Discovering COVID-19 Death Patterns from Deceased Patients: A Case Study in Saudi Arabia

Abdulrahman Alomary[1]
Computer Science Department
Jamoum University College
Umm Al-Qura University
Jamoum, Saudi Arabia

Tarik Alafif[2]
Computer Science Department
Jamoum University College
Umm Al-Qura University
Jamoum, Saudi Arabia

Abdulmohsen Almalawi[3]
Computer Science Department
King Abdulaziz University
Jeddah, Saudi Arabia

Anas Hadi[4]
Computer Science Department
Arab Open University
Riyadh, Saudi Arabia

Faris Alkhilaiwi[5]
Natural Products and Alternative Medicine Department
King Abdulaziz University
Jeddah, Saudi Arabia

Yasser Alatawi[6]
Pharmacy Practice Department
University of Tabuk
Tabuk, Saudi Arabia

***Abstract*—COVID-19 is a serious infection that cause severe injuries and deaths worldwide. The COVID-19 virus can infect people of all ages, especially the elderly. Furthermore, elderly who have co-morbid conditions (e.g., chronic conditions) are at an increased risk of death. At the present time, no approach exists that can facilitate the characterization of patterns of COVID-19 death. In this study, an approach to identify patterns of COVID-19 death efficiently and systematically is applied by adapting the Apriori algorithm. Validation and evaluation of the proposed approach are based on a robust and reliable dataset collected from Health Affairs in the Makkah region of Saudi Arabia. The study results show that there are strong associations between hypertension, diabetes, cardiovascular disease, and kidney disease and death among COVID-19 deceased patients.**

Keywords*—*COVID-19; association rules; Apriori algorithm; patterns; death; chronic diseases

I. Introduction

COVID-19 first appeared in the Chinese city of Wuhan in December 2019, and it was the starting point of the disease spreading to the world. It was officially announced by the World Health Organization (WHO) on March 11, 2020 a pandemic [1]. Many attempts were involved to engage latest technologies for diagnosing and treating COVID-19 patients [2].

The Kingdom of Saudi Arabia (KSA) was not immune from that virus, as the virus spread rapidly. The first confirmed recorded case appeared in the KSA on March 2, 2020 [3]. There were several factors that contributed to the spread of the virus in the KSA, including its geographic location, trade exchanges with China, and its religious and recreational tourism industries [4], [5]. The KSA had previous experiences with epidemics. Viruses such as the Middle East Respiratory Syndrome (MERS-CoV), helped to control and deal well with the pandemic [6]. Since the start of the pandemic until 10 Nov 2022, there have been 284,273 cases and 9,426 deaths of COVID-19 in the KSA [7].

As a precautionary measure, the KSA had implemented strict curfews, shut down all non-essential services, required everyone to wear masks, and closed its air, sea, and land borders. [8], [9]. The government had also transferred education to distance education in all public and private schools, universities, institutes and training centers [10]. Furthermore, Saudi Arabia intensified health care and established field hospitals to cope with the likely increase in COVID-19 patients, and relied on WHO and Ministry of Health protocols when treating COVID-19 patients [11], which significantly decreased the rate of death and hospitalization [12]. In the latest informed statistics during the writing of the last chapter of this research on December 29, 2022, the confirmed cases exceeded 816,470 cases, and 9,610 deaths of COVID-19 patients [13].

Most of the previous studies showed that the large number of people who died from COVID-19 had one or more chronic diseases in a remarkable way. Hence, the need to study those patterns of death from chronic diseases that are more frequent. Understanding these patterns helps to identify the factors that increase the risk of death in this group of people, and thus can help in developing better strategies to control the pandemic and provide a priority care. The issue of obtaining the database of the deceased patients from COVID-19 is difficult due to confidentiality, privacy, and sensitivity of medical data. In this proposed work, the medical data is officially obtained through health protocols from the Health Affairs in Makkah Al-Mukarramah region in the Kingdom of Saudi Arabia.

Our work is motivated by Apriori Association Rules (ARA) algorithm that was introduced in [14], which extracts the most frequent and appearing patterns from a specific database according to the degree of support and confidence [15]. Our work is similar to the work presented in [16]. However, we, in this study, discover COVID-19 death patterns from deceased patients rather than discovering COVID-19 infection patterns.

The remainder of the paper is organized as follows: In Section II, we present related works. In Section III, we describe and analyze the dataset and discuss our proposed approach for detecting death patterns among COVID-19 patients. In Section IV, we describe the experimental procedure based on the algorithm used. Section V discusses the Weka program while Section VI details implementation of Weka program in our study. Section VII presents the results and discussion. In Section VIII, we conclude our study and present our future work.

II. Related Work

Several research works thoroughly studied the COVID-19 patterns. In this section, we review the most recent related works.

Robert et al. [17] introduced the present COVID-19 situation among diabetics, newly diagnosed diabetics, diabetic ketoacidosis, and programmatic initiatives including immunizations. The study method involved performing a literature study through the use of PubMed, Google, and Scopus Up until July 15, 2021. The study results had shown that most research conducted in the KSA had shown diabetes as one of the most common comorbidities among COVID-19 patients. There had been few works conducted in the KSA on COVID-19-induced diabetic ketoacidosis and newly diagnosed diabetes. The Saudi ministry had implemented a number of steps, including thorough recommendations and prioritized immunizations, to reduce the impact of COVID-19 among individuals with diabetes. Telehealth services were heavily utilized in Saudi Arabian diabetes clinics during the COVID-19 pandemic.

Geng et al. [18] attempted to comprehend the risk variables for symptom deterioration and death in COVID-19 patients. By establishing the predictive value of chronic diseases for COVID-19 severity and mortality, this systematic review intended to fill the gap. The study results had revealed that hypertension was a fairly prevalent illness among COVID-19 patients, linked to mortality, admission to the intensive care unit (ICU), acute respiratory distress syndrome, and higher severity. Asthma was linked to a lower likelihood of COVID-19 death, while chronic obstructive pulmonary disease was the best predictor of COVID-19 severity, ICU admission, and fatality. Instead of mortality, the fat patients were more likely to develop severe COVID-19 symptoms. The patients were more likely to develop severe COVID-19 instances and die if they had cancer, chronic liver disease, chronic renal illness, or cerebrovascular disease.

Wang et al. [19] conducted searches in several databases for articles published until April 6, 2020, and after selecting a very large number of published articles. Only 34 articles were reached after sorting. Among the pre-existing chronic diseases were having high blood pressure, cardiovascular disease, kidney disease and diabetes associated with the risk of infection with the virus. The results confirmed that people with pre-existing chronic diseases had an increased risk of developing more serious complications in COVID-19 patients, and severe organ damage or dysfunction was linked to an increased death rate. The study showed that acute kidney disease and heart injury were closely associated with a 3-4-fold higher risk of death associated with COVID-19. The study suggested that healthcare providers need to put those who have a history of high blood pressure and cardiovascular diseases under observation and be subject to continuous follow-up.

