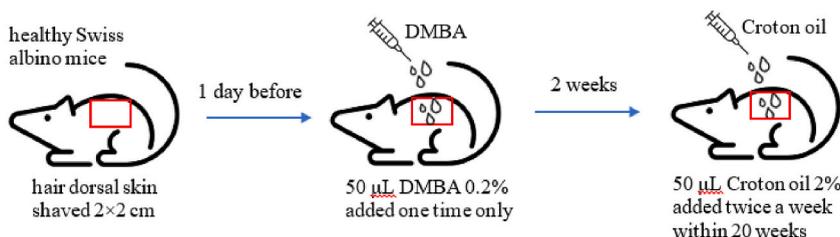$$x = \frac{x - \min(X)}{\max(X) - \min(X)} \tag{4}$$

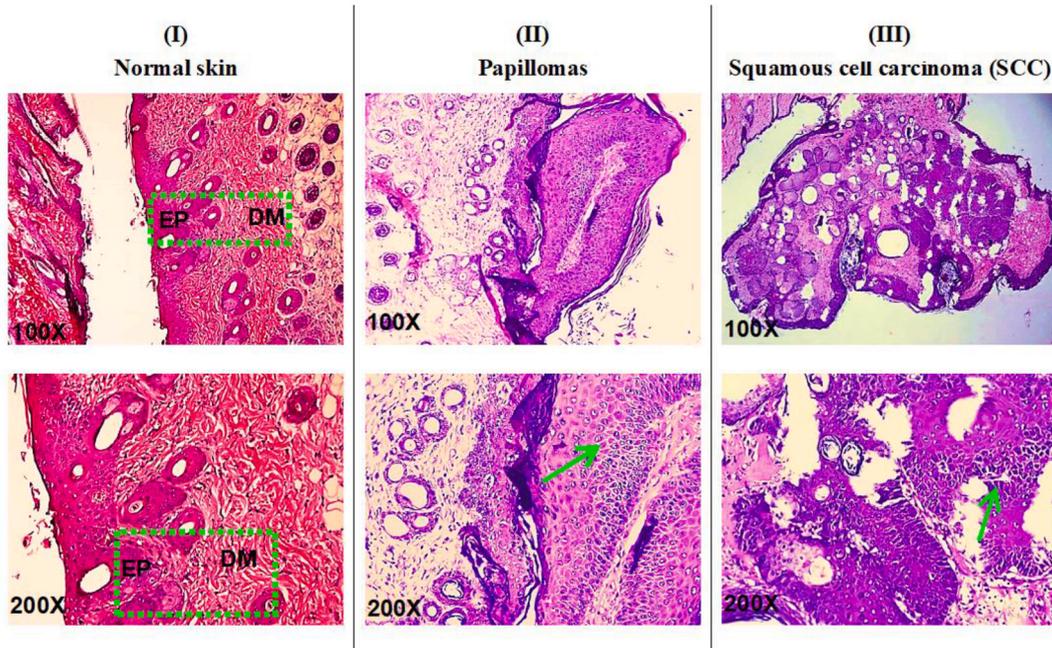**Fig. 2.** Two-stage chemical carcinogenesis protocol used to induce non-melanoma tumors.

4

Fig. 3. Microscopic histopathological results for three classes of H&E-stained mice skin tissue samples. Nuclei are stained blue/black by H&E, whereas keratin and cytoplasm are coloured red. Infiltration of stromal cells is shown by the arrow: (I) Normal skin was stained as control. The dermis (DM) and epidermis (EP) are separate from each other. (II) Mouse skin has epidermal hyperplasia (hyperkeratosis and acanthosis), papillary, and vascular-associated axis. (III) The squamous cells in mouse skin have hyperplasia, forming keratin pearls and plasmodesmata, breaking the basement membrane, and infiltrating stromal cells. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
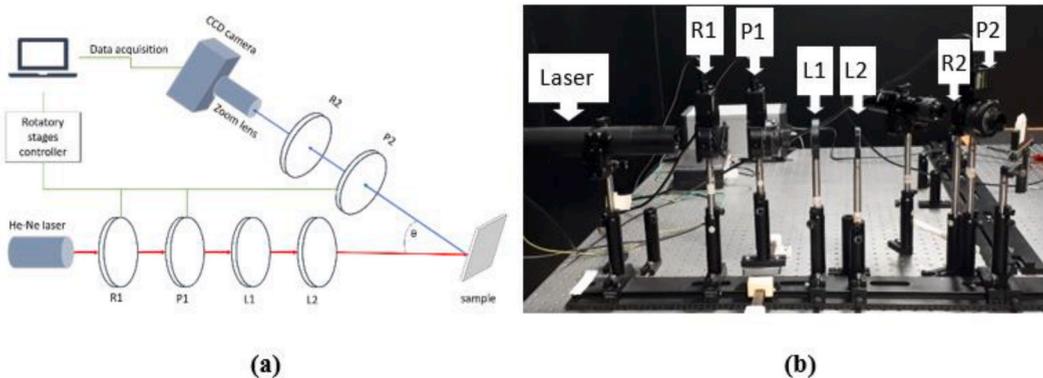

Fig. 4. Experimental setup: (a) schematic illustration, and (b) real-world configuration.

where x is an intensity of current pixel, and X is an input image.

