



# CanSuR: a robust method for staining pattern recognition of HEp-2 cell IIF images

Ankita Mandal<sup>1</sup> · Pradipta Maji<sup>1</sup>

Received: 16 August 2018 / Accepted: 19 February 2019 / Published online: 18 March 2019  
© Springer-Verlag London Ltd., part of Springer Nature 2019

## Abstract

The recognition of staining patterns present in human epithelial type 2 (HEp-2) cells helps to diagnose connective tissue disease. In this context, the paper introduces a robust method, termed as CanSuR, for automatic recognition of antinuclear autoantibodies by HEp-2 cell indirect immunofluorescence (IIF) image analysis. The proposed method combines the advantages of a new sequential supervised canonical correlation analysis (CCA), introduced in this paper, with the theory of rough hypercuboid approach. While the proposed CCA efficiently combines the local textural information of HEp-2 cells, derived from various scales of rotation-invariant local binary patterns, the relevant and significant features of HEp-2 cell for staining pattern recognition are extracted using rough hypercuboid approach. Finally, the support vector machine, with radial basis function kernel, is used to recognize one of the known staining patterns present in IIF images. The effectiveness of the proposed staining pattern recognition method, along with a comparison with related approaches, is demonstrated on MIVIA, SNP and ICPR HEp-2 cell image databases. An important finding is that the proposed method performs significantly better than state-of-the-art methods, on three HEp-2 cell image databases with respect to both classification accuracy and F1 score.

**Keywords** HEp-2 cell staining pattern recognition · Local binary pattern · Canonical correlation analysis · Support vector machine

## 1 Introduction

Autoimmune diseases are a group of disorders where the immune system of the affected individual malfunctions and the tissues are attacked by autoantibodies. One aspect of these diseases is the formation of self-antigens or autoantibodies. There are two general groups of autoimmune diseases, namely, organ specific and non-organ specific. In organ specific, a specific organ can be attacked, while in non-organ specific, multiple organ systems can be attacked by the autoantibodies. Hashimoto's thyroiditis is an example of organ-specific autoimmune disease where the thyroid gland is damaged by autoantibodies. An example of

a systemic autoimmune disease is systemic lupus erythematosus where the autoantibodies can attack any organ in the body. One of the standard examples of autoimmune disorders is connective tissue diseases (CTDs), which are characterized by a chronic inflammatory process concerning connective tissues.

The antinuclear antibody (ANA) test is used to predict the presence of autoantibodies in nucleus of the cell. The ANAs are used as markers to detect certain chronic immuno-inflammatory diseases. The detection as well as quantitation of ANAs is pivotal to the diagnosis of many autoimmune diseases [2, 36]. The presence of ANAs can be determined by indirect immunofluorescence (IIF) [11]. The knowledge about the localization of autoantigens is provided by the IIF, which has become a standard method to predict the presence of ANA in patient serum. The most used cell substrate for IIF demonstration of ANA is human epithelial type 2 (HEp-2) cell [21, 38, 44]. Various staining patterns of ANAs, namely, centromere, cytoplasmic, golgi, homogeneous, nucleolar, speckled and other mixed

---

✉ Pradipta Maji  
pmaji@isical.ac.in  
Ankita Mandal  
amandal@isical.ac.in

<sup>1</sup> Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India

patterns, can provide useful information to diagnose CTDs, as different patterns are associated with different autoimmune diseases and/or autoantigens [1, 26, 38]. So, accurate classification of these staining patterns is very essential. In real-life analysis, this process suffers from low-throughput and inter and/or intra-laboratory variance. Visual evaluation of IIF test is quite time-consuming, and subjective to the skill of experts [16]. At least two experts must examine each ANA specimen under a fluorescence microscope.

Computer-aided diagnosis (CAD) systems are used to overcome the shortcomings of manual test procedure by determining the staining patterns of a given HEp-2 cell image automatically [10, 40]. Both time and effort may be reduced by the automation of staining pattern classification of HEp-2 cell images. The automated method can make IIF analysis faster, easier and more reliable. Machine learning techniques are widely used to develop the CAD systems in the field of medicine. The goal of the CAD systems is to recognize one of the known staining HEp-2 patterns present in the IIF images. In recent past, some efforts on IIF image analysis using CAD systems have been done. However, a fully automated system to serve the purpose is yet to be developed [10, 16]. There are several approaches present in the literature, either to automate individual stages or the entire IIF diagnostic procedure. Such procedures consist of mainly five steps, namely, acquisition of images, segmentation of images, detection of mitosis, classification of fluorescence intensity and recognition of staining patterns. Soda et al. [35] reported an autofocus function to deal with photobleaching effect during image acquisition. Using a set of textural and shape features obtained from the images, the quality of fluorescence images has been evaluated in [17]. Based on statistical features, fluorescence intensity classification has been performed in [34]. Rough segmentation of HEp-2 cells from IIF images has been addressed in [3, 4, 33]. In [22], an automatic pattern recognition system using fully convolutional network has been proposed to simultaneously address the segmentation and classification problem of HEp-2 specimen images. Another framework has been developed in [12] for the classification of HEp-2 cell images by utilizing deep convolutional neural networks. A superpixel-based Hep-2 cell classification technique has been introduced in [8], based on sparse codes of image patches.

Textural features can be used to capture the information of surface of the HEp-2 cells, which have unpredictably ambiguous texture. The inherent textures in different HEp-2 cell types are quite different from each other. Moreover, the visual resemblance among the cells of different classes increases the ambiguity of IIF image analysis. These difficulties create limitations of HEp-2 pattern recognition using CAD systems. To characterize a cell image, several

local texture descriptors, namely, local binary pattern (LBP) [29], rotation-invariant LBP (LBP<sup>ri</sup>) [30], completed LBP [15], co-occurrence of adjacent LBPs [28] and rotation-invariant co-occurrence of adjacent LBPs [27], are used in [5], while the concept of gradient-oriented co-occurrence of LBP has been introduced in [40]. In [9], a combination of morphological features and textural features extracted using LBPs is used for automatic mitotic cell recognition. In [19], the similarity-based watershed algorithm with marker techniques has been used to segment HEp-2 cells, while learning vector quantization has been used to identify the patterns. Strandmark et al. [37] have developed an automatic method, based on random forests, which classifies an HEp-2 cell image. On the other hand, Cordelli et al. [6] have proposed a method which is independent from the color model used and showed that a gray-scale representation based on the HSI model better exploits information for IIF image analysis. In [39], an automated HEp-2 cell staining pattern recognition method has been proposed by using morphological features. To capture local textural information, a modified version of uniform LBPs descriptor is incorporated in this research work. Wiliem et al. [45] have proposed a system for cell classification, comprising of nearest convex hull classifier and a dual-region codebook-based descriptor. In [5], an HEp-2 cell classification approach has been reported based on subclass discriminant analysis. In order to encode gradient and textural characteristics of the depicted HEp-2 patterns, Theodorakopoulos et al. [40] have proposed a new descriptor, based on co-occurrence of uniform LBPs along directions dictated by the orientation of local gradient. Nosaka et al. [27] have developed a method, which integrates the advantages of both support vector machine (SVM) and rotation-invariant co-occurrence among adjacent LBPs image features. In [32], an automatic HEp-2 cell classification approach has been introduced by combining multiresolution co-occurrence texture and large regional shape information.

The integration of multiple scales of same local textural features may provide better recognition of HEp-2 patterns. Due to the drastic variation of different scales and noisy nature of input HEp-2 cell images, the naive integration usually gives poor performance, which reflects in the insufficient and inaccurate staining pattern representation of these images. Multiple scales of unique cell image, on the other hand, may contain complementary information. The linkages between attributes of each HEp-2 cell images can be made by using these multiple scales of unique sample cell. The combination of multiple scales of a unique HEp-2 cell image would have more discriminatory and complete information of the inherent properties of that cell by generating improved system performance than

individual scale. Hence, a proper integration method is needed to incorporate the information of local textural descriptors obtained at multiple scales. In this background, canonical correlation analysis (CCA) [18] or its several variants [7, 14, 42] provide an effective way of capturing the correlation among different multidimensional data sets. Recently, some new supervised CCA have been proposed [24, 25], which perform significantly better than the existing CCA-based approaches. However, these methods are known to have high computational complexity, which renders their application in many real-life data analysis such as classification of HEP-2 cell patterns.

In this regard, the paper presents a new method, termed as CanSuR, for automatic recognition of antinuclear autoantibodies by HEP-2 cell IIF image analysis. It judiciously integrates the advantages of canonical correlation analysis, support vector machine and rough hypercuboid approach. The proposed staining pattern recognition method has two main steps, namely feature extraction and classification. The feature extraction step consists of two levels. At first, the local textural features are extracted from HEP-2 cell images, and then combined using a novel sequential canonical correlation analysis (CCA). To integrate two multidimensional variables, consisting of local features at different scales, the proposed supervised CCA extracts features sequentially by maximizing the correlation between canonical variables, significance among the features and their individual relevance with respect to HEP-2 cell images. The theory of rough hypercuboid approach is used for computing both significance and relevance measures. Finally, the SVM with radial basis function kernel is used to recognize one of the known staining patterns present in IIF images. In this context, it should be mentioned that both proposed CCA and supervised CCA, introduced in [25] and termed as FaRoC, can extract features sequentially from two multidimensional variables. But, the proposed method explores large search space with significantly lesser amount of time than FaRoC. The effectiveness of the proposed staining pattern recognition method, along with a comparison with related approaches, is demonstrated on MIVIA, SNP and ICPR HEP-2 cell image databases.

## 2 A novel sequential supervised CCA

A novel feature extraction algorithm for multimodal data sets, termed as sequential CCA, is introduced in this section. It extracts maximally correlated and most significant as well as relevant latent features sequentially from two multidimensional variables  $\mathcal{X} \in \mathbb{R}^{\mu \times n}$  and  $\mathcal{Y} \in \mathbb{R}^{q \times n}$ . If the number of features  $\mu$  and  $q$  of  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, is

larger than the number of samples  $n$ , that is,  $n < (\mu, q)$ , then the covariance matrices  $\mathcal{C}_{xx}$  and  $\mathcal{C}_{yy}$  become ill-conditioned. In the current work, a fast and supervised sequential CCA is proposed to address the singularity issue of covariance matrices while integrating the information of two multidimensional data sets. The proposed CCA obtains two directional basis vectors  $w_x \in \mathbb{R}^\mu$  and  $w_y \in \mathbb{R}^q$  such that the correlation between canonical variables  $\mathcal{U} = w_x^T \mathcal{X}$  and  $\mathcal{V} = w_y^T \mathcal{Y}$  is maximum, and the extracted feature  $\mathcal{A} = \mathcal{U} + \mathcal{V}$  is most relevant and significant. The correlation coefficient between the canonical variables is given as follows:

$$\begin{aligned} \mathcal{J}(w_x, w_y) &= \max_{w_x, w_y} \mathcal{U}\mathcal{V}^T = \max_{w_x, w_y} w_x^T \mathcal{C}_{xy} w_y \\ &\text{subject to, } \mathcal{U}\mathcal{U}^T = w_x^T \mathcal{C}_{xx} w_x = 1; \\ &\text{and } \mathcal{V}\mathcal{V}^T = w_y^T \mathcal{C}_{yy} w_y = 1. \end{aligned} \tag{1}$$

To incorporate the above two constraints into correlation coefficient-based objective function, the Lagrange multiplier is used in (1), which leads to

$$\begin{aligned} \mathcal{L}(\lambda_1, \lambda_2) &= w_x^T \mathcal{C}_{xy} w_y - \frac{\lambda_1}{2} (w_x^T \mathcal{C}_{xx} w_x - 1) \\ &\quad - \frac{\lambda_2}{2} (w_y^T \mathcal{C}_{yy} w_y - 1); \end{aligned} \tag{2}$$

where  $\lambda_1$  and  $\lambda_2$  are Lagrange multipliers. Differentiating  $\mathcal{L}$  with respect to  $w_x$  and  $w_y$  and setting the vectors of derivative to zero, we obtain

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta w_x} &= \mathcal{C}_{xy} w_y - \frac{\lambda_1}{2} (\mathcal{C}_{xx} w_x + \mathcal{C}_{xx}^T w_x) = 0 \\ &\Rightarrow \mathcal{C}_{xy} w_y = \lambda_1 \mathcal{C}_{xx} w_x; \\ \text{and } \frac{\delta \mathcal{L}}{\delta w_y} &= \mathcal{C}_{xy}^T w_x - \frac{\lambda_2}{2} (\mathcal{C}_{yy} w_y + \mathcal{C}_{yy}^T w_y) = 0 \\ &\Rightarrow \mathcal{C}_{yx} w_x = \lambda_2 \mathcal{C}_{yy} w_y. \end{aligned} \tag{3}$$

Multiplying (3) and (4) by  $w_x^T$  and  $w_y^T$ , respectively, we obtain

$$\begin{aligned} w_x^T \mathcal{C}_{xy} w_y &= \lambda_1 w_x^T \mathcal{C}_{xx} w_x = \lambda_1; \\ \text{and } w_y^T \mathcal{C}_{yx} w_x &= \lambda_2 w_y^T \mathcal{C}_{yy} w_y = \lambda_2. \end{aligned} \tag{4}$$

The correspondence between the Lagrange multipliers  $\lambda_1$  and  $\lambda_2$  can be established using (5) and (6),

$$\lambda_1 = (w_x^T \mathcal{C}_{xy} w_y)^T = w_y^T \mathcal{C}_{yx} w_x = \lambda_2. \tag{5}$$

From (7), it can be seen that the correlation coefficient between two canonical variables  $\mathcal{U} = w_x^T \mathcal{X}$  and  $\mathcal{V} =$