Al Mutair et al. [20] collected information from several private hospitals in the KSA between April 2020 to June 2020. The study was based on descriptive and inferential analysis of the results and on the data analysis of COVID-19 patients' information. The study was classified into two categories: survivors (recovered) and non-survivors (deceased). The inferential reading showed that 31.8% of the survivors were in the age group of 30-40 years, and 24% were in the age group of 21-30 years, and among the non-survivors, the study showed that 66% of the non-survivors over 50 years old, and that 86% of the non-survivors were males. The results showed that 63.8% had no history of chronic diseases Hypertension (HTN), that 19.8% had one of HTN or diabetes mellitus (DM), and that 16.4% had a history of chronic diseases HTN and DM together. The study indicated a strong relationship between differences in age, gender, chronic diseases and deaths associated with COVID-19 in general, and that advanced age and males and the presence of chronic diseases such as high blood pressure and diabetes among COVID-19 patients are prevalent among the deceased patients. The study indicated that there were poor predictive factors that led to the possibility of an increase in the death rate that may reach 9 times among males over the age of 50 years and who have a history of chronic diseases. There is an association between HTN and mortality, and there is an increasing number of deaths among males.

Similar to our work, Alafif et al. [16] explored the most common infection patterns for patients who had recovered from COVID-19, and they were able to collect 131 records of people who recovered from COVID-19 using the survey method through a questionnaire by communicating with them or with their relatives by direct contact. The ARA algorithm was used on the manually collected data to explore the most common infection patterns. The study concluded that there were strong association rules with high confidence scores among males, weight above 70 kg, height above 160 cm, and fever patterns. Also, Alafif et al. [21] predicted the status of COVID-19 patients using their patterns as COVID-19 treatments and diagnosis were reviewed in [22].

From the previous studies, we find most of these studies confirm that the age factor, males, non-Saudis, and those with chronic diseases are the most at risk of death among people infected with the COVID-19, with differences in ratios, study sample and place of study. It was also found that the most prevalent chronic diseases among those infected and deceased from the COVID-19 virus are hypertension, diabetes, cardiovascular disease and kidney disease.

We did not find any previous study that dealt with the same topic and finding of the most frequent chronic diseases found among the deceased due to the COVID-19 death pattern, as we are in this study to discover the death patterns of chronic diseases that were among the deceased and the most frequent.

We discuss the results of our study and a comparison with previous studies in the discussion section, clarifying the compatibility with previous studies and the contradiction of their results.

III. Methodology

Fig. 1 shows the course of the study stages, from defining the study sample, obtaining the database, the stages of cleaning, coordinating and arranging the database, and analyzing the data, then the stage of selecting the study tools, which is the ARA algorithm, followed by the use of two methods to apply the algorithm through manual calculations and the Weka program to verify the conformity of the outputs and finally the stage of writing the results.

A. Association Rules Algorithms

In our modern era, the uses of artificial intelligence are abundant to utilize as much as possible. These can be employed in several fields, including data science, data mining, data analysis, and serial cases that in the past took effort a long and hard time. The most classic algorithm is the ARA algorithm, which is considered the originator of the recursive element mining algorithm [23].

The amount of information or data being maintained in databases has dramatically increased in recent years. Latent knowledge mining is crucial to enhance decision-making when the size of available databases grows exponentially. The critical phase of the knowledge discovery process is data mining. Predictive and descriptive tasks are the two broad categories into which data mining's primary tasks are typically categorized. While the descriptive tasks aim to extract previously undiscovered and significant information from massive databases, such as patterns, associations, changes, anomalies, and substantial structures, the predictive tasks aim to forecast the value of a given attribute based on the standards of other characteristics. These goals of data mining can be met in many ways [24].

There has been more than one algorithm whose works are based on association rules, such as the ARA algorithm, the ECLAT algorithm, and the FP Growth algorithm, but the ARA algorithm is the most famous. The ARA algorithm appeared by Agrawal et al. in 1993 as one of the tools for mining databases of association data. The ARA algorithm extracts recurring patterns are based on the strength of the degree of support and confidence and provide services through their analysis to planning and marketing managers to facilitate the administrative and marketing process and help make appropriate decisions [25].

The main task of the ARA algorithm is to extract the duplicate elements with iterative operations by scanning the database more than once according to the minimum support limit, excluding other data and filtering [26].

Association rules that explore recurring, relational, or causal patterns between groups of itemset emerged from the database. The rules of association were used through the exploration and analysis of shopping cart data at the beginning, and their use has evolved in several areas due to their efficiency and effectiveness in results. The science of data mining is of great importance to obtain information from large or small sources of various categories and to present it for optimal use. The rules of the association are one of the most important methods of data mining. There are several measures by which the strength of the degree of correlation of the algorithm outputs is measured, and we have relied on two measures support and confidence. The method of knowledge discovery in database, is called Data mining. It is based on extracting important patterns from large databases and then analyzing that data to give beneficial information, ideas, and solutions. Through what the discovery of knowledge in databases provides in terms of methods and methods for extracting interesting remarkable patterns in the shortest possible and appropriate period and taking paths at the intersection point of data stores and machine learning, and statistics systems. Data mining is known as applying algorithms to desired patterns found in large databases [27].

The ARA algorithm is an intriguing way to determine what we need to buy or get suggestions for what we need. We are all aware that the e-commerce platform has a variety of approaches. Flipkart, Amazon, Snapdeal, and other online retailers, are what it exactly is. The application proposes items to buy when we attempt to purchase in the online store. It anticipates additional clients who usually make purchases together. This method also enables us to understand how various things are predicted [28].

B. Apriori Algorithm

The Apriori algorithm is an analogue algorithm for association base mining (association base mining) and ranks among the top ten data mining algorithms. Association rule mining is a significant research direction in data mining. It is also a long-standing topic whose main task is to find the inner connection between things [29]. The association rules approach can be implemented using various methods to draw out relevant links between variables or items in a large database by identifying more frequent item combinations. As long as those itemsets appear in the dataset sufficiently and frequently, it recognizes the frequently occurring items in the database and expands them to larger and larger itemsets. If an itemset is frequent, then all of its subgroups' items will also be frequent, according to the Apriori algorithm. In other words, any subset of a frequent itemset must also be frequent [30].

C. History of the ARA

The first algorithm to be suggested for frequent itemset mining was the ARA algorithm. Later, it was enhanced by R Agarwal and R Srikant [14]. The result was known as Apriori. The join and prune steps of this algorithm are used to condense the search space. Finding the most common itemset is done iteratively. The term ARA refers to the algorithm that determines the rules of association between items. It refers to the relationship between two or more objects [15]. To put it another way, we can say that the ARA algorithm is a leaning association rule that examines whether customers of product (A) also purchase product (B) [31].

D. The ARA Algorithm Template

To find the huge l-items, the algorithm initially counts the occurrences of each item. These stages make up the second step. The candidate itemsets C_k are first created using the large itemsets $L_k - 1$. After that, the database is scanned to determine the support of candidates in C_k, which involves increasing the count of all candidates included in a certain transaction t. After that, *supp_min* is used to compare the support of each candidate *k-itemset*. Only when the support is more than the minimal support given by the user, supp_ min, is the *k-itemset* frequent. The algorithm then deletes all rules that do not satisfy the criterion of *supp_conf*. The next stage is to identify those rules with confidence greater than or equal to the user-specified *supp_conf* [32].