The optical parameters of the backscattered polarization light captured by the CCD camera were extracted using the Stokes-Mueller method, which decomposes the Mueller matrix of the polarization light into seven discrete polarization parameters, namely the orientation angle of linear birefringence (α), the linear birefringence (β), the orientation angle of linear dichroism (θ_d), the linear dichroism (D), the circular birefringence or optical rotation angle (γ), the circular dichroism (R), and the depolarization index (Δ). From the 16 images of Mueller matrix in Eq. (3), each Mueller matrix element image was calculated the average value based on the values of pixels in the image and then normalized from the range of [0, 255] to [−1, 1] accordingly. These values of the Mueller matrix are put into the equations for extracted optical parameters that have been presented in previous studies by the current group [35,36]. The mathematical definitions of the seven parameters are shown in Table 1, where l is the path length through the medium (equivalent to the sample thickness); λ_0 is the wavelength of the incident light in a vacuum; n_s and n_f are the refractive indexes of the slow and fast linearly polarized waves, respectively; and μ_s and μ_f are the absorption coefficients of the slow and fast linearly polarized waves, respectively. Finally, n^+, μ^+ and n^-, μ^- are the refractive indexes and absorption coefficients of the right-hand circular polarized light

and left-hand circular polarized light, respectively. It is noted that the optical parameters presented for mainly optical properties of a biomedical sample including birefringence, dichroism and depolarization index. Thus, in this study, these optical parameters were used as the inputs of the classification model.

4.2. Feature extraction

For each sample, a feature vector of length 22, i.e., $[m_{12}; m_{13}; m_{14} \ldots m_{44}; \alpha; \beta; D; \theta_d; \gamma; R; \Delta]$, was constructed consisting of average values from fifteen Mueller matrix images (normalized with respect to m_{11}) and seven optical parameters. The feature vectors were compiled into a dataset consisting of 90 samples including 20 feature vectors for SCC, 33 feature vectors for normal and 37 feature vectors for papilloma. The Linear discriminant analysis (LDA) transformation was applied to transform the training and testing datasets for reducing the data dimensions and increasing the interpretability. The principle of LDA is the determination of new axes for data points projected onto and gernerate new corresponding data points that maximize the separation of different classes and minimize distance within classes [37]. The number of components (axes) are determined as shown in Eq. (5):

$$n\_components \leq \min\big((n\_classes-1), n\_features\big) \tag{5}$$

where $n\_components$ is number of remained components after applying LDA transformation; $n\_classes$ and $n\_features$ are number of classes and features in dataset, respectively.

After applying LDA transformation, the dataset is reduced from 22-dimensonal space (features) to 2-dimensional space for all samples with the new size of 90×2. Fig. 5 (a) & (b) show input data of SCC (red), normal (green) and papilloma (blue) in training dataset before and after applied LDA transformation.

4.3. Classification

In the present study, three common machine learning algorithms (i.e., K-nearest neighbors, Decision tree and Support vector machine) are used to classify three classes of skin tissues from the collected dataset.

4.3.1. K-nearest neighbors

K-nearest neighbors (KNN) is a simple learning algorithm that assigns a label for each instance x by majority voting of its known k nearest neighbors labels [38]. To determine which data points are the K-nearest neighbor of x, the distance between x and other data points, represented as y, is calculated by the Euclidean or Manhattan distance measurement function as shown in Eq. (6) and Eq. (7), respectively.

$$d(x,y) = \sqrt{\sum_{i=1}^{M} (x_i - y_i)^2} \tag{6}$$

$$d(x,y) = \sum_{i=1}^{M} (x_i - y_i) \tag{7}$$

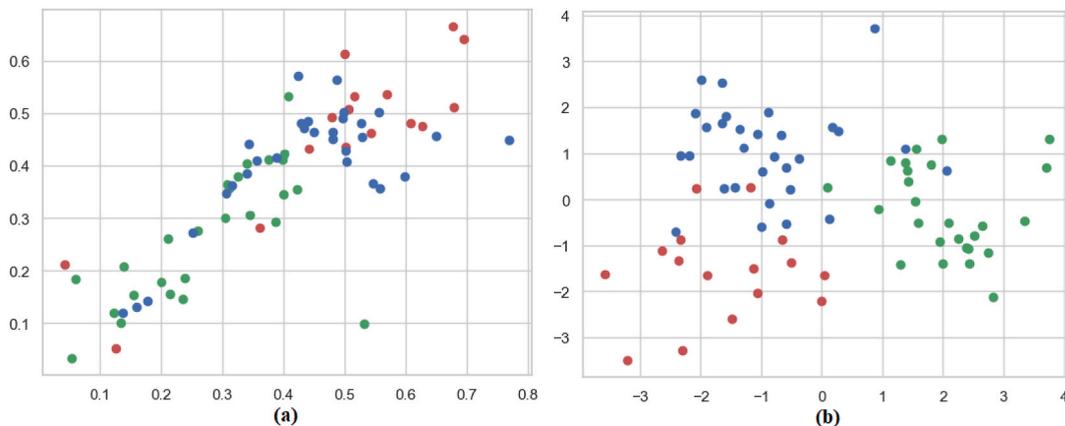where x and y are data points; and M is a length of a feature vector.

Fig. 5. Training dataset (a) before and (b) after applied LDA transformation with SCC (red), normal (green) and papilloma (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4.3.2. Decision tree

Decision tree (DT) algorithm is constructed following tree structure, that implemented by a series of questions that splitting the given data into certain leaf nodes then represented as class labels [39]. A decision tree starts with a root node, related to the first decision rule (denoted as d(t)) based on the feature vectors. Answering d(t), they are divided into two groups (known as branches) of nodes. New decision rules will be generated to continue splitting data. For DT, number of features used to find the split at each node, whereas the maximum depth of tree relates to how much information a classifier captures. If the depth of tree is large, the DT classifier can become complex and it can be the causes of overfitting. Based on the previous study [28], the range of maximum depth of tree is selected for searching the best estimator and avoiding the overfitting problem.

4.3.3. Support vector machine

Support vector machine (SVM) is a supervised learning algorithm that generates and optimizes the hyperplane in N-dimensional space in which N is the number of features that separate training data into defined categories [40]. In the present study, the dataset has two features after applying LDA transformation, the hyperplane (decision boundary) can be a simple line defined as $w^T x + b = 0$. Distance from data point x_0 to the plane $w^T x + b = 0$ is calculated by Eq. (8).

$$d = \frac{|w^T x_0 + b|}{|w|} \tag{8}$$

where d is a distance from x_0 to hyperplane, w and b are parameters of line. Data point x_0 is assigned belongs to the specific category if $w^T x + b > 0$, and otherwise. In training stage, SVM tries to determine the value of w and b that maximizes the distance of support vectors to the hyperplane.