$\omega_y^T \mathcal{Y}$  is same with the Lagrange multipliers  $\lambda_1 = \lambda_2 = \lambda$ . Hence, (3) and (4) can be rewritten as

$$\mathcal{C}_{xy}\omega_y = \lambda\mathcal{C}_{xx}\omega_x \Rightarrow \omega_y = \lambda\mathcal{C}_{xy}^{-1}\mathcal{C}_{xx}\omega_x; \tag{8}$$

$$\text{and } \mathcal{C}_{yx}\omega_x = \lambda\mathcal{C}_{yy}\omega_y \Rightarrow \omega_x = \lambda\mathcal{C}_{yx}^{-1}\mathcal{C}_{yy}\omega_y. \tag{9}$$

Using (4) and (8), we obtain

$$\begin{aligned} \mathcal{C}_{yx}\omega_x &= \lambda\mathcal{C}_{yy}\omega_y = \lambda^2\mathcal{C}_{yy}\mathcal{C}_{xy}^{-1}\mathcal{C}_{xx}\omega_x \\ &\Rightarrow \mathcal{C}_{xx}^{-1}\mathcal{C}_{xy}\mathcal{C}_{yy}^{-1}\mathcal{C}_{yx}\omega_x = \lambda^2\omega_x. \end{aligned} \tag{10}$$

Similarly, using (3) and (9), we obtain

$$\begin{aligned} \mathcal{C}_{xy}\omega_y &= \lambda\mathcal{C}_{xx}\omega_x = \lambda^2\mathcal{C}_{xx}\mathcal{C}_{yx}^{-1}\mathcal{C}_{yy}\omega_y \\ &\Rightarrow \mathcal{C}_{yy}^{-1}\mathcal{C}_{yx}\mathcal{C}_{xx}^{-1}\mathcal{C}_{xy}\omega_y = \lambda^2\omega_y. \end{aligned} \tag{11}$$

Hence, (10) and (11) lead to the following generalized eigenvalue problem:

$$\begin{aligned} &\begin{bmatrix} \mathcal{C}_{xx}^{-1}\mathcal{C}_{xy}\mathcal{C}_{yy}^{-1}\mathcal{C}_{yx} & 0 \\ 0 & \mathcal{C}_{yy}^{-1}\mathcal{C}_{yx}\mathcal{C}_{xx}^{-1}\mathcal{C}_{xy} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} \\ &= \lambda^2 \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix}; \tag{12} \\ &\Rightarrow \begin{bmatrix} \mathcal{H} & 0 \\ 0 & \tilde{\mathcal{H}} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} = \lambda^2 \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} \text{where } \mathcal{H} &= \mathcal{C}_{xx}^{-1}\mathcal{C}_{xy}\mathcal{C}_{yy}^{-1}\mathcal{C}_{yx} \\ \text{and } \tilde{\mathcal{H}} &= \mathcal{C}_{yy}^{-1}\mathcal{C}_{yx}\mathcal{C}_{xx}^{-1}\mathcal{C}_{xy}. \end{aligned} \tag{13}$$

### 2.1 Proposed CCA

In real-life high-dimensional multimodal data analysis, the number of features is larger than the number of samples, which makes the covariance matrices  $\mathcal{C}_{xx}$  and  $\mathcal{C}_{yy}$  non-invertible. Regularized CCA (RCCA) [42] adds regularization parameters  $r_x$  and  $r_y$  to the diagonal elements of the covariance matrices  $\mathcal{C}_{xx}$  and  $\mathcal{C}_{yy}$  to make them invertible. On the other hand, in fast RCCA (FRCCA) [7], the values of off-diagonal elements of the matrices  $\mathcal{C}_{xx}$  and  $\mathcal{C}_{yy}$  are reduced by using shrinkage parameters  $s_x$  and  $s_y$  to deal with the singularity issue of these matrices. Both regularization parameters ( $r_x$  and  $r_y$ ) and shrinkage parameters ( $s_x$  and  $s_y$ ) are used to correct the noise present in  $\mathcal{X}$  and  $\mathcal{Y}$ . In the proposed method, both regularization and shrinkage are done simultaneously to take care of the singularity problem of  $\mathcal{C}_{xx}$  and  $\mathcal{C}_{yy}$ . Here,  $r_x$  and  $r_y$  are varied within a range  $[r_{min}, r_{max}]$ , where the common differences are  $d_x$  and  $d_y$  for  $r_x$  and  $r_y$ , respectively, and  $r_{min} \leq r_x, r_y \leq r_{max}$ . On the other hand,  $s_x$  and  $s_y$  can be computed as

$$s_x = \frac{\sum_{i \neq j} \hat{\mathcal{V}}([\mathcal{C}_{xx}]_{ij})}{\sum_{i \neq j} [\mathcal{C}_{xx}]_{ij}^2}; \quad \text{and} \quad s_y = \frac{\sum_{i \neq j} \hat{\mathcal{V}}([\mathcal{C}_{yy}]_{ij})}{\sum_{i \neq j} [\mathcal{C}_{yy}]_{ij}^2}; \tag{14}$$

where  $\hat{\mathcal{V}}([\mathcal{C}_{xx}]_{ij})$  and  $\hat{\mathcal{V}}([\mathcal{C}_{yy}]_{ij})$  are the unbiased empirical variance of  $[\mathcal{C}_{xx}]_{ij}$  and  $[\mathcal{C}_{yy}]_{ij}$ , respectively. Hence, to deal with the singularity issue, the covariance matrices  $\tilde{\mathcal{C}}_{xx}$  and  $\tilde{\mathcal{C}}_{yy}$  can be redefined as follows:

$$\begin{aligned} \tilde{\mathcal{C}}_{xx} &= \hat{\mathcal{C}}_{xx} + (r_x + (k - 1)d_x)I \\ \text{and } \tilde{\mathcal{C}}_{yy} &= \hat{\mathcal{C}}_{yy} + (r_y + (\ell - 1)d_y)I \end{aligned} \tag{15}$$

$$\text{where } [\hat{\mathcal{C}}_{xx}]_{ij} = \begin{cases} (1 - s_x)[\mathcal{C}_{xx}]_{ij}; & \text{where } i \neq j \\ [\mathcal{C}_{xx}]_{ij}; & \text{where } i = j \end{cases} \tag{16}$$

$$\text{and } [\hat{\mathcal{C}}_{yy}]_{ij} = \begin{cases} (1 - s_y)[\mathcal{C}_{yy}]_{ij}; & \text{where } i \neq j \\ [\mathcal{C}_{yy}]_{ij}; & \text{where } i = j \end{cases} \tag{17}$$

where  $\forall k \in \{1, 2, \dots, t_x\}$  and  $\forall \ell \in \{1, 2, \dots, t_y\}$ . The parameters  $t_x$  and  $t_y$  denote the number of possible values of  $r_x$  and  $r_y$ , respectively. As  $r_x$  and  $r_y$  are varied within a range and followed arithmetic progression, there exists a relation between the eigenvalues of covariance matrices corresponding to the first and the  $\ell$ -th regularization parameters for both multidimensional variables  $\mathcal{X}$  and  $\mathcal{Y}$  as follows [24]:

$$[\hat{\mathcal{C}}_{xx} + (r_x + (k - 1)d_x)I] = \Psi_x[A_x + (k - 1)d_x I]\Psi_x^T; \tag{18}$$

$$[\hat{\mathcal{C}}_{yy} + (r_y + (\ell - 1)d_y)I] = \Psi_y[A_y + (\ell - 1)d_y I]\Psi_y^T. \tag{19}$$

Here,  $A_x$  and  $A_y$  are the diagonal matrices, where diagonal elements are the eigenvalues of  $[\hat{\mathcal{C}}_{xx} + r_x I]$  and  $[\hat{\mathcal{C}}_{yy} + r_y I]$ , respectively, and the corresponding orthonormal eigenvectors are in the columns of  $\Psi_x$  and  $\Psi_y$ . The theoretical analysis, reported in [25], gives assistance to compute the matrices  $\mathcal{H}$  and  $\tilde{\mathcal{H}}$  with  $(k, \ell)$ -th regularization parameters of  $r_x$  and  $r_y$ , that is,  $\mathcal{H}_{kl}$  and  $\tilde{\mathcal{H}}_{kl}$ , as follows:

$$\begin{aligned} \mathcal{H}_{kl} &= \Psi_x[A_x + (k - 1)d_x I]^{-1}\Psi_x^T\mathcal{C}_{xy} \\ &\Psi_y[A_y + (\ell - 1)d_y I]^{-1}\Psi_y^T\mathcal{C}_{yx}; \end{aligned} \tag{20}$$

$$\text{and } \tilde{\mathcal{H}}_{kl} = \Psi_y[A_y + (\ell - 1)d_y I]^{-1}\Psi_y^T\mathcal{C}_{yx} \Psi_x[A_x + (k - 1)d_x I]^{-1}\Psi_x^T\mathcal{C}_{xy}. \tag{21}$$

The inverse of the diagonal matrices  $[A_x + (k - 1)d_x I]$  and  $[A_y + (\ell - 1)d_y I]$  can be computed as

$$\begin{aligned}
 & [A_x + (\kappa - 1)d_x I]^{-1} \\
 &= A_x^{-1} - A_x^{-1}(\kappa - 1)d_x I A_x^{-1} (I + (\kappa - 1)d_x I A_x^{-1})^{-1} \\
 &= A_x^{-1} - A_x^{-1}(\kappa - 1)d_x A_x^{-1} (I + (\kappa - 1)d_x A_x^{-1})^{-1} \\
 &= A_x^{-1} - A_x^{-1} \nabla_\kappa;
 \end{aligned} \tag{22}$$

$$\begin{aligned}
 & [A_y + (\ell - 1)d_y I]^{-1} \\
 &= A_y^{-1} - A_y^{-1}(\ell - 1)d_y I A_y^{-1} (I + (\ell - 1)d_y I A_y^{-1})^{-1} \\
 &= A_y^{-1} - A_y^{-1}(\ell - 1)d_y A_y^{-1} (I + (\ell - 1)d_y A_y^{-1})^{-1} \\
 &= A_y^{-1} - A_y^{-1} \nabla_\ell;
 \end{aligned} \tag{23}$$

where  $\nabla_\kappa = (\kappa - 1)d_x A_x^{-1} (I + (\kappa - 1)d_x A_x^{-1})^{-1}$ ; (24)

and  $\nabla_\ell = (\ell - 1)d_y A_y^{-1} (I + (\ell - 1)d_y A_y^{-1})^{-1}$ . (25)

Hence, using (22) and (23), the matrix  $\mathcal{H}$  of (20) becomes

$$\begin{aligned}
 \mathcal{H}_{kl} &= \Psi_x [A_x^{-1} - A_x^{-1} \nabla_\kappa] \Psi_x^T \mathcal{C}_{xy} \Psi_y [A_y^{-1} - A_y^{-1} \nabla_\ell] \Psi_y^T \mathcal{C}_{yx} \\
 &= \Psi_x A_x^{-1} \Psi_x^T \mathcal{C}_{xy} \Psi_y A_y^{-1} \Psi_y^T \mathcal{C}_{yx} + \mathcal{L}_{kl} = \mathcal{H}_{11} + \mathcal{L}_{kl};
 \end{aligned} \tag{26}$$

where  $\mathcal{L}_{kl} = \Psi_x A_x^{-1} [\nabla_\kappa \Psi_x^T \mathcal{C}_{xy} \Psi_y A_y^{-1} \nabla_\ell - \nabla_\kappa \Psi_x^T \mathcal{C}_{xy} \Psi_y A_y^{-1} - \Psi_x^T \mathcal{C}_{xy} \Psi_y A_y^{-1} \nabla_\ell] \Psi_y^T \mathcal{C}_{yx}$ . (27)

Similarly, using (22) and (23), the matrix  $\tilde{\mathcal{H}}$  of (21) becomes

$$\begin{aligned}
 \tilde{\mathcal{H}}_{kl} &= \Psi_y [A_y^{-1} - A_y^{-1} \nabla_\ell] \Psi_y^T \mathcal{C}_{yx} \Psi_x [A_x^{-1} - A_x^{-1} \nabla_\kappa] \Psi_x^T \mathcal{C}_{xy} \\
 &= \Psi_y A_y^{-1} \Psi_y^T \mathcal{C}_{yx} \Psi_x A_x^{-1} \Psi_x^T \mathcal{C}_{xy} + \tilde{\mathcal{L}}_{kl} = \tilde{\mathcal{H}}_{11} + \tilde{\mathcal{L}}_{kl};
 \end{aligned} \tag{28}$$

where  $\tilde{\mathcal{L}}_{kl} = \Psi_y A_y^{-1} [\nabla_\ell \Psi_y^T \mathcal{C}_{yx} \Psi_x A_x^{-1} \nabla_\kappa - \nabla_\ell \Psi_y^T \mathcal{C}_{yx} \Psi_x A_x^{-1} - \Psi_y^T \mathcal{C}_{yx} \Psi_x A_x^{-1} \nabla_\kappa] \Psi_x^T \mathcal{C}_{xy}$ . (29)

The nonzero eigenvalues of  $\mathcal{H}_{kl}$  and  $\tilde{\mathcal{H}}_{kl}$  are same [13], and the  $t$ -th eigenvector  $\omega_{yt}$  of  $\tilde{\mathcal{H}}_{kl}$  can be obtained from the  $t$ -th eigenvector  $\omega_{xt}$  of  $\mathcal{H}_{kl}$  as  $\omega_{yt} = \tilde{\mathcal{C}}_{yy}^{-1} \mathcal{C}_{yx} \omega_{xt}$ . So, one of the matrices  $\mathcal{H}_{kl}$  and  $\tilde{\mathcal{H}}_{kl}$  is sufficient to compute the eigenvectors of both of them. Hence, either  $\mathcal{H}_{kl}$  or  $\tilde{\mathcal{H}}_{kl}$  will be computed by using (26) or (28), respectively, depending on whether  $\mu \leq q$  or  $\mu > q$ , where  $\forall \kappa \in \{1, 2, \dots, t_x\}$  and  $\forall \ell \in \{1, 2, \dots, t_y\}$ .

