Skin Melanoma Detection in Microscopic Images Using HMM-Based Asymmetric Analysis and Expectation Maximization

Rozita Rastghalam, Habibollah Danyali, Mohammad Sadegh Helfroush, M. Emre Celebi, and Mojgan Mokhtari

***Abstract*—Melanoma is one of the deadliest types of skin cancer with increasing incidence. The most definitive diagnosis method is the histopathological examination of the tissue sample. In this paper, a melanoma detection algorithm is proposed based on decision-level fusion and a Hidden Markov Model (HMM), whose parameters are optimized using Expectation Maximization (EM) and asymmetric analysis. The texture heterogeneity of the samples is determined using asymmetric analysis. A fusion-based HMM classifier trained using EM is introduced. For this purpose, a novel texture feature is extracted based on two local binary patterns, namely local difference pattern (LDP) and statistical histogram features of the microscopic image. Extensive experiments demonstrate that the proposed melanoma detection algorithm yields a total error of less than 0.04%.**

***Index Terms*—Melanoma, Local Binary Pattern, Hidden Markov Model-based EM, Decision-level fusion.**

I. Introduction

SKIN MELANOMA is considered to be the most aggressive type of skin cancer, given its rising occurrence and mortality in recent decades [1]. Skin is the largest organ, consisting of two main layers. The upper layer near the surface of the skin is called "epidermis," while the layer under the epidermis near the skin's fat tissue is called "dermis." The epidermis consists of melanin-producing cells known as melanocytes, which remain in the epidermis under normal conditions. The movement of melanocytes to lower layers of the skin indicates a nevus or melanoma [2].

Manuscript received June 1, 2020; revised September 26, 2020, December 26, 2020, and March 7, 2021; accepted April 20, 2021. Date of publication May 18, 2021; date of current version September 3, 2021. This work was supported by the National Science Foundation under Grant No. 1946391. *(Corresponding author: M. Emre Celebi.)*

Rozita Rastghalam, Habibollah Danyali, and Mohammad Sadegh Helfroush are with Elcetrical Engineering, Shiraz University of Technology, Shiraz, Fars, 71557-13876 (e-mail: rrozita_rrastghalam@yahoo.com; danyali@sutech.ac.ir; ms_helfroush@sutech.ac.ir).

M. Emre Celebi is with the Department of Computer Science, University of Central Arkansas, Conway, AR 72035 USA (e-mail: ecelebi@uca.edu).

Mojgan Mokhtari is with the Isfahan University of Medical Sciences, Isfahan 8174673441, Iran (e-mail: mojganmokhtari80@gmail.com).

Digital Object Identifier 10.1109/JBHI.2021.3081185

Histopathological examination is the most definitive technique to distinguish between nevi and melanomas. To diagnose melanoma, pathologists investigate numerous features such as mitosis, melanocyte locations, nucleus size, and morphology [3]. The distance between the epidermis and the deepest melanocytes permeated into the dermis determines the disease's Breslow depth [4]. Investigating microscopic images is time-consuming. Furthermore, the multiplicity of features, Breslow staging, and incomplete biopsies make the process more challenging [5]. Individual experience dependency and disagreement among pathologists on some samples are the main reasons for using CAD systems to classify microscopic images [6]–[9]. Numerous studies evaluate microscopic images to assist in melanoma diagnosis [10]–[22]. These studies can be generally divided into two major groups: (1) those aiming to detect cells or identify regions of interest in microscopic images [10]–[13], and (2) those aiming to distinguish between nevi and melanomas.

The halo around the melanocytes is used in most studies to distinguish melanocytes from other keratinocytes [13]–[15]. The algorithms in the first group of studies contain two main steps. All cells in the image are identified, and then melanocytes are recognized from the detected nucleus. In [10], a bank of generalized Laplacian of Gaussian kernels is proposed for detecting all cells in the image. In addition, adaptive thresholding [12], mean-shift [12], and multi-pass adaptive voting are applied to detect all cells in the image[16]. In these studies, special gradient intensity and shape information [10], radial line scanning [13], template matching [12], special descriptor [11], and multi-scale radial line scanning [13] are proposed to differentiate melanocytes from other keratinocytes, based on the halo region. In the second group of studies, various methods are proposed for pathologists to distinguish between nevi and melanomas [15]–[22]. Most of these methods include three main steps: segmentation, feature extraction, and classification. For example, Otsu's thresholding method and k-means are proposed for epidermis segmentation [14], [15], [17].

Feature extraction is the most crucial component of a melanoma detection algorithm [14], [15]. Histograms [17], Gray Level Co-occurrence Matrix (GLCM) [15], [17], morphological operation [14], [15], color features [18], [19] and statistical parameters [14], [15] are used to extract features from pathology images. In the current literature, various methods are proposed

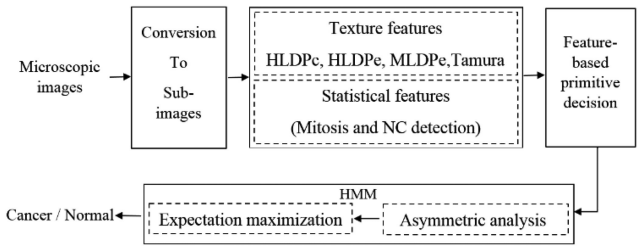

Fig. 1. Block diagram of the proposed melanoma detection (SMTD-LDP-HMM) approach.

to distinguish between nevi and melanomas. These include classification and regression trees (CART) [17], minimum distance, and support vector machine (SVM) [14], [15]. Deep learning techniques have also been used. For instance, in [20], a Fully Convolutional Network is proposed for melanoma detection. In [21], an epidermis segmentation technique is proposed based on a Convolutional Neural Network (CNN). CNNs are utilized for comparison of deep learning with the classification results for randomly cropped images, with the ground truth determined by 11 practicing histopathologists [22]. Finally, a sparse Convolutional Autoencoder is proposed for nucleus detection in histopathology tissue images in [23]. It should be noted that deep learning is also used for automatic nuclei segmentation in microscopic images of various organs [24]–[26]. For example, in [24], [26], the authors propose a convolutional neural network for nuclei segmentation. Despite their promising results in recent studies, however, deep learning techniques require large amounts of data to achieve sufficient accuracy [27].

The abovementioned studies mainly aim to detect melanocytes or classify images into normal and abnormal groups. However, tumor location determination in microscopic images is also vital in Breslow staging and choosing an appropriate treatment. For this purpose, we propose a melanoma detection algorithm based on decision level fusion using a Hidden Markov Model (HMM). In the proposed algorithm, called skin melanoma tumor detection using local difference pattern (LDP) and HMM (SMTD-LDP-HMM(, microscopic images are examined based on blocks or sub-images, and novel texture features are defined based on the local binary pattern (LBP) descriptor. For each sub-image, the probability of malignancy is computed based on the extracted features and using an HMM approach that is trained by expectation maximization (EM). Finally, cancerous blocks and, consequently, melanoma tumor locations are determined.

The remainder of this paper is organized as follows. Section 2 explains the proposed skin melanoma tumor detection (SMTD-LDP-HMM), providing details about feature extraction, primitive decision, and decision-level fusion-based classification using HMM. The experimental results are presented in Section 3. Finally, Section 4 concludes the paper.

II. The Proposed Melanoma Detection Algorithm

Fig. 1 shows the block diagram of the proposed SMTD-LDP-HMM. As the microscopic images are too large, each image is divided into several sub-images. This increases both classification accuracy and processing speed. These sub-images have the same size: $600 \times 1,300$ pixels. Since the full sizes of the microscopic images vary, different numbers of sub-images are extracted from each microscopic image. All sub-images that exist in each microscopic image are extracted and often chosen without overlaps. The proposed LDP-based texture features, as well as some statistical features, are extracted from each sub-image. The other parts of the proposed SMTD-LDP-HMM algorithm involve asymmetric analysis to identify homogeneous textures as well as an EM-based HMM to categorize images based on fusion.

