A Hybrid Fuzzy Clustering Approach for Diagnosing Primary Headache Disorder

Svetlana Simić[1], Zorana Banković[2], José R. Villar[3], Dragan Simić[4], Svetislav D. Simić[4]

[1]University of Novi Sad, Faculty of Medicine, Hajduk Veljkova 1–9, 21000 Novi Sad, Serbia
svetlana.simic@mf.uns.ac.rs
[2]Frontiers Media SA, Paseo de Castellana 77, Madrid, Spain
zbankovic@gmail.com
[3]University of Oviedo, Campus de Llamaquique, 33005 Oviedo, Spain
villarjose@uniovi.es
[4]University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia
dsimic@eunet.rs, dsimic@uns.ac.rs, simicsvetislav@uns.ac.rs

Abstract. Clustering is one of the most fundamental and essential data analysis tasks with broad application. It has been studied in various research fields: data mining, machine learning, pattern recognition, and in engineering, economic, and biomedical data analysis. Headache is not a disease which typically shortens one's life, but it can be a serious social as well as a health problem. Approximately 27 billion euros per year are lost through reduced work productivity in the European Community. This paper is focused on a new strategy based on a hybrid model for combining fuzzy partition method and maximum likelihood estimates clustering algorithm for diagnosing primary headache disorder. The proposed hybrid system is tested on two data sets for diagnosing headache disorder collected from Clinical centre of Vojvodina in Serbia.

Keywords. Data Clustering, Number of Clusters, Maximum Likelihood Estimates Clustering, Fuzzy Partition Method, Calinski-Harabasz index

1 Introduction

Clustering is one of the most fundamental and essential data analysis tasks with broad applications. It is a process in which a group of unlabeled patterns are partitioned into several sets so that similar patterns are assigned to the same cluster, and dissimilar patterns are assigned to different clusters. The purpose of clustering is to identify natural groupings of data from a large data set to produce a concise representation of a system's behavior.

The unsupervised nature of the problem implies that its structural characteristics are not known, except in case of domain knowledge available in advance. There are some goals for clustering algorithms: (1) estimate the optimal number of clusters, (2) determining good clusters and (3) doing so efficiently. One of the main difficulties for cluster analysis is estimating the optimal and correct number of clusters of different types of datasets.

Modern medicine generates a great deal of information stored in the medical database. Extracting useful knowledge and making scientific decision for diagnosis and treatment of disease from the database increasingly becomes necessary. Medical field is primarily directed at patient care activity and only secondarily as research resource. The only justification for collecting medical data is to benefit the individual patient.

Headache disorders are the most prevalent of all the neurological conditions and are among the most frequent medical complains seen in a general practice. More than 90% of the general population report experiencing a headache during any given year, which is a lifetime history of head pain [1]. Headache is not a disease which typically shortens one's life, but it can be a serious social as well as a health problem. Approximately 27 billion euros per year are lost through reduced work productivity in the European Community. The diagnostic criteria developed by the International Headache Society (IHS) have been extensively used in the epidemiological research [2], and some automatic methods, expert systems, knowledge–base systems such as the tools which help physicians to make diagnoses are developed.

This research is focused on diagnosing certain primary headache types in different population: age, type of employment, hospitalized or outpatients. Two different data sets for diagnosing headache disorder from Clinical centre of Vojvodina in Serbia are collected. This paper presents hybrid clustering approach for diagnosing primary headache disorder combining fuzzy partition method and maximum likelihood estimates clustering algorithm. Also, *Calinski-Harabasz index* is used to estimate the optimal and correct number of clusters. The proposed hybrid system is tested on these data sets and facilitated by the application of the IHS criteria for diagnosing primary headache disorder.

This paper is an extension of our previous research [3], and continuous the authors' previous research in computer-assisted diagnosis methods [4] [5] [6] [7] and applications for clustering methods presented in [8] [9] [10].

The rest of the paper is organized in the following way: Section 2 provides an overview of the basic idea on clustering and related work. Primary headache classification is shown in Section 3. Section 4 presents model for fuzzy clustering approach for diagnosing primary headache. The preliminary experimental results are presented in Section 5. Section 6 provides conclusions and some points for future work.

2 Clustering, Classification and Related Work

Clustering and classification are basic scientific tools used to systematize knowledge and analyze the structure of phenomena. Both techniques refer to the process of partitioning a set of objects into groups as dissimilar as possible from one another.

The conventional distinction made between clustering and classification is the following. Clustering is a process of partitioning a set of items into set of categories. Classification is a process of assigning a new item or observation to its proper place in an established set of categories [11]. In clustering, little or nothing is known about category structure, and the objective is to discover a structure that fits the observations. Classification is used mostly as a supervised learning method, but on the other side clustering is used for unsupervised learning. The goal of clustering is descriptive, that of classification is predictive.