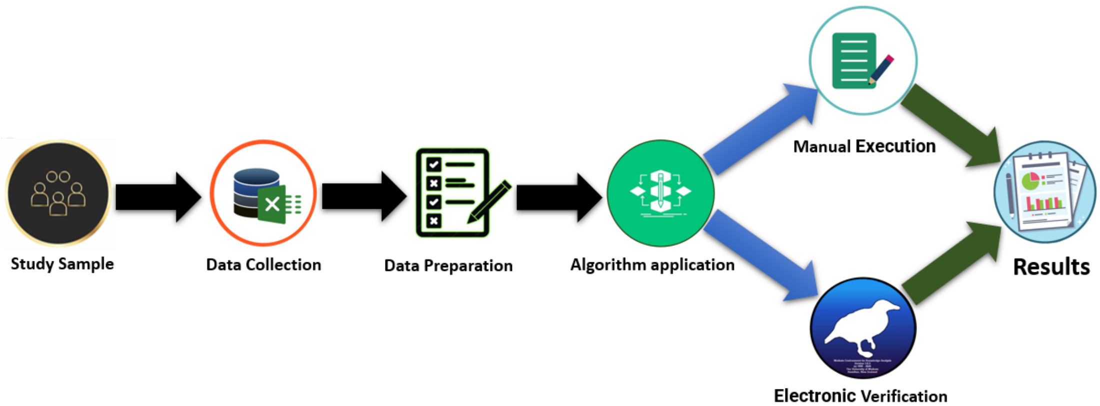

Fig. 1. The pipeline methodology of the proposed approach.

E. Components of the ARA Algorithm

The ARA algorithm consists of more than one component, there are several measures by which the strength of the degree of correlation of the algorithm outputs is measured [33], and we have relied on two criteria; support and confidence. The term support refers to the frequency of itemset in the dataset on all parameters and is defined as follows:

$$Support(X \Rightarrow Y) = \frac{x \cup y}{TotalTransactions} \tag{1}$$

The term (confidence) refers to the frequency of the dataset of utmost *Y* in transactions containing *X*, which are defined as:

$$Confidence(X \Rightarrow Y) = \frac{supp(x \cup y)}{supp(x)} \tag{2}$$

We employed the ARA algorithm to compute the probability of the appearance of the most frequent chronic diseases along with other chronic diseases in the database of the deceased patients from COVID-19 to provide the results to medical providers to take the necessary precautions or provide appropriate health care. In an iterative process, the data is filtered to extract the itemset that achieves the minimum support, and then it is re-read to extract a new itemset, considering *k+1*. As appears in Fig. 2.

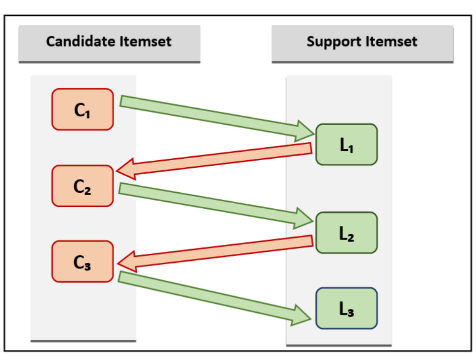

Fig. 2. Filtering iterations.

Fig. 3 shows a simple example of the work of the ARA algorithm in analysing the shopping cart, where the database is completely read, and then each item and its number of repetitions is separately determined. Then, the itemset is filtered according to the minimum support specified in advance by the user, which we call *C1* to produce a grouping of the dataset that meets the minimum support as a single item *1-Itemset* with the exclusion or deletion of any item that does not meet the threshold of the minimum support and this is what we call *L1*.

With *L1* we iterated the process but with two items together *K+1* or 2-Itemset which is the process of filtering the groups of itemset *C2* comparing the number of iterations with the minimum support, and then trimmed the elements that did not meet the threshold of the minimum support. The process was repeated according to the repetition of the itemset with the increase of *K+1* each time, meaning that in the third time, the itemset was three with each other 1-Itemset and compared the number of iterations with the threshold of the minimum support until we reached that it was not possible to create a new group of itemset because the minimum support was not achieved. There was a logical consequence that a non-frequent set cannot be in a set of frequent itemsets. Next, we computed the specific confidence value for those potential itemsets that were most frequent, Accordingly, we saw the strength of that confidence and the correlation of the group of elements whether appearing next to each other or not.

F. Dataset

Our study is based on an Excel database of those who died due to COVID-19 and had a disease or chronic diseases that contributed to the death. After taking approval, the database was obtained from the information centre at Al-Noor Specialist Hospital in Makkah. We processed, audited, revised, and corrected the database, as it initially contained 18,785 records, most of which were duplicates, incomplete, or invalid. The reason for its large size is due to the placement of more than one record for the same patient, where each one contains its patient's daily and weekly condition development independently, but neither an update nor a placement of a cell for the new case, as well as the patient's data, such as medical

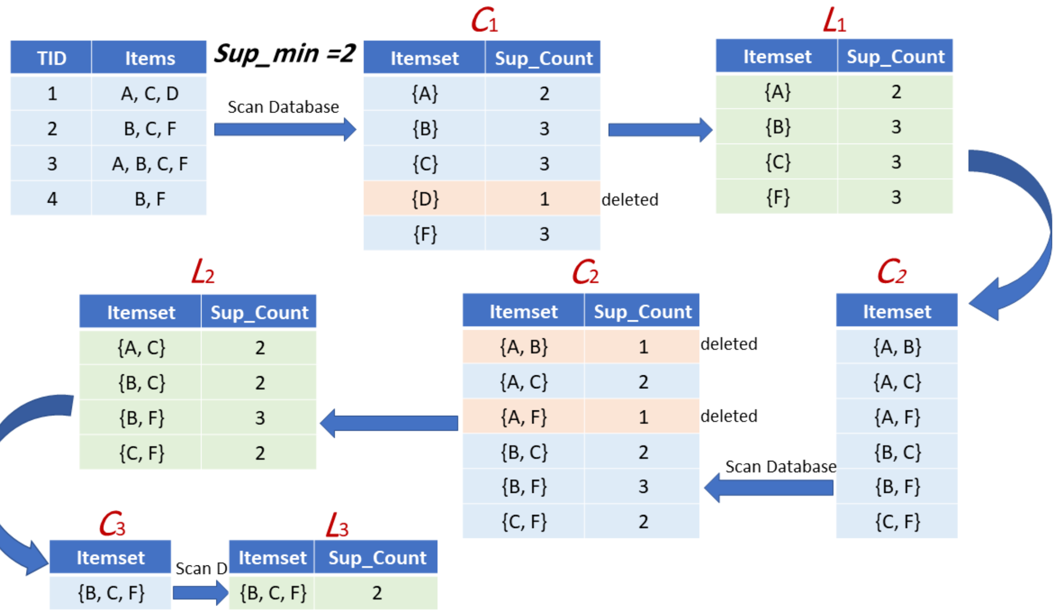

Fig. 3. A simple example of the workflow of the ARA algorithm.

number, age, gender, and nationality. Some patients had more than 43 records for each of them individually. This is a large number and major reason for the records to search 18,785. This is if we know that every patient data has a recurrence. At the end of the corrections and audits, the number of valid records for the study was only 349 for those who died due to COVID-19 and were accompanied by one or more chronic diseases such as diabetes, hypertension, cardiovascular, kidney diseases, immunodeficiency, blood diseases, cancer, immunodeficiency, cardiovascular, and obesity. The dataset is publicly available on[1] .

