ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2022, 23(1), pp. 37-47, DOI: 10.18038/estubtda.877029

RESEARCH ARTICLE

SEGMENTATION of COVID-19 POSITIVE PATIENTS REGARDING SYMPTOMS AND COMPLAINTS

Gökhan SİLAHTAROĞLU [1, *], Kevser ŞAHİNBAŞ [2]

[1] Managament Information System, Medipol Business School, Istanbul Medipol University, Istanbul, Turkey
[2] Managament Information System, Medipol Business School, Istanbul Medipol University, Istanbul, Turkey

ABSTRACT

The COVID-19 has spread rapidly among people living in all around the world and become a global threat. COVID-19 is approaching approximately 46 million cases worldwide according to the World Health Organization (WHO). There are limited number of COVID-19 test kits because of the rapid increasing cases daily. The fatality rate of ill patients with COVID-19 is very high in all around the world. Therefore, it is critical to cluster COVID-19 cases by applying clustering methods and provide the features of each. In this paper, we present symptom statistics of COVID-19 diagnosed patients to be used to foresee whether a patient will suffer through the illness severely or not. A clustering model by applying Fuzzy C-Means and PCA data reduction and visualization of data in a scatter diagram is also presented in the study. Clustering results shows patients may be segmented as risky or not in terms of the symptoms observed. We used the complaints and symptoms of 1.313 PCR-confirmed COVID-19 positive patients admitted to a university hospital in Istanbul. The findings from clustering method suggest that weakness, cough and sore throat were the most common COVID-19 symptoms and all of symptoms are separated into 3 clusters. Herein we report which symptoms are serious that may lead patients to critical situation.

Keywords: COVID-19, Coronavirus, Fuzzy C-Means, Chi-square, Segmentation

1. INTRODUCTION

The World Health Organization (WHO) reported cases of pneumonia of unknown etiology in Wuhan, China's Hubei province, and declared it a pandemic on 31 December 2019 [1]. As of November 3, 2020, there are 46.463.562 laboratory-confirmed cases of the virus and 1.198.569 deaths reported globally [2]. In terms of coronavirus disease (COVID-19) outbreak situation, so far, more than 43 million people have been infected, and more than one million people have died in more than two hundred countries, areas or territories [1]. Thus, detection of symptoms and clustering the patients in a pandemic have critical role to reach an accurate predictive model. Common symptoms of COVID-19 infection are fever, respiratory symptoms, dry cough and dyspnea, sore throat, fatigue, shortness of breath, myalgia, headache, abdominal pain and diarrhea [3].

In this paper, we have presented a descriptive statistics study and patients' segmentation with general complaints and symptoms observed. The study has been conducted for inpatients and outpatients. Firstly, we compared difference between the mean of age of inpatients and outpatients. Then, we compared the symptom proportions observed for both inpatient and outpatient groups. We applied Fuzzy C-Means algorithm to determine the best number of clusters for all patients. Lastly, we ensured that clusters were observed with the scatter diagram with PCA and this also supported the results of Fuzzy C-Means algorithm.

We provide some of the surveyed researches in the literature in relation to the symptom statistics of COVID-19 diagnosed patients. Yifan at al. [4] present a study enrolled a total of 140 ICU nurses work

*Corresponding Author: ksahinbas@medipol.edu.tr
Received: 08.02.2021 Published: 30.03.2022

in Wuhan. The findings from study showed that there were five main symptoms: dizziness (17.9%), headache (19.3%), nausea (21.4%), dyspnea (30.7%) and chest discomfort and palpitation (31.4%). Besides, the symptoms were categorized into three clusters: Cluster A of breathing and sleep disturbances such as dizziness, sleepiness, and dyspnea, Cluster B of gastrointestinal complaints and pain such as nausea and headache, and Cluster C of general symptoms such as xerostomia, fatigue, as well as chest discomfort and palpitation. Geva et al. [5] present an unsupervised clustering on data from 1,526 patients <21 years old hospitalized with COVID-19-related illness. They defined features that clustered patients in a group most likely to have multisystem inflammatory syndrome in children by consensus (MIS-C) using a data-driven, unsupervised approach. Afzal et al. [6] propose a model of clustering of COVID-19 data that applied c-Means and Fuzy c-Means algorithms. They clustered the data from January 2020 in terms of longitude and latitude of the globe, deaths, recoveries, and the confirmed daily cases. Their study identified five clusters as the major centroid.