In practical analysis,  $\mathcal{K} = \min(\mu, q)$  is large enough, and a small subset of extracted features is sufficient to perform a certain task. Hence, the objective of the proposed work is to find out a reduced subset of most significant and relevant extracted features. The sequential generation of

each eigenvector of  $\mathcal{H}_{kl}$  and  $\tilde{\mathcal{H}}_{kl}$  would help to evaluate the quality of each extracted feature individually. The theoretical analysis reported next, based on deflation method [43], gives assistance to compute  $(t + 1)$ -th basis vectors, which are the eigenvectors of matrices  $\mathcal{H}_{kl}(t + 1)$  and  $\tilde{\mathcal{H}}_{kl}(t + 1)$ . Thus, the matrices  $\mathcal{H}_{kl}(t + 1)$  and  $\tilde{\mathcal{H}}_{kl}(t + 1)$  can be computed as follows:

$$\begin{aligned}
 \mathcal{H}_{kl}(t + 1) &= \mathcal{H}_{kl}(1) - \sum_{m=1}^t \rho_{mkl} \omega_{xmk} \omega_{xmk}^T \\
 &= \mathcal{H}_{11} + \mathcal{L}_{kl} - \sum_{m=1}^t \rho_{mkl} \omega_{xmk} \omega_{xmk}^T;
 \end{aligned} \tag{30}$$

and  $\tilde{\mathcal{H}}_{kl}(t + 1) = \tilde{\mathcal{H}}_{kl}(1) - \sum_{m=1}^t \rho_{mkl} \omega_{ymk} \omega_{ymk}^T = \tilde{\mathcal{H}}_{11} + \tilde{\mathcal{L}}_{kl} - \sum_{m=1}^t \rho_{mkl} \omega_{ymk} \omega_{ymk}^T$ ; (31)

where  $\rho_{tkl}$  denotes eigenvalue of  $\mathcal{H}_{kl}(t + 1)$  or  $\tilde{\mathcal{H}}_{kl}(t + 1)$ ,  $\forall t \in \{1, 2, \dots, \mathcal{K} - 1\}$ ,  $\mathcal{K} = \min(\mu, q)$ ,  $\forall \kappa \in \{1, 2, \dots, t_x\}$ , and  $\forall \ell \in \{1, 2, \dots, t_y\}$ . Finally, the most correlated  $\mathcal{D}$  ( $<< \mathcal{K}$ ) features  $\mathcal{A}_{tkl} = \omega_{xtkl}^T \mathcal{X} + \omega_{ytkl}^T \mathcal{Y}$  are extracted from two multidimensional variables  $\mathcal{X}$  and  $\mathcal{Y}$ .

The objective of this work is to extract not only the correlated features but also the most relevant and significant features. Let us assume that the set  $\mathbb{C}$  contains all the  $t$ -th correlated features which are extracted using  $(\kappa, \ell)$ -th regularization parameters of  $r_x$  and  $r_y$ . If  $t = 1$ , the most relevant feature from the set  $\mathbb{C}$  is selected and is included into another set  $\mathbb{S}$ , initially which is empty, that is,  $\mathbb{S} \leftarrow \emptyset$ . If  $t > 1$ , the feature which has maximum relevance (among the features of  $\mathbb{C}$ ) and significance (with respect to the features of  $\mathbb{S}$ ) is selected from the set  $\mathbb{C}$  and is included to  $\mathbb{S}$ . Following objective function is used to select the most relevant and significant feature from the set  $\mathbb{C}$  when  $t > 1$ :

$$\gamma_{\mathcal{A}_{tkl}}(\mathbb{D}) + \frac{1}{t - 1} \sum_{\mathcal{A}_i \in \mathbb{S}} \sigma_{\{\mathcal{A}_{tkl}, \mathcal{A}_i\}}(\mathbb{D}, \mathcal{A}_{tkl}). \tag{32}$$

To compute both significance and relevance of an extracted feature, rough hypercuboid-based hypercuboid equivalence partition matrix [23] is used in the current work. The relevance  $\gamma_{\mathcal{A}_{tkl}}(\mathbb{D})$  of a feature  $\mathcal{A}_{tkl}$  with respect to the class label or decision attribute  $\mathbb{D}$  is determined as follows [23]:

$$\gamma_{\mathcal{A}_{tkl}}(\mathbb{D}) = 1 - \frac{1}{n} \sum_{j=1}^n v_j(\mathcal{A}_{tkl}); \tag{33}$$

$$\gamma_{\mathcal{A}_{tkl}}(\mathbb{D}) = \begin{cases} 1 & \text{if } \mathcal{A}_{tkl} \text{ totally depends on } \mathbb{D}; \\ 0 & \text{if } \mathcal{A}_{tkl} \text{ does not depend on } \mathbb{D}; \end{cases} \quad (34)$$

and  $\gamma_{\mathcal{A}_{tkl}}(\mathbb{D}) \in (0, 1)$  if  $\mathcal{A}_{tkl}$  partially depends on  $\mathbb{D}$ . The confusion vector for the feature  $\mathcal{A}_{tkl}$  is denoted as

$$\mathbb{V}(\mathcal{A}_{tkl}) = [v_1(\mathcal{A}_{tkl}), \dots, v_j(\mathcal{A}_{tkl}), \dots, v_n(\mathcal{A}_{tkl})] \quad (35)$$

where  $v_j(\mathcal{A}_{tkl}) = \min\{1, \sum_{i=1}^c h_{ij}(\mathcal{A}_{tkl}) - 1\}$ , (36)

the matrix  $\mathbb{H}(\mathcal{A}_{tkl}) = [h_{ij}(\mathcal{A}_{tkl})]_{c \times n}$  is termed as hypercuboid equivalence partition matrix of the feature  $\mathcal{A}_{tkl}$  [23], where

$$h_{ij}(\mathcal{A}_{tkl}) = \begin{cases} 1 & \text{if } \mathcal{L}_i \leq \mathcal{O}_j(\mathcal{A}_{tkl}) \leq \mathcal{U}_i \\ 0 & \text{otherwise} \end{cases} \quad (37)$$

represents the membership of sample  $\mathcal{O}_j$  in the  $i$ -th class  $\beta_i$ , and  $c$  denotes the number of classes. According to the decision attribute  $\mathbb{D}$ , the interval of the  $i$ -th class  $\beta_i$  is denoted by  $[\mathcal{L}_i, \mathcal{U}_i]$ . Each row of the  $c \times n$  hypercuboid equivalence partition matrix  $\mathbb{H}(\mathcal{A}_{tkl})$  denotes a hypercuboid equivalence class. The joint relevance  $\gamma_{\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}}(\mathbb{D})$  of the feature set  $\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}$  can be used to compute the significance of the feature  $\mathcal{A}_{tkl}$  as follows:

$$\sigma_{\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}}(\mathbb{D}, \mathcal{A}_{tkl}) = \gamma_{\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}}(\mathbb{D}) - \gamma_{\mathcal{A}_{\bar{i}}}(\mathbb{D}). \quad (38)$$

The joint relevance  $\gamma_{\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}}(\mathbb{D})$  of feature set  $\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}$  is depended on  $c \times n$  hypercuboid equivalence partition matrix  $\mathbb{H}(\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\})$  corresponding to the set  $\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}$ . Thus, the matrices  $\mathbb{H}(\mathcal{A}_{tkl})$  and  $\mathbb{H}(\mathcal{A}_{\bar{i}})$  are used to compute the matrix  $\mathbb{H}(\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\})$ , where

$$\mathbb{H}(\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}) = \mathbb{H}(\mathcal{A}_{tkl}) \cap \mathbb{H}(\mathcal{A}_{\bar{i}}); \quad (39)$$

and  $h_{ij}(\{\mathcal{A}_{tkl}, \mathcal{A}_{\bar{i}}\}) = h_{ij}(\mathcal{A}_{tkl}) \times h_{ij}(\mathcal{A}_{\bar{i}})$ . (40)

### 2.2 Computational complexity

Let us assume that  $\mathcal{K} = \min(\mu, q)$  and  $\mathcal{M} = \max(\mu, q)$ , where the number of extracted features  $\mathcal{D} \ll \mathcal{K}$ . The total time complexity to compute two covariance matrices  $\mathcal{C}_{xx}$  and  $\mathcal{C}_{yy}$  is  $(\mathcal{O}(\mathcal{K}^2n + \mathcal{M}^2n) =) \mathcal{O}(\mathcal{M}^2n)$ , whereas the computational cost to calculate cross-covariance matrix  $\mathcal{C}_{xy}$  is  $\mathcal{O}(\mathcal{K} \mathcal{M} n)$ . The shrinkage parameters  $s_x$  and  $s_y$  can be computed with time complexity  $(\mathcal{O}(\mathcal{K}^2n + \mathcal{M}^2n) =) \mathcal{O}(\mathcal{M}^2n)$ . On the other hand, the eigenvalues  $\Lambda_x$  and  $\Lambda_y$ , along with corresponding eigenvectors  $\Psi_x$  and  $\Psi_y$ , can be calculated with computational complexity  $(\mathcal{O}(\mathcal{K}^3 + \mathcal{M}^3) =) \mathcal{O}(\mathcal{M}^3)$ . Hence, the total

time complexity to compute  $\mathcal{C}_{xx}^{-1}$  and  $\mathcal{C}_{yy}^{-1}$  is  $(\mathcal{O}(\mathcal{K}^3 + \mathcal{M}^3) =) \mathcal{O}(\mathcal{M}^3)$ .

Depending on  $\mu \leq q$  or  $\mu > q$ , one of the matrices  $\mathcal{H}_{11}$  or  $\tilde{\mathcal{H}}_{11}$  has to be computed; hence, computational cost to compute  $\mathcal{H}_{11}$  or  $\tilde{\mathcal{H}}_{11}$  is  $(\mathcal{O}(\mathcal{K}^3 + \mathcal{K}^2 \mathcal{M} + \mathcal{K} \mathcal{M}^2 + \mathcal{M}^3) =) \mathcal{O}(\mathcal{M}^3)$ . Similarly, the computational complexity of the matrix  $\nabla_k$  or  $\nabla_l$  is  $\mathcal{O}(\mathcal{K})$ . Hence, the time complexity to compute matrix  $\mathcal{L}_{kl}$  or  $\tilde{\mathcal{L}}_{kl}$  is  $(\mathcal{O}(\mathcal{K}^3 + \mathcal{K}^2 \mathcal{M} + \mathcal{K} \mathcal{M}^2) =) \mathcal{O}(\mathcal{K} \mathcal{M}^2)$ . Now, the time complexity to calculate the first basis vector of first multidimensional variable is  $\mathcal{O}(\mathcal{K}^2)$ , whereas the  $(t + 1)$ -th basis vector of first multidimensional variable can be computed with computational cost  $(\mathcal{O}(\mathcal{M}^3 + \mathcal{K} \mathcal{M}^2 + t \mathcal{K}^2) =) \mathcal{O}(\mathcal{M}^3)$ , where  $\forall t \in \{1, 2, \dots, \mathcal{K} - 1\}$ . On the other hand, each basis vector of the second multidimensional variable can be computed with time complexity  $(\mathcal{O}(\mathcal{K} \mathcal{M}^2 + \mathcal{K} \mathcal{M}) =) \mathcal{O}(\mathcal{K} \mathcal{M}^2)$ . The total time complexity for computing canonical variables is  $(\mathcal{O}(\mathcal{K} n + \mathcal{M} n) =) \mathcal{O}(\mathcal{M} n)$ . The computational complexity to extract a feature is  $\mathcal{O}(n)$ . The time complexity to compute both relevance and significance of a feature is same, which is  $\mathcal{O}(cn)$ . Hence, the total complexity to extract all highest relevant features among all  $(t_x \times t_y)$  combination of regularization parameters is  $(\mathcal{O}(t_x t_y (\mathcal{K}^2 + \mathcal{K} \mathcal{M}^2 + \mathcal{M} n + n + cn)) =) \mathcal{O}(t_x t_y \mathcal{K} \mathcal{M}^2)$ .

The complexity to extract all the  $(t + 1)$ -th features, by maximizing both relevance and significance, among all  $(t_x \times t_y)$  candidate features is  $(\mathcal{O}(t_x t_y (\mathcal{M}^3 + \mathcal{K} \mathcal{M}^2 + \mathcal{M} n + n + cn)) =) \mathcal{O}(t_x t_y \mathcal{M}^3)$ . The selection of a feature from  $(t_x \times t_y)$  candidate features by maximizing relevance and significance has complexity  $\mathcal{O}(t_x t_y)$ . So, total complexity to compute all  $\mathcal{D}$  maximally relevant and significant features is  $(\mathcal{O}(t_x t_y (\mathcal{K} \mathcal{M}^2 + \mathcal{M}^3 (\mathcal{D} - 1)) + t_x t_y \mathcal{D}) =) \mathcal{O}(t_x t_y \mathcal{M}^3 \mathcal{D})$ . Hence, the overall computational complexity of the proposed sequential supervised CCA algorithm is  $(\mathcal{O}(\mathcal{M}^2 n + \mathcal{K} \mathcal{M} n + \mathcal{M}^3 + \mathcal{K} + \mathcal{K} \mathcal{M}^2 + t_x t_y \mathcal{M}^3 \mathcal{D}))$ , that is  $\mathcal{O}(t_x t_y \mathcal{M}^3 \mathcal{D})$ .