G. Database Analysis

The database included 349 records of those who died of COVID-19 and had one or more chronic diseases, and the number of registered diseases amounted to 10 different ones, as shown in Fig. 4.

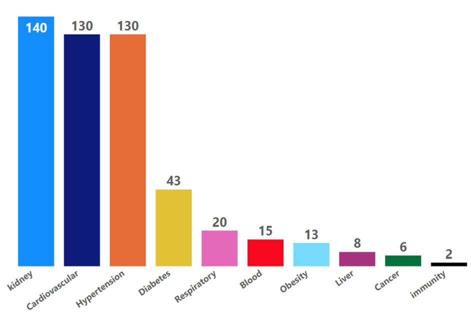

Fig. 4. Chronic diseases and how often they appear in the database in the study population.

[1] http://www.kaggle.com/datasets/abdulrahmanalomary/discovering-covid-19-death-patterns

The percentage of the non-Saudis was more than of the Saudis, according to the results of the database analysis of the deceased patients. The non-Saudis number was 217 and the Saudis was 132, as shown in Fig. 5.

The number of non-Saudis constitutes from 20 different countries. The countries are Jordan, Afghanistan, Indonesia, Pakistan, Burma, Bangladesh, Thailand, Chad, Tunisia, Syria, Sudan, Philippines, Palestine, Malaysia, Egypt, Morocco, Mauritania, Nigeria, India, and Yemen.

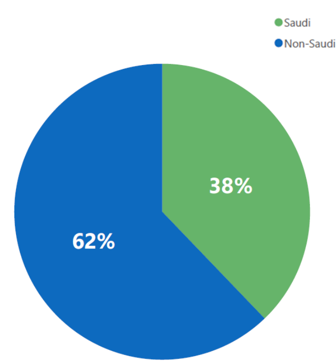

Fig. 5. The ratio of Saudi to non-Saudi COVID-19 patients in the study population.

Fig. 6 shows that the analysis results of the database also showed that the number of males exceeds the number of females in deaths. According to the previous studies we reviewed, they were fully compatible with the fact that the number of males exceeded the number of females since the onset of the disease until the date of writing the study.

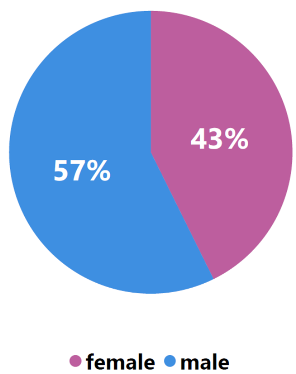

Fig. 6. The ratio of males to females in the study population.

Also, the database records of the deceased patients also show a diversity of blood groups. The blood group O+ is more prominent, followed by A+, then B+, and then the rest of the blood groups, as shown in Fig. 7. However, there are many records in the database of unknown blood types (not recorded) from the source and written in the field (unknown). Fortunately, we do not rely on blood type to explore the most frequent patterns of the cause of death, as the type of chronic diseases and their nomenclature are the study subjects for the implementation of treatment operations.

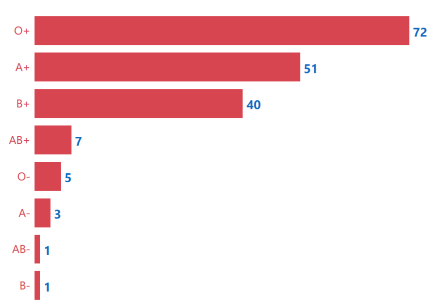

Fig. 7. Types of blood groups in the study population.

The strength of youth appears in all aspects, as the nature of youth is strength, vitality, activity, and endurance of hardships, as well as the case when diseases are infected, we find the strength of immunity or resistance, which is different from the elderly.

The study showed the disparity in the ages of the deceased patients with COVID-19. They had chronic diseases, as the ages of young people age range from 15 to 45 which represented only 13% of the total of deceased patients according to the study's database.

On the other hand, we find that the elderly, whose age ranges from 46 to over 80, are the highest percentage of deaths, reaching 87%. Fig. 8 shows the distribution of ages by age groups into groups.

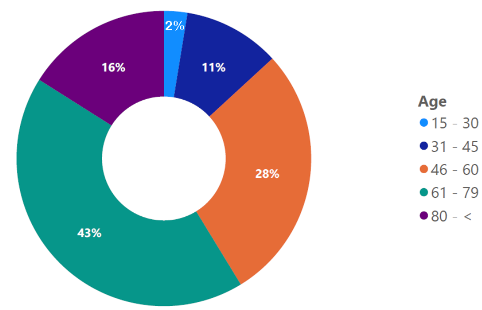

Fig. 8. Distribution of age groups in the study population.

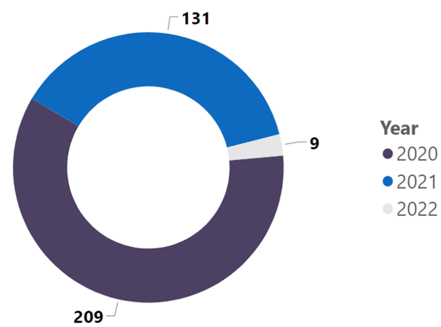

Fig. 9. Total deaths by years in the study population.

COVID-19 appeared at the end of the year 2019 and spread in the year 2020, the disease spread largely and frighteningly, as well as the number of deaths, then the percentage decreased in the following year 2021 until the percentage decreased significantly in the year 2022. Fig. 9 shows the number of deaths during the three consecutive years 2020, 2021 and 2022, and the significant difference between them.

IV. Implementation of Algorithm

In our study, we relied on the ARA algorithm to discover (patterns of death) according to strong association rules through the database of deceased COVID-19 patients. In this section, we manually applied the ARA algorithm at first, the database file (Excel) was manually emptied and written as in Table IV. The medical file number for each deceased patient has been replaced by serial numbers. The names of chronic diseases (patterns) were abbreviated with corresponding letters for ease of writing. We had 10 different types of chronic diseases (patterns) in the database, and their names have been abbreviated with Latin letters for ease of reading, writing, and dealing with them during the manual analysis, as explained in Table I.

TABLE I. LIST OF TYPES THE DISEASES AND THEIR ABBREVIATIONS

Disease Name	Abbreviation
Kidney Diseases →	**K**
Hypertension →	**H**
Cardiovascular →	**C**
Diabetes →	**D**
Respiratory →	**R**
Blood Diseases →	**B**
Obesity →	**O**
Liver Diseases →	**L**
Cancer →	**N**
Immunodeficiency →	**I**

The minimum support was determined according to our reading and estimation of the size of the records and the presence and frequency of patterns in the database. The minimum support is 7.

TABLE II. RESULTS OF THE INITIAL FILTERING PROCESS FOR THE DATASET (L1)

Patterns	Supp_Count
{B}	15
{C}	130
{D}	43
{K}	140
{L}	8
{O}	13
{R}	20
{H}	130

We first performed the process of reading all the patterns, an element of an element from the database (1- frequent), and the number of supporting each element. This is called a filtering process C1 as in Table III. Depending on the predetermined minimum support, the patterns were pruned (deleted) and excluded for the items that did not meet the minimum support. Table II embodies L1.

