

# Classification of healthcare data using hybridised fuzzy and convolutional neural network

Balamurugan Ramasamy<sup>1</sup> ✉, Abdul Zubair Hameed<sup>2</sup>

<sup>1</sup>Department of Mechanical Engineering, M Kumarasamy College of Engineering, Karur, TN 639113, India

<sup>2</sup>Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia

✉ E-mail: balaie2012@gmail.com

Published in Healthcare Technology Letters; Received on 5th July 2018; Revised on 2nd January 2019; Accepted on 27th March 2019

Healthcare performs a key role in the health of humans in the world. While gathering a huge amount of medical data, the problems will appear on the classification of healthcare data. In this work, a fuzzy hybridised convolutional neural network (FCNN) model is stated to guess the class of healthcare data. This model collects the information from the data set and builds the decision table based on the collected features from data sets. The attributes that are unrelated are deleted by using principal component analysis algorithm. The decision of normal and cardiac disease is described by using FCNN classifier. Using the data sets from UCI (University of California Irvine) repository the estimation of the presented model is carried on. The performance of the authors' classification technique is measured by various metrics such as accuracy, *F*-measure, *G*-mean, precision, and recall. The experimental results while compared with some of the existing machine learning methods such as probabilistic neural network, support vector machine and neural network, show the higher performance of FCNN. This model presented in this work acts as a decision support pattern in healthcare for therapeutic specialists.

**1. Introduction:** In healthcare systems, Information and Communication Technology (ICT) takes part in offering an opportunity in developing not only the value and adeptness of the healthcare systems but also improves the quality of health services. Healthcare around the world has its significance in general progressing of the physical condition and the welfare of human beings. Sometimes the healthcare data are found hard to classify because of the uncertain and nature of medical data which are supposed to have increased dimensionality. Healthcare data have details such as patient-centric facts, reports about treatments and resource management information. These statistics are huge and filled with information. In the healthcare data mining methods can be probably used in finding hidden relationship and trends in data. These techniques are found to be very effective in healthcare research Taneja *et al.* [1]

Cardiac disease and breast cancer diagnosis done in the automatic procedure are considered as an important process of real-world medical problems. Wu *et al.* [2] found mostly old people are prone to heart disease; it affects the working performance of all patients. By conducting some of the medical examinations, the great disease could be found in the initial stage itself. These tests are somewhat expensive and face some difficulties. The cardiac related researchers based on the medical history present some of the simple tests which are treated as an economical solution.

The data mining process involves taking out the information, which is needed from a massive amount of data. This data gathered helps in carrying out strategic operations also supports in active decision making. Chen and Zhang [3] decision making seeks new and innovative approaches for updating its process concerning with accurate and valid decision making. Mostly large volumes of data that are collected from the healthcare profession are not so simple and by using some of the old techniques processing these data are not possible. Medical diagnosis and prognosis, which usually contain complexity and uncertainty are considered as the complications in decision making. To improve decision making, the data mining process has been used in various healthcare platforms. In the healthcare domain, the extraction of basic information from the diagnosis data comprises some difficult process. Making smart decisions automatically in machine learning, and to achieve accuracy and to save power this process builds

an automated text classifying system by gathering information from a group of classified files.

In Big data, the context related to the data set is assigned to organised and un-organised data having some of the feature extraction and reduction characteristics. Sampling, transformation, denoising, and normalisation strategies could be treated by the selection of representative subclass, receive single input and the noise elimination and extraction of features can be made. From the large data set, some of the sampling procedures are mentioned; they are the random sampling and stratified sampling. When the trained data set divides this leads to the existence of a multi-class problem, to avoid this decomposition method is used. Through this method used the new concern will be shared into sub concerns based on the feature-interrelated model and place. Feature selection in data mining provides classifiers based on decomposition original feature sets are decomposed into several subsets. In the case when the data become bigger they are hard to attain, save and handle. The decomposed subsets that are collected mutually by clustering this process get the data sets' hidden connection. Immense data clustering methodology is categorised into distinct and several machine clusterings. Some of the selected optimal subsets are used in a single machine, and parallel computing is used in multiple data to attain results at the minimum time also increases the calculation speed with scalability multilayer perception, neural network and support vector machine are some classifiers involved and united to produce a multiphase classifier structure to increase classification performance. The classified data which is combined with the original data set shows the maximum shortage to form a new data set for training.

