
Investigating New Patterns in Symptoms of COVID-19 Patients by Association Rule Mining (ARM)

Anju Singh¹, Divakar Singh², Kamal Upreti^{3,*}, Vaibhav Sharma⁴,
Bhawani Singh Rathore² and Jagdish Raikwal⁵

¹*Sage University, Bhopal, Madhya Pradesh, India*

²*UIT, Barkatullah University, Bhopal, Madhya Pradesh, India*

³*Dr Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India*

⁴*SGRR School of CA &IT, SGRR University Dehradun, India*

⁵*Institute of Engineering and Technology, Devi Ahilya University, Indore, India*

E-mail: singhanju0123@gmail.com; dsingh0123@gmail.com;

kamalupreti1989@gmail.com; vsdeveloper10@gmail.com;

rathore.charles@gmail.com; jraikwal@ietdavv.edu.in

**Corresponding Author*

Received 22 February 2022; Accepted 24 April 2022;

Publication 25 August 2022

Abstract

Background: COVID-19 is a major public health emergency wreaking havoc on public health, happiness, and liberty of travel, as well as the worldwide economy. Scientists from all over the world are working to develop treatments and vaccines; the WHO has given emergency approval to eight vaccines from around the world. However, it is also seen that the efficiency of vaccines is not up to the mark in different age groups. COVID-19 symptoms come in many different shapes and sizes, so it's important to learn about them as soon as possible so that medical attention and management can be easier.

Method: The GitHub Data Repository-made COVID-19 patient data is available on the internet, which is used in this investigation. We have used the

Journal of Mobile Multimedia, Vol. 19_1, 1–28.

doi: 10.13052/jmm1550-4646.1911

© 2022 River Publishers

association rule mining method to look for common patterns in a targeted class or segment and then look at the symptoms based on them.

Result: The result is that this study involves individuals with a median age of 52 years old. Few frequent symptoms like respiratory failure (1%), septic shock (1.4%), respiratory distress syndrome (1.8%), diarrhoea (1.8%), nausea (2%), sputum (3%), headache (5%), sore throat (8%), pneumonia (8%), weakness (7%), malaise/body pain (11%), cough (37%), fever (67%) and remaining diseases like myocardial infarction, cardiac failure, and renal illness (less than 1%) were present. If a patient had chronic disease, respiratory failure, and pneumonia, there was a higher risk of death; if a patient had a combination of chronic disease, respiratory failure, and pneumonia, respiratory failure in the age range of 45 to 84 years there was a higher risk of death. Patients having chronic conditions like pneumonia or renal disease symptoms that died as a result of the corona virus had more serious indication patterns than those without chronic diseases.

Keywords: COVID-19, hotspot, segments, association rule mining (ARM), customer behavior analysis, market basket analysis, pattern, world health organization (WHO).

1 Introduction

The COVID-19 pandemic is a world disaster in terms of health [1]. Globally, 26,54,23,384 cases and 5248,669 deaths had been registered as of December 07, 2021, as per the WHO report. Some European countries have “flattened” the curve; others, such as the United States, Brazil, India, and Russia, are still battling it [2]. In October 2020, the UK public was affected by the delta variant of COVID-19, and in December 2020, the delta variant was officially identified as per the WHO report. The Gamma variant, or lineage P.1, was detected in Tokyo, Japan on January 6, 2021 by the National Institute of Infectious Diseases (NIID). It has been labelled as a gamma variant by WHO. Meanwhile, a second wave has hit the United Kingdom, Germany, Spain, Poland, and Japan. Scientists from all over the world are hard at work developing new treatments and vaccinations. The first findings of the RECOVERY research revealed that steroids were effective in COVID-19 patients on breathing assistance who were hospitalised [3].