TABLE III. SCANNING THE DATABASE AND READING ALL PATTERNS AND THEIR FREQUENCY (C1)

Patterns	Supp_Count	
{B}	15	
{C}	130	
{D}	43	
{I}	2	deleted
{K}	140	
{L}	8	
{N}	6	deleted
{O}	13	
{R}	20	
{H}	130	

According to the results of the first filter C1 in Table III, it is clear that the pattern set I and N did not achieve the minimum support 7, and pruning (deletion) was implemented. Table V shows the second filtering process C2 which was performed by reading two 2-frequent items for all patterns from Table II (L1).

In the next step, the filter results L2 were extracted for every two items (2-Itemset) according to the minimum support from reading Table V (C2), and the patterns that did not meet the minimum support were excluded. Many patterns for which the number of supporting was below the threshold were excluded. 7 frequent patterns 2-itemset were obtained as it appears in Table VI.

TABLE VI. RESULTS OF THE SECOND FILTRATION PROCESS (L2) 2-ITEMSET

Patterns	Supp_Count
{B, H}	7
{C, D}	9
{C, K}	30
{C, H}	31
{D, K}	15
{D, H}	28
{K, H}	38

After knowing the number of support and obtaining a successful reading for each of the two elements (L2), and since there is a repetition that fulfilled the minimum condition for support, we moved to the next step through which all the three elements (patterns) read 3-Itemset, i.e., forming new triple patterns based on the results of Table VI (L2).

In Table VII, (C3), the data was filtered with a 3-Itemset for every three items. The number of repetitions was computed based on the results of Table V.

Through the results of the 3-Itemset reading of the three repetitive patterns, only two groups that met the threshold of support are shown C, K, H and D, K, H. The first group of elements was repeated 13 times, and the second 10 times, as shown in Table VIII.

We continued filtering the data to generate a new set of items (Patterns) for every four 4- Itemset, where there were two sets, and the number of supporting for each set was greater than the threshold minimum support.

In Table IX (L4), 4- Itemset data is filtered for every four items, and the number of repetitions is computed based on the results of Table VIII (L3), which achieved the threshold minimum support. The pattern set shows C, D, K, and H.

TABLE IV. LIST OF TYPES (CHRONIC DISEASES) DATABASE RECORDS OF COVID-19 DEATHS

ID	Patterns (Diseases)	ID	Patterns (Diseases)	ID	Patterns (Diseases)	ID	Patterns (Diseases)
1	H,D	89	H,R	177	K,H	265	C
2	K	90	C	178	H	266	B
3	K	91	K,I	179	H	267	L
4	C	92	H,D,R	180	C	268	K
5	K	93	C	181	H	269	O
6	H	94	H	182	H	270	K
7	K,H,C	95	K	183	H,C	271	K,C
8	K,C,R	96	L	184	H	272	C
9	K	97	C	185	C	273	H,D
10	C	98	H,D	186	H,C	274	K
11	H,D	99	K	187	C	275	C
12	H,D	100	K	188	K	276	B
13	H	101	K,H,C,D,R	189	H	277	K,C,N
14	D	102	C	190	H	278	H
15	D	103	K,H,D	191	H	279	C
16	C	104	K,H	192	H,C	280	C
17	K	105	H,C,R	193	K	281	K
18	B	106	C	194	K,D	282	D
19	K	107	K,C	195	H	283	C
20	C	108	K,H	196	K	284	C
21	C	109	H	197	K,H,C	285	H,C,B
22	L	110	H	198	I	286	K
23	K,C	111	H,C	199	H	287	K,D
24	C	112	H,D	200	O	288	K
25	K,H	113	K,N	201	H	289	K
26	C,B	114	K	202	H	290	H
27	C	115	H	203	N	291	H
28	K,H,C,D ,B	116	C	204	K	292	K,H,C
29	R	117	N	205	C	293	K,H,C
30	K,C	118	C	206	H	294	K
31	H	119	C,D	207	C	295	H
32	H	120	H	208	L	296	B
33	H	121	H,D	209	H	297	K,H
34	D	122	K	210	H	298	H,C
35	K,H,D	123	K	211	K,C,O	299	H
36	R	124	R	212	H,B	300	K
37	K,H	125	K	213	K	301	C
38	C	126	K	214	R	302	K
39	C	127	K,R	215	K,D	303	H
40	C	128	K,O	216	H	304	K,N
41	K,C	129	D	217	K	305	D
42	O	130	H	218	K	306	K
43	C,R	131	H,C	219	K,C	307	H
44	C	132	H,R	220	K,H	308	H,D
45	K	133	K,C	221	C,O	309	K
46	K,C	134	C	222	C	310	K,H
47	K,C	135	N	223	K,H,C	311	K,H,C,B
48	K	136	B	224	K	312	H,C
49	O	137	K	225	K,O	313	K,C
50	C	138	K	226	C	314	K
51	K	139	H,D	227	K,H,L	315	K,H,D
52	H	140	K,R	228	K	316	H,D
53	K	141	K	229	B	317	K,H,B
54	K	142	K,H,C	230	K,H	318	D
55	K,C	143	H	231	K,C	319	K,C
56	O	144	K	232	K	320	H,C
57	K	145	H,C	233	K	321	K,C
58	C,D	146	C	234	C	322	H,D
59	C	147	K	235	K	323	R
60	C	148	K	236	C	324	H,D
61	C	149	K	237	K,D	325	K,H,D
62	L	150	K	238	H	326	H,D
63	H	151	O	239	H,C	327	C
64	K	152	K	240	H,C	328	C
65	K,H,R	153	K,H	241	H	329	K,H,D,C
66	K,H,B	154	H	242	C	330	H
67	K	155	K	243	C	331	K
68	H	156	K,C	244	C	332	K
69	C	157	H	245	C	333	K,H
70	K	158	H,C,D	246	H,C	334	C,O
71	K,H	159	H	247	H,C	335	C
72	H	160	H,D	248	C	336	K,C
73	K,C	161	C	249	R	337	K,B
74	C,O	162	K,D	250	C	338	K
75	K	163	H,C	251	C	339	C
76	K	164	H	252	H	340	K,H,C,D,O
77	K	165	C	253	C	341	K,C
78	C	166	H,D	254	H,C	342	C
79	K	167	H	255	R	343	H,D
80	C	168	H	256	R	344	K
81	K,H,C,D,B	169	K,H	257	C	345	K,H,D
82	C	170	C,D	258	R	346	H
83	K	171	K	259	L	347	K,H
84	C	172	K,H	260	C	348	R
85	K,H,C	173	H	261	C	349	C
86	C	174	K,H	262	C,L		
87	H	175	H	263	C		
88	C	176	H	264	C		

TABLE V. Reading 2-Itemset from the Dataset (C2)

Patterns	Supp_Count		Patterns	Supp_Count	
{B, C}	5	**deleted**	{D, L}	0	**deleted**
{B, D}	2	**deleted**	{D, O}	1	**deleted**
{B, K}	6	**deleted**	{D, R}	2	**deleted**
{B, L}	0	**deleted**	{D, H}	28	
{B, O}	0	**deleted**	{K, L}	1	**deleted**
{B, R}	0	**deleted**	{K, O}	4	**deleted**
{B, H}	7		{K, R}	5	**deleted**
{C, D}	9		{K, H}	38	
{C, K}	30		{L, O}	0	**deleted**
{C, L}	1	**deleted**	{L, R}	0	**deleted**
{C, O}	5	**deleted**	{L, H}	1	**deleted**
{C, R}	4	**deleted**	{O, R}	0	**deleted**
{C, H}	31		{O, H}	1	**deleted**
{D, K}	15		{R, H}	6	**deleted**