### 3 CanSuR: a new method for HEP-2 cell classification

This section presents a new method, termed as CanSuR, for staining pattern recognition of HEP-2 cell images. The proposed method has two stages, namely, extraction of features from HEP-2 cell images and classification of cells based on the extracted features using support vector machine (SVM). The feature extraction step of the proposed method has two levels. In the first level, rotation-

invariant local binary patterns (LBP<sup>ri</sup>s) [30] are extracted directly from the input cell image. The sets of LBP<sup>ri</sup> feature vectors, corresponding to scales  $S_x$  and  $S_y$  of all HEp-2 cells, form two modalities  $\mathcal{X} \in \mathbb{R}^{\rho \times n}$  and  $\mathcal{Y} \in \mathbb{R}^{\rho \times n}$ , respectively. Here, each row in  $\mathcal{X}$  and  $\mathcal{Y}$  represents one of the LBP<sup>ri</sup> feature vectors, and each column represents one of the  $n$  samples. These extracted feature sets  $\mathcal{X}$  and  $\mathcal{Y}$  are integrated in the second level using the proposed sequential CCA method. The proposed sequential CCA extracts maximally correlated, most relevant and significant features from two multidimensional data sets  $\mathcal{X}$  and  $\mathcal{Y}$ . In the proposed HEp-2 cell staining pattern recognition method, henceforth termed as CanSuR, a set of maximally correlated and most significant as well as relevant latent local features is first extracted for each set of training data, and then, SVM is trained using this feature set. After the training, the information of training feature set is used to generate the test set and then the class label of the test sample is predicted using the SVM. The basic steps of the proposed methodology are outlined next:

1. Compute LBP<sup>ri</sup>s using scale  $S_x$  and scale  $S_y$  for each HEp-2 cell image to record the local textural features and generate two data sets  $\mathcal{X}$  and  $\mathcal{Y}$ .
2. Apply the proposed sequential CCA on  $\mathcal{X}$  and  $\mathcal{Y}$  to extract maximally correlated, most significant and relevant latent local features.
3. Recognize the staining pattern of HEp-2 cell images using SVM based on extracted features set.

A graphical representation of the proposed HEp-2 cell classification methodology is shown in Fig. 1. The detail description of each step is presented next.

### 3.1 Step I: generation of local textural features

To generate the local textural features of each HEp-2 cell images, gray scale and rotation-invariant local binary texture operator LBP<sup>ri</sup> is used. For a monochrome texture image, the local binary texture  $\mathcal{T}$  can be defined as [30]

$$\mathcal{T} = t(g_c, g_0, \dots, g_{\mathcal{L}-1}); \tag{41}$$

where  $g_c$  and  $g_\ell, \forall \ell \in \{0, 1, 2, \dots, \mathcal{L} - 1\}$  denote the gray values of the center and neighbor pixels of the circularly symmetric neighborhood with radius  $\mathcal{R}$ . To achieve gray-scale invariance, the gray value of the center pixel  $g_c$  has to be subtracted from the gray values of neighbor pixels  $g_\ell (\ell = 0, \dots, \mathcal{L} - 1)$ , so

$$\mathcal{T} = t(g_c, g_0 - g_c, g_1 - g_c, \dots, g_{\mathcal{L}-1} - g_c). \tag{42}$$

It is assumed that the difference  $(g_\ell - g_c)$  does no depend on  $g_c$ . Hence, (42) can be factorized as follows:

$$\mathcal{T} \approx t(g_c)t(g_0 - g_c, g_1 - g_c, \dots, g_{\mathcal{L}-1} - g_c). \tag{43}$$

The overall luminance of the image  $t(g_c)$  does not relate to local image texture and does not provide any useful information for texture analysis. Therefore, the original joint gray level distribution  $\mathcal{T}$  is dependent on joint difference distribution  $t(g_0 - g_c, g_1 - g_c, \dots, g_{\mathcal{L}-1} - g_c)$  [31],

$$\mathcal{T} \approx t(g_0 - g_c, g_1 - g_c, \dots, g_{\mathcal{L}-1} - g_c). \tag{44}$$

The occurrence of various patterns in the neighborhood of each pixel in a  $\mathcal{L}$ -dimensional histogram is recorded by this operator. In all directions, the differences are zero for constant regions and high for a spot. The highest difference is recorded in the gradient direction for a slowly sloped edge. On the other hand, any change in mean luminance does not affect the signed difference  $(g_\ell - g_c)$ . Hence, the joint difference distribution is invariant against gray-scale shifts. Thus, invariance with respect to the scaling of gray scale is achieved by considering the signs of the differences only,

$$\mathcal{T} \approx t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{\mathcal{L}-1} - g_c));$$

where  $s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases} \tag{45}$

For each sign  $s(g_\ell - g_c)$ , a binomial factor  $2^\ell$  is assigned to produce a unique number that characterizes the spatial structure of the local image texture. The gray values  $g_\ell$  will move along the perimeter of the circle around  $g_0$ , when the image is rotated, and produce different binary patterns for the same texture. To assign a unique identifier to each rotation-invariant local binary pattern, the following formula is used:

$$\text{LBP}^{\text{ri}} = \min \left\{ \mathcal{ROR} \left( \sum_{\ell=0}^{\mathcal{L}-1} s(g_\ell - g_c) 2^\ell, i \right) \mid i = 0, 1, \dots, \mathcal{L} - 1 \right\}; \tag{46}$$

where  $\mathcal{ROR}(x, i)$  performs a circular bit-wise right shift on the  $\mathcal{L}$ -bit number  $x$   $i$  times.

### 3.2 Step II: integration of local textural features

The sets of LBP<sup>ri</sup> feature vectors, corresponding to scales  $S_x$  and  $S_y$  of all HEp-2 cells, form two modalities  $\mathcal{X} \in \mathbb{R}^{\rho \times n}$  and  $\mathcal{Y} \in \mathbb{R}^{\rho \times n}$ , respectively. To integrate the information of these rotation-invariant local binary feature vector sets  $\mathcal{X}$  and  $\mathcal{Y}$ , the proposed supervised sequential CCA is used. Here, the number of LBP<sup>ri</sup> features of  $\mathcal{X}$  and  $\mathcal{Y}$  is same, that is,  $\rho = q$  and  $n$  is the number of HEp-2 cell images. The proposed data integration method extracts latent features those are maximally correlated and most significant as well as relevant from LBP<sup>ri</sup> feature vector

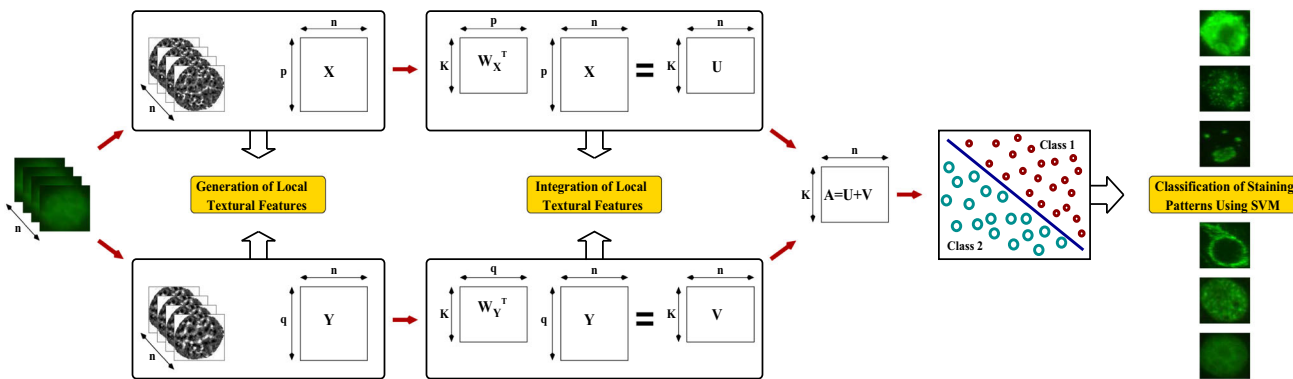


Fig. 1 Block diagram of the proposed method for HEP-2 cell classification

sets  $\mathcal{X}$  and  $\mathcal{Y}$ . It obtains two directional basis vectors  $\omega_x$  and  $\omega_y$  such that the correlation between canonical variables  $\mathcal{U} = \omega_x^T \mathcal{X}$  and  $\mathcal{V} = \omega_y^T \mathcal{Y}$  is maximum, and the extracted feature  $\mathcal{A} = \mathcal{U} + \mathcal{V}$  is most relevant and significant.

### 3.3 Step III: SVM for staining pattern classification

The support vector machine (SVM) [41] is used, in the present work, to classify the staining patterns of HEP-2 cell images by drawing an optimal hyperplane in the feature vector space. This hyperplane defines a boundary that maximizes the margin between data samples belonging to different classes. In the current study, the SVM uses radial basis function kernel to generate nonlinear decision boundary among different classes. The SVM is trained with the correlated, relevant and significant features generated for the training HEP-2 cell set and using the information of training feature set, and the staining pattern present in test HEP-2 cell is predicted.

## 4 Performance analysis and discussions

The performance of the proposed HEP-2 cell image classification method, termed as CanSuR, is presented in this section, along with a comparison with related approaches. The proposed method CanSuR consists of mainly three stages, namely, generation of local features using rotation-invariant LBP ( $LBP^{ri}$ ), integration of multiple scales of  $LBP^{ri}$  using the proposed sequential supervised CCA, and classification of HEP-2 cell images using SVM with radial basis function kernel ( $SVM_R$ ). So, the performance of  $LBP^{ri}$  is compared with that of several local texture descriptors, namely, LBP [29] and rotation-invariant uniform LBP ( $LBP^{riu2}$ ) [30]. Similarly, the performance of proposed sequential CCA is compared with that of several

multimodal data integration methods, namely, CCA [18], regularized CCA (RCCA) [42] and CuRSaR [24]. The comparative performance analysis between  $SVM_R$  and other classifiers, such as extreme learning machine (ELM) [20], SVM with linear kernel ( $SVM_L$ ) and SVM with polynomial kernel ( $SVM_P$ ), is also done in the current study.

The performance of different methods is studied with respect to various scales of LBP 8-neighborhood such as 1 ( $S_1$ ), 2 ( $S_2$ ) and 3 ( $S_3$ ). Both classification accuracy and F1 score on test sets are reported in this section to establish the performance of proposed method as well as existing approaches. To analyze the statistical significance of the proposed method, with respect to different existing methods, both tenfold cross-validation and training–testing are performed. For each data set, LBP,  $LBP^{ri}$  and  $LBP^{riu2}$  provide 256, 36 and 10 features, respectively. So, in case of naive integration between two scales of LBP,  $LBP^{ri}$  and  $LBP^{riu2}$ , the total number of features used for the analysis is twice the number of features of individual scale. On the other hand, fifteen top-ranked extracted features are considered for different data integration methods, namely, CCA, RCCA, CuRSaR and proposed CCA. In the proposed data integration method, the value of two shrinkage parameters  $\delta_x$  and  $\delta_y$  is considered as 0.0001.

### 4.1 Description of data sets

This section provides some basic information about the data sets used for benchmarking the activity. These data sets are (i) ICPR 2012 HEP-2 cell classification contest data, termed as MIVIA image database [10]; (ii) ICPR 2014 HEP-2 cell classification contest data, termed as ICPR image database; and (iii) SNP HEP-2 database [45].

- (i) *MIVIA database* This data set contains 1455 cells from 28 images among which four images are belonging to cytoplasmic, fine speckled and nucleolar patterns, five images are belonging to



coarse speckled and homogeneous, and six centromere images are there. The images have 24 bits color depth with uncompressed  $1388 \times 1038$  pixels resolution. The data set contains 721 and 734 cell images in training and test sets, respectively.

- (ii) *ICPR database* This data set contains 13596 cell images almost equally distributed with respect to six different patterns, namely, centromere, homogeneous, nucleolar, speckled, nuclear membrane and golgi. The set contains 6797 and 6799 cell images in training and test sets, respectively.
- (iii) *SNP database* This database has five pattern classes, namely, centromere, homogeneous, coarse speckled, fine speckled and nucleolar. From 40 specimen images, there are 1884 cell images extracted. The specimen images are divided into training and testing sets with 20 images each (4 images for each pattern). In total, there are 869 and 937 cell images extracted for training and testing.

The cell images from the above data sets were captured in different laboratory settings (e.g., different assays and microscope configurations). For instance, SNP HEP-2 used objective lens magnitude  $20\times$ , while MIVIA HEP-2 used  $40\times$ . The robustness of the proposed method as well as existing approaches is studied through the evaluation on these three data sets. Table 1 indicates the number of training and testing cells with respect to different HEP-2 patterns of these three data sets, which are used to evaluate the performance of the proposed algorithm as well as existing methods.

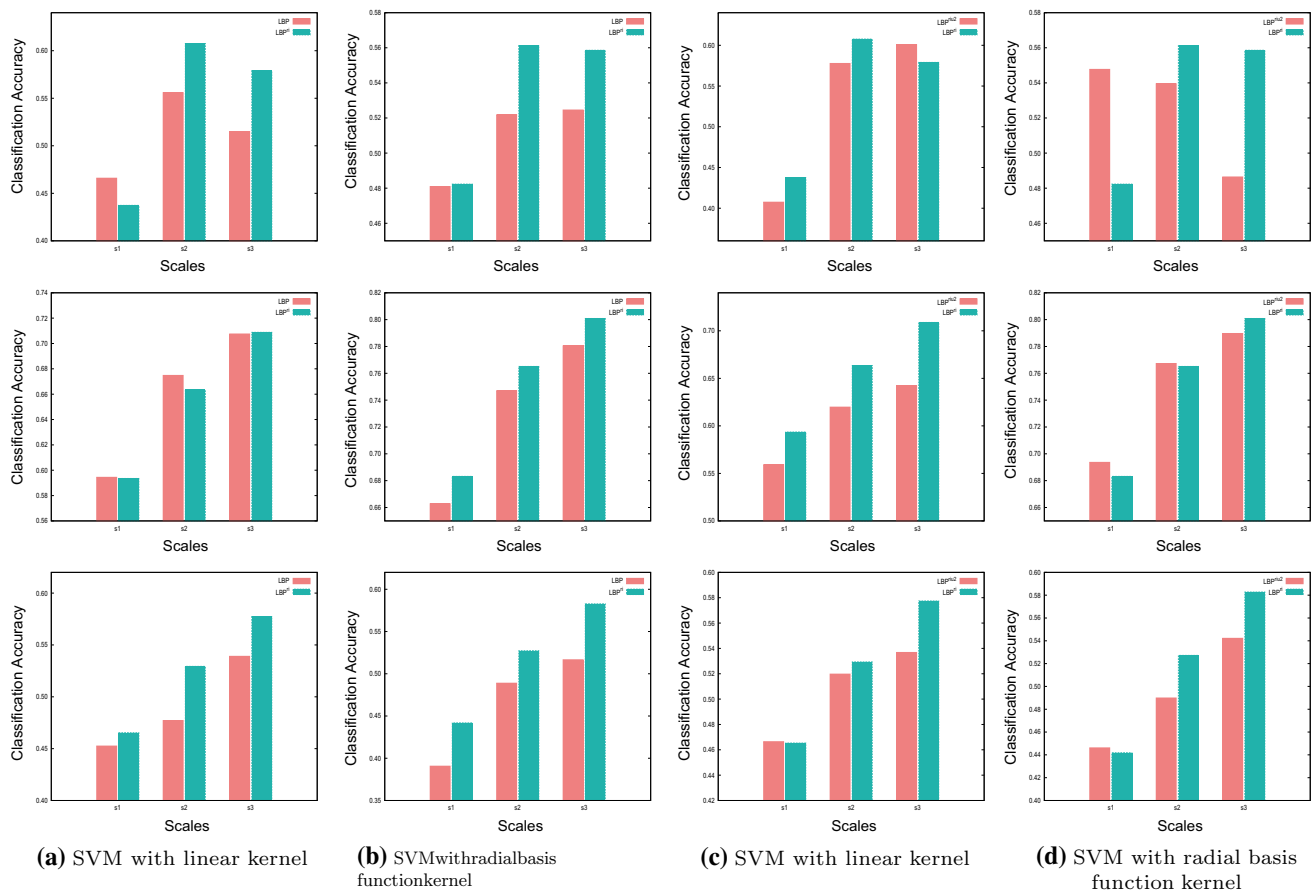
## 4.2 Importance of rotation-invariant LBP