The dataset was randomly partitioned into a training dataset and a testing dataset in a ratio of 80:20, respectively. Fig. 6 shows the number of feature vectors belonging to each class of skin tissue (i.e. 20 feature vectors for SCC, 33 feature vectors for normal and 37 feature vectors for papilloma) in the two datasets. The training dataset was used to perform cross-validation training of three classification algorithms (KNN, DT, and SVM) with the grid search method used to determine the best parameters for the estimator in every case (see Fig. 7 (a) & (b)).

Overfitting is a common problem in the training of machine learning models and occurs when the trained estimator learns too many details about the training dataset and thus performs poorly when applied to unseen samples in the testing stage. One of the most common solutions for preventing overfitting is to employ a cross-validation technique. In particular, the training dataset is split into a certain number of folds (k folds), and the classifier is then fed the data belonging to (k-1) folds in sequence, leaving one fold out for validation in every case, as shown in Fig. 7(b). The performance of the classifier is measured throughout the training folds and validation process, and the performance data are then evaluated and compared in order to determine the best hyperparameters for the model. In the present study, the three classifiers were trained using a 5-fold cross-validation technique. To ensure the reproducibility of the trained model, the samples of each fold were determined randomly using a random number generator. Having completed all 5 folds, the correct predictions were identified by comparing the fold predictions with the correct labels and then the validation accuracy was computed as the mean of the correct predictions of all the folds.
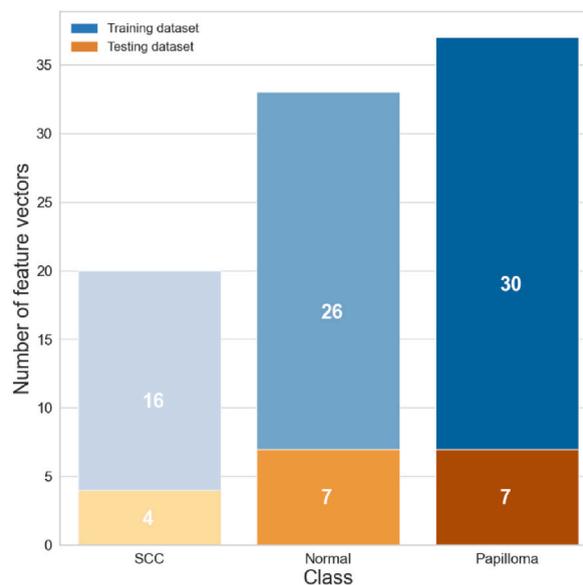

Fig. 6. Feature vector compositions of each type of skin tissue sample in training (orange) and testing (blue) datasets. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
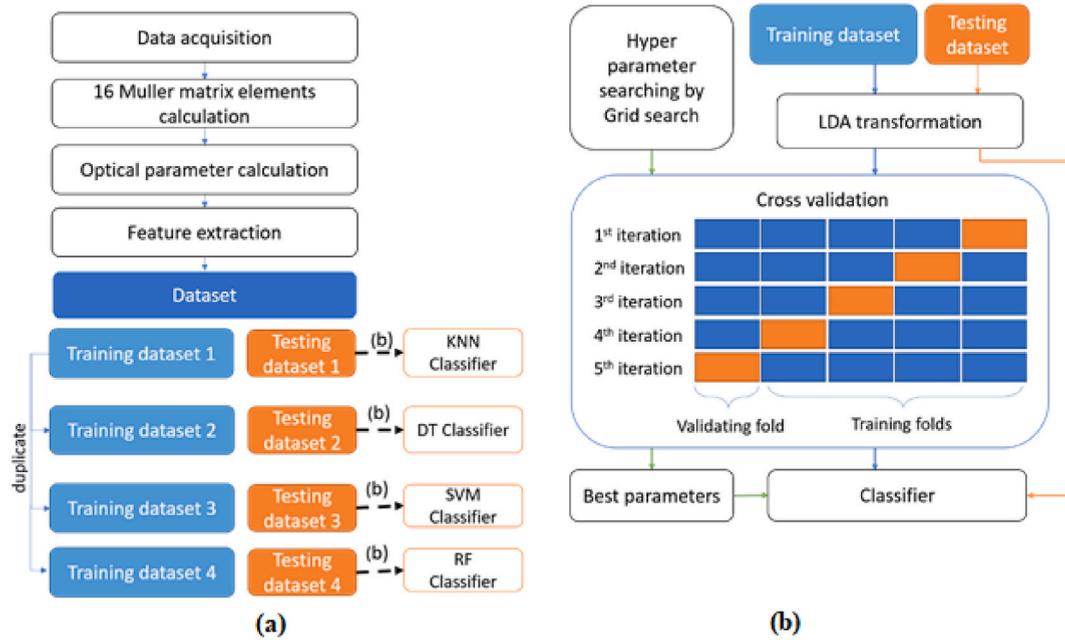
Fig. 7. (a) Workflow of experimental procedure and predictor training; and (b) training and testing process scheme using cross-validation and grid search method.

The hyperparameters for each classifier were chosen using the grid search method. In particular, for each classifier (KNN, DT and SVM), the ranges of the corresponding hyperparameters were fitted using the grid search technique during the training and validation processes in order to determine the parameter values which yielded the optimal classification performance. Table 2 lists the hyperparameters and corresponding ranges of each of the three classifiers.

4.4. Performance metrics

The performance of the three classifiers in the training, validation and testing stages was evaluated using four metrics, namely the accuracy, precision, recall, and F1 score, defined respectively as shown in Eq. (9) – Eq. (12).