The organization of the paper is as follows: Clustering methods is expressed in detail in Section 2. Fuzzy C-Means is introduced in Section 2.1. In Section 3, dataset, proposed model and findings from of the study are presented in detail. Finally, in Section 4, we conclude the paper.

2. CLUSTERING METHOD

Clustering has a crucial task in data mining [7]. Clustering methods are divided according to how the data is assigned to clusters, in other words, what type of partitions they create. In classical clustering methods, each unit must be assigned to a cluster. These methods produce splits that decompose the data set into non-null and binary discrete subgroups. Assigning data to the cluster in such rigid fashion may be insufficient in the presence of data points (units) equidistant from two or more clusters. This rigid partitioning forces the units to belong to only one cluster, while equally members of two or more clusters at the same time. The methods developed to improve the significance of clusters, in the case of such overlapping clusters, allow this “sharp” clustering method to assign units to both clusters simultaneously. Nevertheless, classical clustering methods do not show how precisely or uncertainly the unit is assigned to different clusters [8].

The "membership degree" explained by the fuzzy set theory developed by Zadeh (1965) gave an idea about the uncertainty of this state of belonging. This use of fuzzy sets provides information about indefinite set memberships. The applications of fuzzy set theory in cluster analysis were first suggested in the studies by Bellman, Kalaba, Zadeh and Ruspini. These studies have opened the door to fuzzy cluster research [9].

Diagnostic systems, image processing, classification, missing value management and attribution, bioinformatics, machine learning can be given as examples of applications performed with clustering [10].

2.1. Fuzzy C-Means Algorithm (FCM)

One of the most frequently used algorithms among clustering methods based on objective function is FCM algorithm. The first form of the fuzzy-c algorithm was presented by Dunn in 1973 and later completed by Bezdek in 1981 [11]. Most of fuzzy clustering algorithms are built based on this minimizing function of the fuzzy c means function formulated by Dunn and Bezdek.

The idea behind FCM clustering algorithm is to divide n samples $x_i (i = 1, 2, 3, \dots, n)$ into c fuzzy clusters and minimize the objective function to find the clustering center of each cluster. The formula of the objective function of the FCM algorithm is as follows [12]:

$$J_m = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m (d_{ik})^2 \tag{1}$$

where, J_m is the objective function, $m > 1$ is a constant of controlling the degree of ambiguity of the clustering result, $d_{ik}{}^2 = ||x_k - v_i||2$ is the distance from the sample x_k to the $i - th$ cluster center v_i , μ_{ik} is the membership function of k sample for $i - th$ category, which requires that the sum of membership degree of a sample for each cluster is 1, as shown in Eq. (2).

$$\sum_{i=1}^{c} u_{ik} = 1, \forall k = 1, 2, \dots, n \tag{2}$$

The membership degree μ_{ik} and cluster center v_i are updated in terms of Eqs. (3) and (4) to obtain the minimum J_m,. When updating μ_{ik}, there are two cases

Whether d_{ik} has a value of 0. Define l_k and $\bar{l}_k$ as $l_k = \{i | 1 \leq i \leq c, d_{ik} = 0\} \bar{l}_k = \{1, 2, \dots, c\} - l_k$, use b to represent the number of iterations, so that μ_{ik}^bvalue is

$$\begin{cases} \mu_{ik}^b = \dfrac{1}{\sum_{j=1}^{c} \left(\dfrac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}}, if\ l_k = \varphi \\ \mu_{ik}^b = 0, \forall\ i \in\ \bar{l}_k, and \displaystyle\sum_{i \in l_k} \mu_{ik}^b = 1, if\ l_k \neq \varphi \end{cases} \tag{3}$$

$$v_i^b = \frac{\sum_{k=1}^{n} (\mu_{ik}^b)^2 x_k}{\sum_{k=1}^{n} (\mu_{ik}^b)^m} \tag{4}$$

FCM algorithm decides which unit will be assigned to which cluster by using membership degrees. Each unit is assigned to which cluster its membership is the largest. However, each unit can also be a member of other clusters with certain membership degrees [13].

