ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTIVE DATA MINING RULE-BASED CLASSIFIERS MODEL FOR NOVEL CORONAVIRUS (COVID-19) INFECTED PATIENTS' RECOVERY IN THE KINGDOM OF SAUDI ARABIA

HUSSAIN MOHAMMAD. ABU-DALBOUH [1], SULAIMAN ABDULLAH. ALATEYAH [2]

[1,2] Department Of Computer Science, College Of Science And Arts, Qassim University, Unaizah, SAUDI ARABIA

E-mail: [1]hussainmdalbouh@yahoo.com.,[2]dr.salateyah@gmail.com

ABSTRACT

The coronavirus disease (COVID-19) pandemic, which appeared in Wuhan, China, in December 2019, is quickly spreading worldwide, with over 56 million cases as of mid-November 2020. There is no scientifically validated vaccine or drug for COVID-19; however, patients have recovered with the help of antibiotic drugs, anti-viral drugs, chloroquine, and supplements such as vitamin C. It is now evident that the world needs a quicker and better way to contain and handle the further spread of COVID-19 worldwide with the assistance of non-clinical methods including data mining approaches, augmented intelligence, and other artificial intelligence techniques in order to alleviate the enormous burden on the healthcare system, and also provide the most promising means for the patients' diagnoses. The first objective of this research was to consider a real dataset of coronavirus patients, which included regular statistical reports and also clinical data on patients, which could bring about crucial collaborations within the global research community and the discovery of new insights into tackling the outbreak. Then, using the epidemiological dataset of COVID-19 Kingdom of Saudi Arabia patients, data mining models were constructed for predicting the recovery of COVID-19 infected patients. The Cross-Industry Standard Process for Data Mining was used as the framework for the data mining classification of patients' health care data. The process for generating the classification rules was based on the decision tree algorithm and the created rules were evaluated for use by health care administration for predicting the maximum and minimum number of days for the recovery of COVID-19 patients, the age group of patients at high risk of not recovering from the COVID-19 disease, those expected to recover from the COVID-19 disease, and those likely to quickly recover from the COVID-19 disease. Three different classification methods were tested, i.e., Bayes Net-D, naive Bayes, and J48. As a percentage of the correctly identified cases using the three separate algorithms, the overall accuracies of the evaluation results were 74.7748%, 81.0811%, and 93.6937%, respectively.

Keywords: *Artificial Intelligance, Machine Learning, Classification; Clinical Data; Algorithm, Disease, Healthcare; Coronavirus Dataset*

1. INTRODUCTION

China's 2019 coronavirus disease (COVID-19) epidemic is a worldwide threat to healthcare [1,2] and by some standards the largest outbreak of pneumonia, despite the 2003 outbreak of the severe acute respiration syndrome (SARS). The overall number of cases and deaths exceeded that of SARS within weeks of the initial outbreak [3]. In November and December 2019, the outbreak was first discovered when clusters of pneumonia cases of unidentified etiology were epidemiologically linked to a seafood market and untraced exposures within the city of Wuhan of Hubei Province [4]. Since then, within and outside of Wuhan, the range of cases has continued to increase rapidly, expanding to all 34 parts of China by 30 January 2020, which was the same day that the world health organization announced that the COVID-19 disease was a worldwide public health crisis [5]. The coronavirus epidemic, code-named COVID-19, is a virus-induced infectious disease, a member of the beta coronavirus family, called severe acute respiratory syndrome coronavirus 2, previously referred to as the new 2019 coronavirus [6–8]. The disease is an extremely infectious disease that has attracted global public attention. The modeling of such diseases is extremely important for predicting the impact of a disease. Although traditional statistical modeling may provide effective models, they do not understand the intricacies found within the data.

An exponential rise in the number of cases of COVID-19 is expected to easily over-whelm healthcare systems, exposing the state of health facilities in developing countries [9–11]. An awareness of the need to control the rate of infections has resulted in lockdowns in different economies around the world, resulting in travel restrictions, social distancing, the deployment of rapid testing platforms, and communication tracing. It is important to classify and isolate infected cases in order to avoid rapid transmission and to flatten the transmission curve [12]. Therefore, detecting cases and determining the rate of disease transmission among those infected individuals remains crucial in order to control the spread and to reduce the death rate.

However, no scientifically validated drug or vaccine is available to treat COVID-19, and therefore other non-clinical or non-medical therapeutic strategies, such as data mining techniques, machine learning, and expert systems, among other artificial intelligence techniques, are urgently required to control and avoid further outbreaks of the COVID-19 epidemic.

Data mining is an advanced technique of artificial intelligence which is used to detect novel, useful, and effective hidden patterns or knowledge from a dataset [13–16]. The technique identifies relationships and knowledge or patterns among the datasets in multiple or single datasets [5,17]. It has also been widely used for predicting and diagnosing many diseases, which include coronavirus severe acute respiratory syndrome and coronavirus Middle East respiratory syndrome that were initially identified in 2003 and 2012, respectively. The huge worldwide dataset linked daily to the 2019-nCoV pandemic is a treasured resource to be mined and analysed for valuable, valid, and novel knowledge or pattern extraction for better decision-making to control the spread of the COVID-19 pandemic. Data mining has been widely used in many different applications in the healthcare field, such as projecting patient performance, modeling health outcomes, hospital rating, and measuring the efficacy of treatment and infection prevention, stability, and recovery [18–20].

1.1. Motivation and Study Questions

The world health organization clinical data platform, as well as international and national databases, are vital for understanding this virus and how we can collectively tackle its devastating effects. Many health care sector organizations have requested that their members submit data so that researchers can learn as much as possible about the natural history of the virus, its prognostic factors, and any interventions that may influence the outcome. The coronavirus dataset includes data from more than 90 countries constructed from various reliable sources, reflecting each country's geographical, climate, health, economic, and demographic factors that contribute to accelerating/slowing the spread of COVID-19. Each month, the dataset is updated with the latest number of COVID-19 cases, deaths, and tests. In addition, the dataset is freely available and updated regularly with new case numbers and information on latitude and longitude. Unfortunately, these datasets are statistical daily reports and many researchers do not consider them to be as valuable as patient clinical data.

The following list of comments and inquiries are examples from researchers:

- *"Where can we find a shared dataset of coronavirus patients? Are there any databases/websites that share COVID-19 patients' data?"*
- *"I need the clinical symptom data of COVID-19 confirmed cases or suspected cases."*
- *"I have the same question, specifically, I'm looking for high-frequency clinical data of COVID-19 patients (e.g., blood oxygen level and blood pressure)."*
- *"Do you know if a Covid-19 dataset is available somewhere? I have only found daily statistical data but I would like access to single patient data. Does anyone know about it?"*
- *"… but these three datasets are about statistical data. I need patients' data for machine learning classification, for example, a patient's symptoms, specific blood data, age, sex, pneumonia, cough."*
- *"… These are general statistics and not patients' data. Do you know other sources?"*
- *"Did you talk to the Italian authorities? Do you think they could give that kind of data?"*
- *"The Italian Radiology Society has posted limited clinical information on their website."*
- *"They are not about patients' clinical data but only statistical daily reports."*
- *"Is any laboratory test result dataset available?"*
- *"In Turkey, a few cases have been reported so far, and it is a bit problematic to use them due to ethical considerations."*

- *"I'm currently also looking for patient information for my project. Were you able to find a dataset?"*
- *"I'm looking for more personal data such as age, sex, travel history, previous symptoms, etc., or some combination."*
- *"Does anyone have a clinical dataset of COVI-19 patients?"*
- *"In many cases, these data are not exposed to the public due to policy reasons. I think the respective departments in the AI community would be better for finding patterns which may help doctors to prioritize treatment."*
- *"I am looking for laboratory test report data."*
- *"Is there any dataset that includes 'fever, travel history, sore throat, contact history, etc.? We are doing research on the diagnosis system of COVID-19 using supervised learning."*
- *"I am looking for a dataset that includes cough samples of Covid-19 positive patients."*

As researchers search various databases to tackle the coronavirus threat, it has be-come important to have timely access to accurate data. As the threat intensifies, open access to reliable public data has become imperative and is necessary for a deeper understanding of the current crisis. The aim of this study is to provide a real dataset of coronavirus patients' clinical data which, in turn, could result in crucial collaborations within the global research community and the discovery of new insights for tackling the outbreak. In addition, data mining models are developed using epidemiological datasets of COVID-19 patients from the Kingdom of Saudi Arabia for predicting their recovery.

There is much worldwide publicity about the coronavirus and the stakes are high for all mankind. Nevertheless, there are still important questions that are unanswered, for example, "What is coronavirus?" "Is there any effective model for prediction?" "What are the datasets motivating researchers for modeling in a machine learning algorithm?" and "How can clinical data about coronavirus patients be collected? The aim of this study is to contribute to our current understanding of the impact of coronavirus, which is in-conclusive.

1.2. Purpose of the Study

Coronaviruses are a vast family of viruses that can cause diseases such as Middle East respiratory syndrome and severe acute respiratory syndrome ranging from the common cold to more severe diseases. The common COVID-19 symptoms include fever, cough, shortness of breath, and often pneumonia. In persons with immunodeficiencies, elderly people, and people with chronic diseases such as cancer, diabetes, and lung diseases, COVID-19 can cause serious complications.

The first objective in this study is to provide a real dataset of coronavirus patients' clinical data which could bring about crucial collaborations within the global research community and the discovery of new insights for tackling the outbreak. The second objective is to develop a data mining model to predict the recovery of COVID-19 infected patients using the COVID-19 Kingdom of Saudi Arabia epidemiological dataset.

The structure of this paper is as follows: The background and literature review are provided in Section II; the proposed model is described in Section III; the experiments and evaluation of this study are discussed in Section IV; and conclusions are presented in the last section.

