Automatic RNA virus classification using the Entropy-ANFIS method

Esin Dogantekin [a], Engin Avci [b,*], Oznur Erkus [c]

[a] Zirve University, Emine Bahaeddin Nakiboglu Medical Faculty, Department of Microbiology, 27260 Gaziantep, Turkey
[b] Firat University, Software Engineering, 23119, Elazig, Turkey
[c] Republic Anatolia High School, Elazig, Turkey

ARTICLE INFO

ABSTRACT

Innovations in the fields of medicine and medical image processing are becoming increasingly important. Historically, RNA viruses produced in cell cultures have been identified using electron microscopy, in which virus identification is performed by eye. Such an approach is time consuming and depends on manual controls. Moreover, detailed knowledge about RNA viruses is required. This study introduces the Entropy-Adaptive Network Based Fuzzy Inference System (Entropy-ANFIS method), which can be used to automatically detect RNA virus images. This system consists of four stages: pre-processing, feature extraction, classification and testing the Entropy-ANFIS method with respect to the correct classification ratio. In the pre-processing stage, a center-edge changing method is used, in which the Euclidian distances are calculated from the center pixels to the edges of the imaged object. In this way, the distance vector is obtained. This calculation is repeated for each RNA virus image. In feature extraction, stage norm entropy, logarithmic energy and threshold entropy values are calculated to form the feature vector. The obtained feature vector is independent of the rotation and scale of the RNA virus image. In the classification stage, the feature vector is given as input to the ANFIS classifier, ANN classifier and FCM cluster. Finally, the test stage is performed to evaluate the correct classification ratio of the Entropy-ANFIS algorithm for the RNA virus images. The correct classification ratio has been determined as 95.12% using the proposed Entropy-ANFIS method.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Technological developments have progressed rapidly in recent years. Image-processing techniques are widely used in the medical field due to their advantages. Feature extraction using image-processing techniques to identify RNA viruses is advantageous for the following reasons:

- To facilitate diagnosis of diseases caused by RNA viruses;
- To minimize the time spent on disease diagnosis;
- To determine the type of RNA viruses that cause disease;
- To minimize the financial costs of ordering diagnostics;
- To minimize recognition errors.

Many studies exist in the medical identification literature that use image processing. Sammouda et al. [1] used color imaging of pathological lung cancer to study cell nucleus determination. In Ref. [2], image-processing techniques were applied to large intestine cancer and diagnostics. Such image-processing techniques have never previously been used for the automatic identification of RNA viruses. The main purpose of this study is to use effective image-processing methods to automatically classify RNA viruses because, as shown by the medical imaging literature, image-processing techniques can give accurate and fast results.

Viruses can be separated according to their life cycles, disease groups and genome makeup (RNA/DNA). Molecules of RNA virus are executives. RNA viruses can replicate in animal and plant cells. For example, the viruses that cause tobacco mosaic disease, influenza, polio, measles, rabies, mumps, and yellow fever all have RNA genomes.

The enzyme inside the virion responsible for replication is RNA polymerase. Parental RNA serves as mRNA. The virus does not enter cell protein synthesis. The virus first makes the viral RNA polyribosome, which is a fairly large protein molecule. This protein is then cleaved to create functional proteases with small pieces. New negative strands are synthesized during positive strand replication [2–4].

Each strand with a negative polarity is located inside the virion RNA and is translated into a positive strand by polymerase enzymes. The same negative strand can lead to more than one positive strand synthesis [2–4].

Several types of RNA viruses found in the literature are summarized in the following paragraphs. Picornaviruses are approximately

25–30 nm in diameter, single-stranded, enveloped, and have positive polarity RNA and an icosahedral protein sheath. This family of viruses causes diseases such as aseptic meningitis, paralytic poliomyelitis, myopericarditis, RAS, acute hepatitis and the common cold.

Togaviruses are approximately 60–70 nm in diameter, single-stranded, enveloped, and have positive polarity RNA and an icosahedral protein sheath. This family includes the Alphavirus and Rubivirus subfamilies. This family of viruses causes diseases such as mosquito-borne encephalitis, rubella, mild rash and congenital fetal defects.

Flaviviruses are approximately 45–55 nm in diameter, single-stranded, enveloped, and have positive polarity RNA and an icosahedral protein sheath. These viruses replicate in the cytoplasm. This family includes the Flavivirus and Hepacivirus subfamilies and causes diseases such as mosquito-borne fever, hepatitis, yellow fever, mosquito-borne hemorrhagic fever, Dengue fever, encephalitis and hepatitis C (acute and chronic) and liver cancer.