3. MATERIALS AND METHODS

3.1. Dataset

In this paper, we have used general complaints of 1.292 COVID-19 patients' data that have been obtained from Istanbul Medipol Mega University Hospital. The study started on 11th May 2020 and data collection ended on 7th August 2020. Covid-19 symptoms were also benched-marked with the ones from studies in the literature.

According to the literature, data were obtained from respiratory symptoms, fever, dry cough and dyspnea, shortness of breath, fatigue, myalgia, headache, sore throat, abdominal pain and diarrhea. Age and gender were also added to variables for clustering. It is labelled as 1 if the patient has the relevant symptom and 0 if there is no.

3.2. System Overview

The main aim of this paper is to present symptom statistics of patients diagnosed with COVID-19 to foresee whether a patient will have a severe illness by gathering a hospital's corresponding patient records. It is illustrated in Figure 1 that the patients records were collected from hospital. The Medical Ethics Committee of the Istanbul Medipol University reviewed and approved the study protocol. Authors also received a letter of approval from the Ministry of Health, Turkey to use Covid-19 patients' data for scientific studies. The first stage starts with data-preprocessing. Then, fuzzy cluster centers are initialized randomly. In the next step, the new centers are re-calculated by FCM. This iteration continues until a pre-defined convergence is met. The memberships, distances and mean parameters are updated, clusters are reassigned in terms of new parameters. At the end of the stage, segmentation results are illustrated by scatter diagram.

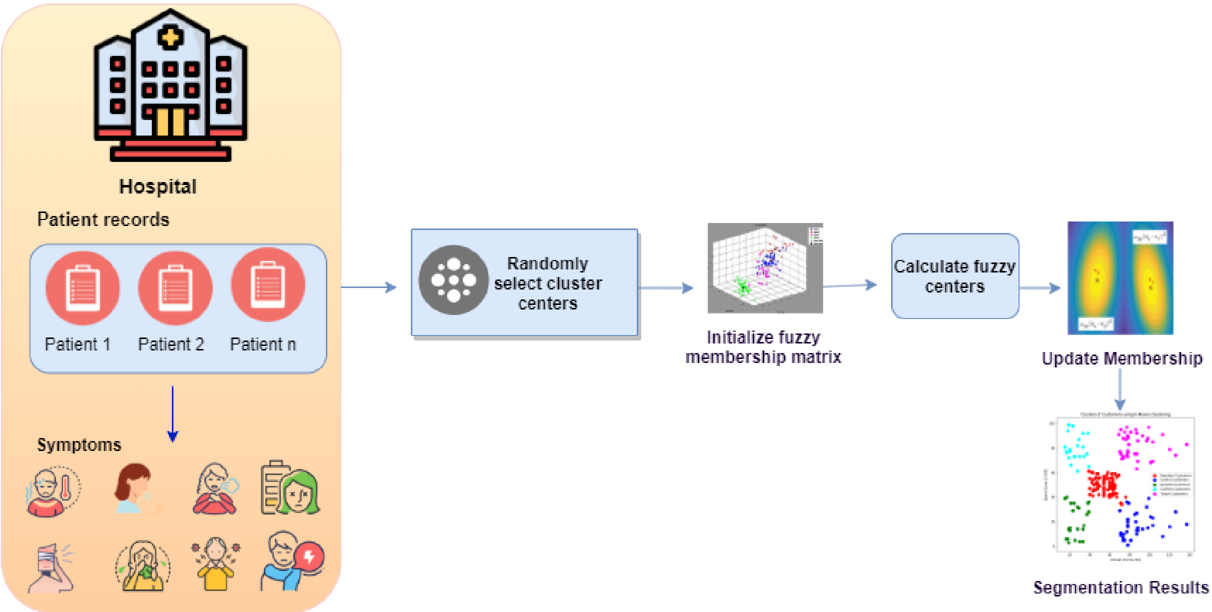

Figure 1. System Overview

3.3. Analysis Results

For the analysis 1.292 patients' data have been used. 57 of them are inpatients and 1.242 of them are outpatients. 8 patients passed away in the hospital during Covid-19 treatment.