2. LITERATURE REVIEW

In [21], the authors proposed a decision tree algorithm trained with different training and testing datasets. Machine learning and artificial intelligence methods were used to solve complex tasks, including many application domains, for example, computer vision, image processing, natural language processing, or market analysis and numerous transcript datasets [22,23]. Information technology methods play an important role for supporting management systems and for shaping the performance of the organization as a whole [24]. In order to allow cybersecurity experts to address the ever-evolving challenge posed by opponents, machine learning and deep learning demonstrate promise [25]. A study by [26] used artificial intelligence to create and deliver available museum and cultural heritage site perspectives [27] and proposed a prediction method based on a cluster algorithm for predicting when a failure would occur, based on data from a time series of bearings. The proposed classifier system was used for training support vector machines for automated animal audio classification [28].
They hone your psyche and train tolerance, which is of value to all. Machine learning and its methods can be used to analyze emotions [29]. Artificial have become very popular in many fields [30,31]. The classification is the most commonly used, [32] proposes system based in classification to test the Heart Rate Variability. In [33] the authors used naive bayes classifier to evaluate the money

laundering risk. In [34] used classification method to detect voice activity.
Coronavirus disease (COVID-19), which occurred in Wuhan, Hubei Province, China, at the end of December 2019, is a quickly evolving infectious disease caused by a new series of coronaviruses called severe acute respiratory syndrome (SARS) coronavirus 2 [35]. SARS was first discovered in 2003, however, COVID-19 surpassed the number of instances and deaths related to SARS within a few weeks, which led the World Health Organization (WHO) to declare the outbreak of COVID-19 to be a global health epidemic of worldwide concern, or pandemic [36].
Coronavirus disease 2019 (COVID-19) has motivated numerous researchers in different fields to contribute to this global paradigm. They have used different methods to address the issues including data mining algorithms and techniques which have been-shown to be effective for making predictions and identifications.
The term data mining or knowledge discovery from data was introduced in the late 1980s [37,38]. The data mining is a method for discovering knowledge by revealing new knowledge from huge databases [38]. Data mining refers to an essential process which applies intelligent methods to extract patterns from raw data [37]. Algorithms and data mining techniques are well-known methods for creating predictive models and data analysis [39]. In addition, Geographic Information System (GIS) and social media data mining have become vital instruments for analyzing the global spread of infectious diseases [40]. In healthcare, for instance, data mining has been widely used to predict and diagnose diseases such as coronavirus (COVID-19) [18].
Many researchers have used the enormous data of COVID-19 and data mining to develop predictive, diagnostic, and therapeutic strategies against pandemics of diseases including COVID-19 and similar diseases in the future [41]. In addition, they have used data mining methods to classify, identify, and predict the coronavirus series [42]. More-over, data mining is an efficient technique for rapidly identifying and repurposing approved therapeutics for COVID-19 patients [43]. For instance, researchers have used an advanced artificial intelligence technique such as data mining to discover valid and novel patterns from a dataset [44]. There are many methods applied in data mining which include simple logistic, spatial data mining, decision tree, random forests, logistic model trees, naive bayes, support vector machine, logistic regression random forest, multilayer perceptron , classification and regression trees, and k-nearest neighbors [36,44,45]. According to [45], a mix of data mining techniques including simple logistic, multilayer perceptron, naive Bayes classifier, and classification and regression trees have been used by [46] to enhance the diagnosis of neonatal jaundice amongst new-born babies. [47] reported that decision tree and neural network have been used by [48] and [49] to predict the performance and the efficiency of the students' results. Moreover, the naive Bayes classifier and the decision tree algorithm were used by [18] to predict recovery from the Middle East respiratory syndrome coronavirus. The data mining techniques were applied in a study to predict the role and impact of environmental factors and the spread of COVID-19 disease with regard to the latitude and longitude effects [45]. Furthermore, [47] developed a model to predict and analyze a solution that reduced students' depression during COVID-19 by using different data mining algorithms such as random forest, decision tree, support vector machine, logistic regression, k-nearest neighbors, and naive Bayes. [36] Introduced an application which used data mining techniques, namely, spatial data mining with a satellite dataset to predict the spread of COVID-19 using the statistics for India. Likewise, [44] used various data mining techniques in order to predict the recovery of COVID-19 patients; however, this research was conducted using a dataset which had been collected indirectly and only five of eight attributes were used in the study [44]. A shortage of attributes can lead to inaccurate results. The dataset for our research was collected by the researchers directly with realistic data. The collected data has 27 attributes for predicting patients' recovery time from COVID-19, patients with a high risk of not recovering, and possible treatments for patients with the disease. Finally, this study could be a reference for future studies with real datasets and adequate attributes.

3. THE PROPOSED MODEL

The CRISP-DM (Cross-Industry Standard Process for Data Mining) strategy was adopted to construct a trust classification model [50]. Fundamentally, the technique consists of the following five stages: (1) collecting the relevant characteristics of the problem under investigation; (2) preparing the data; (3) constructing the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classification model; (4) evaluating the model using one of the methods of evaluation; (5) and finally, using the coronavirus patients' potential prediction model. In the next subsections, these steps are presented.

3.1. Description of the Dataset

The features and factors are separated into the following five groups: personal information, clinical information, comorbidities, hospitalization, and management. The at-tributes for the five groups are shown in Figure 1. Details of the personal information data are listed in Tables 1, 2, 3, 4 and 5.

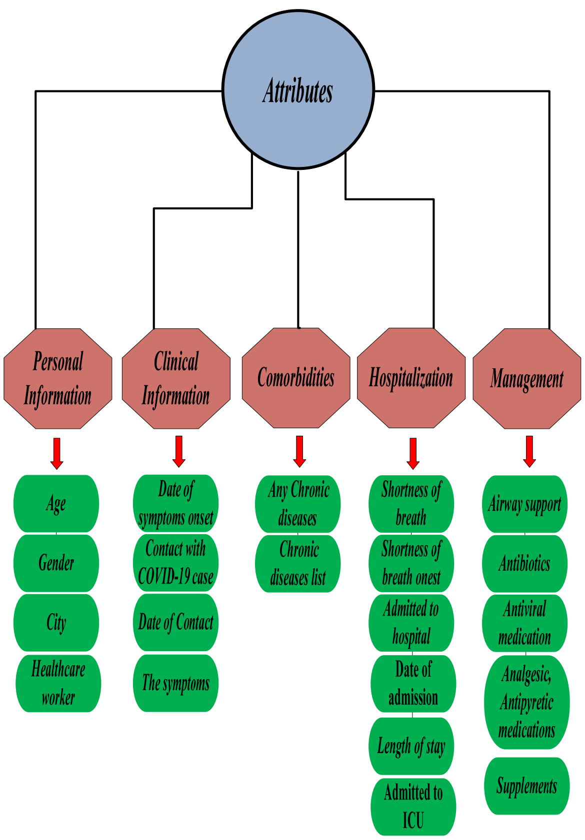


Figure 1: Influence features and attributes.

Table 1: Attribute 1: Age

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
16	2	1.8	1.8	1.8
19	3	2.7	2.7	4.5
20	8	7.2	7.2	11.7
21	3	2.7	2.7	14.4
24	6	5.4	5.4	19.8
25	5	4.5	4.5	24.3
26	10	9.0	9.0	33.3
27	8	7.2	7.2	40.5
28	7	6.3	6.3	46.8
29	6	5.4	5.4	52.3
30	5	4.5	4.5	56.8
31	5	4.5	4.5	61.3
32	1	.9	.9	62.2
33	1	.9	.9	63.1
34	2	1.8	1.8	64.9
35	3	2.7	2.7	67.6
36	5	4.5	4.5	72.1
38	5	4.5	4.5	76.6
39	6	5.4	5.4	82.0
40	1	.9	.9	82.9
45	2	1.8	1.8	84.7
46	2	1.8	1.8	86.5
48	2	1.8	1.8	88.3
49	1	.9	.9	89.2
51	1	.9	.9	90.1
53	1	.9	.9	91.0
54	2	1.8	1.8	92.8
56	1	.9	.9	93.7
61	1	.9	.9	94.6
62	1	.9	.9	95.5
65	2	1.8	1.8	97.3
68	1	.9	.9	98.2
70	1	.9	.9	99.1
75	1	.9	.9	100.0
Total	111	100.0	100.0	

Table 2: Attribute 2: Gender

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
M	55	49.5	49.5	49.5
F	56	50.5	50.5	100.0
Total	111	100.0	100.0	

Table 3: Attribute 3: City

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Unaizah	30	27.0	27.0	27.0
Buraydah	10	9.0	9.0	36.0
Riyadh	54	48.6	48.6	84.7
Other	17	15.3	15.3	100.0
Total	111	100.0	100.0	

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4: Healthcare worker

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	40	36.0	36.0	36.0
No	71	64.0	64.0	100.0
Total	111	100.0	100.0	

Table 5: Contact with COVID-19 case

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	81	73.0	73.0	73.0
No	30	27.0	27.0	100.0
Total	111	100.0	100.0	

3.2. Description of the Dataset

The relevant characteristics are gathered in this phase. Initially, 41 attributes were gathered and some attributes, deemed irrelevant to the report, were eliminated such as "date of symptoms onset, date of contact with COVID 19 case, and after how many days from the date of diagnosis the shortness of breath began". Finally, only 29 conditional attributes and one class attribute were taken into account. Table 1 presents a description of the attributes and possible representation values of the attributes. The class attribute is the patient recovery. The relevant attributes and data view are shown in Figures 2 and 3, respectively.