Ideally, when an HEP-2 cell image is rotated, the pattern operator should produce a unique binary pattern for the same texture. In other words, the pattern operator should be rotation invariant. Figure 2a, b compares the performance of rotation-invariant LBP ( $LBP^{ri}$ ) with that of LBP, considering three scales, namely,  $S_1$ ,  $S_2$  and  $S_3$  using  $SVM_L$  and  $SVM_R$ , respectively. The results reported in Fig. 2a show that  $LBP^{ri}$  provides better classification accuracy of the  $SVM_L$  than LBP in most of the cases. For SNP data set, the  $LBP^{ri}$  attains classification accuracy of 0.465, 0.529 and 0.577, with respect to three scales, namely,  $S_1$ ,  $S_2$  and  $S_3$ , respectively, which are better than that obtained using LBP. On the other hand, for MIVIA data set,  $LBP^{ri}$  performs better than LBP for scales  $S_2$  and  $S_3$ , where it achieves classification accuracy of 0.608 and 0.579, respectively. However, for scale  $S_1$ , LBP performs better

**Table 1** Description of data sets used

Different data sets	Different staining patterns	Number of cells	
		Training	Test
MIVIA	Homogeneous	150	180
	Cytoplasmic	58	51
	Coarse speckled	109	101
	Fine speckled	94	114
	Centromere	208	149
	Nucleolar	102	139
ICPR	Homogeneous	1247	1247
	Speckled	1415	1416
	Nuclear membrane	1104	1104
	Nucleolar	1299	1299
	Centromere	1370	1371
	Golgi	362	362
SNP	Homogeneous	172	188
	Coarse speckled	166	187
	Fine speckled	188	191
	Nucleolar	194	188
	Centromere	149	183

than  $LBP^{ri}$ . The  $SVM_L$  classification accuracy obtained using LBP for scale  $S_1$  is 0.466, whereas the accuracy obtained using  $LBP^{ri}$  is 0.437. In case of ICPR data set,  $LBP^{ri}$  performs better than LBP only for scale  $S_3$ , while the performance is comparable in other two cases. The classification accuracy obtained using  $LBP^{ri}$  with respect to scales  $S_1$ ,  $S_2$  and  $S_3$  is 0.5935, 0.664 and 0.709, respectively. On the other hand, the  $SVM_L$  classification accuracy obtained using LBP with respect to scales  $S_1$ ,  $S_2$  and  $S_3$  is 0.5944, 0.675 and 0.708, respectively. The results reported in Fig. 2b show that  $LBP^{ri}$  achieves better classification accuracy using  $SVM_R$  than LBP in all the cases. For MIVIA data set, the  $LBP^{ri}$  attains classification accuracy of 0.482, 0.561 and 0.559 with respect to three scales, namely,  $S_1$ ,  $S_2$  and  $S_3$ , respectively, which are better than that obtained using LBP. In case of ICPR data set, the  $SVM_R$  classification accuracy obtained using  $LBP^{ri}$  with respect to scales  $S_1$ ,  $S_2$  and  $S_3$  is 0.683, 0.765 and 0.801, which are better than that obtained using LBP. On the other hand, for SNP data set,  $LBP^{ri}$  performs better than LBP for all the scales  $S_1$ ,  $S_2$  and  $S_3$ , where it achieves classification accuracy of 0.442, 0.527 and 0.583, respectively. The results reported in Fig. 2a, b demonstrate that  $LBP^{ri}$ , which is rotation-invariant LBP, performs better than LBP in 6 and 9 cases out of total 9 cases each using linear kernel and radial basis function kernel of SVM, respectively. All the results reported here thus establish the fact that  $LBP^{ri}$



**Fig. 2** Comparative performance analysis of LBP-LBP<sup>ri</sup> and LBP<sup>riu2</sup>-LBP<sup>ri</sup> (top: MIVIA; middle: ICPR; and bottom: SNP)

performs better than LBP, irrespective of the datasets, scales and SVM kernels used.

Similarly, Fig. 2c, d compares the performance of LBP<sup>ri</sup> with that of LBP<sup>riu2</sup> at different scales using linear kernel and radial basis function kernel, respectively, considering three HEP-2 cell data sets. At scale  $S_1$ , the LBP<sup>riu2</sup> provides SVM<sub>L</sub> classification accuracy of 0.407, 0.559 and 0.466, on MIVIA, ICPR and SNP data sets, respectively, while it attains SVM linear kernel accuracy of 0.578, 0.620 and 0.520, respectively, for scale  $S_2$ . On the other hand, LBP<sup>riu2</sup> achieves 0.601, 0.642 and 0.537 classification accuracy at scale  $S_3$  using linear kernel. Similarly, the SVM<sub>R</sub> classification accuracy obtained using LBP<sup>riu2</sup> with respect to scale  $S_1$  are 0.548, 0.694 and 0.446 on MIVIA, ICPR and SNP data sets, respectively. On the other hand, for scale  $S_2$ , LBP<sup>riu2</sup> achieves 0.540, 0.767 and 0.490 accuracy using SVM radial basis function kernel on MIVIA, ICPR and SNP data sets, respectively. For scale  $S_3$ , the SVM<sub>R</sub> classification accuracy obtained using LBP<sup>riu2</sup> with respect to MIVIA, ICPR and SNP data sets is 0.486, 0.790 and 0.542, respectively. All the results reported in Fig. 2c, d validate that the LBP<sup>ri</sup> performs better than LBP<sup>riu2</sup> in 7 cases and 5

cases out of total 9 cases each, using linear kernel and radial basis function kernel of SVM, respectively. Hence, these results can establish the fact that LBP<sup>ri</sup> performs better than LBP<sup>riu2</sup>, irrespective of the scales, datasets and SVM kernels used. Based on the analysis reported in Fig. 2, the LBP<sup>ri</sup> is considered as the textural feature extraction operator in the current research work, as it can extract rotation-invariant local binary patterns more accurately from HEP-2 cell images than both LBP and LBP<sup>riu2</sup>.

### 4.3 Optimum scale for proposed method

In order to find out the optimum value of scale parameter for the proposed method, extensive experimentation is performed on each data set and corresponding results are reported in Fig. 3. Figure 3 compares the performance of different scales of LBP<sup>ri</sup> operator using both linear kernel and radial basis function kernel of SVM. The results reported in Fig. 3 show that for both ICPR and SNP data, LBP<sup>ri</sup> provides highest SVM<sub>L</sub> and SVM<sub>R</sub> classification accuracy with respect to scale  $S_3$ . It achieves highest SVM<sub>L</sub> classification accuracy of 0.709 and 0.577 with respect to

scale  $S_3$  for ICPR and SNP data, respectively. Similarly, the highest SVM<sub>R</sub> classification accuracy with respect to scale  $S_3$  is 0.801 and 0.583 using ICPR and SNP data, respectively. On the other hand, the highest SVM<sub>L</sub> and SVM<sub>R</sub> classification accuracy obtained using LBP<sup>ri</sup> with respect to scale  $S_2$  is 0.608 and 0.561, respectively, for MIVIA data set. All the results reported here also confirm that LBP<sup>ri</sup> attains better classification accuracy with respect to scales  $S_2$  and  $S_3$  than scale  $S_1$  for both linear kernel and radial basis function kernel. Hence, integration of these two scales, namely,  $S_2$  and  $S_3$ , may provide better recognition of HEp-2 patterns. In this regard, the optimum scale for the proposed method is considered as  $S_2 + S_3$ .

Using the naive integration of scales  $S_2$  and  $S_3$ , that is  $S_2 + S_3$ , the SVM<sub>L</sub> achieves classification accuracy of 0.584, 0.770 and 0.545, for MIVIA, ICPR and SNP data sets, respectively, whereas the classification accuracy using radial basis function kernel is 0.582, 0.826 and 0.589, for MIVIA, ICPR and SNP data sets, respectively. The results reported in Table 2 establish the fact that both SVM<sub>L</sub> and SVM<sub>R</sub> classification accuracy on ICPR data set is increased from 0.709 to 0.770 and 0.801 to 0.826, respectively. But, for MIVIA and SNP data sets, the SVM<sub>L</sub> classification accuracy of  $S_2 + S_3$  lies in between that of scales  $S_2$  and  $S_3$ , although the SVM<sub>R</sub> classification accuracy of  $S_2 + S_3$  is increased from 0.561 to 0.582 and 0.583 to 0.589, for MIVIA and SNP data sets, respectively. Similar results can also be found in Table 2 for both LBP and LBP<sup>riu2</sup>. All these results indicate that naive integration, which is direct concatenation, of these two scales is not enough for integrating the information of two scales. Due to the drastic variation of two scales and noisy nature of the input HEp-2 cell images, naive integration usually gives poor performance, which causes insufficient and inaccurate staining pattern representation of the images. Moreover, multiple scales of unique cell image may contain complementary information. The linkages between attributes of each HEp-2 cell images can be made by using these multiple scales of unique sample cell. The combination of multiple scales of a unique HEp-2 cell image would have more discriminatory

and complete information of the inherent properties of that cell by generating improved system performance than individual scale. Hence, a proper integration method is needed to incorporate the local textural feature information obtained at multiple scales.

To integrate the information of two scales, namely,  $S_2$  and  $S_3$ , of LBP<sup>ri</sup> operator, the proposed sequential CCA is used in the current research work. Figure 4 demonstrates the results obtained using integrated scales of  $S_1 + S_2$ ,  $S_1 + S_3$  and  $S_2 + S_3$ , using both linear kernel and radial basis function kernel of SVM, where the proposed sequential CCA is used to integrate the information of two scales. All the results reported in Fig. 4 establish the fact that the integrated information of scales  $S_2$  and  $S_3$  performs significantly better than that of other integrated scales, irrespective of the number of extracted features, SVM kernels and data sets used. The results reported in Fig. 4 also show that the proposed data integration method provides highest SVM<sub>L</sub> classification accuracy of 0.507, 0.594 and 0.610, respectively, with respect to three aforementioned integrated scales on MIVIA data set. On the other hand, the highest SVM<sub>R</sub> classification accuracy of 0.564, 0.616 and 0.636 is obtained using integrated scales  $S_1 + S_2$ ,  $S_1 + S_3$  and  $S_2 + S_3$  on MIVIA data set. Similarly, the highest F1 score for both SVM<sub>L</sub> and SVM<sub>R</sub> is 0.510, 0.584, 0.613 and 0.562, 0.625, 0.651, respectively, for all the combined scales of  $S_1 + S_2$ ,  $S_1 + S_3$  and  $S_2 + S_3$ . For ICPR data set, the integrated scales of  $S_1 + S_2$ ,  $S_1 + S_3$  and  $S_2 + S_3$  obtain highest SVM<sub>L</sub> classification accuracy and F1 score of 0.539, 0.692, 0.742 and 0.525, 0.663, 0.708, respectively. Similarly, the highest SVM<sub>R</sub> classification accuracy and F1 score obtained using scales  $S_1 + S_2$ ,  $S_1 + S_3$  and  $S_2 + S_3$  are 0.702, 0.766, 0.851 and 0.681, 0.742, 0.825, respectively, on ICPR data set. On the other hand, the above-mentioned three integrated scales attain highest SVM<sub>L</sub> classification accuracy and F1 score of 0.486, 0.572, 0.593 and 0.482, 0.586, 0.592, respectively, for SNP data set. Likewise, the highest SVM<sub>R</sub> classification accuracy and F1 score achieved using scales  $S_1 + S_2$ ,  $S_1 + S_3$  and  $S_2 + S_3$  are 0.541, 0.557, 0.631 and 0.539, 0.579, 0.637, respectively, on SNP data set. However, Table 2 reports the classification accuracy and F1 score for fifteen top-ranked features, which may be different from the best values shown in Fig. 4.