TABLE VII. Reading 3-Itemset (C3)

Patterns	Supp_Count	
{B, C, D}	2	deleted
{B, C, K}	2	deleted
{B, C, H}	4	deleted
{B, D, K}	2	deleted
{B, D, H}	2	deleted
{B, K, H}	4	deleted
{C, D, K}	5	deleted
{C, D, H}	6	deleted
{C, K, H}	13	
{D, K, H}	10	

TABLE VIII. Filtering the Results for 3-Itemset (L3)

Patterns	Supp_Count
{C, K, H}	13
{D, K, H}	10

V. Using Weka Program

Weka is an acronym for Waikato Environment for Knowledge Analysis and is developed by the University of Waikato in New Zealand [34]. Weka is an open-source software that contains a set of algorithms and graphics for data analysis and predictive modeling, which is easy to access and use with a graphical interface. The third version of WEKA was fully developed through the Java language in 1997. It is used in applying several tools and is freely available according to the GNU General License. The program can run on any operating system because it is implemented on the Java platform to data pre-processing, classification, regression, clustering, correlation base mining, visualization, and modeling [35], [36].

TABLE IX. The Results of the Filtration Process (L4)

Patterns	Supp_Count
{C, D, K, H}	13

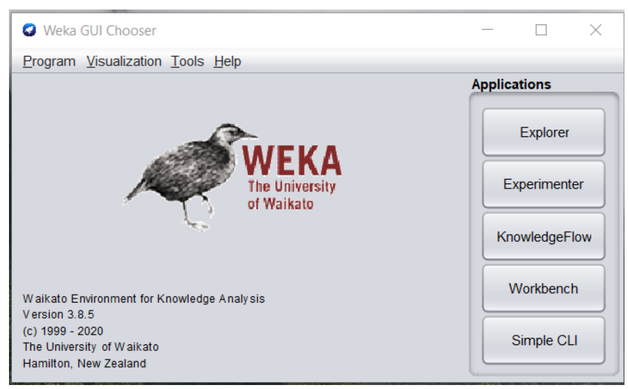

Fig. 10. Weka main interface [31].

VI. Data preparation and implementation Using Weka

To use the Weka program, there is a command that must be modified on the database, a copy of the database (Excel) to convert it to a system (YES - NO) with the extension (.csv) with values separated by commas so that the program read it and deals with it, according to the program settings and apply the ARA algorithm. After opening the file in the program, it is converted and saved with the program extension (.arff) to apply the algorithm and extract the results.

Fig. 10 shows the program's main graphic interface, and in Fig. 12 shows the window for selecting working files and selecting the classification type and algorithm. In Fig. 11, we can see the settings for the algorithm, such as the amount of support and confidence.

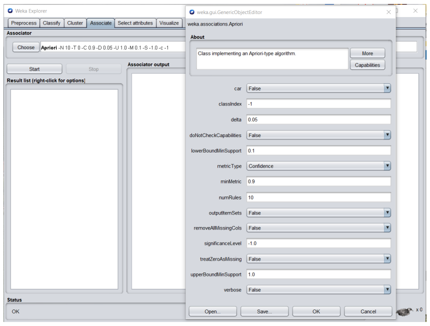

Fig. 11. Apriori algorithm properties settings window for support and confidence.

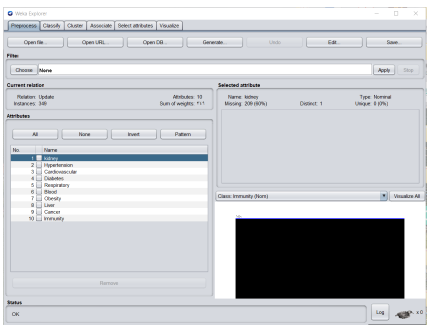

Fig. 12. The window for selecting rules and categories (Associate).

VII. Results and Discussion

A. Manual Results

Through our manual calculations of the ARA algorithm, we explore a set of 4-itemset recurring patterns, which are the most frequent patterns and have a strong correlation base for the probability of appearing next to each other, which are C, D, K, H.

Each pattern has a certain frequency with one or more other patterns explored C, D, K, H, where we find more than one association rule that we read with differences in confidence. According to the rule of Apriori algorithm that states, if an itemset is frequent, then all of its subgroups' items will also be frequent. In other words, any subset of a frequent itemset must also be frequent. Here is a reading of more than one association rule for the set of detected patterns that have high confidence:

$\{K, D\} \Longrightarrow \{H\}$ — $\{D\} \Longrightarrow \{K\}$

$\{D\} \Longrightarrow \{H\}$ — $\{K, H\} \Longrightarrow \{C\}$

$\{H, C\} \Longrightarrow \{K\}$ — $\{H\} \Longrightarrow \{K\}$

$\{H, D\} \Longrightarrow \{K\}$ — $\{K\} \Longrightarrow \{H\}$

$\{K, C\} \Longrightarrow \{H\}$ — $\{C\} \Longrightarrow \{K\}$

Table X shows the number of appeared of the most frequent patterns (support) in the 349 records of the database, where we counted each pattern individually. Then, every two patterns were counted together, counted the number of recurrences for every three patterns, and finally, every four patterns were computed together in the database.

TABLE X. Support Computations Frequency for All Most Frequent Patterns in Database

Patterns	Appearance (frequency)	Patterns	Appearance (frequency)
{C}	130	{D, K}	15
{D}	43	{D, H}	28
{K}	140	{K, H}	39
{H}	131	{C, K, H}	13
{C, D}	9	{D, K, H}	10
{C, K}	34	{C, D, K, H}	13
{C, H}	31	—	—

TABLE XI. Computation of Support for the Most Frequent Patterns

Association Rule	Compute of Support
{Kidney, Diabetes} ⟹ {Hypertension}	146/349 =0.418 = (41.8%)
{Diabetes} ⟹ {Hypertension}	174/349 =0.498 = (49.8%)
{Hypertension, Cardiovascular} ⟹ {Kidney}	171/349 =0.489 = (48.9%)
{Hypertension, Diabetes} ⟹ {Kidney}	168/349 =0.532 = (53.2%)
{Kidney, Cardiovascular} ⟹ {Hypertension}	165/349 =0.472 = (47.2%)
{Diabetes} ⟹ {kidney}	183/349 =0.524 = (52.4%)
{Kidney, Hypertension} ⟹ {Cardiovascular}	169/349 =0.484 = (48.4%)

Before we compute the degree of confidence for the results, we need to compute the support for all groups of itemsets frequent in the results according to the predefined confidence rule.

Table XI shows the computation of the support for all the patterns explored by applying the ARA algorithm, which appeared as the highest results with a strong association rule. In Table XII, we computed the confidence for these results and verified them according to the confidence rule of the ARA algorithm. The results showed a strong association rule between the patterns.

Table XII above reveals that we established seven clear rules and discovered that two of the four patterns were the most frequent among patients who died from the Covid-19 virus, as the confidence rate exceeded 60%:

- {kidney, Diabetes} ⟹ {Hypertension} (66.6%).
- {Diabetes} ⟹ {Hypertension} (65.1%).