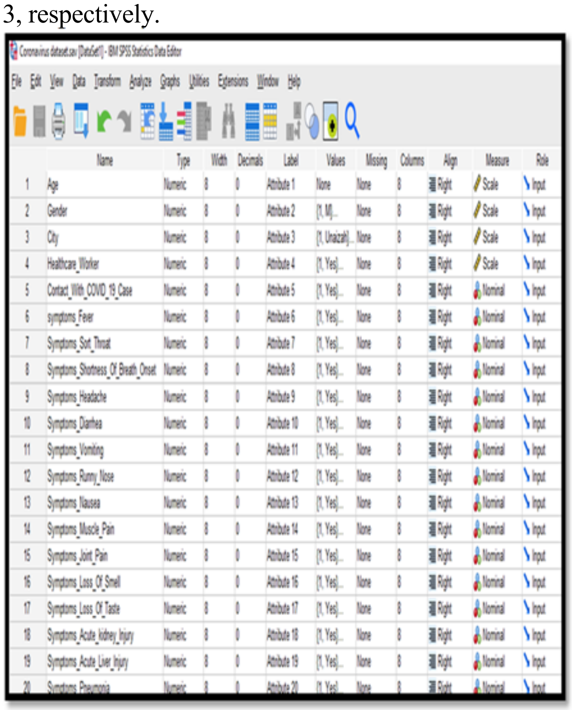


Figure 2: Relevant attributes

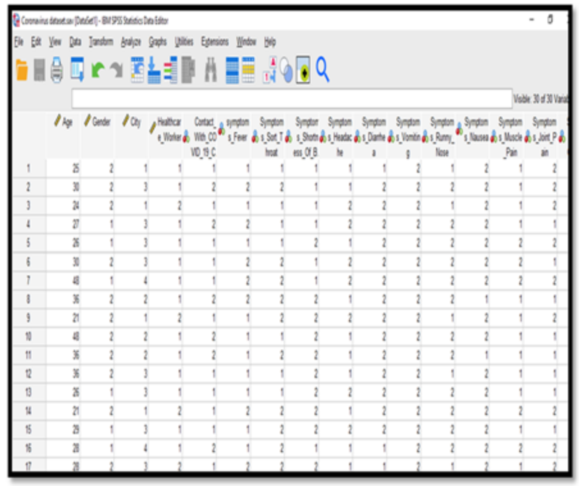


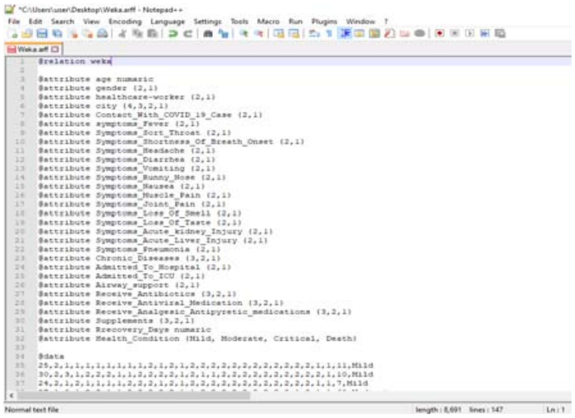
Figure 3: Data view

3.3. Preparing the Data and Selecting the Relevant Attributes

For this step, the collected data was organized in tables in a format appropriate for the data mining algorithms used. In fact, irrelevant attributes could degrade the proposed classification model, therefore, in this study, feature selection was used to select the best set of features. In addition, the same standard value was used for all data and the data were cleaned. A significant proportion of the time and energy involved in the data mining process is the preparation of the input for the data mining investigation. The Weka system takes input in the form of the attribute-relation file format. Figure 4 shows the attribute-relation file format and Table 1 shows the symbolic attribute description. Finally, the list of the most relevant attributes is comprised of the following: Age, Gender, City, Healthcare_Worker, Contact_With_COVID_19_Case, Symptoms_Fever, Symptoms_Sore_Throat, Symptoms_Shortness_Of_Breath_Onset, Symptoms_Headache, Symptoms_Diarrhea, Symptoms_Vomiting, Symptoms_Runny_Nose, Symptoms_Nausea, Symptoms_Muscle_Pain, Symptoms_Joint_Pain, Symptoms_Loss_Of_Smell, Symptoms_Loss_Of_Taste, Symptoms_Acute_kidney_Injury, Symptoms_Acute_Liver_Injury, Symptoms_Pneumonia, Chronic_Diseases, Admitted_To_Hospital, Admitted_To_ICU, Airway_Support, Receive_Antibiotics, Receive_Antiviral_Medication, Receive_Analgesic_Antipyretic_medications, Supplements, Recovery_Days, Health_Condition.

Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



```
@relation weka

@attribute age numeric
@attribute gender (2,1)
@attribute healthcare-worker (2,1)
@attribute city (4,3,2,1)
@attribute Contact_With_COVID_19_Case (2,1)
@attribute symptoms_Fever (2,1)
@attribute Symptoms_Sort_Throat (2,1)
@attribute Symptoms_Shortness_Of_Breath_Onset (2,1)
@attribute Symptoms_Headache (2,1)
@attribute Symptoms_Diarrhea (2,1)
@attribute Symptoms_Vomiting (2,1)
@attribute Symptoms_Runny_Nose (2,1)
@attribute Symptoms_Nausea (2,1)
@attribute Symptoms_Muscle_Pain (2,1)
@attribute Symptoms_Joint_Pain (2,1)
@attribute Symptoms_Loss_Of_Smell (2,1)
@attribute Symptoms_Loss_Of_Taste (2,1)
@attribute Symptoms_Acute_kidney_Injury (2,1)
@attribute Symptoms_Acute_Liver_Injury (2,1)
@attribute Symptoms_Pneumonia (2,1)
@attribute Chronic_Diseases (3,2,1)
@attribute Admitted_To_Hospital (2,1)
@attribute Admitted_To_ICU (2,1)
@attribute Airway_support (2,1)
@attribute Receive_Antibiotics (3,2,1)
@attribute Receive_Antiviral_Medication (3,2,1)
@attribute Receive_Analgesic_Antipyretic_medications (3,2,1)
@attribute Supplements (3,2,1)
@attribute Rrecovery_Days numeric
@attribute Health_Condition (Mild, Moderate, Critical, Death)

@data
25,2,1,1,1,1,1,1,1,1,2,1,2,1,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1,11,Mild
30,2,3,1,2,2,2,1,1,2,2,2,2,1,2,1,1,2,2,2,2,2,2,2,2,2,2,1,10,Mild
24,2,1,2,1,1,1,1,2,2,2,1,2,1,2,2,2,2,2,2,2,2,2,2,2,2,1,1,7,Mild
```

Figure 4: The Attribute-Relation File Format

Table 6: The symploic attribute description

Attribute No	Description	Possible Values
1	Age	Integer
2	Gender	M =1, F =2
3	City	Unaizah =1, Buraydah =2, Riyadh =3, Other =4
4	Healthcare_Worker	Yes =1, No =2
5	Contact_With_COVID_19_Case	Yes =1, No =2
6	symptoms_Fever	Yes =1, No =2
7	Symptoms_Sore_Throat	Yes =1, No =2
8	Symptoms_Shortness_Of_Breath_Onset	Yes =1, No =2
9	Symptoms_Headache	Yes =1, No =2
10	Symptoms_Diarrhe	Yes =1, No =2
11	Symptoms_Vomiting	Yes =1, No =2
12	Symptoms_Runny_Nose	Yes =1, No =2
13	Symptoms_Nausea	Yes =1, No =2
14	Symptoms_Muscle_Pain	Yes =1, No =2
15	Symptoms_Joint_Pain	Yes =1, No =2
16	Symptoms_Loss_Of_Smell	Yes =1, No =2
17	Symptoms_Loss_Of_Taste	Yes =1, No =2
18	Symptoms_Acute_kidney_Injury	Yes =1, No =2
19	Symptoms_Acute_Liver_Injury	Yes =1, No =2
20	Symptoms_Pneumonia	Yes =1, No =2
21	Chronic_Diseases	Yes =1, No =2
22	Admitted_To_Hospital	Yes =1, No =2
23	Admitted_To_ICU	Yes =1, No =2
24	Airway_support	Yes=1,No =2,Don't Know =3
25	Receive_Antibiotics	Yes=1,No =2,Don't Know =3
26	Receive_Antiviral_Medication	Yes=1,No =2,Don't Know =3
27	Receive_Analgesic_Antipyretic_medications	Yes=1,No =2,Don't Know =3
28	Supplements	Yes=1,No =2,Don't Know =3
29	Recovery_Days	Yes=1,No =2,Don't Know =3
30	Health_Condition*	Class attribute

*Health Condition: 1, mild; 2, moderate; 3, critical; 4, death.

For some cases, attribute datatypes must be changed to numeric attributes. Some AI algorithms, which are proficient in handling small datasets, such as linear discriminant analysis (LDA) [51] and multiple perceptron artificial neural network (MLP-NN) [52,53] require numerical attributes for calculations. Furthermore, the support vector machine algorithm, which was also utilized, was intended to effectively work with numerical at-tributes. In addition, as a best practice in the management of the MLP-NN algorithm, at-tributes must also be in numerical form and standardized to obtain the best classification results.

3.4. Building the Classification Model

The next stage is to use the decision tree technique to develop the classification model. The decision tree is an outstanding and useful approach, as it is moderately quick, and thus can be easily transformed into simple classification rules. The strategy of the decision tree relies primarily upon the use of a data benefit metric that defines the most useful attribute in general. The gain of data relies on the entropy measure. Building the decision tree is based on the gain ratio which is ranked and locates the attribute according to its gain ratio.

Recovery Days was the attribute with the highest gain ratio. In the decision tree, therefore, the Recovery Days attribute is classified as the root node. This method is, then, followed for the remaining attributes, and the set of classification rules are produced by following all the paths of the tree where interesting classification rules have been generated by the decision tree. In Table 2, some of the rules created are given in a form that is meaningful.

In Table 2, the first column indicates the rule number, the second column lists the produced rules, the third column gives the number of cases successfully satisfying the rules, and the last column gives the number of attributes contained in the rule. Depending on the number of attributes contained in the rule, the table illustrates the rules in descending order. The longest rule among the generated rules consist of 16 attributes, whereas only two attributes are included in the shorter rule. A system that encourages the use of the produced rules is designed to achieve the goals set by this study, enabling healthcare workers to predict the status of patients with coronavirus. Figures 5 and 6 show the preprocess for some attributes. Finally, Figures 7 and 8 show some of the classification rules and the size of the tree.