3.3.1. Statistical analysis

Table 1. Proportion of inpatients and outpatients.

Symptom Observed	Proportion	Proportion	Proportion	Proportion Test Value
smell and taste loss	0.043	0.118	0.115	0.082
chest pain	0.043	0.034	0.034	0.093
Chills	0.065	0.063	0.063	0.162
back pain	0.022	0.219	0.211	0.177
shortness of breath	0.217	0.024	0.032	0.581
abdominal pain	0.022	0.004	0.005	0.782
covid-contact in the family	0.022	0.063	0.061	0.793
burning sensation in the body	0.000	0.179	0.172	0.823
Diarrhea	0.022	0.014	0.014	0.877
respiratory distress	0.109	0.004	0.009	1.139
Anorexia	0.022	0.008	0.009	1.390
weakness	0.152	0.531	0.516	1.438
joint pain	0.022	0.064	0.062	1.720
nasal congestion	0.000	0.005	0.005	2.155
Sore throat	0.087	0.297	0.289	2.664
Sputum	0.022	0.043	0.042	2.748
Fatigue	0.022	0.185	0.179	2.791
Cough	0.109	0.407	0.395	3.190
runny nose	0.000	0.014	0.014	3.561
headache	0.109	0.199	0.195	3.913
Fever	0.174	0.271	0.267	4.974
muscle / bone pain	0.022	0.078	0.076	5.501
asthenia	0.065	0.210	0.204	6.935
nausea / vomiting	0.087	0.031	0.033	7.377
numbness	0.000	0.004	0.003	11.388
sweating	0.000	0.014	0.014	12.318
Gender-Female: 0, Male: 1	0.587	0.561	0.562	344.951

H_0: There is no difference between inpatients and outpatients in terms of proportions of symptoms observed in the groups.

H_1: There is a significant proportion difference in terms of symptoms observed between inpatients and outpatients.

With 95% confidence, t statistics is taken as 1.96.
The test results in Table suggests that we reject H_0 for those symptoms: smell and taste loss, chest pain, chills, back pain, shortness of breath, abdominal pain, Covid-contact in the family, burning sensation in the body, diarrhea, respiratory distress, anorexia, weakness and joint pain.

We fail to reject H_0 for all other symptoms on behalf of H_1. This means that there are some symptoms which may tell us about the severity of the disease for some patients. Bearing in mind the fact that patients are admitted to healthcare centers with one or more symptoms, this information may be used to predict the likelihood of patients to be admitted as 'inpatient' and will be a workload or burden for the healthcare personnel, if they are not already so.

For example, statistics suggests that if a patient has sore throat, s/he is more likely to be an inpatient than those who has not sore throat. Furthermore, if the patient has sore throat and fatigue, probability of suffering through the illness will increase. More examples may be generated like those mentioned.

Another finding of the study is that most frequently observed symptoms are weakness, cough and sore throat whereas numbness, abdominal pain and nasal congestion are the least frequently observed symptoms. Surprisingly enough, all these symptoms are observed with inpatients than with outpatients.

Table 2. Observation proportion all patients.

Symptom Observed	**Observation proportion all**
Weakness	0.516
Cough	0.395
sore throat	0.289
Fever	0.267
back pain	0.211
Asthenia	0.204
Headache	0.195
Fatigue	0.179
burning sensation in the body	0.172
smell and taste loss	0.115
muscle / bone pain	0.076
Chills	0.063
joint pain	0.062
covid-contact in the family	0.061
Sputum	0.042
chest pain	0.034
nausea / vomiting	0.033
shortness of breath	0.032
Diarrhea	0.014
runny nose	0.014
Sweating	0.014
respiratory distress	0.009
Anorexia	0.009
abdominal pain	0.005
nasal congestion	0.005
Numbness	0.003