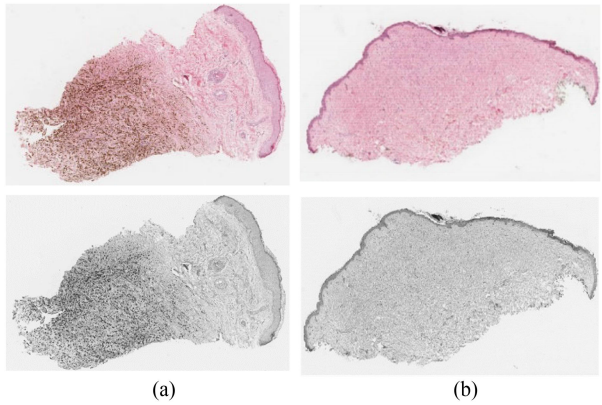

Fig. 2. Microscopic images, (a) Melanoma case, (b) Normal case.

Microscopic images of a melanoma biopsy and a healthy skin sample are shown in Fig. 2(a) and Fig. 2(b), respectively. The increase in pigmentation, mitosis, and cell size create heterogeneous images. Pigmentation is seen as brown spots, and black points are mitosis or large nuclei. Our proposed approach utilizes graylevel microscopic images. Thus, all images are converted to graylevel using the Rec. 709 formula [28]:

$$gray\ scale\ =\ 0.21R + 0.71G + 0.07B \qquad (1)$$

where R, G, and B represent the values of the red, green, and blue bands of the input color image, respectively.

A. Feature Extraction

Feature extraction is the most important component of CAD systems. By extracting numerous features from a microscopic image, we aim to identify characteristics such as mitosis, large nuclei, nuclear-cytoplasmic ratio, and nuclei count. For extracting these characteristics, which may indicate the presence of a tumor, two different types of features are extracted, as described below.

1) Textural Features: Texture feature extraction from microscopic images has proven to be effective for classification purposes [14], [15], [17]. The present study uses three effective techniques to extract texture features from microscopic images. Additionally, some modifications are proposed to improve the performance of these tools, as discussed in detail below.

- Local graylevel difference pattern (LDP)

Introduced in 2002 by Ojala et al. [30], the LBP is a popular texture descriptor. A specific neighborhood is considered for

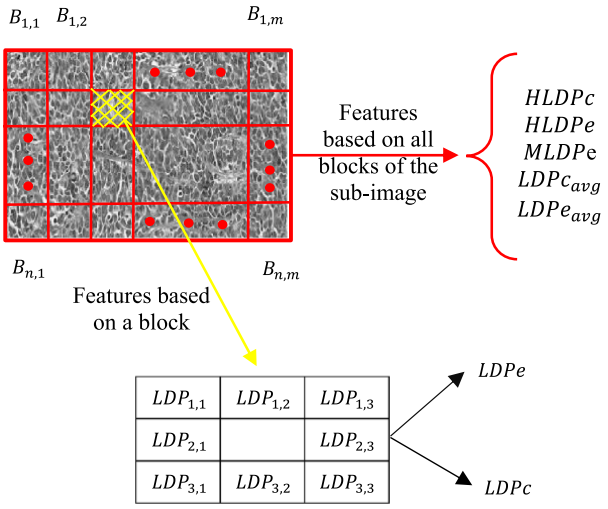

Fig. 3. Extracted features based on the proposed LDP.

each pixel, and the center pixel is specified as the index in each neighborhood in the LBP operator. Based on the following equation, a binary pattern is obtained from each neighborhood.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right) 2^p \; s\left(x\right) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{2}$$

where g_c is the graylevel of the central pixel, $g_p$is the graylevel of one of its neighbors, P is the number of neighbors, R is the neighborhood radius, and $s(.)$ is the standard binary output function [31].

In texture classification, LBP spectrum histograms are used for training a classifier and then classifying textures. However, the LBP model in equation 2 is not rotation-invariant, and its spectrum is also high-dimensional. Direct use of LBP does not usually yield a good texture classification performance. Hence, researchers have been working to improve LBP for texture classification [31]–[33]. For instance, in [31], using median sampling, four descriptors are used to encode a pixel (central gradient, radial gradient, magnitude gradient, and tangent gradient) to generate the initial gradient LBP pattern.

In this study, we introduce a novel method based on the LBP, which is more effective on microscopic images than the original LBP formulation. A novel method with a special focus on image edge and graylevel variations is proposed for extracting texture features from microscopic images. High-frequency features or image edges exist in normal as well as cancerous microscopic images. Therefore, the graylevel differences in neighbor pixels should be explored more rigorously. The edges in normal images only have slight gray level differences whereas in the cancerous cases, the edges often exhibit significant gray level changes. In the primary LBP descriptor, if a pixel in a neighborhood exceeds its index (even by a margin of 1), the corresponding binary number of that pixel will be 1 (as shown in Fig. 4). However, this approach is not effective for extracting texture features from microscopic images because the differences in the graylevels of the pixels in a neighborhood are not determined with sufficient accuracy.

Therefore, in our proposed algorithm, we consider the difference of each pixel relative to the central pixel in a neighborhood. As such, a local graylevel difference pattern (LDP) is extracted from each neighborhood, as opposed to an LBP.

Fig. 3 shows the extracted features based on the proposed LDP. As shown, the sub-image is first partitioned into 3×3 non-overlapping blocks. Let $B_{1,1}$ to $B_{n,m}$ represent the blocks in a sub-image. The LDP is then computed for each block as shown in Fig. 4. Two local features, namely LDPe (edge LDP) and LDPc (color LDP), are extracted from the LDP of each block. Some global features such as histogram of the LDPc (HLDPc), histogram of the LDPe (HLDPe), $LDPc_{avg}$ (average of LDPcs), $LDPe_{avg}$ (avrage of LDPes), and MLDPe (most frequent LDPe) are defined based on the LDP of all blocks in the sub-image (see Fig. 3). Each block has to go through two steps to determine the values of $LDP_{1,1}$ to $LDP_{3,3}$. The LDP is determined for each block of the sub-image, as follow, (3) shown at bottom of next page.

In Fig. 4, we introduce our eight categories and illustrate how to determine the LDP for each block in two steps. The measurement differences are divided into eight categories covering the graylevel range of 0 to 256. As shown, the absolute value of the difference between each pixel and the central pixel is placed in a new matrix. Each new pixel is then assigned to the smallest category whose value exceeds the pixel's graylevel. For instance, pixel 50 in the new matrix belongs to category 6, since it is the first category with a value greater than 50. Likewise, pixel 25 corresponds to category 5 because 32 is the first number that is greater than 25. After determining the LDP values, we can now compute the LDPc and LDPe.

As stated in Eq. (4), the maximum difference between the LDP values in a block is called the local graylevel difference pattern-edge (LDPe). The LDPe extracts high-frequency image information in each block. As shown in Fig. 4, the maximum and minimum values of the LDP in that block are 6 and 1, respectively. Therefore, based on Eq. (4), the LDPe for this block equals 5.