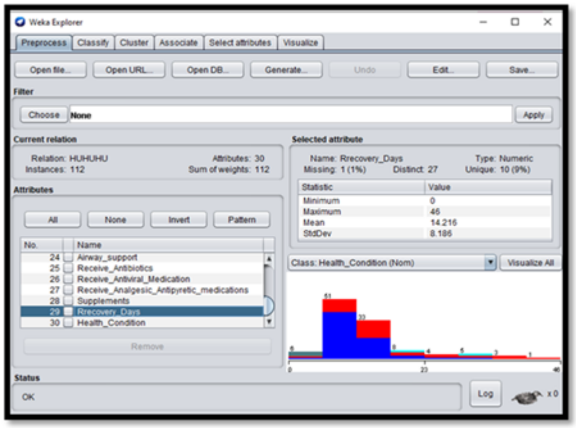


Figure 5: Preprocess Weka explorer for recovery days attribute

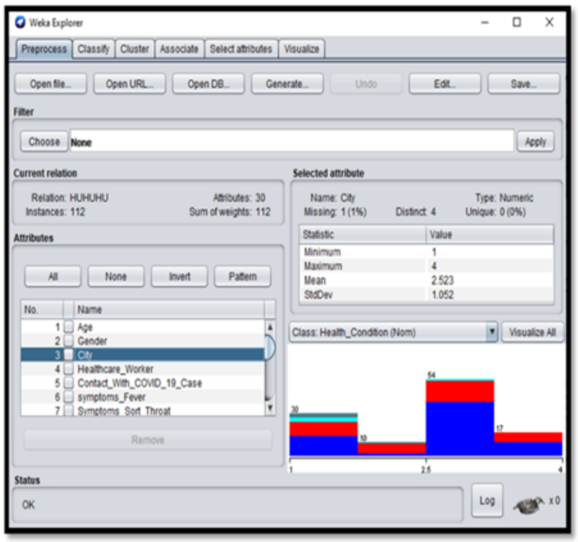


Figure 6: Preprocess Weka explorer for the city attribute

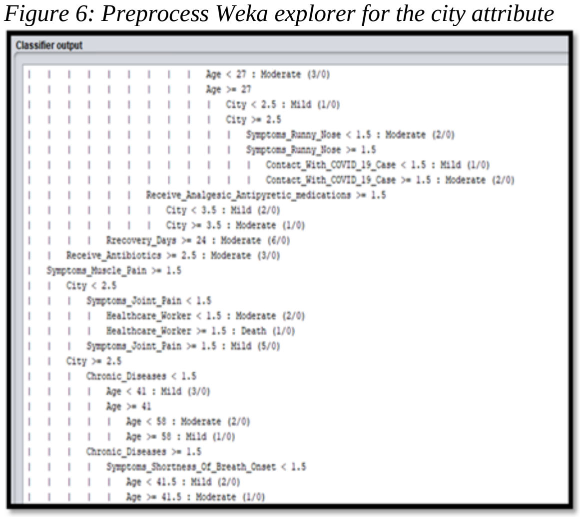


Figure 7: Sample of classification rules

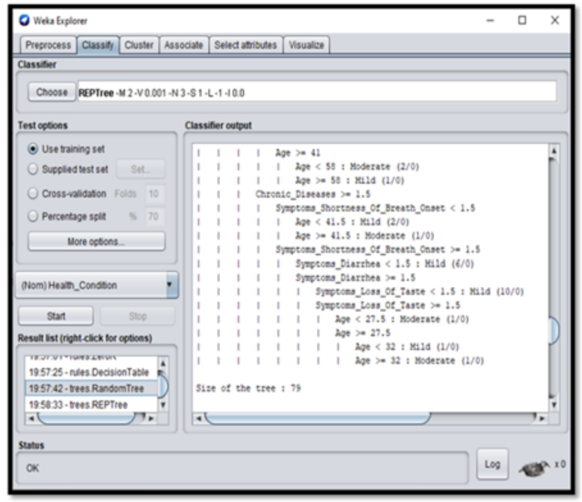


Figure 8: Size of the tree

Some of the interesting rules discovered are:

- IF Recovery_Days > 0 AND Admitted_To_Hospital > 1 AND Symptoms_Joint_Pain > 1 AND Symptoms_Shortness_Of_Breath_Onset > 1: Mild
- IF Recovery_Days > 0 AND Admitted_To_Hospital > 1 AND Symptoms_Pneumonia > 1 AND Receive_Antibiotics <= 2 AND Receive_Antiviral_Medication > 1 AND Symptoms_Nausea <= 1: Moderate
- IF Recovery_Days > 0 AND Admitted_To_Hospital > 1 AND Symptoms_Diarrhea > 1 AND Symptoms_Vomiting > 1 AND Recovery_Days <= 26 AND Airway_Support > 1 AND Gender > 1 AND Symptoms_Shortness_Of_Breath_Onset <= 1: Mild
- IF Recovery_Days > 0 AND Admitted_To_Hospital > 1 AND Contact_With_COVID_19_Case <= 1 AND Symptoms_Shortness_Of_Breath_Onset <= 1: Moderate
- IF Recovery_Days > 0 AND Symptoms_Pneumonia > 1 AND Contact_With_COVID_19_Case <= 1 AND Receive_Antibiotics > 1 AND Recovery_Days <= 14 AND Symptoms_Sore_Throat > 1: Mild
- IF Recovery_Days > 0 AND Symptoms_Fever <= 1: Critical

Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 7: Sample of the generated rules

Rule No	Rules	Instance	No of Attributes
1	IF the Gender is Female AND Age <= 40 AND is Healthcare_Worker is Yes AND Symptoms_Fever is Yes AND Symptoms_Joint_Pain is Yes AND Symptoms_Sore_Throat is No AND Symptoms_Diarrhea is No AND Symptoms_Vomiting is Yes AND Symptoms_Runny_Nose is Yes or No AND Symptoms_Nausea is No AND Receive_Analgesic_Antipyretic_Medications is Don't know AND Admitted_To_Hospital is No AND Airway_Support is No AND Supplements is Yes AND Recovery_Days <= 11 AND Symptoms_Acute_Kidney_Injury is Yes THEN the *Health_Condition* is Moderate	29	16
2	IF the City is Unaizah AND Healthcare_Worker is No AND Age <= 32 AND Symptoms_Loss_Of_Taste is Yes AND Contact_With_COVID_19_Case is No AND Symptoms_Fever is No AND Admitted_To_Hospital is Yes AND Admitted_To_ICU is No AND Receive_Analgesic_Antipyretic_Medications is Yes AND Receive_Antiviral_Medication is No AND Supplements is Yes AND Recovery_Days <= 20 AND Symptoms_Acute_Kidney_Injury is No THEN the *Health_Condition* is Mild	18	13
3	IF Recovery_Days >= 15 AND Admitted_To_ICU is Yes AND Symptoms_Diarrhea is Yes AND Symptoms_Vomiting is No AND City is Buraydah or Riyadh AND Airway_Support is Yes AND Gender is Female AND Symptoms_Shortness_Of_Breath_Onset AND Healthcare_Worker is Yes THEN Moderate	12	9
4	IF Recovery_Days > 0 AND Symptoms_Pneumonia is Yes AND Contact_With_COVID_19_Case is No AND Receive_Antibiotics is No AND Symptoms_Sore_Throat is Yes Then THEN the *Health_Condition* is Mild	4	5
5	IF Symptoms_Loss_Of_Smell is Yes AND Symptoms_Loss_Of_Taste is Yes AND Admitted_To_Hospital is No AND Admitted_To_ICU is No THEN the *Health_Condition* is Mild	36	4
6	IF Recovery_Days > 30 AND Symptoms_Fever is Yes THEN the *Health_Condition* is Critical	3	2

4. EXPERIMENTS AND EVALUATION

Predicting the recovery of coronavirus disease patients is essential for helping healthcare workers and ensuring their retention, improving performances of hospitals, and managing recovery resources. Obviously, predicting coronavirus disease patients is an essential need to help healthcare worker develop plans for overcoming the difficulties their patients may face during their recoveries. In this study, we use the decision tree and classification algorithms, and define key indicators in a small dataset that is used to construct a prediction model. For more accuracy of the proposed model, we use several machine learning algorithms to evaluate the key indicators. Among the algorithms picked, the results demonstrated that the classification algorithm in small datasets is able to identify key indicators. Importantly, we also demonstrated the efficacy of using data mining algorithms and machine learning to analyze and train a small dataset and to produce an acceptable classification with accurate and reliable test rates.

According to the results obtained, we found that the classification accuracy for the three different classification algorithms is high, which could indicate that the collected samples and attributes are adequate to produce a high-quality model of classification. In order to evaluate the performance of a classification model on a test set, the classification accuracy or error rate are commonly used. From the test set, the accuracy of the classification model is processed where it can be used to evaluate the overall performance of different classifiers in the same domain. However, the class labels of the test records should be known, and the assessment methodology is expected to assess the order and process the classification accuracy. In order to achieve the consistency of the classification model Weka software was used.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Three main classification methods, BayesNet-D, Naive Bayes, and J48 were tested. The evaluation results of the proposed model, as shown in Figure 3, describe the percentage of the correctness of classified instances. Figures 9, 10, and 11 show the classification methods, classifier output, and test option.