To sum up, with 95% confidence, t statistics is taken as 1.96, thus proportion test suggests that symptoms nasal congestion, sore throat, sputum, fatigue, cough, runny nose, headache, fever, muscle / bone pain, asthenia, nausea / vomiting, numbness, sweating are observed the same in two patient groups. It is also clear that gender distribution is also the same in groups. In terms of symptoms, two groups differ from each other by the following: smell and taste loss, chest pain, chills, back pain, shortness of breath, abdominal pain, Covid-contact in the family, burning sensation in the body, diarrhea, respiratory distress, anorexia, weakness and joint pain. Here we report that patients who show these symptoms are more likely to suffer from the disease more severely than the others may.

3.3.2. Implementation of clustering

The objective of clustering is to see weather inpatients and outpatients reside in different clusters or not. The reason to choose fuzzy C means rather than other frequently used clustering algorithms such as K-Means, DBSCAN or hierarchical algorithms is that patients are highly intermixed in terms of symptoms.

That means it is possible to encounter all possible symptom combinations in any cases. Fuzzy C-means takes all possibilities when performing clustering and perform the segmentation regarding all possible options. When Fuzzy- C means algorithm has been used for clustering, fuzziness has been taken as 2 delta 0.2 and lambda as 0.1. Delta is the fixed distance from every datapoint to the noise cluster. Delta is updated in each iteration, based on the average interpoint distances. However, a lambda parameter has to be set, according to the shape of the clusters. In and between cluster variations, Partition Coefficient, Partition Entropy and Xie Beni Index have been used to determine the number of clusters.

Table 3. Parameters in Cluster.

	3 Clusters	**4 Clusters**	**5 Clusters**
Between Cluster Variation	3.010	77.22	91.64
Partition Coefficient	0.585	0.574	0.463
Partition Entropy	-0.764	-0.826	-0.890
Xie Beni Index	1.614	1.582	1.586
Noise %	0%	25%	34%

After trying with various (3 to 5) cluster numbers we found that 3 clusters to segment the data is the best regarding the following statistics generated.

Table 1. Within Cluster Variations.

Cluster	**Variations**
Cluster 1	3.010
Cluster 2	3.380
Cluster 3	3.438

Table 5. Within Cluster Variations.

Cluster	**Values**
Between Cluster Variation	3.010
Partition Coefficient	0.585
Partition Entropy	-0.764
Xie Beni Index	1.614

Table 6. Fuzzy Hypervolumes.

Cluster	**Variations**
Cluster 1	1.327 E-64
Cluster 2	4.988 E-64
Cluster 3	2.633 E-64

We also applied chi-square test to see how three clusters are separated from each other. Firstly, we calculated the proportion of each symptom in the clusters and then by applying chi square test we discovered how each cluster differ from each other.

Chi square test suggests that some symptoms separate the clusters with p values which are less than 0.05. Table 7 shows that cluster 1 is distinguished from other two clusters in terms of some symptoms such as cough, fever, weakness, asthenia, fatigue, sore throat, numbness, loss of smell and taste, sputum, nasal congestion, back pain, and muscle pain / bone. Not all but most of these symptoms are the ones which also separates inpatients from outpatients. Nevertheless, Cluster 1 has more members than other two. This is the confusing part to distinguish patients.