2.1 Clustering

Clustering groups data instances into subsets in such a manner that similar instances are grouped together, while different instances belong to different groups. The instances are thereby organized into an efficient representation that characterizes the population being sampled.

Formally, the clustering structure is represented as a set of subsets $C = C_1, ..., C_k$ of S, such that: $S = \bigcup_{i=1}^{k} C_i$ and $C_i \cap C_j = 0$ for $i \neq j$. Consequently, any instance in S belongs to exactly one and only one subset.

Clustering of objects is as ancient as the human need for describing the salient characteristics of men and objects and identifying them with a type. Therefore, it embraces various scientific disciplines: from mathematics and statistics to biology and genetics, each of which uses different terms to describe the topologies formed using this analysis. From biological ”taxonomies”, to medical ”syndromes” and genetic ”genotypes” to manufacturing ”group technology” — the problem is identical: forming categories of entities and assigning individuals to the proper groups within it [12].

Cluster analysis, an important technology in data mining, is an effective method of analyzing and discovering useful information from numerous data. Cluster algorithm groups the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another but are very dissimilar to objects in other clusters.

General references regarding data clustering is presented in [13]. A very good presentation of contemporary data mining clustering techniques can be found in the textbook [14].

2.2 Related Work in Primary Headache

In the past decades, many approaches have been proposed to solve clustering problem in medical data to help physicians to make decision regarding patient illness and future treatments. Structured diagnostic interviews were conducted on 443 headache sufferers from a community sample and *hierarchical cluster analysis* of symptoms in both sub-samples revealed two distinct clusters: (1) unilateral pulsating pain, photophobia and phonophobia; (2) bilateral pressing/tightening pain, mild to moderate intensity, and absence of nausea/vomiting [15]. These clusters indicate that headache symptoms cluster empirically in a manner consistent with IHS criteria for migraine and tension-type headaches, respectively. Also, criterion overlap problems regarding pain intensity and duration were identified.

A new migraine analysis method was proposed by using electroencephalography (EEG) signals under flash stimulation in time domain. These types of signals are commonly pre-processed before the analysis procedure, and pre-processing techniques affect the analysis results. *Histogram differences* in the case of flash stimulation calculated and used as features for the healthy subjects and migraine patients. These features are applied to a *k-means* clustering algorithm to see clustering results of the proposed technique. *Silhoutte* clustering results show that, a good clustering performance is evaluated as 86.6% *correct clustering rate* (CCR) in migraine patients [16].

In the research [17], it was a goal to evaluate the *classification* accuracy of the ant colony optimization (ACO) algorithm for the diagnosis of primary headaches using a website questionnaire expert system on headache diagnosis that was completed by patients. The cross-sectional study was conducted in 850 headache patients who randomly applied to hospitals from three cities in Turkey. Finally, neurologists' diagnosis results were compared with the classification results. The ACO for diagnosis classified patients with 96.9% overall accuracy. Diagnosis accuracies of migraine, tension-type headache (TTH), and cluster headaches were 98.2%, 92.4%, and 98.2% respectively. The headache diagnosis using a website-based algorithm is useful for neurologists in order to gather quick and precise results as well as tracking patients for their headache symptoms.

The use of machine learning is recruited for the *classification* of primary headache disorders, for which a dataset of 832 records of patients with primary headaches was considered, originating from three medical centers located in Turkey is presented [18]. Three main types of primary headaches were derived from the data set including TTH in both episodic and chronic forms, migraine without aura, migraine with aura, and cluster headache. Six popular machine-learning based classifiers, including linear and non-linear ensemble learning, in addition to one regression-based procedure, have been evaluated for the classification of primary headaches within a supervised learning setting, achieving highest aggregate performance outcomes of *Area Under the ROC Curve* 92.3 %, *sensitivity* 89.7 %, and overall classification *accuracy* of 84.3 %.

To evaluate diagnosis accuracy of Artificial Immune Systems (AIS) algorithms for *classification* of migraine, tension-type and cluster-type of headaches by using the website-based diagnosis survey expert system. The headache diagnoses of 850 patients from three different cities in Turkey were evaluated by using AIS algorithms and it is presented in [19]. It is possible to classify primary headaches with AIS algorithm and helpful for neurologist in order to obtain precise results as well as easy information sharing. According to the results, AIS algorithms for diagnosis have the maximum *accuracy* of 71%.

One of the leading reasons that make migraine a bigger issue is that it cannot be diagnosed easily by physicians because of the numerous overlapping symptoms with other diseases, such as epilepsy and tension-type headache [20]. Flash stimulation is used during the recording of EEG signals. To achieve this, different machine learning algorithms on the EEG signals features extracted by using discrete wavelet transform are tested. The real-world dataset, recorded in the laboratory, show that the flash stimulation can improve the classification accuracy for more than 10%.

