An Efficient Method for Automatic Recognition of Virus Particles in TEM Images

Debamita Kumar[(✉)] and Pradipta Maji

Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit,
Indian Statistical Institute, Kolkata, India
{debamita_r,pmaji}@isical.ac.in

Abstract. The conventional approach for virus detection is based on analyzing the negative stain Transmission Electron Microscopy (TEM) images. In this regard, a new method is presented here based on judicious integration of the theory of rough sets and the merits of local texture descriptors. The proposed method identifies the relevant local texture descriptor for representing the intrinsic properties of each pair of virus classes. It also selects important features from each of the relevant descriptors, which can suitably describe significant characteristics of the virus particles present in TEM images. The hypercuboid equivalence partition matrix of rough sets is employed to evaluate the relevance of texture descriptors. Finally, support vector machine (SVM) is used to categorize the virus samples into one of the known virus classes. The efficiency of the proposed method, along with a comparison with related approaches, is demonstrated on publicly available Virus data set.

Keywords: Virus recognition · Transmission Electron Microscopy (TEM) images · Rough sets · Feature selection · Local texture descriptor

1 Introduction

Analysis of negative stain TEM images has long been used for the discovery and detection of virus particles. The visualization of virus structures has been possible only after the development of TEM. The study of virus structures through TEM images allows organism-specific reagents for the recognition of pathogenic agent, without any a priori knowledge of pathogens present in a sample [3]. This enables broader examination of viruses, making TEM an inevitable method for virus diagnosis. Negative staining is a preferred technique for observing small particles, such as viruses, in fluids. However, high maintenance costs, level of expertise and time required for manual inspection of the TEM images, and advancement in automated image acquisition process entail the development of automated analysis of TEM images for virus detection.

TEM images provide important information regarding the shape and size of the virus particles, which can be used to group the particles into various classes, for example, viruses like Adeno, Astro and Rota form icosahedral structure, Cowpox, Dengue and Lassa viruses reflect regularity in their structures, whereas Ebola, Influenza and Marburg viruses are highly irregular in shape. Early approaches for virus identification were mostly dependent on the morphology of the particles [1,2]. But, the morphological information obtained from the TEM images does not proved to be sufficient for classification of the viruses.

Apart from morphology, different viruses exhibit different surface textures when imaged using TEM. This information can be efficiently utilized using various texture analysis techniques for automatic classification of the virus particles from TEM images. Matuszewski et al. [12] have transformed the spatial information of the images into frequency domain using discrete Fourier transform and then, extracted features from the multiple spectral rings formed at the magnitude spectrum of TEM images to differentiate the four icosahedral structured viruses, namely, Adeno, Astro, Rota and Calici. In [17], a method has been proposed based on higher order spectral features, to capture the contour and texture information from the same four icosahedral virus images for the diagnosis. The radial density profile (RDP) has been used in [19] to discriminate the intensity variation between three maturation stages of human cytomegalovirus capsids present in the cell sections of TEM images. The RDP has also been computed based on Fourier magnitude spectrum (FRDP) in [9], which can be interpreted as a generalization of the spectral rings considered in [17]. Moreover, Harandi et al. [6] have mapped the original data to a high-dimensional Hilbert space and then computed the Covariance Descriptors (CovDs) in infinite-dimensional spaces, to address the virus classification problem. Finally, several Bregman divergences have been used to compute the dissimilarities between the resulting CovDs in Hilbert space, among which Jeffreys and Stein divergences based on the SVM have been found to achieve the best result. Recently, a new method, based on convolutional neural network, has been proposed in [7], which transforms a TEM image to a probabilistic map for virus particle detection.

The concepts of local binary pattern (LBP) [16] and RDP (LBP+RDP) have been used in [9] to identify different viruses from TEM images using random forest classifier. Nanni et al. [13] have presented two methods, namely, NewH and Fusion; the former focuses on extracting descriptors from the co-occurrence matrix with the goal of enhancing the performance of Haralick's descriptors [5], while the latter is an ensemble of local phase quantization variants with ternary encoding. From the results reported in literature, it can be noticed that different virus classes are well described by different textures, which hinders the overall virus detection ability. One of the main problems in virus detection is uncertainty, which may have been originated due to incompleteness in class definition and overlapping class boundaries. An effective paradigm to deal with incompleteness and uncertainty is the theory of rough sets [18]. It provides a mathematical framework to model uncertainties related to the given data.