Low-frequency image information in each block, on the other hand, is extracted based on the local graylevel difference pattern-color (LDPc). The LDPc value is determined based on the LDP value and the eight categories introduced earlier. As shown in Fig. 4, the $LDP_{1,1}$ is 3, thus the corresponding LDPc is 8. The LDPe and LDPc are computed for each block of the sub-image as follows:

$$LDPe \;=\; MAX\left(LDP\right) - MIN\left(LDP\right) \tag{4}$$

$$LDPc \;=\; \sum_{i,j} 2^{LDP\;i,j} \tag{5}$$

where i, j represent the indices of the eight LDPs shown in Fig. 3. Thus, the LDPe and LDPc are computed for each block of the sub-image. The values of the LDPe and LDPc can range from 0 to 7 and 16 to 2048, respectively. After determining the LDPe and LDPc, the histograms of these two features are computed. In this study, the LDPc and LDPe histograms are abbreviated as the HLDPc and HLDPe, respectively. The number of HLDPe bins is 8, as the range of the LDPe is 0 to 7. LDPc values range

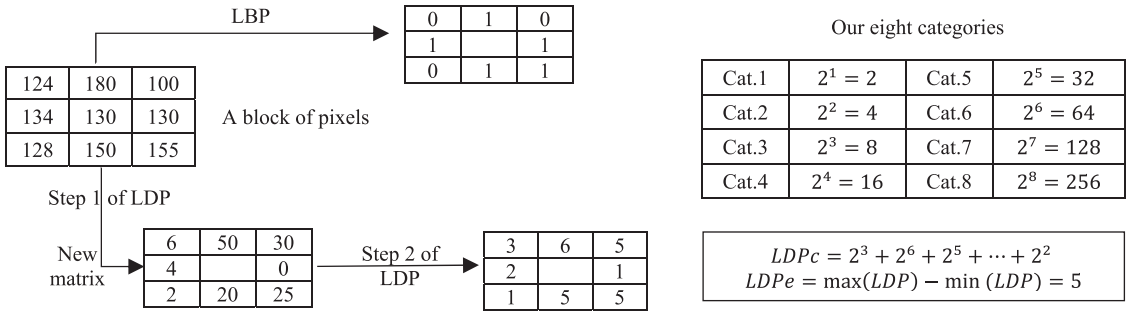

Fig. 4. Illustration of the LDPe and LDPc computations.

from 16 to 2048. To construct the HLDPc, we divide the whole range of LDPc values into 254 bins (from 1 to 254), where each bin includes successive values of LDPc. For example, bin "1" includes the LDPc values between 16 to 23. LDPc between 24 and 31 are counted and shown in the second HLDPc element (bin "2"). In the same manner, intervals of 8 are considered so that the histogram range can be mapped to the range of 1 to 254. Therefore, the HLDPc has 254 bins. Based on the HLDPe and HLDPc, we propose two effective texture features. The $LDPe_{avg}$ is computed based on the HLDPe using the following equation, which shows the average of LDPe for each sub-image, numerically:

$$LDPe_{avg} = \sum_{k=0}^{7} HLDPe\,(k) \cdot k \tag{6}$$

where $k = 0, 1, 2, \ldots, 7$ represents the index of an HLDPe bin in the sub-image. The $LDPc_{avg}$, on the other hand, is computed in each sub-image using the following equation:

$$LDPc_{\,avg} = \sum_{l=1}^{254} HLDPc\,(l)\,.\,(l) \tag{7}$$

where l is the the index of an HLDPc bin in the sub-image. The MLDPe is formulated to detect the most frequent LDPe in each sub-image. MLDpe is defined based on the histogram peak of LDPe, that is, the most frequent element of the HLDPe:

$$MLDPe = argmax_k\ (HLDPe) \tag{8}$$

where k is the index of an HLDPe bin, as before.

- Tamura and GLCM as texture features

The GLCM is a tabulation of the different combinations of pixel brightness values occurring in an image. It is a two-dimensional array representing second-order statistics based on neighboring pixels [34]. GLCM considers the specific position of a pixel relative to the other pixels. This 8×8 matrix is constructed at a distance of $d = 1$ for various directions given as $0^\circ,\ 45^\circ, 90^\circ, 180^\circ$. Contrast, Correlation, Entropy, Energy, and Homogeneity are the five features related to the GLCM matrix in this paper [35]. As the pixels of the normal sub-image are similar to each other, and the same is true for the pixels in the cancer sub-image, correlation is high in both normal and cancer sub-images. As shown in algorithm 1, Contrast and correlation are merged, and G1 is the result of the feature-based fusion of these two textural features.

As described in Algorithm 1 (ContCorr Merge), if the contrast is less than t_1, the image is normal; if the contrast exceeds t_2, the image is classified as cancerous regardless of the amount of correlation. If the contrast value of an image falls between t_1 and t_2, it is necessary to check the correlation value. In this case, if the correlation is smaller than the T threshold, the image is normal, otherwise the image is cancerous. t_1, t_2 and T are three thresholds determined using the training database. Thus, contrast and correlation are applied to each image simultaneously in order to minimize correlation errors. Therefore, G1, G2, and G3 are the three features based on GLCM, of which G2 and G3 represent homogeneity and energy, respectively.

Tamura [36] proposed six textural features and gave descriptions common over all the texture patterns in Brodatz's photographic images. These texture features include Coarseness, Contrast, Directionality, Line-Likeness, Regularity, and Roughness.

Coarseness is an effective feature for the purpose of the present study. Coarseness called T is the last texture feature. Coarseness is related to the distance in graylevels of spatial variations, which is implicitly related to the size of the primitive elements forming the texture. It directly depends on scale and repetition rates, and most fundamental texture features. An image contains repeated texture patterns at different scales. Coarseness aims to identify the largest size of repeated texture patterns [37].

At each pixel, $p(x, y)$, six averages are computed for the windows of size $k = 0, 1, 2, \ldots, 5$ around the pixel, as displayed in Fig. 5(a). Then, the absolute difference is determined for each pixel at each scale, i.e., $E_k(x, y)$, as shown in Fig. 5(b) and the following equations:

$$E_{k,a}\ (P) = \left|A_k^1 - A_k^2\right| \tag{9}$$

$$E_{k,b}\ (P) = \left|A_k^3 - A_k^4\right| \tag{10}$$

$$LDP = [LDP_{1,1}, LDP_{1,2}, LDP_{1,3}, LDP_{2,1}, LDP_{2,3}, LDP_{3,1}, LDP_{3,2}, LDP_{3,3}] \tag{3}$$

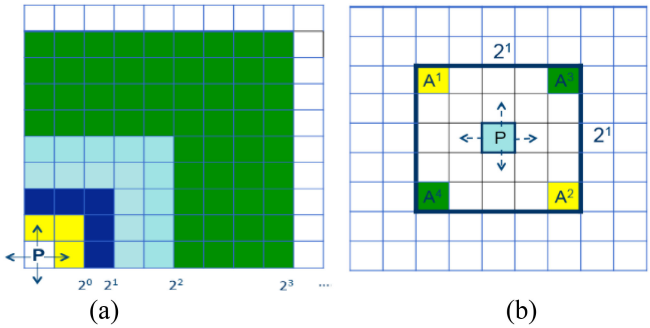

Fig. 5. Coarseness computation, (a) different window size, (b) E(x,y) computation between non- overlapping pairs on opposite sides.

Algorithm 1: ContCorr Merge (*contrast, correlation*, *T*, t_1,t_2).

initialize status $\leftarrow \theta$.
if *contrast* $<t_1$
status $\leftarrow$"normal"
end if
if t_1 $<$*contrast*$<t_2$
  if correlation$<T$
    Status $\leftarrow$"normal"
  else
    Status $\leftarrow$*"melanoma"*
  end if
end if
if *contrast* $\geq t_2$,
Status $\leftarrow$*"melanoma"*
end if
return *Status*.

The window with size *k* that maximizes $E_k(x, y)$ in either direction is identified such that it determines the largest variation between *k* existing scales for each pixel. The best pixel window size, which produces the largest variation in each scale (k), is $S_{(best)}$, which equals 2^k. Finally, coarseness is computed by averaging $S_{(best)}$ over the entire image.

2) Statistical Features: Nuclear-cytoplasmic ratio (NC), nucleus size, and cytoplasm volume are significant features investigated in pathology science for disease detection. Pathologists check these features at multiple magnifications to determine nucleus size and NC. However, we extracted the above features from microscopic images with a single magnification. As mentioned earlier, microscopic images are divided into sub-images in order to maximize classification accuracy. Nucleus size, NC, and cytoplasm volume are determined in each sub-image. The nucleus becomes larger and darker during mitosis and cell division [8]. A normal cell nucleus is very small and insignificant [36]. Therefore, prominent mitosis feature showing cancer in pathological images is darker than the small and pale nucleus in normal cells. Increasing nucleus size or mitosis will maximize NC ratio. Various studies [24], [25], [37], focus on nucleus segmentation in pathology images. However, in this paper, we propose the use of graylevel features and histogram statistical moments to detect nucleus size, NC, and mitosis presence in each sub-image.