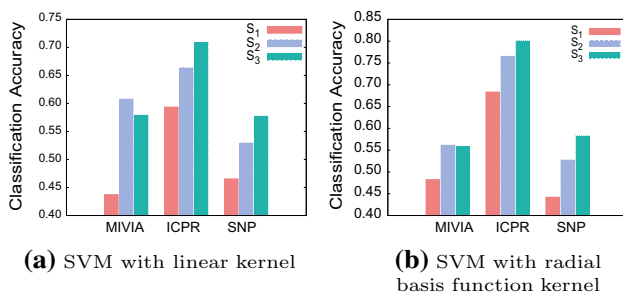


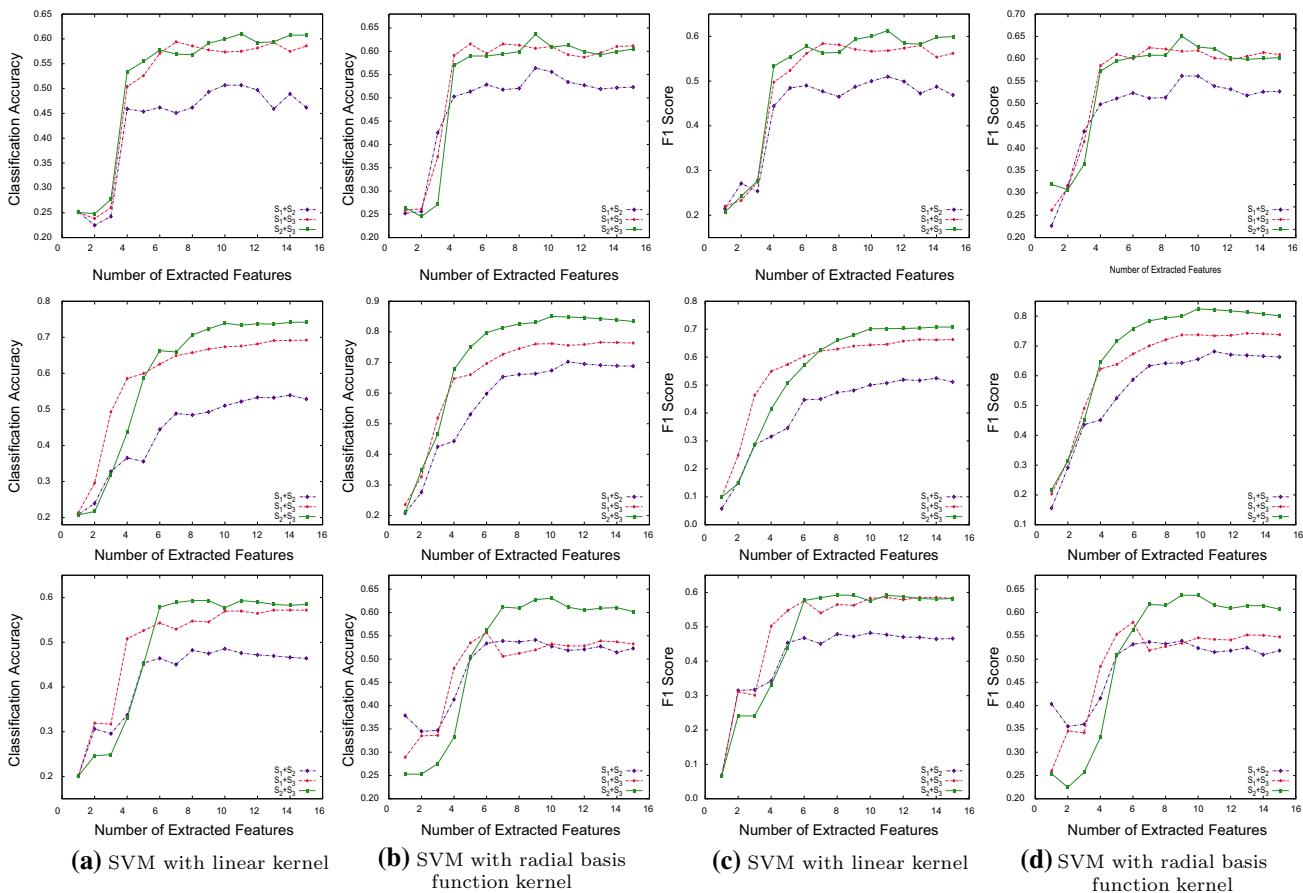
Fig. 3 Performance analysis of LBP<sup>ri</sup> for MIVIA, ICPR and SNP data sets at different scales

#### 4.4 Performance of different methods

This section provides the comparative performance analysis, in terms of classification accuracy and F1 score, of different existing data integration methods considering several classifiers. Results are reported on three HEp-2 cell image databases. To incorporate the information of

**Table 2** Classification accuracy and F1 score on different data sets for training–testing

Different classifiers	Different methods	ELM			SVM <sub>L</sub>			SVM <sub>P</sub>			SVM <sub>R</sub>		
		MIVIA	ICPR	SNP	MIVIA	ICPR	SNP	MIVIA	ICPR	SNP	MIVIA	ICPR	SNP
Accuracy	LBP	0.475	0.724	0.492	0.549	0.726	0.537	0.312	0.656	0.347	0.538	0.798	0.525
	LBP <sup>ri</sup>	0.598	0.811	0.519	0.584	0.770	0.545	0.564	0.753	0.592	0.582	0.826	0.589
	LBP <sup>riu2</sup>	0.549	0.809	0.512	0.584	0.702	0.530	0.302	0.769	0.466	0.546	0.819	0.525
	CCA	0.357	0.590	0.498	0.381	0.533	0.522	0.409	0.563	0.470	0.448	0.638	0.546
	RCCA	0.567	0.664	0.433	0.579	0.385	0.439	0.575	0.596	0.455	0.413	0.680	0.481
	CuRSaR	0.462	0.631	0.453	0.422	0.426	0.501	0.553	0.589	0.449	0.360	0.664	0.505
	Proposed CCA	0.549	0.799	0.561	<b>0.608</b>	0.742	0.585	0.545	0.743	0.560	0.605	<b>0.834</b>	<b>0.602</b>
F1 score	LBP	0.453	0.701	0.484	0.539	0.710	0.532	0.395	0.655	0.400	0.506	0.774	0.524
	LBP <sup>ri</sup>	0.589	0.800	0.507	0.591	0.755	0.542	0.571	0.722	0.592	0.593	0.799	0.590
	LBP <sup>riu2</sup>	0.563	0.792	0.508	0.585	0.684	0.533	0.352	0.740	0.470	0.570	0.800	0.518
	CCA	0.359	0.562	0.509	0.378	0.515	0.529	0.378	0.541	0.483	0.458	0.597	0.546
	RCCA	0.562	0.633	0.446	0.576	0.376	0.454	0.580	0.571	0.467	0.452	0.649	0.483
	CuRSaR	0.478	0.596	0.460	0.450	0.409	0.518	0.563	0.564	0.457	0.411	0.633	0.508
	Proposed CCA	0.514	0.770	0.567	0.600	0.708	0.582	0.527	0.708	0.559	<b>0.602</b>	<b>0.801</b>	<b>0.607</b>



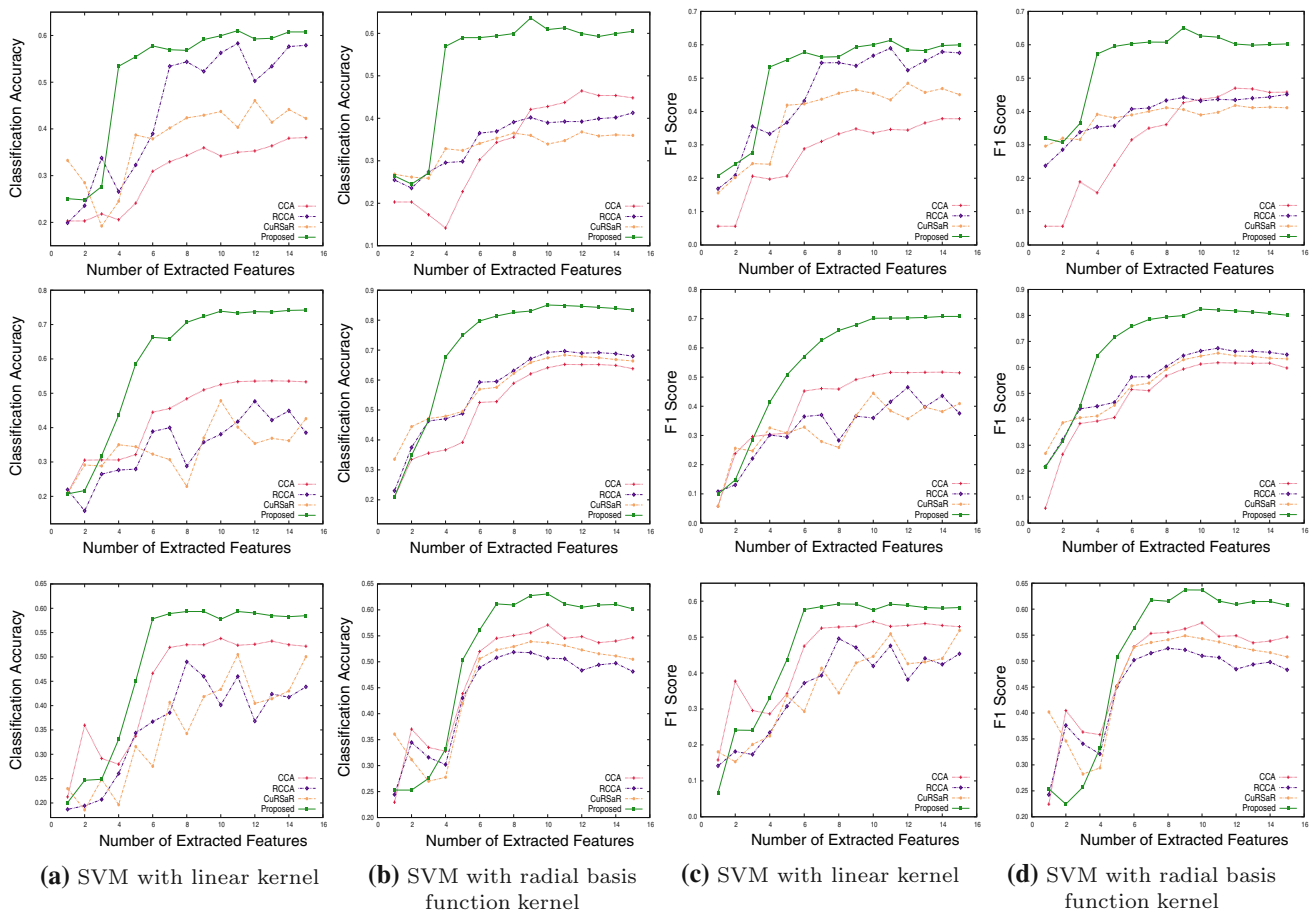
**Fig. 4** Performance of proposed method at different scales for training–testing (top: MIVIA; middle: ICPR; and bottom: SNP)

rotation-invariant local binary pattern ( $LBP^{ri}$ ) at scales  $S_2$  and  $S_3$ , the CCA, RCCA, CuRSaR and the proposed sequential CCA method are used. Figure 5 depicts the variation of F1 score and classification accuracy obtained using both linear kernel and radial basis function kernel of SVM over different number of extracted features. From the results reported in Fig. 5, it can be seen that the proposed data integration method provides highest classification accuracy as well as highest F1 score for training–testing, integrating the local textural features of  $LBP^{ri}$  at scales  $S_2$  and  $S_3$ , irrespective of the number of extracted features, SVM kernels and data sets used.

Table 2 compares the HEP-2 cell staining pattern F1 score and classification accuracy of the proposed as well as different existing methods for training–testing. The results corresponding to naive integration of scales  $S_2$  and  $S_3$  for LBP,  $LBP^{ri}$  and  $LBP^{riu2}$  are also studied in this table, where the number of features which are extracted by the naive integration of LBP,  $LBP^{ri}$  and  $LBP^{riu2}$  is 512, 72 and 20, respectively. On the other hand, the classification accuracy and F1 score of top fifteen features, which are extracted by

using CCA, RCCA, CuRSaR and the proposed sequential CCA method, are reported in this table. All the results confirm that the proposed data integration method provides highest classification accuracy and F1 score of the SVM with radial basis function kernel on ICPR and SNP data sets. For MIVIA data set, the proposed data integration method achieves highest classification accuracy and F1 score of the SVM with linear kernel and radial basis function kernel, respectively.

All the results reported in Fig. 5 and Table 2 establish the fact that the proposed method can extract most relevant and significant features for HEP-2 cell staining pattern classification, which are also maximally correlated between two scales  $S_2$  and  $S_3$ . To optimize the regularization parameters in the proposed data integration method, the rough hypercuboid approach is used. It helps to extract more relevant and significant local features than other multimodal data integration methods. In effect, the proposed method can facilitate higher HEP-2 cell pattern classification accuracy than other methods. Moreover, it can explore large search space to extract features



**Fig. 5** Comparative performance analysis of different data integration methods for training–testing (top: MIVIA; middle: ICPR; and bottom: SNP)

sequentially from two multidimensional scales with significantly lesser amount of time than FaRoC [25]. Both FaRoC and proposed CCA extract features sequentially from two multidimensional variables, but the proposed CCA takes lesser amount of time than FaRoC. Figure 6 compares the execution time of FaRoC with proposed data integration method. It confirms that the execution time of the proposed sequential feature extraction method, presented in Sect. 2, is significantly lower than that of existing FaRoC.

#### 4.5 Statistical significance analysis

This section presents the statistical significance analysis of staining pattern recognition of HEP-2 cell images using different multimodal data integration methods. To study the statistical significance of both accuracy and F1 score, tenfold cross-validation is performed. Figure 7 presents the variation of both tenfold accuracy and F1 score, for fifteen top-ranked extracted features. From the results presented in Fig. 7, it is clearly observed that both F1 score and classification accuracy for the proposed method increase while the number of generated features increases. Also, the F1 score and classification accuracy of the proposed algorithm, obtained using tenfold cross-validation, are significantly higher as compared to existing CCA, RCCA and CuRSaR, irrespective of the number of extracted features, classifiers and data sets used.

