
A hybrid fuzzy clustering approach for diagnosing primary headache disorder

SVETLANA SIMIĆ*, *University of Novi Sad, Faculty of Medicine, Hajduk Veljkova 1–9, 21000 Novi Sad, Serbia.*

ZORANA BANKOVIĆ**, *Frontiers Media SA, Paseo de Castellana 77, Madrid, Spain.*

JOSÉ R. VILLAR†, *University of Oviedo, Campus de Llamaquique, 33005 Oviedo, Spain.*

DRAGAN SIMIĆ††, *University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia.*

SVETISLAV D. SIMIĆ§, *University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia.*

Abstract

Clustering is one of the most fundamental and essential data analysis tasks with broad applications. It has been studied in various research fields: data mining, machine learning, pattern recognition and in engineering, economics and biomedical data analysis. Headache is not a disease that 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 estimation 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 estimation 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 behaviour.

*E-mail: svetlana.simic@mf.uns.ac.rs

**E-mail: zbankovic@gmail.com

†E-mail: villarjose@uniovi.es

††E-mail: dsimic@eunet.rs, dsimic@uns.ac.rs

§E-mail: simicsvetislav@uns.ac.rs

2 A hybrid fuzzy clustering approach for diagnosing primary headache disorder

The unsupervised nature of the problem implies that its structural characteristics are not known, except in the case of domain knowledge available in advance. There are some goals for clustering algorithms: (1) estimating 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 the correct number of clusters of different types of data sets.

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 becomes increasingly necessary. Medical field is primarily directed at patient care activity and only secondarily as a 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 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 that 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-based systems such as the tools that 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 estimation 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 continues the authors' previous research in computer-assisted diagnosis methods [4–6] and industrial applications for clustering methods presented in [7, 8].

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 a set of categories. Classification is a process of assigning a new item or observation to its proper place in an established set of categories [9]. 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 hand 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 $C = C_1, \dots, C_k$ of S , such that: $S = \bigcup_{i=1}^k C_i$ and $C_i \cap C_j = \emptyset$ 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 [10].

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 [11]. A very good presentation of contemporary data mining clustering techniques can be found in the textbook [12].

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. One of the first studies that contains the analyses of 726 headache patients dates back to 1982, and it is presented in [13]. In that research paper, cluster analysis was used to find groups (clusters) of patients with similar symptoms. After the clusters were formed, the frequency of each symptom in each cluster was tabulated. Twelve physicians: eight internists, a neurologist, a pediatrician, a cardiologist and a pathologist were then asked to use the symptom frequencies to give a name to each cluster and to prescribe a typical therapy for each cluster. The ‘IF-THEN’ rule algorithm was developed and tested, and 92.3% of the headache sufferers are in the correct cluster. Only 56 (7.7%) patients are misclassified.

The estimation of the prevalence of headaches and relationship of headache symptoms with severity and duration of attacks are presented in research paper [14]. The data is collected by telephone interviews that were carried out among 10,169 subjects and analyzed to determine whether symptoms clustered into specific types of headache. The reports of symptoms are all binary (present/absent); therefore, the statistical requirements for parametric analyses precluded strategies based on correlations and commonly used clustering algorithms. An iterative solution was used to represent proximity clustering matrices for solution in n dimensional space, and the proximity clustering matrices are described in detail in [15].

Responses of 150 headache-prone subjects: 49 migraine and 101 tension-type headache (TTH) were examined [16]. Using a cluster analysis, the adjectives were grouped into 7 clusters including 5 sensory and 2 affective sub-groups. Headache was most commonly described in terms of clusters that reflected discomfort and aching pain sensations. Migraine and TTH sufferers did not differ markedly in pain quality but the intensity of pain differentiated these groups. A complete linkage cluster analysis, one of several methods of agglomerative hierarchical clustering, was carried out to examine the relationship between descriptors. Experimental result shows correspondence of migraine and

4 A hybrid fuzzy clustering approach for diagnosing primary headache disorder

TTH in two new cluster groups derived using cluster analysis: migraine, 93 of 101 (92.1%); TTH, 49 of 49 (100%); and misclassified, 8 (5.33%) patients, and diagnosis average accuracy is 94.67%.