According to the SOLIDARITY trial conducted by the World Health Organization, early pandemic drugs such as hydroxyl chloroquine and others

are not able to stop mortality for hospitalised patients [4]. Simultaneously, remdesivir showed potential efficacy [5]. Even though these studies had methodological flaws and were designed before scientists had a strong grasp of illness progression, they were nonetheless successful [6]. In the starting face of COVID-19, 3 medicines are approved by the WHO for the treatment of patients [7]. In the USA and India, remdesivir is used; in the United Kingdom and northern countries, dexamethasone is used; some countries treat patients with failover. The effectiveness of these drugs in individuals with obvious contraindications, such as high diabetes, psychosis, underlying malignancy, immune suppression, or situations where steroids may have a very adverse effect, is yet unknown [8, 9]. In the same way, vaccine development has made significant advances. Two of the vaccines, Pfizer and BioNTech's BNT162b2 and Moderna's mRNA-1273, were given emergency approval for use in the United States, Canada, and a few other countries because they showed 95 percent efficacy [10]. As of December 19, 2020, 1.6 million individuals in 4 nations (the USA, China, the UK, and Russia) have received their first doses of COVID-19 shots, according to Bloomberg [11].

However, concerns concerning the vaccine's safety profile in certain groups, like the elderly and individuals with constant comorbidities, have remained unanswered. Moreover, it is unclear whether antibody makers will actually want to fulfil a need when the whole total populace will be vaccinated and secured against COVID-19. According to the Center for Infectious Disease Research and Policy [12], roughly 60–70 percent of the human population should be safe for the COVID-19 pandemic to end. As a result, they estimate that the pandemic will persist for at least another 18–24 months, with hot spots resurfacing in various geographic places on a recurrent basis, assuming a few levels of ongoing reduction measures [12]. Based on this knowledge, we can easily deduce the requirement for taking proper health precautions, like screening all members of the public who have those symptoms, selecting people for testing, and providing hospital facilities and quarantine if needed. COVID-19 containment and symptomatic management are dependent on these approaches. COVID-19 has been linked to various types of symptoms, from the common cold to chronic problems [13]. Identification of those symptom patterns aids physicians and care assistants in providing effective supportive and therapeutic treatment by allowing them to make better clinical decisions.

Much research has arisen identifying associated clinical illness features, comorbidities, and epidemiological drivers because of the remarkable growth

in global COVID-19 cases [14, 15]. On the other hand, modelling research on COVID-19 that addresses relationships between diverse illness factors is insufficient. Current figuring capacities have made organised information extraction and mining conceivable, permitting us to carry out an assortment of information-related assignments like consecutive information characterization, grouping, summing up, and similitude investigation, which can be utilised to build up a connection between various clinical boundaries and foresee likely results [16].

The COVID-19 breakout has posed a considerable challenge to clinicians and public health officials. In our review, we used an association rule mining technique to filter out symptoms of COVID-19 indications. These manifestation design mining tools can be used in conjunction with other techniques to accurately identify disease patterns in clinical settings. Man-made reasoning (AI) has enormous potential in medication. Organizations like Alibaba constructed AI answers to assist China with fighting COVID-19 and anticipate the episode's pinnacle, degree, and length, with high execution precision in certifiable tests the nation over [17]. An AI-based CT picture examination can possibly precisely analyse COVID-19 patients and resolve an assortment of respiratory infections [18]. There are a lot of different AI and man-made reasoning innovations that could speed up the development of COVID-19 vaccines [19, 20]. These include genomic sequence analysis and molecular docking.

In November 2021, variant B.1.1.529 (Omicron) was discovered in Botswana and South Africa. As of December 1, there were 365 cases reported in many countries, including the United States. The new variety is being studied by public health officials, but it is unclear whether it poses a bigger threat than existing COVID-19 forms. Because of the new variation, the United States has issued a travel ban to South Africa and seven neighbouring countries. Many other countries have enacted travel bans as well.

2 Background

Although there has been some computational study on COVID-19, the majority of predictions have relied on complicated techniques (like the NN-technique) [21, 22]. Techniques that are simple and easy to understand are underrated. Straightforward association rules can be utilised to identify each example in an informational index, which is valuable for clinical information investigation. It also enables experts to make well-informed decisions, gather critical data, and create basic data bases in a timely and efficient manner. This

examination identifies side effects in COVID-19 patients and separates them by age, sex, persistent illness, and mortality.