Orthomyxoviruses are approximately 80–120 nm in diameter, enveloped, have 8 segments of single-stranded negative polarity RNA, and a helical protein sheath. One subfamily of this family is called Influenza viruses. This family of viruses causes influenza, fever, myalgia, fatigue, coughing and pneumonia.

Bunyaviruses are approximately 100 nm in diameter, enveloped and have a helical protein sheath. This family of viruses causes pulmonary diseases, rodent-borne diseases, hemorrhagic fever with renal syndrome (HUS), mosquito-borne encephalitis, and hemorrhagic fever.

Astroviruses are approximately 110–130 nm in diameter, enveloped and have a helical protein sheath. Bergel ribosomes are located in the viral envelope. The most important human pathogen in this family is the astrovirus. This family of viruses causes gastroenteritis diseases.

Retroviruses are approximately 90–120 nm in diameter, enveloped and have single-stranded RNA genomes. Oncovirinae is part of this family, as is the Lentivirinae subfamily, which includes Spumavirinae. This family contains important human pathogens, including Human T Lenfotropik viruses (HTLV) and AIDS Human Immunodeficiency Virus (HIV). The most important feature of this virus is the existence of the reverse transcriptase enzyme, the RNA-dependent DNA polymerase.

Coronaviruses are approximately 120 nm in diameter. They are single-stranded and contain positive polarity RNA (mRNA). Peplomers on the envelope can reach lengths of 20 nm. These peplomers are visible and lead to the illusion of a crown on the virus. This family of viruses causes mild infections [4].

Rotaviruses have a diameter of 60–80 nm. The center contains a two-layer capsid. Rotaviruses have a three-layer structure. Rotavirus envelopes are made of 132 capsomeres. The fully infectious virus is formed by double-layered particles. Reoviruses are double-stranded, partial RNA viruses. These viruses cause the death of an estimated 3.5 million preschool children per year in developing countries and acute gastroenteritis in the United States [5–10].

The rhabdoviridae are 180×75 nm in size. This family includes Vesiculovirus and Lyssavirus. The most important members of this family are Rabies virus, Marburg virus and Ebolavirus. This family of viruses causes rabies and hemorrhagic fever diseases [11].

Filoviridae family viruses are 80 nm in diameter. The helical proteins are sheathed. This family comprises important human pathogens. This family of viruses causes hemorrhagic fever disease.

Paramyxoviridae family viruses are 150 nm in diameter. The RNA genomes are single-stranded and have a negative polarity. This family includes the Paramyxovirus, Morbillivirus and Pneumovirus subfamilies. The most important viruses in this family are human parainfluenza virus type 1, 2, 3, 4, 5, mumps, and the measles virus. This family of viruses causes diseases such as bron-

chiolitis, pneumonia (4 types), mumps, aseptic meningitis (rarely orchitis and encephalitis), measles, fever, rash, pneumonia (rarely, encephalitis, SSPE), colds (adult) and pneumonia (YD).

Until now, RNA viruses have been produced in cell culture and detected by electron microscopy, with identification decisions made by eye. With this approach, the error probability rate of specialists is very high, as shown by some studies in microbiology pattern recognition literature [1–9,32]. Identifying viruses takes time in a lab environment. Moreover, expert knowledge, experience and detailed information are required. In this study, Entropy-Adaptive Network Based Fuzzy Inference System (Entropy-ANFIS method) is introduced for automatic identification of RNA viruses by imaging.

2. Adaptive Network Based on Fuzzy Inference System Classifier

The classifier structure of the Adaptive Network Based on Fuzzy Inference System (ANFIS) uses both artificial neural network and fuzzy logic [12–14]. The ANFIS classifier forms if–then rules, couples of input–output and learning algorithms of neural network. These are used to train the ANFIS classifier [15–18].

For example, the ANFIS classifier has 7 inputs ($x_1, x_2, x_3, x_4, x_5, x_6, x_7$) and 1 output ($y$).

For a first-order Sugeno fuzzy model, a typical rule set with base fuzzy if–then rules can be expressed as:

If x_1 A_1 and x_2 B_1 and x_3 C_1 and x_4 D_1 and x_5 E_1 and x_6 F_1 and x_7 G_1 then

$$f_1 = px_1 + ppx_2 + qx_3 + qqx_4 + sx_5 + ssx_6 + rx_7 + u \qquad (1)$$

where p, pp, q, qq, s, ss, r, and u are linear output parameters. The structure of this ANFIS classifier is formed using 5 layers and 256 if–then rules:

Layer-1: Every node i in this layer is a square node with a node function.

