Journal of New Theory
ISSN: 2149-1402

39 (2022) 54-83
Journal of New Theory
https://dergipark.org.tr/en/pub/jnt
Open Access

Diagnosing COVID-19, Prioritizing Treatment, and Planning Vaccination Priority via Fuzzy Parameterized Fuzzy Soft Matrices

Zeynep Parla Parmaksız[1], Burak Arslan[2], Samet Memiş[3], Serdar Enginoğlu[4]

Article Info
Received: 09 Jun 2022
Accepted: 30 Jun 2022
Published: 30 Jun 2022
doi:10.53570/jnt.1128289
Research Article

Abstract – In the fight against the COVID-19 pandemic, it is vital to rapidly diagnose possible contagions, treat patients, plan follow-up procedures with correct and effective use of resources and ensure the formation of herd immunity. The use of machine learning and statistical methods provides great convenience in dealing with too many data produced during research. Since access to the PCR test used for the diagnosis of COVID-19 may be limited, the test is relatively too slow to yield results, the cost is high, and its reliability is controversial; thus, making a symptomatic classification before the PCR is timesaving and far less costly. In this study, by modifying a state-of-the-art classification method, namely Comparison Matrix-Based Fuzzy Parameterized Fuzzy Soft Classifier (FPFS-CMC), an effective method is developed for a rapid diagnosis of COVID-19. The paper then presents the accuracy, sensitivity, specificity, and F1-score values that represent the diagnostic performances of the modified method. The results show that the modified method can be adopted as a competent and accurate diagnosis procedure. Afterwards, a tirage study is performed by calculating the patients' risk scores to manage inpatient overcrowding in healthcare institutions. In the subsequent section, a vaccine priority algorithm is proposed to be used in the case of a possible crisis until the supply shortage of a newly developed vaccine is over if a possible variant of COVID-19 that is highly contagious is insensitive to the vaccine. The accuracy of the algorithm is tested with real-life data. Finally, the need for further research is discussed.

Keywords – *Medical diagnosing, prioritizing treatment, planning vaccination priority, fpfs-matrices, soft decision-making*

Mathematics Subject Classification (2020) – 03E72, 68Q32

1. Introduction

1.1. Diagnosis of COVID-19

The severe acute respiratory syndrome coronavirus (SARS-CoV-2) has affected our lives for the past two years. Rapid diagnosis of possible contagions, planning of follow-up treatment, and effective use of resources have vital importance in the fight against the COVID-19 pandemic. The use of machine learning and statistical methods provides great convenience to deal with these difficulties. In the literature, there are several common

[1]zeynepparlaparmaksiz@gmail.com (Corresponding Author); [2]tburakarslan@gmail.com; [3]samettmemis@gmail.com; [4]serdarenginoglu@gmail.com
[1]Beşiktaş Arts and Sciences Centre, İstanbul, Turkiye
[2,4]Department of Mathematics, Faculty of Arts and Sciences, Çanakkale Onsekiz Mart University, Çanakkale, Turkiye
[3]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, İstanbul Rumeli University, İstanbul, Turkiye

classifiers, such as Support Vector Machine (SVM) [1], Fuzzy k-Nearest Neighbour (Fuzzy kNN) [2], AdaBoost [3], Decision Tree (DT) [4], Fuzzy Soft Set Classifier (FSSC) [5], Fuzzy Soft Set Classifier Using Distance-Based Similarity Measure (FussCyier) [6], and Hamming Distance-Based Fuzzy Soft Set Classifier (HDFSSC) [7]. Recently, a novel classifier, i.e., Compare-Matrix Based Fuzzy Parameterized Fuzzy Soft Classifier (FPFS-CMC) [8,9] that produces high scores in "Breast Cancer", "Parkinsons[sic]", and "Parkinson's Diseases" datasets provided in UCI Machine Learning Repository [10], has been prominent among the aforesaid classifiers in medical diagnosis. However, it has not been applied to COVID-19 yet. Therefore, it is worth studying to diagnose COVID-19 via the classifier. This study, firstly, detects whether the individual is COVID-19 positive by utilizing a state-of-the-art classification method FPFS-CMC.

COVID-19: SARS-CoV-2 was first reported in Wuhan, China, in December 2019 and spread rapidly worldwide. COVID-19 formed a clade within the subgenus sarbecovirus, Orthocoronavirinae subfamily. This virus is a droplet infection [11]. The available research on COVID-19 has been increasing [12-18]. Moreover, several datasets related to COVID-19 have been shared in data repositories, such as UCI and Kaggle. This study uses the datasets titled "Symptoms and COVID Presence (May 2020 data)", "Covid-19 Symptoms", and "Brazilian Covid Symptomatic Patients Data" [19-21], provided in Kaggle Data Repository to diagnose COVID-19 by using a classification method (classifier).

Classifiers: Supervised learning is a sub-field of machine learning which is commonly used in various fields, particularly defense industry, meteorology, psychology, finance, medicine, astronomy, and space sciences. Classification is a supervised learning technique that learns a predictive model from the training data to make an accurate prediction of a datum's label [22]. Classifiers utilize the information of the training set, whose labels are known, and predict the class label of a sample with an unknown label. So far, many classifiers have been produced, such as SVM, Fuzzy kNN, AdaBoost, DT, FSSC, FussCyier, and HDFSSC.

Lately, a state-of-the-art classifier FPFS-CMC, which employs the modeling capability of fuzzy parameterized fuzzy soft matrices (*fpfs*-matrices) [23] in real-world problems containing uncertainties, has been proposed. FPFS-CMC produces high scores than the aforesaid classifiers in "Breast Cancer", "Parkinsons[sic]", and "Parkinson's Disease" datasets provided in UCI Machine Learning Repository [10]. Therefore, this study utilizes FPFS-CMC to diagnose COVID-19.

1.2. Follow-Up Treatment Priority in COVID-19 Patients

Designing an algorithm to calculate each patient's risk score is crucial for hospitals to provide better follow-up methods and treatment services. These unique risk scores for patients who have tested positive for COVID-19 have significance to compare the severity levels of the disease in patients. This study secondly determines how severely the patient will recover from the virus by comparing risk scores.

Comorbidities: Comorbidities can negatively affect patients' conditions during COVID-19 [13,14,17,18]. This study focuses on the common comorbidities during COVID-19 – namely, hypertension, cardiovascular diseases, cancer, chronic kidney failure, and diabetes.

1.3. Vaccination Priority Planning

Furthermore, the COVID-19 vaccination priority planning is essential to overcoming a possible crisis until the supply shortage of a newly developed vaccine is over in the case a possible highly contagious variant of COVID-19 is insensitive to the vaccine. Besides, planning for booster doses is another issue that needs to be considered when an inadequate number of vaccines are available. This study thirdly produces a ranking order among individuals who are willing to get vaccinated based on their vaccination priority scores. Finally, it discusses the need for further research.

Current Vaccination Applications: Currently, several criteria are being utilized in the vaccination process. While planning for vaccination, the Turkish Ministry of Health [16] focused on systemic diseases, age, and occupation of an individual who will be vaccinated. This study expands these criteria to make a more specific analysis of the vaccination process. It considers seven criteria, i.e., systemic disease, age, presence of risk group individuals in the immediate vicinity, presence of COVID-19 history, province-district, transportation preference, and occupation, to make a better priority planning. The framework in this study is shown in Fig. 1.

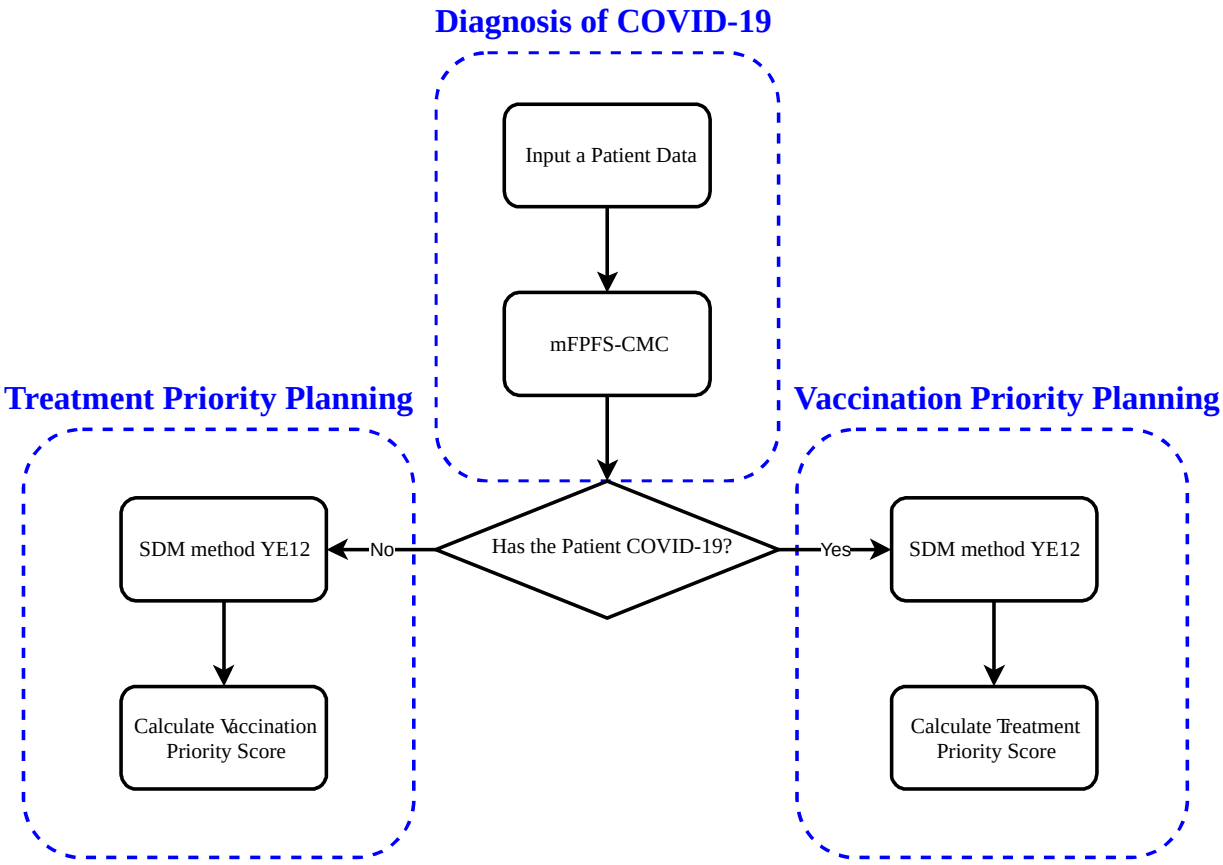

Fig. 1. Flowchart of the proposed work

2. Hypotheses:

This study considers the following hypotheses:

i. If FPFS-CMC is used in medical diagnosis, then whether the individual has COVID-19 can be determined.

ii. If the risk score of the patient can be calculated by looking at the patient's age and systemic diseases, then the follow-up treatment priorities of patients can be compared.

iii. If such parameters as systemic disease history, age, presence of risk-group individuals in the immediate vicinity, presence of COVID-19 history, province-district, transportation preference, and occupation can be obtained, then individuals' vaccination priority scores can be calculated.

3. Preliminaries

This study presents some of the basic definitions required in the next sections. Throughout this study, let E be a parameter set, U be a universal set, $F(E)$ be the set of all the fuzzy sets over E, and $\mu \in F(E)$. Here, $\mu \coloneqq \left\{{}^{\mu(x)}x : x \in E\right\}$.

Definition 1. [24] Let U be a universal set, $\mu \in F(E)$, and α be a function from μ to $F(U)$. Then, the set $\left\{\left({}^{\mu(x)}x, \alpha\left({}^{\mu(x)}x\right)\right) \mid x \in E\right\}$, being the graphic of α, is called a fuzzy parameterized fuzzy soft set (*fpfs*-set) parameterized via E over U (or briefly over U).

