Symptoms-Based Fuzzy-Logic Approach for COVID-19 Diagnosis

Maad Shatnawi
Department of Electrical Engineering Technology
Higher Colleges of Technology
Abu Dhabi, UAE

Anas Shatnawi
Berger-Levrault
Montpellier, France

Zakarea AlShara
Department of Software Engineering
Jordan University of Science and Technology
Irbid, Jordan

Ghaith Husari
Department of Computer Science
East Tennessee State University
Johnson City, Tennessee

Abstract—The coronavirus (COVID-19) pandemic has caused severe adverse effects on the human life and the global economy affecting all communities and individuals due to its rapid spreading, increase in the number of affected cases and creating severe health issues and death cases worldwide. Since no particular treatment has been acknowledged so far for this disease, prompt detection of COVID-19 is essential to control and halt its chain. In this paper, we introduce an intelligent fuzzy inference system for the primary diagnosis of COVID-19. The system infers the likelihood level of COVID-19 infection based on the symptoms that appear on the patient. This proposed inference system can assist physicians in identifying the disease and help individuals to perform self-diagnosis on their own cases.

Keywords—COVID-19; coronavirus diagnosis; fuzzy inference system; fuzzy logic; fuzzy rules; expert systems

I. INTRODUCTION

The Coronavirus disease 2019 (COVID-19) pandemic has seriously affected all aspects of our life including health, education, economy, travel, and entertainment. Spreading rapidly across borders, the coronavirus disease has created a global health crisis and caused numerous death cases all over the world. Coronaviruses are a wide group of viruses that cause sickness starting from the common cold and up to very severe infections leading to death in several situations [1-3]. The common COVID-19 symptoms that are normally seen within 2 to 14 days are cold, dry cough, fever, flu, breathing difficulties, throat sore, and headache [4-8].

Since no therapeutic drug has been confirmed for COVID-19 till this date, the early diagnosis and precautions are essential to control and break down the chain of COVID-19 by immediate isolation of the infected person from the healthy population [9] [10]. The most common methods that global healthcare systems are currently using for Covid-19 identification are Real-Time Polymerize Chain Reaction (RT-PCR) tests in addition to chest Computerized Tomography (CT) scan and X-ray imaging. However, PCR testing requires several hours to get the results and suffers from high false positive rates and false negative rates which means it does not identify all infections, and therefore, PCR should not be used as the only criterion for detecting COVID-19 patients [11-13].

A number of studies reported that chest CT scan has considerably higher COVID-19 diagnosis sensitivity than RT-PCR. On the other hand, CT scans and X-rays have the following limitations. First, CT scans have high false negative rates, as they are unable to distinguish coronary tissue from non-coronary tissue. A large number of COVID-19 patients have normal chest CTs or X-rays. Second, CT scans are unable to discriminate between cancerous tissue, cysts, and coronary tissue. Third, the nonappearance of an anomaly on either a chest X-ray or CT scan does not necessarily eliminate being COVID-19 infected. Fourth, chest CT scans and X-ray cannot precisely differentiate between COVID-19 and other respiratory infections such as seasonal flu. Fifth, CT scanning machines are complex equipment that should be carefully sanitized between potential COVID-19 patients, and there is a risk that the virus remains on the surfaces of CT scanning rooms. Finally, moving potential COVID-19 patients to and from a CT scanning room increases the hazard of spreading the virus within healthcare centers [13-18].

As a result, integrating these methods with a symptom-based diagnosis method will lead to more accurate identification results. In this work, we propose a smart fuzzy inference system to diagnose the COVID-19 based on the symptoms that appear on the patient.

The rest of the paper is organized as follows. Section II presents a background of the fuzzy inference systems and its applications in medical diagnosis. In Section III, we describe the design of our COVID-19 inference system. We evaluate the effectiveness of our approach in Section IV. Section V presents concluding remarks and highlights future directions.