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 [17]. These clusters indicate that headache symptoms cluster empirically in a manner consistent with IHS criteria for migraine and TTH, respectively. Also, criterion overlap problems regarding pain intensity and duration were identified.

A new migraine analysis method was proposed by using electroencephalography 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. *Silhouette* clustering results show that a good clustering performance is evaluated as 86.6% *correct clustering rate* in migraine patients [18].

The ant colony optimization classification algorithm for the diagnosis of primary headaches using a website questionnaire expert system has shown overall diagnosis accuracy with 96.9% [19]. To evaluate diagnosis accuracy of artificial immune systems algorithms for classification of migraine, TTH and cluster headache have the maximum accuracy of 71% [20].

On the other hand, there are many applied statistical methods in diagnosis primary headache disorder for analyzing different features and variables: for using descriptive statistics with categorical data presented using frequencies and percentages, continuous descriptive data are presented using means and standard deviations, which were compared using the Chi-square test or Fisher's exact test to compare two groups categorical variables, the Student 't' test, Mann-Whitney 'U' test for continuous variables, logistic regression analysis to determine binary dependent variables [21]–[24].

According to previous related work, it can be concluded that accuracy is higher in some implementation classification methods than clustering methods, but in the real-world setting, the physicians do not know the 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 presented in detail 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 subforms. Subsequent digits permit more specific diagnosis for some headache types.

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.

TABLE 1. The International Classification of Headache Disorders—the Third Edition [2].

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

The study [25] analyzes different studies that are all based on IHS recommendations. These studies deal with different approaches to attribute selection based on automatic methods, expert systems, knowledge-based 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 [26]. The study [25] shows that the most important features are: (4), (5), (6), (7), (8), (10), (12), (13) and (15) that are signed with **black bold** colour. These 9 features are used in the rest of the research.

4 Modelling the fuzzy clustering approach

In general, clustering algorithms can be grouped on given data set into clusters in 2 main different approaches:

- *Hard clustering*: each object belongs to a 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 and *k*-means++. On the other hand, the representative *fuzzy partitioning methods* are: *fuzzy c-means clustering* method, *fuzzy Gustafson–Kessel clustering* method and *fuzzy Gath–Geva clustering* method (FGGC). In this research, *fuzzy maximum likelihood estimation (FMLE) with a direct distance norm* based on the FGGC is used. *FMLE with a direct distance norm* belongs to fuzzy partitioning methods [27].

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 [28]. When *N* the total number of

TABLE 2. Comparison of a selection attribute for primary headache based on IHS diagnostic criteria: **1.** consistency measure filter, **2.** relief greedy, **3.** relief 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 the 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 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? | + | + | + | + | | | + | + |
| 16. 12. Do you have any body numbness or weakness? | | | | | | | TTH | |
| 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. | | | | | | | | |

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 variations) is defined as:

$$B(c) = \sum_t^N n_t (\mu - \mu_i)^T (\mu - \mu_j) \quad (1)$$

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:

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

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

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

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 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 [28].

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 that 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 7 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* [29].