Moreover, the set of all the *fpfs*-sets parameterized via E over U is denoted by $FPFS_E(U)$.

Definition 2. [23] Let $\alpha \in FPFS_E(U)$. Then, $[a_{ij}]$ is called *fpfs*-matrix of α and is defined by

$$[a_{ij}] = \begin{bmatrix} a_{01} & a_{02} & a_{03} & \dots & a_{0n} & \dots \\ a_{11} & a_{12} & a_{13} & \dots & a_{1n} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

such that for $i \in \{0,1,2,\cdots\}$ and $j \in \{1,2,\cdots\}$,

$$a_{ij} \coloneqq \begin{cases} \mu(x_j), & i = 0 \\ \alpha\left({}^{\mu(x_j)}x_j\right)(u_i), & i \neq 0 \end{cases}$$

Here, if $|U| = m - 1$ and $|E| = n$, then $[a_{ij}]$ has order $m \times n$. Moreover, the set of all the *fpfs*-matrices parameterized via E over U is denoted by $FPFS_E[U]$.

Definition 3. Let $u, v \in \mathbb{R}^n$. Then, the function $P: \mathbb{R}^n \times \mathbb{R}^n \to [-1,1]$ defined by

$$P(u,v) \coloneqq \frac{n\sum_{i=1}^n u_i v_i - (\sum_{i=1}^n u_i)(\sum_{i=1}^n v_i)}{\sqrt{\left[n\sum_{i=1}^n u_i^2 - \left(\sum_{i=1}^n u_i\right)^2\right]\left[n\sum_{i=1}^n v_i^2 - \left(\sum_{i=1}^n v_i\right)^2\right]}}$$

Definition 4. [25] Let D_{train} with $m_1 \times n$ and $C_{m_1 \times 1}$ be a training matrix and the class column vector of D_{train}. Then, fw is called the feature weight vector based on the Pearson correlation coefficient of D_{train} and is denoted by

$$fw_{1j} \coloneqq \left|P\left(D_{train-j}, C\right)\right|, \quad j \in I_n \coloneqq \{1,2,3,\cdots,n\}$$

Definition 5. Let $u \in \mathbb{R}^n$. Then, the vector $\hat{u} \in \mathbb{R}^n$ defined by

$$\hat{u}_i \coloneqq \begin{cases} \dfrac{u_i - \min\limits_{k \in I_n}\{u_k\}}{\max\limits_{k \in I_n}\{u_k\} - \min\limits_{k \in I_n}\{u_k\}}, & \max\limits_{k \in I_n}\{u_k\} \neq \min\limits_{k \in I_n}\{u_k\} \\ 1, & \max\limits_{k \in I_n}\{u_k\} = \min\limits_{k \in I_n}\{u_k\} \end{cases}, \qquad i \in I_n$$

is called normalizing vector of u.

Definition 6. Let $u \in \mathbb{R}^n$. Then, the standard deviation of u is defined by

$$\mathrm{std}(u) \coloneqq \sqrt{\frac{\sum_{i=1}^{n}\left(u_i - \frac{1}{n}\sum_{i=1}^{n} u_i\right)}{n-1}}$$

Definition 7. [25] Let $D = \left[d_{ij}\right]_{m\times(n+1)}$ be a data matrix, $i \in I_m$, and $j \in I_n$. Then, the matrix $\widetilde{D} = \left[\tilde{d}_{ij}\right]_{m\times n}$ defined by

$$\tilde{d}_{ij} \coloneqq \begin{cases} \dfrac{d_{ij} - \min\limits_{k\in I_m}\{d_{kj}\}}{\max\limits_{k\in I_m}\{d_{kj}\} - \min\limits_{k\in I_m}\{d_{kj}\}}, & \max\limits_{k\in I_m}\{d_{kj}\} \neq \min\limits_{k\in I_m}\{d_{kj}\} \\ 1, & \max\limits_{k\in I_m}\{d_{kj}\} = \min\limits_{k\in I_m}\{d_{kj}\} \end{cases}$$

is called column normalized matrix (feature-fuzzification matrix) of D.

Definition 8. [25] Let $(D_{train})_{m_1\times n}$ be a training matrix obtained from $D = \left[d_{ij}\right]_{m\times(n+1)}$. Then, the matrix $\widetilde{D}_{train} = \left[\tilde{d}_{ij-train}\right]_{m_1\times n}$ defined by

$$\tilde{d}_{ij-train} \coloneqq \begin{cases} \dfrac{d_{ij-train} - \min\limits_{k\in I_m}\{d_{kj}\}}{\max\limits_{k\in I_m}\{d_{kj}\} - \min\limits_{k\in I_m}\{d_{kj}\}}, & \max\limits_{k\in I_m}\{d_{kj}\} \neq \min\limits_{k\in I_m}\{d_{kj}\} \\ 1, & \max\limits_{k\in I_m}\{d_{kj}\} = \min\limits_{k\in I_m}\{d_{kj}\} \end{cases}, \qquad i \in I_{m_1} \text{ and } j \in I_n$$

is called column normalized matrix (feature-fuzzification matrix) of D_{train}.

Definition 9. [25] Let $(D_{test})_{m_2\times n}$ be a training matrix obtained from $D \coloneqq \left[d_{ij}\right]_{m\times(n+1)}$. Then, the matrix $\widetilde{D}_{test} = \left[\tilde{d}_{ij-test}\right]_{m_2\times n}$ defined by

$$\tilde{d}_{ij-test} \coloneqq \begin{cases} \dfrac{d_{ij-test} - \min\limits_{k\in I_m}\{d_{kj}\}}{\max\limits_{k\in I_m}\{d_{kj}\} - \min\limits_{k\in I_m}\{d_{kj}\}}, & \max\limits_{k\in I_m}\{d_{kj}\} \neq \min\limits_{k\in I_m}\{d_{kj}\} \\ 1, & \max\limits_{k\in I_m}\{d_{kj}\} = \min\limits_{k\in I_m}\{d_{kj}\} \end{cases}, \qquad i \in I_{m_2} \text{ and } j \in I_n$$

is called column normalized matrix (feature-fuzzification matrix) of D_{test}.

Definition 10. [25] Let $\left[a_{ij}\right]_{m\times n}, \left[b_{ij}\right]_{m\times n} \in FPFS_E[U]$ and $p \in \mathbb{Z}^+$. Then, the mapping $s_M^p : FPFS_E[U] \times FPFS_E[U] \to \mathbb{R}$ defined by

$$s_M^p\left(\left[a_{ij}\right], \left[b_{ij}\right]\right) = 1 - \frac{1}{\sqrt[p]{(m-1)n}}\left(\sum_{i=1}^{m-1}\sum_{j=1}^{n}\left|a_{0j}a_{ij} - b_{0j}b_{ij}\right|^p\right)^{\frac{1}{p}}$$

is a pseudo-similarity over $FPFS_E[U]$ and is called Minkowski pseudo-similarity. Here, s_M^1 is referred to as Hamming pseudo-similarity and is denoted by s_H. Moreover, s_M^2 is referred to as Euclidean pseudo-similarity and is denoted by s_E.

Definition 11. [25] Let $[a_{ij}]_{m\times n}, [b_{ij}]_{m\times n} \in FPFS_E[U]$ and $p \in \mathbb{Z}^+$. Then, the mapping $s_{Hs}^p: FPFS_E[U] \times FPFS_E[U] \to \mathbb{R}$ defined by

$$s_{Hs}^p([a_{ij}],[b_{ij}]) = 1 - \frac{1}{\sqrt[p]{m-1}}\left(\sum_{i=1}^{m-1} \max_{j\in I_n}\left\{\left|a_{0j}a_{ij} - b_{0j}b_{ij}\right|^p\right\}\right)^{\frac{1}{p}}$$

is a pseudo-similarity over $FPFS_E[U]$ and is called p-Hausdorff pseudo-similarity. Here, s_{Hs}^1 is referred to as Hausdorff pseudo-similarity and is denoted by s_{Hs}.

Definition 12. [25] Let $[a_{ij}]_{m\times n}, [b_{ij}]_{m\times n} \in FPFS_E[U]$. Then, the mapping $s_C: FPFS_E[U] \times FPFS_E[U] \to \mathbb{R}$ defined by

$$s_C([a_{ij}],[b_{ij}]) = 1 - \max_{i\in I_{m-1}}\left\{\max_{j\in I_n}\left\{\left|a_{0j}a_{ij} - b_{0j}b_{ij}\right|\right\}\right\}$$

is a pseudo-similarity over $FPFS_E[U]$ and is called Chebyshev pseudo-similarity.

4. Method

This study

- diagnoses COVID-19 by employing the classification method mFPFS-CMC,
- calculates the follow-up treatment priority of individuals with COVID-19 by risk scores,
- computes COVID-19 vaccination priority scores.

4.1. Safety

This study does not include vertebrate animals, potentially hazardous biological agents (microorganisms, rDNA, and tissues, including blood and blood products), and hazardous substances and devices.

The survey studies provided herein do not require any personal data while gathering information about individuals' views about the vaccination process, and their criteria points. By not recording any name, gender, or e-mail address, the respective patient is anonymized. Thus, it stores the received data anonymously so that they cannot be associated with real people.

All the participants have explicit consent to the following to be used for academic reasons:

- participants' views on vaccination priority planning
- their information about their systemic disease, age, and occupation
- whether they live in the immediate vicinity
- presence of their COVID-19 history
- the population of their current province and district
- their transportation preferences

Participants are aware of the potential risks of the study:

- an outside source may be tampered when using the internet for collecting information.
- there is always a possibility of hacking or other security breaches that could threaten the confidentiality of their responses while the confidentiality of participants' responses will be protected once the data are downloaded from the internet.

The surveys declare that the participants are free not to answer any questions. All the responses are deleted from the online survey. No personal or electronic identifier is kept. The data file is stored on a password-protected computer.

4.2. Experimentation

4.2.1. Diagnosis of COVID-19

This subsection presents a classification algorithm to diagnose COVID-19.

