Pattern Recognition 117 (2021) 107982

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Rough-Bayesian approach to select class-pair specific descriptors for HEp-2 cell staining pattern recognition

Debamita Kumar, Pradipta Maji*

Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata, 700108, West Bengal, India

ARTICLE INFO

Article history:
Received 30 September 2019
Revised 13 March 2021
Accepted 31 March 2021
Available online 8 April 2021

Keywords:
HEp-2 cell images
Staining pattern recognition
Texture analysis
Rough sets
Bayes decision theory

ABSTRACT

One of the important problems in computer-aided diagnosis of connective tissue disease is automatic recognition of staining patterns present in HEp-2 cells. In this regard, the paper introduces a novel approach for the recognition of staining patterns by HEp-2 cell indirect immunofluorescence image analysis. The proposed method assumes that a fixed set of local texture descriptors or scales may not be effective for classifying staining patterns into multiple classes. A particular set of descriptors or scales may be significant for classifying a pair of classes, but may not be relevant for other pairs of classes. The proposed approach, therefore, first selects a set of local texture descriptors under appropriate scales for each class-pair, and then forms the final feature set for multiple classes from the relevant descriptors of all possible pairs of classes. A novel framework, termed as Rough-Bayesian model, is introduced to evaluate the relevance of a descriptor and/or a scale. It is based on the merits of rough sets and Bayes decision theory. During the selection of relevant descriptor and/or scale, the proposed method takes care of the presence of both noisy pixels in an HEp-2 cell image and noisy HEp-2 cell images in a staining pattern class. The support vector machine is used to predict the staining patterns present in HEp-2 cell images. The performance of the proposed method, along with a comparison with state-of-the-art methods, is demonstrated on several HEp-2 cell image databases. An important finding is that the accuracy for classifying HEp-2 cell images is significantly increased if class-pair specific descriptors under appropriate scales are considered, instead of selecting a uniform set of descriptors and scales for multiple classes.

1. Introduction

Connective tissue disease (CTD) is a class of disorders, where target of pathology is the connective tissue of the body. The CTDs such as scleroderma, rheumatoid arthritis, systemic lupus erythematosus, and mixed CTDs are characterized by the presence of spontaneous overactivity of the immune system that results in the production of extra antinuclear antibodies (ANAs) in blood of patients [1]. The indirect immunofluorescence (IIF)-ANA test is commonly used to detect ANAs in the patient serum, for the diagnosis of several CTDs. The most used substrate for ANA test is human epithelial type 2 (HEp-2) cells. In IIF, images of a patient's serum are acquired by staining a biological tissue with antibodies that are tied to a fluorescent chemical compound; they bind to other specific antibodies (which are target of the assay), when illuminated producing a fluorescence image observable through a microscope [1]. Each image contains many cells, and represents the distribution of target antibodies inside the tissue. The patterns observable in this distribution are the main information used for the diagnosis, since each pattern can be related to a specific CTD. As the staining pattern is independent of each cell nucleus or cell including cytoplasm, each cell needs to be evaluated separately [2].

In general, the analysis of HEp-2 images includes classification of fluorescence intensity [3], mitotic cell detection [4], segmentation of HEp-2 cells [5], as well as staining pattern recognition of HEp-2 cells [2]. The HEp-2 cell allows identification of more than thirty different nucleolar, nucleoplasmic and cytoplasmatic staining patterns, which are given by upwards of hundred different ANAs. Usually, Homogeneous, Nucleolar, Centromere, Cytoplasmic, Fine and Coarse Speckled, Nuclear Membrane and Golgi are the common IIF-ANA patterns, considered for the diagnosis of CTDs [6]. Most of the distinctions between HEp-2 patterns are based on texture, as it carries important information about the surface of the HEp-2 cells. For HEp-2 cell pattern classification, it is necessary to understand how physicians discriminate between different texture types of HEp-2 cell images. The textural properties of an HEp-2 cell image are likely to provide valuable information for classification, where different cell regions are treated as different texture classes. The classification of these cell images is necessary in order to identify different ANAs.

* Corresponding author.
E-mail addresses: debamita_r@isical.ac.in (D. Kumar), pmaji@isical.ac.in (P. Maji).

https://doi.org/10.1016/j.patcog.2021.107982
0031-3203/

In this background, a significant amount of research has been undertaken in last few years, to recognise the staining patterns present in HEp-2 cell images [2,7]. Different global texture descriptors based on gray level co-occurrence matrix are used in [8–10], while edge orientation histograms, rotation-invariant Gabor features, and modified Zernike moments are also found to provide important information of an HEp-2 cell image [8]. In [10–12], the well-known SIFT descriptor [13] has been used, while the DCT based descriptor has been applied in [9,12]. Faraki et al. [14]. used covariance descriptor obtained from a bank of Gabor filters, while Kong et al. [15]. applied maximum-response filter banks and histogram of gradients for HEp-2 cell texture analysis. The concept of bag of features model is used in [12,14], for the generation of cell signature. In [8], it has been shown that the local variations of intensity patterns are more effective than global information, obtained from the HEp-2 cell images, in differentiating various staining pattern classes. Several local texture descriptors, namely, local binary pattern (LBP) [16], rotation-invariant LBP (LBP^{ri}) [17], rotation-invariant uniform LBP (LBP^{riu2}) [17], co-occurrence of adjacent LBP (CoALBP) [18], completed LBP (CLBP) [19], rotation-invariant CoALBP (RICLBP) [20], median robust extended LBP [21], noise tolerant LBP [22], pairwise rotation-invariant co-occurrence LBP [23] and local ternary pattern [24], can be used to characterize a cell image. Some of them have already been found to be successful for HEp-2 cell pattern analysis [8]. In order to encode gradient and textural characteristics of the HEp-2 patterns, the concept of gradient-oriented co-occurrence of LBP has been introduced in [11], while the dominant LBP has been presented in [25] for texture analysis.

Recently, there has been a growing interest in the development of deep learning models for feature extraction from HEp-2 cell images [26,27]. Several deep neural network models have also been developed [28–32] for HEp-2 cell staining pattern recognition, due to their ability to achieve high classification accuracy. In [28], a deep convolutional residual in residual network has been proposed for simultaneous HEp-2 specimen image segmentation and classification. The deep residual network has been applied in [29] for classifying HEp-2 cell images, while a deep feature extraction method, based on convolutional auto-encoders, has been introduced in [30] for staining pattern recognition. A deeper architecture with convolutional neural network (CNN) is developed in [31], and several preprocessing techniques have been applied in [32] to enhance the performance of deep CNN in identifying staining patterns present in HEp-2 cell images. However, the training of these deep learning models is computationally very expensive due to the large parameter space. To alleviate this problem, various deep architectures have recently been introduced in [33,34], which enable building of deep architectures with comparatively smaller parameter space. In [33], the deep CNN has been used to classify HEp-2 cell staining patterns, whereas the features extracted by shallow and deep multi-scale convolutional component are fused in [34] for performance improvement. However, the limited number of training images causes inappropriate learning that may result into inaccurate classification of the HEp-2 cell images. Hence, data augmentation is necessary for these models to identify staining patterns properly.

The intrinsic textures of various HEp-2 cell types are fairly different from each other. The HEp-2 cell stained images also have unpredictably ambiguous texture. This difficulty exists in both inter-class and intra-class examples. Moreover, due to the nature of human cell, the shared visual similarity of inter-class samples, like the texture of certain parts of the tissues, further increases the ambiguity. These difficulties limit the accuracy of HEp-2 cell classification. Also, there exists uncertainty due to inexactness, vagueness and incompleteness in HEp-2 class definition, as well as overlapping characteristics of HEp-2 cell classes. In this regard, the theory of rough sets can be used to handle uncertainties associated with HEp-2 cell pattern classes. It provides an effective means for the analysis of data by constructing upper and lower approximations of set concepts from the acquired data, based on information granules [35]. The probabilistic modeling, on the other hand, aims to measure the belongingness of an object to a specific class, based on a probability distribution. As rough sets and probabilistic models are complementary in nature, they can be integrated for modeling uncertainty inherent in classification of staining pattern classes of HEp-2 cells.

Another important aspect of texture is scale. The appearance of most textures is changed when viewed at different scales. According to psychovisual studies, the human visual system processes images in multiple scales; thus capable of preserving both local and global information. The HEp-2 cell images may contain information over a large range of scales. In medical imaging perspective, the scale of the imagery may be different in many cases, and so it is important to understand how information changes over different scales of imagery. Hence, a fixed set of scales or descriptors may not be effective for classifying staining patterns present in HEp-2 cell images into multiple classes. A particular set of descriptors may be significant for classifying a pair of classes, but may not be relevant for other pairs of classes. Similarly, a uniform set of scales may be effective for classifying a pair of classes, but may not be significant for another pair of classes. However, the existing approaches reported in [8,11,20,25] consider a uniform set of local texture descriptors and/or scales for all staining pattern classes, which restricts the recognition rates of HEp-2 cell classification. Recently, using an ensemble of classifiers, a classification approach has been introduced in [36], based on one-against-one scheme, to classify HEp-2 cell images, while both feature extraction and classification approaches are differentiated in [37], based on one-against-all scheme. However, both of these approaches suffer from training a large number of classifiers.

In this background, the paper introduces a novel method to select a set of relevant descriptors under appropriate scales for the recognition of staining patterns present in HEp-2 cell images. The proposed method assumes that a uniform set of descriptors and/or scale is not effective for all HEp-2 cell classes. Therefore, a set of relevant local texture descriptors is first identified under appropriate scales for a pair of classes, and then the final feature set for multiple classes is formed from the relevant descriptors of all possible pairs of classes selected under appropriate scales. Judiciously integrating the merits of rough sets and Bayes decision theory, a new framework, termed as Rough-Bayesian model, is introduced to evaluate the relevance of a local texture descriptor and/or a scale for a pair of classes. While rough set theory deals with the uncertainty due to inexactness, vagueness, or incompleteness in HEp-2 class definition, the probabilistic model addresses the uncertainty due to overlapping classes by measuring the belongingness of an HEp-2 cell to a specific class, based on some presumed intensity distribution of the class. During the computation of the relevance of each descriptor and/or scale, the proposed method takes care of the presence of both noisy pixels in an HEp-2 cell image and noisy HEp-2 cell images in a staining pattern class. Finally, support vector machine (SVM) is used to predict the staining patterns present in HEp-2 cell images. The performance of the proposed method, along with a comparison with other related methods, is demonstrated on a set of real-life HEp-2 cell image databases.

2. Selection of relevant descriptors and scales

This section presents a new method to find a set of effective descriptors, along with the corresponding scales, for classification of staining patterns in HEp-2 cell images.

Table 1
Confusion Matrix for ICPR Test Data Set at Scale 4.

Actual/Predict	Centromere	Homogeneous	Nucleolar	Speckled	Nuclear Mem.	Golgi
			Predicted Class Using **LBP**			
Centromere	**79.94%(1096)**	2.26%(31)	3.72%(51)	13.42%(184)	0.29%(4)	0.36%(5)
Homogeneous	1.44%(18)	**69.77%(870)**	3.69%(46)	18.93%(236)	5.37%(67)	0.80%(10)
Nucleolar	4.31%(56)	6.39%(83)	**67.82%(881)**	10.08%(131)	4.08%(53)	7.31%(95)
Speckled	9.04%(128)	20.76%(294)	6.29%(89)	**60.38%(855)**	1.62%(23)	1.91%(27)
Nuclear Mem.	0.18%(2)	14.58%(161)	4.26%(47)	4.26%(47)	**71.74%(792)**	4.98%(55)
Golgi	0.83%(3)	9.12%(33)	**38.40%(139)**	9.67%(35)	11.60%(42)	**30.39%(110)**
			Predicted Class Using **LBP**^{ri}			
Centromere	**83.08%(1139)**	1.97%(27)	4.74%(65)	9.99%(137)	0.07%(1)	0.15%(2)
Homogeneous	1.68%(21)	**81.96%(1022)**	4.17%(52)	7.22%(90)	3.93%(49)	1.04%(13)
Nucleolar	3.00%(39)	6.16%(80)	**81.45%(1058)**	4.46%(58)	3.00%(39)	1.92%(25)
Speckled	15.89%(225)	**38.56%(546)**	12.43%(176)	**30.23%(428)**	1.84%(26)	1.06%(15)
Nuclear Mem.	0.18%(2)	11.14%(123)	3.35%(37)	1.36%(15)	**80.98%(894)**	2.99%(33)
Golgi	1.66%(6)	10.50%(38)	**46.41%(168)**	4.70%(17)	8.84%(32)	**27.90%(101)**
			Predicted Class Using **LBP**^{riu2}			
Centromere	**85.34%(1170)**	1.53%(21)	3.06%(42)	9.85%(135)	0.22%(3)	0.00%(0)
Homogeneous	1.68%(21)	**69.53%(867)**	3.69%(46)	19.97%(249)	5.13%(64)	0.00%(0)
Nucleolar	1.77%(23)	3.93%(51)	**73.29%(952)**	15.78%(205)	5.23%(68)	0.00%(0)
Speckled	9.89%(140)	20.06%(284)	7.84%(111)	**59.25%(839)**	2.90%(41)	0.07%(1)
Nuclear Mem.	0.82%(9)	12.23%(135)	6.43%(71)	1.81%(20)	**78.71%(869)**	0.00%(0)
Golgi	0.55%(2)	15.47%(56)	**59.94%(217)**	8.84%(32)	14.92%(54)	**0.28%(1)**
			Predicted Class Using **CoALBP**			
Centromere	**79.36%(1088)**	4.38%(60)	6.71%(92)	8.90%(122)	0.29%(4)	0.36%(5)
Homogeneous	0.88%(11)	**68.64%(856)**	6.26%(78)	19.09%(238)	4.01%(50)	1.12%(14)
Nucleolar	7.85%(102)	7.16%(93)	**73.98%(961)**	7.62%(99)	1.46%(19)	1.92%(25)
Speckled	7.06%(100)	21.75%(308)	4.66%(66)	**64.83%(918)**	0.99%(14)	0.71%(10)
Nuclear Mem.	0.91%(10)	5.71%(63)	1.99%(22)	3.89%(43)	**77.63%(857)**	9.87%(109)
Golgi	6.35%(23)	5.25%(19)	15.19%(55)	6.35%(23)	30.94%(112)	**35.91%(130)**
			Predicted Class Using **Class-Pair Specific Descriptor**			
Centromere	**83.81%(1149)**	2.12%(29)	5.18%(71)	8.53%(117)	0.07%(1)	0.29%(4)
Homogeneous	0.80%(10)	**75.70%(944)**	3.29%(41)	16.36%(204)	3.37%(42)	0.48%(6)
Nucleolar	5.00%(65)	5.16%(67)	**79.06%(1027)**	7.85%(102)	1.00%(13)	1.92%(25)
Speckled	6.50%(92)	17.80%(252)	4.59%(65)	**69.70%(987)**	0.99%(14)	0.42%(6)
Nuclear Mem.	0.27%(3)	6.88%(76)	1.45%(16)	2.54%(28)	**84.69%(935)**	4.17%(46)
Golgi	3.04%(11)	2.49%(9)	18.51%(67)	5.25%(19)	14.64%(53)	**56.08%(203)**

2.1. Motivation

In general, a single scale of same local textural features is used for the recognition of HEp-2 patterns. Due to the drastic variation of different scales and noisy nature of the input HEp-2 cell images, the single scale and/or descriptor usually gives poor performance, which is reflected in the insufficient and inaccurate staining pattern representation of these images.