According to previous related work, it can be concluded that *accuracy* is higher in implementation classification methods then clustering methods, but in the real-world setting, the physicians do not know type of primary headache in advance.

3 Primary Headache Classification

The International Classification of Headache Disorders – The Third Edition (ICHD-3) established the uniform terminology and consistent operational diagnostic criteria for a wide range of the headache disorders around the world [2]. The ICHD-3 provides a hierarchy of diagnoses with varying degrees of specificity. Headache disorders are identified with three or sometimes five-digit codes which is, in details, presented in short identification for just two important digit codes in Table 1. All headache disorders are classified into two major groups: **A**) *Primary headaches from ICHD-3 code 1. to 4.* and **B**) *Secondary headaches ICHD-3 code from 5. to 12*. The first digit specifies the major diagnostic categories (i.e. *Migraine*). The second digit indicates a disorder within the category (i.e. *Migraine without aura*). Each category is then subdivided into groups, types, subtypes and sub-forms. Subsequent digits permit more specific diagnosis for some headache types.

Table 1. The International Classification of Headache Disorders – the Third Edition [2]

	ICHD-3 code		Diagnosis - Primary headache disorders
A		1.1	Migraine without aura
		1.2	Migraine with aura
		¦	
		1.6	Episodic syndromes that may be associated with migraine
	2.		**Tension-type headache (TTH)**
		2.1	Infrequent episodic tension-type headache
		¦	
		2.4	Probable tension-type headache
	3.		**Trigeminal autonomic cephalalgias (TACs)**
	4.		**Other primary headache disorders**
B	**5.**		**Secondary headache disorders**
	¦		
	12.		

When first meeting a patient, physicians who are more concerned with the detailed anamnesis and clinical examinations, apply ICHD-3 criteria and can easily establish the primary headache diagnosis. If the criteria are not satisfied, the physicians will have to suggest an additional examination to a patient.

The study [21] analyzes different studies which are all based on IHS recommendations. These studies deal with different approaches to attribute selection based on automatic methods, expert systems, knowledge–base systems and physicians' expert knowledge as well, as shown in Table 2. Feature selection could be divided on *Stochastic* and *no-Stochastic Feature Selection* methodology, a refinement of an initial stochastic feature selection task with a no-stochastic method to reduce a bit more the subset of features to be retained [22]. The study [21] shows that the most important features are: (4), (5), (6), (7), (8), (10), (12), (13) and (15) which are signed with **red bold** colour. These nine features are used in the rest of the research.

Table 2. Comparison of a selection attribute for primary headache based on IHS diagnostic criteria: **1.** Consistency measure filter, **2.** ReliefF Greedy, **3.** ReliefF top10, **4.** Genetic algorithm wrapper, **5.** ACO based classification algorithm, **6.** Rule Based Fuzzy Logic System, **7.** Physician's expert choice, **8.** column RES – final decision for important attributes selection

	Attributes	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**RES**
1.	Sex								
2.	1. How old were you when the headache occurred for the first time?								
3.	2. How often do you have headache attacks?	+							
4.	3. How long do the headache attacks last?	+	+	+	+		+	+	+
5.	4. Where is headache located?	+	+		+	+	+	+	+
6.	**5. How intense is the pain?**	+	+	+	+	+	+	+	+
7.	6. What is the quality of the pain you experience	+	+	+	+	+	+	+	+
8.	**7. Do your headaches worsen after physical activities such as walking?**	+	+	+	+		+	+	+
9.	8. Do you avoid routine physical activities because you are fear they might trigger your headache?		+	+				+	
10.	9.a) Are the headaches accompanied by? a) Nausea	+	+	+	+		+	+	+
11.	9.b) Are the headaches accompanied by? b) Vomiting		+	+			+	+	
12.	9.c) Are the headaches accompanied by? c) Photophobia	+	+	+	+	+	+	+	+
13.	9.d) Are the headaches accompanied by? d) Phonophobia		+	+		+	+	+	+
14.	10. Do you have temporary visual, sensory or speech disturbance?								
15.	11. Do you, during a headache attack, have tension and/or heightened tenderness of head or neck muscles?	+	+	+	+			+ TTH	+
16.	12. Do you have any body numbness or weakness?								
17.	13. Do you have any indications of oncoming headache?	+	+		+				
18.	14. Headache is usually triggered by: Menstrual periods	+							
19.	15. In the half or my visual field, lasting 5 minutes to an hour, along with the headache attack or an hour before.								
20.	16. Along with the headache attack or an hour before one I have sensory symptoms.								