FPFS-CMC: FPFS-CMC firstly utilizes the Pearson correlation coefficient between each column, corresponding to parameters, and the last column, manifesting the class labels, in the considered dataset to calculate feature weights based on the impact of parameters on classification. Then, using feature fuzzification of the training and testing samples and feature weights, it creates two *fpfs*-matrices: a training *fpfs*-matrix and a testing *fpfs*-matrix. It then creates a comparison matrix based on the pseudo-similarities between the training and testing *fpfs*-matrices. After that, it calculates the standard deviation of each column of the comparison matrix to produce the parameter weights and then merges the parameter weights and the matrix to generate the comparison *fpfs*-matrix. The ideal training sample is obtained by applying the soft decision-making (SDM) method sMBR01 on the comparison *fpfs*-matrix. Finally, the testing sample is given the class label of the optimum training sample. The same procedures are applied for all the test samples. However, FPFS-CMC has a disadvantage in terms of running time compared to the aforesaid classifiers herein. To overcome this drawback, this study modifies FPFS-CMC by employing the SDM method EMK19 [26,27] instead of sMBR01 [28]. Thus, the modified FPFS-CMC (mFPFS-CMC) produces a running time advantage of up to 70% over FPFS-CMC. The pseudocode of mFPFS-CMC is as follows:

Algorithm 1. Pseudocode of mFPFS-CMC

Input: $(D_{train})_{m_1\times n}$, $C_{m_1\times 1}$, and $(D_{test})_{m_2\times n}$
Output: $T'_{m_2\times 1}$

procedure mFPFS-CMC(D_{train}, C, D_{test})
  Compute fw using D_{train} and C
  Compute feature fuzzification of D_{train} and D_{test}, namely $\widetilde{D}_{train}$ and $\widetilde{D}_{test}$
  for i from 1 to m_2 **do**
    Compute the training *fpfs*-matrix $[a_{ij}]$ using fw and $\widetilde{D}_{j-train}$
    for j from 1 to m_1 **do**
      Compute the testing *fpfs*-matrix $[b_{ij}]$ using fw and $\widetilde{D}_{i-test}$
      $F_{j1} \leftarrow s_H([a_{ij}],[b_{ij}])$
      $F_{j2} \leftarrow s_C([a_{ij}],[b_{ij}])$
      $F_{j3} \leftarrow s_E([a_{ij}],[b_{ij}])$
      $F_{j4} \leftarrow s_{Hs}([a_{ij}],[b_{ij}])$
      $F_{j5} \leftarrow s_M^3([a_{ij}],[b_{ij}])$
    end for
    for j from 1 to 5 **do**
      $sd_j \leftarrow \text{std}(F_{.j})$
    end for
    $pw \leftarrow 1 - \frac{\widehat{sd}}{}$
    Compute comparison *fpfs*-matrix $[g_{ij}]$ using pw and F
    $[[s_{k1}],[dm_{k1}],[op_{k1}]] \leftarrow \text{EMK19}([g_{ij}])$
    $t'_{i1} \leftarrow C(op_{11}, 1)$
  end for
return $T'_{m_2\times 1}$
end procedure

SDM Methods: The related literature offers many SDM methods operating *fpfs*-matrices [23,29-34]. The SDM methods employ single, double, or multiple *fpfs*-matrix/matrices. In FPFS-CMC, sMBR01 working with a single *fpfs*-matrix is used. Another SDM method EMK19, employing a single *fpfs*-matrix, provides the best advantage in running time of FPFS-CMC over the others. For this reason, this study has chosen EMK19 to modify FPFS-CMC.

Real-Life Interpretation: This study firstly applies FPFS-CMC to the datasets Symptoms and COVID Presence (May 2020 data) [19], Covid-19 Symptoms [20], and Brazilian Covid Symptomatic Patients Data [21] provided in Kaggle Data Repository utilizing MATLAB R2021b software and a laptop with I(R) Core(TM) i5-10210U CPU @ 1.60GHz 2.11 GHz and 8.00 GB. Moreover, it compares them with kNN, Fuzzy kNN, and SVM. In the simulation process, to split the datasets as training and testing, 5-fold cross-validation is used (for more details about k-fold cross-validation, see [35-37]). The simulation results in Table 1 show that FPFS-CMC can be successfully applied to diagnose COVID-19. However, the results also manifest that the classifier has a running time disadvantage. To overcome this difficulty, this study employs the SDM method EMK19 instead of sMBR01 used in a step of the classifier FPFS-CMC. Here, FPFS-CMC with EMK19 is denoted by mFPFS-CMC.

Table 1. Simulation results of the classifiers for the considered dataset

Datasets' References	Classifiers	Acc∓SD	Sen∓SD	Spe∓SD	F1∓SD	RT ∓SD
[19]	kNN	96.8273∓0.0044	83.7574∓0.0220	99.9566∓0.0010	91.0564∓0.0134	1.7714∓0.6284
	Fuzzy kNN	96.8237∓0.0042	83.6718∓0.0211	99.9772∓0.0005	91.0489∓0.0129	0.5669∓0.2851
	SVM	96.7667∓0.0046	85.8218∓0.0388	99.3908∓0.0081	91.1028∓0.0138	0.8290∓0.5427
	FPFS-CMC	96.8182∓0.0041	83.6434∓0.0207	99.9772∓0.0005	91.0326∓0.0127	964.4258∓122.9805
	mFPFS-CMC	96.8182∓0.0041	83.6434∓0.0207	99.9772∓0.0005	91.0326∓0.0127	286.2255∓40.8747
[20]	kNN	84.8850∓0.0351	96.9331∓0.0312	30.0278∓0.1534	91.3102∓0.0203	0.0460∓0.0060
	Fuzzy kNN	84.7990∓0.0296	95.5320∓0.0368	35.8611∓0.1578	91.1349∓0.0181	0.0009∓0.0004
	SVM	86.8589∓0.0366	95.9630∓0.0318	45.3333∓0.1828	92.2945∓0.0212	0.0214∓0.0045
	FPFS-CMC	87.3923∓0.0314	94.9915∓0.0370	52.6944∓0.1460	92.4903∓0.0195	0.2757∓0.0159
	mFPFS-CMC	87.0406∓0.0318	94.9915∓0.0370	50.7222∓0.1589	92.3014∓0.0195	0.2902∓0.0223
[21]	kNN	96.1174∓0.0081	92.1834∓0.0215	98.0516∓0.0064	93.9858∓0.0129	0.5669∓0.0101
	Fuzzy kNN	96.5564∓0.0070	94.1915∓0.0187	97.7187∓0.0065	94.7402∓0.0110	0.1055∓0.0018
	SVM	86.5061∓0.0116	71.7255∓0.0247	93.7736∓0.0103	77.7876∓0.0200	0.2278∓0.0184
	FPFS-CMC	96.4052∓0.0071	94.4757∓0.0189	97.3536∓0.0081	94.5396∓0.0110	130.4978∓0.1860
	mFPFS-CMC	96.4052∓0.0071	94.4757∓0.0189	97.3536∓0.0081	94.5396∓0.0110	50.4621∓0.2595

Acc, Sen, Spe, and F1 and their standard deviations (SD) are presented in percentage. Running time and its SD are presented in seconds.

In medical diagnosis, accuracy, sensitivity, specificity, and F1-score are vital and expected to occur close to 100%. According to the results in Table 1, although mFPFS-CMC's result of sensitivity for the dataset in [19] and the results of accuracy and specificity for the dataset in [20] are less than 90%, its other results are

above 90%. As seen in Table 2, since the considered datasets are imbalanced, some results are lower than 90%. Moreover, $O((m-1)^2 n)$ and $O\big((m-1)nt\big)$ represent the computational complexities of sMBR01 and EMK19, respectively, such that $m-1$, n, and t denote the number of samples, the number of attributes, and the number of matrices, respectively. Here, since mFPFS-CMC utilizes one matrix, $t=1$. Because the computational complexity of sMBR01 is higher than the computational complexity of EMK19, mFPFS-CMC has a running time advantage of up to 70% over FPFS-CMC. To this end, improving mFPFS-CMC is worth studying. Consequently, mFPFS-CMC is reliable and practical in medical diagnosis.

Table 2. Details of the considered datasets (# represents "the number of")

No.	Reference	Sample #	Attribute #	Class #	Class Labels	Samples' Distribution	Balanced/Imbalanced
1.	[19]	5434	20	2	No and Yes	1051 (No) 4383 (Yes)	Imbalanced
2.	[20]	227	31	2	0 and 1	214 (0) 13 (1)	Imbalanced
3.	[21]	2779	10	2	0 and 1	916 (0) 1863 (1)	Imbalanced

Here, the mathematical notations of the performance metrics, namely accuracy (acc), sensitivity (sen), specificity (spe), and F1-score (F1) [38,39], are as follows: Let $D_{test}=\{x_1,x_2,\cdots,x_n\}$, $T=\{T_1,T_2,\cdots,T_n\}$, $T'=\{T_1',T_2',\cdots,T_n'\}$, and $I_n:=\{1,2,3,\cdots,n\}$ be the set of n samples to be classified, the set of ground truth classes of the samples, the set of prediction class of the samples, and an index set, respectively. Then,

$$\text{Accuracy}(T,T'):=\frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Sensitivity}(T,T'):=\frac{TP}{TP+FN}$$

$$\text{Specificity}(T,T'):=\frac{TN}{TN+FP}$$

$$\text{F1}-\text{Score}(T,T'):=\frac{2TP}{2TP+FP+FN}$$

where TP, TN, FP, and FN are the number of true positive, true negative, false positive, and false negative, respectively, and their mathematical notations are as follows:

$$TP:=|\{x_k:1\in T_k\wedge 1\in T_k',k\in I_n\}|$$

$$TN:=|\{x_k:0\in T_k\wedge 0\in T_k',k\in I_n\}|$$

$$FP:=|\{x_k:0\in T_k\wedge 1\in T_k',k\in I_n\}|$$

$$FN:=|\{x_k:1\in T_k\wedge 0\in T_k',k\in I_n\}|$$

Furthermore, Table 3 shows the mean values TP, FN, TN, and FP for each fold of cross-validation obtained in ten runs by classifiers for the considered datasets.

Table 3. Mean of TP, FN, TN, and FP values obtained in ten runs

		[19]				[20]				[21]			
Classifiers	k-fold	TP	FN	TN	FP	TP	FN	TN	FP	TP	FN	TN	FP
kNN	Fold 1	176.9	33.1	875.3	0.7	36.3	0.7	1.7	6.3	169.5	13.5	364.9	7.1
	Fold 2	178.7	32.3	875.8	0.2	37	1	2.3	5.7	169.7	14.3	365.4	6.6
	Fold 3	175.7	34.3	876.6	0.4	35.8	1.2	3.5	5.5	168.1	14.9	365.8	7.2
	Fold 4	173.7	36.3	876.8	0.2	35.5	1.5	2.4	5.6	167.4	15.6	364.5	8.5
	Fold 5	175.3	34.7	876.6	0.4	35.7	1.3	2.5	5.5	169.7	13.3	366.1	6.9
Fuzzy kNN	Fold 1	176.5	33.5	875.9	0.1	35.8	1.2	2.1	5.9	173.4	9.6	363.4	8.6
	Fold 2	178.5	32.5	875.8	0.2	36.7	1.3	2.7	5.3	174.4	9.6	363.1	8.9
	Fold 3	175.5	34.5	876.6	0.4	34.8	2.2	4.1	4.9	171.8	11.2	364.4	8.6
	Fold 4	173.5	36.5	876.8	0.2	35.2	1.8	3.3	4.7	170.8	12.2	364.3	8.7
	Fold 5	175.4	34.6	876.9	0.1	35.2	1.8	2.6	5.4	172.4	10.6	365.3	7.7
SVM	Fold 1	181.3	28.7	870.4	5.6	35.5	1.5	3	5	132.6	50.4	349.7	22.3
	Fold 2	183.7	27.3	870.4	5.6	36.8	1.2	4.2	3.8	131	53	348.7	23.3
	Fold 3	179.5	30.5	871.9	5.1	35.2	1.8	5.1	3.9	130.1	52.9	352.3	20.7
	Fold 4	178.5	31.5	870.7	6.3	35.6	1.4	3.2	4.8	132.3	50.7	346.9	26.1
	Fold 5	179	31	872.9	4.1	35.4	1.6	3.2	4.8	131	52	349.4	23.6
FPFS-CMC	Fold 1	176.5	33.5	875.9	0.1	35.5	1.5	3.7	4.3	173.9	9.1	361.2	10.8
	Fold 2	178.2	32.8	875.8	0.2	36.7	1.3	4.5	3.5	174.3	9.7	362.4	9.6
	Fold 3	175.5	34.5	876.6	0.4	34.5	2.5	5.6	3.4	172.5	10.5	363.1	9.9
	Fold 4	173.5	36.5	876.8	0.2	35.1	1.9	4.1	3.9	171.4	11.6	364	9
	Fold 5	175.4	34.6	876.9	0.1	34.9	2.1	3.8	4.2	173.3	9.7	363	10
mFPFS-CMC	Fold 1	176.5	33.5	875.9	0.1	35.5	1.5	3.4	4.6	173.9	9.1	361.2	10.8
	Fold 2	178.2	32.8	875.8	0.2	36.7	1.3	4.2	3.8	174.3	9.7	362.4	9.6
	Fold 3	175.5	34.5	876.6	0.4	34.5	2.5	5.5	3.5	172.5	10.5	363.1	9.9
	Fold 4	173.5	36.5	876.8	0.2	35.1	1.9	3.9	4.1	171.4	11.6	364	9
	Fold 5	175.4	34.6	876.9	0.1	34.9	2.1	3.9	4.1	173.3	9.7	363	10

4.2.2. Follow-Up Treatment Priority in COVID-19 Patients

This subsection proposes a treatment priority algorithm to utilize in follow-up treatment priority in COVID-19 patients.