II. FUZZY INFERENCE SYSTEMS: BACKGROUND AND RELATED WORK

Fuzzy Inference Systems (FIS) use fuzzy reasoning in order to represent the knowledge of experts about certain problems in human-like decision-making. These systems are based on fuzzy logic modeling and allow attaining solutions based on
linguistic terms. They are principally useful in cases where human knowledge is available but there is no sufficient information to feed traditional mathematical model variables [19-25]. The fuzzy inference system is made up of four main modules; fuzzification module, knowledge base, inference engine, and defuzzification module as shown in Fig. 1.

The most commonly used fuzzy inference technique is the Mamdani model [26, 27]. The Mamdani fuzzy inference process is performed in four consequent stages; fuzzification, rule evaluation, rule output aggregation, and defuzzification. The fuzzification module maps the crisp input value into a degree of membership of fuzzy sets by applying fuzzification membership functions. A membership function returns a value between zero (for non-membership) and one (for full-membership). The Knowledge base includes the IF-THEN rules that are provided by field experts. The rules are in the form [28]:

IF (A is x) AND (B is y) AND (C is z) … THEN (R is m)  (1)

where A, B, and C represent the input variables while x, y, and z represent the corresponding linguistic terms (e.g., yes, no), R represents the rule output variable and m represent the corresponding linguistic term (e.g., high risk, medium risk, low risk). The defuzzification module converts the output of the inference engine into a crisp output value. The Centroid or the Center of Gravity (COG) method is the most popular defuzzification technique where the weighted average of the area bounded by the aggregated membership function curve of the output variable is considered the crisp output value [13, 29, 30].

The final defuzzified output value using the centroid method is calculated by the following equation:

\[ d_{COG}(M) = \frac{\int_{M}^{D} M(x) x dx}{\int_{M}^{D} M(x) dx} \]  (2)

where M represents the membership function of the output variable.

Fuzzy inference systems have been widely used in the medical diagnosis of different diseases. Lee and Wang [31] presented a fuzzy expert system based on fuzzy ontology as a decision support model for diabetes. Mayilvaganan and Rajeswari [32] proposed high blood pressure fuzzy logic classifier. Ekong et al. [33], Djam et al. [34] and Sharma et al. [35] proposed fuzzy expert systems for malaria diagnosis. Chandra [36] suggested a fuzzy expert system for migraine analysis and diagnosis.

Several fuzzy inference models have been proposed for heart disease diagnosis such as Kumar [37], Adeli and Neshat [38], Kumar and Kaur [39], Kasbe and Pippal [40], Allahverdi et al. [41], Oad et al. [42] and Subbulakshmi et al. [43]. A number of fuzzy inference models were proposed for cancer detection [44] such as Keles et al. [45], Balanica et al. [46] and Latha et al. [47] for breast cancer diagnosis, Lavanya et al. [48] for lung cancer diagnosis, and Saritas et al. [49] for prostate cancer diagnosis. Kolhe et al. [50] presented a fuzzy-logic based approach for disease-diagnosis in crops.

Patel et al. [51], Anand et al. [52], Mishra et al. [53] and Sundararaman et al. [54] proposed Asthma fuzzy diagnosis systems. Damirchi-Darasi et al. [55] proposed fuzzy rule-based expert system for the diagnosis of spinal cord disorders. Kadhim et al. [56] suggested a fuzzy expert system for back pain diagnosis. Zarei et al. [57] suggested a fuzzy modeling and control of HIV infection and [58] proposed diagnosis of HIV/AIDS using Fuzzy Cluster Means Algorithm.


### III. COVID-19 INFERENCE SYSTEM

In this work, a smart fuzzy inference system is proposed for the early detection of COVID-19 based on the patient symptoms including cold, cough, fever, flu, breathing difficulties, throat infection and headache [8]. The proposed system infers the likelihood level of COVID-19 infection based on the symptoms that appear on the patient. The COVID-19 fuzzy inference system is designed by identifying the input and output variables in addition to the fuzzy sets and membership functions of each variable. Afterward, a set of fuzzy rules that are connecting input variables with output variables are set. The proposed inference system is aims at diagnosing the COVID-19 based on the patient data.