TABLE XII. Computation of Confidence for the Most Frequent Patterns

Association Rule	Compute of Confidence
{Kidney, Diabetes} $\Longrightarrow$ {Hypertension}	10/15 =0.651 = (66.6%)
{Diabetes} $\Longrightarrow$ {Hypertension}	28/43 =0.651 = (65.1%)
{Hypertension, Cardiovascular} $\Longrightarrow$ {Kidney}	13/31 =0.419 = (41.9%)
{Hypertension, Diabetes} $\Longrightarrow$ {Kidney}	10/29 =0.357 = (35.7%)
{Kidney, Cardiovascular} $\Longrightarrow$ {Hypertension}	13/34 =0.382 = (38.2%)
{Diabetes} $\Longrightarrow$ {kidney}	15/43 =0.348 = (34.8%)
{Kidney, Hypertension} $\Longrightarrow$ {Cardiovascular}	13/39 =0.333 = (33.3%)

Through the previous two patterns with a high degree of confidence, we noted that hypertension and diabetes were common and present in both patterns. In the first reading of these results among the most frequent patterns, we found that the deceased patients who had chronic diseases such as kidney disease and diabetes also had hypertension with a high confidence level of 66.6%. We also found that the deceased patients who had diabetes also had hypertension with a high confidence level of 65.1%.

This confirms that chronic diseases have significantly impacted the incidence of death among those infected with COVID-19, especially when we found that there were deceased patients who had more than one chronic disease or more. In this case, we found that there was a strong correlation between the diseases of the kidney and diabetes and the emergence of hypertension with a confidence degree of 66.6%.

In the second degree, diabetes led to the emergence of hypertension disease with a confidence level of 65.1%. In other words, the deceased patients who were suffering from kidney disease and diabetes together were also likely to have high blood pressure. In degrees less than 50% confidence, we found that the emergence of hypertension and cardiovascular diseases led to the emergence of kidney disease.

B. Weka Results

After the database has been converted and modified to suit the settings of the Weka program. The program was implemented through the (Associate) tab, and some modifications were made to the graphical algorithm options by opening the algorithm options window (Apriori) to implement and show the algorithm results in the program in Fig. 13, The following is a display of the top three results:

- kidney=**K** Diabetes=**D** $\Longrightarrow$ Hypertension=**H** (0.67).
- Diabetes=**D** $\Longrightarrow$ Hypertension=**H** (0.65).
- Hypertension=**H** Cardiovascular=**C** $\Longrightarrow$ kidney=**K** (0.42).
- kidney=**K** Cardiovascular=**C** $\Longrightarrow$ Hypertension=**H** (0.38).
- Hypertension=**H** Diabetes=**D** $\Longrightarrow$ kidney=**K** (0.36).
- Diabetes=**D** $\Longrightarrow$ kidney=**K** (0.35).
- kidney=**K** Hypertension=**H** $\Longrightarrow$ Cardiovascular=**C** ((0.33).

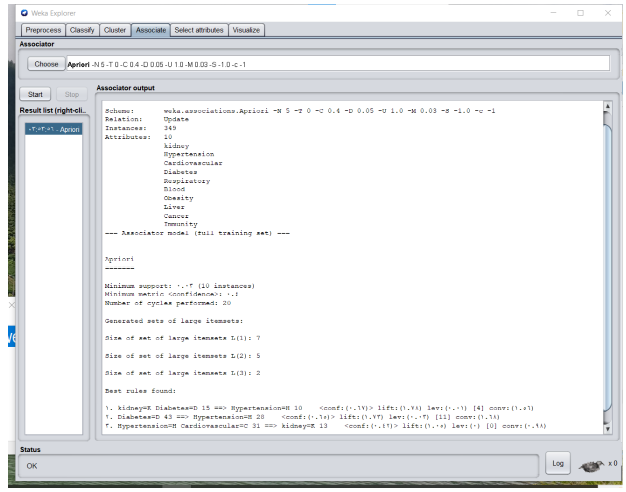

Fig. 13. Program outputs of Weka.

It is noted in the Weka results that it rounds the results to the nearest number, and thus there is a very slight difference in some numbers. They indicate that the pattern set C, D, H, K is the most frequent pattern in the database of COVID-19 deaths. They are cardiovascular - disease - diabetes - hypertension - kidney disease

The occurrences of each of the two patterns were binarily repeated according to the following numbers:

- Cardiovascular with diabetes 9 times.
- Cardiovascular with kidney disease 30 times.
- Cardiovascular with hypertension 31 times.
- Diabetes with kidney disease 15 times.
- Diabetes with hypertension 28 times.
- Kidney disease with hypertension 38 times.

We conclude that those who had chronic diseases (kidney disease and diabetes) often had a hypertension with a high confidence level of 67%. This is if we know that the number of occurrences of kidney diseases and diabetes together in the database of the study is 15 times and that of hypertension is 10 times next to (kidney diseases and diabetes).

Also, those who had diabetes often had hypertension with a high confidence degree of 65%. Where diseases appeared and repeated 43 times, hypertension appeared 131 times, and the two diseases together diabetes appeared next to hypertension in the database 29 times. There were who carried hypertension and cardiovascular diseases and often had kidney disease, but this confidence level was less than 50%.

VIII. Conclusion

Discovering death patterns contributes to enhancing the capabilities of healthcare decision-makers in order to know the most frequent and prevalent chronic diseases among the deceased from COVID-19. In this research, we proposed to use the ARA algorithm to detect death patterns from the data of deceased COVID-19 patients. A non- clinical COVID-19 database, consisting of COVID-19 death patterns was presented and analyzed. The results of manual calculations and our experimental results showed strong association rules with high confidence scores between hypertension, diabetes, cardiovascular, and kidney disease. The results are confirmed by a study of Al Mutair et al. [20] that there is an increase in deaths due to high blood pressure and diabetes. The results are also confirmed by the study of Geng et al. [18], which indicated that cardiovascular diseases and kidney diseases were frequent among the deaths from COVID-19 patients at a high rate. The study of Wang et al. [19] matches with our results and stated that there is a link between high blood pressure and heart disease among patients who died of COVID-19.

Studying the death patterns of people with COVID-19 who have chronic diseases contributes to the medical knowledge in the medical domain, as it helps to understand the most factors that affect the severity of the disease to identify patients who need more special care. Understanding these patterns helps to develop better strategies to control the epidemic by giving more attention and care and priority care to COVID-19 patients. However, this study is only limited to a small number of deceased COVID-19 patients due to challenges of obtaining their medical data from one medical institution. However, in the future, more data will be collected from different medical institutions. Also, this study can be extended to include the analysis of environmental genetic factors that may affect the risk of death with COVID-19.

REFERENCES

[1] WHO. Coronavirus disease (covid-19) reports (2020). [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[2] T. Alafif, A. M. Tehame, S. Bajaba, A. Barnawi, and S. Zia, “Machine and deep learning towards covid-19 diagnosis and treatment: survey, challenges, and future directions,” *International journal of environmental research and public health*, vol. 18, no. 3, p. 1117, 2021.