Table 7. Chi Square Statistics

	Cluster 01 689 patients.	Cluster 02 121 patients	Cluster 03 544 patients	Chi Square C2 vs C3		
	Proportion	Proportion	Proportion	P	P	P
Gender-Female: 0, Male: 1	54% Male	60% Male	52% Male	0.18	**0.00**	**0.00**
Covid-contact in the family	9%	5%	3%	**0.03**	**0.00**	**0.00**
cough	38%	42%	38%	0.36	**0.00**	**0.00**
fever	28%	24%	30%	0.49	**0.00**	**0.00**
weakness	48%	58%	45%	**0.01**	**0.00**	**0.00**
asthenia	20%	24%	14%	0.36	**0.00**	**0.00**
fatigue	16%	20%	18%	0.22	**0.00**	**0.00**
sore throat	29%	33%	20%	0.56	**0.00**	**0.00**
headache	17%	25%	14%	**0.01**	**0.00**	**0.00**
numbness	1%	0%	0%	0.57	0.13	0.55
Loss of smell and taste	13%	13%	6%	0.98	**0.00**	**0.00**
sputum	4%	4%	5%	0.99	**0.00**	**0.00**
nasal congestion	1%	0%	0%	0.67	0.06	0.30
chills / chills	5%	7%	8%	0.44	**0.00**	**0.00**
chest pain	3%	3%	6%	0.97	**0.00**	**0.00**
back pain	20%	22%	22%	0.65	**0.00**	**0.00**
muscle pain / bone	7%	7%	9%	1.00	**0.00**	**0.00**
joint pain	4%	8%	7%	0.12	**0.00**	**0.00**
burning sensation in the body	16%	19%	17%	0.40	**0.00**	**0.00**
nausea / vomiting	2%	4%	4%	0.12	**0.02**	**0.00**
runny nose	1%	2%	1%	0.77	0.11	**0.03**
diarrhea	1%	1%	3%	0.92	0.07	0.11
abdominal pain	1%	0%	0%	0.87	0.22	0.44
anorexia	0%	1%	2%	0.76	0.28	0.19
shortness of breath	2%	4%	4%	**0.04**	**0.03**	**0.00**
sweating	1%	2%	1%	0.63	0.09	**0.01**
respiratory distress	0%	3%	0%	4.79	0.11	
Hospitalization day	1%	114%	16%	265.20	**0.00**	
Discharge Status Discharge: 0 Death: 1	0%	3%	0%	5.67	**0.06**	

In order to provide the reader with a visual chart and show the distribution of patients in cluster, we applied Principal Component Analysis (PCA) technique to reduce dimension to 2. Doing so, we had a chance to draw a scatter diagram of patients so that we could see the clusters if possible. PCA dimension 1 and dimension 2 represent 99% of all variables. Indeed, 3 clusters and the place of inpatients in all data could be identified in the scatter diagram. In this way it is possible to label inpatients, outpatients and those who lost their lives. It can be seen from the Figures 2-4.

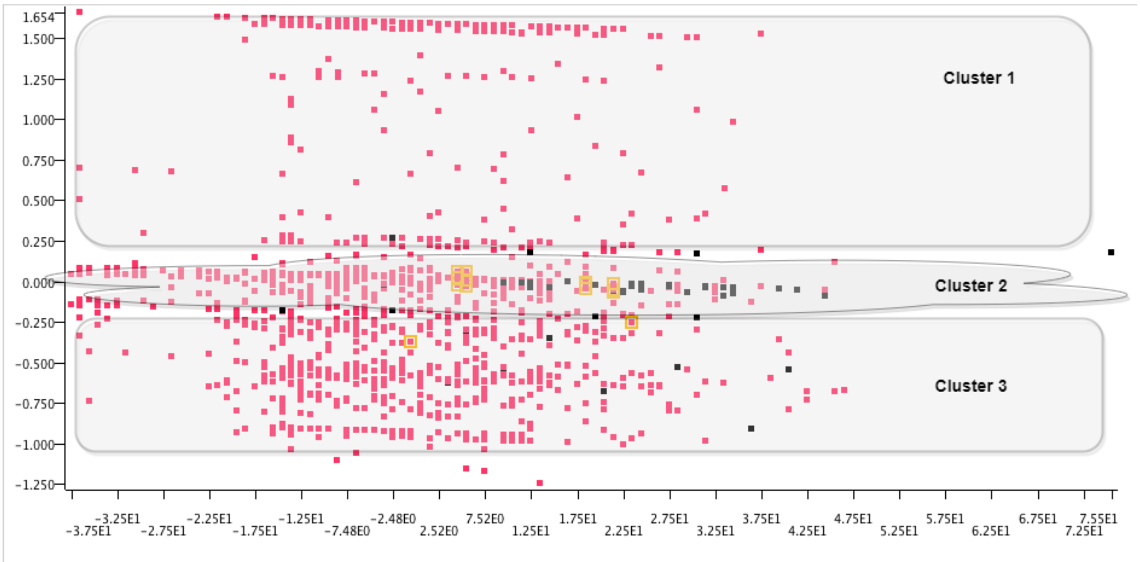

Figure 2. Cluster of all patients, blue dots represent the patients who have died.