In this regard, the paper presents a new method for automatic virus classification from TEM images, by judiciously integrating the merits of local texture descriptors and the theory of rough sets. Given a set of virus TEM images, the proposed method first identifies a subset of important features from the original feature set of the local texture descriptors, computed at a specific scale and then forms the relevant feature set corresponding to each class-pair which can appropriately differentiate the pair of virus classes. The theory of rough hypercuboid approach is employed to evaluate the relevance of a texture descriptor in categorizing virus samples into multiple classes. In the proposed approach, the important feature set is selected in such a way that it not only characterizes the virus samples present in the data set, but also the classes to which the samples belong. The set of relevant texture descriptors, selected from the important feature sets, corresponding to each the class-pair is considered to form the feature set for multiple virus classes. The SVM with linear kernel is used to evaluate the performance of the proposed method. The effectiveness of proposed method, along with a comparison with related approaches, is demonstrated on a real-life Virus image database.

2   Proposed Algorithm

The inherent texture properties of virus TEM images are quite different from each other. The inter-class structural similarities, intra-class variations in appearance, uncertainty in virus class definitions, and presence of noise further amplify the difficulty of texture identification. In this context, the proposed method introduces a new approach for classifying viruses from TEM images. It first identifies important features, which effectively capture significant characteristics of each pair of virus classes, from the relevant texture descriptors evaluated at a particular scale, and finally forms the feature set for multiple virus classes.

Suppose, $\mathbb{U} = \{y_1, \cdots, y_k, \cdots, y_n\}$ denote the set of n training images of virus particles, where each image $y_k \in \Re^m$. Let $\mathbb{C} = \{F_1, \cdots, F_j, \cdots, F_m\}$ represent the set of m features. So, a virus TEM image y_k can be represented by a set $H_k = \{H_{k1}, \cdots, H_{kj}, \cdots, H_{km}\}$, containing m feature values corresponding to y_k, where $H_{kj} = y_k(F_j)$ represents the value of feature F_j obtained for image y_k. In the proposed approach, four local texture descriptors are considered, which are LBP: local binary pattern [16], LBP^{ri}: rotation-invariant LBP [15], LBP^{riu2}: rotation-invariant uniform LBP [15], and CoALBP: co-occurrence of adjacent LBP [14]. So, H_k represents the normalized histogram, obtained from either of the four local descriptors, for the image y_k. Let, H_k be arranged in descending order and denoted by $\tilde{H}_k = \{\tilde{H}_{k1}, \cdots, \tilde{H}_{kj}, \cdots, \tilde{H}_{km}\}$ such that $\tilde{H}_{k1} \geq \tilde{H}_{k2} \geq \cdots \geq \tilde{H}_{km}$, and the corresponding feature indices of $\tilde{H}_k$ is preserved in the set $J_k = \{J_{k1}, \cdots, J_{kj}, \cdots, J_{km}\}$. Assume, each of the TEM images of $\mathbb{U}$ belongs to one of the c virus classes $\mathbb{U}/\mathbb{D} = \{B_1, \cdots, B_i, \cdots, B_c\}$, where $\mathbb{D}$ denotes the set containing class label information.

Primarily, significant properties of an image y_k is represented by the values of H_k, which represents the normalized histogram of y_k. However, the proposed

method assumes that not all the feature values of H_k contribute uniformly in representing the characteristics of the image y_k. Indeed, only a subset of features of $\mathbb{C}$ can precisely illustrate important properties of y_k, which is defined as the important feature set of y_k and denoted by I_k, where $I_k \subseteq \mathbb{C}$. However, each image has it's own characteristics, which can be reflected by a specific set of important features.