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN} \tag{9}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{F1score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

where TP, TN, FP and FN denote true positive, true negative, false positive and false negative, respectively. Moreover, the performance of the classifier has further evaluated by confusion matrix.

Table 2
Hyperparameters and ranges of three classifiers.

Classifier	Parameter	Hyperparameter
KNN	Number of neighbors	[3, 4, 5, …, 17, 18, 19]
	Weight metric	Uniform/Distance
	Distance function	Euclidean distance/Manhattan distance
DT	Maximum percentage of features used in each splitting	0.5/1.0
	Maximum depth of tree	[2, 3, 4, …, 8]
SVM	Type of kernel	Linear/radial basis function (rbf)
	Gamma value	$[10^{-9}, 10^{-8}, 10^{-7} \ldots 10^2, 10^3]$

Fig. 8 shows the confusion matrix which TP and TN indicate the classifier correctly predict the positive and negative cases, whereas FP and FN indicate the classifier wrongly predict the positive and negative cases, respectively.

5. Results and discussion

5.1. Mueller matrix images

Fig. 9 shows the typical Mueller matrix images obtained for the three classes of skin tissue. Note that, for each sample, the element images are normalized to the corresponding m_{11} element image for comparison purposes. It is seen that there are different visualization among three classes of skin tissue. Particularly, Mueller matrix elements m_{22}, m_{23}, m_{42} in Fig. 9(b) and elements $m_{22}, m_{24}, m_{33}, m_{34}, m_{43}$ and m_{44} in Fig. 9(c) contain distinct green and (to a lesser extent) red color areas than that corresponding elements for the normal skin tissue (Fig. 9(a)). It means that the papilloma and SCC skin tissue samples have a greater intensity than the normal skin sample. Notably, the difference in intensity and color of the element images increases as the cancer stage increases. Hence, the feasibility for utilizing the features extracted from the Mueller matrix elements for the classification of the skin cancer stage is confirmed.

The difference between the element images of the three different classes was evaluated statistically by performing Analysis of Variance (ANOVA) tests to compute the corresponding F-scores and p-values. The corresponding results are shown in Fig. 10. The F-score represents the ratio of the variance between groups to that within groups. In other words, a larger F-score indicates that the variability between groups is relatively larger than that within groups, and hence the distributions of the individual groups seldom (if ever) overlap. The results presented in Fig. 10(a) thus suggest that $m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24}, m_{32}, m_{34}, m_{42}, m_{44}$, linear birefringence α, linear dichroism D, and circular dichroism R are particularly suitable features for classification purposes. On the other hand, to reject the null hypothesis of the dataset is equal, a high F-score is required, equivalently, a low p-value (see Fig. 10(b)). The results presented in Fig. 10(b) thus confirm that elements $m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24}, m_{32}, m_{34}, m_{42}, m_{44}$, linear birefringence α, linear dichroism D, and circular dichroism R, each with a p-value lower than 0.05, are promising candidates for classification purposes.

5.2. Classification performance

As described in Section 4.3, the optimal hyperparameters for each classifier were identified using the grid search technique. The results are shown in Table 3.

Each classifier was trained using the optimal hyperparameters indicated in Table 3. As described in Section 4, the training dataset was split into 5 folds, and each classifier was trained 5 times individually. In each training round, one fold was omitted for validation purposes and to prevent the model from overfitting. Fig. 11 shows the performance metrics of the three classifiers in the training and validation stages in combination of 15 Mueller matrix images and 7 optical parameters dataset (see Fig. 11 (a)), and 15 Mueller matrix images dataset (see Fig. 11 (b)), respectively. It obviously seen that when combining the two types of dataset (i.e. Mueller matrix images and optical parameters) the models achieve the performance metrics better than applying only Mueller matrix images in both the training and validation stages. Particularly, in Fig. 11 (a), the KNN classifier achieved a 100 % performance for all four evaluation metrics on the training dataset However, the performance metrics reduced by approximately 10 % on the validation dataset. The DT classifier achieved a poorer performance than the KNN classifier for all four evaluation metrics, and also exhibited a slight reduction in performance (~5 %) when applied to the training dataset. Overall, the results thus suggest that a slight overfitting effect occurred for both estimators. The performance of the SVM classifier was very similar to that of the DT model on the training dataset. However, in contrast to the KNN and DT models, the SVM model performance improved when applied to the validation dataset and was identical to that of the KNN classifier.

Fig. 12 shows the decision surfaces of the three classifiers of 15 Mueller matrix images and 7 optical parameters dataset. The results
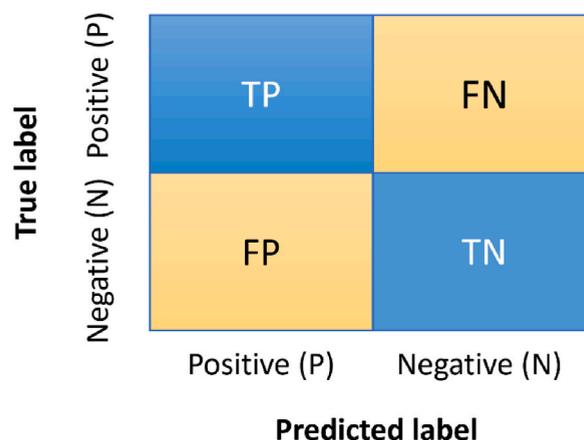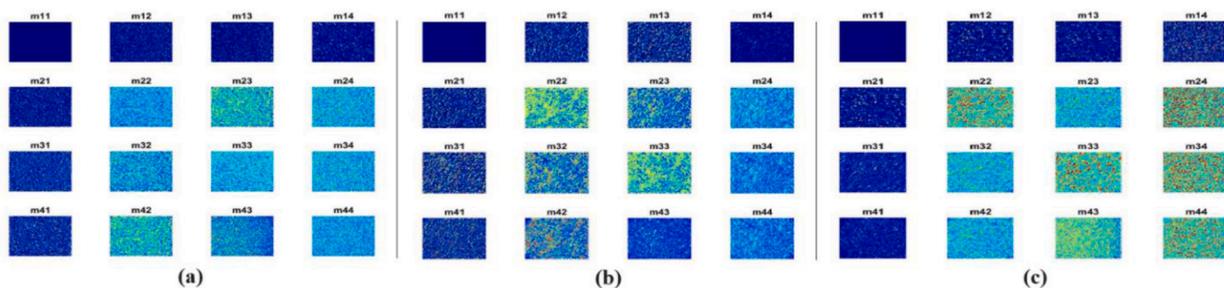

Fig. 8. Confusion matrix.