Sahare and Gupta [4] show a classifier based on the future query is constructed with the data set and regarding the data collected from patients have several challenges also is time-consuming and also the data obtained might be large. Like artificial intelligence, the neural networks are treated, and this method provides steps leading to the final decision making and neurons are then connected to the nerve cells, which collect the information from the environment.

To improve decision making, data mining has been employed in various healthcare fields. The extraction of understandable knowledge from diagnosis data remains as one of the major

challenges in the healthcare domain. Rokach [5] in this Letter, proposed a classification problem with the healthcare data. The main contributions of the proposed work are summarised below:

- To perform an extensive study on the state-of-art classification model based on data mining.
- To design an active and proficient modelling system for future decision making that can describe the data of the class.
- To lower the computational complexity and intensify the overall accuracy for classification.
- Result analysis, comparison, and validation of the proposed system.

The remaining part of this Letter is structured as follows. The literature survey of relevant concepts is discussed in Section 2. Section 3 specifies our proposed method. The comparison analysis and performance assessment are explained in Section 4. This Letter concludes in Section 5 and then gives the outlook for future work.

**2. Related work:** The GSAM (genetic standard additive model) has been proposed by Nguyen *et al.* [6], which is the combination of the fuzzy standard additive model with genetic algorithm (GA), presented to handle the uncertainty and computational problems. Three frequent steps which are comprised of GSAM learning process are the unsupervised learning-based rule initialisation; the evolutionary rule optimisation based on GA and gradient descent supervised learning-based parameter. To extract the discriminative features in the high-dimensional data sets for this the wavelet transformation is used. GSAM is made talented with its reduced computational features when minimised the number of wavelets is used. The suggested method is improved by the frequent usage of Wisconsin breast cancer and Cleveland heart disease data sets.

Ludwiget *et al.* [7] examine a fuzzy-based decision tree algorithm which is used for sorting out the genetic expression data. The proposed approach is compared with the standard decision tree algorithm which includes the classification tasks, and some other familiar data mining algorithms are usually applied.

Morente-Molinera *et al.* [8] obtained a linguistic representation by using linguistic modelling methods. The data with the assistance of a multi-granular linguistic scheme could be changed and conveyed using many morphological label sets. Instead of representing with a number, expressing the data with morphological descriptions intensifies the reading ability, decreases the complication and the precision level could be controlled manually by the data recovering methods.

Using numerous supervised learning algorithms, the data sets can be transformed and used for classification tasks. Subsequently, when the testing processes are conducted determines that, in certain circumstances, the data complication is reduced which results in improved results on the subject of classification.

Deng *et al.* [9] made a study to avoid the defects of static depiction using fuzzy learning ideas into DL (deep learning). The hierarchical deep neural network is the greater part of the fuzzy system proposed, and this gets information from both fuzzy and neural representations. Then, these two respective views, the knowledge gained are joined totally making the ultimate representation of data which has to be considered. The efficiency of the pattern is exhibited on the sensible image classification tasks, data prediction using the frequency which is greater than before and MRI (magnetic resonance image) class of brain these all increase the uncertainties in the raw data.

Fisher *et al.* [10] explained a new data-determined workflow particularly to the area, possibly would be simplified for observing some of the approaches using time sequence data. Internal erosion events are identified with the help of the sensors which are placed on the levee surfaces. To make sure whether the workflow is suitably strong to perform using various data sets and several strange

events, various data sets from tentative test site earth banks are used. Nine spectral features are obtained from wavelet-denoising methods from disintegrated sections of the time-series data.

**3. Proposed method:** The challenges in training are solved using fuzzy systems when there is a possibility of having many inputs of the data. Osoba *et al.* [11] examine the exclamations and the dimensionality of the fuzzy convolutional neural network (CNN). Fig. 1 demonstrates the architecture of the proposed classification. In common, the fuzzy system performance and the convergence rate of the learning process could be weakened by the high-dimensional data. So, there is a need for a tool which reduces the dimension also the feature extraction which is to be used before the performance of fuzzy CNN. This is considered very crucial in terms of medical data because these data's are collected with high-dimensional type.

**3.1. Principal component analysis (PCA):** The data estimation can be reduced using PCA. To process high-dimensional data sets, dimensionality reduction is seen as a better way. That is the 2D data set will be reduced to 1D. It also may be beneficial for dealing with the noisy elements in the data sets. The grouping calculation is done after dimension reduction is performed.

PCA algorithms pre-processing steps are shown below:

- Pre-processing is determined before executing PCA.
- Compute the covariance matrix.
- Determine the eigenvectors.
- Creating segments and setting up a module vector.
- Creating main modules.