3 Literature Review

Machine learning algorithms for prediction and knowledge discovery have lately been prominent in biomedical research [23]. Genomic examination [24], infection quality investigation [25], mortality prediction [26], customised medication [27], drug revelation [28], and forecast of unfavourable medication events [29], patient likeness [30], and reasonable AI techniques in medication [31] are all examples of how AI is being used in biomedicine. The mining of association rules is one use of machine learning in medicine (ARM). R. Agrawal was the first to suggest ARM [32]. It was first used to analyse sales data, with the goal of identifying relevant patterns that may define the occurrence of an event, considering the different items in each “set of transactions.” ARM’s basic concept is a brute-force technique. This method lists all potential rules first, then prunes those that do not satisfy the supplied condition. However, due to the many possible combinations, this strategy is computationally prohibitive. R. Agrawal [32] devised the Apriori approach to decrease the number of choices. The Apriori approach has two significant flaws. First, it generates a huge number of candidate item sets in a larger data set while also creating frequent itemsets. Second, it necessitates several database scans, resulting in increased computing expenses. Han et al. suggested the Frequent Pattern Growth technique to solve these limitations (FP growth) [33]. This is a review assessment of COVID-19 patients’ results provided by Meera Tandan et al. [34, 41].

The major side effects can be classified into two categories according to the changes in their frequencies. The primary symptoms include fever, des-titute craving, cough, expectoration, shortness of breath, chills, pharyngeal distress, and myalgia. Their frequencies are the most elevated right now since their onset and have diminished steadily afterwards. The moment incorpo-rates weakness, queasiness, vomiting, diarrhea, and stomach distress. Their frequencies have increased since the onset of COVID-19 and then decreased after coming to top values. To begin with, the category of indications has a place for the systemic and nearby tissue harm and provocative reactions associated with side effects [41], primarily systemic and respiratory side effects. The moment category of symptoms has a place in respiratory and stomach-related organ dysfunction indications such as queasiness, heaving, diarrhea, and stomach inconvenience. The information allows the illness

related to these manifestations to be inside the side effect variable for a composite yet clinically identifiable symptom pattern. For example, the term ‘pneumonia’ is used for symptom factors for patients who have a cluster of clinical symptoms that are compatible with a chest infection. The frequency of SARS-CoV-2 disease is seen normally in adult male patients whose age is 34 and above. Additionally, SARS-CoV-2 is more likely to contaminate individuals with persistent co-morbidities such as heart-related illness and diabetes. In a ponder of 425 COVID-19 patients in Wuhan, there were no cases in the children’s age group up to 15. In any case, 28 child patients have been detailed as of January 2020. The clinical features of contaminated paediatric patients change, but most have mild symptoms with no fever or pneumonia and have a great prognosis. Another ponder found that although a child had radiological ground-glass lung opacities, the understanding was asymptomatic. In summary, children can be less likely to be contaminated or, in the case of being infected, present milder signs than grown-ups; subsequently, it is possible that their guardians will not look for treatment, leading to underestimates of the COVID-19 rate in this age group [40].

The ARM calculation contains side-effect exchanges plans to develop continuous thing sets, having somewhere around a client-determined edge. In this way, they followed a similar methodology to Sultana N. Nahar et al. [36] by fixing a “certainty” edge of 90 percent. This was because the “assurance” metric is used to rank the rules [37]. For doing this, it is essential to set up an edge worth at least 0.001 and a “lift” more prominent than 1.0 for decidedly connected principles. In this case, the idea was to filter the best 10 guidelines with the most elevated help scores. To catch uncommon or inconsistent things, they picked low help and high assurance measures. This idea is obtained from the survey by McCormick et al. [38] in mining clinical indications. When an indication that infrequently happens is unequivocally connected with another uncommon manifestation, it is fundamental to not bar the guidelines describing these side effects. Such principles give clinicians significant understanding of a clever sickness like COVID-19. In different areas, like business, the limit with low help and high certainty will create many standards that may not be fascinating for client examination. So low help and high certainty accumulate few principles. However, the results can be of mind-blowing interest to clinicians, as they could explain lesser-known peculiarities [39]. It is generally expected to be valid in clinical determination where numerous indication blends will just show up in a few patient cases. Thus, a method like this one for looking for examples and rules that are similar will help us look more closely at manifestation revelation.