Table 8: Classification accuracy of the 3 different algorithms

Algorithm	Percentage of correctly classified instances
BayesNet-D	74.7748%
Naive Bayes	81.081%
J48	93.6937%

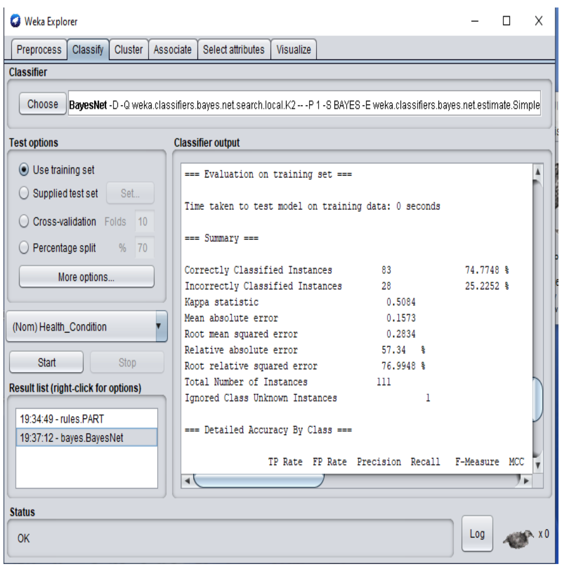


Figure 9: BayesNet-D validation

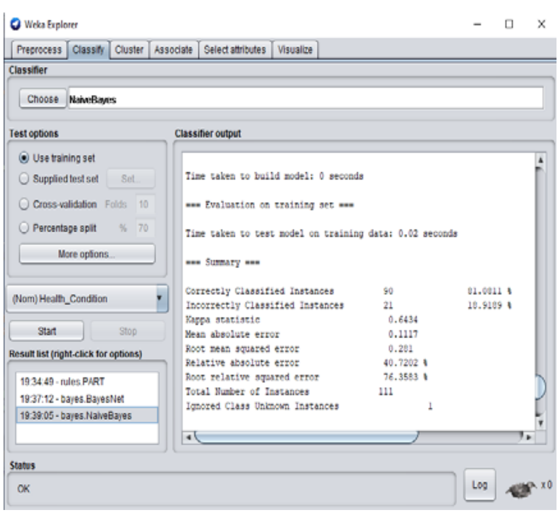


Figure 10: Naïve Bayes validation

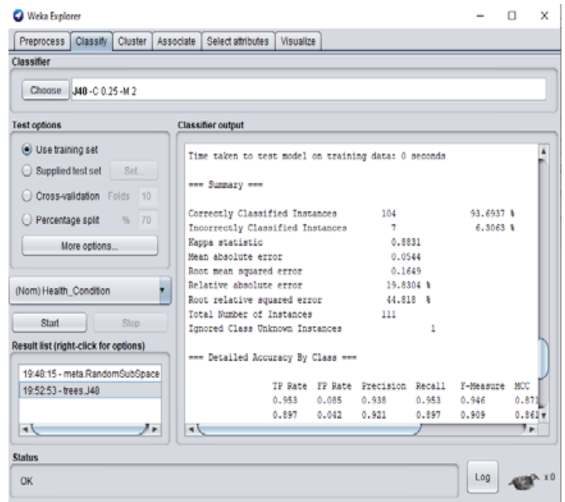


Figure 11: J48 validation

5. LIMITATION AND CONSTRAINTS

When fitting machine learning models to smaller datasets, there are inherent limitations. The models have fewer examples to learn from as the training datasets get smaller, increasing the risk of overfitting. As well as, difficulty obtaining information related to Coronavirus patients.

6. CONCLUSIONS

As researchers search various databases to confront the coronavirus threat, timely access to accurate data has become essential. As the risk severe illness increases, access to reliable public data is necessary for a better understanding of the current crisis. The first objective of this research was to provide a real dataset of clinical data on coronavirus patients, not just statistical daily reports. Then, a data mining model was developed using the epidemiological dataset of COVID-19 patients from the Kingdom of Saudi Arabia to predict the recovery of COVID-19 infected patients. Data mining is an extremely powerful technique that can be used for identifying new useful information from a dataset. Most coronavirus datasets summarize statistical data. These statistics are not about clinical data for patients but rather daily statistical reporting, therefore, the data are not relevant to all researchers.

In this study, we use data mining to investigate and evaluate coronavirus patients. The classification model can be using by healthcare systems for improving patient out-comes. The data mining algorithms extract information for a deeper understanding of patients' status, and therefore can assist healthcare providers with making decisions regarding essential actions needed. In addition, healthcare mangers and management sys-tem can update and improve their decisions and policies using the rules generated from the proposed model and the patterns revealed from it, as well as review and enhance their strategies, and advance the quality of their management system.

The extracted knowledge can also be using by healthcare management systems to improve their organizations, upgrade their methodologies, and improve the nature of the management board framework. Certain classification methods can be used to verify the most effective and accurate classification approach to use with patient data. In this study, data from coronavirus patients were assessed for the attributes most affected by this pandemic and a set of features and attributes was identified which can be used for improving the quality of patient care. Advancements in technology have a rapid impact on every field of life, whether it be medical or some other field. By analyzing the data, artificial intelligence has demonstrated promising outcomes in healthcare decision-making.

ACKNOWLEDGMENTS: Researchers would like to thank the Deanship of Scientific Research, Qassim University for funding publication of this project.

REFRENCES:

[1] C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern, Lancet. 395 (2020) 470–473.

[2] M. Jamshidi, A. Lalbakhsh, J. Talla, Z. Peroutka, F. Hadjilooei, P. Lalbakhsh, M. Jamshidi, L. La Spada, M. Mirmozafari, M. Dehghani, Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment, IEEE Access. 8 (2020) 109581–109595.

[3] L. Hawryluck, W.L. Gold, S. Robinson, S. Pogorski, S. Galea, R. Styra, SARS control and psychological effects of quarantine, Toronto, Canada, Emerg. Infect. Dis. 10 (2004) 1206.

[4] H. Nishiura, S. Jung, N.M. Linton, R. Kinoshita, Y. Yang, K. Hayashi, T. Kobayashi, B. Yuan, A.R. Akhmetzhanov, The extent of transmission of novel coronavirus in Wuhan, China, 2020, (2020).

[5] L.J. Muhammad, S.S. Usman, Power of artificial intelligence to diagnose and prevent further COVID-19 Outbreak: a short communication, ArXiv Prepr. ArXiv2004.12463. (2020).

[6] A. Banerjee, K. Kulcsar, V. Misra, M. Frieman, K. Mossman, Bats and coronaviruses, Viruses. 11 (2019) 41.

[7] D. Yang, J.L. Leibowitz, The structure and functions of coronavirus genomic 3′ and 5′ ends, Virus Res. 206 (2015) 120–133.

[8] A.J. Jinia, N.B. Sunbul, C.A. Meert, C.A. Miller, S.D. Clarke, K.J. Kearfott, M.M. Matuszak, S.A. Pozzi, Review of sterilization techniques for medical and personal protective equipment contaminated with SARS-CoV-2, IEEE Access. 8 (2020) 111347–111354.

[9] P. Ozili, COVID-19 in Africa: socio-economic impact, policy response and opportunities, Int. J. Sociol. Soc. Policy. (2020).

[10] E.J. Emanuel, G. Persad, R. Upshur, B. Thome, M. Parker, A. Glickman, C. Zhang, C. Boyle, M. Smith, J.P. Phillips, Fair allocation of scarce medical resources in the time of Covid-19, (2020).

[11] W. Dattilo, A.C. e Silva, R. Guevara, I. MacGregor-Fors, S.P. Ribeiro, COVID-19 most vulnerable Mexican cities lack the public health infrastructure to face the pandemic: a new temporally-explicit model, MedRxiv. (2020).

[12] J. Hellewell, S. Abbott, A. Gimma, N.I. Bosse, C.I. Jarvis, T.W. Russell, J.D. Munday, A.J. Kucharski, W.J. Edmunds, F. Sun, Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, Lancet Glob. Heal. 8 (2020) e488–e496.

[13] S. Hussain, L.J. Muhammad, F.S. Ishaq, A. Yakubu, I.A. Mohammed, Performance evaluation of various data mining algorithms on road traffic accident dataset, in: Inf. Commun. Technol. Intell. Syst., Springer, 2019: pp. 67–78.

[14] L.J. Muhammad, A.A. Haruna, I.A. Mohammed, M. Abubakar, B.G. Badamasi, J.M. Amshi, Performance evaluation of classification data mining algorithms on coronary artery disease dataset, in: 2019 9th Int. Conf. Comput. Knowl. Eng., IEEE, 2019: pp. 1–5.

[15] H. Abu-Dalbouh, N.M. Norwawi,

Bidirectional agglomerative hierarchical clustering using AVL tree algorithm, Int. J. Comput. Sci. Issues. 8 (2011) 95.

[16] H.A. Dalbouh, N.M. Norwawi, Improvement on agglomerative hierarchical clustering algorithm based on tree data structure with bidirectional approach, in: 2012 Third Int. Conf. Intell. Syst. Model. Simul., IEEE, 2012: pp. 25–30.

[17] L.J. Muhammad, S. Sani, A. Yakubu, M.M. Yusuf, T.A. Elrufai, I.A. Mohammed, A.M. Nuhu, Using decision tree data mining algorithm to predict causes of road traffic accidents, its prone locations and time along Kano–Wudil highway, Int J Database Theory Appl. 10 (2017) 197–208.

[18] I. Al-Turaiki, M. Alshahrani, T. Almutairi, Building predictive models for MERS-CoV infections using data mining techniques, J. Infect. Public Health. 9 (2016) 744–748.

[19] A. Rahaman, M.M. Islam, M.R. Islam, M.S. Sadi, S. Nooruddin, Developing IoT Based Smart Health Monitoring Systems: A Review., Rev. d'Intelligence Artif. 33 (2019) 435–440.

[20] M.M. Islam, A. Rahaman, M.R. Islam, Development of smart healthcare monitoring system in IoT environment, SN Comput. Sci. 1 (2020) 1–11.

[21] A. Ostreika, M. Pivoras, A. Misevičius, T. Skersys, L. Paulauskas, Classification of Objects by Shape Applied to Amber Gemstone Classification, Appl. Sci. . 11 (2021). https://doi.org/10.3390/app11031024.

[22] N.M. Jebreel, J. Domingo-Ferrer, D. Sánchez, A. Blanco-Justicia, KeyNet: An Asymmetric Key-Style Framework for Watermarking Deep Learning Models, Appl. Sci. . 11 (2021). https://doi.org/10.3390/app11030999.

[23] R.P. Bonidia, J.S. Machida, T.C. Negri, W.A.L. Alves, A.Y. Kashiwabara, D.S. Domingues, A. De Carvalho, A.R. Paschoal, D.S. Sanches, A Novel Decomposing Model With Evolutionary Algorithms for Feature Selection in Long Non-Coding RNAs, IEEE Access. 8 (2020) 181683–181697.

[24] A. Bieńkowska, K. Tworek, The Moderating Role of IT in Process of Shaping Organizational Performance by Dynamic Capabilities of Controlling, Appl. Sci. . 11 (2021). https://doi.org/10.3390/app11020889.

[25] S. Zeadally, E. Adi, Z. Baig, I.A. Khan, Harnessing artificial intelligence capabilities to improve cybersecurity, Ieee Access. 8 (2020) 23817–23837.