As it is depicted in Figure 2 one cluster has fewer patients than the other two. This is probably the one which is labeled Cluster 2 by Fuzzy – C means. One cluster has more inpatients (represented with black dots) than others. This is the cluster that chi- square separated from others. That means clustering patients via their symptoms with PCA /Visual Graphics and Fuzz C- Means have yielded very similar results. From this finding on, one can say that early symptoms of COVID-19 patients may reveal whether they will be admitted as inpatients sooner or later, and also if they will need more special care than any others.

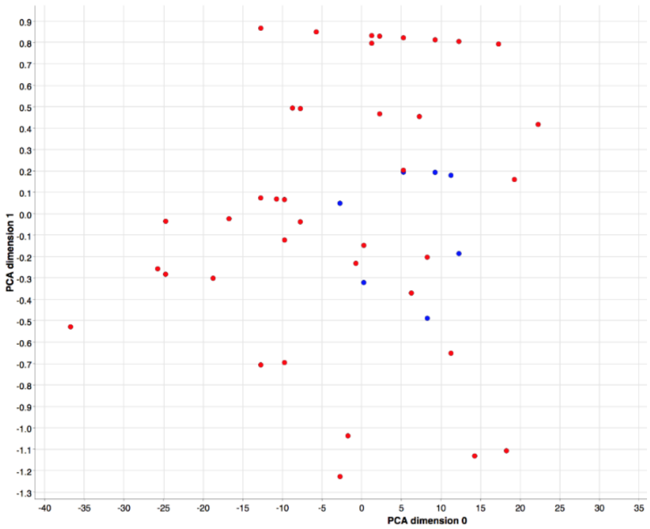

Figure 3. Cluster of inpatients blue dots represent the patients who have died.

Figure **3** and
Figure **4** show a detailed view of Cluster 2 which has inpatients and those who passed away. In the figures blue dots represent the patients who died of Covid-19. Figures depict that, all patients and in

patients scatter diagrams are very similar and outpatients' data are embedded in all patients scatter diagram. This confirms that PCA data reduction and visualization worked well.

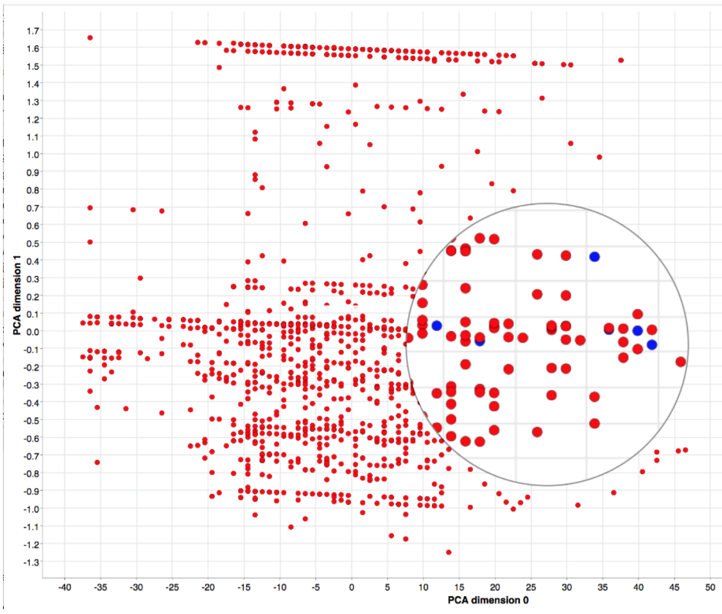

Figure 4. All patients and inpatients marked, *blue dots represent the patients who have died.*

4. CONCLUSION

Patients arrive at healthcare centers with some symptoms sever or not. However, they may show no symptoms, albeit they may only have COVID-19 contact which is enough to request a PCR test. It is reported that only few of them stay at hospital right after learning the positive test result [14]. More often, their situations get worse in time and they go back to the hospital to stay in services or intensive care units. This study has aimed to discover if symptoms and their combinations are any helpful to determine the severeness of the illness and foresee if a patient will come back to the hospital in a worse situation.