Let us assume that $\mathcal{M} = \{\mathcal{M}_1, \cdots, \mathcal{M}_p, \cdots, \mathcal{M}_t\}$ represents a set of t modalities, where each modality $\mathcal{M}_p$ corresponds to a local descriptor considered under a particular scale. In the current example, four types of local descriptor, namely, LBP [16], LBP^{ri} [17], LBP^{riu2} [17], and CoALBP [18], are used, while four scales of LBP neighborhood such as 1 (S_1), 2 (S_2), 3 (S_3), and 4 (S_4) are considered. The 4-neighborhood is used for CoALBP, while for others 8-neighborhood is considered. Tables 1 and 2 present several confusion matrices obtained for ICPR image database [38], which was used in HEp-2 cell classification contest organized by ICPR 2014. In particular, 68,429 cell images constitute the database, out of which 20% of data set, that is, 13,596 cell images are publicly available. For experimental purposes, the set of 13,596 images is partitioned into training and test sets, consisting of 6,797 and 6,799 cell images, respectively. In both training and test sets, the images are almost equally distributed with respect to six different pattern classes, namely, Centromere, Homogeneous, Nucleolar, Nuclear Membrane, Speckled, and Golgi.

2.1.1. Performance at fixed scale

Each confusion matrix of Table 1 corresponds to each of the four local descriptors at scale S_4. The SVM [39], with linear kernels, is used to evaluate the performance of different feature descriptors in the present work. While the first matrix corresponds to LBP, next three matrices, namely, second, third, and fourth, are obtained for LBP^{ri}, LBP^{riu2}, and CoALBP, respectively. The LBP, LBP^{ri}, LBP^{riu2}, and CoALBP, respectively, achieve 85.83%, 70.21%, 69.56%, and 100% classification accuracy on training cell images of ICPR database, while 67.72%, 68.27%, 69.10%, and 70.75% accuracy on test images. However, from the confusion matrices, it can be seen that the class labels of 38.40%, 46.41%, and 59.94% of the cell images of Golgi class are predicted as Nucleolar class for LBP, LBP^{ri}, and LBP^{riu2}, respectively, while the predicted class of 30.94% of the cell images of Golgi class is Nuclear Membrane for CoALBP. Similarly, the class label of 38.56% cell images of Speckled class is predicted as Homogeneous class for LBP^{ri}. Moreover, for LBP^{riu2}, only 1 image, out of 362 cell images, of Golgi class is correctly identified, while the class label is correctly predicted for 85.34% of the cell images of Centromere class. For LBP, LBP^{riu2}, and CoALBP, nearly 20% of Homogeneous cells is predicted as Speckled cells, while almost same percentage of Speckled cells is classified as Homogeneous cells. So, a particular descriptor may be important for classifying a class-pair, but may not be effective for another pair of classes.

2.1.2. Performance at fixed descriptor

Each confusion matrix of Table 2 corresponds to each scale of LBP^{ri}. While the first two matrices correspond to scales S_1 and S_2, next two matrices are obtained for scales S_3 and S_4. The LBP^{ri} attains 60.67%, 68.56%, 73.30%, and 70.21% classification accuracy on training cell images of ICPR database for scales S_1, S_2, S_3, and S_4, respectively, while 59.35%, 66.35%, 70.88%, and 68.28% of accuracy on test images. From the first four confusion matrices, it is seen that the class labels of 45.30%, 55.25%, 48.07%, and 46.41% of the cell images of Golgi class are predicted as Nucleolar class for scales S_1, S_2, S_3, and S_4, respectively, while the class of 34.04% and

Table 2
Confusion Matrix for ICPR Test Data Set for LBP^{ri}.

Actual/Predict	Centromere	Homogeneous	Nucleolar	Speckled	Nuclear Mem.	Golgi
			Predicted Class Using **Scale 1**			
Centromere	**75.27%(1032)**	0.22%(3)	7.59%(104)	16.92%(232)	0.00%(0)	0.00%(0)
Homogeneous	0.08%(1)	**50.12%(625)**	5.21%(65)	30.07%(375)	14.35%(179)	0.16%(2)
Nucleolar	4.85%(63)	5.23%(68)	**60.20%(782)**	18.78%(244)	10.39%(135)	0.54%(7)
Speckled	5.44%(77)	15.89%(225)	12.50%(177)	**63.28%(896)**	2.12%(30)	0.78%(11)
Nuclear Mem.	0.00%(0)	13.86%(153)	14.58%(161)	8.24%(91)	**63.04%(696)**	0.27%(3)
Golgi	1.38%(5)	18.51%(67)	**45.30%(164)**	20.17%(73)	13.54%(49)	**1.10%(4)**
			Predicted Class Using **Scale 2**			
Centromere	**84.32%(1156)**	1.68%(23)	5.91%(81)	7.80%(107)	0.29%(4)	0.00%(0)
Homogeneous	0.88%(11)	**72.49%(904)**	2.73%(34)	14.51%(181)	9.22%(115)	0.16%(2)
Nucleolar	2.54%(33)	4.39%(57)	**79.14%(1028)**	7.85%(102)	5.70%(74)	0.38%(5)
Speckled	15.61%(221)	34.04%(482)	8.90%(126)	**38.91%(551)**	2.26%(32)	0.28%(4)
Nuclear Mem.	0.36%(4)	11.87%(131)	7.88%(87)	2.26%(25)	**77.08%(851)**	0.54%(6)
Golgi	2.21%(8)	13.81%(50)	**55.25%(200)**	11.33%(41)	11.60%(42)	**5.80%(21)**
			Predicted Class Using **Scale 3**			
Centromere	**86.14%(1181)**	1.39%(19)	3.36%(46)	8.53%(117)	0.22%(3)	0.36%(5)
Homogeneous	1.28%(16)	**65.12%(812)**	6.50%(81)	18.36%(229)	7.94%(99)	0.80%(10)
Nucleolar	2.62%(34)	3.23%(42)	**77.14%(1002)**	11.93%(155)	3.39%(44)	1.69%(22)
Speckled	10.17%(144)	18.29%(259)	5.72%(81)	**62.71%(888)**	2.19%(31)	0.92%(13)
Nuclear Mem.	0.09%(1)	13.77%(152)	3.80%(42)	2.72%(30)	**76.99%(850)**	2.63%(29)
Golgi	0.00%(0)	8.29%(30)	**48.07%(174)**	9.67%(35)	10.22%(37)	**23.76%(86)**
			Predicted Class Using **Scale 4**			
Centromere	**83.08%(1139)**	1.97%(27)	4.74%(65)	9.99%(137)	0.07%(1)	0.15%(2)
Homogeneous	1.68%(21)	**81.96%(1022)**	4.17%(52)	7.22%(90)	3.93%(49)	1.04%(13)
Nucleolar	3.00%(39)	6.16%(80)	**81.45%(1058)**	4.46%(58)	3.00%(39)	1.92%(25)
Speckled	15.89%(225)	**38.56%(546)**	12.43%(176)	**30.23%(428)**	1.84%(26)	1.06%(15)
Nuclear Mem.	0.18%(2)	11.14%(123)	3.35%(37)	1.36%(15)	**80.98%(894)**	2.99%(33)
Golgi	1.66%(6)	10.50%(38)	**46.41%(168)**	4.70%(17)	8.84%(32)	**27.90%(101)**
			Predicted Class Using **Class-Pair Specific Scale**			
Centromere	**91.76%(1258)**	1.17%(16)	2.77%(38)	4.16%(57)	0.07%(1)	0.07%(1)
Homogeneous	0.16%(2)	**81.40%(1015)**	2.65%(33)	12.51%(156)	2.89%(36)	0.40%(5)
Nucleolar	1.92%(25)	4.39%(57)	**83.99%(1091)**	6.62%(86)	1.39%(18)	1.69%(22)
Speckled	4.10%(58)	15.11%(214)	3.67%(52)	**75.56%(1070)**	1.48%(21)	0.07%(1)
Nuclear Mem.	0.09%(1)	7.52%(83)	1.36%(15)	1.72%(19)	**85.24%(941)**	4.08%(45)
Golgi	1.66%(6)	3.59%(13)	21.82%(79)	5.52%(20)	6.91%(25)	**60.50%(219)**

38.56% of the cell images of Speckled class is predicted as Homogeneous class for scales S_2 and S_4, respectively. On the other hand, the class labels of 30.07% of cell images of Homogeneous class are predicted as Speckled class for scale S_1. So, a specific scale may be significant in discriminating a class-pair, while it may not be important for another pair of classes.

From all the confusion matrices reported in Tables 1 and 2, it is seen that a unique scale and/or a unique local descriptor may not be effective for all staining pattern classes. In other words, a scale or a descriptor may be significant for classifying a pair of classes, while may not be effective for another pair of classes. For a pair of classes, if the appropriate descriptors and corresponding scales can be identified properly, better classification accuracy can be achieved for classifying HEp-2 cells into multiple classes, as evident from last confusion matrices of Tables 1 and 2. These two matrices correspond to the cases when descriptors and scales, respectively, are class-pair specific given fixed scale (S_4) and fixed descriptor (LBP^{ri}). In case of class-pair specific descriptor, as reported in Table 1, the classification accuracy on test images is 77.14%, while it is 82.28% if class-pair specific scale (Table 2) is considered. When both descriptor and scale are considered to be class-pair specific, the classification accuracy on test images is increased to 85.54% and corresponding confusion matrix is presented in Table 3. More than 50% cells of Golgi class are correctly identified in these three cases, which is not possible if a single descriptor or a single scale is considered for the extraction of features of all classes.

2.2. Proposed method

The proposed method selects important features from the relevant modalities for each pair of classes, and forms the resultant feature set for multiple staining pattern classes. The block diagram of the proposed method for HEp-2 cell staining pattern classification is depicted in Fig. 1. At first, the class specific feature set is formed for each of the HEp-2 cell staining pattern classes, considering the histograms of feature values for each of the input HEp-2 cell images, under a specific modality. Then, the class-pair specific feature sets are obtained for all possible class-pairs, from each of the class specific feature sets. The relevance of each class-pair specific feature set under a particular modality is evaluated based on rough sets and Bayes decision theory. After selecting a set of most relevant feature sets for each class-pair, the final feature set for multiple staining pattern classes is formed, which is used to train support vector machine. Each of these steps is elaborated next one by one.

Let us assume that $X = \{x_1, \cdots, x_j, \cdots, x_n\}$ be a set of n number of HEp-2 cell images, where each image $x_j \in \Re^m$ is represented by a set $h_j = \{h_{j1}, \cdots, h_{jk}, \cdots, h_{jm}\}$, consisting of feature values of m features $\{\mathcal{A}_1, \cdots, \mathcal{A}_k, \cdots, \mathcal{A}_m\}$. In the proposed work, four local descriptors, namely, LBP, LBP^{ri}, LBP^{riu2}, and CoALBP, are considered. So, h_j is the normalized histogram of x_j corresponding to either LBP, LBP^{ri}, LBP^{riu2} or CoALBP. Let $\tilde{h}_j = \{\tilde{h}_{j1}, \cdots, \tilde{h}_{jk}, \cdots, \tilde{h}_{jm}\}$ be the sorted normalized histogram of x_j, in descending order, and $I_j = \{I_{j1}, \cdots, I_{jk}, \cdots, I_{jm}\}$ represents the feature order of $\tilde{h}_j$. Fig. 2(a) presents the normalized LBP histogram at S_2 of an HEp-2 cell image belonging to Golgi class of ICPR data, while Fig. 2(b) depicts the corresponding sorted histogram.

2.2.1. Class dependent features

In order to find out the set of dominant features present in $\tilde{h}_j$, corresponding to each HEp-2 cell image x_j, the following function

Table 3
Confusion Matrix for ICPR Test Data Set Using Class-Pair Specific Descriptors and Scales.