Initially, as of  $n$ -aspect to  $k$ -aspect, the data is reduced. The typical data are  $\{x_1, x_2, \dots, x_n\}$ . Then compute the mean by using

$$\mu = \left(\frac{1}{n}\right) \sum_{i=1}^n x_i \quad (1)$$

Once the mean is found, compute the covariance by using the resulting equation

$$C = \left(\frac{1}{n}\right) \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (2)$$

Here,  $n$  is the number of data items.

The eigenvectors are organised by the condition mentioned below:

$$ce_i = \lambda_i e_i \quad (3)$$

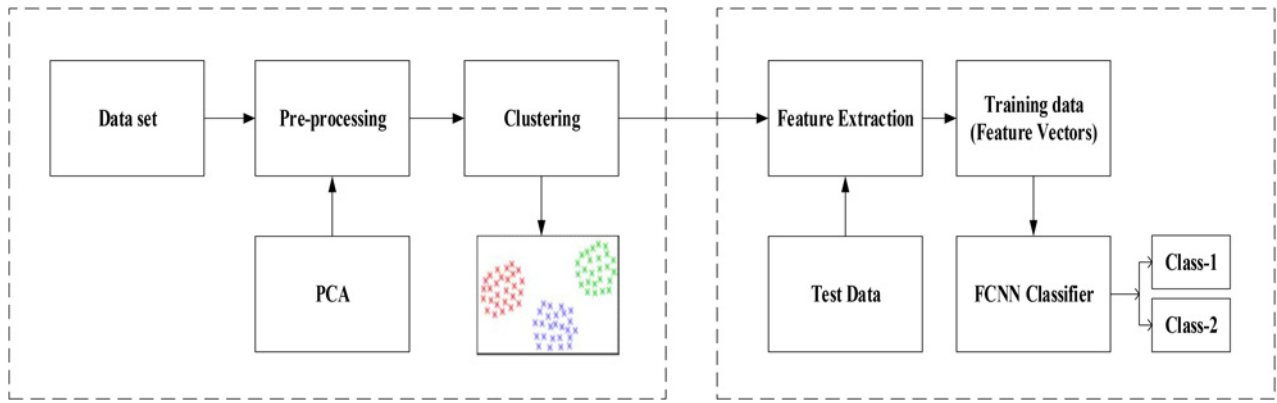
The covariance matrix's eigenvalues and eigenvectors are to be found. The eigenvalues are computed by the following equation:

$$\det(c - \lambda I) = 0 \quad (4)$$

The principal components are dealt with the eigenvector with most important eigenvalues.

**3.2. Clustering:** The reduced data sets are clustered with the input  $X = \{f_1, f_2, f_3, \dots, f_n\}$  is needed to be grouped into some clusters. The below the formula optimises centres of the  $k$  clusters computed by  $k$ -means and the error of the each cluster

$$\min \sum_{j=1}^n \sum_{i=1}^{nk} \|f_i^j - C_j\|^2 \quad (5)$$



**Fig. 1** Schematic representation of the proposed model

where  $\|f_i^j - C_j\|^2$  is the distance between a data point  $f_i^j$  of the cluster  $j$  and the cluster centre  $C_j$ . A set of  $k$ -initial centres are accessed by the following steps:

- (1) The initial point from the input by a uniform random variable called  $C_1$ .
- (2) Calculate the distance  $D(X)$  between  $f_i$  and the centre  $C_1$  for every data.

$$D(x) = \sqrt{\sum_{i=1}^n (f_i - c_i)^2} \quad (6)$$

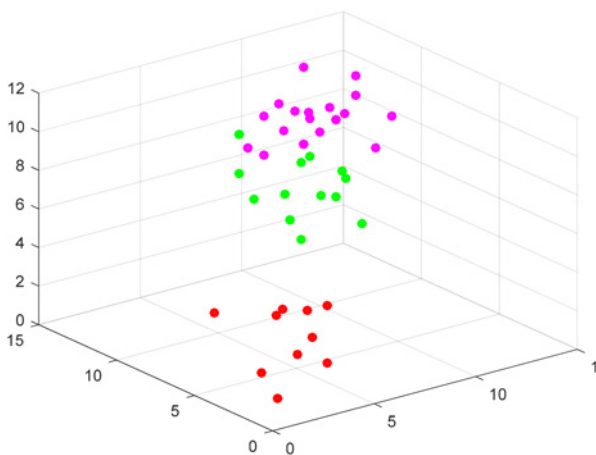
- (3) Then, probability weighted distribution randomly select a new candidate to become a centre

$$\frac{D(x)^2}{\sum_{i=0}^k D(x_i)^2} \quad (7)$$

- (4) Continue steps 2 and 3 until the selection of  $k$  initial centres.
- (5)  $K$ -means algorithm is applied after selecting an initial centre.