We applied the Mamdani Fuzzy model to build the COVID-19 inference system. We define 9 symptoms as the input variables to the inference system. We group these variables into two categories; most common symptoms and less common symptoms. The most common symptoms category includes fever, tiredness, and dry cough while the less common symptoms category includes diarrhea, sore throat, headache, conjunctivitis, loss of taste or smell, and breathing difficulties. The output variable is risk of being COVID-19 infected. The COVID-19 inference system is illustrated in Fig. 2.
A. Membership Functions

Each input variable has two Gaussian membership functions as shown in Table I. The Gaussian membership function gaussmf $[\sigma \mu]$ is defined by its mean $\mu$ and standard deviation $\sigma$. The fever variable is represented by the body temperature which ranges between 36.5 and 42°C as presented in Fig. 3 while each of the remaining input variables has a level in the range from 0 to 5 as indicated in Fig. 4 [62, 63]. The output variable ranges from 0 to 100 and it has four Gaussian membership functions; low risk, medium risk, high risk and very high risk as presented in Fig. 5.

B. Fuzzy Rules

We define the following linguistic fuzzy rules:

- Rule 1: If 3 symptoms of Category-1 AND at least 4 symptoms of Category-2 are present $\rightarrow$ Very high risk of COVID-19 infection.
- Rule 2: If 3 symptoms of Category-1 AND less than 4 symptoms of Category-2 are present $\rightarrow$ High risk of COVID-19 infection.
- Rule 3: If the 3 symptoms of Category-1 are present $\rightarrow$ High risk of COVID-19 infection.
- Rule 4: If 2 symptoms of Category-1 AND at least 2 symptoms of Category-2 are present $\rightarrow$ High risk of COVID-19 infection.
- Rule 5: If 1 symptom of Category-1 AND at least 4 symptoms of Category-2 are present $\rightarrow$ High risk of COVID-19 infection.
- Rule 6: If 2 symptoms of Category-1 AND 1 symptom of Category-2 are present $\rightarrow$ Medium risk of COVID-19 infection.
- Rule 7: If 1 symptom of Category-1 AND 3 symptoms of Category-2 are present $\rightarrow$ Medium risk of COVID-19 infection.
- Rule 8: If 1 symptom of Category-1 AND 2 symptoms of Category-2 are present $\rightarrow$ Medium risk of COVID-19 infection.
- Rule 9: If no symptoms of Category-1 AND 6 symptoms of Category-2 are present $\rightarrow$ Medium risk of COVID-19 infection.
- Rule 10: If 2 symptoms of Category-1 AND no symptoms of Category-2 are present $\rightarrow$ Low risk of COVID-19 infection.
• Rule 11: If only 1 symptom of Category-1 AND 1 symptom of Category-2 are present \( \rightarrow \) Low risk of COVID-19 infection.
• Rule 12: If only 1 symptom of Category-1 is present AND no symptoms of Category-2 \( \rightarrow \) Low risk of COVID-19 infection.
• Rule 13: If no symptoms of Category-1 AND less than 6 symptoms of Category-2 are present \( \rightarrow \) Low risk of COVID-19 infection.

C. Defuzzification of the Output

Based on the input patient symptoms, the inference system initiates a set of fuzzy rules where each rule produces an output. Fuzzy operator “min” was used for generating the output fuzzy set by taking every rule that satisfied the AND operational logic for a given set of input values. Then the output fuzzy set of each rule was combined into a single fuzzy set by the aggregation process. The single fuzzy set was defuzzified into a single numeric output value using the Centroid method to determine the percentage risk level of being COVID-19 infected.

IV. System Testing and Evaluation

This section presents the evaluation of our approach in the following terms: (a) system validation based on the feedback of field experts and (b) system testing using generated mock patient data.

A. System Validation

For the evaluation of the proposed system, we define two main research questions (RQs) and received feedback from the field experts in the healthcare domain using a survey. The RQs are:

• RQ1: Are the COVID-19 Symptoms considered by our approach correct? The goal of this research question is to evaluate the list of COVID-19 symptoms that are used to build our approach.
• RQ2: Are the fuzzy rules correct? The goal of this research question is to evaluate the correctness of the set of fuzzy rules that are used by our approach to decide whether a person is infected by COVID-19 or not.