Actual/Predict	Centromere	Homogeneous	Nucleolar	Speckled	Nuclear Membrane	Golgi
Centromere	**92.27%(1265/1371)**	1.02%(14/1371)	2.26%(31/1371)	4.30%(59/1371)	0.07%(1/1371)	0.07%(1/1371)
Homogeneous	0.08%(1/1247)	**83.88%(1046/1247)**	1.04%(13/1247)	13.79%(172/1247)	0.96%(12/1247)	0.24%(3/1247)
Nucleolar	2.00%(26/1299)	2.77%(36/1299)	**88.22%(1146/1299)**	4.31%(56/1299)	0.92%(12/1299)	1.77%(23/1299)
Speckled	3.04%(43/1416)	13.28%(188/1416)	3.11%(44/1416)	**79.31%(1123/1416)**	1.06%(15/1416)	0.21%(3/1416)
Nuclear Mem.	0.00%(0/1104)	4.89%(54/1104)	1.45%(16/1104)	1.36%(15/1104)	**89.04%(983/1104)**	3.26%(36/1104)
Golgi	1.66%(6/362)	2.21%(8/362)	13.81%(50/362)	2.76%(10/362)	9.67%(35/362)	**69.89%(253/362)**

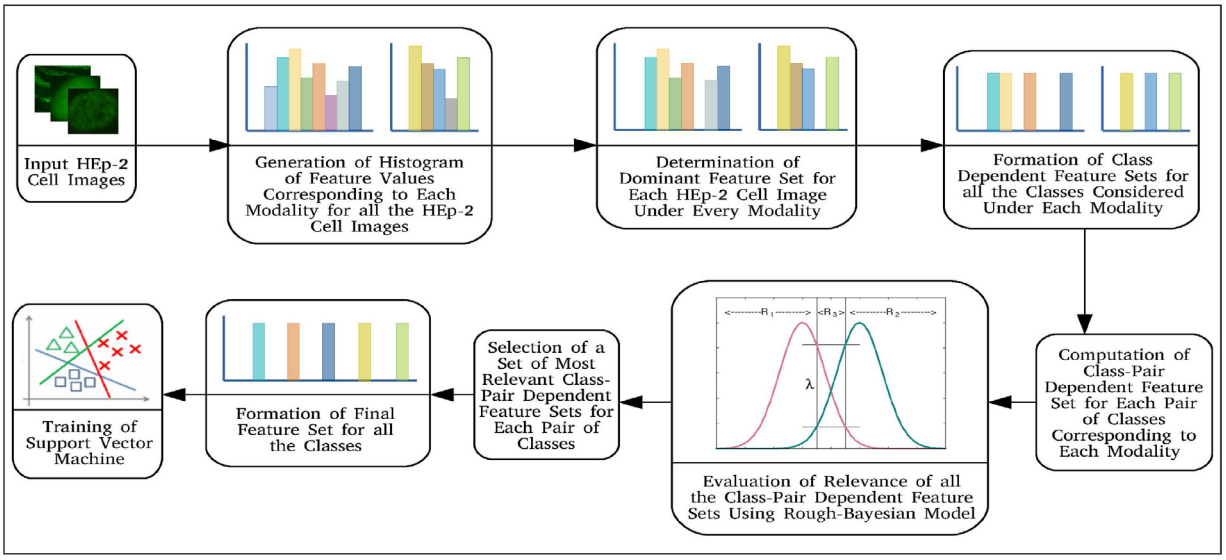
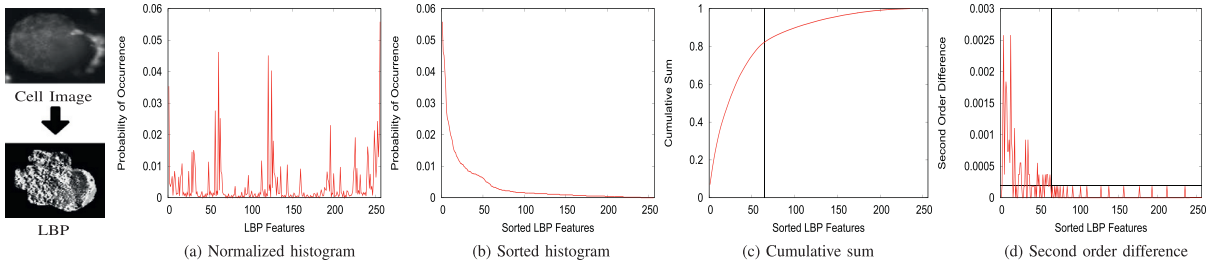


Fig. 1. Block diagram of the proposed HEp-2 cell staining pattern recognition method.



(a) Normalized histogram (b) Sorted histogram (c) Cumulative sum (d) Second order difference

Fig. 2. Illustrative example to find out the dominant features of an HEp-2 cell image.

is defined:

$$\mathcal{S}(r) = \sum_{l=1}^{r} \tilde{h}_{jl}. \tag{1}$$

The function $\mathcal{S}(r)$ represents the cumulative sum of first r normalized feature values of $\tilde{h}_j$. As each normalized feature value $\tilde{h}_{jk}$ is positive, the function $\mathcal{S}(r)$ is single-valued and monotonically increasing in the interval $1 \leq r \leq m$. Moreover, $0 \leq \mathcal{S}(r) \leq 1$ for $1 \leq r \leq m$. The second-order derivative of this function is the difference

$$\frac{\partial^2 \mathcal{S}(r)}{\partial r^2} = \mathcal{S}(r+1) + \mathcal{S}(r-1) - 2\mathcal{S}(r). \tag{2}$$

Fig. 2 (c) presents the variation of cumulative sum $\mathcal{S}(r)$ of first r normalized feature values of $\tilde{h}_j$, corresponding to Fig. 2(b). In this figure, the number of dominant features of x_j is denoted by the point in X-axis, where the slope of the graph changes abruptly. This is due to the fact that further addition of features into the dominant feature set does not offer any significant increase in the cumulative sum of the normalized histogram of the object x_j. In order to identify this point, the difference operator is applied on $\mathcal{S}(r)$. Now, from the definition, it is known that the first-order difference yields a non-zero response along a ramp. With the change of the slope of the ramp, only the magnitude of the response of first-order difference operator changes. Hence, it is not possible to identify the point of interest with first-order difference response. On the other hand, in case of second-order difference, a non-zero response is only obtained at the beginning and end of the ramp, that is, where the slope changes, and a zero response is obtained along the ramp. Since the graph of $\mathcal{S}(r)$ can be considered as piece-wise linear after the slope changes, the second-order difference produces almost zero response in this region. Fig. 2(d) depicts the second-order difference of $\mathcal{S}(r)$, corresponding to Fig. 2(c). Thus, the number of dominant features for a particular object x_j is defined as the feature index beyond which second-order difference yields near to zero response, indicating no further significant change in slope of the graph. As the second-order difference produces double sided response, modulus of the response is considered for computational advantage. The number of dominant features d_j, present in $\tilde{h}_j$, is thus defined as $d_j = r$, if $\left|\frac{\partial^2 \mathcal{S}(l)}{\partial l^2}\right| < \varepsilon_1$,

$\forall l \in \{r+1, r+2, \cdots, m\}$. So, the average number of dominant features for the set X is given by

$$\overline{d} = \frac{1}{n}\sum_{j=1}^{n} d_j. \tag{3}$$

The set of dominant features, denoted by V_j, corresponding to sample x_j, is defined as

$$V_j = \{\mathcal{A}_k \mid I_{jq} = k \quad \text{and} \quad q \leq \overline{d}\}. \tag{4}$$

The features, which are dominant in sample x_j, belong to the set V_j, while rest of the features those are considered to be insignificant do not belong to V_j. Algorithm 1 presents the basic steps to determine the set of dominant features for each training cell image $x_j \in X$.

Algorithm 1 Determination of Set of Dominant Features for Each Cell Image $x_j \in X$.

Input: Set of training cell images X.
Output: Set of dominant features for each cell image $x_j \in X$.

1: **for** each $x_j \in X$ **do**
   (i) Compute the normalized histogram h_j under a particular modality.
   (ii) Sort the normalized histogram and compute cumulative sum of the sorted histogram.
   (iii) Determine the second-order difference of the cumulative sum of sorted histogram.
   (iv) Find out the number of dominant features d_j of the sample x_j, upto which the second-order difference exhibits a non-zero response.
2: Compute the average number of dominant features $\overline{d}$ over the entire set X.
3: **for** each $x_j \in X$ **do**
   • Select the first $\overline{d}$ features from the sorted histogram $\tilde{h}_j$ as the set of dominant features for the sample x_j.

Let us assume that each sample $x_j \in X$ belongs to one of the c classes of the set $\{\omega_1, \cdots, \omega_i, \cdots, \omega_c\}$. It is likely that the samples of the same class will have similar dominant feature sets, while samples from different classes will have different sets of dominant features. Presence of a noisy sample in the set X may introduce certain features in the dominant set which definitely do not bear any significant characteristics of the class, say ω_i, to which the noisy sample belongs. Rather, presence of those features in the dominant set can cause erroneous decision regarding classification of the samples of class ω_i. Hence, it is necessary to identify the features from the dominant set of a sample which correctly represent the characteristics of the class to which it belongs.

In order to address the aforementioned problem, the probability of occurrence of a feature $\mathcal{A}_k$ in the dominant sets of the samples of a particular class ω_i is defined as follows:

$$p(\mathcal{A}_k|\omega_i) = \frac{1}{|\omega_i|}\sum_{x_j \in \omega_i} v_{jk}; \tag{5}$$

$$\text{where} \quad v_{jk} = \begin{cases} 1, & \text{if } \mathcal{A}_k \in V_j; \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Here, $|\omega_i|$ represents the number of samples belonging to class ω_i. Fig. 3(a) depicts the probability of occurrence of LBPs in Golgi class. The features, which have occurred in the dominant sets of the samples belonging to class ω_i inadvertently, will have low probability of occurrence. Hence, the features having $p(\mathcal{A}_k|\omega_i)$ values below a certain threshold $T(\omega_i)$ can be considered to be insignificant for the class ω_i and can be discarded without loss of much information. Moreover, if the samples of a class ω_i show a large degree of variations among them, the sets of dominant features of those samples will also vary to a great extent. So, there will be a large number of features $\mathcal{A}_k$s having low $p(\mathcal{A}_k|\omega_i)$ values. Here, $T(\omega_i)$ is expected to have a low value. On the other hand, if the samples of a class, say ω_r, exhibit similar properties, then most of the features $\mathcal{A}_k$s will have high $p(\mathcal{A}_k|\omega_r)$ values. In this case, $T(\omega_r)$ should have a high value. Hence, the values of probability of occurrence of features in the dominant sets of a class are not enough to determine the threshold of that class. The frequency of probability of occurrence values also needs to be considered. Hence, the histogram $H(\omega_i)$ of the probability of occurrence values is calculated as follows:

$$H_l(\omega_i) = H_l(\omega_i) + 1 \quad \text{if} \quad p(\mathcal{A}_k|\omega_i) \in \left(\frac{l-1}{L}, \frac{l}{L}\right], \quad \forall k; \tag{7}$$

where $1 \leq \sum_{l=1}^{L} H_l(\omega_i) \leq m$, L is the number of bins in histogram and $l \in \{1, 2, \cdots, L\}$ denotes an index of the bins. The maximum value of l, for which

$$\sum_{q=1}^{l} H_q(\omega_i) < \varepsilon_2 \sum_{q=1}^{L} H_q(\omega_i), \tag{8}$$

determines the threshold for the class ω_i as follows:

$$T(\omega_i) = \frac{l}{L}. \tag{9}$$

The parameter ε_2, which is considered to be 0.1 in the current study, depicts the amount of acceptable loss. Fig. 3(b) represents the histogram corresponding to probability of occurrence values of Fig. 3(a). The threshold value for Golgi class of ICPR data, obtained using (9), is 0.2.

Based on the threshold $T(\omega_i)$, the class dependent feature set $\mathcal{N}(\omega_i)$ for the class ω_i is defined as

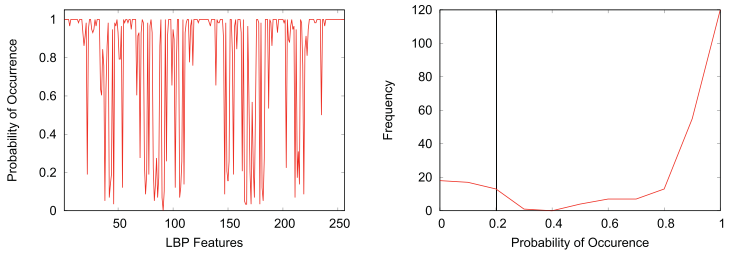
$$\mathcal{N}(\omega_i) = \{\mathcal{A}_k \mid p(\mathcal{A}_k|\omega_i) \geq T(\omega_i)\}. \tag{10}$$

So, the k-th feature $\mathcal{A}_k$ belongs to the feature set $\mathcal{N}(\omega_i)$ only if $\mathcal{A}_k$ is dominant for most of the samples belonging to class ω_i as well as represents the important characteristics of the class ω_i. The main steps to generate class dependent feature set $\mathcal{N}(\omega_i)$, corresponding to the class ω_i, are outlined in Algorithm 2.

Algorithm 2 Determination of Class Dependent Feature Set.

Input: Set of dominant features of each cell image $x_j \in X$.
Output: Class specific feature set $\mathcal{N}(\omega_i)$; $i = 1, \cdots, c$.

1: **for** each class ω_i **do**
   (i) Calculate the probability of occurrence of features present in the dominant sets of samples belonging to the class ω_i.
   (ii) Compute histogram of the probability of occurrence values in the class ω_i.
   (iii) Determine class specific threshold $T(\omega_i)$ from the histogram of the class ω_i.
   (iv) Form a feature set $\mathcal{N}(\omega_i)$, corresponding to the class ω_i, with the features having probability of occurrence values greater than $T(\omega_i)$.
2: Stop.



(a) Golgi (b) $T(\text{Golgi}) = 0.2$

Fig. 3. Illustration to find out dominant features of an HEp-2 cell class.

2.2.2. Pairwise class dependent features

Let $\mathcal{N}(\omega_i)$ and $\mathcal{N}(\omega_r)$ be two feature sets corresponding to two classes ω_i and ω_r, respectively. These two sets represent the unique characteristics of two classes. The resultant feature set for a pair of classes ω_i and ω_r, representing the characteristics of both the classes, can be defined as

$$\mathcal{N}(\{\omega_i, \omega_r\}) = \mathcal{N}(\omega_i) \cap \mathcal{N}(\omega_r). \tag{11}$$

Hence, the feature set $\mathcal{N}(\{\omega_i, \omega_r\})$ contains those features which are dominant or significant with respect to both the classes ω_i and ω_r.

Given a set of modalities $\mathcal{M} = \{\mathcal{M}_1, \cdots, \mathcal{M}_p, \cdots, \mathcal{M}_t\}$, the proposed method evaluates the relevance of each feature set $\mathcal{N}_p(\{\omega_i, \omega_r\})$, corresponding to the p-th modality $\mathcal{M}_p$, for a pair of classes $\{\omega_i, \omega_r\}$. Let $\Gamma_p(\{\omega_i, \omega_r\})$ be the relevance of the feature set $\mathcal{N}_p(\{\omega_i, \omega_r\})$. The relevance $\Gamma \in [0, 1]$ of each feature set is computed based on Bayes decision theory and the concept of rough sets, introduced in Section 3. If the value of $\Gamma_p(\{\omega_i, \omega_r\})$ is 1, then all the samples belonging to two classes ω_i and ω_r can be classified correctly using the set of features selected under the p-th modality $\mathcal{M}_p$. If the value of $\Gamma_p(\{\omega_i, \omega_r\}) = 0$, then no sample of either classes can be correctly classified with the selected set of features. On the other hand, if $\Gamma_p(\{\omega_i, \omega_r\}) \in (0, 1)$, then some of the samples of two classes can be accurately classified, while others are misclassified. So, the value of $\Gamma_p(\{\omega_i, \omega_r\})$ refers to the relevance of the p-th modality $\mathcal{M}_p$ for differentiating a pair of classes $\{\omega_i, \omega_r\}$. After selecting a set $\tilde{\mathcal{N}}_{ir} = \{\mathcal{N}_p(\{\omega_i, \omega_r\})\}$ of $\tilde{t}$ most relevant feature sets for a pair of classes $\{\omega_i, \omega_r\}$ from t modalities, the final feature set $\mathcal{D}$ for all possible pairs of classes is obtained as follows:

$$\mathcal{D} = \bigcup \tilde{\mathcal{N}}_{ir}. \tag{12}$$

The basic steps to select a set of best $\tilde{t}$ feature sets, for a pair of classes $\{\omega_i, \omega_r\}$, are outlined in Algorithm 3; while Algorithm 4 presents the main steps of proposed method for HEp-2 cell image classification.