$$O_{1,i} = \mu_{A_i}(x_1), \quad \text{for } i = 1, 2,$$
$$O_{1,i} = \mu_{B_i}(x_2), \quad \text{for } i = 3, 4,$$
$$O_{1,i} = \mu_{C_i}(x_3), \quad \text{for } i = 5, 6,$$
$$O_{1,i} = \mu_{D_i}(x_4), \quad \text{for } i = 7, 8,$$
$$O_{1,i} = \mu_{E_i}(x_5), \quad \text{for } i = 9, 10,$$
$$O_{1,i} = \mu_{F_i}(x_6), \quad \text{for } i = 11, 12,$$
$$O_{1,i} = \mu_{G_i}(x_7), \quad \text{for } i = 13, 14, \qquad (2)$$

where $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ are inputs to node i and A_i, B_i, C_i, D_i, E_i, F_i and G_i are linguistic labels associated with this node function. In other words, $O_{1,i}$ is the membership function of A_i, B_i, C_i, D_i, E_i, F_i and G_i. Usually, $\mu_{A_i}(x_1)$, $\mu_{B_i}(x_2)$, $\mu_{C_i}(x_3)$, $\mu_{D_i}(x_4)$, $\mu_{E_i}(x_5)$, $\mu_{F_i}(x_6)$ and $\mu_{G_i}(x_7)$ are chosen to be bell-shaped with maxima equal to 1 and minima equal to 0, such as

$$\mu_i(x_i) = \exp\big(\big(-(x_i - c_i)/(a_i)\big)^2\big) \qquad (3)$$

where a_i, c_i indicate the parameter sets. The parameters in this layer are referred to as premise parameters.

Layer-2: Every node in this layer is a circle node labeled Π that multiplies the incoming signals and sends out the product. For instance,

$$O_{2,i} = w_i$$
$$= \mu_{A_i}(x_1).\mu_{B_i}(x_2).\mu_{C_i}(x_3).\mu_{D_i}(x_4).\mu_{E_i}(x_5).\mu_{F_i}(x_6).\mu_{G_i}(x_7)$$
$$i = 1, 2, 3, \ldots, 256. \qquad (4)$$

Each node output represents the firing strength of a rule. (In fact, other T-norm operators performing generalized functions AND can be used as the node function in this layer.)

Layer-3: Every node in this layer is a circle node labeled N. The ith node calculates the ratio of the ith rule's firing strength to the sum of the firing strengths of all of the rules:

$$O_{3,i} = \bar{w}_i = w_i/(w_1 + w_2 + \cdots + w_{256}),$$
$$i = 1, 2, 3, \ldots, 256 \tag{5}$$

Layer-4: Every node i in this layer is a square node with a node function

$$O_{4,i} = \bar{w}_i.f_i$$
$$= w_i.(px_1 + ppx_2 + qx_3 + qqx_4 + sx_5 + ssx_6 + rx_7 + u),$$
$$i = 1, 2, 3, \ldots, 256 \tag{6}$$

where w_i is the output of layer 3, and $\{p_i, pp_i, q_i, qq_i, s_i, ss_i, r_i, u_i\}$ is the parameter set. Parameters in this layer are referred to as consequent parameters.

Layer-5: The single node in this layer is a circle node labeled $\sum$ that computes the overall output as the summation of all incoming signals:

$$O_{5,i} = \text{overall output} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \tag{7}$$

3. Fuzzy c-means clustering algorithm

Fuzzy c-means (FCM) is a method of clustering that allows one piece of data to belong to two or more clusters [19]. This method (developed by Dunn and improved by Bezdek) is frequently used in pattern recognition. It is based on the minimization of the following objective function:

$$Tm = \sum_{i=1}^{M} \sum_{j=1}^{K} u_{ij}^m \|x_i - l_j\|^2, \quad 1 \leqslant m < \infty \tag{8}$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, l_j is the d-dimensional center of the cluster, and $\| * \|$ is any norm expressing similarity between any measured data and the center. Fuzzy partitioning is performed through an iterative optimization of the objective function shown above in which the membership u_{ij} and cluster centers l_j are updated, thus:

$$u_{ij} = 1 / \sum_{k=1}^{K} \left(\frac{\|x_i - l_j\|}{\|x_i - l_k\|} \right), \qquad l_j = \frac{\sum_{i=1}^{M} u_{ij}^m.x_i}{\sum_{i=1}^{M} u_{ij}^m}. \tag{9}$$

FCM is a clustering method. It permits one piece of data to belong to two or more clusters [19]. This FCM method is commonly used in pattern recognition. In this paper, the application of this FCM version can be given as below [19]:

---
1. *Matrix $U = [u_{ij}]$ is initialized.*
2. *The center vectors $L(k) = [lj]$ with $U(k)$ are calculated, and $U(k)$ is updated to $U(k + 1)$*

$$u_{ij} = 1 / \sum_{k=1}^{K} \left(\frac{\|x_i - l_j\|}{\|x_i - l_k\|} \right), \qquad l_j = \frac{\sum_{i=1}^{M} u_{ij}^m.x_i}{\sum_{i=1}^{M} u_{ij}^m}$$

3. *If $\|U(k + 1) - U(k)\| < \varepsilon$ then STOP; else return to step 2.*

In this study, the FCM algorithm mentioned above was used to cluster the RNA virus images.