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 $x_k = [x_{k1}, x_{k2}, \dots, x_{kn}]^T$, $x_k \in \mathbb{R}^n$. A set of N observations is denoted by $X = \{x_k \mid k = 1, 2, \dots, N\}$ and is represented as an $N \times n$ 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 $[0, 1]$. An $N \times c$ matrix $U = [\mu_{ik}]$ represents the fuzzy partitions, its conditions are given by:

$$\mu_{ij} \in [0, 1], 1 \leq i \leq N, 1 \leq k \leq c \quad (4)$$

$$\sum_{k=1}^c \mu_{ik} = 1, 1 \leq i \leq N \quad (5)$$

$$0 < \sum_{i=1}^N \mu_{ik} < N, 1 \leq k \leq c \quad (6)$$

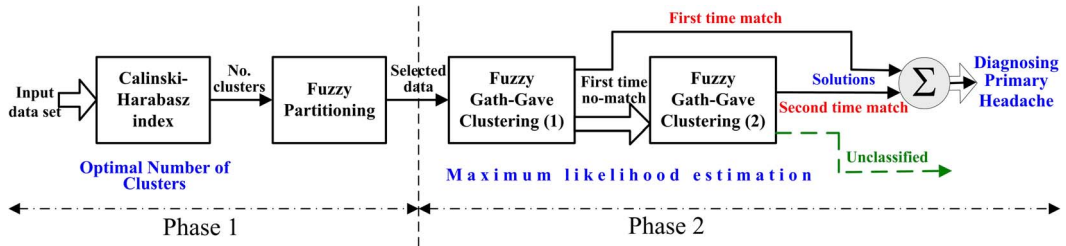


FIGURE 1. A hybrid model for *diagnosing primary headache disorder*.

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

$$M_{fc} = \left\{ U \in \mathbb{R}^{N \times c} \mid \mu_{ik} \in [0, 1], \forall i, k; \sum_{k=1}^c \mu_{ik} = 1, \forall i; 0 < \sum_{k=1}^c \mu_{ik} < N, \forall k \right\}. \quad (7)$$

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 FMLE clustering algorithm

The basic steps of the proposed hybrid algorithm for the FMLE clustering algorithm, which employs a distance norm based on FMLE proposed in [30], 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.

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 Figure 1. It has 2 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 into 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 FMLE fits much better when there are only 2 clusters to distinguish. Therefore, it could be considered that *selected data* has only 2 clusters.

Selected data are given according to the appropriate value in questions 6 and 8, which are marked with **black bold**, from Table 2. The second phase is realized in 2 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 with **red bold** on Figure 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

Algorithm 1: The algorithm—fuzzy Gath–Geva clustering method—for **Fuzzy maximum likelihood estimation**

Begin

Step 1: — **Initialization.**

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

Step 2: — **Calculate the cluster centres.**

Repeat for $l = 1, 2, \dots$

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

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

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

— **The distance function is chosen as**

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

— **with the a priori probability**

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

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

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

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

End.

number of these patients is much smaller than the number of patients in the first phase, and they are called *second time match patients* marked with **red bold**. The rest of patient are ‘unclassified’ and they are marked with **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 FMLE clustering algorithm was further on, in our research, tested on 2 data sets for diagnosing headache disorder collected from Clinical Centre of Vojvodina in Serbia. *Headache data set 1* is a part of large study [31] encompassing adult working population. *Headache data set 2* presents a part of the study encompassing student population [32].

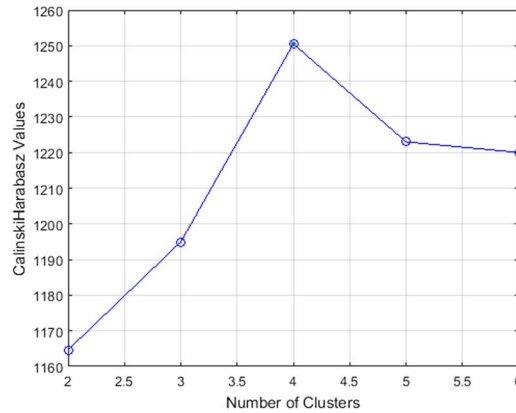


FIGURE 2. Calinski–Harabasz index for headache data set 1 to estimate the optimal number of clusters.