The cumulative sum of first q features of the sorted normalized histogram $\tilde{H}_k$ is computed to select the important features of y_k and denoted by $E(y_k, q)$, which is defined as the energy function in the current study. Certainly, $E(y_k, m) = 1$ and $E(y_k, q) \in [0, 1], \forall y_k \in \mathbb{U}$. It signifies the fraction of total energy, present in $\tilde{H}_k$, which is retained by the first q features of y_k. Thus, relevant information regarding the properties of the image y_k can be suitably represented by the energy of y_k, evaluated from $\tilde{H}_k$. In order to retain a given fraction of energy E_0 of the sample y_k, the required number of important features d_k is obtained as the minimum value of q for which $E(y_k, q)$ attains E_0 value. The average number of important features $\overline{d}$ corresponding to the entire set of samples $\mathbb{U}$ is computed from the individual d_ks. As defined in [10], the set of important features I_k of the sample y_k can be obtained from the first $\overline{d}$ features of sorted histogram $\tilde{H}_k$, as follows:

$$I_k = \{F_j \mid J_{kq} = j \quad \text{and} \quad q \leq \overline{d}\}. \tag{1}$$

Thus, the important feature set $I_k \subseteq \mathbb{C}$ of the image y_k consists of only those features which can efficiently describe the inherent properties of y_k.

Now, it is most likely that the samples from a specific virus class will be represented by similar sets of important features, whereas samples from different classes will be represented by different important feature sets. Therefore, the probability of occurrence $P(F_j|B_i)$ of a feature F_j in the important sets of samples belonging to a particular class B_i is obtained and a threshold parameter ϵ is introduced to differentiate noisy features from important features. The feature set $\mathcal{C}(B_i)$ representing the class B_i is formed with only those features F_j's which have $P(F_j|B_i)$ values greater than or equal to ϵ. So, the set $\mathcal{C}(B_i)$ will only contain a feature F_j if it is found to be important for most of the samples of B_i and also reflects the relevant characteristics of the virus class B_i. Let, the significant characteristics of a pair of classes, say $\{B_i, B_r\}$, be represented by the set $\mathcal{C}(\{B_i, B_r\})$. So, the set should contain those features which can efficiently reflect the properties of both the classes B_i and B_r. Hence, $\mathcal{C}(\{B_i, B_r\})$ is formed by taking intersection of the two sets $\mathcal{C}(B_i)$ and $\mathcal{C}(B_r)$.

Let us consider a set of t modalities $\mathcal{M} = \{\mathcal{M}_1, \cdots, \mathcal{M}_p, \cdots, \mathcal{M}_t\}$, where each modality corresponds to a specific texture descriptor evaluated at a particular scale. Given the set of modalities, the proposed method computes the relevance $\Gamma_p(\{B_i, B_r\})$ of the feature set $\mathcal{C}_p(\{B_i, B_r\})$ under modality $\mathcal{M}_p$ to assess the efficacy of the set in describing important properties of the class-pair $\{B_i, B_r\}$. In order to quantify relevance, the concept of hypercuboid equivalence partition matrix of rough hypercuboid approach [11] is employed. The relevance

$\Gamma_p(\{B_i, B_r\})$ of the feature set $\mathcal{C}_p(\{B_i, B_r\})$ with respect to the pair of classes $\{B_i, B_r\}$ is obtained as follows [11]:

$$\Gamma_p(\{B_i, B_r\}) = 1 - \frac{1}{n_{ir}} \sum_{k=1}^{n_{ir}} v_k(\mathcal{C}_p(\{B_i, B_r\})) \tag{2}$$

where, n_{ir} is the number of samples which belongs to class-pair $\{B_i, B_r\}$ and

$$v_k(\mathcal{C}_p(\{B_i, B_r\})) = \begin{cases} 1 & \text{if } h_{ik}(\mathcal{C}_p) = 1 \text{ and } h_{rk}(\mathcal{C}_p) = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

$$\text{where } h_i(\mathcal{C}_p) = [h_{ik}(\mathcal{C}_p)]_{1 \times n_{ir}} = \bigcap_{F_j \in \mathcal{C}_p} h_i(F_j). \tag{4}$$

Here, $h_{ik}(F_j) \in \{0, 1\}$ represents the membership of sample y_k in i-th equivalence partition induced by the feature F_j and is determined as follows:

