Pairwise FCM based feature weighting for improved classification of vertebral column disorders

Yavuz Unal a, Kemal Polat b,⁎, H. Erdinc Kocer c

a Education Faculty, Amasya University, Amasya, Turkey
b Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Abant Izzet Baysal University, 14280 Bolu, Turkey
c Technology Faculty, Department of Electrical and Electronics, Selcuk University, Konya, Turkey

ARTICLE INFO

Article history:
Received 7 October 2013
Accepted 6 December 2013

Keywords:
Vertebral column
Pairwise Fuzzy C-means clustering based feature weighting
Classification
Data pre-processing

ABSTRACT

In this paper, an innovative data pre-processing method to improve the classification performance and to determine automatically the vertebral column disorders including disk hernia (DH), spondylolisthesis (SL) and normal (NO) groups has been proposed. In the classification of vertebral column disorders’ dataset with three classes, a pairwise fuzzy C-means (FCM) based feature weighting method has been proposed. In this method, first of all, the vertebral column dataset has been grouped as pairwise (DH-SL, DH-NO, and SL-NO) and then these pairwise groups have been weighted using a FCM based feature set. These weighted groups have been classified using classifier algorithms including multilayer perceptron (MLP), k-nearest neighbor (k-NN), Naive Bayes, and support vector machine (SVM). The general classification performance has been obtained by averaging of classification accuracies obtained from pairwise classifier algorithms. To evaluate the performance of the proposed method, the classification accuracy, sensitivity, specificity, ROC curves, and f-measure have been used. Without the proposed feature weighting, the obtained f-measure values were 0.7738 for MLP classifier, 0.7021 for k-NN, 0.7263 for Naive Bayes, and 0.7298 for SVM classifier algorithms in the classification of vertebral column disorders’ dataset with three classes. With the pairwise fuzzy C-means based feature weighting method, the obtained f-measure values were 0.9509 for MLP, 0.9313 for k-NN, 0.9603 for Naive Bayes, and 0.9468 for SVM classifier algorithms. The experimental results demonstrated that the proposed pairwise fuzzy C-means based feature weighting method is robust and effective in the classification of vertebral column disorders’ dataset. In the future, this method could be used confidently for medical datasets with more classes.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Lower back pain is among the most frequent ailments in the world. It is the second commonest illness after headaches. Besides lowering quality of life, it also causes labor loss and early retirement [1]. This ailment brings huge burdens to the relatives of ill people and to the economies of countries. The fact that the imaging techniques that are used in the diagnosis of this ailment are high-cost and the radiologists who can diagnose this ailment are few in number make it more important for this ailment to be treated in a computer-assisted manner [2].

Lower back pain is a very common ailment in adults. About 90% of all adults have at least one lower back pain attack in their whole lives. If all adults were questioned at the same time, it would become obvious that 15% of them have lower back pain [3]. Lower back pain may be due to trauma, lifting heavy objects or a reverse action; however, it may not be due to a certain cause. Moreover, some cancer types may also cause lower back pain; however, this is rare. Again, brucella and some similar infections may also appear together with lower back pain. For all these reasons, lower back pain should be taken serious and the diagnosis should be fast and accurate [4].

Today, the plain radiography in the research of lower back pain has become less important. The patient does not receive any X-rays and the superior imaging ability in various plans; and its being able to image the spinal cord and other soft tissues clearly makes magnetic resonance become more important day by day. The magnetic resonance imaging method makes regular diagnosis and discriminatory diagnosis possible. Although magnetic resonance is a very useful method, the assessment of the images requires great experience. Wrong comments on the images lead to wrong diagnoses and discriminatory diagnosis possible. Although magnetic resonance is a very useful method, the assessment of the images requires great experience. Wrong comments on the images lead to wrong diagnoses and discriminatory diagnosis possible. Although magnetic resonance is a very useful method, the assessment of the images requires great experience. Wrong comments on the images lead to wrong diagnoses and discriminatory diagnosis possible. Although magnetic resonance is a very useful method, the assessment of the images requires great experience. Wrong comments on the images lead to wrong diagnoses and discriminatory diagnosis possible. Although magnetic resonance is a very useful method, the assessment of the images requires great experience. Wrong comments on the images lead to wrong diagnoses and discriminatory diagnosis possible.
In this study, we have used the vertebral column dataset taken from UCI (University of California, Irvine) machine learning [7]. In this dataset, there are six biomechanical features and 310 data including three classes including disk hernia (DH), spondylolisthesis (SL), and normal (NO). These biomechanical features are derived from the shape and orientation of the pelvis and lumbar spine. The names of these features are pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius and grade of spondylolisthesis [8].