Fig. 9. Mueller matrix images of: (a) normal skin tissue, (b) papilloma skin tissue, and (c) SCC skin tissue.
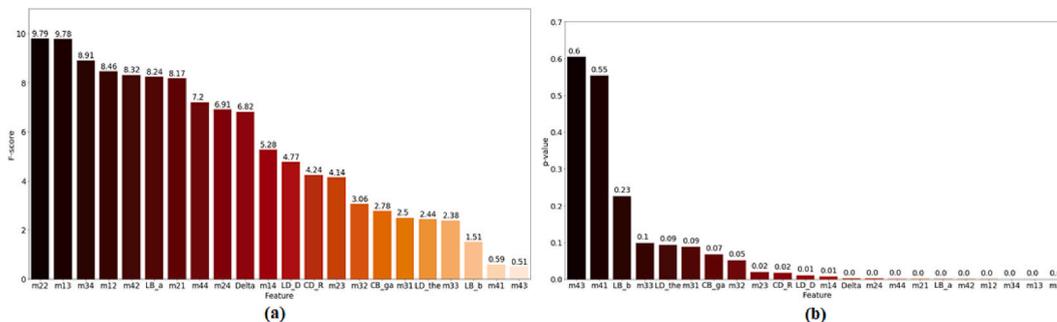

Fig. 10. Results of ANOVA test for: (a) F-Score and (b) p-value.

Table 3

Grid search results for optimal hyperparameters of each classifier.

Classifier	Parameter	Hyperparameter	Best parameter
KNN	Number of neighbors k	[3, 4, 5, …, 17, 18, 19]	k = 9
	Weight metric	Uniform/Distance	Distance
	Distance function	Euclidean distance/Manhattan distance	Euclidean distance
DT	Maximum percentage of features used in each splitting	0.5/1.0	0.5
	Maximum depth of the tree	[2, 3, 4, …, 8]	3
SVM	Type of kernel	Linear/Radial basis function (rbf)	Linear
	Gamma value	$[10^{-9}, 10^{-8}, 10^{-7}, …, 10^2, 10^3]$	None

presented in Fig. 12(a) confirm that the KNN model shows slight overfitting of some of the papilloma (green circle) and SCC (blue square) datapoints. Similarly, the DT model suffers overfitting of some of the normal, SCC and papilloma datapoints (red triangles, blue square and green circle in Fig. 12(b)). In general, overfitting causes the decision surface to contain small regions with only one data point. Observing the decision boundaries for the SVM model (Fig. 12(c)), it is seen that although the boundaries do not discriminate perfectly between the different classes, they contain no such regions. Thus, it is inferred that no overfitting occurs in the SVM model. It is further noted that all three models fail to distinguish between the normal and papilloma classes or SCC and papilloma classes in some cases.