$$h_{ik}(F_j) = \begin{cases} 1 & \text{if } L(B_i) \leq y_k(F_j) \leq U(B_i) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The interval $[L(B_i), U(B_i)]$ is the value range of the feature F_j for the class B_i. So, the feature value $y_k(F_j)$ of each object y_k belonging to the class B_i corresponding to the feature $F_j \in \mathcal{C}_p(\{B_i, B_r\})$ falls within the interval $[L(B_i), U(B_i)]$, which implies that an equivalence partition is nonempty. The intersection between every two such intervals corresponding to the features of set $\mathcal{C}_p(\{B_i, B_r\})$ may form an implicit hypercuboid, as indicated by the shaded rectangle in Fig. 1. It encloses the misclassified objects that belong to more than one equivalence partitions with respect to the attribute set $\mathcal{C}_p(\{B_i, B_r\})$. The relevance $\Gamma_p(\{B_i, B_r\})$ of the set $\mathcal{C}_p(\{B_i, B_r\})$, with respect to the pair of classes $\{B_i, B_r\}$, depends on the cardinality of the implicit hypercuboid.
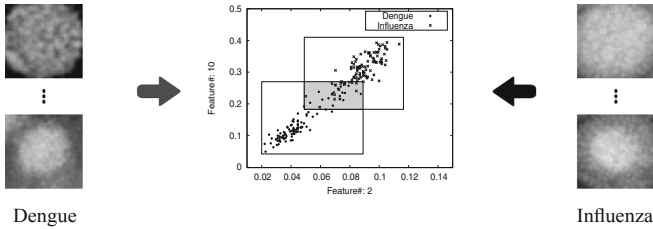

Fig. 1. Example of rough hypercuboids in two dimensions: two class hypercuboids corresponding to upper approximations of Dengue and Influenza virus classes, along with implicit hypercuboid (boundary region) denoted by the shaded region.

So, it can be observed from (2) that the relevance increases with decrease in the cardinality of implicit hypercuboids. If $\Gamma_p(\{B_i, B_r\}) = 1$, then no implicit

hypercuboid is formed and both the classes B_i and B_r can be defined precisely using the knowledge of $\mathcal{C}_p(\{B_i, B_r\})$. On the other hand, if $\Gamma_p(\{B_i, B_r\}) = 0$, then both the classes cannot be defined using the information of $\mathcal{C}_p(\{B_i, B_r\})$. However, if $\Gamma_p(\{B_i, B_r\}) \in (0, 1)$, then B_i and B_r can be approximated using the feature set $\tilde{\mathcal{C}}_{ir}$ for the class-pair $\{B_i, B_r\}$ is formed by considering the relevance value of the feature set $\mathcal{C}_p(\{B_i, B_r\})$, corresponding to each of the t modalities, as follows:

$$\tilde{\mathcal{C}}_{ir} = \underset{\mathcal{C}_p(\{B_i, B_r\})}{\arg\max} \{\Gamma_p(\{B_i, B_r\})\}. \tag{6}$$

Finally, the feature set $\mathcal{D}$, corresponding to all the virus classes, is formed as

$$\mathcal{D} = \bigcup \tilde{\mathcal{C}}_{ir}. \tag{7}$$

The SVM is used to predict the texture patterns present in TEM images of test samples of virus particles, based on the final feature set $\mathcal{D}$ obtained using (7).

3 Performance Analysis

In order to establish the efficacy of the proposed method in classifying the virus particles present in TEM images, extensive experiment is conducted on Virus image database and the corresponding results are presented in this section.

3.1 Algorithms Compared

The proficiency of the proposed method is validated through comparison with several texture descriptors, which are LBP [16], LBP$^{\mathrm{ri}}$ [15], LBP$^{\mathrm{riu2}}$ [15] and CoALBP [14], obtained at different scales. Here, $\mathcal{S}_1$: scale 1, $\mathcal{S}_2$: scale 2, $\mathcal{S}_3$: scale 3, $\mathcal{S}_4$: scale 4, $\mathcal{S}_{123}$: concatenation of $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$, and $\mathcal{S}_{124}$: concatenation of $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_4$ are considered. In the current study, the descriptors along with the corresponding scales, are chosen arbitrarily and therefore, any other sets of descriptors are equally compatible to be used in the proposed descriptor selection method. In case of CoALBP, 4-neighborhood is considered, while 8-neighborhood is considered for the rest.