4 Modeling the Fuzzy Clustering Approach

In general, clustering algorithms can be grouped on given data set into clusters in two main different approaches:

- *Hard clustering*: each object belongs to specific cluster or not
- *Soft clustering* also named - *fuzzy clustering* - each object belongs to each cluster to a certain degree.

The primary representative *hard clustering partitioning methods* are: *k*-means, *k*-medoids, *k*-medians, *k*-means++. On the other hand, the representative *fuzzy partitioning methods* are: *fuzzy c-means clustering* method, *fuzzy* Gustafson-Kessel clustering method, *fuzzy* Gath-Geva clustering method. In this research *fuzzy maximum likelihood estimates with a direct distance norm* based on the *fuzzy Gath-Geva clustering method* (FGGC) is used. *Fuzzy maximum likelihood estimates with a direct distance norm* belongs to fuzzy partitioning methods [23].

4.1 Optimal Number of Clusters

The concept of dense and well-separated clusters, *Calinski-Harabasz index* is used to estimate the optimal number of clusters, by using two measures known as: the *Variance Ratio Criterion* and *Total within Sum of Squares,* for choosing the suitable *c*, number of clusters. To build *Calinski-Harabasz index*, it is first necessary to define the inter cluster dispersion [24]. When *N* the total number of observations is known, (data points), *c* number of clusters with their relative centroids and the global centroid, the *inter-cluster dispersion B(c)* (between cluster variation) is defined as:

$$B(c)=\sum_{t}^{N} n_t\,(\mu-\mu_i)^T\,(\mu-\mu_j) \tag{1}$$

In the above expression, n_t is the number of elements belonging to the cluster *c*, μ is the global centroid, μ_i is the centroid of cluster *i*, and μ_j is the centroid of cluster *j*.

The intra-cluster dispersion *W(c)* is defined as, within cluster variation:

$$W(c)=\sum_{t=1}^{C}\sum_{x\in C_N}^{N}(x-\mu_t)^T(x-\mu_t) \tag{2}$$

The Calinski-Harabasz index is defined as the ratio between *B(c)* and *W(c)*:

$$CH(c)=\frac{N-c}{c-1}\times\frac{B(c)}{W(c)} \tag{3}$$

The *Calinski–Harabasz index* is based on comparing the weighted ratio of the between cluster sum of squares (the measure of cluster separation) and the within cluster sum of squares (the measure of how tightly packed the points are within a cluster). For a low intra-cluster dispersion and a high inter-cluster dispersion, it is needed to find the number of clusters that maximizes this index. Ideally, the clusters should be well separated, so the between cluster sum of squares value should be large, but points within a cluster should be as close as possible to one another, resulting in smaller values of the within cluster sum of squares measure [24].

The decision to assign a point to a cluster depends only on its features and sometimes on the position of a set of other points. But also, there are different algorithms which are based on alternative strategies to solve this problem and can yield very different results. The technique *Improved Covariance Estimation for Gustafson-Kessel Clustering* algorithm is employed in the extraction of the rules from data and estimation of the *optimal number of clusters* for *fuzzy partitioning methods*. It calculates seven different coefficients to estimate the optimal number of clusters: *Partition*

Coefficient, *Classification Entropy*, *Partition Index*, *Separation Index*, *Xie and Beni Index*, *Dunn Index*, *Alternative Dunn Index* [25].

4.2 Fuzzy Partition Method

The data set is typically an observation of some physical process. Each observation consists of *n* measured variables, grouped into an *n*-dimensional row vector $x_k = [x_{k1}, x_{k2}, \ldots, x_{kn}]^T$, $x_k \in \mathbb{R}^n$. A set of *N* observations is denoted by $X = \{ x_k \mid k = 1, 2, \ldots, N\}$, and is represented as an *N* x *n* matrix, a data set. Since clusters can formally be viewed as subsets of the data set, the number of subsets (clusters) is denoted by *c*. Fuzzy partition can be seen as a generalization of hard partition, it allows μ_{ik} to attain real values in [0, 1]. A *N* x *c* matrix $U = [\mu_{ik}]$ represents the fuzzy partitions, its conditions are given by:

$$\mu_{ij} \in [0, 1],\ 1 \le i \le N,\ \ 1 \le k \le c \tag{4}$$

$$\sum_{k=1}^{c} \mu_{ik} = 1,\ 1 \le i \le N \tag{5}$$

$$0 < \sum_{i=1}^{N} \mu_{ik} < N,\ 1 \le k \le c \tag{6}$$

Let $X = [x_1, x_2, \ldots, x_N]$ be a *finite set* and let $2 \le c < N$ be an integer. The *fuzzy partitioning space* for *X* is the set

$$M_{fc} = \left\{ U \in \Re^{N x c} \mid \mu_{ik} \in [0,1], \forall i,k; \sum_{k=1}^{c} \mu_{ik} = 1, \forall i;\ 0 < \sum_{k=1}^{c} \mu_{ik} < N, \forall k \right\}. \tag{7}$$

The *i*-th column of *U* contains values of the *membership function* of the *i*-th fuzzy subset of *X*. The equation (5) constrains the sum of each column to 1, and thus the total membership of each x_k in *X* equals one. The distribution of memberships among the *c* fuzzy subsets is not constrained.