The performance of the trained classifers was further evaluated on the testing dataset. Fig. 13 (a) – (c) show the confusion matrixes obtained for the three classifiers. It is seen that all three classifiers perform well in identifying the SCC samples, but misclassify one (KNN and SVM) or two (DT) normal skin tissue samples as papilloma tissue. As discussed above, the KNN and DT models suffer several instances of overfitting in the decision surface images, whereas the SVM does not. Table 4 shows the accuracies of the three models in the training, validation, and testing stages. Overall, the SVM classifier provides the best tradeoff between the prediction performance and robustness to overcome toward overfitting. Accordingly, Table 5 shows the detailed performance metrics of the SVM classifier for the three classes of skin tissue. It is seen that the classifier achieves an excellent classification performance for the SCC class (precision = 1; recall = 1). However, the classification performance for the normal class is slightly degraded (precision = 1; recall = 0.86) due to the erroneous classification of one normal sample as papilloma (precision = 0.88; recall = 1). Despite this performance loss, the SVM model still has a good potential for cancer detection and diagnosis. SVM is efficient with the clear margin of separation between three samples classes. Furthermore, the number of features for each data point is smaller than the number of training data samples enhanced the performance of SVM classifier. The present study shows that using features extracted from Mueller matrix images (2D) are used as input of AI models as good as the performance of AI models using Mueller matrix values (1D) [29] or additional information for effective optical parameters (1D) [28] as inputs. It is noted that when compared the results of the present study to the results in
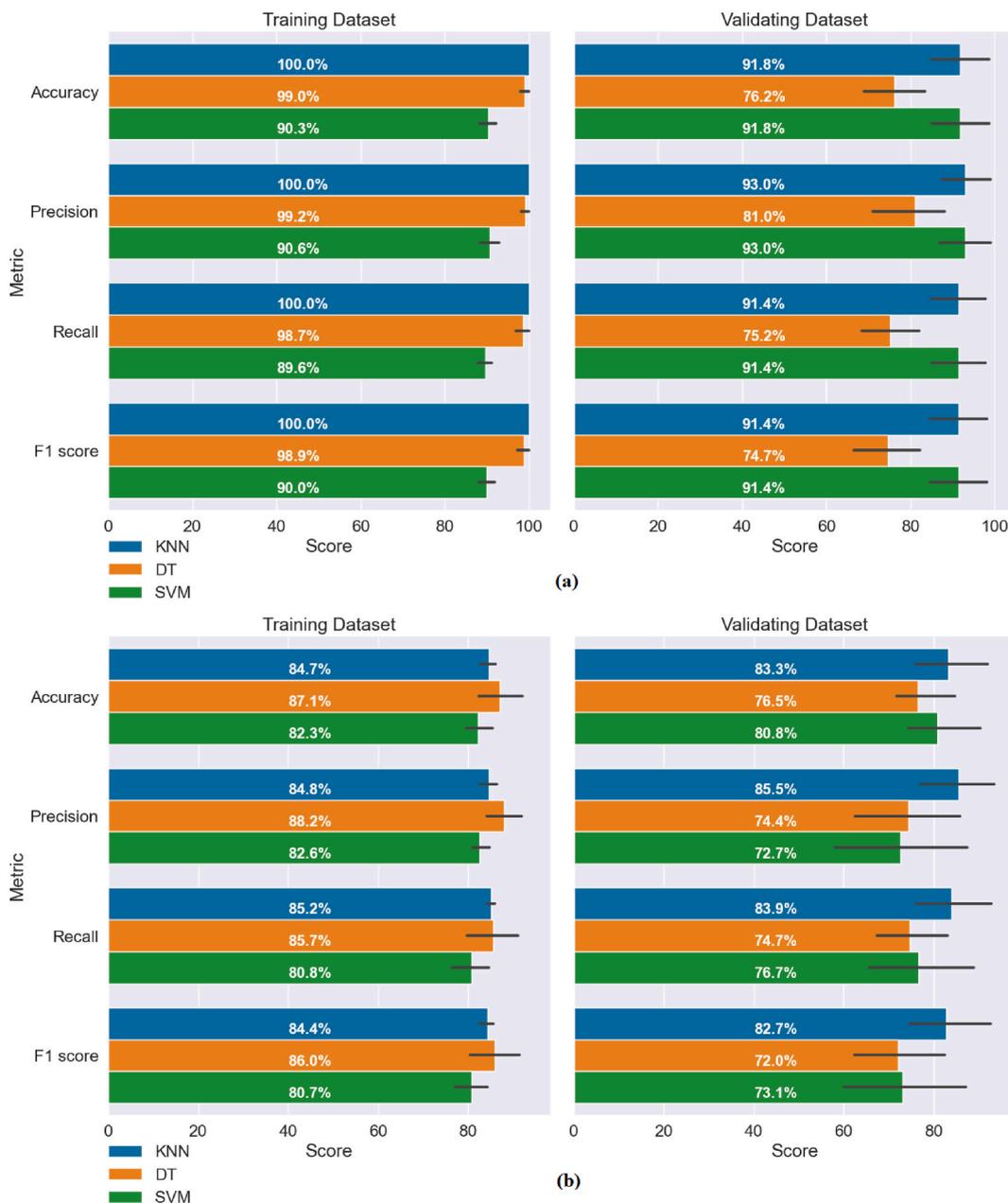
Fig. 11. Cross-validation results for three classifiers with (a) combination of 15 Mueller matrix images and 7 optical parameters dataset, and (b) 15 Mueller matrix images dataset, respectively.

Ref. [28], SVM classifier is more efficient and stable than KNN and DT classifier.

In general, the results of the proposed study confirm that by combination of polarimetric imaging and machine learning can support diagnosis of skin cancer. The polarimetric imaging provide a comprehensive data of skin cancer. The AI classification is employed to detect the skin cancer with high accuracy and less time consuming. The rapid detection results provide a fast and reliable post-information for decision-making process in skin cancer diagnosis. However, the study is still remain some limitations such as small dataset, using a cancer mimic mice skin tissue instead of human skin tissue, and the accuracy of classification not high enough. In the future, a classification will be performed based on human tissue, and thus provides a more comprehensive information for human skin cancer detection. While using human skin tissue for classification, the strictly samples preparation process is required. The errors from the polarimetric imaging and the operations of user are going to eliminated. Besides, some other uncertainty factors of cancer characteristics have to be careful considered. Finally, a larger and diverse dataset are collected to obtain the higher classification accuracy.
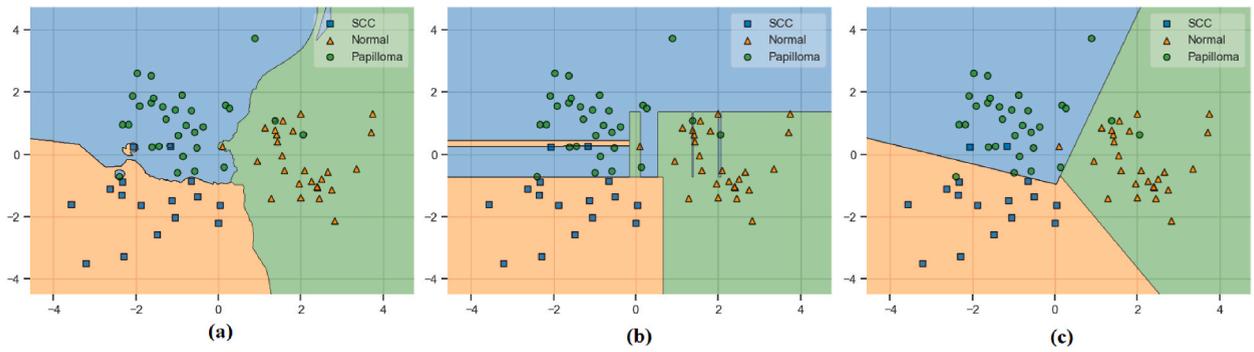
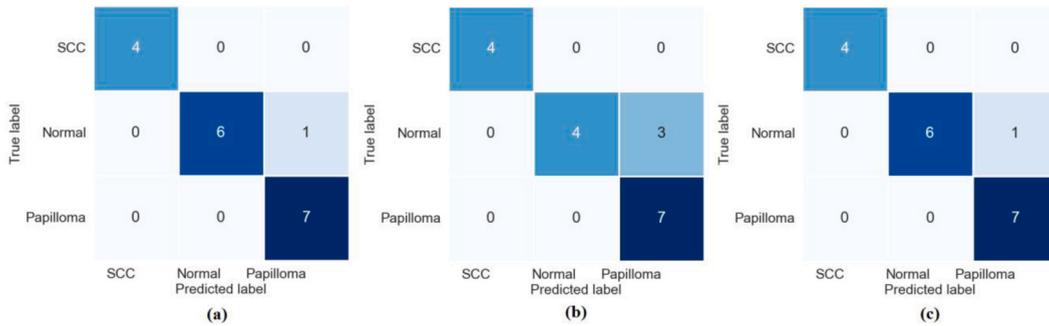**Fig. 12.** Decision surfaces of different classifiers on training dataset: (a) KNN, (b) DT, and (c) SVM.