Algorithm 3 Selection of Relevant Modality $\mathcal{M}_p \in \mathcal{M}$ for a Pair of Classes $\{\omega_i, \omega_r\}$.

Input: Class specific feature sets $\mathcal{N}_p(\omega_i)$ and $\mathcal{N}_p(\omega_r)$ for each modality $\mathcal{M}_p \in \mathcal{M}$.
Output: A set of class-pair specific feature sets $\tilde{\mathcal{N}}_{ir} = \{\mathcal{N}_p(\{\omega_i, \omega_r\})\}$, along with corresponding modalities $\{\mathcal{M}_p\}$.

1: **for** each modality $\mathcal{M}_p \in \mathcal{M}$ **do**

(i) Compute class-pair specific feature set $\mathcal{N}_p(\{\omega_i, \omega_r\})$ using two class specific feature sets $\mathcal{N}_p(\omega_i)$ and $\mathcal{N}_p(\omega_r)$ of the classes ω_i and ω_r, respectively.
(ii) Compute the relevance $\Gamma_p(\{\omega_i, \omega_r\})$ of the class-pair specific feature set $\mathcal{N}_p(\{\omega_i, \omega_r\})$ using **Algorithm 5**.

2: Select the set $\tilde{\mathcal{N}}_{ir} = \{\mathcal{N}_p(\{\omega_i, \omega_r\})\}$ of best $\tilde{t}$ feature sets, based on the relevance values.

Algorithm 4 Proposed Method for HEp-2 Cell Classification.

Input: Set of training cell images X.
Output: Final feature set $\mathcal{D}$.

1: **for** each modality $\mathcal{M}_p \in \mathcal{M}$ **do**

(i) Obtain the set of dominant features for each of the training cell images $x_j \in X$ by **Algorithm 1**.
(ii) Determine class specific feature set $\mathcal{N}_p(\omega_i)$, where $i = 1, \cdots, c$, using **Algorithm 2**.

2: **for** each class-pair $\{\omega_i, \omega_r\}$ **do**

- Select a set of class-pair specific feature sets $\tilde{\mathcal{N}}_{ir} = \{\mathcal{N}_p(\{\omega_i, \omega_r\})\}$, using **Algorithm 3**.

3: Form the final feature set $\mathcal{D}$, from all the selected class-pair specific feature sets $\tilde{\mathcal{N}}_{ir}$'s, corresponding to all pairs of classes.

Algorithm 5 Computation of Relevance of Feature Set.

Input: A set of objects $\mathbb{U} = \{x_1, \cdots, x_j, \cdots, x_n\}$, a set of condition attributes $\mathbb{C} = \{\mathcal{A}_1, \cdots, \mathcal{A}_k, \cdots, \mathcal{A}_m\}$, decision attribute set $\mathbb{D}$, and admissible classification error α.
Output: $\text{Rel}_{\mathbb{C}}(\mathbb{D})$, relevance of condition attribute set $\mathbb{C}$ with respect to decision attribute set $\mathbb{D}$.

1: **for** each condition attribute or feature $\mathcal{A}_k \in \mathbb{C}$ **do**

- Form equivalence classes or information granules.
- Construct α-lower approximation for each class $\omega_i \in \mathbb{D}$ using (31).
- Construct the matrix $\mathbb{L}(\mathcal{A}_k)$ using (35).

2: Construct the matrix $\tilde{\mathbb{L}}(\mathbb{C})$, corresponding to the set $\mathbb{C}$, from m number of $\mathbb{L}(\mathcal{A}_k)$ matrices using (37).
3: Calculate the relevance $\text{Rel}_{\mathbb{C}}(\mathbb{D})$ of $\mathbb{C}$ with respect to $\mathbb{D}$ from $\tilde{\mathbb{L}}(\mathbb{C})$ using (36).

3. Rough-Bayesian approach for computation of relevance

The problem of generation of final feature set using (12) for c number of HEp-2 cell staining pattern classes boils down to the evaluation of each feature set, corresponding to a pair of classes considered under a particular modality. This section introduces a novel framework, termed as Rough-Bayesian model, to compute the relevance of a set of features, integrating judiciously the merits of rough sets and Bayes decision theory.

Let $\mathbb{U} = \{x_1, \cdots, x_j, \cdots, x_n\}$ be the set of n objects, $\mathbb{C} = \{\mathcal{A}_1, \cdots, \mathcal{A}_k, \cdots, \mathcal{A}_m\}$ be the set of m condition attributes or features, and $\mathbb{D}$ is the decision attribute set in $\mathbb{U}$. Let $\mathbb{U}/\mathbb{D} = \{\omega_1, \cdots, \omega_i, \cdots, \omega_c\}$ denote c equivalence classes or information

granules of $\mathbb{U}$ generated by the equivalence relation induced from the decision attribute set $\mathbb{D}$. The proposed method assumes that each class ω_i can be approximated by a pair of sets, namely, lower approximation $\underline{A}(\omega_i)$ and boundary region $B(\omega_i)$, based on the theory of rough sets. The upper approximation $\overline{A}(\omega_i) = [\underline{A}(\omega_i) \cup B(\omega_i)]$ [35]. If an object belongs to the lower approximation $\underline{A}(\omega_i)$ of a class ω_i, it definitely belongs to that class. On the other hand, if an object belongs to the boundary region $B(\omega_i)$, it possibly belongs to that class and potentially belongs to another class. Both lower and upper approximations of a class ω_i can be defined by the equivalence classes of $\mathbb{U}$ generated by the equivalence relation induced from each condition attribute $\mathcal{A}_k \in \mathbb{C}$. The equivalence classes or information granules, corresponding to the condition attribute $\mathcal{A}_k$ or the set $\mathbb{C}$, are constructed based on Bayes decision theory, considering two-class problem.

Let ω_1 and ω_2 be the two classes in which the objects belong. Let us assume that the a priori probabilities $P(\omega_1)$ and $P(\omega_2)$ of two classes are either known or they can be estimated from the available training objects. Further, the class-conditional probability density functions (pdf) $p(x|\omega_i), i = 1, 2$; representing the distribution of the feature vectors in each of the classes, are assumed to be known. The pdf is also referred to as likelihood function of ω_i with respect to x. Given the above inputs, the task is to compute the conditional probabilities $P(\omega_i|x), i = 1, 2$; each of them represents the probability that the object x belongs to the respective class ω_i. Integrating the theory of rough sets and Bayes decision theory, the classification rules for the proposed method are formulated as follows:

- **R1:** x is classified to ω_1 if $P(\omega_1|x) - P(\omega_2|x) > \lambda$;
- **R2:** x is classified to ω_2 if $P(\omega_1|x) - P(\omega_2|x) < -\lambda$;
- **R3:** x is unclassified if $|P(\omega_1|x) - P(\omega_2|x)| \leq \lambda$.

The parameter $\lambda(\geq 0)$ controls the region of uncertainty due to inexactness in class definitions of ω_1 and ω_2, and overlapping boundaries between ω_1 and ω_2. According to Bayes rule,

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}; \quad i = 1, 2. \tag{13}$$

$$\text{Also,} \quad P(\omega_1|x) + P(\omega_2|x) = 1. \tag{14}$$

Combining rule **R1**, (13), and (14), we get

$$x \text{ is classified to } \omega_1 \text{ if } \frac{P(\omega_1)p(x|\omega_1)}{P(\omega_2)p(x|\omega_2)} > \frac{1+\lambda}{1-\lambda}. \tag{15}$$

Let us assume that the likelihood functions $p(x|\omega_i)$ of ω_i with respect to x are multivariate normal distributions, that is,

$$p(x|\omega_i) = \frac{1}{\sqrt{(2\pi)^m|\Sigma_i|}} \exp\left\{-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right\};$$

where μ_i and Σ_i are the mean of class ω_i and corresponding $m \times m$ covariance matrix, while $|\Sigma_i|$ denotes the determinant of Σ_i. So, the classification rule of **R1** reduces to:

$$x \text{ is classified to } \omega_1 \text{ if } d_M^2(x, \mu_2, \Sigma_2) - d_M^2(x, \mu_1, \Sigma_1) - \ln\frac{|\Sigma_1|}{|\Sigma_2|} > \Delta_{12}; \tag{16}$$

$$\text{where} \quad \Delta_{12} = 2\left[\ln\left(\frac{1+\lambda}{1-\lambda}\right) - \ln\frac{P(\omega_1)}{P(\omega_2)}\right]; \tag{17}$$

$$\text{and} \quad d_M(x, \mu_i, \Sigma_i) = \sqrt{(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)} \tag{18}$$

is the Mahalanobis distance between x and μ_i.

Assuming the individual features, constituting the feature vector, are mutually uncorrelated, we get

$$\Sigma_i = \text{diag}(\sigma_{i1}^2, \cdots, \sigma_{ik}^2, \cdots, \sigma_{im}^2); \tag{19}$$

where σ_{ik}^2 is the variance of the k-th feature. So,

$$x \text{ is classified to } \omega_1 \text{ if } \sum_{k=1}^{m}\left[\frac{(x_k-\mu_{2k})^2}{\sigma_{2k}^2} - \frac{(x_k-\mu_{1k})^2}{\sigma_{1k}^2} - 2\ln\left(\frac{\sigma_{1k}}{\sigma_{2k}}\right)\right] > \Delta_{12}. \tag{20}$$

Similarly, from rules **R2** and **R3**, we get

$$x \text{ is classified to } \omega_2 \text{ if } \sum_{k=1}^{m}\left[\frac{(x_k-\mu_{2k})^2}{\sigma_{2k}^2} - \frac{(x_k-\mu_{1k})^2}{\sigma_{1k}^2} - 2\ln\left(\frac{\sigma_{1k}}{\sigma_{2k}}\right)\right] < -\Delta_{12}; \tag{21}$$

$$\text{and } x \text{ is unclassified if } \left|\sum_{k=1}^{m}\left[\frac{(x_k-\mu_{2k})^2}{\sigma_{2k}^2} - \frac{(x_k-\mu_{1k})^2}{\sigma_{1k}^2} - 2\ln\left(\frac{\sigma_{1k}}{\sigma_{2k}}\right)\right]\right| \leq \Delta_{12}. \tag{22}$$

Based on the concept of positive approximation accelerator, introduced in [40,41], the computing performance of a condition attribute set $\mathbb{C}_{k+1} = \{\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_{k+1}\}$ can be improved using the following recursive expression principle:

$$POS_{\mathbb{C}_{k+1}}^{\mathbb{U}_k}(\mathbb{D}) = POS_{\mathbb{C}_k}^{\mathbb{U}_k}(\mathbb{D}) \cup POS_{\mathcal{A}_{k+1}}^{\mathbb{U}_{k+1}}(\mathbb{D}); \tag{23}$$

$$\text{where} \quad \mathbb{U}_{k+1} = \mathbb{U}_k \setminus POS_{\mathbb{C}_k}^{\mathbb{U}_k}(\mathbb{D}); \quad \mathbb{U}_1 = \mathbb{U}; \quad \text{and}$$
$$POS_{\mathbb{C}}^{\mathbb{U}}(\mathbb{D}) = \bigcup_{\omega_i \in \mathbb{U}/\mathbb{D}} \underline{A}(\omega_i). \tag{24}$$

Here, $POS_{\mathbb{C}}^{\mathbb{U}}(\mathbb{D})$ is termed as the positive region, which contains all objects of $\mathbb{U}$ that can be classified to classes of $\mathbb{U}/\mathbb{D}$ using the knowledge in attribute set $\mathbb{C}$. In effect, the decision attribute set $\mathbb{D}$ can be positively approximated using granulation orders $\mathbb{C}_k$ and $\mathbb{C}_{k+1}$ on the gradually reduced universe, respectively. So, instead of considering all m features of $\mathbb{C}$ to generate information granules or equivalence classes, equivalence classes corresponding to each individual feature $\mathcal{A}_k \in \mathbb{C}$ can be generated separately, and then the decision attribute $\mathbb{D}$ can be positively approximated by the equivalence classes of individual features.