The nucleus and cytoplasm pixels are counted separately in each sub-image. To determine graylevel features, the cytoplasm and nucleus areas are identified in training samples with the help of an expert. The nuclei colors of the training images are identified and the range of color changes was used as a color feature to detect the nuclei. Therefore, the pixels of each sub-image whose graylevel was below 55 were counted in order to determine the number of nuclei and the presence of mitoses. The presence of mitoses enlarges the nucleus, resulting in an increase in the number of dark pixels in the image. Furthermore, the presence of pigments in melanoma images leads to an increase in the number of dark pixels in these images. Moreover, the numerical range of 200 to 255 is identified based on cytoplasm color in the training images. Thus, to determine the amount of cytoplasm in each sub-image, pixels with a graylevel exceeding 200 were counted. The NC criterion is calculated by dividing the number of dark pixels with a graylevel of less than 55 by the number of pixels in the sub-image with a graylevel of more than 200. Thus, the NC ratio is computed based on the proportion of nucleus pixels to cytoplasm pixels:

$$NC = \frac{Nuc}{Cyt} \tag{11}$$

where Nuc and Cyt represent the number of the nucleus and cytoplasm pixels, respectively.

In addition, histogram entropy and its peak are computed to determine the distribution of pixels of each sub-image between the nucleus and cytoplasm. For this purpose, histogram entropy of each sub-image is merged with its histogram peak as feature-level fusion. Shannon entropy is computed from a suitably normalized histogram as follows:

$$\mathrm{E} = \sum -p_i \log p_i \tag{12}$$

B. Feature-Based Primitive Decision

Using a classifier, each of the extracted features can be used to determine whether the images are cancerous or normal. However, this decision is not reliable as it is made using only one feature of the image. Each feature is evaluated before entering the classification stage, using a simple classifier, as well as training samples. Our Decision-level, Fusion-based classification merges all the features and makes the best decision for each sub-image, as described in the next section. As such, a primitive linear decision boundary is determined for each feature, called (T_i), where i is the number corresponding to each feature. Initially, sub-images are divided into testing and training datasets to select a primitive linear decision boundary for each feature. The success and error rates of each feature in the training dataset are specified using primitive linear boundary classification. Therefore, the number of normal samples predicted as normal (TN), the number of samples wrongly detected as cancerous (FP), the number of cancer cases diagnosed as cancerous (TP), and the number of cancerous cases incorrectly detected as normal (FN) are determined by each feature and its corresponding primitive boundary (T_i). Each extracted feature from the sub-image is

compared to its corresponding primitive decision boundary, and each feature represents a normal or cancerous decision for that sub-image based on its primitive decision boundary. Next, this feature vector, all of whose elements can be normal, cancerous, or a combination of normal and cancerous, is ready for fusion and classification, based on our Decision-level Fusion-based HMM Classification.

C. Decision-Level Fusion-Based Classification Using HMM

Pathologists investigate heterogeneous portions of a sample with higher accuracy because heterogeneity and asymmetry in a sample texture increase malignancy probability. Hence, an HMM approach is proposed, which considers both homogeneity and similarity between each sub-image and their neighborhoods, and the features extracted from each sub-image. In [38], an HMM-based approach is proposed for Decision-level Fusion-based classification with a simpler scenario. While in this study, a more comprehensive and modified version, called DFC-EM, is proposed to perform Decision-level Fusion-based classification (DFC), using an HMM trained using EM.

HMM is defined by state probabilities, transition probabilities, emission probabilities (B_h and B_m matrices), and initial probabilities [39]. In our model, transition and emission probabilities are trained based on asymmetric analysis and expectation maximization, respectively, while transition probabilities in [38] are kept constant, and the emission probabilities are not trained. Thus, the advantages of DFC_EM over simple HMM are as follows:

- Sub-images are compared to their neighborhoods based on asymmetric analysis, and the transition matrix is developed for each sub-image. Thus, homogeneous segments are identified for each microscopic image, while the transition matrix in [38] was ineffective in HMM classifier.
- By training the elements of B_h and B_m using the EM approach, it is possible to identify more effective features, which will have a higher impact on the outcome of the proposed DFC-EM. However, this was not considered in [38], and all features contributed equally in the final decision, as evident in the random forest method.

The N states of the model and the M observation symbols per state are defined by $s = \{s_1, s_2, \ldots, s_N\}$ and $V = \{v_1, v_2, \ldots, v_M\}$, respectively. If the observations are continuous, M is infinite. The state transition probability distribution is $A = \{a_{ij}\}$, where a_{ij} are defined as follows:

$$a_{ij} = p\{\, q_{t+1} = j | \, q_t = i\}, \; 1 \le i, j \le N \tag{13}$$

where q_t, denotes the current state and a_{ij} represent the probability that the state is s_j at time $t+1$, assuming that at time t, the state is s_i. The structure of this stochastic matrix defines the connection structure of the model. Transition probabilities should satisfy the normalization constraints:

$$a_{ij} \ge 0, 1 \le i, j \le N, \sum_{j=1}^{N} a_{ij} = 1, 1 \le i \le N \tag{14}$$

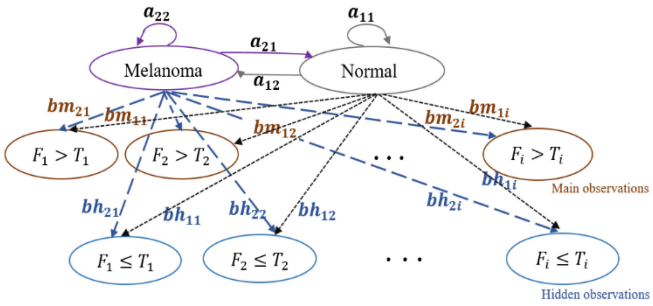

Fig. 6. The proposed DFC-EM model.

The observation symbol probability distribution in each state is shown by $B = \{b_j(k)\}$, where $b_j(k)$ is the probability that the symbol v_k is emitted in the state s_j. The proposed DFC-EM model consists of two normal and melanoma states, as well as a number of observations, whose count equals that of the extracted features from the image. Thus, i, which describes the number of features, is set to 10 in this study. Our proposed DFC-EM model is shown in Fig.6. The next step involves determining the relevant parameters.

1) Asymmetric Analysis-Based Transition Matrix: As mentioned earlier, asymmetric analysis is used to determine the transition matrix elements. Each training sample consists of 5 or more sub-images, each of which is compared to its neighborhood using asymmetric analysis. Features of each sub-image are compared to the corresponding features extracted from the neighboring sub-image. The similarity of adjacent sub-image features is determined based on different similarity measures. While a_{11} and a_{22} indicate the number of similar features in each neighboring sub-image, a_{12} and a_{21}show the number of dissimilar features in matrix *A*. Therefore, the diagonal elements of matrix *A* determine homogeneity, and the remaining elements show heterogeneity.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{15}$$

Therefore, inhomogeneous segments of each microscopic image are identified using the transition matrix. Inhomogeneous segments of each sample are identified in this step, while completely homogeneous samples without any changes or marking will move forward to the next classifier step. The next HMM classifier determines whether these marking segments are cancerous or normal. For instance, a melanoma sample consists of normal and cancerous sub-images. The location of the tumor can be approximated using matrix *A* based on the heterogeneity and asymmetry of these sub-images.