In this Letter, the relevant documents are considered as appoints. When  $k=3$ , these points are partitioned into three clusters with the  $k$ -means algorithm.

The  $k$ -clusters are produced by the  $k$ -means clustering technique for whole files. The set of clusters  $C = \{C_1, C_2, \dots, C_k\}$ , where  $C_k (k = 1, 2, \dots, k)$  are consisting of a group of related documents are belonging to a distinct cluster  $C_i$ . Finally, clustered documents



**Fig. 2** 3D projection of  $k$ -means

are optimised accurately from the cluster by ABC algorithm. The 3D projection view of the  $k$ -means clustering is illustrated in Fig. 2.

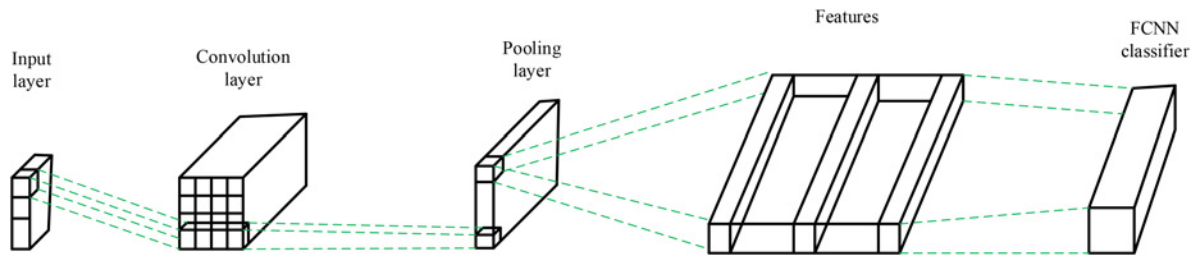
**3.3. Feature selection:** This scheme intends to decide on the feature subsets, which can be utilised to get a mapping function from samples to classes that are ‘in the same class as conceivable’ as per some measure. Streaming features are characterised as features that flow in one by one over time, whereas the quantity of training examples remains fixed. In this work, the features are chosen by the heuristic-based genetic algorithm.

**3.4. Heuristic-based GA for feature selection:** GA is a technique using the populations. For a problematic domain, it utilises a binary string of static length to signify a reasonable result. In this situation, let ‘F’ is an aggregate sum of features; there occurs  $2^F$  features subset. Every entity is spoken by an F-bit string. ‘1’ or ‘0’ representation of any bit denotes the availability or un-availability of the resultant feature. Probable answers or populace with a stable quantity of the populace and its magnitude is created by chance. Evaluation of every individual is related to its fitness using a fitness function. The population consists of three genetic operators, and they are (i) selection, (ii) crossover, and (iii) mutation. Based on the individual’s fitness, the selection operator picks out individuals in the population. Compared with the lower-fitness individual the higher-fitness individual selects more probability. A pair of certain entities or parentages at that time crossover, they interchange their information using one another to produce renewed entities or offspring. Then, GA does modification by reversing the rate 0/1 of the portion of a particular entity. This procedure functions regularly until an ending condition is attained.

Every phase is termed as a generation. To design the mentioned fitness function, various correlations are presented as a further condition. The fitness function is segmented into two grades: based-cost and benefit-cost. The based-cost term is designed by the following condition (1)

$$BASED\_COST = acy(x) * \rho_{x,y} \quad (8)$$

In the training set,  $acy(x)$  is termed as the accuracy concerning classification with entity  $x$ .  $\rho_{x,y}$  is the multiple correlation factors among  $x$  and decision class  $y$ . Based on this period, high accuracy is being produced by the individual and obtaining a high multiple correlation coefficient results in achieving a good score. Like a gain cost, the benefit-cost is also considered in the circumstance in which some of the multiple correlation coefficients values among the certain set and the complete set is better compared to a distinct constant ( $-0.025$  in this research) then accuracy for training with the certain set is higher than them by means of the complete set. The sum of feature in the total set is  $F$  and the total feature in  $x$



**Fig. 3** Hybrid structure of CNN and fuzzy

is  $e(x)$ , when the state is found correct at that time benefit-cost is known by

$$BENIFIT\_cost = f - e(x)/f \quad (9)$$

Or else, it turns out to be zero. Then removing any feature will be fulfilled when executed in this method. So, a basic method of the stated fitness function based on the novel standard concerning with the performance of classification is mentioned below:

$$fitness(x) = [BASED\_cost + benefit\_cost]/2.0 \quad (10)$$

where single  $x$  fitness is said as  $(x)$ .