1) Study design: To ease the accessibility to the survey, we created a web-based survey using Google forms\(^1\). To test the relevance of the survey’s questions before publishing, we conducted a pilot with five candidate participants from the healthcare domain. Each tester practitioner evaluated all questions and their related answers. As a result, they propose minor revisions of the survey. The survey was prepared based on three main sections as follows:

• The first section aims to allow us to describe the participants of this survey by collecting general information about them such as their ages, levels of experience in healthcare and medicine, professions, organizations and countries.
• The second section aims to evaluate the correctness of the set of symptoms related to COVID-19. To this end, each practitioner is asked to select a set (subset) of COVID-19 symptoms among the ones used in our approach. In order to identify COVID-19 symptoms that are not used by our approach, we make it also possible for a practitioner to add new COVID-19 symptoms.
• The last section includes questions related to the evaluation of the fuzzy rules defined in our approach. We asked the participants to evaluate each rule based on three options: Totally Agreed, Partially Agreed and Not Agreed. Totally agreed means that participants confirm our rule are correct following their experiences. Partially agreed refers to the case where participants agreed with this rule, but they do not consider it as correct in all cases. Meaning, the rules are correct for most cases, but not all compared to the COVID-19 patients based on their experiences. Not agreed means that participants do not agree with this rule. That means the rule should be modified.

To avoid prejudice, the survey was distributed to diverse participants from different health professions, levels of experience, organizations and countries. This distribution is based on social media and direct contact of health organizations such as hospitals and medical centers. One hundred participants have been invited to participate in the survey. They also have been requested to forward the survey to their networks.

2) Results

a) Participants: We have received 90 responses in total from participants from 11 different countries on four continents. The participants are also from different professions as presented in Fig. 6 where they cover almost all health domains that are related to COVID-19. Following their experience in the health domain, 58.9\%, 28.8\% and 12.2\% of the participants have respectively more than 10 years, between 5 and 10 years and less than 4 years. The results show that 56.2\%, 30\%, 4.5\% and 3.4\% of the participants work respectively in Hospitals, Medical Centers, Universities and Pharmacies. As a result, the participants are diverse in their professions, type of health organizations, levels of experiences and geographical areas which means that they represent a good-enough sample that does not include prejudice in the answers to the survey questions.

b) RQ1: Are the COVID-19 Symptoms considered by our approach correct?

The results of the survey show that 90\% of the participants agreed with us that these symptoms can be strongly used in COVID-19 diagnoses.

\(^1\) This survey is accessible through: https://docs.google.com/forms/d/e/1FAIpQLScDp2a2uBCJg6366l-7vOFjVd9tXEi_nh8k4QXc5k9bZ2xv6Q/viewform
Fig. 7 shows the results of the evaluation of COVID-19 symptoms we used in our approach. The results show that the participants agreed with us for most of these symptoms. 6 of these symptoms (i.e., fever, breathing difficulties, loss of taste or smell, headache, dry cough, and tiredness) have been selected by more than 82% of the participants where fever is ranked as number one by 92.2% of the participants. The sore throat and diarrhea symptoms have been also selected by a quite number of participants. This means that the symptoms considered in our approach are representative compared to real COVID-19 cases based on the experience of the participants.

As it is allowed for the participants to add extra COVID-19 symptoms, we received only two extra symptoms: the stress and the body pain where each has been selected by one practitioner. These are rare symptoms. Thus, their absence will not negatively impact our approach.

Further, the results show that 85.5% of the participants agreed with us that Fever, Tiredness and Dry cough have more correlation with COVID-19 than the other Symptoms (from Question 8 in the survey). This confirms our decision to include these three symptoms in Category-1 that have higher weights in the fuzzy rules.

c) RQ2: Are the fuzzy rules correct?

Fig. 8 shows the results of evaluating the fuzzy rules. These results show that all of the rules are either partially or totally accepted by more than 80% of the participants. For example, Rule 1 has been accepted by 97.88% (77.88% + 20%) of the participants. This means that the fuzzy rules can be used to build the COVID-19 inference system.