Risk Scores: This study calculates risk scores for age and the aforesaid comorbidities by using the following functions:

Age: The odds ratios of individuals' ages provided in [14,17] show that the age criterion affects the death rates caused by COVID-19. Therefore, this study considers the death rates of COVID-19 patients per 100,000 people in the last 7 days by age group in Table 4 provided in [15]. It then determines a priority score for individuals according to their ages. Hence, the age risk score function f_A is as follows:

$$f_A : P \to [0,1]$$

$$x \to f_A(x) = \chi(x,i)$$

such that $\chi(x,i) = the\ value\ corresponding\ to\ i^{th}\ age\ range\ that\ x\ belongs\ in.$ Here, $A = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ is a set of age ranges such that x_1 = "0-14", x_2 = "15-24", x_3 = "25-49", x_4 = "50-64", x_5 = "65-79", and x_6 = ">80" and P is a set of patents. To illustrate, if a patient x belongs in the age range 15-24, then the priority score $f_A(x) = \chi(x,2) = 0.0013$.

Table 4. Normalized death rates according to age range

Age Ranges	0-14	15-24	25-49	50-64	65-79	>80
Scores	0.0130	0.0013	0.0100	0.1037	0.4552	1

The scores are obtained by normalizing and merging the ranges < 2, 2-4, and 5-14 as 0-14.

Hypertension: In Turkiye, approximately 80% of hypertension patients have primary hypertension and the remaining 20% have secondary hypertension [40]. Therefore, this study considers the basic risk scores 0.8 and 0.2 for primary and secondary hypertensions, respectively. Moreover, it adds the effect of transmitting factors to basic risk scores, i.e., excessive alcohol intake, smoking, sedentary life, polysystem, non-steroidal anti-inflammatories, and low potassium intake. Hence, the hypertension risk score function is as follows:

$$f_H : P \to [0,1]$$

$$x \to f_H(x) = \begin{cases} 0.8 + \frac{1}{30}\sum_{i=1}^{6} \chi(i), & x\ has\ primary\ hypertension \\ 0.2 + \frac{1}{30}\sum_{i=1}^{6} \chi(i), & x\ has\ secondary\ hypertension \end{cases}$$

such that $\chi(i) = \begin{cases} 1, & x\ has\ h_i \\ 0, & x\ has\ not\ h_i \end{cases}$. Here, $H = \{h_1, h_2, h_3, h_4, h_5, h_6\}$ is a set of transmitting factors such that h_1 = "excessive alcohol intake", h_2 = "smoking", h_3 = "sedentary life", h_4 = "polysystem", h_5 = "non-steroidal anti-inflammatories", and h_6 = "low potassium intake" and P is a set of patients. To illustrate, if the patient x has primary hypertension and three transmitting factors h_1, h_2, and h_6, then the hypertension risk score of x is $f_H(x) = 0.8 + \frac{1}{30}(1+1+0+0+0+1) = 0.9$.

Cardiovascular Diseases: In Turkiye, the mortality rate because of cardiovascular disease is 42% [41]. Therefore, this study calculates the cardiovascular risk score by using the rate 42% and interactive risk score

$R(x)$, obtained by the interactive form provided in [42,43]. Hence, the cardiovascular risk score function is as follows:

$$f_{Cr} : P \to [0,1]$$
$$x \to f_{Cr}(x) = \begin{cases} 0.42 + R(x), & R(x) \leq 0.58 \; doesn't \; have \; cardiovascular \; disease \\ 1, & otherwise \end{cases}$$

Here, P is a set of patients.

Cancer**:** This study determines cancer risk scores by using Table 5 provided as cited in [44, 45]. Hence, the cancer risk score function is as follows:

$$f_C : P \to [0,1]$$
$$x \to f_C(x) = 1 - \chi(x,i,j)$$

such that $\chi(x,i,j) = the \; 5 - year \; lifetimes \; rate \; of \; i^{th} \; cancer \; type \; and \; j^{th} \; stage \; that \; x \; has.$ Here, $C = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ is a set of cancer types such that $x_1 =$ "breast cancer", $x_2 =$ "colorectal cancer", $x_3 =$ "non-Hodgkin lymphoma cancer", $x_4 =$ "lung cancer ", $x_5 =$ "testicular cancer ", $x_6 =$ "bladder cancer", and $x_7 =$ "uterine cancer ", $S = \{s_1, s_2, s_3\}$ is a set of stages such that $s_1 =$ "early", $s_2 =$ "local forward", and $s_3 =$ "metastatic", and P is a set of patients. To illustrate, if a patient x has lung cancer in the metastatic stage, then the cancer risk score $f_C(x) = 1 - \chi(x,4,3) = 1 - 0.04 = 0.96$.

Table 5. Five-year lifetime rates for different cancer types and stages [as cited in 44]

Cancer Types/Stages	Early	Local Forward	Metastatic
Breast Cancer	0.99	0.85	0.26
Colorectal Cancer	0.90	0.71	0.13
Non-Hodgkin Lymphoma Cancer	0.82	0.74	0.62
Lung Cancer	0.55	0.27	0.04
Testicular Cancer	0.99	0.96	0.74
Bladder Cancer	0.70	0.34	0.05
Uterine Cancer	0.95	0.68	0.17

Chronic Kidney Failure: This study determines chronic kidney failure risk scores by using the stage number of the disease in Table 6 provided in [46]. Hence, the chronic kidney failure risk score function is as follows:

$$f_{Ch} : P \to [0,1]$$
$$x \to f_{Ch}(x) = \frac{S(x)}{5}$$

such that $S(x)$ is the stage of the disease shown in Table 6. Here, P is a set of patients. To illustrate, if the patient is in the third stage, then his/her chronic kidney failure risk score is $\frac{3}{5} = 0.6$.

Table 6. Stages of chronic kidney failure [46]

Stage	Glomerular Filtration Rate	Description	Treatment Stage
1	90+	Normal kidney function but urine findings or structural abnormalities or genetic trait point to kidney disease	Observation, control of blood pressure
2	60-89	Mildly reduced kidney function, and other findings (as for stage 1) point to kidney disease	Observation, control of blood pressure and risk factor
3	30-59	Moderately reduced kidney function	Observation, control of blood pressure and risk factor
4	15-29	Severely reduced kidney function	Planning for end-stage renal failure
5	<15 or on dialysis	Very severe or end-stage kidney failure (sometimes call established renal failure)	Treatment choices

Diabetes: This study determines diabetes risk scores by using the group number of the disease $G(x)$ in Table 7 provided in [47]. Moreover, it considers whether the individual has a cardiovascular disease because cardiovascular diseases significantly increase the risk of diabetes [48]. Hence, the diabetes risk score function is as follows:

$$f_D : P \to [0,1]$$

$$x \to f_D(x) = \frac{9 - G(x)}{10} + 0.2\,\chi(x)$$

such that

$$\chi(x) = \begin{cases} 1, & x \text{ has a cardiovascular disease} \\ 0, & \text{otherwise} \end{cases}$$

To illustrate, if the patient is in the second group and he/she has a cardiovascular disease, then his/her diabetes risk score is $\frac{9-2}{10} + 0.2 = 0.9$.

Table 7. Types of diabetes [47]

Group	Description
1	Classic type-1 diabetes, a severe immune system disease
2	A type of diabetes caused by severe insulin deficiency
3	Severe insulin resistance
4	A type of diabetes caused by obesity
5	Moderate diabetes

COVID-19 Death Correlation Score: This study calculates the death correlation score, for the aforesaid comorbidities, using mean death rates in Table 8. Moreover, the death correlation score for age criterion is obtained to be 0.26 by the mean of values in Table 4. The death rates of the patients with comorbidities who died from COVID-19 are obtained from [11,12,18,49].

Table 8. COVID-19-induced death rates of the patients with comorbidities

Group	Chen et al., 2020	Çoktaş, 2020	Erol, 2020	Zhou et al., 2020	Mean
Hypertension	0.58	N\A	0.76	0.45	0.60
Cardiovascular diseases	0.70	N\A	0.63	N\A	0.67
Cancer	N\A	0.89	N\A	0	0.45
Chronic kidney failure	N\A	N\A	0.93	1	0.97
Diabetes	N\A	N\A	0.73	0.47	0.60

N\A: Not Available

Calculation of risk score: This study computes the total risk score of a patient with COVID-19 via age and comorbidities risk scores with the death correlation scores corresponding to each risk score. Hence, the treatment priority score function is as follows:

$$f_{TPS} : P \to [0,1]$$
$$x \to f_{TPS}(x) = \frac{1}{\sum_{i=1}^{6} r_i}\left(f_A(x)r_1 + f_H(x)r_2 + f_{Cr}(x)r_3 + f_C(x)r_4 + f_{Ch}(x)r_5 + f_D(x)r_6\right)$$

Here, $MDR = \{r_1, r_2, r_3, r_4, r_5, r_6\}$ is a set of the mean death rates in Table 7 such that $r_1 = 0.26$, $r_2 = 0.60$, $r_3 = 0.67$, $r_4 = 0.45$, $r_5 = 0.97$, and $r_6 = 0.60$ and P is a set of patients. To illustrate, if the risk scores of a patient x are $f_A(x) = 0.65$, $f_H(x) = 0.4$, $f_{Cr}(x) = 0.52$, $f_C(x) = 0.34$, $f_{Ch}(x) = 0.18$, and $f_D(x) = 0$, then the treatment priority score of x is as follows:

$$f_{TPS}(x) = \frac{0.65 \cdot 0.26 + 0.4 \cdot 0.60 + 0.52 \cdot 0.67 + 0.34 \cdot 0.45 + 0.18 \cdot 0.97 + 0 \cdot 0.60}{0.26 + 0.60 + 0.67 + 0.45 + 0.97 + 0.60} = \frac{1.085}{3.55} = 0.3056$$

A Hypothetical Scenario: This study considers scores provided in Table 9 for ten patients to illustrate the aforesaid treatment priority score function's performance and performance of the SDM method YE12 employing a single matrix [50,51]. Let $P = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$ be a set of patients and $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ be a set of parameters such that e_1 = "age", e_2 = "hypertension", e_3 = "cardiovascular diseases", e_4 = "cancer", e_5 = "chronic kidney failure", and e_6 = "diabetes". Here, the weights of parameters are $r_1 = 0.26$, $r_2 = 0.60$, $r_3 = 0.67$, $r_4 = 0.45$, $r_5 = 0.97$, and $r_6 = 0.60$, respectively. Then, the *fpfs*-matrix $[a_{ij}]_{11\times 6}$ constructed by these weights and the data in Table 9 is as follows:

$$[a_{ij}] = \begin{bmatrix} 0.26 & 0.60 & 0.67 & 0.45 & 0.97 & 0.60 \\ 0.0130 & 0.84 & 0.54 & 0.39 & 0.6 & 0.7 \\ 0.4552 & 0.26 & 0.72 & 0.73 & 0.2 & 0.3 \\ 0.0100 & 0.90 & 0.64 & 0.66 & 0.4 & 0.9 \\ 0.0130 & 0.87 & 1 & 0.26 & 0.6 & 0.4 \\ 1 & 0.93 & 0.48 & 0.87 & 0.2 & 0 \\ 0.1037 & 0.27 & 0.82 & 0.96 & 0.8 & 0.3 \\ 0.0130 & 0.35 & 0.74 & 0.38 & 0.8 & 0.6 \\ 0.4552 & 0.36 & 0.42 & 0.83 & 0.2 & 0.4 \\ 0.0013 & 0.24 & 0.60 & 0.66 & 0.4 & 0.8 \\ 0.0013 & 0.34 & 0.80 & 0.01 & 0.4 & 0.1 \end{bmatrix}$$