**3.5. Fuzzy hybridised CNN (FCNN):** The information gathered from the fuzzy and CNN is combined with the fusion layer to build the complete representation for information ordering. On original data, the uncertainties will be minimised by fuzzy and the noise will be reduced by the neural statement. The exposed new method FCNN uses both fuzzy and neural representations to create the fused representation for the last classification. The FCNN method is used on challenging classification task like data ambiguity and noise. The experiments of FCNN are tested with different difficult data classification tasks. The uncertainties and noises in the original data can easily be removed by the proposed FCNN method. Finally, the two different representations are used for the classification. Fig. 3 describes the hybrid architecture of CNN and fuzzy. For the input, the collected raw data is applied. At the initial stage of training, the raw data will be extracted from the applied input data, and it is converted into a multi-dimensional matrix, this extracted matrix is given as the input for the next layer. The max-pooling rule is used by the pooling layer that is present next to the convolution layer, and it minimises the dimensions by choosing the coefficients over every sampling cell. By using the same process, the other pooling and convolution layers are built. The output that is obtained from the pooling layer is used as an input for the fully connected layer. The output from the fully connected layer is used to extract features from the inputted raw data set. This layer-by-layer process can be done several times based on the requirements. On our proposed FCNN technique, to minimise the amount of parameters and the complexity while processing on layers, on the last CNN part, a sparse regularisation penalty is additionally included. The sparse regularisation penalty helps to create sparseness on the distribution of the weights. Subsequently, on support vector regression classifier, the feature map will be given as a in it includes much valuable data.

**3.6. FCNN training:** There on the training process of the FCNN method includes two main processes called parameter initialisation and fine-tuning. Because of the convex of a complete learning system in deep learning the starting stage is much complicated. Efficient neural network coverage will be earned by initialising the superior strategy. The initialisation process will perform for parts such as the fuzzy and

convolutional neural. In the deepest part, the weights among every layer are present. The hybridisation and the classification function will be prepared. Biases  $b$  for every node is set as zero. After this, based on the rule, the weight among the layers is initialised, and it is given as

$$Un \left[ -\frac{1}{\sqrt{m^{(l-1)}}}, \frac{1}{\sqrt{m^{(l-1)}}} \right] \quad (11)$$

The  $Un$  is the even dispersal and  $m^{l-1}$  is the  $(l-1)$ th level. For the hybridisation level,  $m^{l-1}$  calculates the number of nodules on the final levels of the fuzzy and convolutional features.

**4. Experimental evaluation:** Results regarding our method and experimental setups are explained in this section. The implementation process of our method is explained in Section 3. To check the result, different classification methods are used with the four healthcare data sets collected from the UCI repository. The measured ratio of error and accuracy of classification methods is used for the investigation as the measurement parameters. These parameters indicate the higher accuracy rate and minimal error rate on the data set used for the classification process with our technique. On the other side, less accuracy rate and higher error rate are used with the data set on classification technique and classifier obtains that the data set is minimal correctly classified. Initially, on our work, the input data is separated as testing and training data. For the creation of classifier, the training set data is used and to validate the test set will be used. The rates that are allotted for the data training and testing are 66 and 34%. The tenfold cross-validation technique is used for the generation of classifiers for classification which are included through the machine learning tools. The results of our process are measured with respect to the rates of accurateness and error. In this unit, the data set details have been described, evaluation and results are discussed. First, the section describes the data sets and the hardware-software used for the experiments. The next section describes the performance metric used for machine learning. Finally, the results obtained using different data sets are demonstrated.

**4.1. Data set description:** The proposed classifier experiments on three benchmark data sets collected from the UCI Machine Learning Repository for performance evaluation. The data sets are randomly generated for training, testing and validation in simulation time. The considered data sets from the UCI for performance evaluation are Heart-Statlog, Wisconsin Breast Cancer and Cleveland Heart Disease. The proposed work is implemented in the MATLAB platform mentioned in Table 1.