4.3 Fuzzy Maximum Likelihood Estimates Clustering Algorithm

The basic steps of the proposed hybrid algorithm for the *fuzzy maximum likelihood estimates* (FMLE) clustering algorithm which employs a distance norm based on fuzzy maximum likelihood estimates proposed in [26], are summarized by the pseudo code shown in Algorithm 1.

In consistence with the theory, notice in previous subsection, *Fuzzy Partition Method*, there is a set of data *X* specify *c*, choose a *weighting exponent* $m > 1$ and a termination tolerance $\varepsilon > 0$.

Initialize the partition matrix with a more robust method. It is important to mention that in *Step 3*. the distance to the cluster center (centroid) is calculated on the basis of the fuzzy covariance matrices of the cluster.

Algorithm 1: *The algorithm – fuzzy Gath-Geva clustering method – for* ***Fuzzy Maximum Likelihood Estimates***

Begin

Step 1: --- ***Initialization.***

X; c; $m > 1$; $\varepsilon > 0$

Step 2: --- ***Calculate the cluster centers.***

Repeat for *l=1, 2, ...*

$$v_i^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^{w x_k}}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^{w}}, 1 \le i \le c$$

Step 3: --- ***Compute the distance measure*** D_{ik}^2

$$F_i^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^{w} (x_k - v_i^{(l)})(x_k - v_i^{(l)})^T}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^{w}}, 1 \le i \le c$$

--- ***The distance function is chosen as***

$$D_{ik}^2(x_k, v_i) = \frac{(2\pi)^{(\frac{n}{2})} \sqrt{\det(F_i)}}{\alpha_i} \exp(\frac{1}{2}(x_k - v_i^l)^T \ F_i^{-1} (x_k - v_i^l))$$

--- **with the *a priori* probability**

$$\alpha_i = \frac{1}{N} \sum_{k=1}^{1} \mu_{ik}$$

Step 4: --- ***Update the partition matrix***

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} (D_{ik}(x_k, v_i) / (D_{jk}(x_k, v_j))^{2/(m-1)}}, 1 \le i \le c, 1 \le k \le N$$

Step 5: ***Until*** $\| U^{(l)} - U^{(l-1)} \| < \varepsilon$

End.

4.4 Implemented Model for Diagnosing Primary Headache Disorder

The proposed hybrid model for Diagnosing Primary Headache Disorder implemented in this research is presented in Fig. 1. It has of two phases. First phase includes: (1) estimate the optimal and correct number of clusters of input data set; (2) fuzzy partitioning step where input data set is divided in two classes, but only one of them is appropriate for further analysis, and it is called *Selected data*.

Our previous research [3] has shown that FGGC *method – for Fuzzy Maximum Likelihood Estimates* fits much better when there are only two clusters to distinguish. Therefore, it could be considered that *Selected data* has only two clusters.

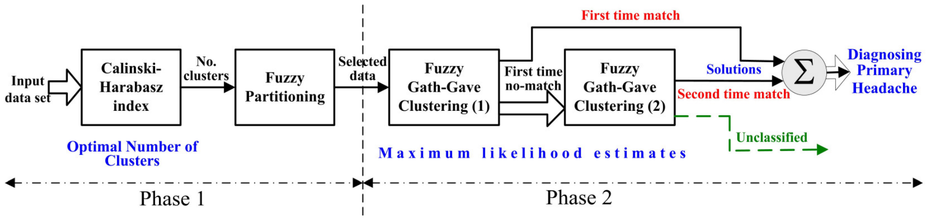


Fig. 1. A hybrid model for *Diagnosing Primary Headache Disorder*

Selected data are given according to the appropriate value in questions 6 and 8, which are marked in **black bold**, from Table 2. The second phase is realized in two steps of *fuzzy Gath-Geva clustering*. In the first step, the patients whose diagnosis undoubtedly confirms types of primary headache are selected and they are called *First Time Match Patients* marked in **red bold** on Fig. 1. On the other hand, there are *First Time No-Match Patients* and it is input in second step of *fuzzy Gath-Geva clustering*. The second step also creates new clusters for same types of primary headaches, but number of these patients is much smaller then the number of patients in the first phase, and they are called *Second Time Match Patients* marked in **red bold**. The rest of patient are "*unclassified*" and they are marked in **green bold**. Both *Match Patients* are "summarized" and they present *Diagnosed patients with Primary Headache* with an appropriate type.