The performance of several existing approaches for virus classification is analyzed with reference to the proposed method, which include Haralick textural features [5], Fourier RDP (FRDP) [9], dominant LBP (DLBP) [10], discriminative features for texture description (DFTD) [4], LBP+RDP [9], NewH [13], Fusion [13] and Jeffreys/Stein [6]. Here, 10-fold cross-validation is performed to evaluate the classification accuracy of related approaches as well as proposed method. The comparative performance of different algorithms is studied through box-and-whisker plots, tables of means, medians, standard deviations, and p-values computed through both paired-t and Wilcoxon signed-rank tests, with 95% confidence level. In box-and-whisker plots, the central line on each box represents the median, the upper and lower quartiles are depicted by upper and lower boundaries of the box, respectively. The whiskers are drawn from mean to three standard deviations, so that extreme data points are also included. The outliers are plotted individually, denoted as '+'.
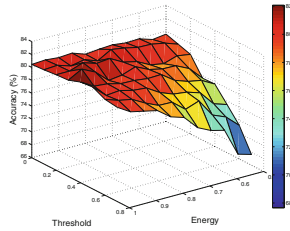
Fig. 2. Effect of energy E and threshold ϵ on classification accuracy.

3.2 Description of Data Set

The effectiveness of the existing approaches as well as proposed method is studied through evaluation on the real-life Virus data set [8]. The samples of the database are imaged using negative stain TEM. The set includes 15 different virus classes, each of which is represented by 100 TEM images. Different viruses exhibit different structural properties. However, the diameter or cross-section remains almost constant within a particular virus class. Usually, the diameter varies from 25 nm to 270 nm depending on the morphology of the viruses. The virus particles, which are presented in the data set, include Dengue, Cowpox, Ebola, Influenza, Adenovirus, Astrovirus, Rotavirus, Norovirus, Crimean-Congo Haemorrhagic Fever, Lassa, Marburg, Orf, Papilloma, Rift Valley, and WestNile.
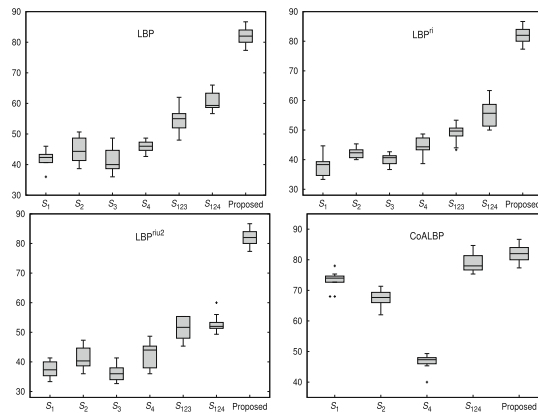

Fig. 3. Performance of different local texture descriptors and proposed method.

3.3 Optimum Values of Different Parameters

In the current study, the feature set $\mathcal{C}(B_i)$, representing the important properties of the class B_i, is obtained for a particular value of Energy E and threshold ϵ.

Table 1. Classification accuracy of various local descriptors and proposed method

Descriptors	Scales	Mean	Median	StdDev	Paired-t:p	Wilcoxon:p
LBP	$\mathcal{S}_1$	42.00	42.33	2.70	2.35E−11	2.52E−03
	$\mathcal{S}_2$	44.60	44.33	4.10	2.75E−10	2.50E−03
	$\mathcal{S}_3$	41.20	40.00	3.91	2.72E−10	2.52E−03
	$\mathcal{S}_4$	45.73	46.00	**1.78**	2.24E−11	2.50E−03
	$\mathcal{S}_{123}$	55.00	55.00	4.21	1.22E−08	2.53E−03
	$\mathcal{S}_{124}$	60.53	59.33	3.20	4.80E−08	2.52E−03
LBP$^{\text{ri}}$	$\mathcal{S}_1$	37.87	38.33	3.45	5.71E−13	2.53E−03
	$\mathcal{S}_2$	42.33	42.33	1.87	2.02E−11	2.52E−03
	$\mathcal{S}_3$	40.00	40.67	2.04	4.12E−11	2.52E−03
	$\mathcal{S}_4$	44.53	44.33	2.96	1.62E−12	2.53E−03
	$\mathcal{S}_{123}$	48.87	49.67	3.19	4.90E−10	2.53E−03
	$\mathcal{S}_{124}$	55.67	55.67	4.36	4.53E−08	2.52E−03
LBP$^{\text{riu2}}$	$\mathcal{S}_1$	37.53	37.33	3.06	4.62E−11	2.53E−03
	$\mathcal{S}_2$	41.13	40.33	3.61	3.86E−10	2.52E−03
	$\mathcal{S}_3$	36.47	36.00	3.19	1.14E−11	2.52E−03
	$\mathcal{S}_4$	42.40	44.00	4.67	2.04E−09	2.53E−03
	$\mathcal{S}_{123}$	51.20	51.67	3.73	9.55E−11	2.46E−03
	$\mathcal{S}_{124}$	52.93	52.00	3.03	1.40E−10	2.50E−03
CoALBP	$\mathcal{S}_1$	73.20	74.00	3.14	1.82E−05	2.52E−03
	$\mathcal{S}_2$	67.47	67.67	2.81	5.54E−09	2.47E−03
	$\mathcal{S}_4$	46.73	47.33	2.64	2.07E−11	2.52E−03
	$\mathcal{S}_{124}$	78.87	78.00	2.92	2.75E−03	7.49E−03
Proposed		**82.07**	**82.00**	2.62		