More information about the FCM clustering method can be found in Ref. [19].

4. The automatic classification method of RNA viruses using ANFIS

The RNA virus images used in this study were taken from the web site of the University of Pompeu Fabra, Division of Biology [3]. The images can be used for educational and non-commercial purposes without permission, as noted on the related web site [3].

The following 12 typical RNA virus images were chosen for use in this study:

1. Picornaviruses (Pico),
2. Togaviruses (Toga),
3. Flaviviruses (Flavi),
4. Orthomyxoviruses (Orthom),
5. Bunyaviruses (Bunya),
6. Astroviruses (Astro),
7. Retroviruses (Retro),
8. Coronaviruses (Corona),
9. Reoviruses (Reo),
10. Rhabdoviruses (Rhabdo),
11. Filoviruses (Filo),
12. Paramyxoviruses (Param).

Using ANFIS to automatically classify RNA viruses involves many stages, including pre-processing, feature extraction, classification and testing. In the first stage, the RNA virus images used in this study are pre-processed.

The pre-processing stage uses many digital image processes. The first process is to obtain the digital RNA virus images used in this study via image processing. A sensor and sensor output must be digitized in the mark to do this. In the pre-processing stage, the processes are performed to obtain better results in the RNA virus images. This pre-processing stage contains the contrast, expansion, and noise elimination processes.
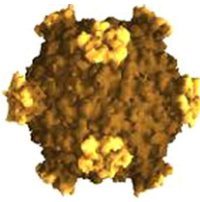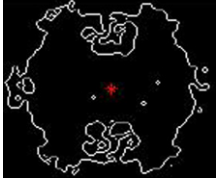
The autonomous segmentation of digital RNA virus image processing is one of the most difficult processes. The segmentation process outputs raw data, which can be processed using a computer. At this point, the data should indicate whether the RNA virus can be identified [20–22]. There are many techniques for pre-processing stages, including the following:

1. RNA virus images are translated from color into gray images. The RNA virus images are thus converted into gray-level images.
2. The gray-level image histograms of these gray-level RNA virus images are obtained.
3. Threshold values are determined utilizing the image histogram for each of these RNA virus images.
4. This value is determined using a value above the threshold value for which the pixels in the image output value in the bottom 1 have been assigned a value of 0.
5. The picture is thus separated from the background. We applied the Canny edge extraction algorithm to separate the background image. With this approach, the edges of objects in these images are determined.
6. Finally, the center-edge changing method is applied to the RNA virus images. The center-edge changing method can be defined as a drawing. In the function, r is the shape, and this shape of the center of gravity is located on the border between any points that represent the Euclidian distance. This calculation of the Euclidian distance process is also expressed as Eq. (10) below [21–25]:

$$r = \sqrt{(x - x_m)^2 + (y - y_m)^2}. \tag{10}$$

Table 1
Some of the calculated entropy values of RNA viruses image.

Before pre-processing	After pre-processing	The norm entropy value (E1)	The logaritmic energy entropy value (E2)	The threshold entropy value (E3)
CORONAVIRUSES		119.3178	4.0817e+003	439
REOVIRUSES		509.4558	1.0894e+004	960
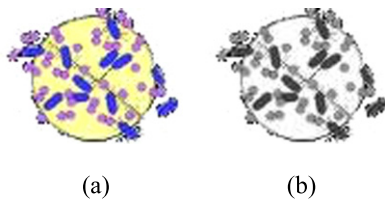
Fig. 1. RNA virus images. (a) RNA viruses (Rhabdovirus), (b) the gray image converted into a virus state.
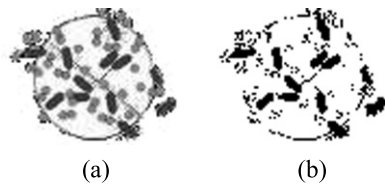
Fig. 2. RNA virus images. (a) RNA virus (Rhabdovirus), (b) the segmented viruses.

Fig. 3. RNA virus images. (a) RNA virus (Rhabdovirus), (b) Canny edge detection process after the image is extracted.

More information about the center-edge changing method can be found in previous research [21–25].