Let $\mathbb{U}/\mathcal{A}_k = \{\beta_1, \cdots, \beta_i, \cdots, \beta_{\tilde{c}}\}$ denote $\tilde{c}$ equivalence classes or information granules of $\mathbb{U}$ induced by the condition attribute $\mathcal{A}_k$. Based on Bayes decision theory, three information granules, namely, β_1, β_2 and β_3, are constructed for two classes corresponding to the condition attribute $\mathcal{A}_k$, as follows:

$$x_j \in \beta_1 \quad \text{if} \left[\frac{(x_{jk}-\mu_{2k})^2}{\sigma_{2k}^2} - \frac{(x_{jk}-\mu_{1k})^2}{\sigma_{1k}^2} - 2\ln\left(\frac{\sigma_{1k}}{\sigma_{2k}}\right)\right] > \tilde{\Delta}_{12}; \tag{25}$$

$$x_j \in \beta_2 \quad \text{if} \left[\frac{(x_{jk}-\mu_{2k})^2}{\sigma_{2k}^2} - \frac{(x_{jk}-\mu_{1k})^2}{\sigma_{1k}^2} - 2\ln\left(\frac{\sigma_{1k}}{\sigma_{2k}}\right)\right] < -\tilde{\Delta}_{12}; \tag{26}$$

$$\text{and } x_j \in \beta_3 \quad \text{if} \left|\left[\frac{(x_{jk}-\mu_{2k})^2}{\sigma_{2k}^2} - \frac{(x_{jk}-\mu_{1k})^2}{\sigma_{1k}^2} - 2\ln\left(\frac{\sigma_{1k}}{\sigma_{2k}}\right)\right]\right| \leq \tilde{\Delta}_{12}; \tag{27}$$

$$\text{where} \quad \tilde{\Delta}_{12} = \frac{\Delta_{12}}{m} = 2\left[\ln\left(\frac{1+\delta}{1-\delta}\right) - \ln\frac{P(\omega_1)}{P(\omega_2)}\right]. \tag{28}$$

This can be viewed as a supervised granulation process, which utilizes class information. The parameter δ has the same interpretation as that of λ in (17). Combining (17) and (28), the following relation holds between λ and δ:

$$\lambda = 1 - 2\left[1 + \left(\frac{1+\delta}{1-\delta}\right)^{m}\left(\frac{P(\omega_1)}{P(\omega_2)}\right)^{1-m}\right]^{-1}. \tag{29}$$

Let $\omega_i \subseteq \mathbb{U}$, and $E(\beta_p, \omega_i)$ be the relative degree of misclassification of the set β_p with respect to the set ω_i, which is as follows [42]:

$$E(\beta_p, \omega_i) = \begin{cases} 1 - \frac{|\beta_p \cap \omega_i|}{|\beta_p|}, & \text{if } |\beta_p| > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

The quantity $E(\beta_p, \omega_i)$ is also referred to as the relative classification error. The number of misclassified objects is given by the product of $E(\beta_p, \omega_i)$ and $|\beta_p|$, which is referred to as an absolute classification error. According to the theory of variable precision rough sets [42], each class ω_i can be approximated using the measure $E(\beta_p, \omega_i)$ by constructing the α-lower and α-upper approximations of ω_i, where α is the admissible classification error and $\alpha \in [0.0, 0.5)$ as per the majority requirement. The α-lower approximation $\underline{A}_\alpha(\omega_i)$, α-upper approximation $\overline{A}_\alpha(\omega_i)$, and α-boundary region $B_\alpha(\omega_i)$ of the class ω_i can be defined as follows [42]:

$$\underline{A}_\alpha(\omega_i) = \bigcup \{\beta_p \mid E(\beta_p, \omega_i) \leq \alpha\}; \tag{31}$$

$$\overline{A}_\alpha(\omega_i) = \bigcup \{\beta_p \mid E(\beta_p, \omega_i) < 1 - \alpha\}; \tag{32}$$

$$B_\alpha(\omega_i) = \bigcup \{\beta_p \mid \alpha < E(\beta_p, \omega_i) < 1 - \alpha\}. \tag{33}$$

Hence, the lower approximation $\underline{A}_\alpha(\omega_i)$ is the collection of those elements of $\mathbb{U}$ that can be classified into ω_i with the classification error not greater than α.

Using (24) and (31), the α-dependency $\gamma_{\mathcal{A}_k}(\mathbb{D})$, also known as degree of dependency, of decision attribute $\mathbb{D}$ on the condition attribute $\mathcal{A}_k$ can be computed as follows [35,42]:

$$\gamma_{\mathcal{A}_k}(\mathbb{D}) = \frac{|POS^{\mathbb{U}}_{\mathcal{A}_k}(\mathbb{D})|}{|\mathbb{U}|} = \frac{1}{n}\left|\bigcup_{\omega_i \in \mathbb{U}/\mathbb{D}} \underline{A}_\alpha(\omega_i)\right|; \tag{34}$$

where $\gamma_{\mathcal{A}_k}(\mathbb{D}) \in [0, 1]$. However, as α-lower approximation of a class contains both correctly classified and misclassified objects, the γ measure of (34) fails to capture discriminative characteristics of a feature properly. In order to circumvent the above problem, a relevance measure is introduced next by discarding the adverse effect of misclassified objects.

The correctly classified objects belonging to the α-lower approximations of c classes can be arrayed as a $(c \times n)$ matrix $\mathbb{L}(\mathcal{A}_k) = [\mathrm{L}_{ij}(\mathcal{A}_k)]$, where i-th row of the matrix

$$\mathrm{L}_i(\mathcal{A}_k) = \{\mathrm{L}_{i1}(\mathcal{A}_k)/x_1 + \mathrm{L}_{i2}(\mathcal{A}_k)/x_2 + \cdots + \mathrm{L}_{in}(\mathcal{A}_k)/x_n\}$$

corresponds to α-lower approximation $\underline{A}_\alpha(\omega_i)$ of the i-th class ω_i, "+" means the operator of union in this case, and

$$\mathrm{L}_{ij}(\mathcal{A}_k) = \begin{cases} 1, & \text{if } x_j \in \underline{A}_\alpha(\omega_i) \text{ and } x_j \in \omega_i; \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

So, $\mathrm{L}_i(\mathcal{A}_k) \subseteq \underline{A}_\alpha(\omega_i)$. Since a correctly classified object belongs to the lower approximation of only one class, each column of $\mathbb{L}(\mathcal{A}_k)$ must contain at most one 1. Based on the above definition, the relevance of a condition attribute $\mathcal{A}_k$, with respect to c classes $\{\omega_1, \cdots, \omega_i, \cdots, \omega_c\}$, is defined as

$$\mathrm{Rel}_{\mathcal{A}_k}(\mathbb{D}) = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} \mathrm{L}_{ij}(\mathcal{A}_k); \tag{36}$$

where $0 \leq \mathrm{Rel}_{\mathcal{A}_k}(\mathbb{D}) \leq 1$. Higher value of relevance measure indicates better attribute or feature for classification. In this regard, it should be noted that, for $\alpha = 0$, the definitions of set or class approximations, given in (31), (32) and (33) based on variable precision rough sets, reduce to that of Pawlak's rough sets [35]. So, for $\alpha = 0$, $\mathrm{L}_i(\mathcal{A}_k) = \underline{A}_0(\omega_i) = \underline{A}(\omega_i)$, and the relevance measure $\mathrm{Rel}_{\mathcal{A}_k}(\mathbb{D})$ defined in (36) is exactly same with the degree of dependency $\gamma_{\mathcal{A}_k}(\mathbb{D})$ reported in (34).

Given a set $\mathbb{C} = \{\mathcal{A}_1, \cdots, \mathcal{A}_k, \cdots, \mathcal{A}_m\}$ of m condition attributes, in general, the equivalence classes corresponding to the set $\mathbb{C}$ can be formed from that of individual attributes. Based on the resultant equivalence classes of $\mathbb{C}$, the α-lower approximations of c classes can be found out using (31), and accordingly $\gamma_{\mathbb{C}}(\mathbb{D})$ or $\mathrm{Rel}_{\mathbb{C}}(\mathbb{D})$ can be computed using (34) or (36), respectively. However, as both $\gamma_{\mathbb{C}}(\mathbb{D})$ and $\mathrm{Rel}_{\mathbb{C}}(\mathbb{D})$, computed this way, consider only dependency in multidimensional feature space and do not take into account the relevance of individual attributes, they fail to identify relevant features [41,43]. In order to identify relevant as well as complementary features, the union of m individual $(c \times n)$ matrices $\{\mathbb{L}(\mathcal{A}_k)\}$, corresponding to m condition attributes, is considered, which is as follows:

$$\tilde{\mathbb{L}}(\mathbb{C}) = \bigcup_{\mathcal{A}_k \in \mathbb{C}} \mathbb{L}(\mathcal{A}_k); \tag{37}$$

and corresponding relevance $\mathrm{Rel}_{\mathbb{C}}(\mathbb{D})$ is computed from $\tilde{\mathbb{L}}(\mathbb{C})$ using (36). Obviously,

$$\mathrm{Rel}_{\{\mathcal{A}_k, \mathcal{A}_l\}}(\mathbb{D}) \geq \max\{\mathrm{Rel}_{\mathcal{A}_k}(\mathbb{D}), \mathrm{Rel}_{\mathcal{A}_l}(\mathbb{D})\}. \tag{38}$$

4. Experimental results and discussions

This section presents the performance of the proposed descriptor selection method for HEp-2 cell image classification, along with a comparison with related approaches. The methods compared are several local texture descriptors, namely, LBP [16], $\mathrm{LBP}^{\mathrm{ri}}$ [17], $\mathrm{LBP}^{\mathrm{riu2}}$ [17], CoALBP [18], CLBP [19], and RICLBP [20], as well as several state-of-the-art methods for HEp-2 cell classification. The performance of different methods is also compared with respect to various scales of LBP neighborhood such as 1 (S_1), 2 (S_2), 3 (S_3), 4 (S_4), S_{123}: concatenation of (S_1, S_2, S_3), and S_{124}: concatenation of (S_1, S_2, S_4). The 4-neighborhood is used for CoALBP and RICLBP, while for others 8-neighborhood is considered. The SVM [39] is used to evaluate the performance of different local texture descriptors. Both training-testing and ten-fold cross-validation (CV) are performed to compute the classification accuracy. The comparative performance analysis of different algorithms is also studied using tables of means, medians, standard deviations, and p-values computed through paired-t and Wilcoxon signed-rank tests, considering 95% confidence level. All the methods are run in Ubuntu 18.04.2 LTS environment having machine configuration Intel core i7 64-bit processor, 3.40 GHz, and 64 GB DDR4 RAM, and without any GPU. The source code of the proposed method is available at http://www.isical.ac.in/~bibl/results/rough-bayes/roughbayes.html.

4.1. Description of data sets

Three HEp-2 cell image databases, namely, MIVIA database (ICPR 2012 HEp-2 cell classification contest data set) [44], ICPR image database (ICPR 2014 HEp-2 cell classification contest data set) [38], and SNP HEp-2 database [7], are considered for the evaluation of the proposed method as well as related existing approaches. The cell images of the above three data sets were captured in different laboratory settings. For instance, SNP HEp-2 used objective lens magnitude 20×, while MIVIA HEp-2 used 40×. The MIVIA database contains 1455 cells obtained from 28 images: 721 training cell images and 734 test cell images. Six staining pattern classes,

namely, Cytoplasmic, Fine Speckled, Nucleolar, Coarse Speckled, Homogeneous, and Centromere, are considered in this case. The ICPR database contains the set of 13,596 HEp-2 cell images, which is partitioned into 6797 training and 6799 test cell images. The cell images belong to six different patterns, namely, Centromere, Homogeneous, Nucleolar, Speckled, Nuclear Membrane, and Golgi. The SNP database contains 1806 labelled cell images obtained from forty slide images: 869 training images and 937 test images. The samples are grouped into five pattern classes, namely, Centromere, Homogeneous, Fine Speckled, Coarse Speckled and Nucleolar. Each of the three sets of 1455, 13596, and 1806 cell images of MIVIA, ICPR, and SNP databases, respectively, is also split into ten separate folds for ten-fold CV. In both training and test sets as well as in ten folds, the cell images are almost equally distributed with respect to different staining pattern classes.

4.2. Optimum values of different parameters

The performance of the proposed modality selection method depends on the values of both δ and α. While δ controls the region of uncertainty due to overlapping class boundaries between two pattern classes and inexactness in class definition, α imposes an explicit limitation on the admissible level of classification error. To find out the optimum value of δ, extensive experiment is carried out on three HEp-2 image databases by varying the values of δ from 0.00 to 0.95 and corresponding results are presented in first graph of Fig. 4. From the results reported in Fig. 4, it is seen that the performance of the proposed method increases with the increase in value of δ upto 0.40, irrespective of the databases and experimental setup used. For $\delta > 0.40$, the performance remains almost constant in all the cases. So, the optimum value of δ is considered as 0.40. It ensures that an object definitely belongs to one of the staining pattern classes if its probability of belongingness in that class is at least 0.70.

Keeping the value of δ at 0.4, α is varied from 0.00 to 0.45 and the corresponding results are plotted in middle graph of Fig. 4. All the results reported in Fig. 4 reveal that the stable performance of the algorithm is obtained for medium values of α and considerable classification accuracy is achieved at $\alpha = 0.30$ for all the three databases, irrespective of experimental setup used. Hence, the value of α is considered to be 0.30 for the rest of the study. The optimum values of $\alpha = 0.3$ and $\delta = 0.4$ ensure that an object can be classified to one of the staining pattern classes with classification error not greater than 0.3, if its probability of belongingness in that class is at least 0.7.

The SVM [39] with linear kernel is used to analyze the performance of the proposed method. It is based on supervised statistical learning theory which tries to obtain an optimized hyperplane in the kernel space by maximizing the margin and minimizing the classification error between the given classes. The robust performance of the SVM is achieved by properly adjusting the parameter C, known as the penalty parameter in the kernel function. For large value of C, error minimization is predominant, that is, the optimization will choose a hyperplane with smaller margin so that all the training samples are correctly classified. Conversely, if the value of C is small, then margin maximization is emphasized, even if more training samples are misclassified. So, in the proposed method, exhaustive search is conducted on a finite set of values of C, ranging from 10^{-5} to 10^{5}, and the corresponding results are reported in last graph of Fig. 4. From the results, it can be observed that for $C > 0.01$, the performance of the proposed method remains constant. So, the optimum value of C is set at an intermediate value, which is 1 in this case.

Table 4
Performance of Proposed Method With and Without $\mathcal{S}$ Function and Threshold for 10-Fold CV.

Different	Different	MIVIA					SNP					ICPR				
Measures	Approaches	Mean	Median	StdDev	Paired-t:p	Wilcoxon:p	Mean	Median	StdDev	Paired-t:p	Wilcoxon:p	Mean	Median	StdDev	Paired-t:p	Wilcoxon:p
Accuracy	Method_1	**94.90**	**95.58**	1.67	6.21E-01	7.03E-01	81.91	81.15	6.79	4.24E-01	3.37E-01	**84.67**	84.63	3.13	9.98E-01	9.97E-01
	Method_2	**94.90**	95.24	**1.61**	6.61E-01	6.60E-01	**82.19**	80.87	7.02	6.09E-01	5.24E-01	84.56	**84.67**	**2.90**	9.93E-01	9.91E-01
	Proposed	94.83	94.90	1.93			82.02	**81.42**	**6.32**			84.06	83.75	3.24		
No of Features	Method_1	4188.00	4204.00	22.49	1.00E+00	9.98E-01	4044.80	4096.00	107.94	1.00E+00	9.98E-01	4145.20	4173.00	82.52	1.00E+00	9.98E-01
	Method_2	3690.20	3691.00	20.69	1.00E+00	9.98E-01	4017.20	4043.50	80.41	1.00E+00	9.98E-01	3682.80	3684.00	18.40	1.00E+00	9.98E-01
	Proposed	**2736.40**	**2737.50**	**15.77**			**2655.80**	**2674.50**	**68.10**			**2549.80**	**2547.00**	**15.28**		

Method_1: Without applying second-order derivative of $\mathcal{S}$ function and threshold; Method_2: Without applying threshold.