B. System Testing

Based on a certain input of patient symptoms, the inference system initiates a set of fuzzy rules where each rule produces an output. Then, aggregation and defuzzification is performed to generate a single overall output through the process of Centroid calculation. This final output represents the percentage risk of being COVID-19 infected. The proposed system is tested on some mock patient cases and the results are presented in Table II. Fig. 9 illustrates the rule evaluation process for a high-risk case while Fig. 10 and Fig. 11 illustrate the rule evaluation process for medium-risk and low-risk cases, respectively.

The 3D surface view for the rule that relates the COVID-19 infection risk to both fever and tiredness symptoms is demonstrated in Fig. 12. The dark blue surface represents a very low infection risk (less than 54%) when both symptoms are low. The green surface represents a higher risk (between 54% and 60%) when one of the two symptoms is high. The yellow surface represents a 65% risk when both symptoms are high. The 3D surface view for the rule that relates the COVID-19 infection risk to both breathing difficulties and sore throat symptoms is demonstrated in Fig. 13. Unfortunately, the 3D surface viewer can show the relationship of the output variable with only two input variables and since high infection risk (more than 65%) exists when at least 3 variables are high, this case cannot be viewed through this tool.

![Fig. 6. Professions of the Survey Participants.](image)

![Fig. 7. The Results of Evaluating the Symptoms of COVID-19.](image)

![Fig. 8. Evaluation of Fuzzy Rules.](image)
# Table II: System Testing with Possible Values for Patient Symptoms

<table>
<thead>
<tr>
<th>Case</th>
<th>Fever (36-42°C)</th>
<th>Tiredness (0-5)</th>
<th>Dry cough (0-5)</th>
<th>Diarrhea (0-5)</th>
<th>Sore throat (0-5)</th>
<th>Headache (0-5)</th>
<th>Conjunctivitis (0-5)</th>
<th>Loss of taste or smell (0-5)</th>
<th>Breathing difficulties (0-5)</th>
<th>Output (Risk Percentage)</th>
<th>Risk level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38.5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>84.84</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>92.33</td>
<td>Very high</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>73.35</td>
<td>Medium</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>67.52</td>
<td>Medium</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>51.50</td>
<td>Low</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>87.14</td>
<td>High</td>
</tr>
<tr>
<td>7</td>
<td>39.5</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>71.41</td>
<td>Medium</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>89.37</td>
<td>Very high</td>
</tr>
<tr>
<td>9</td>
<td>38</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>61.77</td>
<td>Medium</td>
</tr>
<tr>
<td>10</td>
<td>38</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>51.73</td>
<td>Low</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>73.22</td>
<td>Medium</td>
</tr>
<tr>
<td>12</td>
<td>37</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>54.52</td>
<td>Low</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>87.10</td>
<td>High</td>
</tr>
</tbody>
</table>

![Fig. 9. Rule Evaluation for a High-Risk Case.](image)

![Fig. 10. Rule Evaluation for a Medium-Risk Case.](image)
Fig. 11. Rule Evaluation for a Low-Risk Case.

Fig. 12. COVID-19 Infection Risk vs. Fever and Tiredness.

Fig. 13. COVID-19 Infection Risk vs. Breathing Difficulties and Sore Throat.

V. CONCLUSIONS AND FUTURE DIRECTIONS

We proposed a smart fuzzy inference system for the initial identification of COVID-19. The system infers the risk level of being COVID-19 infected based on the symptoms that appear on patients. The symptoms considered are fever, tiredness, and dry cough, diarrhea, sore throat, headache, conjunctivitis, loss of taste or smell, and breathing difficulties. This inference system can assist physicians in identifying the disease. Although the proposed system cannot provide a very accurate COVID-19 identification, it can be integrated with other identification techniques such as PCR test and CT scan to work together to confirm infected cases.

In future work, we are planning to implement this diagnosis system into a web application to allow individuals to perform self-diagnosis on their own cases. The work can be extended to include other patient data such as blood pressure, breathing air peak-flow-rate, and having a chronic disease. One of the interesting future directions is to apply data mining techniques to generate fuzzy rules from patient data.

REFERENCES