5 Experimental Results

The proposed hybrid fuzzy maximum likelihood estimates clustering algorithm was further on, in our research, tested on two data sets for diagnosing headache disorder collected from Clinical centre of Vojvodina in Serbia. *Headache Data Set 1*, is a part of large study [27], encompassing adult working population. *Headache data set 2* presents a part of the study encompassing student population [28].

As mentioned before, this research uses the most important features: (4), (5), (6), (7), (8), (10), (12), (13) and (15) defined in Table 2. All headache data sets have nine features – attributes; four classes – types of primary headache: *Migraine without aura* (MWoA), *Migraine with aura* (MWA), *Tension-type headache* (TTH), *Other primary headaches* (Other); missing data – *No*.

5.1 Experimental results *Headache Data Set 1*

The *Input Data Set* is *Headache Data Set 1* and consists of 579 instances. Calculating maximum *Calinski-Harabasz value,* is 1250 and optimal number of clusters is 4 (Fig. 2). After *Fuzzy Partition* process there are 289 *Selected Data*. Pairwise comparison classes (MWoA – Other) for *Headache Data Set 2* is used in the first step of FGGC. After the first step FGGC, 205 instances are *First Time Match*, 78 patients suffer from MWoA, 127 are categorized as Other, and the remaining 84 are *First Time No-Match*. After 104 iterations, Centroid 1: 2.26, 1.22; Centroid 2: 1.65, 1.71 given in Fig. 3 (a).

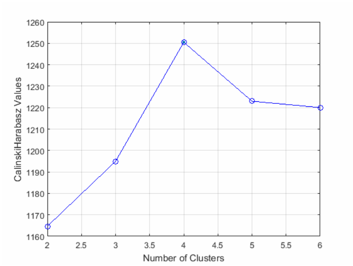
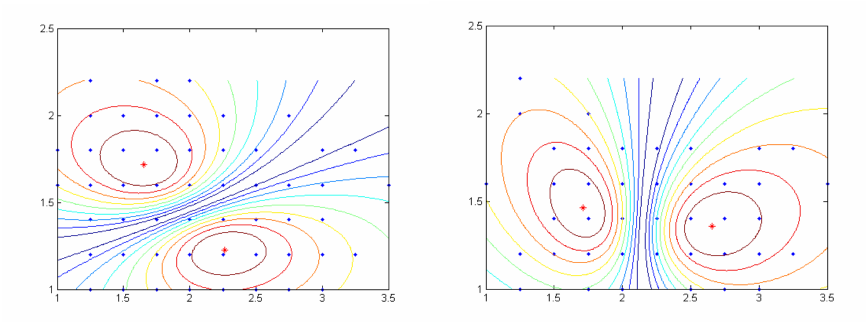


Fig. 2. Calinski-Harabasz *Index* for *Headache Data Set 1* to estimate the optimal number of clusters



(a) (b)

Fig. 3. *Hybrid fuzzy maximum likelihood estimates clustering* presented with two clusters; after first step of *fuzzy Gath-Geva clustering*; (b) after second step of *fuzzy Gath-Geva clustering*

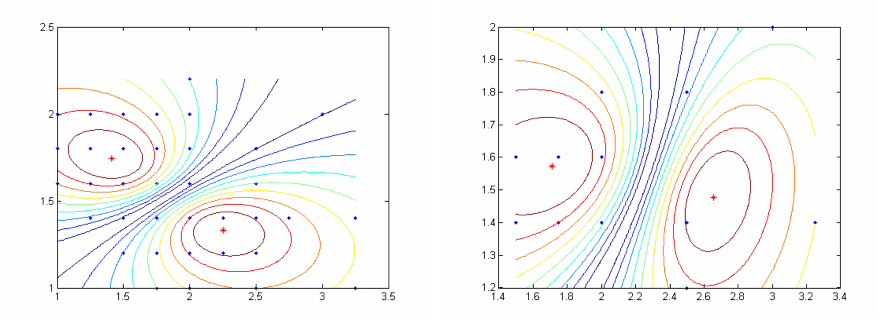
Table 3. Pairwise comparison for *Headache Data Set* 1

Com.	MWoA	MWA	TTH	Other
1	(78+11)=89/103 **83.39 %**			(127+25)=152/186 **81.72 %**
2	(91+1)=92/103 **89.32 %**		(151+24)=175/224 **78.13 %**	
3		(56+0)=56/66 **84.85 %**		(130+56)=147/186 **79.03 %**
4		(54+3)=57/66 **84.85 %**	(153+35) = 188/224 **83.94 %**	
Total	**103**	**66**	**224**	**186**
Average Accuracy			**82.87 %**	

After the second step of FGGC, 36 instances are *Second Time Match Patients*, 11 patients are found to have MWoA, 25 are categorized as Other, and the remaining 48 instances are *Unclassified*. After 16 iterations, Centroid 1: 2.65, 1.35; Centroid 2: 1.71, 1.46; Fig. 3 (b). Average accuracy for whole *Headache Data Set 1* in parwise comparison is 82.87 %. These presented experimental results are only for parwise comparison between MWoA and Other headache, and the rest could be discussed in the same manner.