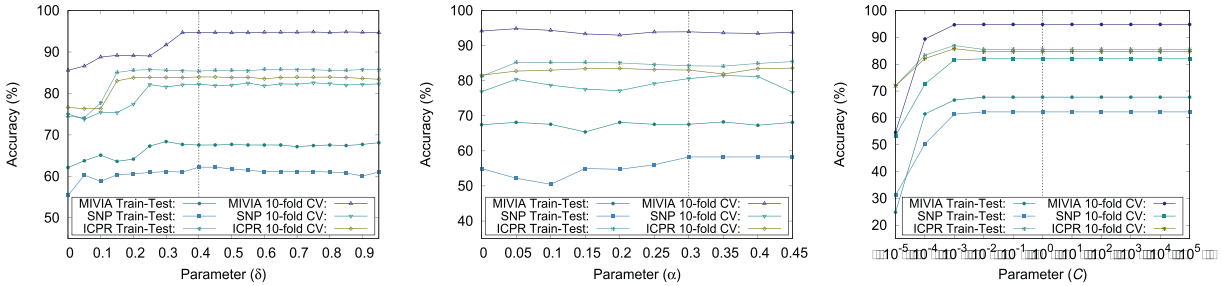


Fig. 4. Variation of classification accuracy for different values of δ, α and C.

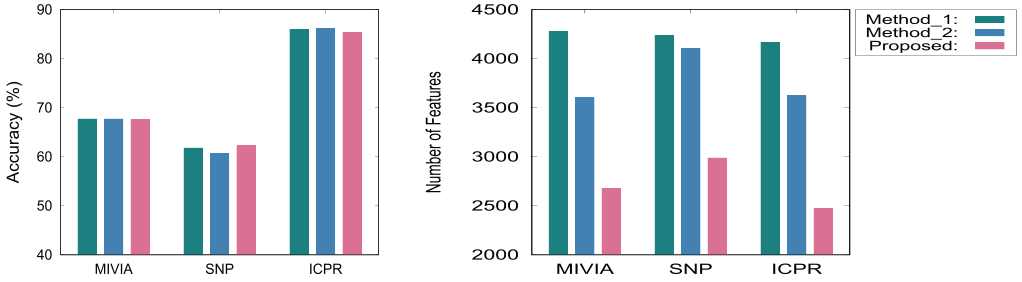


Fig. 5. Performance of proposed method with and without $\mathcal{S}$ function and threshold for training-testing.

4.3. Effectiveness of difference operator and threshold

Prior to selecting relevant local texture descriptors under appropriate scales, the proposed method takes care of the presence of noisy pixels in an HEp-2 cell image as well as noisy images in a staining pattern class. The proposed approach considers the second-order derivative of $\mathcal{S}$ function and class specific threshold, introduced in (1) and (9), respectively, to make itself insensitive to noisy pixels and noisy images. In order to establish the importance of both $\mathcal{S}$ function and threshold, extensive experiment is carried out on three HEp-2 cell image databases. Fig. 5 and Table 4 compare the performance of the proposed method with and without $\mathcal{S}$ function and class specific threshold. The Method_1 is the proposed method without using both second-order derivative of $\mathcal{S}$ function and threshold, while Method_2 is the proposed method without using only class specific threshold. All the results reported in Fig. 5 and Table 4 establish the fact that the introduction of both $\mathcal{S}$ function and threshold helps to identify class specific relevant features. While the second-order derivative of $\mathcal{S}$ function helps to eliminate the effect of noisy pixels present in HEp-2 cell images, the class specific threshold takes care of the presence of noisy images in staining pattern classes. In effect, the proposed method obtains comparable classification accuracy, in both training-testing and 10-fold CV, with significantly lesser number of features compared to both Method_1 and Method_2.

4.4. Importance of class-Pair specific modalities

In general, the existing approaches, based on local texture descriptors, consider a fixed set of modalities for all the HEp-2 cell classes, where each modality corresponds to a specific local texture descriptor considered under a particular scale. However, the proposed method assumes that a fixed set of modalities may not be useful for all the classes. Rather, class-pair specific modalities should be considered while analyzing HEp-2 cell images.

To establish the effectiveness of class-pair specific modalities over uniform modalities for all the classes, extensive experiment is carried out on three HEp-2 cell image databases, considering fifteen modalities corresponding to LBP, LBP^{ri}, LBP^{riu2}, and CoALBP, along with their four scales. Fig. 6 and Table 5 present the comparative performance analysis for single and multiple modalities. From the results reported in top row of Fig. 6 for single modality, it can be seen that the proposed method attains highest classification accuracy of training-testing in most of the cases. Similarly, all the results reported in bottom row of Fig. 6, corresponding to three modalities, confirm that the proposed method achieves highest classification accuracy in all the cases. The results reported in Table 5, corresponding to 10-fold CV, show that the proposed method obtains highest mean and median values, irrespective of the databases and number of modalities used. Also, it attains significantly better results in all 132 cases, irrespective of statistical significance tests.

In order to establish the effectiveness of the proposed approach under other sets of modalities, Fig. 7 and Table 6 compare the performance of the proposed method with that of individual modalities, where the proposed method signifies selection of class-pair relevant modality from the following different sets of modalities: (a) a set of six modalities - Gabor features, gray level co-occurrence matrix (GLCM), histogram of oriented gradients (HOG), local phase quantization (LPQ), binarized statistical image features (BSIF), and deep features extracted using CNN [27]; (b) a set of four modalities - LPQ under four scales or window sizes; (c) a set of four modalities - BSIF under four scales; (d) a set of eight modalities - sign (S) and magnitude (M) components of LBP^{ri} under four scales; (e) a set of three modalities - deep features extracted at three layers of CNN [27]; and (f) a set of seven modalities - CLBP under four scales and RICLBP at three scales. From all the results reported in Fig. 7 and Table 6, it is seen that the proposed method is able to identify single relevant modality for each class-pair, irrespective of descriptors and scales, as well as sign and magnitude components of a descriptor. Also, the accuracy is significantly increased if class-pair specific modality is considered instead of considering a fixed modality for all the classes. Out of total 192 cases, the proposed method achieves statistically significant p-values in 176 cases and better but not significant p-values in 11 cases.

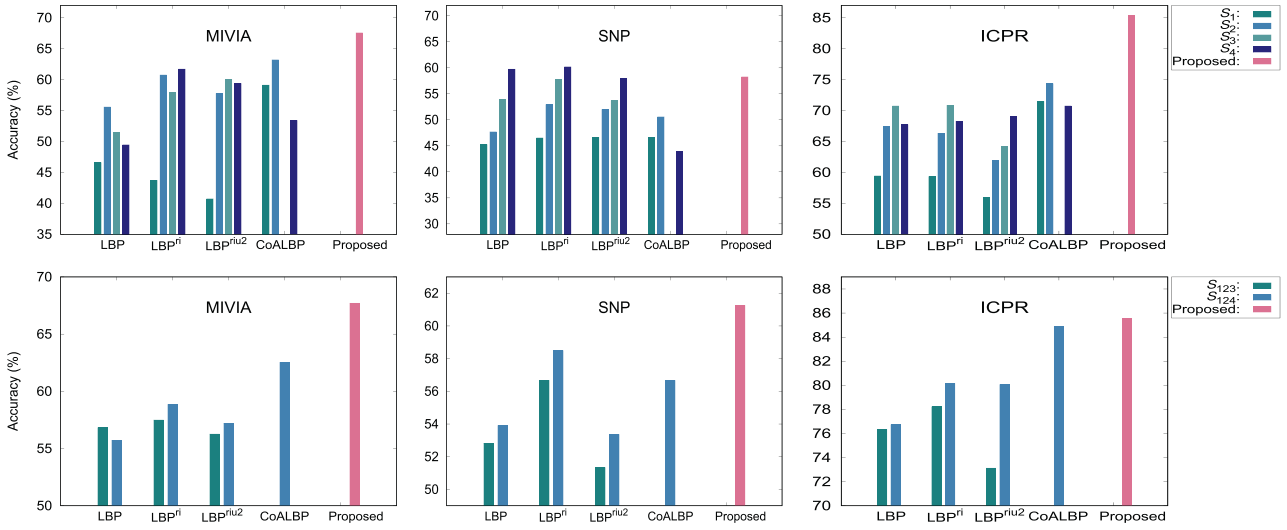


Fig. 6. Performance of several local texture descriptors under different scales and proposed method at single (top row) and multiple (bottom row) modalities for training-testing.

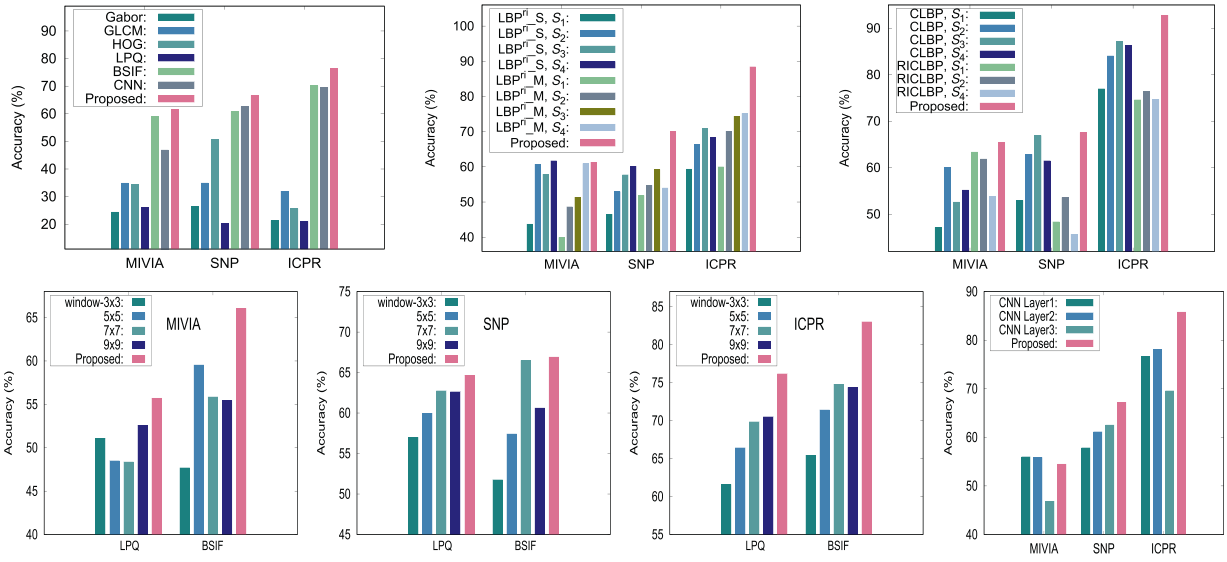


Fig. 7. Performance of the proposed method at single modality considering different sets of descriptors.

4.5. Comparative performance analysis

Finally, the performance of the proposed method is extensively compared with that of several state-of-the art methods for texture classification and HEp-2 cell staining pattern recognition. For the proposed method, single modality corresponding to each class-pair is considered from the set of seven modalities, namely, CLBP under four scales and RICLBP at three scales. Results are reported with respect to overall classification accuracy (OCA), mean class accuracy (MCA) and execution time (in second).

4.5.1. Texture classification methods

This section compares the performance of the proposed descriptor selection method with that of several state-of-the art texture classification methods, namely, dominant LBP (DLBP) [25], discriminative features for texture description (DFTD) [45], restricted Boltzmann machine (RBM) [46], discriminative deep Belief networks (disDBN) [26], CNN [27], deep encoding pooling (DEP) [47], and CNN_mRMR - a deep architecture combining CNN and minimum redundancy-maximum relevance (mRMR) framework [48]. Results are reported in Table 7 on three HEp-2 cell databases for training-testing. The performance of both DLBP and DFTD is evaluated for S_{123} and S_{124}. While DLBP considers dominant features of LBP histogram [25], DFTD takes into account LBP^{ri} features to form the discriminative set [45]. The SVM with linear kernel is used for staining pattern classification in all the cases. All the results reported in Table 7 establish the fact that the proposed method attains highest values of both OCA and MCA, irrespective of the databases used. Also, the execution time required by the proposed method is significantly lower than that of state-of-the-art deep learning based texture classification methods.

4.5.2. HEp-2 Cell classification methods

This section compares the performance of the proposed HEp-2 cell staining pattern recognition method with that of several state-of-the art staining pattern classification methods, namely, RICLBP [20], Fisher tensor (FT) [14], automated pattern recognition system (APRS) [49], deep CNN model of Gao et al. (deepCNN) [33], dual convolutional auto-encoder (DCAE) [30], CNN using several pre-processing techniques (CNNPT) [32], deep CNN model of Jia et al. (DCNN) [31], and deep cross residual network (DCRNet_v2) [50]. Corresponding results are reported in Table 8 with respect to OCA,

Table 5
Comparative Performance Analysis at Single and Multiple Modalities for 10-Fold CV.