Body weakness, cough, sore throat and fever are the most commonly observed symptoms of COVID-19 patients. In the study, patients are segmented into three clusters by Fuzzy C-means algorithm. One segment has not got any inpatients and other two include both inpatients and outpatients. Only one of them include 8 patients who died. This cluster may be considered as the most serious cluster. The symptoms which separate this cluster from other two are cough, weakness, fatigue, asthenia, headache, sore throat, and burning sensation.

In one cluster, cough, weakness, fatigue, asthenia, headache, sore throat, burning sensation in the body observations are higher than that of other clusters but *fever* is slightly less observed. Most of them are male patients. Patients who show if not all but most of these symptoms may be defined critical patients. The least serious group come with Covid-19 contact in the family, headache and weakness complaints. The cluster of patients in between the serious one and the other differs from the serious group by fever, sweating and asthenia symptoms and differs from the least serious group by showing fewer headache and weakness symptoms.

Physicians should ensure that patients with these symptoms seek care early to reduce the risk of being patient associated with COVID-19. Numbness, abdominal pain and nasal congestion are predictive symptoms of poor outcomes. To sum up, we report that the patients having most of the following symptoms at the same time should be taken as serious cases:

Cough, weakness, fatigue, asthenia, headache, sore throat, burning sensation in the body. The risk increases if the patient is male.

CONFLICT OF INTEREST

The authors stated that there are no conflicts of interest regarding the publication of this article.

REFERENCES

[1] World Health Organization (WHO) Homepage, https://www.who.int/emergencies/diseases/novel-coronavirus-2019, last accesses 2020/11/03.

[2] Huang C, Wang Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 2020; 395 (10223), 497–506.

[3] Singhal T. A review of coronavirus disease-2019 (COVID-19), Indian J. Pediatr 2020; 87, 281–286.

[4] Yifan T, Ying L, Chunhong G et all. Symptom Cluster of ICU Nurses Treating COVID-19 Pneumonia Patients in Wuhan, China, Journal of Pain Pain and Symptom Management, 2020; 60.

[5] Geva A, Patel MM, Newhams MM, et all. Data-driven clustering identifies features distinguishing multisystem inflammatory syndrome from acute COVID-19 in children and adolescents, EclinicalMedicine, 2021; 40.

[6] Afzal A, Ansari Z, Alshanrani S, et all. Clustering of COVID-19 data for knowledge discovery using c-means and fuzzy c-means, Results in Physics, 2021; 29.

[7] Nejatian S, Parvin H, Faraji E. Using sub-sampling and ensemble clustering techniques to improve performance of imbalanced classification, Neurocomputing, 2018; 276, pp.55–66.

[8] Rashidi F, Nejatian S, Parvin H, Rezaie V. Diversity based cluster weighting in cluster ensemble: an information theory approach, Artif Intell, 2019; 52(2), 1341–68.

[9] Niu H, Khozouie N, Parvin H, Alinejad-Rokny H, Beheshti A, Mahmoudi MR. An ensemble of locally reliable cluster solutions, Appl Sci, 2020; 10(5), 1891.

[10] Bagherinia A, Minaei-Bidgoli B, Hossinzadeh M, Parvin H. Reliability-based fuzzy clustering ensemble, Fuzzy Sets Syst, 2020.

[11] Mojarad M, Nejatian S, Parvin H, Mohammadpoor M. A fuzzy clustering ensemble based on cluster clustering and iterative fusion of base clusters, Appl Intell, 2019; 7, 2567–81.

[12] Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, J Cybern, 1973; 3(3), 32-57.

[13] Silahtaroğlu G. Veri Madenciliği Kavram ve Algoritmaları (4. Basım). Papatya Bilim Yayıncılık. 2019.

[14] Lam C, Calvert J, Siefkas A, et all. Personalized stratification of hospitalization risk amidst COVID-19: A machine learning approach, Health Policy and Technology, 2021; 10.