There have been several studies on classification and determination of vertebral column diseases in the literature. These are as follows: Neto and Barreto applied standalone support vector machine (SVM), multiple layer perception (MLP) and generalized regression neural network (GRNN) which are among machine learning methods to UCI pathologies of the vertebral column dataset (VCP dataset). They then applied a combination of these machine learning (ML) methods and released the results in a comparative method after classifying their work [8]. Mattos and Barreto introduced two new ensemble methods in their work. One of these methods is fuzzy adaptive resonance theory (FA), and the other one is self-organizing map (SOM) neural networks based classifiers. Using many datasets including VCP dataset, they released comparative results [9]. Neto et al. called their study as “Reject Option” and applied VCP dataset in this study where they introduced a new technique. They used several different ML classifiers in their study [10]. Abdurabou designed a hybrid system in his study and joined the two machine learning (ML) techniques: case-based reasoning (CBR) and artificial neural network. This system was applied to the UCI vertebral column dataset. The hybridization of CBR and ANN showed that the classification accuracy increased [11]. Ansari et al. diagnosed disease through machine learning classifiers to vertebral column dataset taken from UCI. These classifiers are: feed forward back propagation neural network, generalized regression neural network and support vector machine. They achieved 93.87% classification accuracy with feed forward back propagation neural network [12].

In this study, a novel data pre-processing method called the pairwise fuzzy C-means based feature weighting method has been proposed both to determine of type of vertebral column disorders including disk hernia group, spondylolisthesis group, and normal group to improve the classification accuracy of used classifier algorithms in the classification of vertebral column disorders’ dataset. As seen from the literature review, the proposed feature weighting method is firstly proposed and also applied to this dataset by us. The proposed hybrid system consists of two stages: in the first stage (feature weighting procedure), each feature in the vertebral column dataset has been weighted to transform from a non-linearly separable case to a linearly separable case by pairwise fuzzy C-means based feature weighting. In the second stage, the weighted vertebral column dataset that has a linearly separable distribution is classified by classifier algorithm including multi-layer perceptron (MLP), k-nearest neighbor (k-NN), Naive Bayes, and support vector machine (SVM). Before applied the pairwise fuzzy C-means based feature weighting (PFCMBFW) to vertebral column dataset, the obtained classification accuracies were 78.387, 85.4839, 83.2258, and 81.6129 by k-NN, MLP, Naive Bayes, and SVM classifiers, respectively. After applied the PFCMBFW to this dataset, the obtained classification accuracies were 95.4839, 96.7742, 97.4194, and 96.4516 by k-NN, MLP, Naive Bayes, and SVM classifiers, respectively. The obtained results have shown that the proposed feature weighting method could be confidently used to classify the vertebral column disorders.

The organization of the paper is as follows: the dataset used in this paper is explained in Section 2. A detailed description of the methodology adopted in this paper is described in Section 3 which includes feature extraction, feature weighting methods and classifiers. Section 4 explains the interpretation of the results and discussion. Section 5 concludes the paper.

### 2. Material: vertebral column dataset

In this study, the used vertebral column with three classes (pathologies of the vertebral column dataset) was formed by Dr. Henrique Mota et al. [7]. In this dataset, there are 310