[3] MOH. Ministry of health, saudi moh reports first case of infection 2021. [Online]. Available: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-03-02-002.aspx

[4] S. Yezli and A. Khan, “Covid-19 social distancing in the kingdom of saudi arabia: Bold measures in the face of political, economic, social and religious challenges,” *Travel medicine and infectious disease*, vol. 37, p. 101692, 2020.

[5] A. A. Alahmari, A. A. Khan, A. Elganainy, E. L. Almohammadi, A. M. Hakawi, A. M. Assiri, and H. A. Jokhdar, “Epidemiological and clinical features of covid-19 patients in saudi arabia,” *Journal of Infection and Public Health*, vol. 14, no. 4, pp. 437–443, 2021.

[6] A. M. Al-Khani, M. A. Khalifa, A. Almazrou, and N. Saquib, “The sars-cov-2 pandemic course in saudi arabia: A dynamic epidemiological model,” *Infectious Disease Modelling*, vol. 5, pp. 766–771, 2020.

[7] OWD. Covid-19 data explorer.” our world in data. [Online]. Available: https://ourworldindata.org/coronavirus/country/saudi-arabia

[8] J. A. Al-Tawfiq and Z. A. Memish, “Covid-19 in the eastern mediterranean region and saudi arabia: prevention and therapeutic strategies,” *International Journal of Antimicrobial Agents*, vol. 55, no. 5, p. 105968, 2020.

[9] S. Ebrahim and Z. Memish, “Saudi arabia’s measures to curb the covid-19 outbreak: temporary suspension of the umrah pilgrimage, 2020,” *J Travel Med*.

[10] M. Hassounah, H. Raheel, and M. Alhefzi, “Digital response during the covid-19 pandemic in saudi arabia,” *Journal of medical Internet research*, vol. 22, no. 9, p. e19338, 2020.

[11] M. Barry, M. Al Amri, and Z. A. Memish, “Covid-19 in the shadows of mers-cov in the kingdom of saudi arabia,” *Journal of epidemiology and global health*, vol. 10, no. 1, p. 1, 2020.

[12] MOH. Ministry of health saudi moh records 4,211 new covid-19 cases and 5,162 new recoveries, [cited 31 january 2022]. [Online]. Available: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2022-01-31-006.aspx.

[13] MoH. Kingdom of saudi arabia ministry of health. covid19 command and control center ccc. the national health emergency operation center december,2022. [Online]. Available: https://COVID19.moh.gov.sa

[14] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, 1993, pp. 207–216.

[15] K. E. Heraguemi, N. Kamel, and H. Drias, “Association rule mining based on bat algorithm,” in *Bio-Inspired Computing-Theories and Applications*. Springer, 2014, pp. 182–186.

[16] T. Alafif, A. Etaiwi, Y. Hawsawi, A. Alrefaei, A. Albassam, and H. Althobaiti, “Discovid: discovering patterns of covid-19 infection from recovered patients: a case study in saudi arabia,” *International Journal of Information Technology*, vol. 14, no. 6, pp. 2825–2838, 2022.

[17] A. A. Robert, A. Al Saeed, and M. A. Al Dawish, “Covid-19 among people with diabetes mellitus in saudi arabia: Current situation and new perspectives,” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 15, no. 5, p. 102231, 2021.

[18] J. Geng, X. Yu, H. Bao, Z. Feng, X. Yuan, J. Zhang, X. Chen, Y. Chen, C. Li, and H. Yu, “Chronic diseases as a predictor for severity and mortality of covid-19: a systematic review with cumulative meta-analysis,” *Frontiers in Medicine*, p. 1442, 2021.

[19] X. Wang, X. Fang, Z. Cai, X. Wu, X. Gao, J. Min, and F. Wang, “Comorbid chronic diseases and acute organ injuries are strongly correlated with disease severity and mortality among covid-19 patients: a systemic review and meta-analysis,” *Research*, vol. 2020, 2020.

[20] A. Al Mutair, A. Al Mutairi, S. Alhumaid, S. Maaz Abdullah, A. R. Zia Zaidi, A. A. Rabaan, and A. Al-Omari, “Examining and investigating the impact of demographic characteristics and chronic diseases on mortality of covid-19: Retrospective study,” *PloS one*, vol. 16, no. 9, p. e0257131, 2021.

[21] T. Alafif, R. Alotaibi, A. Albassam, and A. Almudhayyani, “On the prediction of isolation, release, and decease states for covid-19 patients: A case study in south korea,” *ISA transactions*, vol. 124, pp. 191–196, 2022.

[22] T. Alafif, A. M. Tehame, S. Bajaba, A. Barnawi, and S. Zia, “Machine and deep learning towards covid-19 diagnosis and treatment: survey, challenges, and future directions,” *International journal of environmental research and public health*, vol. 18, no. 3, p. 1117, 2021.

[23] H. Wu, Z. Lu, L. Pan, R. Xu, and W. Jiang, “An improved apriori-based algorithm for association rules mining,” in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 2, 2009, pp. 51–55.

[24] F. H. AL-Zawaidah, Y. H. Jbara, and A. Marwan, “An improved algorithm for mining association rules in large databases,” *World of Computer science and information technology journal*, vol. 1, no. 7, pp. 311–316, 2011.

[25] M. Kaur, U. Garg, and S. Kaur, “Advanced eclat algorithm for frequent itemsets generation,” *International Journal of Applied Engineering Research*, vol. 10, no. 9, pp. 23 263–23 279, 2015.

[26] R. Zhong and H. Wang, “Research of commonly used association rules mining algorithm in data mining,” in *2011 International Conference on Internet Computing and Information Services*. IEEE, 2011, pp. 219–222.

[27] S. Rao and P. Gupta, “Implementing improved algorithm over apriori data mining association rule algorithm 1,” 2012.

[28] J. Singh, H. Ram, and D. J. Sodhi, “Improving efficiency of apriori algorithm using transaction reduction,” *International Journal of Scientific and Research Publications*, vol. 3, no. 1, pp. 1–4, 2013.

[29] J. Yabing, “Research of an improved apriori algorithm in data mining association rules,” *International Journal of Computer and Communication Engineering*, vol. 2, no. 1, p. 25, 2013.

[30] T. A. Kumbhare and S. V. Chobe, "An overview of association rule mining algorithms," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 927–930, 2014.

[31] K. Vanitha, "Using hash based apriori algorithm to reduce the candidate 2-itemsets for mining association rule," *Journal of Global Research in computer science*, vol. 2, no. 5, pp. 78–80, 2011.

[32] K. Dahdouh, A. Dakkak, L. Oughdir, and A. Ibriz, "Association rules mining method of big data for e-learning recommendation engine," in *International Conference on Advanced Intelligent Systems for Sustainable Development*. Springer, 2018, pp. 477–491.

[33] C. Li, N. Ding, G. Zhang, and L. Li, "Association analysis of serial cases based on apriori algorithm," pp. 137–140, 2019.

[34] Waikato. University of waikato, new zealand, web site:. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/index.html

[35] I. H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques with java implementations," *Acm Sigmod Record*, vol. 31, no. 1, pp. 76–77, 2002.

[36] G. Holmes, A. Donkin, and I. H. Witten, "Weka: A machine learning workbench," in *Proceedings of ANZIIS'94-Australian New Zealnd Intelligent Information Systems Conference*. IEEE, 1994, pp. 357–361.