[26] G. Pisoni, N. Díaz-Rodríguez, H. Gijlers, L. Tonolli, Human-Centered Artificial Intelligence for Designing Accessible Cultural Heritage, Appl. Sci. . 11 (2021). https://doi.org/10.3390/app11020870.

[27] P. Park, M. Jung, P. Di Marco, Remaining Useful Life Estimation of Bearings Using Data-Driven Ridge Regression, Appl. Sci. . 10 (2020). https://doi.org/10.3390/app10248977.

[28] L. Nanni, S. Brahnam, A. Lumini, G. Maguolo, Animal Sound Classification Using Dissimilarity Spaces, Appl. Sci. . 10 (2020). https://doi.org/10.3390/app10238578.

[29] T. Bikku, K.S. Sree, Deep Learning Approaches For Classifying Data: A review, J. Eng. Sci. Technol. 15 (2020) 2580–2594.

[30] B.H.K. Al-obaidi, S.K. Ali, D.T. Jassim, Influence Of A River Water Quality On The Efficiency Of Water Treatment Using Artificial Neural Network, J. Eng. Sci. Technol. 15 (2020) 2610–2623.

[31] N.Z. Dina, W.I. Sabilla, Kartono, The Impact Of Using Visual Learning Environment On Student Programming Course Learning Achievement: A Case Study Of Universitas Airlangga, J. Eng. Sci. Tech. 14 (2019) 712–725.

[32] H. Kirti, S.J. Sohal, Multistage Classification Of Arrhythmia And Atrial Fibrillation On Long-Term Heart Rate Variability, J. Eng. Sci. Technol. 15 (2020) 1277–1295.

[33] M.D.A. Islam, M.K. Nasir, Evaluation Of Money Laundering Risk Of Bank Accounts Using Naive Bayes Classification, J. Eng. Sci. Technol. 15 (2020) 3481–3493.

[34] S.N. Mohammed, A.K. Hassan, A. Yulianur, T. Saidi, B. Setiawan, S. Sugianto, M. Rusdi, H.R. Farhan, M.S. Kod, H.I. Shahadi, Automatic voice activity detection using fuzzy-neuro classifier, J. J. Eng. Sci. Technol. 15 (2020) 2854–2870.

[35] J. Li, Q. Xu, R. Cuomo, V. Purushothaman, T. Mackey, Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study, JMIR Public Heal.

Surveill. 6 (2020) e18700.

[36] M. Rajathi, R. Arumugam, An Application of Spatial Data mining in the study of Corona Virus (COVID-19) Pandemic through Statistical Approach, 8 (2020).

[37] J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, Elsevier, 2011.

[38] K.J. Cios, W. Pedrycz, R.W. Swiniarski, Data mining methods for knowledge discovery, Springer Science & Business Media, 2012.

[39] S.M. Ayyoubzadeh, S.M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, S.R.N. Kalhori, Predicting COVID-19 incidence through analysis of google trends data in iran: data mining and deep learning pilot study, JMIR Public Heal. Surveill. 6 (2020) e18828.

[40] D. Li, H. Chaudhary, Z. Zhang, Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining, Int. J. Environ. Res. Public Health. 17 (2020) 4988.

[41] A. Alimadadi, S. Aryal, I. Manandhar, P.B. Munroe, B. Joe, X. Cheng, Artificial intelligence and machine learning to fight COVID-19, (2020).

[42] K. Ghosh, S.A. Amin, S. Gayen, T. Jha, Chemical-informatics approach to COVID-19 drug discovery: Exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors, J. Mol. Struct. 1224 (2020) 129026.

[43] D. Gurwitz, Repurposing current therapeutics for treating COVID-19: A vital role of prescription records data mining, Drug Dev. Res. (2020).

[44] L.J. Muhammad, M.M. Islam, U.S. Sharif, S.I. Ayon, Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients Recovery, (2020).

[45] A. Keshavarzi, Coronavirus Infectious Disease (COVID-19) Modeling: Evidence of Geographical Signals, Available SSRN 3568425. (2020).

[46] D. Ferreira, A. Oliveira, A. Freitas, Applying data mining techniques to improve diagnosis in neonatal jaundice, BMC Med. Inform. Decis. Mak. 12 (2012) 143.

[47] Krishna, K. V Praveen, Prediction and Analysis of Data Mining Models for Students Underlying Issues during Novel Coronavirus (COVID-19), Int. J. Eng. Res. Technol. NCAIT. 8 (2020).

[48] M. Sharma, M. Mavani, Accuracy comparison of predictive algorithms of data mining: Application in education sector, in: Int. Conf. Adv. Comput. Commun. Control, Springer, 2011: pp. 189–194.

[49] M. Wook, Y.H. Yahaya, N. Wahab, M.R.M. Isa, N.F. Awang, H.Y. Seong, Predicting NDUM student's academic performance using data mining techniques, in: 2009 Second Int. Conf. Comput. Electr. Eng., IEEE, 2009: pp. 357–361.

[50] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0: Step-by-step data mining guide, SPSS Inc. 9 (2000) 13.

[51] A. Sharma, K.K. Paliwal, Linear discriminant analysis for the small sample size problem: an overview, Int. J. Mach. Learn. Cybern. 6 (2015) 443–454.

[52] S. Ingrassia, I. Morlini, Neural network modeling for small datasets, Technometrics. 47 (2005) 297–311.

[53] A. Pasini, Artificial neural networks for small dataset analysis, J. Thorac. Dis. 7 (2015) 953.

Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

APPENDIX A

One of the aims of the research is to have a real collection of data on coronavirus patients, not only a statistical regular report, but also clinical data on patients. This study uses the first 30 attributes, (Attribute No, Description and Possible Values), as shown in table 6. The sample size is 111, the 41 attributes are found in the full data. In the following tables, the complete data collection of coronavirus patients is shown.

Table 9: The Overall Attributes

att1:Age	att21:Chronic_Diseases
att2:Gender	at22:Admitted_To_Hospital
att3:City	att23:Admitted_To_ICU
att4:Healthcare_Worker	att24:Airway_support
att5:Contact_With_COVID_19_Case	att25:Receive_Antibiotics
att6:symptoms_Fever	att26:Receive_Antiviral_Medication
att7:Symptoms_Sort_Throat	att27: Receive_Analgesic_Antipyretic_medications
att8: Symptoms_Shortness_Of_Breath_Onset	att28:Supplements
att9:Symptoms_Headache	att29:Rrecovery_Days
att10:Symptoms_Diarrhea	att30:Health_Condition
att11:Symptoms_Vomiting	att31:Date of contact with COVID-19 Case
att12:Symptoms_Runny_Nose	att32:Date of symptoms onset
att13:Symptoms_Nausea	att33:Date of confirming your covid-19 infection
att14:Symptoms_Muscle_Pain	att34:Date of your recovery/Death
att15:Symptoms_Joint_Pain	att35:Other Symptoms does not mentioned
att16:Symptoms_Loss_Of_Smell	att36:chronic diseases that you suffer from before infected with covid-19
att17:Symptoms_Loss_Of_Taste	att37:Other chronic diseases does not mentioned
att18:Symptoms_Acute_kidney_Injury	att38:Was there shortness of breath that you are hospitalized for it
att19:Symptoms_Acute_Liver_Injury	att39:After how many days from the date of diagnosis the shortness of breath began
att20:Symptoms_Pneumonia	att40:Date of admission to hospital
	att41:Days you stay at the hospital

Table 10: The Full Dataset From Attribute 1 To Attribute 23

No	att1	att2	att3	att4	att5	att6	att7	att8	att9	att10	att11	att12	att13	att14	att15	att16	att17	att18	att19	att20	att21	at22	att23
1	25	2	1	1	1	1	1	1	1	1	2	1	2	1	2	2	2	2	2	2	2	2	2
2	30	2	3	1	2	2	2	1	1	2	2	2	2	1	2	1	1	2	2	2	2	2	2
3	24	2	1	2	1	1	1	1	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2
4	27	1	3	1	2	2	1	1	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2
5	26	1	3	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2
6	30	2	3	1	1	2	2	1	2	2	2	2	2	2	1	1	1	2	2	2	2	2	2
7	48	1	4	1	1	2	2	1	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2
8	36	2	2	1	2	2	2	2	1	2	2	2	1	1	1	1	1	2	2	2	2	2	2
9	21	2	1	2	1	1	2	2	2	2	2	1	2	1	2	1	1	2	2	2	2	2	2
10	48	2	2	1	2	1	1	2	1	2	2	2	2	1	1	2	2	2	2	2	2	2	2
11	36	2	2	1	2	1	2	2	1	2	2	2	1	1	1	1	1	2	2	2	2	2	2
12	36	2	3	1	1	1	1	2	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
13	26	1	3	1	1	1	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2
14	21	2	1	2	1	2	2	2	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2
15	29	1	3	1	1	1	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	2	2
16	28	1	4	1	2	1	2	1	1	1	2	2	2	2	2	2	2	2	2	1	2	1	2
17	28	2	3	2	1	2	2	2	1	1	2	1	2	1	2	1	1	2	2	2	2	2	2
18	29	1	3	1	1	1	2	1	1	2	1	2	2	1	1	2	1	2	2	1	2	2	2
19	26	1	3	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2
20	28	2	3	2	1	2	2	2	1	1	2	1	2	1	2	1	1	2	2	2	2	2	2
21	26	1	3	1	1	1	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2
22	31	1	3	1	2	1	2	2	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2
23	26	1	3	1	1	1	2	2	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
24	68	1	1	1	1	1	1	2	1	1	2	2	2	1	2	2	2	2	2	2	2	2	2
25	31	1	3	1	2	1	2	2	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2
26	56	1	1	2	2	1	2	1	2	2	2	2	1	2	2	2	2	2	2	1	1	1	2