The processes of these morphological and logical operations for RNA viruses are shown in Figs. 1–3.

In the feature extraction stage, the pre-processed RNA virus images from the first stage (i.e., the pre-processing stage) are used. In this feature extraction stage, each of the RNA virus images is rotated 15°, between 0° and 165°. Therefore, 12 rotated images at different angles are obtained from each of the RNA virus images.

These rotated RNA virus images are then scaled to 10 different sizes. The number of obtained total RNA virus images is thus $12 \times 12 \times 10 = 1440$. Moreover, three different entropy values, i.e., the norm, logarithmic energy and entropy threshold values, are calculated for each of these 1440 RNA virus images (Table 1).

Table 2
ANFIS structure and training parameters used in this study.

The number of layers	5 Input: 3 Rules number: 8 Output: 1
Type of input membership functions	Bell-shaped
Learning rule	Hybrid learning algorithm
Sum-squared error	0.0000000001
Reaching epochs number to sum-squared error	834

Entropy-based features give the information about signal definitions. The concept of entropy involves a system that is used to measure the regularity of thermodynamics and is a well-known concept in physics. The entropy measurement method measures the degree of the disorder of any sign [21].

In recent years, entropy as a concept has been widely used in the field of signal processing. The entropy-type equations that have been widely used in signal processing can be found in previous studies [21–25].

The size of the obtained feature vector at the final feature extraction stage is 1440×3. Half of this 1440×3 feature vector is used for the classification stage, i.e., a 720×3 $(12 \times 60 \times 3)$ feature vector is given as input for the ANFIS classifier and FCM cluster. The remainder of the 1440×3 feature vector is used in the testing stage to determine the correct classification and clustering performance of the Entropy-ANFIS and Entropy-FCM methods used in this study.

The structure of the Entropy-ANFIS and Entropy-FCM algorithms used in this study is given in Fig. 4.

In the classification and cluster stage, half of the 1440×3 $(12 \times 120 \times 3)$ feature vector obtained in the feature extraction stage is used for classification. This feature vector is used as input for the ANFIS classifier, ANN classifier and FCM cluster. The rest of this 1440×3 $(12 \times 120 \times 3)$ feature vector obtained in the feature extraction stage is used to test the correct classification and cluster ratios of ANFIS classifier, ANN classifier and FCM cluster in the testing stage. The training parameters of ANFIS classifier used in this study are given in Table 2.
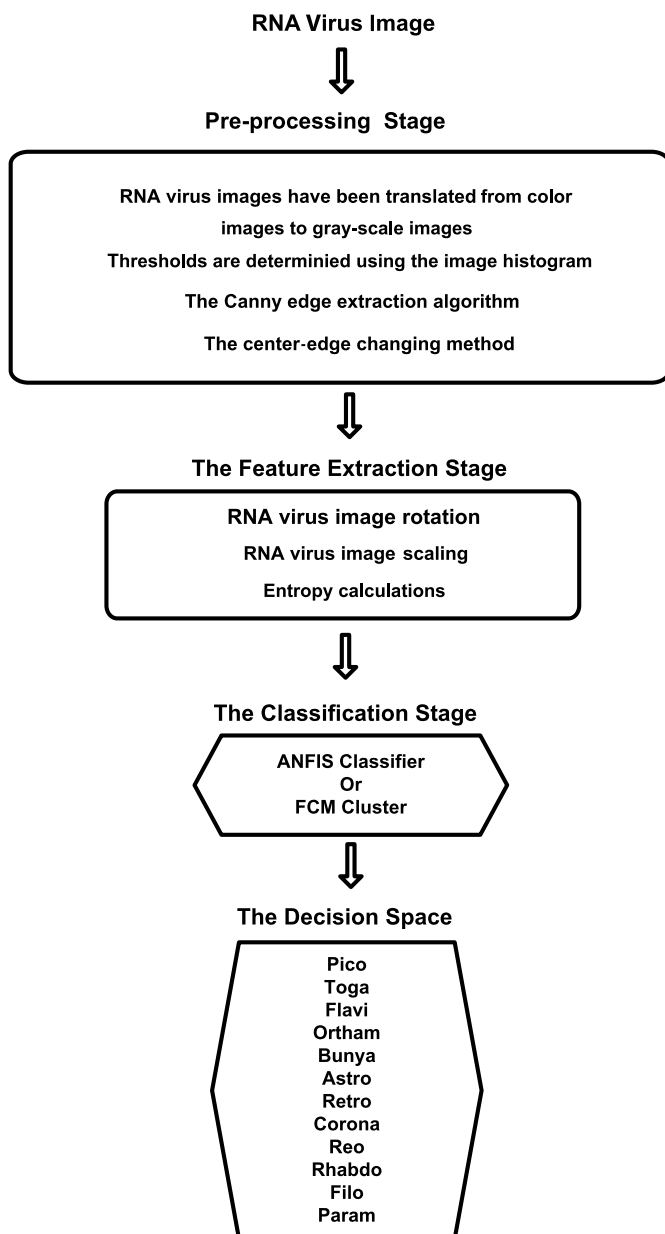
5. The obtained results

The testing stage is conducted to obtain the applied results of Entropy-ANFIS, Entropy-ANN and Entropy-FCM algorithms for RNA

Table 3
The correct classification rates for the proposed Entropy-ANFIS approach.