Then, this study applies the SDM method YE12 to $[a_{ij}]$. Thus, the decision set, score matrix, and ranking order produced by YE12, respectively, are as follows:

$$\{^{0.5765}x_1, ^{0.4111}x_2, ^{0.6187}x_3, ^{0.6012}x_4, ^{0.4859}x_5, ^{0.5990}x_6, ^{0.5679}x_7, ^{0.4009}x_8, ^{0.4821}x_9, ^{0.3360}x_{10}\}$$

$$[s_{i1}] = [0.5765 \quad 0.4111 \quad 0.6187 \quad 0.6012 \quad 0.4859 \quad 0.5990 \quad 0.5679 \quad 0.4009 \quad 0.4821 \quad 0.3360]^T$$

and

$$x_{10} \prec x_8 \prec x_2 \prec x_9 \prec x_5 \prec x_7 \prec x_1 \prec x_6 \prec x_4 \prec x_3$$

Here,

$$s_{41} = f_{TPS}(x_4) = \frac{0.0130 \cdot 0.26 + 0.87 \cdot 0.60 + 1 \cdot 0.67 + 0.26 \cdot 0.45 + 0.6 \cdot 0.97 + 0.4 \cdot 0.60}{0.26 + 0.60 + 0.67 + 0.45 + 0.97 + 0.60} = \frac{2.1344}{3.55} = 0.6012$$

Table 9. Priority scores for 10 individuals

Individuals / Criteria	Age	Hypertension	Cardiovascular Diseases	Cancer	Chronic Kidney Failure	Diabetes
x_1	0.0130	0.84	0.54	0.39	0.6	0.7
x_2	0.4552	0.26	0.72	0.73	0.2	0.3
x_3	0.0100	0.90	0.64	0.66	0.4	0.9
x_4	0.0130	0.87	1	0.26	0.6	0.4
x_5	1	0.93	0.48	0.87	0.2	0
x_6	0.1037	0.27	0.82	0.96	0.8	0.3
x_7	0.0130	0.35	0.74	0.38	0.8	0.6
x_8	0.4552	0.36	0.42	0.83	0.2	0.4
x_9	0.0013	0.24	0.60	0.66	0.4	0.8
x_{10}	0.0013	0.34	0.80	0.01	0.4	0.1

4.2.3. Vaccination Priority Planning

This subsection proposes a vaccine priority algorithm to employ in a possible vaccine crisis and for the planning of booster doses.

Survey Criteria Scores: This study applies an online form, conducted through Google Forms, to 200 people to gain insight into people's understanding of the importance of the aforesaid seven criteria in vaccination priority and planning. This survey asks participants to rank the seven criteria from least to most important and calculates survey criterion points by arithmetic average. For the survey's safety, see Section 4.1. the means of values determined by 200 participants for the aforesaid seven criteria, i.e., systemic disease, age, presence of risk group individuals in the immediate vicinity, presence of COVID-19 history, province-district, transportation preference, and occupation, provided in Table 10 are denoted by S_{Sd}, S_A, S_{IV}, S_{CH}, S_{PD}, S_{TP}, and S_O, respectively.

Table 10. Survey results

Abbreviations	Survey Criteria	Mean Values	Normalized Mean Values
Sd	Systemic disease	$S_{Sd} = 3.0825$	$NS_{Sd} = 1$
A	Age	$S_A = 2.805$	$NS_A = 0.9100$
IV	Presence of risk group individuals in the immediate vicinity	$S_{IV} = 3.0575$	$NS_{IV} = 0.9919$
CH	Presence of COVID-19 history	$S_{CH} = 2.705$	$NS_{CH} = 0.8775$
PD	Province-district	$S_{PD} = 1.765$	$NS_{PD} = 0.5726$
TP	Transportation preference	$S_{TP} = 2.27$	$NS_{TP} = 0.7364$
O	Occupation	$S_O = 2.2725$	$NS_O = 0.7372$

Individual Priority Scores: This study details the aforesaid seven criteria to make a more individual-specific vaccination planning and assigns a score to each criterion.

Systemic Disease: Scientific studies observe that cases of COVID-19 with systemic diseases have a higher death rate than cases without systemic diseases. Therefore, this study considers mean percentage values of mortality rates in Table 11 provided in [11,12,18,49]. Moreover, it determines a priority score for individuals with systemic disease. Hence, the systemic disease priority score function f_{Sd} is as follows:

$$f_{Sd} : I \to [0,1]$$
$$x \to f_{Sd}(x) = \begin{cases} \chi(x,i), & i^{th}\ systemic\ disease\ that\ x\ has \\ 0, & x\ has\ no\ systemic\ disease \end{cases}$$

such that $\chi(x,i) = the\ value\ corresponding\ to\ i^{th}\ systemic\ disease\ that\ x\ has.$ Here, $Sd = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ is a set of systemic diseases such that $x_1 =$ "kidney failure (dialysis)", $x_2 =$ "chronic lung disease", $x_3 =$ "cardiovascular disease", $x_4 =$ "diabetes", $x_5 =$ "cancer", and $x_6 =$ "no" and I is a set of individuals. To illustrate, if an individual x has diabetes, then the systemic disease priority score $f_{Sd}(x) = \chi(x,5) = 0.60$.

Table 11. Systemic diseases and the respective vaccination scores

Systemic Diseases	[44]	[12]	[11]	[18]	Mean
Chronic lung disease	N\A	N\A	0.80	0.67	0.74
Cardiovascular diseases	0.70	N\A	0.63	N\A	0.67
Cancer	N\A	0.89	N\A	0	0.45
Chronic kidney failure	N\A	N\A	0.93	1	0.97
Diabetes	N\A	N\A	0.73	0.47	0.60

Age: This study considers the number of new COVID-19 patients per 100,000 people in the last 7 days by age group in Table 12 provided in [15]. It then determines a priority score for individuals according to their ages. Hence, the age priority score function f_A is as follows:

$$f_A : I \to [0,1]$$
$$x \to f_A(x) = \chi(x,i)$$

such that $\chi(x,i) = the\ value\ corresponding\ to\ i^{th}\ age\ range\ that\ x\ belongs.$ Here, $A = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ is a set of age ranges such that x_1 = "0-14", x_2 = "15-24", x_3 = "25-49", x_4 = "50-64", x_5 = "65-79", and x_6 = ">80" and I is a set of individuals. To illustrate, if an individual x belongs in the age range 15-24, then the age priority score $f_A(x) = \chi(x,2) = 0.8$.

Table 12. Normalized case distribution according to age range

Age Ranges	0-14	15-24	25-49	50-64	65-79	>80
Scores	0.39	0.80	1.00	0.92	0.85	0.76

The scores are obtained by normalizing and merging the ranges < 2, 2-4, and 5-14 as 0-14.

Presence of risk group individuals in the immediate vicinity: This study assigns priority scores in Table 13 according to immediate vicinity levels of individuals that live in the same house. Hence, it sets a priority score for individuals that live in the same house. Hence, the immediate vicinity priority score function f_{IV} is as follows:

$$f_{IV} : I \to [0,1]$$
$$x \to f_{IV}(x) = \begin{cases} \chi(x,i), & i^{th}\ immediate\ vicinity\ type\ that\ x\ lives\ with \\ 0, & x\ does\ not\ live\ with\ any\ immediate\ vicinities \end{cases}$$

such that $\chi(x,i) = the\ value\ corresponding\ to\ i^{th}\ immediate\ vicinity\ type\ that\ x\ lives\ in.$ Here, $IV = \{x_1, x_2, x_3, x_4\}$ is a set of immediate vicinities such that x_1 = "with chronic elderly patients", x_2 = "with chronic young patients", x_3 = "with elders", and x_4 = "with more than ten non-risky individuals" and I is a set of individuals. To illustrate, if an individual x lives with elders, then the immediate vicinity priority score $f_{IV}(x) = \chi(x,3) = 0.4$.

Table 13. Vicinity and risk relation

Immediate Vicinities	Scores
With chronic elderly patients	0.8
With chronic young patients	0.6
With elders	0.4
With > ten non-risky individuals	0.2

Presence of COVID-19 History: How severely an individual with a history of COVID-19 presents symptoms is related to vaccination priority. This study determines priority scores in Table 14 according to their COVID-19 histories. Thus, the COVID-19 histories priority score function f_{CH} is as follows:

$$f_{CH} : I \to [0,1]$$
$$x \to f_{CH}(x) = \chi(x,i)$$

such that $\chi(x,i) = the\ value\ corresponding\ to\ i^{th}\ COVID-19\ history\ type\ that\ x\ has.$ Here, $CH = \{x_1, x_2, x_3, x_4, x_5\}$ is a set of COVID-19 history types such that x_1 = "patient with COVID-19 hospitalized in intensive care", x_2 = "individual who have not had COVID-19", x_3 = "patient with moderate symptoms with COVID-19", x_4 = "patient with mild symptoms of COVID-19", and x_5 = "patient with COVID-19 who did not show any symptoms" and I is a set of individuals. To illustrate, if an individual x has not contracted COVID-19, then the COVID-19 history priority score $f_{CH}(x) = \chi(x,2) = 0.6$.

Table 14. COVID-19 history and risk relation

COVID-19 History	Scores
Patient with COVID-19 hospitalized in intensive care	0.8
Individual with no COVID-19 history	0.6
Patient with moderate symptoms of COVID-19	0.5
Patient with mild symptoms of COVID-19	0.4
Patient with COVID-19 with no symptoms	0.3

Province-District: It is observed that COVID-19 cases increase in direct proportion to the region and the population of the region. Therefore, this study produces priority scores in Table 15 based on two criteria: province and district. Thereby, the province-district priority score function f_{PD} is as follows:

$$f_{PD} : I \rightarrow [0,1]$$
$$x \rightarrow \quad f_{PD}(x) = \chi(x,i,j)$$

such that $\chi(x,i,j) = the\ value\ corresponding\ to\ i^{th}\ and\ j^{th}\ population\ ranges\ of the\ province\ and\ district\ where\ x\ lives.$ Here, $P = \{x_1, x_2, x_3, x_4, x_5\}$ is a set of provinces' population ranges such that x_1 = "$< 10^6$", x_2 = "$10^6 - 3 \cdot 10^6$", x_3 = "$3 \cdot 10^6 - 5 \cdot 10^6$", x_4 = "$5 \cdot 10^6 - 7 \cdot 10^6$", and x_5 = "$\geq 7 \cdot 10^6$", $D = \{d_1, d_2, d_3, d_4, d_5\}$ is a set of districts' population ranges such that d_1 = "$10^4 - 10^5$", d_2 = "$10^5 - 2 \cdot 10^5$", d_3 = "$2 \cdot 10^5 - 3 \cdot 10^5$", d_4 = "$3 \cdot 10^5 - 4 \cdot 10^5$", and d_5 = "$\geq 4 \cdot 10^5$", and I is a set of individuals. To illustrate, if an individual x lives in a province and district with the populations $4 \cdot 10^6$ and $2.5 \cdot 10^5$, respectively, then the province-district priority score $f_{PD}(x) = \chi(x,3,3) = 0.64$.