2) Expectation Maximization-Based Emission Matrix: The main stage of the proposed DFC is based on emission probability matrices. A total of 10 extracted features are possible observations in the proposed DFC-EM. As shown in Fig. 6, there are two types of observations in the proposed DFC-EM model called main and hidden observations. Therefore, two emission probability matrices are introduced. The main observations are greater than their corresponding primitive decision boundaries,

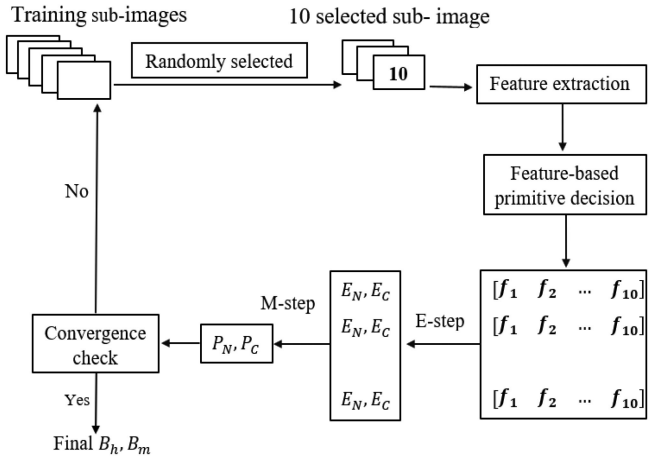

Fig. 7. Estimation of the HMM parameters using EM.

but hidden observations are less than or equal to their corresponding primitive decision boundaries. In training images in [38], the emission probability matrices are set based on the numbers of FN, FP, TP, TN, for each feature.

In this paper, Expectation-Maximization (EM) algorithm is used to learn and determine the parameters of the probability matrices. The EM algorithm is an iterative method to compute the maximum-likelihood estimates for model parameters when the examined data are incomplete, have missing data points, or include unobserved latent variables. EM works by choosing an initial guess for the model parameters in order to create a probability distribution. This is sometimes called the "E-Step" for the "Expected" distribution. In the maximization step (M-step), complete data generated after the expectation (E) step are used to update the parameters. These E- and M-steps are repeated until convergence.

As previously mentioned, the observation symbol probability distribution in each state is in matrix *B* of HMM, known as the emission probabilities matrix. There are two types of observations in our proposed HMM and B_h and B_m are introduced as emission probabilities matrices. The B_h and B_m matrices used in [38] for the final decision are used here as the initial guess in our EM-based HMM. As stated, emission probabilities in B_h and B_m matrices are obtained based on the effectiveness of each feature in the training samples based on the proposed HMM-EM. Fig. 7 displays the iteration of different steps based on the EM approach until convergence of the desired parameters.

B_h and B_m contain the emission probabilities of both states to hidden and main observations, respectively. The first row of these matrices represent the normal state, and the second shows the melanoma state. They are specified as:

$$B_m = \begin{bmatrix} b_{m11} & b_{m12} & b_{m13} & \dots & b_{m1n} \\ b_{m21} & b_{m22} & b_{m23} & \dots & b_{m2n} \end{bmatrix} \tag{16}$$

$$B_h = \begin{bmatrix} b_{h11} & b_{h12} & b_{h13} & \dots & b_{h1n} \\ b_{h21} & b_{h22} & b_{h23} & \dots & b_{h2n} \end{bmatrix} \tag{17}$$

After determining the initial guess for each parameter, the normal and cancerous probability of observations should be computed in each iteration based on the EM algorithm. The following equations describe the process:

$$P\ (N|O) = \frac{P(\mathrm{O}|\mathrm{N}) \cdot P\,(N)}{P(\mathrm{O}|\mathrm{N}) \cdot P\,(N) + P\,(O|C) \cdot P\,(C)}, \quad 0 \le P(N|O) \le 1 \tag{18}$$

$$P\ (C|O) = \frac{P(\mathrm{O}|\mathrm{C}) \cdot P\,(C)}{P(\mathrm{O}|\mathrm{C}) \cdot P\,(C) + P(O|N) \cdot P\,(N)}, \quad 0 \le P\,(C|O) \le 1 \tag{19}$$

where O, $P(C)$, and $P(N)$ are possible observation, melanoma probability, and normal probability, respectively. In the first iteration of the EM algorithm, values of the parameters are determined by initial guess. Thus, the values of $P(C)$ and $P(N)$ are considered based on B_h and B_m matrices, in the first iteration of our HMM classifier training through the EM algorithm. The two main steps of the EM algorithm, i.e., the E and M steps, must be performed for training the B_h and B_m matrices parameters. Therefore, *k* images are randomly selected from the training set in each iteration, and expectation is computed for each image as follows:

$$E_N P\ (N|O) = P\,(N|O) \cdot n \tag{20}$$

$$E_C P\ (C|O) = P\,(C|O) \cdot c \tag{21}$$

where n and c represent the number of observations that are determined to be normal or cancerous based on the primitive decision among the ten possible observations, respectively. sIn the M-step, the updated P_N and P_C are obtained as follows:

$$P_N = \frac{\sum_{i=1}^{k} E_N \cdot P\,(N|O_i)}{\sum_{i=1}^{k} E_N \cdot P\,(N|O_i) + \sum_{i=1}^{k} E \cdot P\,(N|O_i)} \tag{22}$$

$$P_C = \frac{\sum_{i=1}^{k} E_c \cdot P\,(N|O_i)}{\sum_{i=1}^{k} E_C \cdot P\,(N|O_i) + \sum_{i=1}^{k} E_N \cdot P\,(N|O_i)} \tag{23}$$

All the steps mentioned above are repeated until probabilities P_N and P_C converge, as shown in Fig. 7.

III. Experimental Results

Thirty skin microscopic images obtained from [40] and [41] were converted to 280 sub-images and then used to test the proposed melanoma detection algorithm. These microscopic images are of different sizes, with a minimum size of approximately $9,500 \times 12,000$ pixels. Therefore, each microscopic image is converted to different numbers of sub-images. All samples are stained using the Hematoxylin-Eosin (H&E) method. It should be noted that our algorithm can seamlessly separate melanoma cases from normal ones despite the fact that the microscopic images are obtained from different sources. Microscopic images at 20× magnification are investigated in this study. All images are 8-bit, normalized, and converted to graylevel in the first step of the algorithm. Since these images are very large, each image is divided into a number of sub-images of size $600 \times 1,300$ pixels. The number of sub-images in each microscope image

TABLE I
COMPARISON OF THE PROPOSED LDP AND LBP FOR TEXTURE FEATURE EXTRACTION IN MICROSCOPIC IMAGES

	Measures of location/dispersion								
	Mean			Standard deviation			Skewness		
	LBP	LDPc	LDPe	LBP	LDPc	LDPe	LBP	LDPc	LDPe
TP	114	131	130	120	129	127	102	125	121
TN	99	116	112	105	113	113	89	114	116
FP	31	14	18	25	17	17	41	16	14
FN	36	19	20	30	21	23	48	25	29

TP: true positive, TN: true negative, FP: false positive, FN: false negative, LBP: local binary pattern, LDPc: proposed local difference pattern-color, LDPe: proposed local difference pattern-edge
E.o.LBP: the number of error based on LBP, E.o.LDPe: the number of error based on LBP, E.o.LDPc: the number of error based on LBP

depends on the original image size. Partitioning the microscopic image to sub-images has increased the accuracy of the sample examination and decisions, especially in the study of nucleus size and mitosis detection.

A. Local Graylevel Difference Pattern Result

As mentioned earlier, the differences in the graylevels of the pixels are not determined with sufficient accuracy using the primary LBP descriptor. Therefore, the LDP is proposed to perform local graylevel difference pattern extraction. The LDPe and LDPc are proposed to extract high and low-frequency information of each neighborhood, while the edge intensity and graylevel intensity features are represented by their histograms, HLDPe and HLDPc. In Table I, the proposed LDPe and LDPc are compared with the original LBP descriptor. As expected, the results of the LDPe and LDPc are more accurate than the LBP since the LDP captures the textural details better. As shown in Table I, measures of location/dispersion such as mean, skewness, and standard deviation are computed for this evaluation.