<table>
<thead>
<tr>
<th>The case of dataset</th>
<th>The name of feature in dataset</th>
<th>Min. value</th>
<th>Max. value</th>
<th>Mean value</th>
<th>Std. dev. value</th>
<th>The correlation coefficient between feature and class label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>Pelvic incidence</td>
<td>26.148</td>
<td>129.834</td>
<td>60.497</td>
<td>17.237</td>
<td>-0.029</td>
</tr>
<tr>
<td>Weighted</td>
<td></td>
<td>18.58</td>
<td>195.32</td>
<td>69.27</td>
<td>40.75</td>
<td>-0.167</td>
</tr>
<tr>
<td>Raw</td>
<td>Pelvic_tilt</td>
<td>-6.555</td>
<td>49.432</td>
<td>17.543</td>
<td>10.008</td>
<td>-0.211</td>
</tr>
<tr>
<td>Weighted</td>
<td></td>
<td>-9.84</td>
<td>74.26</td>
<td>20.55</td>
<td>16.86</td>
<td>-0.372</td>
</tr>
<tr>
<td>Raw</td>
<td>Lumbar_lordosis_angle</td>
<td>14</td>
<td>125.742</td>
<td>51.931</td>
<td>18.554</td>
<td>-0.036</td>
</tr>
<tr>
<td>Weighted</td>
<td></td>
<td>8.45</td>
<td>212.31</td>
<td>64.69</td>
<td>46.69</td>
<td>-0.131</td>
</tr>
<tr>
<td>Raw</td>
<td>Sacral_slope</td>
<td>13.367</td>
<td>121.43</td>
<td>42.954</td>
<td>13.423</td>
<td>0.12</td>
</tr>
<tr>
<td>Weighted</td>
<td></td>
<td>7.94</td>
<td>182.79</td>
<td>49.81</td>
<td>29.43</td>
<td>0.026</td>
</tr>
<tr>
<td>Raw</td>
<td>Pelvic_radius</td>
<td>70.083</td>
<td>163.071</td>
<td>117.92</td>
<td>13.317</td>
<td>0.234</td>
</tr>
<tr>
<td>Weighted</td>
<td></td>
<td>65.86</td>
<td>153.27</td>
<td>116.63</td>
<td>15.08</td>
<td>0.202</td>
</tr>
<tr>
<td>Raw</td>
<td>Degree_spondylolisthesis</td>
<td>-11.058</td>
<td>418.543</td>
<td>26.297</td>
<td>37.559</td>
<td>-0.119</td>
</tr>
<tr>
<td>Weighted</td>
<td></td>
<td>-0.84</td>
<td>5859</td>
<td>351.58</td>
<td>533.107</td>
<td>-0.120</td>
</tr>
</tbody>
</table>
samples consisting of 100 normal group, 60 disk hernia group, 
150 spondylolisthesis group and six biomechanical features called 
pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, 
pelvic radius and grade of spondylolisthesis. These used features 
were derived from the shape and orientation of the pelvis and 
lumbar spine [8]. The class distribution of raw vertebral column 
with three classes is shown in Fig. 1. As can be seen from Fig. 1, this 
dataset has a non-linearly separable distribution. Therefore, the 
classification of this dataset is a challenge task in terms of classifier 
methods. The data weighting or pre-processing method should 
be used both to transform from non-linearly separable dataset to 
linearly separable dataset and to improve the classification perf 

euance of classifier algorithms. The aims of the feature weighting method are both to map the 
features according to their distributions in a dataset and to transform 
from non-linearly separable dataset to linearly separable dataset. The 
feature weighting method aims to decrease the variance within 
features forming dataset. Thanks to this weighting method, the similar 
data in same feature are gathered and the discrimination ability of the 
classifier is increased. So, the FCM based feature weighting method 
should be used prior to the classification stage [23]. The readers can 
refer to [23] for more information in FCM clustering.

The FCM clustering based feature weighting method works as 
follows:

(i) The cluster centers of each feature are calculated using the 
FCM method.
(ii) After computing the centers of features, the ratios of means of 
features to their cluster centers are calculated.
(iii) These ratios are multiplied with the data of each feature.

The pseudo code of the FCM clustering based feature weighting 
method is shown in Fig. 5.

Fig. 6 denotes the class distribution of raw and weighted 
vertebral column disorders’ datasets. In Fig. 6, after applying the 
PFCMBFW to the vertebral column disorders’ dataset, the class 
distribution of the raw dataset has been transformed from a non-
linearly separable dataset to linearly separable dataset. Also, we 
have given the class distributions of pair three groups including 
DH-SL, DH-NO, and SL-NO in Fig. 7.

3. Method

3.1. The proposed structure

In this paper, an innovative feature weighting method called 
PFCMBFW is proposed and combined with classifier algorithms 
including multilayer perceptron (MLP), k-nearest neighbor (k-NN), 
Naive Bayes, and support vector machine (SVM) to classify the 
vertebral column disorders with three classes. Fig. 3 shows the 
block diagram of the proposed method.

3.1.1. Stage I: pairwise fuzzy C-means based feature weighting 
(PFMBFW)-pre-processing stage

Previously, we have used the FCM clustering based feature 
weighting method proposed by us to diagnose and classify the 
various diseases [23]. In this paper, we have improved the FCM 
based feature weighting method for pair grouping the classes in a 

multi-class dataset and called the pairwise FCM based feature 
weighting method. The proposed scheme is explained in Fig. 4. 
In this scheme, we have grouped the dataset as pairwise and then 
weighted these groups using fuzzy C-means clustering based feature 
weighting method. The reason of using pairwise of this structure is 
twofold. The first reason is to improve the classification performance 
of classifier algorithms in the multi-class datasets. The second reason 
is to propose a novel scheme to classify the multi-class datasets.