RNA virus type	Number of correctly classified RNA viruses	Number of incorrectly classified RNA viruses	Percentage of correct classification rate (%)
Pico	57	3	95
Toga	58	2	96.66
Flavi	57	3	95
Orthom	60	0	100
Bunya	58	2	95
Astro	56	4	93.33
Retro	60	0	100
Corona	59	1	98.33
Reo	57	3	95
Rhabdo	55	5	91.67
Filo	50	10	83.33
Param	59	1	98.33
Total	**686**	**34**	**95.12**

Fig. 4. The structure of the Entropy-ANFIS and Entropy-FCM algorithms used in this study.

virus classification. In the testing stage, the performance of the Entropy-ANFIS, Entropy-ANN and Entropy-FCM methods for automatic RNA virus classification is tested.

For this stage, the rest of the 1440×3 ($12 \times 120 \times 3$) feature vector obtained in the feature extraction stage is used. Thus, a 720×3 ($12 \times 60 \times 3$) feature vector is used in this stage. In this experimental study, 2-fold cross-validation schema was applied, in which 720×3 samples were used to train the Entropy-ANFIS, Entropy-ANN and Entropy-FCM algorithms; the remaining 720×3 samples were used as the test dataset.

The obtained average classification accuracy of the proposed Entropy-ANFIS algorithm is given in Table 3.

The obtained confusion matrix for the proposed Entropy-ANFIS algorithm is given in Table 4.

The obtained clustering accuracy of the Entropy-FCM algorithm using the same feature vector is given in Table 5.

The obtained confusion matrix for the Entropy-FCM algorithm is given in Table 6.

The training parameters for the multi-layer artificial neural network classifier used in this study are given in Table 7.

Several values, such as the number of hidden layers, the number of cells in hidden layers, the learning rates and the value of the activation function for ANN, have been selected to obtain the best performance after several trials.

6. Conclusions and discussion

This study introduces the Entropy-Adaptive Network Based Fuzzy Inference System (Entropy-ANFIS method), which can be used to automatically identify RNA virus images. In the literature, there are many pattern recognition studies [26–32]. In these studies, many different feature extraction methods were used.

In this study, the center-edge changing and the calculation of entropy methods are used for feature extraction. The most significant advantage of this method is its invariance under rotation and scaling operations. Therefore, RNA viruses can be detected from their images even if the images are rotated or scaled.

In this study, the computer simulations in a MATLAB environment are designed to complete the system and test the performance of the proposed automatic RNA virus classification using the Entropy-ANFIS method. As shown in Tables 3–8, the developed system classifies RNA viruses with a high success rate, even after the rotation and scaling operations. The correct classification performance of the proposed Entropy-ANFIS method is 95.12%. The correct clustering performances of the Entropy-ANN and the Entropy-FCM are 94.02% and 89.44%, respectively. Therefore, the proposed Entropy-ANFIS method performs better than the Entropy-ANN and Entropy-FCM methods.

RNA viruses are classified incorrectly mainly due to their similar shapes. Some RNA virus images are of uncertain identity because

Table 4
The obtained confusion matrix of the proposed Entropy-ANFIS algorithm.

	Pico	Toga	Flavi	Ortho	Bun	Astro	Retro	Coro	Reo	Rhabd	Filo	Param
Pico	57	–	–	1	–	–	–	2	–	–	–	–
Toga	–	58	–	–	–	–	1	–	–	–	–	1
Flavi	–	–	57	–	–	2	–	–	1	–	–	–
Ortho	–	–	–	60	–	–	–	–	–	–	–	–
Bun	–	–	–	–	58	–	–	2	–	–	–	–
Astro	–	–	2	–	–	56	–	–	1	–	–	1
Retro	–	–	–	–	–	–	60	–	–	–	–	–
Coro	–	–	1	–	–	–	–	59	–	–	–	–
Reo	–	–	–	2	–	–	–	–	57	–	1	–
Rhabd	2	1	–	–	2	–	–	–	–	55	–	–
Filo	–	3	–	1	–	2	1	–	2	1	50	–
Param	–	–	–	–	–	–	–	–	–	1	–	59

Table 5
The correct clustering rates for the Entropy-FCM approach.