The comparative performance analysis of different data integration algorithms is also studied in Fig. 8, which shows the box and whisker plots for classification accuracy and F1 score obtained using tenfold cross-validation, on each HEP-2 cell data set using both linear kernel and radial basis function kernel of SVM. The lower and upper boundaries of each box represent the 75th percentile or lower quartile and 25th percentile or upper quartile, respectively, whereas the median is denoted by the central

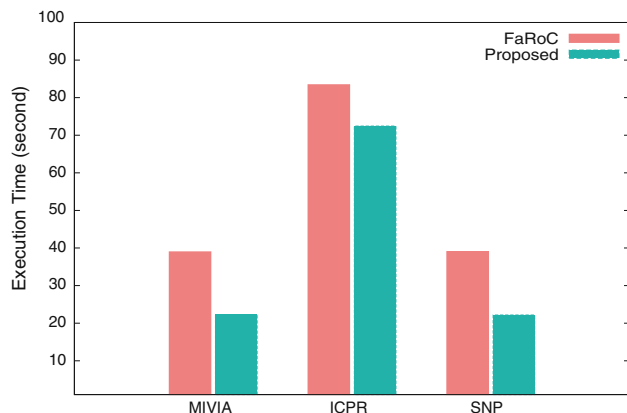


Fig. 6 Execution time of proposed CCA and FaRoC

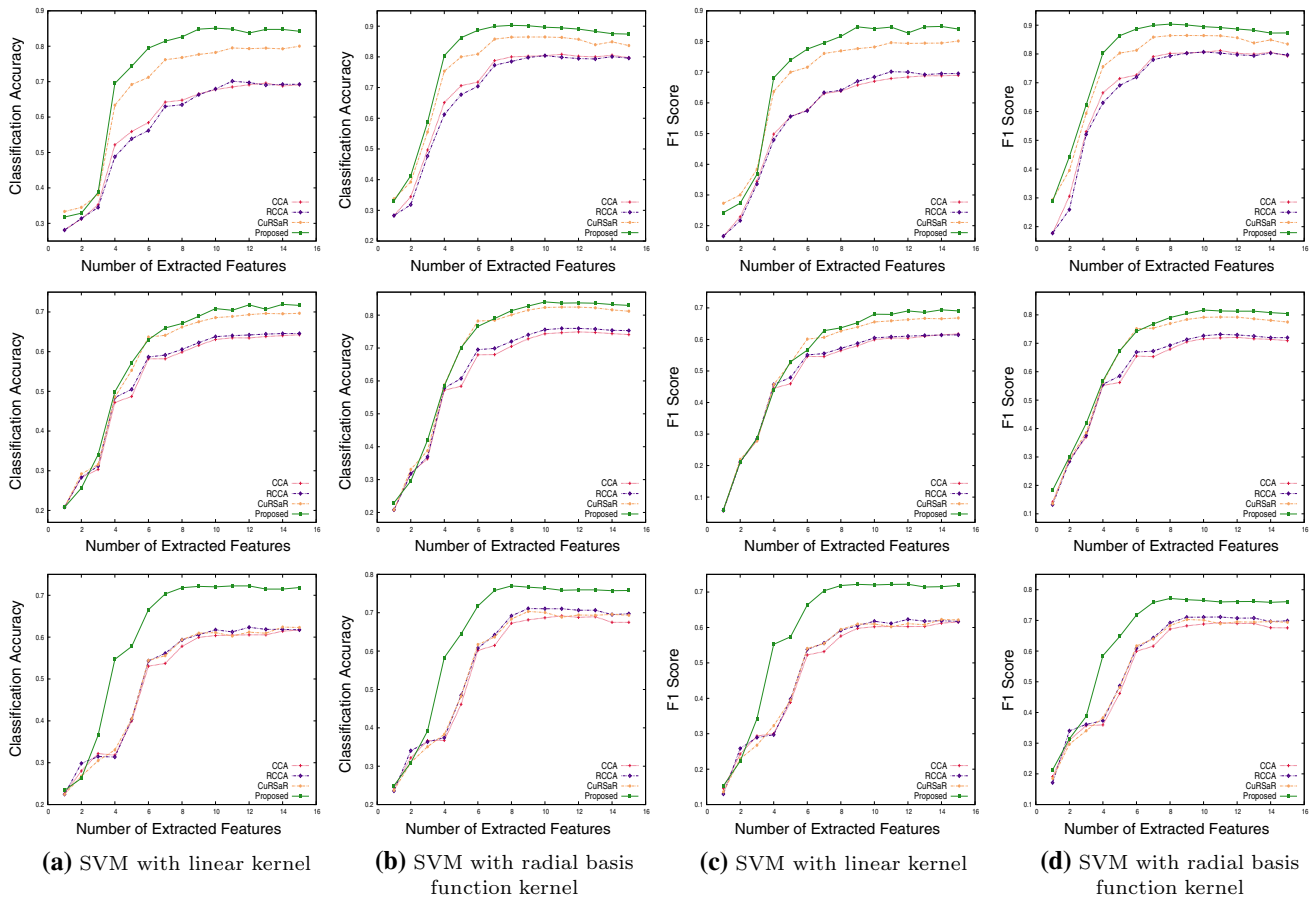
line. The whiskers are the two lines outside the box that extend to the highest and lowest observations. The outliers are plotted individually, and denoted as ‘+’. It represents the test cross-validation set for which a method obtains worse tenfold F1 score or classification accuracy than for the other test cross-validation sets of the same HEP-2 cell data set.

Wilcoxon signed-rank test (one-tailed), paired-*t* test (one-tailed) and Friedman test are used to compute the *p* values for statistical significance analysis. In case of tenfold cross-validation, the means and standard deviations of the classification accuracy and F1 score of the ELM, SVM<sub>L</sub>, SVM<sub>P</sub> and SVM<sub>R</sub> rules are computed for all data sets. Tests of significance are performed for the inequality of means obtained using the proposed method and other related algorithms compared. Since both mean pairs and variance pairs are unknown and different, a generalized version of *t*-test, termed as Behrens–Fisher problem in hypothesis testing, is used here. On the other hand, Wilcoxon signed-rank test computes the differences between the proposed method and other related algorithms using F1 score and accuracy for each observation. The *p* values are computed by using positive and negative ranks of absolute differences. Friedman test is a nonparametric statistical test which is used to detect the differences between the proposed method and other related algorithms by ranking procedure.

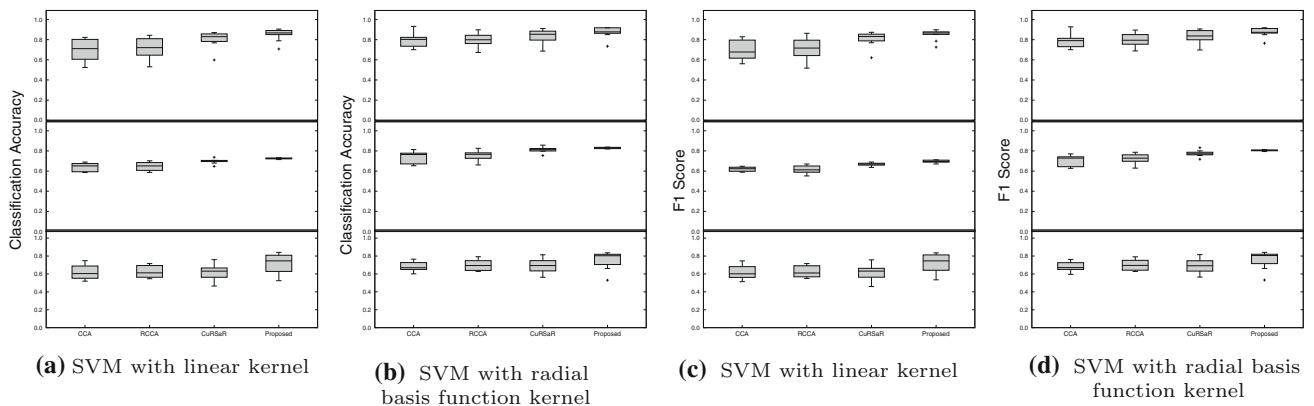
Table 3 reports the means and standard deviations of tenfold cross-validation F1 score and accuracy for all the methods using several classifiers. The *p* values of existing data integration methods are also recorded in this table with respect to the proposed sequential CCA method for three HEP-2 cell data sets. The highest mean values are pointed in bold in this table. All the results presented in Table 3 establish the fact that the proposed HEP-2 cell staining pattern recognition method, which mainly consists of the proposed sequential CCA and SVM with radial basis function kernel, attains best mean classification accuracy as well as F1 score, irrespective of the data integration methods, classifiers and data sets used. Out of total 216 cases, the proposed method attains significantly better *p* values (marked in bold) than other methods in 139 cases and better but not significant *p* values (marked in italics) in 66 cases, considering 95% confidence level.

## 5 Conclusion and discussions

The main contribution of this paper lies in developing a methodology, termed as CanSuR, which can be used to diagnose connective tissue disease by recognizing the staining patterns present in HEP-2 cells. When an HEP-2 cell image is rotated, the pattern operator should be rotation



**Fig. 7** Comparative performance analysis of different data integration methods for tenfold cross-validation (top: MIVIA; middle: ICPR; and bottom: SNP)



**Fig. 8** Box and whisker plots for different data integration methods (top: MIVIA; middle: ICPR; and bottom: SNP)

invariant such that it should produce a unique binary pattern for the same texture. In this context, the proposed method uses rotation-invariant local binary pattern as texture descriptor operator to extract important features from HEP-2 cell images. On the other hand, integration of two scales may provide better recognition of HEP-2 patterns.

But, naive integration, which is the direct concatenation of two scales, may not be effective for integrating important information of two scales. A proper multimodal data integration method is thus needed to incorporate the information of two scales. In this regard, a new supervised CCA algorithm is proposed in the current research work to

**Table 3** Classification accuracy and F1 score of different data integration methods for tenfold cross-validation

Data sets	Different classifiers	Different algorithms	Accuracy				F1 score						
			Mean	SD	Paired-t:p	Friedman:p	Wilcoxon:p	Mean	SD	Paired-t:p	Friedman:p	Wilcoxon:p	
MIVIA	ELM	CCA	0.739	0.096	<b>5.9E-03</b>	2.1E-01	<b>1.4E-02</b>	0.735	0.095	<b>1.3E-02</b>	2.1E-01	<b>1.8E-02</b>	
		RCCA	0.726	0.097	<b>2.9E-03</b>	<b>1.1E-02</b>	<b>4.7E-03</b>	0.729	0.093	<b>8.7E-03</b>	<b>1.1E-02</b>	<b>8.3E-03</b>	
	SVM <sub>L</sub>	CuRSaR	0.795	0.077	8.9E-02	3.2E-01	8.6E-02	0.794	0.072	2.0E-01	5.3E-01	2.9E-01	
		Proposed CCA	0.837	0.045				0.821	0.052				
	SVM <sub>P</sub>	CCA	0.691	0.119	<b>5.8E-03</b>	2.1E-01	<b>1.4E-02</b>	0.690	0.117	<b>5.8E-03</b>	2.1E-01	<b>1.8E-02</b>	
		RCCA	0.693	0.111	<b>4.2E-03</b>	<b>2.0E-02</b>	<b>1.0E-02</b>	0.696	0.106	<b>4.5E-03</b>	<b>1.1E-02</b>	<b>8.3E-03</b>	
	SVM <sub>R</sub>	CuRSaR	0.800	0.078	1.1E-01	5.8E-02	5.1E-02	0.801	0.074	1.3E-01	2.1E-01	7.0E-02	
		Proposed CCA	0.842	0.058				0.841	0.062				
	ICPR	ELM	CCA	0.722	0.096	<b>2.4E-02</b>	2.1E-01	<b>3.0E-02</b>	0.725	0.088	<b>4.6E-02</b>	2.1E-01	<b>3.0E-02</b>
			RCCA	0.739	0.087	<b>3.3E-02</b>	3.2E-01	<b>2.5E-02</b>	0.746	0.082	8.3E-02	5.3E-01	1.0E-01
SVM <sub>L</sub>		CuRSaR	0.802	0.078	3.4E-01	1.0E+00	5.4E-01	0.803	0.072	5.2E-01	5.3E-01	7.5E-01	
		Proposed CCA	0.816	0.055				0.801	0.064				
SVM <sub>P</sub>		CCA	0.797	0.070	<b>7.0E-03</b>	9.6E-02	<b>1.0E-02</b>	0.793	0.069	<b>3.7E-03</b>	5.8E-02	<b>1.1E-02</b>	
		RCCA	0.796	0.065	<b>4.2E-03</b>	<b>1.1E-02</b>	<b>6.3E-03</b>	0.797	0.062	<b>2.1E-03</b>	<b>1.1E-02</b>	<b>6.3E-03</b>	
SVM <sub>R</sub>		CuRSaR	0.837	0.065	<b>4.1E-02</b>	<b>2.0E-02</b>	<b>3.3E-02</b>	0.834	0.061	<b>3.0E-02</b>	5.8E-02	<b>1.8E-02</b>	
		Proposed CCA	<b>0.874</b>	0.055				<b>0.873</b>	0.045				
SNP		ELM	CCA	0.684	0.048	<b>3.3E-05</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	0.659	0.040	<b>1.2E-05</b>	<b>1.6E-03</b>	<b>2.5E-03</b>
			RCCA	0.692	0.052	<b>6.2E-05</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	0.663	0.049	<b>3.2E-05</b>	<b>1.6E-03</b>	<b>2.5E-03</b>
	SVM <sub>L</sub>	CuRSaR	0.743	0.023	<b>1.0E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	0.710	0.021	<b>3.5E-05</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	
		Proposed CCA	0.797	0.007				0.770	0.010				
	SVM <sub>P</sub>	CCA	0.643	0.040	<b>2.0E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	0.617	0.022	<b>2.1E-06</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	
		RCCA	0.646	0.042	<b>6.3E-04</b>	<b>1.1E-02</b>	<b>3.5E-03</b>	0.614	0.037	<b>1.0E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	
	SVM <sub>R</sub>	CuRSaR	0.697	0.022	<b>4.6E-02</b>	2.1E-01	<b>2.3E-02</b>	0.668	0.016	<b>4.4E-03</b>	<b>1.1E-02</b>	<b>3.5E-03</b>	
		Proposed CCA	0.717	0.020				0.690	0.015				
	ELM	CCA	0.697	0.047	<b>1.3E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	0.669	0.037	<b>1.3E-05</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	
		RCCA	0.707	0.061	<b>1.5E-03</b>	<b>1.1E-02</b>	<b>3.5E-03</b>	0.675	0.053	<b>1.6E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	
SVM <sub>L</sub>	CuRSaR	0.767	0.032	5.4E-02	3.2E-01	4.3E-02	0.734	0.032	<b>4.6E-03</b>	<b>1.1E-02</b>	<b>8.3E-03</b>		
	Proposed CCA	0.786	0.007				0.769	0.007					
SVM <sub>P</sub>	CCA	0.741	0.057	<b>4.0E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	0.709	0.053	<b>1.3E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>		
	RCCA	0.753	0.056	<b>8.4E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>	0.720	0.051	<b>1.6E-04</b>	<b>1.6E-03</b>	<b>2.5E-03</b>		
SVM <sub>R</sub>	CuRSaR	0.812	0.027	<b>3.8E-02</b>	5.8E-02	<b>2.3E-02</b>	0.775	0.030	<b>5.7E-03</b>	5.8E-02	<b>8.3E-03</b>		
	Proposed CCA	<b>0.830</b>	0.006				<b>0.805</b>	0.007					
ELM	CCA	0.636	0.074	<b>4.0E-02</b>	2.1E-01	5.7E-02	0.636	0.074	<b>4.1E-02</b>	2.1E-01	5.7E-02		
	RCCA	0.650	0.067	<b>4.4E-02</b>	2.1E-01	5.7E-02	0.650	0.067	<b>4.5E-02</b>	2.1E-01	5.7E-02		