The aims of the feature weighting method are both to map the 
features according to their distributions in a dataset and to transform 
from non-linearly separable dataset to linearly separable dataset. The 
feature weighting method aims to decrease the variance within 
features forming dataset. Thanks to this weighting method, the similar 
data in same feature are gathered and the discrimination ability of the 
classifier is increased. So, the FCM based feature weighting method 
should be used prior to the classification stage [23]. The readers can 
refer to [23] for more information in FCM clustering.

The FCM clustering based feature weighting method works as 
follows:

(i) The cluster centers of each feature are calculated using the 
FCM method.
(ii) After computing the centers of features, the ratios of means of 
features to their cluster centers are calculated.
(iii) These ratios are multiplied with the data of each feature.

The pseudo code of the FCM clustering based feature weighting 
method is shown in Fig. 5.

Fig. 6 denotes the class distribution of raw and weighted 
vertebral column disorders’ datasets. In Fig. 6, after applying the 
PFCMBFW to the vertebral column disorders’ dataset, the class 
distribution of the raw dataset has been transformed from a non-
linearly separable dataset to linearly separable dataset. Also, we 
have given the class distributions of pair three groups including 
DH-SL, DH-NO, and SL-NO in Fig. 7.

3.1.2. Stage II: the classification algorithms

In this work, four different classifiers consisting of multilayer 
perceptron (MLP), k-nearest neighbor (k-NN), Naive Bayes, and support 
vector machine (SVM) were used to study the efficacy of the 
proposed system and their brief description was given in this section.

3.1.2.1. Support vector machine (SVM) classifier algorithm. The SVMs 
are designed to form the following plane in that space and classify 
after carrying the data to high-dimension space:

\[ d(x) = w^T x + b \]  

(1)
Here $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are the parameters that form the multi-dimensional plane, and $\phi(x)$ is the conversion function that is used to convert the data into high-dimensional space from entry space.

The working principle of the SVM consists generally of determining the decision borders (hyper-planes) that separate the data of each class from each other in the most proper way [13].

The classification of a two-class data that is separable in a linear manner is the most basic problem for SVM. To solve this problem, SVM tries to determine an optimum hyper plane where the interclass border is at its maximum level. This plane must be able to make the distinction between the two classes in the best way. The optimum hyper plane carries the distances between the pixels that belong to each class and the hyper plane to their maximum levels [14].

SVMs can be applied to both data groups that can be separated in a linear manner and those that cannot. In the case where they are separated in a linear manner, the purpose of the SVM is to find the hyper-plane that separate the two classes from each other.

In situations where data are not separable in linear ways, the following method is applied: The problem in which education data are left on the other side of the hyper plane is solved by adding a regulatory parameter which has positive values and is shown with C to the problem. This parameter ensures the balance
between the minimization of wrong classification errors and maximization of a positive loose variable ($\xi_i$) and the border. In such a situation, the optimization problem for the data which are inseparable in linear manner is expressed as follows [15]:

$$\min \left[ \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \xi_i \right]$$

In the classification of dataset using SVM, we have used the Pearson VII function based universal kernel proposed by Karl Pearson in 1895 [24]. The mathematical formulation of the Pearson VII function based universal kernel is given in Eq. (3). In this work, the used parameters have been found by a trial and error method. In this way, the best SVM structure is obtained:

$$K_r(x', y') = \frac{1}{1 + \left( \frac{2\|x - y\|^2}{\sqrt{2\|x\|^2 - 1}} \right)^m}$$

3.1.2.2. Multilayer perceptron (MLP) classifier algorithm. Today, these are the multilayer perception networks that have the ability of...
making general analyses and that are used in developing many learning algorithms and are known and used in widespread areas.

These networks are models that contain numerous units organized within multi-layers and are for general purposes, flexible and non-linear. Multi-layer perception networks are valuable in problems where there is little or no knowledge about the relation forms between input vectors and their outputs [16,17]. Fig. 8 shows the used MLP structure with five input layers, two hidden layers, and three output layers. In training the MLP classifier algorithm, the below parameters were used as follows:

- The used activation function: Sigmoid
- The learning rate: 0.3
- The momentum rate: 0.2
- The number of epochs: 500.