5.2 Experimental results *Headache Data Set 2*

The *Headache Data Set 2* consists of 132 instances. After *Fuzzy Partition* process are selected 97 *Selected Data*. Pairwise comparison classes (MWoA – Other) for *Headache Data Set 2* is used in the first step of FGGC, after 190 iterations, 74 instances are *First Time Match*, 33 patients suffer from MWoA, 41 are categorized as Other, and the reamining 23 are *First Time No-Match*. The cluster centroids are: Centroid 1: 2.25, 1.32; Centroid 2: 1.41, 1.74 given in Fig. 4 (a).



(a) (b)

Fig. 4. Hybrid *fuzzy maximum likelihood estimates clustering* presented with two clusters; after first step of *fuzzy Gath-Geva clustering*; (b) after second step of *fuzzy Gath-Geva clustering*

Table 4. Pairwise comparison for *Headache Data Set 2*

Com.	MWoA	MWA	TTH	Other
1	(33+0)=33/46 **71.74 %**			(41+4)=45/51 **88.24 %**
2	(32+1)=33/46 **71.74 %**		(24+0)=24/28 **85.71 %**	
3		(7+0)=7/7 **100.00 %**		(38+6)=42/51 **82.35 %**
4		(7+0)=7/7 **100.00 %**	(18+6) = 24/28 **85.71 %**	
Total	**46**	**7**	**28**	**51**
Average Accuracy			**85.68 %**	

After the second step of FGGC, 4 instances are *Second Time Match Patients*, 0 patients are found to suffer from MWoA, 4 are categorized as Other, and rest 20 instances are *Unclassified*. After 35 iterations, the cluster centroids are: Centroid 1: 2.65, 1.47; Centroid 2: 1.70, 1.57; Fig. 4 (b). Average accuracy for whole *Headache Data Set 2* in parwise comparison is 85.68 %. These presented experimental results are only for parwise comparison between MWoA and Other headache, and the rest could be discussed in the same manner. And finally, 109 instances out of 132 instances in *Headache Data Set 2* have been correctly evaluated. *Total accuracy* is 82.58 %.

6 Conclusion and Future Work

The aim of this paper is to propose the new hybrid strategy for fuzzy clustering approach for diagnosing primary headache disorder. First, the algorithm employs the model to estimate the *optimal number of clusters* using *Calinski-Harabasz Index*. The new proposed hybrid approach is obtained by combining *fuzzy partition* method and *fuzzy Gath-Geva clustering* algorithm with a distance norm based on fuzzy maximum likelihood estimates.

Preliminary experimental results encourage further research by the authors because both experimental data sets in domain of primary headache have accuracy: the first has the *total accuracy* of 83 % and the second has the *total accuracy* of 86 %. Our future research will focus on creating new hybrid model combined with evolutionary techniques which will efficiently solve different well-known data sets and also real-world medical data sets.