27	40	2	4	2	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
28	46	1	3	2	1	1	2	2	1	2	2	2	2	2	1	2	1	2	2	2	2	2	2
29	62	1	3	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2
30	25	2	1	1	1	1	1	1	2	2	2	2	2	1	2	1	1	2	2	2	2	2	2
31	39	1	3	2	1	2	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	2	2
32	54	1	3	2	1	1	2	1	2	2	2	2	2	2	2	1	1	2	2	2	1	2	2
33	26	1	4	1	1	1	2	2	1	1	2	1	1	1	1	1	1	2	2	2	2	2	2
34	27	1	3	1	1	1	2	2	1	1	2	1	2	1	2	2	1	2	2	2	2	2	2
35	38	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2
36	38	1	4	2	2	2	2	1	1	2	2	2	2	1	2	1	1	2	2	2	2	2	2
37	24	1	4	2	1	1	2	2	1	1	2	1	2	2	1	1	1	2	2	2	2	2	2
38	27	1	3	1	1	1	2	2	1	1	2	1	2	1	2	2	1	2	2	2	2	2	2
39	46	1	3	2	1	1	2	2	1	2	2	2	2	2	1	2	1	2	2	2	2	2	2
40	54	1	3	2	1	1	2	1	2	2	2	2	2	2	2	1	1	2	2	2	1	2	2
41	35	2	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
42	27	2	3	2	2	1	2	1	1	2	2	2	2	1	1	1	1	2	2	2	1	2	2
43	26	1	3	1	1	1	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2
44	29	2	4	2	1	2	1	2	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
45	36	1	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
46	39	1	1	1	2	1	2	1	1	2	2	2	2	2	1	1	1	2	2	2	2	2	2
47	25	1	3	2	2	1	1	1	1	2	2	1	1	1	1	1	1	2	2	1	2	1	2
48	26	1	1	2	1	1	2	2	1	1	2	2	2	1	1	1	1	2	2	2	2	2	2
49	36	1	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
50	39	1	1	1	2	1	2	1	1	2	2	2	2	2	1	1	1	2	2	2	2	2	2
51	20	2	1	2	1	1	1	2	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2
52	20	2	4	2	1	1	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2
53	20	2	1	2	1	2	1	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2
54	65	1	1	2	1	1	1	1	1	2	2	1	2	2	2	1	1	2	2	1	1	1	1
55	29	1	3	2	2	2	2	2	2	1	2	2	2	2	2	1	1	2	2	2	2	2	2
56	21	2	2	1	1	2	1	2	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
57	16	2	4	2	2	1	2	2	1	1	2	2	1	2	2	2	2	2	2	2	2	2	2
58	20	2	4	2	1	2	1	1	1	2	2	2	2	1	2	1	1	2	2	2	2	2	2
59	38	2	1	2	1	2	2	2	1	2	2	2	1	2	2	1	1	2	2	1	2	2	2
60	20	2	4	2	2	2	2	1	1	1	2	2	2	1	1	2	2	2	2	2	2	2	2
61	19	2	2	2	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
62	24	1	3	1	1	1	2	2	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
63	20	2	4	2	1	1	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2
64	20	2	4	2	1	1	1	2	1	1	2	1	2	2	1	2	2	2	2	2	2	2	2
65	24	1	3	1	1	1	2	2	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
66	27	2	4	2	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
67	29	2	4	2	1	2	1	2	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
68	35	1	3	2	1	1	1	2	1	2	2	2	2	2	1	2	2	2	2	2	2	2	2
69	27	2	4	2	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
70	19	2	2	2	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
71	24	1	3	1	1	1	2	2	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
72	33	2	3	2	1	1	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2
73	28	2	3	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	1	2	2
74	29	2	3	2	2	2	1	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2
75	35	2	3	2	1	2	1	1	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2
76	31	1	3	2	1	2	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2
77	39	2	3	2	1	2	2	2	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2
78	34	1	3	2	2	1	1	2	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
79	27	1	3	2	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2
80	39	2	3	1	2	1	1	1	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
81	38	2	3	2	1	1	1	2	2	1	2	1	2	2	2	1	1	2	2	2	2	2	2
82	27	1	3	2	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2
83	34	1	3	2	2	1	1	2	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2
84	28	2	3	2	1	2	1	1	1	1	1	1	1	1	1	1	1	2	2	1	2	1	2
85	39	2	3	1	2	1	1	1	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
86	30	2	3	2	1	1	2	1	1	2	2	2	2	1	1	1	1	2	2	2	1	2	2
87	31	1	3	2	1	2	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2
88	38	2	3	2	1	1	1	2	2	1	2	1	2	2	2	1	1	2	2	2	2	2	2
89	16	2	1	2	1	1	1	2	1	1	2	1	1	1	1	1	1	2	2	2	2	2	2
90	45	2	1	2	1	1	2	2	1	1	2	2	2	1	1	2	2	2	2	2	1	1	2
91	61	2	1	2	1	1	2	1	1	2	2	2	2	1	1	2	1	2	2	1	1	1	2
92	31	2	1	2	1	1	2	1	1	1	2	2	1	1	1	2	2	2	2	2	2	1	2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

93	53	1	1	2	1	1	1	2	1	1	2	2	2	1	1	2	2	2	2	2	1	2	2
94	49	2	1	2	1	1	1	1	1	1	2	2	2	1	1	1	1	2	2	2	2	2	2
95	51	1	1	2	2	1	1	2	2	2	2	2	2	1	1	1	1	2	2	2	1	2	2
96	26	1	3	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
97	30	1	3	1	1	1	1	1	1	2	2	2	2	1	1	2	2	2	2	2	1	2	2
98	28	1	3	1	1	2	1	1	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
99	45	2	1	2	1	2	2	2	1	1	2	1	1	1	1	2	2	2	2	2	2	2	2
100	32	1	3	2	1	2	1	1	1	1	2	1	2	2	2	1	1	2	2	1	2	2	2
101	26	1	3	1	1	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2
102	19	2	1	2	1	1	1	1	1	2	1	2	1	1	1	1	1	2	2	1	2	2	2
103	28	2	2	2	1	2	2	2	1	2	2	2	2	1	2	1	1	2	2	2	2	2	2
104	25	2	1	2	1	2	1	2	1	2	2	2	1	1	2	1	2	2	2	2	1	2	2
105	24	2	2	1	1	2	1	2	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2
106	25	2	2	2	1	2	2	2	1	2	2	2	2	1	1	1	1	1	2	2	2	2	2
107	30	2	3	1	1	1	1	2	1	2	2	1	2	1	1	2	2	2	2	2	1	2	2
108	65	1	1	2	2	1	2	1	2	2	2	2	2	1	2	1	2	1	2	1	1	1	1
109	75	2	1	2	1	2	2	1	2	2	2	2	2	1	1	2	2	1	2	1	1	1	1
110	70	1	1	2	1	1	1	1	1	2	1	2	2	1	1	1	1	1	1	1	1	1	1
111	20	1	2	2	1	1	1	2	1	2	2	1	2	2	1	1	1	2	2	2	3	2	2

Table 11: The Full Dataset From Attribute 24 To Attribute 41

No	att 24	att 25	att 26	att 27	att 28	att 29	att30	att31	att32	att33	att34	att35	att36	att37	att 38	att39	att40	att 41
1	2	2	2	1	1	11	Mild	14-07-20	20-07-20	22-07-20	01-08-20				No			
2	2	2	2	2	1	10	Mild		17-06-20	19-06-20	28-06-20				No			
3	2	2	2	1	1	7	Mild	24-10-20	07-11-20	09-11-20	15-11-20		Asthma		No			
4	2	1	2	1	1	40	Moderate		20-05-20	23-05-20	02-07-20	low back pain and diaphoresis			Yes	6		
5	2	2	2	1	1	11	Mild	11-06-20	18-06-20	19-06-20	29-06-20		Diabetes mellitus		No			
6	2	2	2	1	1	14	Mild	06-12-20	10-12-20	16-12-20	29-12-20	Nasal congesion			No	7		
7	2	2	2	2	2	15	Moderate	22-06-20	30-06-20	28-06-20	12-07-20	fatigue			No	1		
8	2	2	2	1	1	12	Moderate		01-11-20	03-11-20	14-11-20				No			
9	2	2	2	2	2	14	Mild	25-08-20	01-09-20	01-09-20	14-09-20				No			
10	2	1	1	1	1	31	Moderate		24-05-20	24-05-20	24-06-20		Diabetes mellitus and Hypertension		No			
11	2	2	2	1	1	12	Moderate		01-11-20	03-11-20	14-11-20				No			
12	2	2	2	1	1	12	Moderate	12-10-20	14-10-20	15-10-20	26-10-20				No			
13	2	2	2	1	1	19	Moderate	17-06-20	19-06-20	19-06-20	07-07-20				No			
14	2	2	2	2	2	14	Mild	10-07-20	16-07-20	16-07-20	29-07-20				No			
15	2	2	2	1	2	6	Mild		06-12-20	12-12-20	17-12-20				No			
16	1	1	3	1	1	14	Mild		06-08-20	10-08-20	23-08-20				Yes		10-08-20	5
17	2	2	2	2	2	11	Mild	26-03-20	01-04-20	07-04-20	17-04-20	Delayed menstruation			No			
18	2	1	2	1	1	7	Moderate	27-06-20	01-07-20	04-07-20	10-07-20				No			
19	2	2	2	1	1	11	Mild	11-06-20	18-06-20	19-06-20	29-06-20		Diabetes mellitus		No			
20	2	2	2	2	2	11	Mild	24-03-20	01-04-20	07-04-20	17-04-20	Delayed menstruation			No			
21	2	2	2	1	3	10	Mild	21-05-20	25-05-20	26-05-20	04-06-20				No			
22	2	2	2	2	2	11	Mild		23-06-20	25-06-20	05-07-20				No			
23	2	1	2	1	1	46	Moderate	12-05-20	14-05-20	16-05-20	30-06-20	chest pain			No			
24	2	2	2	1	1	10	Mild	07-06-20	11-06-20	21-06-20	30-06-20	fatigue			No			
25	2	2	2	2	2	11	Mild		23-06-20	25-06-20	05-07-20				No			