So, the values of both the parameters have an influence on the performance of the proposed method in classifying the virus particles into multiple classes.

The value of E is varied from 0.50 to 1.00 at an interval of 0.05, while the value of ϵ is varied from 0.00 to 0.70 at an interval of 0.10 and the corresponding classification accuracy of the proposed method is noted in order to obtain the optimum values of the parameters. The SVM classifier with linear kernel is applied on virus TEM images and the effect of the values of E and ϵ on 10-fold classification accuracy is reported in Fig. 2. The results depicted in Fig. 2 demonstrate that the accuracy of the proposed method increases with increase in energy E and decrease in threshold ϵ, and highest classification accuracy is obtained at $E = 0.95$ and $\epsilon = 0.30$, which validates the concept of energy and threshold defined in the current study. While low energy provides restricted representation of the important feature sets of samples, high energy signifies more information captured from the images, leading to a descriptive representation of the sets. The features, which have occurred in the important feature sets

inadvertently with low probability of occurrence, are removed by incorporating the threshold parameter ϵ in the algorithm. So, for the proposed approach, the values of parameters E and ϵ are fixed at 0.95 and 0.30. Hence, a precise representation of the set $\mathcal{C}(B_i)$, reflecting the characteristics of the class B_i, is obtained with the features which preserves 95% of total energy by at least 30% samples of each B_i.

3.4   Comparison with Existing Approaches

In general, a specific descriptor, obtained at a particular scale, may be used to describe the textural properties of all the TEM images present in the Virus data set. However, the proposed method identifies class-pair specific modality to capture the inherent characteristics of each of the virus classes.

In order to validate the significance of class-pair relevant modalities over uniform modalities, extensive experiment is carried out on Virus data set, considering fifteen modalities corresponding to four local descriptors LBP, LBP$^{\mathrm{ri}}$, LBP$^{\mathrm{riu2}}$ and CoALBP, evaluated at both individual and concatenated scales. Figure 3 and Table 1 report the classification accuracy achieved by the proposed method as well as various local texture descriptors on the samples of Virus database. The results presented in Fig. 3 and Table 1 reveal that highest mean and median values are achieved by the proposed method, irrespective of descriptors and scales considered. Also, statistical significance analysis reveals that the proposed method attains significantly lower p-values in all the 44 cases. In this regard, it is to be mentioned that the proposed method attains 82.07% accuracy on Virus database with 2327 number of features, selected from the initial feature sets of fifteen modalities corresponding to the four local descriptors, evaluated at four different scales. If the feature sets of all the modalities are concatenated, then 80% accuracy can be achieved with 4280 features. Thus, the proposed method can efficiently identify class-pair specific modalities to describe each pair of classes with lesser number of features.

Finally, the performance of several existing approaches for virus classification is analyzed with reference to the proposed method, which include Haralick textural features [5], FRDP [9], DLBP [10], DFTD [4], LBP+RDP [9], NewH [13], Fusion [13] and Jeffreys/Stein [6]. In these methods, important information is captured from the TEM images to categorize the virus samples into one of the known fifteen virus classes. The performance of the proposed approach with reference to different existing methods is analyzed in Fig. 4. It is evident from Fig. 4 that the methods corresponding to Haralick features, DLBP, DFTD and FRDP exhibit poor performance in recognizing viruses from TEM images. On the other hand, the methods like LBP+RDP, NewH, Fusion, and Jeffreys/Stein show improvement in the performance. However, the proposed method attains highest classification accuracy on Virus data set with respect to all the existing methods.
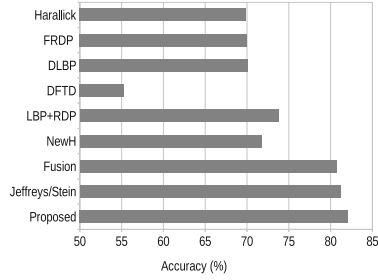
Fig. 4. Analysis of performance of the proposed approach and existing methods.