3.1.2.3. k-nearest neighbor (k-NN) classifier algorithm. k-Nearest neighbor (k-NN) algorithm is one of the simplest pattern recognition methods that classify the objects based on the nearest training samples in the feature space [18]. This algorithm makes classification according to the class of the nearest neighbor as a given “k” value. The classification of a vector in the k-nearest neighbor algorithm is made by using vectors whose classes are already known. The sample to be tested is taken to the procedure one by one with each example in the training set. In order to define the class of the specific sample to be tested, samples that are the nearest to it are chosen in “k” number. The sample to be tested is said to belong to the class that has the most samples chosen in the set [19]. In our study, we have selected the k value of 1 in the training and testing of k-NN classifier since we have obtained the best classification performance for k = 1.

3.1.2.4. Naive Bayes classifier algorithm. This is a statistical and probability based classifier used in the classification process [20]. In order to understand this algorithm better, we need to look at the conditional probability concept. If the realization of an event is dependent on some conditions, we can talk about “conditional probability”. Let us look at A and B events that agree with each other. A and B events have mutual points. This situation can be expressed as A ∩ B ≠ ∅. The B event will be realized in case the A event is known. This means the occurrence of the event B depends on the occurrence of the event A. This probability is shown as p(B/A) and expressed like this

\[ P(B/A) = \frac{P(A \cap B)}{P(A)} \] (4)

The concept “Naive Bayes classifier”, or in short, “Bayes Classifier” may be explained like this.

Let X be the very example whose membership is not known. Then the example X = \( \{x_1, x_2, ..., x_n\} \) is formed from the quality values. Let us assume that there is “m” class in this example class. Let the \( C_1, C_2, ..., C_n \) be the class values. The following is then expressed as

\[ P(C_i|X) = \frac{P(X|C_i)p(C_i)}{P(X)} \] (5)

probabilities are calculated. To reduce the process burden in the calculation, simplifying is possible for the \( P(X|C_i) \) probability. For this, the following formula can be set by accepting that the \( x_i \) values of the sample are independent of each other

\[ P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \] (6)

In order to classify the unknown example X, and since the denominators in \( P(C_i|X) \) are equal to each other, it is sufficient to compare the quanta values. By choosing the highest value among these values, the class of the unknown sample is defined

\[ \arg \max_{C_i} {P(X|C_i)p(C_i)} \] (7)

The above expression that uses the posterior probabilities is also known as maximum a posteriori classification MAP. As a result, the following formula is used as the Bayes classifier [20,21]:

\[ C_{MAP} = \arg \max_{C_i} \prod_{k=1}^{n} P(x_k|C_i) \] (8)

4. Results and discussion

4.1. Performance metrics

In order to test the performance of the proposed method, the classification accuracy, sensitivity, specificity, precision, recall, F-measure, ROC (receiver operating characteristics), and AUC (area under the ROC curve) values have been used and explained as follows.

In the training and testing of classifier models, 10-fold cross validation scheme and 50–50% training-testing data split have
been used. To explain the performance measures, the confusion matrix has been given. The confusion matrix is shown in Table 2 (actual vs. predicted).

For classification accuracy, sensitivity and specificity analysis, precision, recall and f-measure, the following expressions have been used:

Classification accuracy (%) = \( \frac{TP + TN}{TP + FP + TN + FN} \times 100 \) (9)

Sensitivity (%) = \( \frac{TP}{TP + FN} \times 100 \) (10)

Specificity = \( \frac{TN}{FP + TN} \times 100 \) (11)

Precision = \( \frac{TP}{TP + FP} \) (12)

Recall = \( \frac{TP}{TP + FN} \) (13)

\( f - measure = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \) (14)

Table 2
Confusion matrix.

<table>
<thead>
<tr>
<th>Confusion matrix</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True negative (TN)</td>
<td>False positive (FP)</td>
</tr>
<tr>
<td>No</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Yes</td>
<td>TN</td>
<td>FP</td>
</tr>
</tbody>
</table>

TP, TN, FP and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Table 3
The obtained performance metrics in the classification of raw and weighted vertebral column disorders’ dataset using k-NN, MLP, Naive Bayes, and SVM classifier algorithms with 10-fold cross validation.