26	2	1	2	1	1	31	Critical		22-06-20	22-06-20	22-07-20		Diabetes mellitus		No	5		
27	2	2	2	1	1	27	Moderate	17-07-20	22-07-20	22-07-20	17-08-20	anosmia and Loss of sense of taste			No			
28	2	2	2	1	1	13	Mild	04-07-20	07-07-20	08-07-20	20-07-20				No			
29	2	2	2	2	1	21	Mild	01-07-20	20-07-20	20-07-20	09-08-20		hypertension	gout	No			
30	2	2	2	1	1	12	Mild	11-11-20	20-11-20	25-11-20	06-12-20				No	4		
31	2	2	2	1	1	11	Mild	05-06-20	07-06-20	07-06-20	17-06-20	Dry cough			No			
32	2	2	2	1	1	11	Moderate	01-07-20	12-07-20	12-07-20	22-07-20		Diabetes mellitus		No			
33	2	2	2	1	2	11	Moderate	15-06-20	20-06-20	25-06-20	05-07-20				No			
34	2	2	2	2	1	11	Mild	09-08-20	15-08-20	26-08-20	05-09-20				No			
35	2	2	2	1	1	16	Mild		22-08-20	23-08-20	07-09-20	fatigue	Hypertension	Allergic rhinitis	No			
36	2	2	2	2	2	12	Moderate		15-11-20	25-11-20	06-12-20				No	7		
37	2	2	2	1	2	9	Mild	18-11-20	22-11-20	22-11-20	30-11-20				No			
38	2	2	2	2	1	11	Mild	09-08-20	15-08-20	26-08-20	05-09-20				No			
39	2	2	2	1	1	13	Mild	04-07-20	07-07-20	08-07-20	20-07-20				No			
40	2	2	2	1	1	11	Moderate	01-07-20	12-07-20	12-07-20	22-07-20		Diabetes mellitus		No			
41	2	2	2	1	1	11	Mild	01-08-20	05-08-20	10-08-20	20-08-20	fatigue			No	3		
42	2	2	2	1	1	10	Mild		04-10-20	06-10-20	15-10-20		immunocompromised		No	2		
43	2	2	2	1	3	10	Mild	21-05-20	25-05-20	26-05-20	04-06-20				No			
44	2	2	2	1	2	11	Mild	31-07-20	05-08-20	09-08-20	19-08-20	Dizziness			No	3		
45	2	2	2	1	2	10	Mild		26-10-20	29-10-20	07-11-20				No			
46	2	2	2	1	2	6	Moderate		01-11-20	16-11-20	21-11-20				No	1		
47	1	1	1	1	1	21	Critical		25-08-20	26-08-20	15-09-20				Yes	1	27-08-20	1
48	2	2	2	1	1	10	Moderate	25-07-20	03-08-20	05-08-20	14-08-20				No			
49	2	2	2	1	2	10	Mild		26-10-20	29-10-20	07-11-20				No			
50	2	2	2	1	2	6	Moderate		01-11-20	16-11-20	21-11-20				No	1		
51	2	2	2	1	2	5	Mild	10-08-20	14-08-20	14-08-20	18-08-20				No			
52	2	3	3	1	3	11	Mild	01-09-20	02-09-20	05-09-20	15-09-20				No			
53	2	2	2	2	1	10	Mild	30-10-20	30-10-20	08-11-20	17-11-20				No			
54	1	1	3	1	1	33	Critical	05-08-20	07-08-20	07-08-20	08-09-20		Diabetes mellitus and Hypertension		Yes	3	10-08-20	30
55	2	1	2	1	1	11	Mild		03-08-20	04-08-20	14-08-20				No			
56	2	2	2	1	2	13	Moderate	23-08-20	27-08-20	29-08-20	10-09-20	nasal dryness			No			
57	2	3	3	1	3	11	Mild		11-12-20	11-12-20	21-12-20				No			
58	2	3	2	1	1	37	Moderate	13-07-20	23-07-20	25-07-20	30-08-20				No			
59	2	2	2	1	1	21	Mild	05-10-20	09-10-20	10-10-20	30-10-20				No	2		
60	2	2	2	1	2	11	Mild		02-12-20	04-12-20	14-12-20				No	2		
61	1	2	2	1	1	16	Moderate	07-07-20	12-07-20	14-07-20	29-07-20				No			
62	2	2	2	1	1	11	Mild	15-11-20	16-11-20	16-11-20	26-11-20				No	1		
63	2	3	3	1	3	11	Mild	01-09-20	02-09-20	05-09-20	15-09-20				No			

64	2	2	2	1	1	33	Mild	04-10-20	07-10-20	09-10-20	10-11-20	low back pain and diaphoresis			No			
65	2	2	2	1	1	11	Mild	15-11-20	16-11-20	16-11-20	26-11-20				No	1		
66	2	3	3	1	2	13	Moderate	21-11-20	01-12-20	05-12-20	17-12-20	Decreased vision			No			
67	2	2	2	1	2	11	Mild	31-07-20	05-08-20	09-08-20	19-08-20	Dizziness			No	3		
68	2	2	2	1	1	9	Moderate	26-09-20	30-09-20	30-09-20	08-10-20				No			
69	2	3	3	1	2	13	Moderate	21-11-20	01-12-20	05-12-20	17-12-20	Decreased vision			No			
70	1	2	2	1	1	16	Moderate	07-07-20	12-07-20	14-07-20	29-07-20				No			
71	2	2	2	1	1	11	Mild	15-11-20	16-11-20	16-11-20	26-11-20				No	1		
72	2	2	2	1	2	9	Mild	01-06-20	10-06-20	10-06-20	18-06-20				No			
73	2	2	2	2	2	13	Mild		17-06-20	18-06-20	30-06-20		immunocompromised		No			
74	2	1	2	1	1	11	Mild		16-07-20	21-07-20	31-07-20			Pregnant	No			
75	1	2	2	1	1	26	Mild	19-09-20	24-09-20	25-09-20	20-10-20				No	5		
76	2	2	2	2	1	11	Mild	04-08-20	08-08-20	08-08-20	18-08-20				No			
77	2	2	2	1	1	12	Mild	20-10-20	23-10-20	23-10-20	03-11-20	back pain			No			
78	2	2	2	1	1	17	Moderate		02-07-20	02-07-20	18-07-20				No			
79	2	2	2	2	1	14	Mild		23-11-20	23-11-20	06-12-20				No			
80	2	1	2	1	1	15	Mild		01-06-20	04-06-20	20-06-20				No	2		
81	2	1	1	2	1	13	Mild	07-03-20	15-03-20	20-03-20	01-04-20				No			11
82	2	2	2	2	1	14	Mild		23-11-20	23-11-20	06-12-20				No			
83	2	2	2	1	1	17	Moderate		02-07-20	02-07-20	18-07-20				No			
84	1	1	1	1	1	33	Moderate	12-03-20	15-03-20	19-03-20	20-04-20	sever cough			Yes	1	20-03-20	30
85	2	1	2	1	1	17	Mild		01-06-20	04-06-20	20-06-20				No	2		
86	2	2	2	1	1	23	Mild		24-11-20	25-11-20	17-12-20		Asthma		No	4		
87	2	2	2	2	1	11	Mild	04-08-20	08-08-20	08-08-20	18-08-20				No			
88	2	1	1	2	1	13	Mild		15-03-20	20-03-20	01-04-20				No			11
89	2	2	2	1	1	5	Moderate	22-07-20	27-07-20	06-08-20	10-08-20				No			
90	2	1	2	1	1	11	Moderate	24-07-20	27-07-20	28-07-20	08-08-20		Hypertension		No			1
91	1	1	1	1	1	20	Critical	12-07-20	17-07-20	22-07-20	10-08-20		Hypertension and cardiac disease		Yes	5	25-07-20	7
92	1	1	2	1	1	25	Moderate	18-07-20	24-07-20	27-07-20	20-08-20	Dizziness			Yes	1		1
93	2	2	2	1	2	6	Mild	10-03-20	15-03-20	20-03-20	25-03-20		Diabetes mellitus		No			
94	2	1	1	1	1	15	Moderate		08-08-20	10-08-20	24-08-20				No			
95	2	1	2	1	1	10	Moderate		01-09-20	01-09-20	10-09-20		Diabetes mellitus		No			
96	2	2	2	1	1	26	Moderate	19-06-20	21-06-20	21-06-20	16-07-20		Asthma		No			
97	2	2	2	1	1	11	Mild	15-07-20	19-07-20	20-07-20	30-07-20		Asthms		No			
98	2	2	2	1	1	21	Moderate	15-06-20	20-06-20	20-06-20	10-07-20				No			
99	2	2	2	1	1	11	Moderate	19-11-20	22-11-20	23-11-20	03-12-20	Abdominal pain			No			
100	2	1	1	1	1	22	Moderate		12-06-20	20-06-20	11-07-20	Skin rash		rheumatoid arthritis	No			
101	2	2	2	1	2	10	Mild	18-06-20	20-06-20	22-06-20	01-07-20				No			

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

102	2	2	2	1	1	12	Mode rate	18-11-20	24-11-20	24-11-20	05-12-20				No	4		
103	2	2	2	1	1	16	Mild		23-10-20	22-10-20	06-11-20				No			
104	2	2	2	2	2	13	Mild	26-06-20	01-07-20	03-07-20	15-07-20	Dizziness	Diabetes mellitus	breastfeed mother	No			
105	2	2	2	1	1	9	Mild	20-07-20	22-07-20	24-07-20	01-08-20				No			
106	2	2	2	2	1	38	Mode rate	15-10-20	19-10-20	21-10-20	28-11-20				No			
107	1	2	2	1	1	15	Mode rate	14-07-19	18-07-19	19-07-19	02-08-19		Asthma		Yes	2		
108	1	1	1	1	3	0	Death		31-07-20	07-08-20	09-09-20	Dizziness	Hypertension,cardiac disease and diabestes mellitus		Yes	Before test	07-08-20	60
109	1	3	3	1	3	0	Death	04-08-20	07-08-20	07-08-20	01-09-20		Hypertension,chronic kidney disease and diabestes mellitus		Yes	10	19-08-20	13
110	1	1	1	1	1	0	Death	06-07-20	14-07-20	17-07-20	19-10-20		Hypertension, cardiac disease and asthma		Yes	3	14-07-20	41
111	2	2	2	1	1	0	Death	03-12-20	03-12-20	05-12-20	20-12-20				No			