4 Conclusion

In the current study, a new method is developed for automatic recognition of virus particles from negative stain TEM images. The main contribution of the paper lies in considering relevant modality corresponding to each pair of classes present in Virus database, rather than considering uniform modalities for all the classes. The relevance of each modality is computed based on hypercuboid equivalence partition matrix of rough hypercuboid approach. A subset of important features is selected for each of the relevant modalities by reducing the impact of both noisy pixels present in a virus image as well as noisy virus images present in a particular virus class. The proficiency of the proposed approach with respect to different existing method is established on real-life Virus image data set.

References

1. Almeida, J.D.: A classification of virus particles based on morphology. Can. Med. Assoc. J. **89**(16), 787–798 (1963)
2. Almeida, J.D., Waterson, A.P.: Some implications of a morphologically oriented classification of viruses. Arch. Virol. **32**(1), 66–72 (1970)
3. Goldsmith, C.S., Miller, S.E.: Modern uses of electron microscopy for detection of viruses. Clin. Microbiol. Rev. **22**(4), 552–563 (2009)
4. Guo, Y., Zhao, G., Pietikäinen, M.: Discriminative features for texture description. Pattern Recogn. **45**(10), 3834–3843 (2012)
5. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3**(6), 610–621 (1973)
6. Harandi, M., Salzmann, M., Porikli, F.: Bregman divergences for infinite dimensional covariance matrices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1003–1010 (2014)

7. Ito, E., Sato, T., Sano, D., Utagawa, E., Kato, T.: Virus particle detection by convolutional neural network in transmission electron microscopy images. Food Environ. Virol. **10**(2), 201–208 (2018)

8. Kylberg, G., Uppstrom, M., Borgefors, G., Sintorn, I.-M.: Segmentation of virus particle candidates in transmission electron microscopy images. J. Microsc. **245**, 140–147 (2012)

9. Kylberg, G., Uppström, M., Sintorn, I.-M.: Virus texture analysis using local binary patterns and radial density profiles. In: San Martin, C., Kim, S.-W. (eds.) CIARP 2011. LNCS, vol. 7042, pp. 573–580. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25085-9_68

10. Liao, S., Law, M.W.K., Chung, A.C.S.: Dominant local binary patterns for texture classification. IEEE Trans. Image Process. **18**(5), 1107–1118 (2009)

11. Maji, P.: A rough hypercuboid approach for feature selection in approximation spaces. IEEE Trans. Knowl. Data Eng. **26**(1), 16–29 (2014)

12. Matuszewski, B.J., Shark, L.-K.: Hierarchical iterative Bayesian approach to automatic recognition of biological viruses in electron microscope images. In: Proceedings of the International Conference on Image Processing, pp. 347–350 (2001)

13. Nanni, L., Paci, M., Brahnam, S., Ghidoni, S., Menegatti, E.: Virus image classification using different texture descriptors. In: Proceedings of the International Conference on Bioinformatics and Computational Biology, Las Vegas, NV (2013)

14. Nosaka, R., Ohkawa, Y., Fukui, K.: Feature extraction based on co-occurrence of adjacent local binary patterns. In: Ho, Y.-S. (ed.) PSIVT 2011. LNCS, vol. 7088, pp. 82–91. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25346-1_8

15. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)

16. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recogn. **29**(1), 51–59 (1996)

17. Ong, H.C.L.: Virus recognition in electron microscope images using higher order spectral features. Ph.D. thesis, Queensland University of Technology (2006)

18. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991). ISBN 0792314727

19. Sintorn, I.-M., Homman-Loudiyi, M., Söderberg-Nauclér, C., Borgefors, G.: A refined circular template matching method for classification of human cytomegalovirus capsids in TEM images. Comput. Methods Programs Biomed. **76**(2), 95–102 (2004)