<table>
<thead>
<tr>
<th>Performance metrics</th>
<th>10-fold cross validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k-NN classifier (for k = 1)</td>
</tr>
<tr>
<td></td>
<td>In raw dataset</td>
</tr>
<tr>
<td>Classification accuracy (%)</td>
<td>78.39</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>71.82</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>82</td>
</tr>
<tr>
<td>Precision</td>
<td>0.69</td>
</tr>
<tr>
<td>Recall</td>
<td>0.72</td>
</tr>
<tr>
<td>f-measure</td>
<td>0.70</td>
</tr>
<tr>
<td>AUC</td>
<td>0.835</td>
</tr>
<tr>
<td>Computation time (s)</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4
The obtained performance metrics in the classification of raw and weighted vertebral column disorders’ datasets using k-NN, MLP, Naive Bayes, and SVM classifier algorithms with the training-testing portion of 50–50.

<table>
<thead>
<tr>
<th>Performance metrics</th>
<th>With data partition of 50% test–50% training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k-NN classifier (for k = 1)</td>
</tr>
<tr>
<td></td>
<td>In raw dataset</td>
</tr>
<tr>
<td>Classification accuracy (%)</td>
<td>75.48</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>59.18</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>83</td>
</tr>
<tr>
<td>Precision</td>
<td>0.62</td>
</tr>
<tr>
<td>Recall</td>
<td>0.59</td>
</tr>
<tr>
<td>f-measure</td>
<td>0.60</td>
</tr>
<tr>
<td>AUC</td>
<td>0.812</td>
</tr>
<tr>
<td>Computation time (s)</td>
<td>0</td>
</tr>
</tbody>
</table>

4.2. Results

In the present paper, an innovative data weighting method called pairwise fuzzy C-means (FCM) based feature weighting method has been proposed for classifying the vertebral column disorders including disk hernia (DH), spondylolisthesis (SL) and

Table 5
The obtained performance metrics in the classification of pairwise weighted vertebral column disorders’ dataset using k-NN, MLP, Naive Bayes, and SVM classifier algorithms with 10-fold cross validation.

<table>
<thead>
<tr>
<th>The name of pair in dataset</th>
<th>Performance metrics</th>
<th>The used classifier algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification accuracy (%)</td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>DH-NO</td>
<td>91.875</td>
<td>94</td>
</tr>
<tr>
<td>DH-SL</td>
<td>95.71</td>
<td>98</td>
</tr>
<tr>
<td>SL-NO</td>
<td>88.8</td>
<td>88</td>
</tr>
<tr>
<td>Average value</td>
<td>DH-NO</td>
<td>92.12</td>
</tr>
<tr>
<td>SL-NO</td>
<td>90.00</td>
<td>91</td>
</tr>
<tr>
<td>Average value</td>
<td>DH-SL</td>
<td>99.04</td>
</tr>
<tr>
<td>SL-NO</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td>Average value</td>
<td>DH-NO</td>
<td>94.37</td>
</tr>
<tr>
<td>SL-NO</td>
<td>95.71</td>
<td>96</td>
</tr>
<tr>
<td>Average value</td>
<td>DH-SL</td>
<td>92</td>
</tr>
<tr>
<td>SL-NO</td>
<td>92.50</td>
<td>93</td>
</tr>
<tr>
<td>Average value</td>
<td>DH-SL</td>
<td>99.04</td>
</tr>
<tr>
<td>SL-NO</td>
<td>96</td>
<td>91</td>
</tr>
<tr>
<td>Average value</td>
<td>DH-SL</td>
<td>95.84</td>
</tr>
</tbody>
</table>
normal (NO) groups. The aim of this weighting method are both to improve the classification performance of classifier algorithms for multi-class datasets and to transform from non-linearly separable dataset to linearly separable dataset.

The class distribution of vertebral column disorders’ dataset has a non-linearly separable dataset. This case was shown in Fig. 6. To classify this dataset linearly, either the non-linear classifier algorithm should be used or the data or the feature transformation methods should be used prior to classifier algorithms. For an online automatic diagnosis system, the proposed feature weighting method and \( k \)-NN classifier algorithm could be easily combined to detect the disease using symptoms.