Fig. 13. Confusion matrices of: (a) KNN, (b) DT, and (c) SVM models on testing dataset.

Table 4
Accuracy scores of three classifiers.

Classifier	Training accuracy score	Validating accuracy score	Testing accuracy score	Overfitting
KNN	100 %	91.8 %	94 %	Yes
DT	91.3 %	76.2 %	83 %	Yes
SVM	90.3 %	91.8 %	94 %	No

Table 5
Performance metrics of SVM model on testing dataset.

	Precision	Recall	F1 Score
SCC	1.00	1.00	1.00
Normal	1.00	0.86	0.92
Papilloma	0.88	1.00	0.93

6. Conclusion

This study has proposed a method for classifying the class (normal, papilloma and SCC) of non-melanoma mice skin cancer tissue using three classifiers (KNN, DT and SVM) based on the Mueller matrix elements and optical parameters acquired from the back-scattered light using a CCD camera and Mueller matrix imaging technique. The results have shown that, the difference in intensity and color of the element images increases as the cancer stage increases. In addition, the ANOVA results shown that the elements m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24}, m_{32}, m_{34}, m_{42}, m_{44}, and optical parameters α, D, R, each with a p-value lower than 0.05, are significant candidates for classification purposes. Furthermore, among the three models, the SVM classifier has the best robustness to overcome toward overfitting in the training stage and achieves the joint highest classification accuracy of the three schemes (94 %). Moreover, the SVM classifier achieves F1 scores of 0.92, 0.93 and 1.00 for normal, papilloma, and SCC skin tissue samples, respectively, and thus provides a useful tool for cancer staging purposes. Overall, the results presented in this study confirm that the optical properties of skin tissue samples vary in response to the changes brought about in the cell structure by tumor development. This is a premilary step for classification of skin cancer using mouse tissue sample. The results have shown that the machine learning methods which take the

Mueller matrix imaging and optical parameters as features for classification purposes provide an objective and reliable tool for skin cancer. In the future, the experiments on human skin tissue sample will be performed, different machine learning algorithms is explored incorporating additional imaging modalities, and conducting clinical trials for human validation. A more comprehensive results is provided for supporting the early-stage diagnosis of multiple skin cancers, including NMSC.

Availability of data and materials

The data used to generate the results are available in the Zenodo repository: https://doi.org/10.5281/zenodo.10065413.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Thi-Thu-Hien Pham: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Thanh-Ngan Luu:** Data curation, Formal analysis, Investigation, Methodology, Software. **Thao-Vi Nguyen:** Data curation, Validation, Visualization. **Ngoc-Trinh Huynh:** Investigation, Methodology, Resources, Validation. **Quoc-Hung Phan:** Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Thanh-Hai Le:** Data curation, Formal analysis, Investigation, Resources, Software, Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] J. Ferlay, et al., Cancer statistics for the year 2020: an overview, Int. J. Cancer (2021), https://doi.org/10.1002/ijc.33588.
[2] H. Sung, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, Ca - Cancer J. Clin. 71 (3) (2021) 209–249.
[3] A. Rembielak, T. Ajithkumar, Non-melanoma skin cancer – an underestimated global health threat? Clin. Oncol. 31 (11) (2019) 735–737.
[4] N. Eisemann, et al., Non-melanoma skin cancer incidence and impact of skin cancer screening on incidence, J. Invest. Dermatol. 134 (1) (2014) 43–50.
[5] P. Weber, et al., Dermatoscopy of neoplastic skin lesions: recent advances, updates, and revisions, Curr. Treat. Options Oncol. 19 (11) (2018) 56.
[6] C. Ring, N. Cox, J.B. Lee, Dermatoscopy, Clin. Dermatol. 39 (4) (2021) 635–642.
[7] A. Rajabi-Estarabadi, et al., Optical coherence tomography imaging of melanoma skin cancer, Laser Med. Sci. 34 (2) (2019) 411–420.
[8] L. Rey-Barroso, et al., Optical technologies for the improvement of skin cancer diagnosis: a review, Sensors 21 (1) (2021).
[9] M. Mogensen, G.B.E. Jemec, Diagnosis of nonmelanoma skin cancer/keratinocyte carcinoma: a review of diagnostic accuracy of nonmelanoma skin cancer diagnostic tests and technologies, Dermatol. Surg. 33 (10) (2007) 1158–1174.
[10] S. Arroyo-Camarena, et al., Spectroscopic and imaging characteristics of pigmented non-melanoma skin cancer and melanoma in patients with skin phototypes III and IV, Oncol. Ther. 4 (2) (2016) 315–331.
[11] M.R. Famielec, A. Borzecki, D. Krasowska, G. Chodorowska, Clinical usefulness of high-frequency ultrasonography in the monitoring of basal cell carcinoma treatment effects, Postep. Dermatologii i Alergol. 37 (3) (2020) 364–370.
[12] L. Rey-Barroso, et al., Visible and extended near-infrared multispectral imaging for skin cancer diagnosis, Sensors 18 (5) (2018) 1–15.
[13] S. Batz, C. Wahrlich, A. Alawi, M. Ulrich, J. Lademann, Differentiation of different nonmelanoma skin cancer types using OCT, Skin Pharmacol. Physiol. 31 (5) (2018) 238–245.
[14] L.F. di Ruffano, et al., Optical coherence tomography for diagnosing skin cancer in adults, Cochrane Database Syst. Rev. 12 (12) (2018), https://doi.org/10.1002/14651858.CD013189.
[15] A. Levine, O. Markowitz, Introduction to reflectance confocal microscopy and its use in clinical practice, JAAD Case Reports 4 (10) (2018) 1014–1023.
[16] N. Ghosh, I.A. Vitkin, Tissue polarimetry: concepts, challenges, applications, and outlook, J. Biomed. Opt. 16 (11) (2011), 110801.
[17] V.V. Tuchin, Polarized light interaction with tissues, J. Biomed. Opt. 21 (7) (2016), 071114.
[18] R.M.A. Azzam, Stokes-Vector and mueller-matrix polarimetry [invited], J. Opt. Soc. Amer. A 33 (7) (2016) 1396–1408.
[19] H. He, R. Liao, N. Zeng, P. Li, Z. Chen, X. Liu, H. Ma, Mueller matrix polarimetryan emerging new tool for characterizing the microstructural feature of complex biological specimen, J. Lightwave Technol. 37 (11) (2019) 2534–2548.
[20] A. Vahidnia, et al., Quantitative polarimetry Mueller matrix decomposition approach for diagnosing melanoma and non-melanoma human skin cancer, OSA Continuum 4 (11) (2021) 2862–2874.
[21] L. Jütte, et al., Registration of polarimetric images for in vivo skin diagnostics, J. Biomed. Opt. 27 (9) (2022), 096001, https://doi.org/10.1117/1.JBO.27.9.096001.
[22] H. Zhai, et al., Distinguishing tissue structures via polarization staining images based on different combinations of Mueller matrix polar decomposition parameters, Opt. Lasers Eng 152 (2022), 106955.
[23] X. Li, Y. Han, H. Wang, T. Liu, S.C. Chen, H. Hu, Polarimetric imaging through scattering media: a review, Front. Physiol. 10 (2022), 815296, 2022.
[24] D. Tua, R. Liu, W. Yang, L. Zhou, H. Song, L. Ying, Q. Gan, Imaging-based intelligent spectrometer on a plasmonic rainbow chip, Nat. Commun. 14 (1) (2023) 1902.