**Table 1** Simulation parameters

Processor	Intel Core 2 Quad @ 2.3 GHz
RAM	3 GB
OS	Windows 7
Mat Lab version	R 2016a



**Table 2** Classification results

Metric	Data set 1				Data set 2				Data set 3			
	FCNN	GSAM	FDT	FDNN	FCNN	GSAM	FDT	FDNN	FCNN	GSAM	FDT	FDNN
recall	0.972	0.896	0.813	0.789	0.948	0.875	0.807	0.786	0.960	0.867	0.812	0.802
F-measure	0.931	0.914	0.852	0.802	0.973	0.907	0.819	0.810	0.925	0.832	0.856	0.854
precision	0.917	0.762	0.803	0.750	0.950	0.756	0.821	0.762	0.946	0.786	0.876	0.785
G-mean	0.864	0.852	0.816	0.785	0.866	0.856	0.867	0.754	0.872	0.877	0.877	0.752
accuracy	0.925	0.902	0.845	0.854	0.948	0.890	0.823	0.812	0.979	0.888	0.813	0.888

4.2. Results: In this section, results are reported for different methods for given three benchmark data sets. After the iterations, the classification accuracy will be measured. We compared our proposed algorithm with two other classifiers to obtain the performance of the proposed system. Table 2 shows the classification results of the proposed work. On classifying the data set employing original features, it is noted that the classification accuracy is 97.92%.

**5. Conclusion:** For the previous ten years, most of the deaths are caused by heart diseases. Most of the researchers are designing several data mining methods for healthcare doctors to cure these heart diseases. On handling the heart diseases, mostly the machine learning techniques are used, and one of the effective techniques is CNN. The method introduced in this Letter combines the  $k$ -means clustering and FCNN for diagnosing the heart disease. The experimental results of our method show a higher accuracy on diagnosing the heart diseases. The accuracy achieved by our method is 95.5%; it is much better than the other older systems.

**6. Funding and declaration of interests:** Conflict of interest: none declared.

## 7 References

- [1] Taneja S., Suri B., Gupta S., *ET AL.*: 'A fuzzy logic based approach for data classification'. Data Engineering and Intelligent Computing, 2008, pp. 605–616
- [2] Wu X., Zhu X., Wu G.Q., *ET AL.*: 'Data mining with big data', *IEEE Trans. Knowl. Data Eng.*, 2014, **26**, (1), pp. 97–107
- [3] Chen C.P., Zhang C.Y.: 'Data-intensive applications, challenges, techniques and technologies, a survey on big data', *Inf. Sci.*, 2014, **275**, pp. 314–347
- [4] Sahare M., Gupta H.: 'A review of multi-class classification for imbalanced data', *Int. J. Adv. Comput. Res.*, 2012, **2**, (5), pp. 160–164
- [5] Rokach L.: 'Decomposition methodology for classification tasks: a meta decomposer framework', *Pattern Anal. Appl.*, 2006, **9**, (2), pp. 257–271
- [6] Nguyen T., Khosravi A., Creighton D., *ET AL.*: 'Classification of healthcare data using genetic fuzzy logic system and wavelets', *Expert Syst. Appl.*, 2015, **42**, (4), pp. 2184–2197
- [7] Ludwig S.A., Picek S., Jakobovic D., *ET AL.*: 'Classification of cancer data: analyzing gene expression data using a fuzzy decision tree algorithm', in Kahraman C., Topcu Y. (Eds.): 'Operations research applications in health care management' Springer, Cham., Switzerland, (2018), **262**, pp. 327–347
- [8] Morente-Molinera J.A., Mezei J., Carlsson C., *ET AL.*: 'Improving supervised learning classification methods using multigranular linguistic modelling and fuzzy entropy', *IEEE Trans. Fuzzy Syst.*, 2017, **25**, (5), pp. 1078–1089
- [9] Deng Y., Ren Z., Kong Y., *ET AL.*: 'A hierarchical fused fuzzy deep neural network for data classification', *IEEE Trans. Fuzzy Syst.*, 2017, **25**, (4), pp. 1006–1012
- [10] Fisher W.D., Tracy K., Krzhizhanovskaya V.V., *ET AL.*: 'Anomaly detection in earth dam and levee passive seismic data using support vector machines and automatic feature selection', *J. Comput. Sci.*, 2017, **20**, pp. 143–153
- [11] Osoba O., Mitaim S., Kosko B.: 'Triply fuzzy function approximation for hierarchical Bayesian inference', *Fuzzy Optim. Decis. Mak.*, 2012, **11**, (3), pp. 241–268