In this paper, we have proposed a hybrid system with two stages consisting of data pre-processing and classification. In the first stage, the vertebral column disorders’ dataset was weighted using the pairwise fuzzy C-means (FCM) based feature weighting method. And then, the weighted dataset with three classes was linearly classified using classifier algorithms including \( k \)-NN, MLP, SVM and Naive Bayes algorithms. In the training and testing phase of classifier algorithms, the 10-fold cross validation and the training and testing portion of 50–50% were used. Table 3 presents the obtained performance metrics in the classification of raw and weighted vertebral column disorders’ datasets using \( k \)-NN, MLP, Naive Bayes, and SVM classifier algorithms with 10-fold cross validation. Table 4 shows the obtained performance metrics in the classification of raw and weighted vertebral column disorders’ datasets using \( k \)-NN, MLP, Naive Bayes, and SVM classifier algorithms with the training–testing portion of 50–50%. Table 5 gives the obtained performance metrics in the classification of pairwise weighted vertebral column disorders’ dataset using \( k \)-NN, MLP, Naive Bayes, and SVM classifier algorithms with 10-fold cross validation. In the \( k \)-NN classifier, we have used the value of \( k \) as 1 in our experimental studies. The cause of choosing this value in \( k \)-NN classifier is that the best classification performance was obtained for \( k \) = 1 in the classification of vertebral column disorders’ dataset. Without the proposed feature weighting, the obtained \( f \)-measure values were 0.7738 for MLP classifier, 0.7021 for \( k \)-NN, 0.7263 for Naive Bayes, and 0.7298 for SVM classifier algorithms in the classification of vertebral column disorders’ dataset with three classes. With the pairwise fuzzy C-means based feature weighting method, the obtained \( f \)-measure values were 0.9509 for MLP, 0.9313 for \( k \)-NN, 0.9603 for Naive Bayes, and

**Table 6**
The obtained ROC curves for raw and weighted vertebral column disorders’ dataset using \( k \)-NN, MLP, SVM, and Naive Bayes classifier algorithms with 10-fold cross validation.

<table>
<thead>
<tr>
<th>The used classifier algorithm</th>
<th>In the weighted vertebral column disorders’ dataset</th>
<th>In the raw vertebral column disorders’ dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )-NN (for ( k = 1 ))</td>
<td><img src="image1.png" alt="ROC Curve" /></td>
<td><img src="image2.png" alt="ROC Curve" /></td>
</tr>
<tr>
<td>MLP</td>
<td><img src="image3.png" alt="ROC Curve" /></td>
<td><img src="image4.png" alt="ROC Curve" /></td>
</tr>
<tr>
<td>Naive Bayes</td>
<td><img src="image5.png" alt="ROC Curve" /></td>
<td><img src="image6.png" alt="ROC Curve" /></td>
</tr>
<tr>
<td>SVM</td>
<td><img src="image7.png" alt="ROC Curve" /></td>
<td><img src="image8.png" alt="ROC Curve" /></td>
</tr>
</tbody>
</table>

The experimental results demonstrated that the proposed pair- 
wise fuzzy C-means (FCM) based feature weighting proposed.

Table 7 demonstrates the obtained ROC curves for raw and 
weighted vertebral column disorders’ dataset using k-NN, MLP, 
SVM, and Naive Bayes classifier algorithms. In the ROC curves, 
if the area under the ROC curve is closer to 1 or equal to 1, the best 
classification performance will be obtained. The best performance 
for classification of vertebral column disorders’ dataset was with 
Naive Bayes classifier. Table 7 gives the performance comparison 
between our method and other methods used in the literature. 
As seen from the given results, the proposed feature weighting 
method has obtained the very promising results and this method 
could be safely used for detecting the vertebral column disorders 
with three classes. In the future, this method could be used 
confidently in the medical datasets with more classes.

5. Conclusions

Since the vertebral column disorders is a common disease that 
appear in public, to detect the these diseases, we have proposed 
a new hybrid scheme combining pairwise fuzzy C-means (FCM) 
based feature weighting method and classifier algorithms containing 
the k-NN, MLP, SVM, and Naive Bayes classifier algorithms. In this paper, the 
vertebral column disorders including disk hernia (DH), spondylo-
lothesis (SL) and normal (NO) groups have been linearly classified 
by the proposed method. The proposed feature weighting scheme 
is the improved version of FCM based feature weighting proposed. 
The experimental results demonstrated that the proposed pair-
wise fuzzy C-means based feature weighting method is robust and 
effective in the classification of the vertebral column disorders’ 
dataset. The best performance for the classification of vertebral column 
disorders’ dataset was with Naive Bayes classifier. In the 
future, this method could be used confidently in the medical 
datasets with more classes. In the constituting of new features 
regarding the vertebral column disorders, the image based features 
could be used and combined with patient information to 
detect the vertebral column disorders as a future work.

Conflict of interest statement

None.

Table 7
The performance comparison of our method and other methods used in the classification of vertebral column disorders.