**Table 3** (continued)

Data sets	Different classifiers	Different algorithms	Accuracy					F1 score				
			Mean	SD	Paired-t/p	Friedman:p	Wilcoxon:p	Mean	SD	Paired-t/p	Friedman:p	Wilcoxon:p
SVM <sub>L</sub>	CuRSaR		0.649	0.100	1.4E-01	5.3E-01	1.2E-01	0.647	0.102	1.3E-01	5.3E-01	1.0E-01
	Proposed CCA		0.714	0.108				0.714	0.108			
	CCA		0.617	0.077	2.1E-02	2.1E-01	2.3E-02	0.615	0.075	1.9E-02	2.1E-01	1.8E-02
	RCCA		0.617	0.069	1.8E-02	5.8E-02	2.3E-02	0.617	0.068	1.8E-02	5.8E-02	2.3E-02
SVM <sub>p</sub>	CuRSaR		0.623	0.091	3.6E-02	5.8E-02	5.7E-02	0.621	0.091	3.5E-02	5.8E-02	3.7E-02
	Proposed CCA		0.719	0.103				0.719	0.104			
	CCA		0.609	0.073	1.9E-02	5.8E-02	2.3E-02	0.615	0.070	1.8E-02	5.8E-02	2.3E-02
	RCCA		0.618	0.049	8.9E-03	5.8E-02	1.2E-02	0.630	0.047	9.3E-03	5.8E-02	1.4E-02
SVM <sub>R</sub>	CuRSaR		0.629	0.096	7.1E-02	5.3E-01	8.4E-02	0.638	0.094	7.0E-02	5.3E-01	8.4E-02
	Proposed CCA		0.716	0.108				0.723	0.105			
	CCA		0.675	0.055	2.3E-02	5.8E-02	2.3E-02	0.676	0.054	2.1E-02	5.8E-02	2.3E-02
	RCCA		0.697	0.057	3.9E-02	5.8E-02	4.6E-02	0.699	0.057	3.6E-02	5.8E-02	3.7E-02
	CuRSaR		0.693	0.077	9.0E-02	2.1E-01	1.2E-01	0.694	0.077	8.0E-02	2.1E-01	7.0E-02
	Proposed CCA		<b>0.758</b>	<b>0.098</b>				<b>0.761</b>	<b>0.098</b>			

integrate the information of two scales. Finally, support vector machine with radial basis function kernel is used to recognize one of the known staining patterns present in IIF images. The proposed method takes into account the merits of rough hypercuboid approach and supervised CCA. While the proposed CCA helps to integrate local textural descriptors obtained from multimodal sources, the rough hypercuboid facilitates to extract significant and relevant features for HEp-2 pattern recognition. The effectiveness of the proposed method, along with a comparison with related approaches, has been demonstrated on several publicly available HEp-2 cell image databases.

The proposed method is basically the realization of computer-aided diagnosis system for the analysis of IIF images. The results produced by the proposed technique can be used to support the scientists' subjective analysis. In effect, it may lead to prediction accuracy on test samples being consistent across laboratories and more reliable. So, one can use the proposed system to automatically identify the patterns present in the specimen HEp-2 cell images, in order to address the shortcomings of manual test procedure. Both scientific and industrial societies may be interested about the proposed method for automatic IIF image pattern analysis as it will reduce high labor costs and increase the reliability especially in the presence of photo-effect which bleaches the tissues severely in a few seconds. The proposed method may also avoid ambiguous results caused by subjective analysis and provide more efficient analysis report.

**Acknowledgements** This work was partially supported by the Department of Science and Technology, Government of India (Grant No. SB/S3/EECE/050/2015). The authors would like to thank Debamita Kumar of Indian Statistical Institute, Kolkata, for her valuable experimental support.

**References**

1. Agmon-Levin N, Damoiseaux J, Kallenberg C, Sack U, Witte T, Herold M, Bossuyt X, Musset L, Cervera R, Plaza-Lopez A, Dias C, Sousa MJ, Radice A, Eriksson C, Hultgren O, Viander M, Khamashta M, Regenass S, Andrade LEC, Wiik A, Tincani A, Rönnelid J, Bloch DB, Fritzler MJ, Chan EKL, Garcia-De La Torre I, Konstantinov KN, Lahita R, Wilson M, Vainio O, Fabien N, Sinico RA, Meroni P, Shoenfeld Y (2014) International recommendations for the assessment of autoantibodies to cellular antigens referred to as anti-nuclear antibodies. *Ann Rheum Dis* 73(1):17–23
2. Agmon-Levin N, Shapira Y, Selmi C, Barzilai O, Ram M, Szyper-Kravitz M, Sella S, Katz BS, Youinou P, Renaudineau Y, Larida B, Invernizzi P, Gershwin ME, Shoenfeld Y (2010) A comprehensive evaluation of serum autoantibodies in primary biliary cirrhosis. *J Autoimmun* 34(1):55–58
3. Banerjee A, Maji P (2016) Rough-probabilistic clustering and hidden markov random field model for segmentation of HEp-2 cell and brain MR images. *Appl Soft Comput* 46:558–576

4. Banerjee A, Maji P (2017) Stomped-*r*: a novel probability distribution for rough-probabilistic clustering. *Inf Sci* 421:104–125
5. Di Cataldo S, Bottino A, Islam I, Vieira TF, Ficarra E (2014) Subclass discriminant analysis of morphological and textural features for HEp-2 staining pattern classification. *Pattern Recognit* 47(7):2389–2399
6. Cordelli E, Soda P (2011) Color to grayscale staining pattern representation in IIF. In Proceedings of the 24th international symposium on computer-based medical systems, pp 1–6
7. Cruz-Cano R, Lee MT (2014) Fast regularized canonical correlation analysis. *Comput Stat Data Anal* 70:88–100
8. Ensafi S, Lu S, Kassim AA, Tan CL (2016) Accurate HEp-2 cell classification based on sparse coding of superpixels. *Pattern Recognit Lett* 82:64–71
9. Foggia P, Percannella G, Soda P, Vento M (2010) Early experiences in mitotic cells recognition on HEp-2 slides. In: Proceedings of the 23rd IEEE international symposium on computer-based medical systems, pp 38–43
10. Foggia P, Percannella G, Soda P, Vento M (2013) Benchmarking HEp-2 cells classification methods. *IEEE Trans Med Imaging* 32(10):1878–1889
11. Friou GJ, Finch SC, Detre KD, Santarsiero C (1958) Interaction of nuclei and globulin from lupus erythematosus serum demonstrated with fluorescent antibody. *J Immunol* 80(4):324–329
12. Gao Z, Wang L, Zhou L, Zhang J (2017) HEp-2 cell image classification with deep convolutional neural networks. *IEEE J Biomed Health Inf* 21(2):416–428
13. Gladwell GML (1995) On isospectral spring mass systems. *Inverse Probl* 11(3):591–602
14. Golugula A, Lee G, Master SR, Feldman MD, Tomaszewski JE, Speicher DW, Madabhushi A (2011) Supervised regularized canonical correlation analysis: integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery. *BMC Bioinf* 12:483
15. Guo Z, Zhang L, Zhang D (2010) A completed modeling of local binary pattern operator for texture classification. *IEEE Trans Image Process* 19(6):1657–1663
16. Hiemann R, Buttner T, Krieger T, Roggenbuck D, Sack U, Conrad K (2009) Challenges of automated screening and differentiation of non-organ specific autoantibodies on HEp-2 cells. *Autoimmun Rev* 9(1):17–22
17. Hiemann R, Hilger N, Sack U, Weigert M (2006) Objective quality evaluation of fluorescence images to optimize automatic image acquisition. *Cytometry A* 69(3):182–184
18. Hotelling H (1936) Relations between two sets of variates. *Biometrika* 28(3/4):321–377
19. Hsieh TY, Huang YC, Chung CW, Huang YL (2009) HEp-2 cell classification in indirect immunofluorescence images. In: Proceedings of the 7th international conference on information, communications and signal processing, pp 1–4
20. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1):489–501
21. Humbel RL (1993) Detection of antinuclear antibodies by immunofluorescence. *Man Biolog Markers Dis* A2:1–16
22. Li Y, Shen L, Yu S (2017) HEp-2 specimen image segmentation and classification using very deep fully convolutional network. *IEEE Trans Med Imaging* 36(7):1561–1572
23. Maji P (2014) A rough hypercuboid approach for feature selection in approximation spaces. *IEEE Trans Knowl Data Eng* 26(1):16–29
24. Maji P, Mandal A (2017) Multimodal omics data integration using max relevance-max significance criterion. *IEEE Trans Biomed Eng* 64(8):1841–1851
25. Mandal A, Maji P (2018) FaRoC: fast and robust supervised canonical correlation analysis for multimodal omics data. *IEEE Trans Cybern* 48(4):1229–1241
26. Mariz HA, Sato EI, Barbosa SH, Rodrigues SH, Dellavance A, Andrade LE (2011) Pattern on the antinuclear antibody-HEp-2 test is a critical parameter for discriminating antinuclear antibody-positive healthy individuals and patients with autoimmune rheumatic diseases. *Arthritis Rheum* 63(1):191–200
27. Nosaka R, Fukui K (2014) HEp-2 cell classification using rotation invariant co-occurrence among local binary patterns. *Pattern Recognit* 47(7):2428–2436
28. Nosaka R, Ohkawa Y, Fukui K (2012) Feature extraction based on co-occurrence of adjacent local binary patterns. In: Proceedings of the 5th Pacific Rim conference on advances in image and video technology, pp 82–91. Springer, Berlin
29. Ojala T, Pietikainen M, Harwood D (1994) Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Proceedings of the 12th IAPR international conference on pattern recognition, conference on computer vision & image processing, pp 582–585
30. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
31. Ojala T, Valkealahti K, Oja E, Pietikainen M (2001) Texture discrimination with multidimensional distributions of signed gray-level differences. *Pattern Recognit* 34(3):727–739
32. Qi X, Zhao G, Li C, Guo J, Pietikainen M (2017) HEp-2 cell classification via combining multiresolution co-occurrence texture and large region shape information. *IEEE J Biomed Health Inf* 21(2):429–440
33. Roy S, Maji P (2017) Rough-fuzzy segmentation of HEp-2 cell indirect immunofluorescence images. *Int J Data Min Bioinf* 17(4):311–340
34. Soda P, Iannello G (2006) A multi-expert system to classify fluorescent intensity in antinuclear autoantibodies testing. In: Proceedings of the 19th IEEE symposium on computer-based medical systems, pp 219–224
35. Soda P, Rigon A, Afeltra A, Iannello G (2006) Automatic acquisition of immunofluorescence images: algorithms and evaluation. In: Proceedings of the 19th IEEE symposium on computer-based medical systems, pp 386–390
36. Solomon DH, Kavanaugh AJ, Schur PH (2002) Evidence-based guidelines for the use of immunologic tests: antinuclear antibody testing. *Arthritis Rheum* 47(4):434–444
37. Strandmark P, Ulen J, Kahl F (2012) HEp-2 staining pattern classification. In: Proceedings of the 21st international conference on pattern recognition, pp 33–36
38. Tan EM (1989) Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology. *Adv Immunol* 44:93–151
39. Theodorakopoulos I, Kastaniotis D, Economou G, Fotopoulos S (2012) HEp-2 cells classification via fusion of morphological and textural features. In: Proceedings of the 12th IEEE international conference on bioinformatics and bioengineering, pp 689–694
40. Theodorakopoulos I, Kastaniotis D, Economou G, Fotopoulos S (2014) HEp-2 cells classification via sparse representation of textural features fused into dissimilarity space. *Pattern Recognit* 47(7):2367–2378

41. Vapnik V (1995) The nature of statistical learning theory. Springer, New York
42. Vinod HD (1976) Canonical ridge and econometrics of joint production. *J Economet* 4(2):147–166
43. White PA (1958) The computation of eigenvalues and eigenvectors of a matrix. *J Soc Ind Appl Math* 6(4):393–437
44. Wiik AS (2005) Anti-nuclear autoantibodies: clinical utility for diagnosis, prognosis, monitoring, and planning of treatment strategy in systemic immunoinflammatory diseases. *Scand J Rheumatol* 34(4):260–268
45. Wiliem A, Wong Y, Sanderson C, Hobson P, Chen S, Lovell BC (2013) Classification of human epithelial type 2 cell indirect immunofluorescence images via codebook based descriptors. In: Proceedings of the IEEE workshop on applications of computer vision, pp 95–102

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.