Table 15. Province-district and risk relation

Province/District	$10^4 - 10^5$	$10^5 - 2 \cdot 10^5$	$2 \cdot 10^5 - 3 \cdot 10^5$	$3 \cdot 10^5 - 4 \cdot 10^5$	$\geq 4 \cdot 10^5$
$< 10^6$	0.40	0.42	0.44	0.46	0.48
$10^6 - 3 \cdot 10^6$	0.50	0.52	0.54	0.56	0.58
$3 \cdot 10^6 - 5 \cdot 10^6$	0.60	0.62	0.64	0.66	0.68
$5 \cdot 10^6 - 10^7$	0.70	0.72	0.74	0.76	0.78
$\geq 10^7$	0.80	0.82	0.84	0.86	0.88

Transportation Preferences: The transportation preferences of individuals affect the number of COVID-19 cases. Therefore, this study sets a score in Table 16 for each possible transportation preference. Thereafter, the transportation preference priority score function f_{TP} is as follows:

$$f_{TP} : I \rightarrow [0,1]$$
$$x \rightarrow \quad f_{TP}(x) = \chi(x,i)$$

such that $\chi(x,i) = the\ value\ corresponding\ to\ i^{th}\ transporting\ type\ that\ x\ prefers.$ Here, $TP = \{x_1, x_2, x_3, x_4, x_5\}$ is a set of transporting types such that x_1 = "using public transport 4+ per day", x_2 = "using public transport 4 times a day", x_3 = "using public transport 2 times a day", x_4 = "travelling by private vehicle", and x_5 = "not travelling" and I is a set of individuals. To illustrate, if an individual x travels by her/his own private vehicle, then the transportation preference priority score $f_{TP}(x) = \chi(x,4) = 0.2$.

Table 16. Transportation preferences and risk relation

Transportation Preferences	Scores
Using public transport 4+ per day	0.8
Using public transport 4 times a day	0.6
Using public transport 2 times a day	0.4
Using private vehicle	0.2
Not Travelling	0

Occupations: This study considers priority scores corresponding to the classification of occupations in Table 17 provided in [16]. Therefore, the occupation priority score function f_O is as follows:

$$f_O : I \rightarrow [0,1]$$
$$x \rightarrow \quad f_O(x) = \chi(x,i)$$

such that $\chi(x,i) = the\ value\ corresponding\ to\ i^{th}\ occupation\ that\ x\ has.$ Here, $O = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$ is a set of occupations such that x_1 = "health workers", x_2 = "nursing homes and protection homes", x_3 = "Ministry of National Defense, Ministry of Interior, individuals in strategic positions", x_4 = "municipal police, private security personal, Ministry of Justice, correctional facilities", x_5 = "education sector (teachers and faculty), food sector workers and bakeries, caterers, food and beverage processing plans, etc. registered with the Social Security Institution, transportation sector, workers registered with the Social Security Institution", x_6 = "workers in mass, crowded areas", x_7 = "businesses with less than ten employees", x_8 = "other/home-office", and x_9 = "unemployed" and I is a set of individuals. To illustrate, if an individual x is a teacher, then the occupation priority score $f_O(x) = \chi(x,5) = 0.5$.

Table 17. Vaccination group ranking

Type	Occupations	Scores
	Health Workers	1.0
	Nursing homes and Protection homes	0.8
A1, A2, A3	Ministry of National Defense, Ministry of Interior, Individuals in Strategic Positions	0.7
A4, A5, A6	Municipal Police, Private Security Personal, Ministry of Justice, Correctional Facilities	0.6
A7, A8, A9	Education Sector (Teachers and Faculty), Food Sector Workers and Bakeries, Caterers, Food and Beverage Processing Plans, etc. registered with the Social Security Institution, Transportation Sector, Workers registered with the Social Security Institution	0.5
A10	Workers in mass, crowded areas	0.4
A11	Businesses with less than ten employees	0.3
A12	Other/ Home-office	0.2
	Unemployed	0

The scores were assigned based on how crowded individuals' workspace is.

Calculation of Vaccination Priority Score: This study calculates the vaccination priority scores via the aforesaid scores in this subsection. Hence, the vaccination priority score function is as follows:

$$f_{VPS} : I \to [0,1]$$

$$x \to f_{VPS}(x) = \frac{NS_{Sd}f_{Sd}(x) + NS_A f_A(x) + NS_{IV}f_{IV}(x) + NS_{CH}f_{CH}(x) + NS_{PD}f_{PD}(x) + NS_{TP}f_{TP}(x) + NS_O f_O(x)}{NS_{Sd} + NS_A + NS_{IV} + NS_{CH} + NS_{PD} + NS_{TP} + NS_O}$$

Here, I is a set of individuals. To illustrate, if the total risk scores of an individual x are $f_{Sd}(x) = 0.82$, $f_A(x) = 0.92$, $f_{IV}(x) = 0.4$, $f_{CH}(x) = 0.5$, $f_{PD}(x) = 0.4$, $f_{TP}(x) = 0$, and $f_O(x) = 0.2$, then the vaccination priority score is as follows:

$$f_{VPS}(x) = \frac{1 \cdot 0.82 + 0.9100 \cdot 0.92 + 0.9919 \cdot 0.4 + 0.8775 \cdot 0.5 + 0.5726 \cdot 0.4 + 0.7364 \cdot 0 + 0.7372 \cdot 0.2}{1 + 0.9100 + 0.9919 + 0.8775 + 0.5726 + 0.7364 + 0.7372} = 0.4925$$

A Hypothetical Scenario: This study considers scores provided in Table 18 for ten individuals to illustrate the performances of the aforesaid vaccination priority score function and the SDM method YE12. Let $I = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$ be a set of individuals and $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ be a set of parameters such that e_1 = "systemic disease", e_2 = "age", e_3 = "presence of risk group individuals in the immediate vicinity", e_4 = "presence of COVID-19 history", e_5 = "province-district", e_6 = "transportation preference", and e_7 = "occupation". Here, the weights of the parameters are $S_{Sd} = 1$, $S_A = 0.9100$, $S_{IV} = 0.9919$, $S_{CH} = 0.8775$, $S_{PD} = 0.5726$, $S_{TP} = 0.7364$, and $S_O = 0.7372$. Then, the *fpfs*-matrix $[a_{ij}]_{11\times 6}$ constructed by these weights and the data in Table 18 is as follows:

$$[a_{ij}] = \begin{bmatrix} 1 & 0.9100 & 0.9919 & 0.8775 & 0.5726 & 0.7364 & 0.7372 \\ 0.97 & 0.80 & 0.4 & 0.6 & 0.68 & 0.6 & 0.8 \\ 0.45 & 1 & 0.2 & 0.8 & 0.52 & 0.4 & 0.6 \\ 0.60 & 0.39 & 0.2 & 0.5 & 0.54 & 0.6 & 0 \\ 0.60 & 0.76 & 0.6 & 0.5 & 0.76 & 0.8 & 0.4 \\ 0.74 & 0.85 & 0.4 & 0.3 & 0.72 & 0 & 0.4 \\ 0.74 & 0.92 & 0.8 & 0.6 & 0.84 & 0.2 & 0.5 \\ 0.67 & 0.80 & 0.8 & 0.3 & 0.82 & 0.6 & 0.5 \\ 0.97 & 0.76 & 0.2 & 0.8 & 0.68 & 0.4 & 0.8 \\ 0.45 & 0.76 & 0.6 & 0.6 & 0.42 & 0.4 & 1 \\ 0.60 & 0.39 & 0.8 & 0.4 & 0.46 & 0.8 & 0.3 \end{bmatrix}$$

Then, this study applies the SDM method YE12 to $[a_{ij}]$. The decision set, score matrix, and ranking order produced by YE12, respectively, are as follows:

$$\{^{0.6939}x_1, {}^{0.5656}x_2, {}^{0.4022}x_3, {}^{0.6256}x_4, {}^{0.4945}x_5, {}^{0.6684}x_6, {}^{0.6411}x_7, {}^{0.6584}x_8, {}^{0.6069}x_9, {}^{0.5447}x_{10}\}$$

$$[s_{i1}] = [0.6939 \quad 0.5656 \quad 0.4022 \quad 0.6256 \quad 0.4945 \quad 0.6684 \quad 0.6411 \quad 0.6584 \quad 0.6069 \quad 0.5447]^T$$

and

$$x_3 \prec x_5 \prec x_{10} \prec x_2 \prec x_9 \prec x_4 \prec x_7 \prec x_8 \prec x_6 \prec x_1$$

Here,

$$s_{71} = f_{VPS}(x_7) = \frac{1 \cdot 0.67 + 0.9100 \cdot 0.8 + 0.9919 \cdot 0.8 + 0.8775 \cdot 0.3 + 0.5726 \cdot 0.82 + 0.7364 \cdot 0.6 + 0.7372 \cdot 0.5}{1 + 0.9100 + 0.9919 + 0.8775 + 0.5726 + 0.7364 + 0.7372} = 0.6411$$

Table 18. Priority scores for ten individuals

Individuals / Criteria	Systemic Disease	Age	Presence of risk group individuals in the immediate	Presence of COVID-19 History	Province-District	Transportation Preference	Occupation
x_1	0.97	0.80	0.4	0.6	0.68	0.6	0.8
x_2	0.45	1	0.2	0.8	0.52	0.4	0.6
x_3	0.60	0.39	0.2	0.5	0.54	0.6	0
x_4	0.60	0.76	0.6	0.5	0.76	0.8	0.4
x_5	0.74	0.85	0.4	0.3	0.72	0	0.4
x_6	0.74	0.92	0.8	0.6	0.84	0.2	0.5
x_7	0.67	0.80	0.8	0.3	0.82	0.6	0.5
x_8	0.97	0.76	0.2	0.8	0.68	0.4	0.8
x_9	0.45	0.76	0.6	0.6	0.42	0.4	1
x_{10}	0.60	0.39	0.8	0.4	0.46	0.8	0.3

Real-Life Interpretation: This study applies the SDM method YE12 to a real-life data derived from an online survey by Google Forms with 100 participants, 50 of whom are COVID-19 positive and the rest of whom are COVID-19 negative. The participants are asked to give information about their systemic diseases, ages, whether they are living in the immediate vicinity, the presence of their COVID-19 history, province-district they currently live in, their transportation preferences, and occupations. The results are provided in Table 19. To evaluate the performance of the SDM method YE12, this study utilizes the following validity function:

$$V : P \rightarrow [0,1]$$

$$x \rightarrow V(x) = \begin{cases} 1, & x \; is \; COVID-19 \; positive \; and \; the \; order \; of \; x < 75 \\ 1, & x \; is \; COVID-19 \; negative \; and \; the \; order \; of \; x > 25 \\ 0, & otherwise \end{cases}$$

Afterwards, it calculates the validity score $VS = \frac{1}{|P|}\sum_{x \in P} V(x)$. For this survey, the validity score is $VS = 0.96$. For example, for the second participant x_2, $V(x_2) = 1$ because he/she is COVID-19 positive and her/his order is less than 75. Similarly, for the participant x_{24}, $V(x_{24}) = 0$ because he/she is COVID-19 negative and her/his order is not greater than 25.