References

1. Hagen, K., Zwart, J-A., Vatten L., Stovner, L. J., Bovin, G.: Prevalence of migraine and non-migrainous headache – head-HUNT, a large population-based study. Cephalalgia, Vol. 20, No. 10, pp. 900–906 (2000)
2. The International Classification of Headache Disorders 3rd edition. https://www.ichd-3.org/
3. Simić, S., Banković, Z., Simić, D., Simić, S. D.: A hybrid clustering approach for diagnosing medical diseases. LNCS (LNAI), Vol. 10870, pp 741-752 (2018)
4. Krawczyk, B., Simić, D., Simić, S., Woźniak, M.: Automatic diagnosis of primary headaches by machine learning methods. Open Medicine, Vol. 8, No. 2, pp. 157-165 (2013)
5. Simić, S., Simić, D., Slankamenac, P., Simić-Ivkov, M.: Computer-assisted diagnosis of primary headaches. International Workshop on Hybrid Artificial Intelligence Systems, pp. 314–321 (2008)
6. Simić, S., Simić, D., Slankamenac, P., Simić-Ivkov, M.: Rule-based fuzzy logic system for diagnosing migraine. LNCS, Vol. 5138, pp. 383–388, Springer, Berlin, Heidelberg (2008)
7. Jackowski, K., Jankowski, D., Simić, D., Simić, S.: Migraine diagnosis support system based on classifier ensemble. Advances in Intelligent Systems and Computing, Vol, 311, pp 329–339 Springer, Cham (2014)
8. Simić, D., Ilin, V., Tanackov, I., Svirčević, V., Simić, S.: A hybrid analytic hierarchy process for clustering and ranking best location for logistics distribution center. Hybrid Artificial Intelligent Systems, LNCS, Vol. 9121, pp. 477–488, Springer (2015)
9. Simić, D., Ilin, V., Svirčević, V., Simić, S.: A hybrid clustering and ranking method for best positioned logistics distribution centre in Balkan Peninsula. Logic Journal of the IGPL, Vol. 25, No. 6, pp. 991–1005 (2017)
10. Simić, D., Jackowski, K, Jankowski, D., Simić, S.: Comparison of clustering methods in cotton textile industry. Intelligent Data Engineering and Automated Learning, LNCS, Vol. 9375, pp. 501–508, Springer (2015)
11. Anderberg, M. R.: Cluster Analysis for Applications. Academic Press (1973)
12. Oded, M., Lior, R.: Clustering methods. In: Data Mining and Knowledge Discovery Handbook, Oded, M., Lior, R. (Eds.), pp. 321–352, Springer (2005)
13. Jain, A., Dubes, R.: Algorithms for Clustering Data. Prentice-Hall, NJ (1988)

14. Han, J., Kamber, M: Data Mining. Morgan Kaufmann Publishers (2001)
15. Bruehl, S., Lofland, K. R., Semenchuk, E. M., Rokicki, L.A., Penzien, D.B.: Use of cluster analysis to validate IHS diagnostic criteria for migraine and tension-type headache. Headache, Vol. 39, Issue 3, pp. 181–189 (1999)
16. Alkan, A., Akben, S. B.: Use of k-means clustering in migraine detection by using EEG records under flash stimulation. International Journal of the Physical Sciences, Vol. 6, No. 4, pp. 641–650 (2011)
17. Celik, U., Yurtay, N.: An ant colony optimization algorithm-based classification for the diagnosis of primary headaches using a website questionnaire expert system. Turkish Journal of Electrical Engineering and Computer Sciences, Vol. 25, No. 5, pp. 4200–4210 (2017)
18. Kaky, A. J. M. Intelligent Systems Approach for Classification and Management of Patients with Headache. Ph.D. thesis, Liverpool, John Moores University (2017)
19. Celik, U., Yurtay N., Koc E. R., Tope V., Gulluoglu, H., Ertas, M.: Diagnostic accuracy comparison of artificial immune algorithms for primary headaches. Computational and Mathematical Methods in Medicine, Vol. 2015, Article ID 465192 (2015)
20. Subasi, A., Ahmed, A., Aličković, E., Hassan, A. R.: Effect of photic stimulation for migraine detection using random forest and discrete wavelet transform. Biomedical Signal Processing and Control, Vol. 49, March 2019, pp. 231–239 (2019)
21. Simić, S., Banković, Z., Simić, D., Simić, S. D.: Different approaches of data and attribute selection on headache disorder. LNCS, Vol. 11315, pp 241–249, Springer, Cham (2018)
22. Tallon-Ballesteros, A. J., Correia, L., Cho S-B.: Stochastic and non-stochastic feature selection. LNCS, Vol. 10585, pp. 592–598, Springer, Cham (2017)
23. Gath, I., Geva, A. B.: Unsupervised optimal fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, Issue 7, pp. 773–780 (1989)
24. Bonaccorso, G.: Machine Learning Algorithms. Packt Publishing (2nd edition) (2018)
25. Babuška, R., van der Veen, P. J., Kaymak, U.: Improved covariance estimation for Gustafson-Kessel clustering. IEEE International Conference on Fuzzy Systems, pp. 1081–1085 (2002)
26. Bezdek, J. C., Dunn, J. C.: Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions. IEEE Transactions on Computers, Vol. C-24, Issue 8, pp. 835–838 (1975)
27. Simić, S.: The effect of migraine and tension-type headache on life and work of a working population ("*Uticaj migrene i glavobolje tenzionog tipa na život i rad radno aktivne populacije*") (*Serbian*), Ph.D. thesis, University of Novi Sad, Faculty of Medicine (2009)
28. Cakić, R.: Investigating the presence of neuropathic pain component in migraine and tension- type headache in students ("*Ispitivanje prisustva neuropatske komponente bola kod studenata obolelig od migrene i glavobolje tenzionog tipa*") (*Serbian*), Master thesis, University of Novi Sad, Faculty of Medicine (2013)