In this paper, we used the HotSpot Algorithm. A hotspot is a calculation that can straightforwardly mine affiliation rules from genuine information. The HotSpot calculation chosen in this paper can specifically mine affiliation rules and powerfully obtain the extent of genuine number of interims without the discretization of genuine information, hence maintaining a strategic distance from the impact of subjective components in discretization. The handling speed is amazingly quick. But like the Apriori calculation, the HotSpot algorithm has one self-evident deficiency: support selection ought to be set misleadingly based on involvement and cannot be set precisely concordant with the real scale of the issue. On the off chance that the support threshold is set too low, a lumbering and complex tree structure may be created; on the off chance that the support threshold is set too high, some related intervals may exist within the rare target trait values may be disregarded. In this manner, within the handle of support selection, numerous comparison tests are required to decide the optimal support based on the mining comes about. In order to overcome the affectability of the HotSpot algorithm in bolster edge settings and improve the quality of ARM [42], Apriori algorithms can handle only binary values. In our work, any numeric data can be taken as input and we can get results that also find rules for targeted attributes and for targeted class values. In our work, we can find rules for highly co-related data that only obtain a lift value greater than 1. In the base paper, only a few attributes are considered for the conclusion, but in this paper, we have considered 51 attributes to find the good outcomes related to COVID-19. Considering Above all, the proposed work provides unique and useful rules that will help doctors and patients during meditation and may take necessary steps for patients.

3.1 Association Rule Mining (ARM)

ARM discovers the pattern of frequently occurring objects or events in the data collection, as well as the relationship between things or events. It is also known as “if→then,” “if” denotes the antecedent and “then” the consequent. The success of generated rules is typically quantified in terms of (i) Support, (ii) Confidence, and (iii) Lift. $Support(X \rightarrow Y) = ARM$ finds the pattern of frequently occurring objects or occurrences in the data set, including the link between items or events. The pattern reveals the combination of things or events that occur concurrently. It is useful in medicine to understand how one condition is linked to another, such as diabetes and hypertension. In the context of medicine, an association rule between symptom (or disease) is

expressed in the form $X \rightarrow Y$, where X and Y are a disjoint set of symptom (or disease), i.e., $X \cap Y = \varphi$. In other words, X is the rule's antecedent, and Y is the rule's consequent. Also, known as "if→then", "if" represents antecedent, and "then" represents consequent. The success of generated rules is usually quantified in terms of (i) Support, (ii) Confidence, and (iii) Lift. Theoretically, support has been defined as follows: **Support ($X \rightarrow Y$) = (Persons having both X and Y)/(Total number of Persons)**. As a result, support dictates how frequently a rule applies to a specific data collection. Confidence is described as **Confidence ($X \rightarrow Y$) = (Persons having both X and Y)/(Persons having X)**, In this case, confidence defines the frequency with which illness (or disease) Y appears in people who have X . Lift can be stated as follows: **Lift($X \rightarrow Y$) = ((Persons having both X and Y)/(Persons having X))/(Fraction of Persons having Y)**. where the fraction of Persons with Y is the number of Persons with Y divided by total of sufferers Lift indicates how frequently symptom Y appears when symptom X appears while controlling the likelihood of occurrence of symptoms Y . The correlation among both X and Y is determined by the value of lift; independent ($=1$), positive associated (>1), and negative related (<1). The disadvantage of the "Confidence" measures is that they may overestimate an organization's value.