The LDPe histograms (HLDPe) are displayed for the cancerous and normal sub-images in Fig. 8(a). According to Fig. 8, the more frequent edges for normal and cancerous cases are related to LDPe(1) and LDPe(4), respectively. This represents stronger edges in cancer images. Histograms of LDPc (HLDPc) are shown for a normal and a cancerous sub-image in Fig. 8(b). The uniform LDPc histogram for cancerous cases explains the presence of different graylevels in cancerous cases tissue, while HLDPc has limited graylevels in normal sub-images due to cytoplasm multiplicity.

B. Histogram-Based Statistical Feature Result

As mentioned earlier, in order to determine the cytoplasm and number of nuclei in each sub-image, the number of graylevel pixels between 0 to 50 and 200 to 250 is determined using the histogram of each sub-image, respectively. The proportion of pixels with a graylevel of [0–50] to pixels with a graylevel of [200–250] is considered as the CN criterion. In addition to counting pixels related to cytoplasm and nucleus, the histogram peak, along with its entropy, is proposed in this paper. The reason

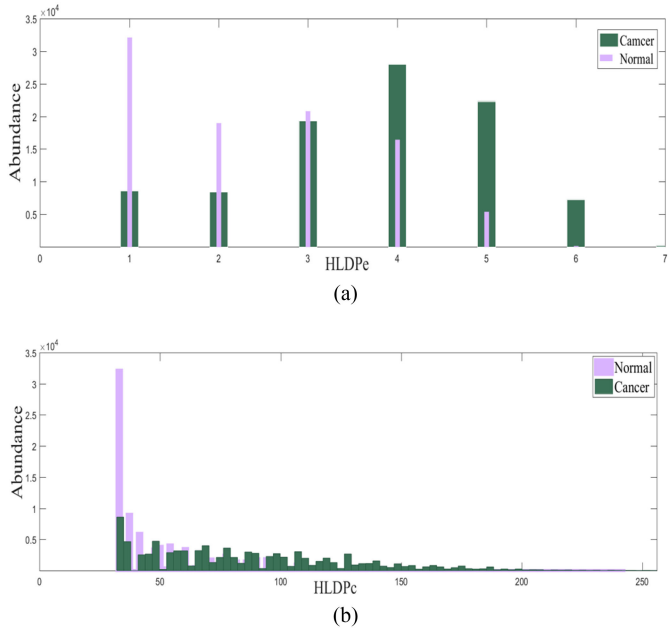

Fig. 8. Histograms of LDPe and LDPc, (a) HLDPe, (b) HLDPc.

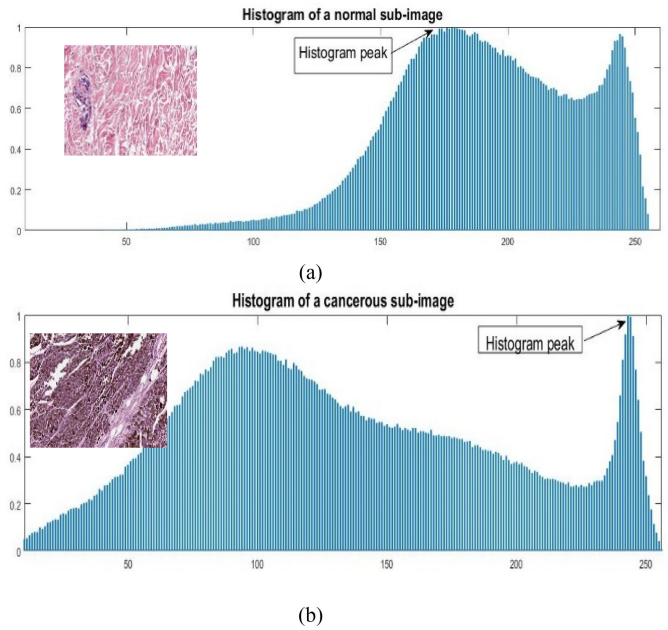

Fig. 9. sub-images and their histogram, (a) normal and (b) melanoma.

for the fusion of histogram and its entropy is displayed in Fig. 9. Sometimes the histogram peak of a melanoma sample occurs in the histogram range of 200 to 250, which represents the cytoplasm abundance (a feature of normal samples), as shown in Fig. 9(b). Thus, histogram entropy and its peak reduce FN, and the total number of errors as displayed in Table II. Since reducing FN errors is more important than FP errors, the simultaneous use of sub-images histogram peak and their entropy has been significant for the training sample. Fig. 9(a) shows a normal sample that has defined melanoma incorrectly through histogram peak alone.

TABLE II
RESULTS OF SIMULTANEOUS USE OF SUB-IMAGE HISTOGRAM AND ITS ENTROPY

Image. no	Hist.peak	Hist.entropy	Hist&entropy
TP	123	118	133
TN	112	109	115
FP	18	21	15
FN	27	32	17
TOE	45	53	32

TABLE III
RESULTS OF THE PRIMITIVE DECISION BY OUR PROPOSED LDP

Performance measures		Textural features (LDP)		
		F1	F2	F3
Training images	TN	51	52	50
	TP	73	71	74
	FN	7	9	6
	FP	9	8	10
	TOE	16	17	16
Test image	TN	60	59	61
	TP	62	61	64
	FN	8	9	6
	FP	10	11	9
	TOE	18	20	15

TABLE IV
RESULTS OF THE PRIMITIVE DECISION BASED ON GLCM& TAMURA

Performance measure		GLCM&Tamura			
		F4	F5	F6	F7
Training images	TN	53	48	54	48
	TP	67	67	71	63
	FN	13	13	9	17
	FP	7	12	6	12
	TOE	20	25	15	29
Test image	TN	62	57	63	59
	TP	58	57	62	58
	FN	12	13	8	12
	FP	8	13	7	11
	TOE	20	26	15	23

C. Feature-Based Primitive Decision Result

In this paper, 14 out of 30 microscopic images are used as training images. 14 training images contain 140 sub-images. Leave-one-out (LOO) cross-validation is used to evaluate the accuracy of primitive boundary selection. This evaluation scheme repeatedly divides the training dataset into several parts. In each iteration, an algorithm is trained on the full dataset, excluding only one part, and then performs the test on the excluded part. Each part contains whole sub-images of a training microscopic image. Therefore, in each iteration, one training image, which contains several sub-images, participates in the algorithm test step. LOO is applied on 14 microscopic images (140 sub-images) of the training dataset to select the primitive decision boundary for each feature. Tables III, IV and V show the classification result by each feature and its corresponding decision boundary on the test and training images.

F1 through F3 in Table III are local graylevel difference pattern-based texture features that are defined based on the LDP. Columns F4 to F7 in Table IV correspond to other texture features described in Section 2.1.1.2. Table V display the result of histogram-based statistical features, which are the number of the nucleus, NC, and fused histogram entropy with histogram peak (F8 to F10).

TABLE V
RESULTS OF THE PRIMITIVE DECISION BY STATISTICAL FEATURES

Performance measure		statistical		
		F8	F9	F10
Training images	TN	51	52	50
	TP	70	70	73
	FN	10	10	7
	FP	9	8	10
	TOE	19	18	17
Test image	TN	61	63	62
	TP	61	64	63
	FN	9	6	7
	FP	9	7	8
	TOE	18	13	15

D. The Overall Result of the Proposed SMTD

As mentioned, there are two decision steps in the proposed DFC-EM for each sub-image. First, the homogeneity of each sample is determined based on the transition matrix (A). The transition matrix describes the similarity of features in each sub-image and its neighborhood. Therefore, matrix (A) is different for each sub-image and is obtained without any previous training for each test sample. In the second step of the proposed DFC, the emission matrix B is computed iteratively through training to determine the probability of being normal and cancerous in each sub-image. The value of k is 10 in each EM iteration. The two emission matrices (B_h and B_m) are modified after EM training as follows (24) and (25) shown at bottom of next page.

The EM algorithm applied to our training dataset has two advantages. Suppose the first observation (or first feature (F1)) is error-free while the other observations have one or more errors in one EM iteration for all ten sub-images. Thus, F1 should be given more importance compared to other observations. This is done in EM. In addition, the number of correctly predicted observations is essential for the final decision for each sub-image, which is considered in the EM training approach. Therefore, the number of anticipated normal or cancerous observations has influenced our HMM classifier, and each observation has a different effect, depending on its performance.