Table 19. Survey results for vaccination priority

Participants No / Criteria	Systemic disease	Age	Immediate vicinity	COVID-19 history	Province/ district	Transportation preference	Occupation	YE12's scores	COVID-19
1	0.97	0.85	0.8	0.8	0.48	0	0.7	0.6918	+
2	0.74	0.8	0.8	0.5	0.58	0.6	0.2	0.6217	+
3	0.6	0.8	0.2	0.6	0.88	0.8	0.4	0.5906	+
4	0.67	1	0	0.4	0.88	0.6	0.7	0.5824	+
5	0.6	0.92	0.8	0.5	0.46	0.2	0.4	0.5793	+
6	0.6	0.92	0.8	0.5	0.68	0.2	0.2	0.5757	+
7	0.67	0.92	0.4	0.6	0.82	0.2	0.4	0.5737	+
8	0.74	0.85	0.4	0.5	0.86	0.2	0.4	0.5636	+
9	0.6	1	0	0.5	0.78	0.6	0.4	0.5376	+
10	0.97	0.8	0	0.5	0.88	0.2	0.4	0.5292	+
11	0.74	1	0.2	0.5	0.58	0.2	0.4	0.5255	+
12	0.6	0.92	0.8	0.4	0.5	0.2	0	0.5176	+
13	0.67	1	0	0.6	0.46	0.6	0.2	0.5080	+
14	0.6	1	0	0.4	0.52	0.6	0.4	0.4970	+
15	0.6	0.85	0.8	0.4	0.4	0.2	0	0.4968	+
16	0.67	0.92	0	0.5	0.5	0.6	0.2	0.4843	+

17	0.6	0.85	0.4	0.4	0.42	0.6	0	0.4813	+
18	0.6	1	0	0.5	0.56	0.2	0.5	0.4781	+
19	0.6	0.85	0	0.4	0.88	0.2	0.5	0.4711	+
20	0.6	0.85	0	0.8	0.88	0.2	0	0.4680	+
21	0.6	1	0	0.5	0.7	0.2	0.3	0.4666	+
22	0.6	1	0	0.5	0.54	0.2	0.4	0.4635	+
23	0.67	0.92	0	0.5	0.54	0.2	0.4	0.4630	+
24	0	0.92	0	0.6	0.88	0.6	0.5	0.4597	-
25	0.6	0.92	0	0.5	0.88	0.2	0.2	0.4591	+
26	0.6	0.92	0.4	0.4	0.58	0.2	0	0.4573	+
27	0.6	0.92	0	0.8	0.4	0.2	0.2	0.4571	+
28	0.6	0.92	0	0.8	0.4	0.2	0.2	0.4571	+
29	0.74	1	0	0.4	0.88	0.2	0	0.4553	+
30	0.6	0.92	0	0.4	0.58	0.2	0.5	0.4525	+
31	0	1	0.4	0.5	0.78	0.2	0.4	0.4522	+
32	0.6	0.92	0	0.5	0.42	0.2	0.5	0.4519	+
33	0	0.85	0.8	0.6	0.88	0	0	0.4459	+
34	0.6	0.85	0	0.4	0.86	0	0.5	0.4438	+
35	0.6	0.92	0	0.8	0.52	0.2	0	0.4436	+
36	0.6	0.92	0	0.5	0.68	0.2	0.2	0.4394	+
37	0	1	0	0.6	0.5	0.6	0.5	0.4348	-
38	0.6	0.85	0	0.8	0.5	0.2	0	0.4307	+
39	0.6	0.92	0	0.3	0.88	0.2	0.2	0.4290	+
40	0	1	0	0.6	0.82	0.6	0.2	0.4283	-
41	0	1	0	0.6	0.4	0.6	0.5	0.4250	-
42	0.6	0.85	0	0.6	0.6	0	0.3	0.4231	+
43	0	0.92	0	0.5	0.78	0.2	0.8	0.4222	+

Table 19. (Continued) Survey results for vaccination priority

Participants No / Criteria	Systemic disease	Age	Immediate vicinity	COVID-19 history	Province/ district	Transportation preference	Occupation	YE12's scores	COVID-19
44	0	0.92	0	0.6	0.88	0.6	0.2	0.4217	-
45	0.74	0.85	0	0.3	0.4	0.6	0	0.4201	+
46	0	0.92	0	0.6	0.82	0.6	0.2	0.4158	-
47	0	1	0	0.6	0.88	0.6	0	0.4089	-
48	0	1	0	0.6	0.88	0.6	0	0.4089	-
49	0	1	0	0.6	0.62	0.6	0.2	0.4087	-
50	0.45	0.92	0	0.5	0.88	0.2	0	0.4080	+
51	0	0.8	0	0.6	0.5	0.6	0.5	0.4036	-
52	0	0.8	0	0.6	0.88	0.6	0.2	0.4030	-
53	0	0.92	0	0.6	0.4	0.6	0.4	0.3999	+
54	0	0.92	0.2	0.6	0.8	0.2	0.2	0.3974	-
55	0	1	0	0.6	0.5	0.6	0.2	0.3969	-
56	0	0.8	0	0.6	0.8	0.6	0.2	0.3951	-
57	0	1	0	0.5	0.86	0.2	0.4	0.3919	+
58	0.45	0.92	0	0.5	0.6	0.2	0	0.3805	+
59	0	1	0	0.6	0.84	0.2	0.2	0.3797	-
60	0	0.8	0	0.6	0.62	0.6	0.2	0.3774	-
61	0	1	0	0.6	0.42	0.2	0.5	0.3764	-
62	0	1	0	0.6	0.8	0.2	0.2	0.3758	-
63	0	0.8	0	0.6	0.84	0.6	0	0.3737	-
64	0	0.8	0	0.6	0.82	0.6	0	0.3718	-
65	0	0.8	0	0.6	0.62	0.2	0.5	0.3648	-
66	0	1	0	0.6	0.42	0.6	0	0.3637	-

67	0	0.92	0	0.4	0.82	0.2	0.4	0.3605	+
68	0	0.85	0	0.5	0.6	0.2	0.5	0.3556	+
69	0	1	0	0.4	0.52	0.2	0.4	0.3435	+
70	0	0.8	0	0.6	0.4	0.2	0.5	0.3432	-
71	0	1	0	0.6	0.62	0.2	0	0.3328	-
72	0	1	0	0.6	0.62	0.2	0	0.3328	-
73	0	1	0	0.6	0.6	0.2	0	0.3308	-
74	0	0.8	0	0.6	0.62	0.2	0.2	0.3269	-
75	0	0.85	0	0.6	0.4	0	0.5	0.3257	+
76	0	1	0.2	0.6	0.4	0	0	0.3199	-
77	0	1	0.2	0.6	0.4	0	0	0.3199	-
78	0	1	0	0.6	0.62	0	0	0.3075	-
79	0	1	0	0.6	0.6	0	0	0.3056	-
80	0	0.8	0	0.6	0.4	0.2	0.2	0.3052	-
81	0	0.92	0	0.6	0.46	0.2	0	0.3046	-
82	0	0.8	0	0.6	0.6	0.2	0	0.2996	-
83	0	0.92	0	0.6	0.4	0.2	0	0.2987	-
84	0	0.8	0	0.6	0.84	0	0	0.2979	-
85	0	0.92	0	0.6	0.62	0	0	0.2950	-
86	0	1	0	0.6	0.46	0	0	0.2918	-
87	0	1	0	0.6	0.46	0	0	0.2918	-
88	0	0.8	0	0.6	0.52	0.2	0	0.2917	-
89	0	0.8	0.2	0.6	0.4	0	0	0.2887	-
90	0	0.8	0	0.6	0.46	0.2	0	0.2858	-
91	0	0.8	0	0.6	0.46	0.2	0	0.2858	-

92	0	0.8	0	0.6	0.46	0.2	0	0.2858	-
93	0	0.92	0	0.6	0.52	0	0	0.2852	-
94	0	0.92	0	0.6	0.42	0	0	0.2754	-
95	0	0.8	0	0.6	0.6	0	0	0.2743	-
96	0	0.92	0	0.6	0.4	0	0	0.2734	-
97	0	0.85	0	0.6	0.5	0	0	0.2723	+
98	0	0.8	0	0.6	0.4	0	0	0.2547	-
99	0	0.8	0	0.6	0.4	0	0	0.2547	-
100	0	0.85	0	0.4	0.42	0	0	0.2343	+

5. Findings & Conclusions

This study successfully dealt with the rapid diagnosis of possible contagions, planning of follow-up methods and more developed treatment services, and planning an individual-specific vaccination priority by machine learning and statistical methods.

Diagnosis of COVID-19: This study employed mFPFS-CMC, the modified FPFS-CMC which is prominent among the well-known classifiers in medical diagnosis to diagnose COVID-19, and the dataset "Symptoms and COVID Presence (May 2020 data)" provided in Kaggle Data Repository. This is the first study to apply this classifier to COVID-19. The simulation results in Table 1 showed that mFPFS-CMC can be successfully applied to diagnose COVID-19 and it has a running time advantage of up to 70% over FPFS-CMC. This study then presented the accuracy, sensitivity, and specificity results of mFPFS-CMC. Although the sensitivity results of mFPFS-CMC are below 90%, its accuracy and specificity results are above 90%. The results showed that mFPFS-CMC is reliable and practical in medical diagnosis. Consequently, it has become more practical, timesaving, and far less costly to diagnose COVID-19 with the help of mFPFS-CMC.

Follow-Up Treatment Priority in COVID-19 Patients: This study constructed six risk score functions related to age, hypertension, cardiovascular disease, cancer, chronic kidney failure, and diabetes using the data provided in [11,12,14,15,17,18,40-47,49]. It then proposed a treatment priority score function to utilize in follow-up treatment priority in the presence of COVID-19 patients using the aforesaid six risk score functions. This study achieved developing a methodology that calculates each patient's risk score to provide better follow-up and treatment services. Afterward, it applied the SDM method YE12 to a hypothetical scenario. The results showed that the method herein is viable to rank COVID-19 patients in terms of treatment priority.

Vaccination Priority Planning: This study examined the vaccination process based on an individual-specific perspective via vaccination priority scores, calculated by considering the aforesaid seven criteria, i.e., systemic disease, age, presence of risk group individuals in the immediate vicinity, presence of COVID-19 history, province-district, transportation preference, and occupation. Moreover, it proposed a multi-dimensional vaccination priority algorithm to be used in a possible vaccine crisis in the case that a new variant that is insensitive to the vaccine or for booster dose planning. This study then presented a hypothetical and a real-life problem obtained by an online survey, conducted by Google Forms. The results manifested that this algorithm has 96% validity.

Suggestions: Although the first section of the project utilized three datasets consisting of 227, 2779, and 5434 patients, the others used the data of up to 200 participants. Increasing the number of participants can positively affect the validity of the results. Since the datasets herein are imbalanced, the sensitivity and specificity results can be improved by balancing the datasets. In general, increasing the number of the considered studies can produce more sensitive results than the results herein. Moreover, the treatment priority method can also be applied to a real-life dataset. All these methodologies can be adapted to reflect more on the abilities of *fpfs*-matrices. A software program can be derived from this study to use the health system and e-Pulse system, the personal health record system used in Turkiye.

Author Contributions

Zeynep Parla Parmaksız produced the main conceptual ideas, developed the theoretical framework, and carried out the simulations. Burak Arslan and Samet Memiş improved the theoretical framework and simulations. Serdar Enginoğlu encouraged the authors to investigate the applications of the soft decision-making via *fpfs*-matrices to machine learning and supervised the findings of this study. All the authors discussed the results and contributed to the final paper.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The authors thank Beşiktaş Arts and Sciences Centre and The Scientific and Technological Research Council of Turkiye (TUBİTAK) for their valuable support. This paper was derived from the study presented by Zeynep Parla Parmaksız at Regeneron International Science and Engineering Fair (ISEF) 2022, Atlanta, GA, USA.

References

[1] C. Cortes, V. Vapnik, *Support-vector Networks*, Machine Learning 20 (3) (1995) 273–297.

[2] J. M. Keller, M. R. Gray, J. A. Givens, *A Fuzzy K-nearest Neighbor Algorithm*, IEEE Transactions on Systems, Man, and Cybernetics 15 (1985) 580–585.

[3] Y. Freund, R. E. Schapire, *A Decision-theoretic Generalization of On-line Learning and an Application to Boosting*, Journal of Computer and System Sciences 55 (1) (1997) 119–139.

[4] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and Regression Trees. 3rd Edn., CRC Press, Wadsworth, 1998.

[5] B. Handaga, H. Onn, T. Herawan, *FSSC: An Algorithm for Classifying Numerical Data using Fuzzy Soft Set Theory*, International Journal of Fuzzy System Applications 2 (4) (2012) 29–46.