RNA virus types	Number of correctly clustering RNA viruses	Number of incorrectly clustering RNA viruses	Percentage of correct clustering rate (%)
Pico	52	8	86.66
Toga	55	5	91.67
Flavi	54	6	90
Orthom	57	3	95
Bunya	50	10	83.33
Astro	52	8	86.66
Retro	60	0	100
Corona	49	11	81.66
Reo	51	9	85
Rhabdo	50	10	83.33
Filo	54	6	90
Param	60	0	100
Total	**644**	**76**	**89.44**

Table 6
The obtained confusion matrix of the Entropy-FCM algorithm.

	Pico	Toga	Flavi	Ortho	Bun	Astro	Retro	Coro	Reo	Rhabd	Filo	Param
Pico	52	–	2	1	–	3	–	–	2	–	–	–
Toga	–	55	–	–	3	–	–	2	–	–	–	–
Flavi	–	–	54	–	4	–	–	–	–	–	2	–
Ortho	1	–	2	57	–	–	–	–	–	–	–	–
Bun	–	4	–	–	50	–	5	–	–	–	1	–
Astro	–	–	–	3	–	52	–	3	–	–	2	–
Retro	–	–	–	–	–	–	60	–	–	–	–	–
Coro	–	–	–	4	–	3	–	49	1	3	–	–
Reo	–	3	–	–	–	–	4	–	51	2	–	–
Rhabd	3	–	–	–	2	–	–	–	–	50	–	5
Filo	–	–	3	–	–	–	–	3	–	–	54	–
Param	–	–	–	–	–	–	–	–	–	–	–	60

Table 7
Multi-layer neural network structure and training parameters.

Number of layers	3
The number of layer neurons	Input: 3, Hidden layer: 20, Output: 12
Initiate weights and biases	The Nguyen–Widrow method
Activation functions	Log-sigmoid
Training parameters	Back propagation
Learning rule	
Mean square error	0.00001

they have similar shapes. As shown in these results, the pre-processing stage is the most important part of the Entropy-ANFIS method for correctly identifying RNA virus images.

References

[1] http://tr.wikipedia.org/wiki/Retroviridae, last accessed: 17 March 2009.
[2] http://tr.wikipedia.org/wiki/Vir%C3%BCs, last accessed: 12 March 2009.
[3] http://complex.upf.es/~ricard/RNA_SITE.html, last accessed: 29 January 2010.
[4] http://www.webturkiyeportal.com/webforum/113459-rna-virusleri.html, last accessed: 12 March 2009.
[5] http://www.apath.com/Virus_Detection.htm, last accessed: 29 January 2010.
[6] http://yunus.hacettepe.edu.tr/~coner/VIR/1/coronaviridae.htm, last accessed: 17 March 2009.
[7] http://web.inonu.edu.tr/~bdurmaz/Reovirusrota.htm, last accessed: 19 March 2009.
[8] http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01567612, last accessed: 24 April 2009.
[9] http://akizilkaya.pamukkale.edu.tr/B%C3%B6l%C3%BCm4_goruntu_isleme.pdf, last accessed: 24 April 2009.
[10] http://www.teknohaber.net/makale.php?id=50801, last accessed: 23 March 2009.
[11] http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1567656, last accessed: 2 April 2009.
[12] E. Avci, I. Turkoglu, Modelling of tunnel diode by adaptive-network-based fuzzy inference system, Int. J. Comput. Intell. 1 (1) (2003) 231–233.
[13] B. Kosko, Neural Networks and Fuzzy Systems, A Dynamical Systems Approach, Prentice Hall, Englewood Cliffs, NJ, 1991.
[14] J.S.R. Jang, C.T. Sun, Neuro-fuzzy modeling and control, Proc. IEEE 3 (3) (1995).
[15] J.S.R. Jang, ANFIS: Adaptive network based fuzzy inference systems, IEEE Trans. Syst. Man Cybern. 23 (May 1993) 665–685.

Table 8
The obtained confusion matrix for the Entropy-ANN algorithm.