Several criteria, including negative predictive value (NPV), positive predictive value (PPV), sensitivity (SEN), specificity (SPC), and Hammoude distance (HM), are used to evaluate the proposed algorithm. These criteria are defined as follows:

$$NPV = \frac{TN}{TN + FN} \tag{26}$$

$$PPV = \frac{TP}{TP + FP} \tag{27}$$

$$SEN = \frac{TP}{TP + FN} \tag{28}$$

$$SPC = \frac{TN}{TN + FP} \tag{29}$$

TABLE VI
DETECTION RESULTS OF OUR DECISION-LEVEL FUSION ALGORITHM BY MEANS OF HMM

	Performance measures									
	TP	TN	FP	FN	TOE	HM	PPV	NPV	SEN	SPC
Fusion by HMM [40]	64	67	3	6	9	0.13	0.95	0.92	0.91	0.96
Fusion by HMM & EM	66	69	1	4	5	0.07	0.98	0.94	0.94	0.98

TABLE VII
RESULTS OF THE PROPOSED DFC ALGORITHM ON FIVE SELECTED IMAGE SAMPLES

No. class	Possible observations										HMM[40]		HMM&EM	
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	$p(N)$	$p(C)$	$p(N)$	$p(C)$
1.N	N	N	N	N	N	N	N	N	N	N	0.84	0.16	100	0
2.C	C	C	C	C	C	C	C	C	C	C	0.14	0.86	0	100
3.N	N	N	N	C	C	C	C	N	N	N	0.57	0.40	0.90	0.09
4.C	C	C	N	C	N	N	N	C	C	N	0.49	0.52	0.07	0.92
5.N	N	N	C	C	C	N	C	C	N	N	0.46	0.49	0.60	0.38

C: Cancer, N: Normal, $**p**(N)$: $**p**(N/O)$, $**p**(C)$: $**p**(C/O)$

$$HM = \frac{FN + FP}{TN} \tag{30}$$

In Table VI, the proposed DFC-EM algorithm is compared with a simple HMM classifier introduced in 2016 [40], based on TP, TN, FP, FN, TOE, NPV, PPV, SEN, and SPC. As seen, the errors are reduced in our proposed algorithm by training HMM parameters using EM.

Furthermore, the proposed method is more effective for simultaneous fusion and classification. According to Tables III, IV and V, the fewest errors are seen in F6, but the proposed decision-level fusion reduces the number of errors to five. The extracted features are merged, and each sub-image is declared normal or cancerous based on both proposed DFC-EM and simple HMM [40]. However, as presented in Table VI, decision-level fusion-based classification using proposed DFC-EM had just five errors, while the number of errors was nine in [40]. Thus, training HMM parameters has had significant effects on our DFC_EM model, leading to more accurate results than untrained HMM [40]. The total error by the proposed method is reduced to less than 4%.

The decision based on the proposed DFC algorithm is obtained for five selected samples in Table VII, and the images of the selected samples are displayed in Fig. 10. The result

TABLE VIII
COMPARISON RESULTS OF OUR DECISION-LEVEL FUSION ALGORITHM BY MEANS OF HMM

	HMM [40]	CNN [23]	HMM&EM
Accuracy	0.93	0.88	0.96

of the proposed DFC-EM and simple HMM [40] are also specified in Table VII. According to this table, DFC-EM is more sensitive and accurate than the untrained simple HMM. Finalizing a decision is difficult for the sub-images in which the number normal and cancer observations are equal (4.C and 5.N). The proposed DFC-EM has performed better than the untrained HMM [40] in difficult conditions. According to case 5.N in Table VII, the final decision is correctly declared normal, based on the proposed DFC-EM, while the untrained HMM [40] is confused about this difficult case. Therefore, the accuracy of decision and classification increases by training the HMM parameters. In Table VIII, the proposed DFC-EM algorithm and simple HMM classifier [40] are compared with a CNN, introduced in 2019 [23]. In [23] the accuracy was 68% based on a CNN for melanoma detection. The first to third columns of Table VIII, display the result of applying three different algorithms to our database, namely, simple HMM, CNN, and the proposed HMM-EM. As specified, the proposed DFC-EM has more accuracy than the other methods. However, this does not mean that CNN or the other deep-learning techniques are ineffective in melanoma detection. Using a much larger database or an effective data augmentation strategy can increase CNN accuracy. This study uses a pretrained ResNet50 convolutional neural network (CNN). The ResNet architecture relies on the concept of skip connections as a mechanism to prevent loss of information when the deep network is being trained. Using this concept, very deep networks can be trained that significantly improve model performance [23]. Residual connections have a considerable advantage in ResNet: the knowledge obtained during the training process is maintained by the connections and the network has a higher capacity, which leads to faster training. ResNet50 has a 3-layer bottleneck block. These layers consist of 1×1, 3×3, and 1×1 convolutions, and the 1×1 layers first decrease and then increase (restore) the dimensions. As a result, the 3×3 layer has a bottleneck with smaller input/output dimensions. Individual learning rates were used for each layer, which is different from other approaches that apply the same rate to all layers in the CNN. More specifically, we used slower learning rates for input-adjacent layers, but increased the rate for output-adjacent layers [23]. Finally, to classify our test set, we used 140 sub-images (80 melanomas and 60 normal) to train

$$B_h = \begin{bmatrix} 0.84 & 0.88 & 0.88 & 0.08 & 0.16 & 0.76 & 0.28 & 0.92 & 0.76 & 0.08 \\ 0.16 & 0.12 & 0.12 & 0.92 & 0.84 & 0.24 & 0.72 & 0.08 & 0.24 & 0.92 \end{bmatrix} \tag{24}$$

$$B_m = \begin{bmatrix} 0.14 & 0.10 & 0.14 & 0.86 & 0.90 & 0.10 & 0.86 & 0.19 & 0.05 & 0.77 \\ 0.86 & 0.90 & 0.86 & 0.14 & 0.10 & 0.9 & 0.14 & 0.81 & 0.95 & 0.23 \end{bmatrix} \tag{25}$$

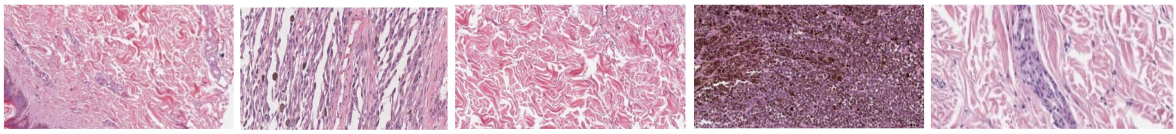

Fig. 10. Image corresponding to Table IV, (a) 1.N, (b) 2.C, (d) 3.N, (e) 4.C, (f) 5.N.

the system. The test set contains 140 sub-images (70 melanomas and 70 normal) and is disjoint from the training set.

IV. Conclusion

This paper presented an automatic technique for melanoma detection in skin microscopic images. The proposed technique first extracts texture and statistical histogram features of the microscopic sub-image and then estimates tumor location based on asymmetric analysis and the extracted features. The final decision is made using a fusion-based HMM classifier, which is trained using the EM method for each sub-image. EM is utilized to optimize the parameters of the classifier, which merges the extracted features and makes an optimal decision for each sub-image simultaneously. According to our findings, melanoma can be detected using texture features and fusion-based HMM classifier in each pathology sample. Nuclei graylevel parameters were determined based on a small number of cases. Future work includes making these parameters adaptable and the use of additional cellular features. We also aim to investigate Long Short-Term Memory (LSTM) networks in microscopic sub-images to improve feature extraction for more effective melanoma detection and Breslow staging.