Different	Different	MIVIA					SNP					ICPR				
Descriptors	Scales	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p
LBP	S_1	74.01	74.49	2.44	1.10E-09	2.53E-03	60.71	60.93	7.82	1.25E-07	2.47E-03	57.44	56.31	8.26	8.50E-08	2.53E-03
	S_2	83.33	83.33	3.52	6.04E-06	2.53E-03	64.48	65.85	8.32	5.23E-07	2.50E-03	67.33	67.35	6.79	7.38E-07	2.53E-03
	S_3	82.45	81.63	2.46	7.20E-07	2.52E-03	66.50	67.49	7.54	2.39E-04	4.67E-03	72.12	72.52	5.02	4.08E-07	2.53E-03
	S_4	81.43	81.29	2.81	1.60E-07	2.47E-03	68.47	66.39	5.55	1.33E-04	3.37E-03	70.01	68.31	5.25	4.59E-07	2.53E-03
LBP^{ri}	S_1	66.80	66.67	3.06	2.93E-12	2.50E-03	62.90	64.75	10.16	3.36E-06	2.52E-03	54.35	55.72	6.70	1.73E-08	2.53E-03
	S_2	79.73	80.61	4.18	2.87E-07	2.52E-03	72.57	74.86	7.09	1.17E-03	4.67E-03	64.82	66.03	7.07	5.43E-07	2.53E-03
	S_3	83.61	82.99	2.45	1.26E-08	2.45E-03	73.88	73.50	6.23	3.81E-03	8.30E-03	67.34	66.69	7.22	3.12E-06	2.53E-03
	S_4	81.63	81.63	2.36	1.10E-07	2.50E-03	73.39	71.59	6.01	8.12E-04	4.67E-03	69.84	68.97	6.48	2.90E-06	2.53E-03
LBP^{riu2}	S_1	68.84	68.03	3.64	6.21E-09	2.52E-03	63.11	63.66	10.38	1.77E-07	2.53E-03	53.40	54.29	7.36	2.37E-08	2.53E-03
	S_2	78.50	78.91	3.71	4.34E-08	2.52E-03	70.00	70.22	8.51	1.83E-05	2.52E-03	60.62	60.60	5.20	1.39E-09	2.52E-03
	S_3	82.52	83.67	4.95	2.96E-05	2.52E-03	69.89	69.67	7.90	1.66E-05	2.53E-03	63.72	64.64	5.53	6.15E-08	2.53E-03
	S_4	83.67	85.03	4.16	7.15E-05	2.52E-03	69.73	69.95	10.13	1.27E-05	2.50E-03	62.52	63.50	7.38	3.86E-06	2.53E-03
CoALBP	S_1	88.23	88.44	2.20	2.15E-06	2.46E-03	69.29	71.31	7.66	1.12E-05	2.53E-03	68.88	67.20	6.86	1.88E-06	2.53E-03
	S_2	89.66	89.12	2.26	1.05E-04	2.50E-03	69.34	69.40	7.15	2.09E-06	2.53E-03	73.55	73.04	4.97	9.97E-07	2.53E-03
	S_4	83.06	83.33	3.18	1.84E-06	2.52E-03	57.16	57.10	**4.94**	1.67E-06	2.53E-03	70.62	69.85	3.79	2.61E-07	2.53E-03
Proposed (Single)		**93.95**	**93.20**	**1.55**			**80.60**	**81.69**	8.48			**83.04**	**81.95**	3.05		
LBP	S_{123}	88.23	88.44	3.29	2.43E-04	2.53E-03	74.15	74.04	5.28	1.44E-04	2.52E-03	74.46	72.74	5.21	1.08E-06	2.53E-03
	S_{124}	88.78	89.12	2.43	9.19E-05	2.47E-03	76.61	77.60	5.69	4.87E-04	3.98E-03	74.83	73.40	5.54	5.30E-06	2.53E-03
LBP^{ri}	S_{123}	86.46	86.73	3.21	1.18E-05	2.52E-03	78.14	78.14	5.83	1.21E-02	1.60E-02	77.19	78.28	3.93	3.50E-05	2.53E-03
	S_{124}	85.37	85.71	**1.85**	8.83E-07	2.52E-03	78.91	77.32	**5.19**	9.92E-03	1.42E-02	74.03	75.46	4.64	5.71E-06	2.53E-03
LBP^{riu2}	S_{123}	85.92	86.05	2.66	1.28E-07	2.52E-03	76.89	77.87	7.56	2.04E-03	6.26E-03	70.13	69.99	4.38	1.81E-08	2.53E-03
	S_{124}	86.67	87.07	3.32	1.49E-06	2.53E-03	78.63	78.14	7.33	5.82E-03	1.09E-02	71.60	71.06	4.38	1.03E-05	2.53E-03
CoALBP	S_{124}	92.04	92.18	2.57	2.49E-03	4.56E-03	80.22	80.87	6.36	2.13E-02	1.82E-02	83.35	83.05	3.70	3.30E-02	4.63E-02
Proposed (Three)		**94.83**	**94.90**	1.93		**82.02**	**82.02**	**81.42**	6.32			**84.06**	**83.75**	**3.24**		

Table 6
Comparative Performance Analysis of Different Descriptors at Single Modality for 10-Fold CV.

Different	Different	MIVIA					SNP					ICPR				
Descriptors	Scales	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p
Gabor		32.04	32.65	3.31	7.43E-12	2.53E-03	31.15	30.60	5.49	1.25E-10	2.53E-03	36.48	43.18	11.23	4.30E-09	2.53E-03
GLCM		53.40	53.74	4.08	1.04E-09	2.53E-03	53.17	53.28	12.06	9.49E-06	2.53E-03	41.30	45.38	11.83	3.73E-08	2.53E-03
HOG		42.72	43.20	4.34	2.37E-09	2.52E-03	54.10	56.01	10.20	4.74E-08	2.52E-03	36.82	38.66	**5.43**	7.29E-12	2.53E-03
LPQ		75.92	76.19	**2.59**	4.99E-06	2.53E-03	65.25	63.66	6.80	1.50E-05	2.53E-03	63.97	65.04	10.38	5.26E-06	2.53E-03
BSIF		79.66	80.95	3.21	2.35E-05	2.53E-03	66.99	68.58	9.63	1.01E-04	2.53E-03	66.02	66.84	8.39	5.48E-07	2.53E-03
CNN		76.53	77.89	4.39	1.75E-06	2.47E-03	68.58	64.75	9.27	8.63E-06	2.53E-03	70.60	70.98	7.64	2.42E-06	2.53E-03
	Proposed	**89.12**	**89.80**	**3.53**			**81.64**	**81.15**	**6.42**			76.52	77.06	6.45		
LPQ	3 × 3	75.92	76.19	2.59	1.63E-08	2.46E-03	65.25	63.66	6.80	9.22E-03	1.09E-02	63.97	65.04	10.38	1.20E-05	2.53E-03
LPQ	5 × 5	78.44	77.55	4.11	1.78E-05	2.53E-03	66.78	66.94	7.54	9.90E-02	1.93E-01	69.05	70.95	7.88	2.72E-06	2.53E-03
LPQ	7 × 7	79.59	80.27	2.62	8.29E-08	2.52E-03	71.86	72.68	8.30	4.78E-01	5.61E-01	71.33	70.80	7.27	5.44E-06	2.53E-03
LPQ	9 × 9	78.98	78.57	3.52	6.21E-08	2.52E-03	69.51	67.76	7.17	3.00E-01	3.23E-01	71.99	71.42	8.13	3.58E-03	6.26E-03
	Proposed	**89.18**	**89.12**	**2.23**			**74.64**	**73.77**	**6.06**			**75.63**	**75.31**	**6.59**		
BSIF	3 × 3	79.66	80.95	3.21	9.31E-08	2.52E-03	66.99	68.58	9.63	7.36E-06	2.52E-03	66.02	66.84	8.39	1.36E-06	2.53E-03
BSIF	5 × 5	85.78	85.37	2.09	5.81E-08	2.47E-03	70.00	68.03	**5.73**	6.60E-06	2.52E-03	72.76	72.63	5.63	1.81E-07	2.53E-03
BSIF	7 × 7	85.92	86.73	3.19	1.94E-05	2.52E-03	74.32	72.13	5.94	6.20E-04	2.50E-03	75.22	73.70	5.46	8.23E-06	2.53E-03
BSIF	9 × 9	83.33	82.99	2.89	5.93E-07	2.52E-03	70.66	67.76	7.23	1.26E-05	2.46E-03	75.55	74.14	6.00	3.14E-04	2.53E-03
	Proposed	**93.95**	**93.54**	**1.85**			**80.00**	**79.78**	6.68			**80.70**	**79.82**	**4.63**		

(*continued on next page*)

Table 6 (*continued*)

Different Descriptors	Different Scales	MIVIA					SNP					ICPR				
		Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p	Mean	Median	StDv	Paired-*t*:p	Wilcoxon:p
LBP^{ri}_S	S_1	66.80	66.67	3.06	1.24E-09	2.52E-03	62.90	64.75	10.16	4.38E-05	2.53E-03	59.61	59.26	2.63	3.67E-08	2.53E-03
LBP^{ri}_S	S_2	79.73	80.61	4.18	2.47E-05	2.50E-03	72.57	74.86	7.09	1.61E-04	2.52E-03	67.55	68.19	**1.82**	5.64E-07	2.53E-03
LBP^{ri}_S	S_3	83.61	82.99	2.45	1.80E-05	2.52E-03	73.88	73.50	6.23	5.47E-05	2.53E-03	69.23	69.15	1.92	9.89E-07	2.53E-03
LBP^{ri}_S	S_4	81.63	81.63	2.36	5.34E-07	2.50E-03	73.39	71.58	6.01	2.72E-05	2.52E-03	72.04	71.76	2.07	2.98E-06	2.53E-03
LBP^{ri}_M	S_1	67.62	68.37	5.01	5.87E-08	2.53E-03	63.17	65.57	11.77	1.49E-04	2.53E-03	54.58	55.58	5.69	7.82E-09	2.53E-03
LBP^{ri}_M	S_2	80.41	79.25	3.22	1.09E-06	2.50E-03	71.86	71.04	7.59	1.29E-04	2.52E-03	65.79	65.22	9.04	7.61E-07	2.53E-03
LBP^{ri}_M	S_3	79.05	77.89	4.10	4.57E-06	2.52E-03	75.25	72.40	8.34	1.08E-03	3.46E-03	68.04	67.46	8.81	2.90E-06	2.53E-03
LBP^{ri}_M	S_4	75.24	74.83	4.75	2.26E-06	2.52E-03	76.94	76.23	4.29	7.89E-05	2.53E-03	69.82	67.98	7.11	4.61E-07	2.53E-03
Proposed		**90.00**	**90.14**	**1.70**			**84.59**	**84.43**	**3.65**			**87.21**	**86.13**	4.10		
CNN	Layer 1	82.79	81.97	3.19	4.85E-07	2.52E-03	75.30	75.41	6.80	1.22E-05	2.53E-03	73.83	72.96	4.33	1.17E-07	2.53E-03
	Layer 2	87.21	87.41	2.72	2.82E-05	2.53E-03	78.20	77.05	6.29	2.18E-06	2.50E-03	74.64	74.17	4.43	8.28E-07	2.53E-03
	Layer 3	76.53	77.89	4.39	2.12E-07	2.52E-03	68.58	64.75	9.27	1.20E-05	2.53E-03	70.60	70.98	7.64	5.45E-06	2.52E-03
Proposed		**94.42**	**94.56**	**1.46**			**87.05**	**87.16**	**4.25**			**84.72**	**84.30**	**2.91**		
CLBP	S_1	88.57	88.44	3.17	8.09E-05	2.53E-03	78.58	77.87	7.28	5.51E-01	8.44E-02	74.87	76.96	7.37	4.29E-06	2.53E-03
CLBP	S_2	90.95	91.16	3.04	6.11E-04	3.46E-03	**82.02**	81.69	5.22	7.38E-01	1.20E-01	81.88	81.33	4.10	1.70E-07	2.53E-03
CLBP	S_3	90.61	90.14	**1.50**	2.99E-07	2.49E-03	81.80	80.60	7.09	7.30E-01	2.22E-01	85.48	84.19	4.11	1.60E-05	2.53E-03
CLBP	S_4	90.34	90.48	1.86	2.63E-06	2.50E-03	80.66	77.87	6.21	6.76E-01	1.31E-01	85.52	84.67	3.32	1.42E-05	2.53E-03
RICLBP	S_1	87.62	87.76	2.58	1.64E-06	2.49E-03	72.62	72.40	8.00	2.60E-01	3.72E-02	70.59	69.88	7.30	4.54E-07	2.53E-03
RICLBP	S_2	88.03	88.10	2.36	7.04E-07	2.49E-03	70.38	68.58	8.06	1.49E-01	3.72E-02	74.95	73.55	4.11	8.35E-10	2.53E-03
RICLBP	S_4	84.22	84.35	2.95	6.43E-07	2.53E-03	57.60	57.10	**3.96**	5.95E-03	3.72E-02	69.99	69.04	4.99	2.01E-08	2.53E-03
Proposed		**96.60**	**96.60**	1.73			77.65	**84.15**	21.15			**90.11**	**89.10**	**2.50**		

Table 7
Performance Analysis of Proposed Approach and Different Texture Classification Methods.

Different	MIVIA			SNP			ICPR		
Methods	OCA	MCA	Time	OCA	MCA	Time	OCA	MCA	Time
DLBP, S_{123}	50.00	51.41	2.33	44.18	44.27	2.76	57.42	53.80	67.30
DLBP, S_{124}	50.95	52.08	2.34	41.30	41.34	2.54	56.45	53.67	67.30
DFTD, S_{123}	55.45	55.61	1.13	49.20	49.19	1.46	71.72	70.21	58.96
DFTD, S_{124}	57.49	57.95	1.14	52.19	52.05	1.48	73.03	71.56	58.96
RBM	33.51	29.08	2826.39	47.49	47.38	3209.72	67.58	67.47	4558.93
disDBN	60.33	57.56	3114.46	54.66	52.69	7195.57	74.75	74.86	8889.01
CNN	46.87	39.39	2538.84	62.54	62.58	2888.41	69.57	69.33	18748.66
DEP	62.40	63.71	18353.64	57.20	56.38	22795.76	80.06	79.41	84428.37
CNN_mRMR	57.36	59.42	24631.54	50.48	49.46	27573.80	81.72	81.24	96925.61
Proposed	65.40	66.73	58.59	67.56	67.49	66.38	92.79	92.18	762.75

Table 8
Performance Analysis of Proposed Approach and Different HEp-2 Cell Classification Methods.

Different	MIVIA			SNP			ICPR		
Methods	OCA	MCA	Time	OCA	MCA	Time	OCA	MCA	Time
RICLBP	63.49	64.67	17.98	63.82	63.82	14.56	83.91	81.70	179.85
FT	46.87	51.31	6374.39	48.03	48.01	3808.18	79.03	75.67	12820.44
APRS	54.90	52.61	-	53.79	53.60	-	*	*	-
deepCNN	64.03	64.17	2527.28	70.44	70.55	2875.60	84.95	84.63	18604.08
DCAE	42.64	46.52	8532.00	41.62	41.58	9120.96	62.49	62.91	57565.31
CNNPT	50.32	48.77	29999.30	47.28	46.41	35537.22	84.80	84.72	97760.40
DCNN	46.19	49.53	5278.31	46.21	45.13	6221.17	68.50	68.93	48444.35
DCRNet_v2	66.49	66.99	12539.98	52.29	51.32	13947.27	87.20	85.60	871304.30
Proposed	65.40	66.73	58.59	67.56	67.49	66.38	92.79	92.18	762.75

MCA and execution time (in second). While RICLBP, FT, and APRS methods rely on handcrafted features, other approaches consider features extracted using deep architectures for HEp-2 cell classification. From the results presented in Table 8, it can be seen that the proposed method for HEp-2 cell staining pattern recognition outperforms several state-of-the-art methods on all the three data sets, irrespective of the quantitative indices used. However, deepCNN [33] performs better than the proposed method only on SNP database, while DCRNet_v2 [50] attains slightly higher accuracy on MIVIA database.