[25] C. Rodríguez, et al., Polarimetric data-based model for tissue recognition, Biomed. Opt Express 12 (2021) 4852–4872.

[26] D. Ivanov, et al., Polarization-based histopathology classification of ex vivo colon samples supported by machine learning, Front. Physiol. 9 (2022), 814787.

[27] K.M. Sindhoora, K.U. Spandana, D. Ivanov, E. Borisova, U. Raghavendra, S. Rai, S.P. Kabekkodu, K.K. Mahato, N. Mazumder, Machine-learning-based classification of Stokes-Mueller polarization images for tissue characterization, J. Phys. Conf. Ser. 1859 (1) (2021), 012045.

[28] T.N. Luu, Q.H. Phan, T.H. Le, T.T.H. Pham, Classification of human skin cancer using Stokes-Mueller decomposition method and artificial intelligence models, Optik 249 (2022), 168239.

[29] N.T. Luu, T.H. Le, Q.H. Phan, T.T.H. Pham, Characterization of Mueller matrix elements for classifying human skin cancer utilizing random forest algorithm, J. Biomed. Opt. 26 (7) (2021), 075001.

[30] J.R. Sambles, Polarized light in optics and spectroscopy, J. Mod. Opt. 38 (6) (1991) 1204–1205.

[31] J.S. Baba, et al., Development and calibration of an automated Mueller matrix polarization imaging system, J. Biomed. Opt. 7 (3) (2002) 341–349.

[32] V.T.T. Quyen, H.N. Thuy, H.M. Hien, H.N. Trinh, Antitumor-promoting effect of ethanolic extract from leaves of perilla frutescens var. crispa (benth.) in mouse skin, Med Pharm Res 2 (3) (2018) 33–39, https://doi.org/10.32895/UMP.MPR.2.3.33.

[33] M.M. Suarez-Alvarez, D.T. Pham, M.Y. Prostov, Y.I. Prostov, Statistical approach to normalization of feature vectors and clustering of mixed datasets, Proc. R. Soc. A Math. Phys. Eng. Sci. 468 (2145) (2012) 2630–2651.

[34] X.H. Cao, I. Stojkovic, Z. Obradovic, A robust data scaling algorithm to improve classification accuracies in biomedical data, BMC Bioinf. 17 (1) (2016) 1–10.

[35] Y.-L. Lo, T.-T.-H. Pham, P.-C. Chen, Characterization on five effective parameters of anisotropic optical material using Stokes parameters—demonstration by a fiber-type polarimeter, Opt Express 18 (9) (2010) 9133.

[36] T.T.H. Pham, Y.L. Lo, Extraction of effective parameters of turbid media utilizing the Mueller matrix approach: study of glucose sensing, J. Biomed. Opt. 17 (9) (2012), 0970021.

[37] G.J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, Wiley Series in Probability and Statistics, 2004, p. 552.

[38] D.L. James, C. Bezdek, Siew K. Chuah, Generalized k-nearest neighbor rules, Fuzzy Sets Syst 18 (3) (1986) 237–256.

[39] S.L. Salzberg, Book review: C4.5: programs for machine learning by J. Ross quinlan. Morgan kaufmann publishers, Inc., 1993, Mach. Learn. 16 (1994) 235–240, https://doi.org/10.1023/A:1022645310020.

[40] V. Kecman, "Basics of machine learning by support vector machines," in *real world app.* Comput. Intell. 103 (2005) 49–103.