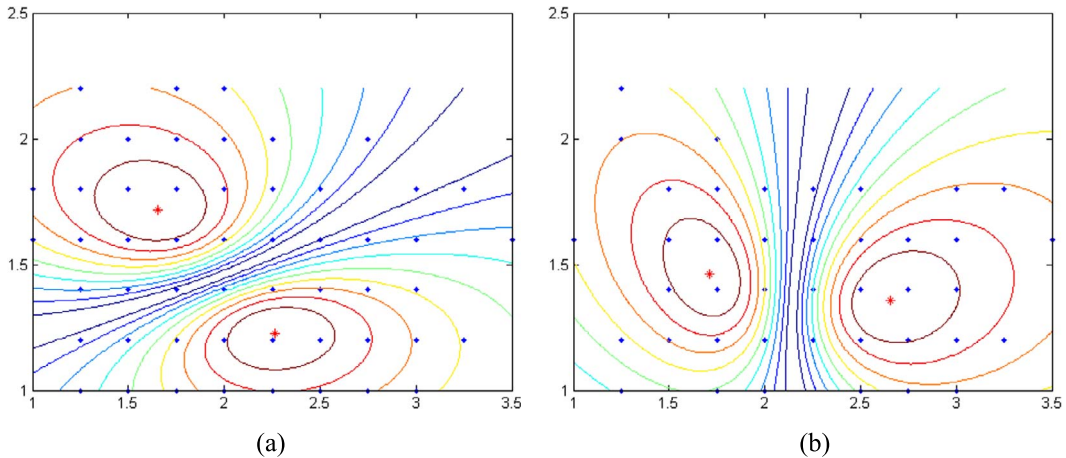


FIGURE 3. Hybrid FMLE clustering presented with two clusters: (a) after first step of fuzzy Gath–Geva clustering, (b) after second step of fuzzy Gath–Geva clustering.

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 9 features—attributes; 4 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. Calculated maximum Calinski–Harabasz value is 1250 and optimal number of clusters is 4 (Figure 2). After fuzzy partition process, there are 289 selected data. Pairwise comparison classes (MWOA—other) for headache data set 1 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 headaches, and the remaining

TABLE 3. Pairwise comparison for headache data set 1.

| Com. | MW _o A | MWA | TTH | Other | Unclassified |
|-------------------------|-------------------------------------|-----------------------------------|---------------------------------------|---------------------------------------|-------------------------------------|
| 1 | (78 + 11) = 89/103 83.39% | | | (127 + 25) = 152/186 81.72% | (14 + 34) = 48/289 16.61% |
| 2 | (91 + 1) = 92/103 89.32% | | (151 + 24) = 175/224 78.13% | | (11 + 49) = 60/327 18.35% |
| 3 | | (56 + 0) = 56/66 84.85% | | (130 + 56) = 147/186 79.03% | (10 + 39) = 49/252 19.44% |
| 4 | | (54 + 3) = 57/66 86.36% | (153 + 35) = 188/224 83.94% | | (9 + 36) = 45/290 15.51% |
| Total | 103 | 66 | 224 | 186 | |
| Average Accuracy | | | | 82.91% | 17.09% |

TABLE 4. Pairwise comparison for headache data set 2.

| Com. | MW _o A | MWA | TTH | Other | Unclassified |
|-------------------------|----------------------------|--------------------------|----------------------------|----------------------------|----------------------------|
| 1 | (33 + 0) = 33/46 71.74% | | | (41 + 4) = 45/51 88.24% | (13 + 6) = 19/97 19.58% |
| 2 | (32 + 1) = 33/46 71.74% | | (24 + 0) = 24/28 85.71% | | (13 + 4) = 17/74 22.97% |
| 3 | | (7 + 0) = 7/7 100.00% | | (38 + 6) = 42/51 82.35% | (0 + 9) = 9/58 15.51% |
| 4 | | (7 + 0) = 7/7 100.00% | (18 + 6) = 24/28 85.71% | | (0 + 4) = 4/35 11.42% |
| Total | 46 | 7 | 28 | 51 | |
| Average accuracy | | | 85.68% | | 14.32% |

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 FGGC algorithm with a distance norm based on FMLE.