It is worth mentioning here that APRS [49] considers the set of all 13,596 images of ICPR data set to train the model, which includes the entire set of 6799 test images considered in the current study. So, the performance of APRS on ICPR data cannot be evaluated, and hence, corresponding entries are denoted by "*" in Table 8. To compare the performance of the proposed method with that of APRS on ICPR data, the HEp-2 cell images of Task 2 set of ICPR 2014 contest are used as test set. For Task 2, 1008 specimen images, belonging to seven HEp-2 cell classes, are made available. In [49], few cells from the specimen images have been extracted. Among them, 5032 cell images, along with the corresponding annotations, are shared by the authors of [49] for comparative performance evaluation. The APRS method achieves 82.05% OCA and 83.04% MCA on Task 2 data set, while the proposed method attains 72.48% OCA and 73.04% MCA. As the training model of APRS is provided by the authors of [49], the corresponding execution time, denoted by "-", is ignored in Table 8.

The better performance of the proposed method is achieved due to the fact that it considers class-pair specific modalities for analyzing HEp-2 cell images, rather than considering a fixed set of modalities for all the staining pattern classes. A set of dominant features is extracted under each modality, based on the probability of texture pattern occurrence, which makes the proposed algorithm insensitive to noisy HEp-2 cell images. Moreover, the proposed approach employs rough sets and Bayes decision theory to evaluate the relevance of each modality. While former deals with the uncertainty due to inexactness, vagueness, and incompleteness in HEp-2 class definition, latter addresses the uncertainty due to overlapping class boundaries. In effect, the proposed method provides significantly better performance as compared to existing methods.

5. Conclusion and future directions

The major contribution of this paper is three-fold, namely, (i) development of a method to select a set of relevant texture descriptors under appropriate scales for HEp-2 cell pattern classification; (ii) introducing a new relevance measure, based on judicious integration of variable precision rough sets and Bayes decision theory; and (iii) demonstrating the effectiveness of the proposed method on several benchmark HEp-2 cell image databases.

The proposed descriptor selection method first selects local texture descriptors under appropriate scales for a class-pair, and then forms the final feature set for multiple classes from the set of descriptors of all possible pairs of classes. To judge the quality of a descriptor and/or scale, the theory of rough sets is judiciously integrated with Bayes decision theory. To make the proposed descriptor selection method insensitive to noisy pixels present in an HEp-2 cell image and noisy HEp-2 cell images in a staining pattern class, the concept of significant descriptors of an HEp-2 cell image and a staining pattern class has been introduced. Finally, the SVM with linear kernel is used to predict the class of HEp-2 cell images. The performance of the proposed approach is exhibited on three benchmark HEp-2 cell image databases, along with a comparative study with state-of-the-art methods. For three HEp-2 cell data sets, significantly better results are found for the proposed method compared to several existing methods.

The results obtained on different HEp-2 cell image databases demonstrate that the proposed method is capable of identifying discriminative and significant textural features under appropriate

scales that may contribute to revealing the underlying staining pattern class structures. The method is only used for HEp-2 cell staining pattern classification problem. In future, this method will be extended to other texture classification tasks, and further its merits and limitations will be evaluated. Other texture descriptors will also be considered to study whether incorporation of new descriptors can improve the staining pattern recognition rate. The proposed method will also be applied on other HEp-2 cell image databases such as AIDA database [51] in near future. A method will be developed based on some quantitative measures to find out the optimum values of different parameters of the proposed method.

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.

Acknowledgment

This publication is an outcome of the R&D work undertaken in the project under the Visvesvaraya PhD Scheme of Ministry of Electronics and Information Technology, Government of India, being implemented by Digital India Corporation.

References

[1] P.L. Meroni, P.H. Schur, ANA Screening: an old test with new recommendations, Autoimmun Rev 69 (8) (2010) 1420–1422.
[2] P. Soda, G. Iannello, Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis, IEEE Trans. Inf. Technol. Biomed. 13 (3) (2009) 322–329.
[3] P. Soda, G. Iannello, M. Vento, A multiple expert system for classifying fluorescent intensity in antinuclear autoantibodies analysis, Pattern Analysis and Applications 12 (3) (2009) 215–226.
[4] G. Iannello, G. Percannella, P. Soda, M. Vento, Mitotic cells recognition in HEp-2 images, Pattern Recognit. Lett. 45 (2014) 136–144.
[5] A. Banerjee, P. Maji, Rough-probabilistic clustering and hidden markov random field model for segmentation of HEp-2 cell and brain MR images, Appl. Soft. Comput. 46 (2016) 558–576.
[6] D.H. Solomon, A.J. Kavanaugh, P.H. Schur, Evidence-Based guidelines for the use of immunologic tests: antinuclear antibody testing, Arthritis Rheum. 47 (4) (2002) 434–444.
[7] A. Wiliem, Y. Wong, C. Sanderson, P. Hobson, S. Chen, B.C. Lovell, Classification of Human Epithelial Type 2 Cell Indirect Immunofluoresence Images Via Codebook Based Descriptors, in: Proceedings of the IEEE Workshop on Applications of Computer Vision, 2013, pp. 95–102.
[8] S.D. Cataldo, A. Bottino, I. Islam, T.F. Vieira, E. Ficarra, Subclass discriminant analysis of morphological and textural features for HEp-2 staining pattern classification, Pattern Recognit. 47 (7) (2014) 2389–2399.
[9] V. Snell, W. Christmas, J. Kittler, HEp-2 Fluorescence pattern classification, Pattern Recognit. 47 (7) (2014) 2338–2347.
[10] R. Stoklasa, T. Majtner, D. Svoboda, Efficient k-NN based HEp-2 cells classifier, Pattern Recognit. 47 (7) (2014) 2409–2418.
[11] I. Theodorakopoulos, D. Kastaniotis, G. Economou, S. Fotopoulos, HEp-2 Cells classification via sparse representation of textural features fused into dissimilarity space, Pattern Recognit. 47 (7) (2014) 2367–2378.
[12] A. Wiliem, C. Sanderson, Y. Wong, P. Hobsone, R.F. Minchin, B.C. Lovell, Automatic classification of human epithelial type 2 cell indirect immunofluorescence images using cell pyramid matching, Pattern Recognit. 47 (7) (2014) 2315–2324.
[13] D.G. Lowe, Distinctive image features from scale-Invariant keypoints, Int. J. Comput. Vis. 60 (2004) 91–110.
[14] M. Faraki, M.T. Harandi, A. Wiliem, B.C. Lovell, Fisher tensors for classifying human epithelial cells, Pattern Recognit. 47 (7) (2014) 2348–2359.
[15] X. Kong, K. Li, J. Cao, Q. Yang, L. Wenyin, HEp-2 Cell pattern classification with discriminative dictionary learning, Pattern Recognit. 47 (7) (2014) 2379–2388.
[16] T. Ojala, M. Pietikinen, D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognit. 29 (1) (1996) 51–59.
[17] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-Scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[18] R. Nosaka, Y. Ohkawa, K. Fukui, Feature Extraction Based on Co-Occurrence of Adjacent Local Binary Patterns, in: Proceedings of the Advances in Image and Video Technology, 2012, pp. 82–91.
[19] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classification, IEEE Trans. Image Process. 19 (6) (2010) 1657–1663.
[20] R. Nosaka, K. Fukui, HEp-2 Cell classification using rotation invariant co-Occurrence among local binary patterns, Pattern Recognit. 47 (7) (2014) 2428–2436.
[21] L. Liu, S. Lao, P.W. Fieguth, Y. Guo, X. Wang, M. Pietikainen, Median robust extended local binary pattern for texture classification, IEEE Trans. Image Process. 25 (3) (2016) 1368–1381.
[22] A. Fathi, A.R. Naghsh-Nilchi, Noise tolerant local binary pattern operator for efficient texture analysis, Pattern Recognit. Lett. 33 (9) (2012) 1093–1100.
[23] X. Qi, R. Xiao, C.G. Li, Y. Qiao, J. Guo, X. Tang, Pairwise rotation invariant co-Occurrence local binary pattern, IEEE Trans. Pattern Anal. Mach. Intell. 36 (11) (2014) 2199–2213.
[24] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Trans. Image Process. 19 (6) (2010) 1635–1650.
[25] S. Liao, M.W.K. Law, A.C.S. Chung, Dominant local binary patterns for texture classification, IEEE Trans. Image Process. 18 (5) (2009) 1107–1118.
[26] Y. Liu, S. Zhou, Q. Chen, Discriminative deep belief networks for visual data classification, Pattern Recognit. 44 (10–11) (2011) 2287–2296.
[27] D. Cascio, V. Taormina, G. Raso, Deep CNN for IIF images classification in autoimmune diagnostics, Applied Sciences 9 (8) (2019) 1618.
[28] Y. Li, L. Shen, S. Yu, HEp-2 Specimen image segmentation and classification using very deep fully convolutional network, IEEE Trans. Med. Imaging 36 (7) (2017) 1561–1572.
[29] H. Lei, T. Han, F. Zhou, Z. Yu, J. Qin, A. Elazab, B. Lei, A deeply supervised residual network for HEp-2 cell classification via cross-Modal transfer learning, Pattern Recognit. 79 (2018) 290–302.
[30] C. Vununu, S.-H. Lee, K.-R. Kwon, A deep feature extraction method for HEp-2 cell image classification, Electronics (Basel) 8 (1) (2019) 20.
[31] X. Jia, L. Shen, X. Zhou, S. Yu, Deep Convolutional Neural Network Based HEp-2 Cell Classification, in: Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 77–80.
[32] L.F. Rodrigues, M.C. Naldi, J.F. Mari, Comparing convolutional neural networks and preprocessing techniques for HEp-2 cell classification in immunofluorescence images, Comput. Biol. Med. 116 (2020) 103542.
[33] Z. Gao, L. Wang, L. Zhou, J. Zhang, HEp-2 Cell image classification with deep convolutional neural networks, IEEE J. Biomed. Health Inform. 21 (2) (2016) 416–428.
[34] Y. Li, L. Shen, HEp-Net: A Smaller And better deep-learning network for HEp-2 cell classification, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 7 (3) (2019) 266–272.
[35] Z. Pawlak, Rough sets: theoretical aspects of reasoning about data, Kluwer Academic Publishers, Dordrecht and Boston and London, 1991. ISBN: 0792314727
[36] D. Cascio, V. Taormina, M. Cipolla, S. Bruno, F. Fauci, G. Raso, A multi-Process system for HEp-2 cells classification based on SVM, Pattern Recognit. Lett. 82 (2016) 56–63.
[37] D. Cascio, V. Taormina, G. Raso, An automatic HEp-2 specimen analysis system based on an active contours model and an SVM classification, Applied Sciences 9 (2) (2019) 307.
[38] P. Hobson, B.C. Lovell, G. Percannella, A. Saggese, M. Vento, A. Wiliem, HEp-2 Staining pattern recognition at cell and specimen levels: datasets, algorithms and results, Pattern Recognit. Lett. 82 (2016) 12–22.
[39] V. Vapnik, The nature of statistical learning theory, Springer, New York, 1995. ISBN: 0-387-94559-8
[40] Y. Qian, J. Liang, W. Pedrycz, C. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (2010) 597–618.
[41] P. Maji, A rough hypercuboid approach for feature selection in approximation spaces, IEEE Trans. Knowl. Data Eng. 26 (1) (2014) 16–29.
[42] W. Ziarko, Variable precision rough set model, J. Comput. Syst. Sci. 46 (1993) 39–59.
[43] P. Maji, S. Paul, Rough set based maximum relevance-Maximum significance criterion and gene selection from microarray data, Int. J. Approximate Reasoning 52 (3) (2011) 408–426.
[44] P. Foggia, G. Percannella, P. Soda, M. Vento, Benchmarking HEp-2 cells classification methods, IEEE Trans. Med. Imaging 32 (10) (2013) 1878–1889.
[45] Y. Guo, G. Zhao, M. Pietikinen, Discriminative features for texture description, Pattern Recognit. 45 (10) (2012) 3834–3843.
[46] H. Larochelle, M. Mandel, R. Pascanu, Y. Bengio, Learning algorithms for the classification restricted boltzmann machine, Journal of Machine Learning Research 13 (Mar) (2012) 643–669.
[47] J. Xue, H. Zhang, K. Dana, Deep Texture Manifold for Ground Terrain Recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 558–567.
[48] M. Toğaçar, B. Ergen, Z. Cömert, Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks, Biocybernetics and Biomedical Engineering 40 (1) (2020) 23–39.
[49] S. Manivannan, W. Li, S. Akbar, R. Wang, J. Zhang, S.J. McKenna, An automated pattern recognition system for classifying indirect immunofluorescence images of HEp-2 cells and specimens, Pattern Recognit. 51 (2016) 12–26.
[50] L. Shen, X. Jia, Y. Li, Deep cross residual network for HEp-2 cell staining pattern classification, Pattern Recognit. 82 (2018) 68–78.
[51] A.B. Elgaaied, D. Cascio, S. Bruno, et al., Computer-Assisted classification patterns in autoimmune diagnostics: the AIDA project, Biomed Res. Int. (2016), doi:10.1155/2016/2073076.

Debamita Kumar has received the B.Tech. degree in Electronics and Communication Engineering from West Bengal University of Technology, India, in 2011, and the

M.Tech degree in Electronics and Telecommunication Engineering from Indian Institute of Engineering Science and Technology, Shibpur, India, in 2016. Currently, she is a Research Fellow in the Machine Intelligence Unit, Indian Statistical Institute, Kolkata. Her research interests include pattern recognition, machine learning, deep learning, medical imaging, and so forth. She has published a few papers in international journals and conferences.

Pradipta Maji is a Professor in the Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. His research interests include pattern recognition, machine learning, computational biology and bioinformatics, medical image processing, and so forth. He is a Fellow of National Academy of Sciences, India. He has received the 2008 Microsoft Young Faculty Award from Microsoft Research Laboratory India Pvt., the 2009 Young Scientist Award from the National Academy of Sciences, India, the 2011 Young Scientist Award from the Indian National Science Academy, India, and the 2015 Young Faculty Research Fellowship from the Ministry of Electronics and Information Technology, Government of India. He has been selected as the 2009 Young Associate of the Indian Academy of Sciences, India and elevated to the 2016 Senior Member of IEEE, USA.