<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Method</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Neto and Barreto [8]</td>
<td>SVM</td>
<td>82</td>
</tr>
<tr>
<td>2</td>
<td>Neto and Barreto [8]</td>
<td>MLP</td>
<td>83</td>
</tr>
<tr>
<td>3</td>
<td>Mattos and Barreto [9]</td>
<td>C-GRNN</td>
<td>81</td>
</tr>
<tr>
<td>4</td>
<td>Neto et al. [10]</td>
<td>C-SVM</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>Abdraabou [1]</td>
<td>C-MLP</td>
<td>88</td>
</tr>
<tr>
<td>6</td>
<td>Ansari et al. [12]</td>
<td>SOM</td>
<td>83</td>
</tr>
<tr>
<td>7</td>
<td>Our study</td>
<td>SVM (different kernels)</td>
<td>85</td>
</tr>
<tr>
<td>8</td>
<td>SVM</td>
<td>Hybrid CBR and ANN</td>
<td>85</td>
</tr>
<tr>
<td>9</td>
<td>SVM</td>
<td>GRNN</td>
<td>93.87</td>
</tr>
<tr>
<td>10</td>
<td>SVM</td>
<td>MLP</td>
<td>87.29</td>
</tr>
<tr>
<td>11</td>
<td>SVM</td>
<td>Naive Bayes</td>
<td>84</td>
</tr>
<tr>
<td>12</td>
<td>SVM</td>
<td>k-NN classifier</td>
<td>95.48</td>
</tr>
<tr>
<td>13</td>
<td>SVM</td>
<td>The combination of pairwise FCM based feature weighting and Naive Bayes</td>
<td>97.42</td>
</tr>
<tr>
<td>14</td>
<td>SVM</td>
<td>The combination of pairwise FCM based feature weighting and SVM</td>
<td>94.65</td>
</tr>
</tbody>
</table>

0.9486 for SVM classifier algorithms. The best method for this 
dataset was the combination of pairwise fuzzy C-means (FCM) 
based feature weighting and Naive Bayes classifier.

Acknowledgments

This study is supported by the Scientific Research Projects (BAP) 
of Amasya University.

References

desinification: a comparison of digital versus subjective assessments and digital 
[7] V. Vapnik, O. Chapelle, Bounds on error expectation for support vector 
disorders using machine learning classifiers, in: International Conference on 
Information Science and Applications (ICISA), 2013, pp. 1–6.
Sanches, and Mario Hernández (Eds.), Proceedings of the Fifth Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA11). Springer-
disorders using machine learning classifiers, in: International Conference on 
Information Science and Applications (ICISA), 2013, pp. 1–6.
[13] V. Vapnik, D. Chapelle, Bounds on error expectation for support vector 
2005.
[16] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 
1998.
Yavuz Unal has received his BSc and MSc degrees in electronic and computer education science program from Selcuk University in 2002 and 2006. He is a Ph.D. student at the Selcuk University in the Department of Computer Engineering. His research interests focus on image processing, machine learning and fuzzy logic.

Kemal Polat has graduated from Electrical–Electronics Engineering Department of Selcuk University with BSc degree in 2002 and from Electrical–Electronics Engineering Department of Selcuk University with MSc degree in 2004. He took his Ph.D. degree in Electrical and Electronic Engineering at Selcuk University in 2008. He is now working as Assistant Professor in Electrical and Electronic Engineering Department, Engineering of Faculty, Abant Izzet Baysal University since September, 2011. His current research interests are biomedical signal classification, statistical signal processing, digital signal processing, pattern recognition and classification. He has published 56 articles in the SCI journals and 16 international conference papers. He is an reviewer in various SCI journals including Information Sciences, Pattern Recognition Letters, Digital Signal Processing, Soft Computing, Artificial Intelligence in Medicine, Applied Soft Computing, Computers in Biology and Medicine, Expert Systems, Computational Statistics & Data Analysis, IEEE Transactions on Biomedical Engineering, IEEE Transactions on Evolutionary Computation, TALANTA, Soft Computing, Journal of Medical Systems and Scientific Research and Essays, so on. He is the member of editorial board of Journal of Neural Computing and Applications (SCI expanded).

Hasan Erdiç Koçer has received his BSc and MSc degrees in electronic and computer education science program from Marmara University and Selcuk University in 1998 and 2002, respectively, and the PhD degree from Selcuk University in 2010. He is also assistant professor at Selcuk University, Department of Computer Engineering. He has published more than five research papers in journals and peer reviewed conferences.