[6] S. A. Lashari, R. Ibrahim, N. Senan, *Medical Data Classification using Similarity Measure of Fuzzy Soft Set-based Distance Measure*, Journal of Telecommunication, Electronic and Computer Engineering 9 (2–9) (2017) 95–99.

[7] I. T. R. Yanto, R. R. Seadudin, S. A. Lashari, Haviluddin *A Numerical Classification Technique Based on Fuzzy Soft Set using Hamming Distance*, in: R. Ghazali, M. M. Deris, N. M. Nawi, J. H. Abawajy (Eds.), Third International Conference on Soft Computing and Data Mining, Johor, Malaysia, 2018, pp. 252–260.

[8] S. Memiş, S. Enginoğlu, U. Erkan, *A Classification Method in Machine Learning Based on Soft Decision-Making via Fuzzy Parameterized Fuzzy Soft Matrices*, Soft Computing 26 (2022) 1165–1180.

[9] S. Memiş, 2021. FPFS-CMC. GitHub Repository. Retrieved from https://github.com/sametmemis/FPFS-CMC.git

[10] D. Dua, C. Graff, 2019. UCI Machine Learning Repository [Database].

[11] A. T. Erol, *Predictive Indicators and Risk Factors in COVID-19 Mortality in Intensive Care: Retrospective Observatory Study*, Master's Thesis in Medicine, University of Health Sciences (2020) İstanbul Turkiye.

[12] F. Çoktaş, *Assessment of Risk Factors Related to COVID-19 Disease in Healthcare Workers*, Master's Thesis in Medicine, University of Health Sciences (2020) İstanbul Turkiye.

[13] M. S. Gold, D. Sehayek, S. Gabrielli, X. Zhang, C. McCusker, M. Ben-Shoshan, *COVID-19 and Comorbidities: A Systematic Review and Meta-Analysis*, Postgraduate Medicine 132(8) (2020) 749–755.

[14] M. Parohan, S. Yaghoubi, A. Seraji, M. H. Javanbakht, P. Sarraf, M. Djalali, *Risk Factors for Mortality in Patients with Coronavirus Disease 2019 (COVID19) Infection: A Systematic Review and Meta-Analysis of Observational Studies*, The Aging Male 23 (5) (2020) 1416–1424.

[15] Turkish Ministry of Health, (2020). COVID-19 Weekly Status Report 12/10/2020 – 18/10/2020 Ankara. Retrieved from https://covid19.saglik.gov.tr/Eklenti/39169/0/covid-19-weekly-situation-report---42-weekpdf.pdf?_tag1=B1A9C0854AC47DADC8FF332CC2CB9A336F580D4B

[16] Turkish Ministry of Health, (2021). Vaccination Group Ranking. Retrieved from https://covid19asi.saglik.gov.tr/EN-80295/list-of-covid-19-vaccination-groups.html

[17] Z. Zheng, F. Peng, B. Xua, J. Zhao, H. Liu, J. Peng, Q. Li, C. Jiang, Y. Zhou, S. Liu, C. Ye, P. Zhang, Y. Xing, H. Guo, W. Tang, *Risk Factors of Critical & Mortal COVID-19 Cases: A Systematic Literature Review and Meta-Analysis*, Journal of Infection 81 (2020) 16–25.

[18] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, B. Cao, *Clinical Course and Risk Factors for Mortality of Adult Inpatients with COVID-19 in Wuhan, China: A Retrospective Cohort Study*, The Lancet 395 (2020) 1054–1062.

[19] H. Harikrishnan, (2020, April). Symptoms and COVID Presence (May 2020 data), Version 1. Retrieved June 21, 2021, from https://www.kaggle.com/datasets/hemanthhari/symptoms-and-covid-presence.

[20] J. Thyadi, (2021, July). Covid-19 Symptoms, Version 1. Retrieved June 21, 2022, from https://www.kaggle.com/datasets/jayasreethyadi/covid19-symptoms.

[21] S. Shahane, (2021, Mar). Brazilian Covid Symptomatic Patients Data, Version 1. Retrieved June 21, 2022, from https://www.kaggle.com/datasets/saurabhshahane/brazilian-covid-symptomatic-patients-data.

[22] M. Mehryar, A. Rostamizadeh, A. Talwalkar, Foundations of Machine Learning. 2nd Edn., The MIT Press, London, 2018.

[23] S. Enginoğlu, N. Çağman, *Fuzzy Parameterized Fuzzy Soft Matrices and Their Application in Decision-making*, TWMS Journal of Applied and Engineering Mathematics 10 (4) (2020) 1105–1115.

[24] N. Çağman, F. Çıtak, S. Enginoğlu, *Fuzzy Parameterized Fuzzy Soft Set Theory and Its Applications*, Turkish Journal of Fuzzy Systems 1 (1) (2010) 21–35.

[25] S. Memiş, S. Enginoğlu, U. Erkan, *Numerical Data Classification via Distance-based Similarity Measures of Fuzzy Parameterized Fuzzy Soft Matrices*, IEEE Access 9 (2021) 88583–88601.

[26] S. Enginoğlu, S. Memiş, F. Karaaslan, *A New Approach to Group Decision-making Method Based on TOPSIS under Fuzzy Soft Environment*, Journal of New Results in Science 8 (2) (2019) 42–52.

[27] S. Memiş, 2021. EMK19. GitHub Repository. Retrieved from https://github.com/sametmemis/EMK19.git

[28] S. Enginoğlu, S. Memiş, *Comment on Fuzzy Soft Sets [The Journal of Fuzzy Mathematics 9(3), 2001, 589–602]*, International Journal of Latest Engineering Research and Applications 3 (9) (2018) 1–9.

[29] T. Aydın, S. Enginoğlu, *A Configuration of Five of The Soft Decision-making Methods via Fuzzy Parameterized Fuzzy Soft Matrices and Their Application to A Performance-based Value Assignment Problem*, in: M. Kılıç, K. Özkan, M. Karaboyacı, K. Taşdelen, H. Kandemir, A. Beram (Eds.), Second International Conferences on Science and Technology: Natural Science and Technology (ICONST-NST) Prizren, Kosovo, 2019, pp. 56–67.

[30] T. Aydın, S. Enginoğlu, *Configurations of SDM Methods Proposed between 1999 and 2012: A Follow-up Study*, in: K. Yıldırım (Ed.) Fourth International Conference on Mathematics: "An İstanbul Meeting for World Mathematicians", İstanbul, Turkiye, 2020, pp. 183–202.

[31] S. Enginoğlu, S. Memiş, *A Configuration of Some Soft Decision-making Algorithms via fpfs-matrices*, Cumhuriyet Science Journal 39 (4) (2018) 871–881.

[32] S. Enginoğlu, T. Öngel, *Configurations of Several Soft Decision-making Methods to Operate in Fuzzy Parameterized Fuzzy Soft Matrices Space*, Eskişehir Technical University Journal of Science and Technology A - Applied Sciences and Engineering 21 (1) (2020) 58–71.

[33] S. Enginoğlu, T. Aydın, S. Memiş, B. Arslan, *Operability-oriented Configurations of The Soft Decision-making Methods Proposed between 2013 and 2016 and Their Comparisons*, Journal of New Theory 2021 (34) (2021) 82–114.

[34] S. Enginoğlu, T. Aydın, S. Memiş, B. Arslan, *SDM Methods' Configurations (2017-2019) and Their Application to A Performance-based Value Assignment Problem: A Follow up Study*, Annals of Optimization Theory and Practice 4 (1) (2021) 41–85.

[35] S. Memiş, B. Arslan, T. Aydın, S. Enginoğlu, Ç. Camcı, *A Classification Method Based on Hamming Pseudo-similarity of Intuitionistic Fuzzy Parameterized Intuitionistic Fuzzy Soft Matrices*, Journal of New Results in Science 10 (2) (2021) 59–76.

[36] M. Stone, *Cross-validatory Choice and Assessment of Statistical Predictions*, Journal of the Royal Statistical Society Series B (Methodological) 36 (1974) 111–147.

[37] D. Üstün, A. Toktaş, A. Akdağlı, *Deep Neural Network–based Soft Computing the Resonant Frequency of E–shaped Patch Antennas*, International Journal of Electronics and Communications 102 (2019) 54–61.

[38] T. Fawcett, *An Introduction to ROC Analysis*, Pattern Recognition Letters 27 (2006) 861–874.

[39] U. Erkan, *A Precise and Stable Machine Learning Algorithm: Eigenvalue Classification (EigenClass)*, Neural Computing and Applications 33 (10) (2021) 5381–5392.

[40] F. Saygılı, Hypertension Diagnosis and Treatment Guide, Turkish Society of Endocrinology and Metabolism, 2019.

[41] Z. Rakipoğlu, (2019, April 4). The Death Rate from Heart Diseases in Turkiye is 42 Percent, Anadolu Agency (2020, July 17). Retrieved from https://www.aa.com.tr/tr/saglik/turkiyede-kalp-hastaliklarinda-olum-orani-yuzde-42/1442301.

[42] TKD, (2020). Cardiovascular Risk Calculator, Turkish Society of Cardiology Web Site. Retrieved on August 18, 2020, from https://tkd.org.tr/cardibil/kalp-damar-sagligi/cardivaskuler-risk-hesaplama.

[43] H. Dülek, Z. Tuzcular Vural, I. Gönenç, *Risk Factors in Cardiovascular Diseases*, The Journal of Turkish Family Physician 9 (2) (2018) 53–58.

[44] M. Özdoğan, (2020, October 6). For Some Common Cancers Lifetime Statistics, Retrieved on June 21, 2021, from https://www.drozdogan.com/kanser-yasam-suresi-istatistikleri/

[45] National Cancer Institute (NCI) (2020) Surveillance, Epidemiology, and End Results (SEER) Program. (2020, October 6). Retrieved from https://seer.cancer.gov/statfacts/html/all.html

[46] F. S. Taş, K. Cengiz, E. Erdem, A. Karataş, C. Kaya, *Causes of Mortality in Acute and Chronic Renal Failure*, Fırat Medical Journal 16 (3) (2011) 120–124.

[47] E. Ahlqvist, P. Storm, A. Käräjämäki, M. Martinell, M. Dorkhan, A. Carlsson, P. Vikman, R. B. Prasad, D. M. Aly, P. Almgren, Y. Wessman, N. Shaat, P. Spégel, H. Mulder, E. Lindholm, O. Melander, O. Hansson, U. Malmqvist, A. Lernmark, K. Lahti, T. Forsén, T. Tuomi, A. H. Rosengren, L. Groop, *Novel Subgroups of Adult-onset Diabetes and Their Association with Outcomes: A Data-driven Cluster Analysis of Six Variables*, Lancet Diabetes Endocrinol 6 (2018) 361–369.

[48] Turkish Diabetes Foundation, Diabetes Diagnosis and Treatment Guide. Armoni Nüans Baskı Sanatları A. Ş., (2017).

[49] T. Chen, D. Wu, H. Chen, W. Yan, D. Yang, G. Chen, K. Ma, D. Xu, H. Yu, H. Wang, T. Wang, W. Guo, J. Chen, C. Ding, X. Zhang, J. Huang, M. Han, S. Li, X. Luo, J. Zhao, Q. Ning, *Clinical Characteristics of 113 Deceased Patients with Coronavirus Disease 2019: Retrospective Study*, British Medical Journal 368 (2020) m1091.

[50] S. Enginoğlu, S. Memiş, A Review on Some Soft Decision-making Methods. Proceedings of The International Conference on Mathematical Studies and Applications 2018 Karamanoğlu Mehmetbey University, Karaman, Turkiye, 4-6 October 2018.

[51] S. Memiş, 2021. YE12. GitHub repository. Retrieved from https://github.com/sametmemis/YE12.git