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 *average accuracy* of 83% and the second has the *average 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] K. Hagen, J.-A. Zwart, L. Vatten, L. J. Stovner and G. Bovin. Prevalence of migraine and non-migrainous headache—head-HUNT, a large population-based study. *Cephalalgia*, **20**, 900–906, 2000. doi:10.1046/j.1468-2982.2000.00145.x.
- [2] Headache Classification Committee of the International Headache Society Headache Classification Committee of the International Headache Society The International Classification of Headache Disorders, 3rd edition. *Cephalalgia*, **38**, 1–211, 2018. doi:10.1177/0333102417738202.
- [3] S. Simić, Z. Banković, D. Simić and S. D. Simić. A Hybrid Clustering Approach for Diagnosing Medical Diseases. Vol. **10870** of *Lecture Notes in Computer Science*, pp. 741–752, Springer, Cham, 2018. doi:10.1007/978-3-319-92639-1_62.
- [4] B. Krawczyk, D. Simić, S. Simić and M. Woźniak. Automatic diagnosis of primary headaches by machine learning methods. *Open Medicine*, **8**, 157–165, 2013. doi:10.2478/s11536-012-0098-5.
- [5] S. Simić, D. Simić, P. Slankamenac and M. Simić-Ivkov. Computer-Assisted Diagnosis of Primary Headaches. Vol. **5271** of *Lecture Notes in Computer Science*, pp. 314–321, Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-87656-4_39.
- [6] S. Simić, D. Simić, P. Slankamenac and M. Simić-Ivkov. Rule-Based Fuzzy Logic System for Diagnosing Migraine. Vol. **5138** of *Lecture Notes in Computer Science*, pp. 383–388, Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-87881-0_37.
- [7] D. Simić, V. Ilin, I. Tanackov, V. Svirčević and S. Simić. A Hybrid Analytic Hierarchy Process for Clustering and Ranking Best Location for Logistics Distribution Center. Vol. **9121** of *Hybrid Artificial Intelligent Systems, Lecture Notes in Computer Science*, pp. 477–488, Springer, Cham, 2015. doi:10.1007/978-3-319-19644-2_40.
- [8] D. Simić, V. Ilin, V. Svirčević and S. Simić. A hybrid clustering and ranking method for best positioned logistics distribution centre in Balkan Peninsula. *Logic Journal of the IGPL*, **25**, 991–1005, 2017. doi: 10.1093/jigpal/jzx047.
- [9] M. R. Anderberg. *Cluster Analysis for Applications*. Academic Press, 1973.
- [10] M. Oded and R. Lior. Clustering methods. In *Data Mining and Knowledge Discovery Handbook*, M. Oded and R. Lior, eds, pp. 321–352. Springer, Boston, MA, 2005. doi:10.1007/0-387-25465-X_15.
- [11] A. Jain and R. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, NJ, 1988.
- [12] J. Han and M. Kamber. *Data Mining*. Morgan Kaufmann Publishers, 2001.

- [30] J. C. Bezdek and J. C. Dunn. Optimal fuzzy partitions: a heuristic for estimating the parameters in a mixture of normal distributions. *IEEE Transactions on Computers*, **C-24**, 835–838, 1975. doi: [10.1109/T-C.1975.224317](https://doi.org/10.1109/T-C.1975.224317).
- [31] S Simić. 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). PhD Thesis, University of Novi Sad, Faculty of Medicine, 2009.
- [32] R. Cakić. Investigating the Presence of Neuropathic Pain Component in Migraine and Tension-Type Headache in Students (*‘Ispitivanje Prisustva Neuropatske Komponente Bola Kod Studenata Obolelog od Migrene i Glavobolje Tenzionog Tipa’*) (Serbian). Master Thesis, University of Novi Sad, Faculty of Medicine, 2013.

Received 1 March 2019