References

[1] J. Smolle, "Computer recognition of skin structures using discriminant and cluster analysis," *Ski. Res. Technol.*, vol. 6, pp. 58–63, 2000.
[2] K. Robertson and J. L. Rees, "Variation in epidermal morphology in human skin at different body sites as measured by reflectance confocal microscopy," *Acta Derm. Venereol.*, vol. 90, pp. 368–373, 2010.
[3] R. King, K. N. Weilbaecher, G. McGill, E. Cooley, M. Mihm, and D. E. Fisher, "Microphthalmia transcription factor: A sensitive and specific melanocyte marker for melanoma diagnosis," *Amer. J. Pathol.*, vol. 155, pp. 731–738, 1999.
[4] A. A. Marghoob, K. Koenig, F. V. Bittencourt, A. W. Kopf, and R. S. Bart, "Breslow thickness and clark level in melanoma," *Cancer*, vol. 88, pp. 589–595, 2000.
[5] P. S. Colloby, K. P. West, and A. Fletcher, "Observer variation in the measurement of breslow depth and clark's level in thin cutaneous malignant melanoma," *J. Pathol.*, vol. 163, pp. 245–250, 1991.
[6] N. Stathonikos, M. Veta, A. Huisman, and P. van Diest, "Going fully digital: Perspective of a dutch academic pathology lab," *J. Pathol. Inform.*, vol. 4, pp. 15, 2013.
[7] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
[8] O. Jimenez-Del-Toro *et al.*, "Analysis of histopathology images: From traditional machine learning to deep learning," in *Biomedical Texture Analysis: Fundamentals, Tools and Challenges*, Academic Press, 2017, pp. 281–314.
[9] M. E. Celebi, N. Codella, and A. Halpern, "Dermoscopy image analysis: Overview and future directions," *IEEE J. Biomed. Heal. Inform.*, vol. 23, no. 2, pp. 474–478, Mar. 2019.
[10] H. Xu, C. Lu, R. Berendt, N. Jha, and M. Mandal, "Automatic nuclear segmentation using multiscale radial line scanning with dynamic programming," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 10, pp. 2475–2485, Oct. 2017.
[11] H. Xu, C. Lu, and M. Mandal, "An efficient technique for nuclei segmentation based on ellipse descriptor analysis and improved seed detection algorithm," *IEEE J. Biomed. Heal. Inform.*, vol. 18, no. 5, pp. 1729–1741, Sep. 2014.
[12] C. Lu, Z. Ma, and M. Mandal, "Automated segmentation of the epidermis area in skin whole slide histopathological images," *IET Image Process,* vol. 9, pp. 735–742, 2015.
[13] C. Lu, M. Mahmood, N. Jha, and M. Mandal, "Detection of melanocytes in skin histopathological images using radial line scanning," *Pattern Recognit,* vol. 46, pp. 509–518, 2013.
[14] C. Lu, and M. Mandal, "Automated analysis and diagnosis of skin melanoma on whole slide histopathological images," *Pattern Recognit,* vol. 48, pp. 2738–2750, 2015.
[15] H. Xu, C. Lu, R. Berendt, N. Jha, and M. Mandal, "Automated analysis and classification of melanocytic tumor on skin whole slide images," *Comput. Med. Imag. Graph.*, vol. 66, pp. pp. 124–134, 2018.
[16] C. Lu, H. Xu, J. Xu, H. Gilmore, M. Mandal, and A. Madabhushi, "Multi-Pass adaptive voting for nuclei detection in histopathological images," *Sci. Rep.*, vol. 6, 2016, Article number: 33985.
[17] M. Wiltgen, A. Gerger, and J. Smolle, "Tissue counter analysis of benign common nevi and malignant melanoma," *Int. J. Med. Inform.*, vol. 69, pp. 17–28, 2003.
[18] M. Mokhtari, M. Rezaeian, S. Gharibzadeh, and V. Malekian, "Computer aided measurement of melanoma depth of invasion in microscopic images," *Micron*, vol. 61, pp. 40–48, 2014.
[19] H. Xu, H. Wang, R. Berendt, N. Jha, and M. Mandal, "Computerized measurement of melanoma depth of invasion in skin biopsy images," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2017, pp. 17–20.
[20] A. Phillips, I. Y. H. Teo, and J. Lang, "Fully convolutional network for melanoma diagnostics," 2018, *arXiv1806.04765*.
[21] K. R. J. Oskal, M. Risdal, E. A. M. Janssen, E. S. Undersrud, and T. O. Gulsrud, "A U-net based approach to epidermal tissue segmentation in whole slide histopathological images," *SN Appl. Sci.*, vol. 1, pp. 1–12, 2019.
[22] A. Hekler *et al.*, "Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images," *Eur. J. Cancer*, vol. 118, pp. 91–96, 2019.
[23] L. Hou *et al.*, "Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images," *Pattern Recognit,* vol. 86, pp. 188–200, 2019.
[24] L. Xie, J. Qi, L. Pan, and S. Wali, "Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei," *Neurocomputing*, vol. 376, pp. 166–179, 2020.
[25] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2421–2433, Oct. 2015.
[26] H. Jung, B. Lodhi, and J. Kang, "An automatic nuclei segmentation method based on deep convolutional neural networks for histopathology images," *BMC Biomed. Eng.*, vol. 1, no. 1, pp. 1–12, 2019.
[27] D. Komura and S. Ishikawa, "Machine learning methods for histopathological image analysis," *Comput. Struct. Biotechnol. J.*, vol. 16. The Authors, pp. 34–42, 2018.
[28] M. Anderson, R. Motta, S. Chandrasekar, and M. Stokes, "Proposal for a standard default color space for the internet - sRGB," 1996.
[29] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[30] L. Ji, Y. Ren, G. Liu, and X. Pu, "Training-Based gradient LBP feature models for multiresolution texture classification," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2683–2696, Sep. 2018.
[31] L. Ji, Y. Ren, X. Pu, and G. Liu, "Median local ternary patterns optimized with rotation-invariant uniform-three mapping for noisy texture classification," *Pattern Recognit,* vol. 79, pp. 387–401, 2018.

[32] L. Ji, M. Chang, Y. Shen, and Q. Zhang, "Recurrent convolutions of binary-constraint cellular neural network for texture recognition," *Neurocomputing*, vol. 387, pp. 161–171, 2020.
[33] R. Rastghalam, and H. Pourghassem, "Breast cancer detection using spectral probable feature on thermography images," in *Proc. Iranian Conf. Mach. Vis. Image Process.*,, 2013, pp. 116–120.
[34] J. C. Kavitha, and A. Suruliandi, "Texture and color feature extraction for classification of melanoma using SVM," in *Proc. Int. Conf. Comput. Technol. Intell. Data Eng.*, 2016.
[35] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst. Man Cybern.*, vol. 8, no. 6, pp. 460–473, Jun. 1978.
[36] G. Paolino, M. Donati, D. Didona, S. Mercuri, and C. Cantisani, "Histology of non-melanoma skin cancers: An update," *Biomedicines*, vol. 5, no. 4, pp. 71, 2017.
[37] A. L. Barbieri, O. Fadare, L. Fan, H. Singh, and V. Parkash, "Challenges in communication from referring clinicians to pathologists in the electronic health record era," *J. Pathol. Inform.*, vol. 9, 2018.
[38] R. Rastghalam, and H. Pourghassem, "Breast cancer detection using MRF-based probable texture feature and decision-level fusion-based classification using HMM on thermography images," *Pattern Recognit,* vol. 51, pp. 176–186, 2016.
[39] Y. Matsuyama, "Hidden markov model estimation based on alpha-EM algorithm: Discrete and continuous alpha-HMMs," in *Proc. Int. Jt. Conf. Neural Netw.*, 2011, pp. 808–816.
[40] "Spectrum webscope: AP residents\Skin\Slide collection," Accessed: Dec. 17, 2019 [Online]. Available: http://fomvss349p4.med.ubc.ca:82/AP residents/Skin/Slide collection/view.apml?
[41] "Skin," Accessed: Dec. 22, 2019 [Online]. Available: https://www.biomax.us/tissue-arrays/Skin