	Pico	Toga	Flavi	Ortho	Bun	Astro	Retro	Coro	Reo	Rhabd	Filo	Param
Pico	58	–	1	1	–	–	–	–	–	–	–	–
Toga	–	57	–	–	1	–	1	1	–	–	–	–
Flavi	1	1	57	–	–	1	–	–	–	–	–	–
Ortho	–	–	–	60	–	–	–	–	–	–	–	–
Bun	–	2	–	–	55	–	–	–	2	–	1	–
Astro	–	–	–	2	–	54	–	1	1	1	–	1
Retro	–	–	–	1	–	–	59	–	–	–	–	–
Coro	–	–	–	–	–	–	–	60	–	–	–	–
Reo	–	1	–	–	–	1	1	–	57	–	–	–
Rhabd	3	–	–	–	2	–	–	–	–	50	–	5
Filo	1	–	2	2	–	1	–	1	1	–	51	1
Param	–	–	–	–	–	–	1	–	–	–	–	59

[16] E. Avci, I. Turkoglu, M. Poyraz, Intelligent target recognition based on wavelet adaptive network based fuzzy inference system, in: Lecture Notes in Comput. Sci., vol. 3522, Springer-Verlag, 2005, pp. 594–601.

[17] E. Comak, A. Arslan, I. Turkoglu, A decision support system based on support vector machines for diagnosis of the heart valve diseases, Comput. Biol. Med. 37 (1) (January 2006) 21–27.

[18] E. Avci, Z.H. ve Akpolat, Speech recognition using a wavelet packet adaptive network based fuzzy inference system, Expert Syst. Appl. 31 (3) (2006) 495–503.

[19] http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html, last accessed: 15 July 2010.

[20] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison–Wesley Publishing Company, 1993.

[21] E. Avci, D. Avci, A novel approach for digital radio signal classification: Wavelet packet energy-multiclass support vector machine (WPE–MSVM), Expert Syst. Appl. 34 (3) (2008) 2140–2147.

[22] A.D. Kulkarni, Computer Vision and Fuzzy-Neural Systems, Prentice Hall PTR, USA, 2001, 509 pp.

[23] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.

[24] S. Tonga, A. Bezerianosa, J. Paula, Y. Zhub, N. Thakora, Nonextensive entropy measure of EEG following braining jury from cardiac arrest, Phys. A 305 (2002) 619–628.

[25] J.C. Principe, N.R. Euliano, W.C. Lefebvre, Neural and Adaptive Systems, 1st ed., John Wiley & Sons, New York, 2000, 656 pp.

[26] M.H.F. Overwijk, D. Reefman, Maximum-entropy deconvolution applied to electron energy-loss spectroscopy, Micron 31 (2000) 325–331.

[27] X. Li, Edge directed statistical inference with applications to image processing, PhD thesis, Princeton University, 2000, 131 pp.

[28] R.R. Coifman, M.V. Wickerhauser, Entropy-based algorithms for best basis selection, IEEE Trans. Inform. Theory 38 (2) (1992) 713–718.

[29] P. Fakhari, E. Vahedi, C. Lucas, Protecting patient privacy from unauthorized release of medical images using a bio-inspired wavelet-based watermarking approach, Digit. Signal Process. 21 (3) (May 2011) 433–446.

[30] Z. Iscan, A. Yüksel, Z. Dokur, M. Korürek, T. Ölmez, Medical image segmentation with transform and moment based features and incremental supervised neural network, Digit. Signal Process. 19 (5) (September 2009) 890–901.

[31] N. Sriraam, R. Shyamsunder, 3-D medical image compression using 3-D wavelet coders, Digit. Signal Process. 21 (1) (January 2011) 100–109.

[32] J. Luo, Y. Zhu, Denoising of medical images using a reconstruction-average mechanism, Digit. Signal Process. 22 (2) (March 2012) 337–347.

Esin Dogantekin was born in 1981. She graduated from Firat University Medicine Faculty in 2005. She worked as an assistant in Department of Microbiology and Clinical Microbiology of the Firat Medicine Center between 2006 and 2010. She works as an expert doctor of Microbiology and Clinical Microbiology in Bingol State Hospital. Her research interests include medical pattern recognition techniques and medical image processing.

Engin Avci was born in Elazig, Turkey, in 1978. He received the B.S. degree from the Firat University, Technical Education Faculty, Department of Electronics and Computer Education in 2000, M.Sc. degree from the Firat University, Technical Education Faculty, Department of Electronics and Computer Education in 2002, and Ph.D. degree from Firat University, Engineering Faculty, Department of Electrical and Electronics Engineering in 2005. He works as Assistance Professor in Firat University, Technical Education Faculty, Department of Electronics and Computer Education. His research interests concern pattern recognition techniques, communication, signal processing, radar target recognition, and intelligent systems.

Oznur Erkus was born in Turkey, in 1985. She received the B.S. degree from the Firat University, Technical Education Faculty, Department of Electronics and Computer Education in 2007, M.Sc. degree from the Firat University, Technical Education Faculty, Department of Electronics and Computer Education in 2010. Her research interest is image processing.