4 Research Methodology

4.1 Proposed Framework

In the proposed work, real and categorical data can be taken as input and can find rules for targeted class value. This work can find rules for highly correlated data by considering lift values greater than 1. In previous work [41], only a few attributes were taken to reach the conclusion, but in this work, we have taken 21 attributes as input to reach the conclusion. We sent out and cleaned information in the input of the hotspot algorithm. The "symptom data" was cleaned before being transformed to a "transaction" format and evaluated with the Hotspot method, which is accessible in Java as the association rule mining technique. The dataset is of 1560 patients, each having 36 attributes, but we have extracted only 21 of them, as shown in Figure 1 and Table 1.

In Figure 2, shows the proposed framework of ARM for COVID-19 with the target attribute set to the target value. The parameters taken for this experiment is: total population: 1560 instances. There are different rules with different support values for target attributes and target values: Yes (125

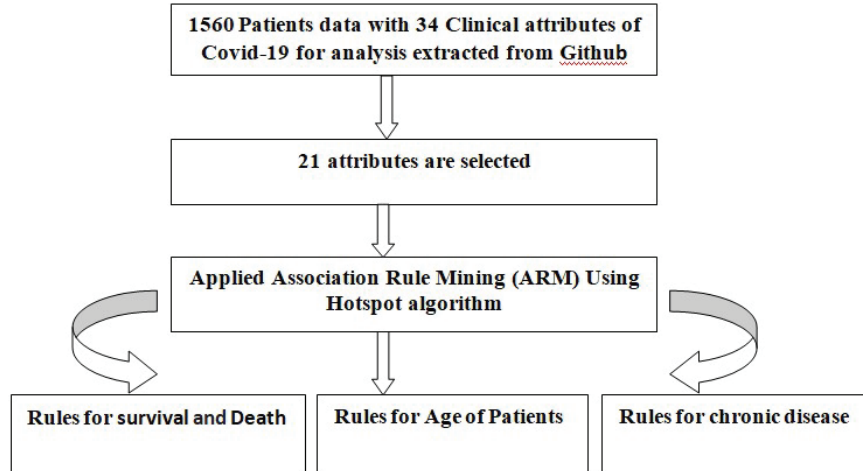


Figure 1 Proposed framework for ARM of COVID-19 patients.

instances (8.01%) in the target population). We used Java programming language to implement a hotspot algorithm. Hotspot is rule-based data mining method to settle the test of consequently identifying novel and significant symptom patterns in COVID-19 information. The measurably critical standards were displayed in an assortment of patient classes, including sex, ongoing illness, mortality, and age. As far as anyone is concerned, this is the main review to extricate the most well-known manifestations of COVID-19 patients utilizing straightforward yet incredible hotspot-based mining. Hotspot learns a bunch of rules (shown in a tree-like design) that augment or limit an objective variable or worth of interest. With an ostensible objective, one should search for sections of the information where there is a high likelihood of minority esteem occurring [35]. By using the Hotspot algorithm, we have found various rules for survival and death, the age of the patient, and chronic disease.

5 Results

5.1 Data Set Attribute Description

As per the data set, 36 attributes with data types such as string, numeric, and categorical (Death values as {0,1}, Sex values as {Female, Male}, Age values as {"20–44 yrs", ">85 yrs", "45–54 yrs", "55–64 yrs", "65–84 yrs", "<=19 yrs"} which are represented as categorical data values as given

Table 1 Covid-19 dataset for analysis

S. No	Attribute Name	Data Type/Categorical Value	Description
1.	serial_no	string	
2.	Age1	numeric	
3.	Age	{“20–44 yrs”, “>85 yrs”, “45–54 yrs”, “55–64 yrs”, “65–84 yrs”, “<=19 yrs”, “20”}	Categorical data
4.	Sex	{Female, Male}	Categorical data
5.	Country	string	
6.	Hospitalization	{0,1}	Categorical data
7.	ChronicDZ_YN	{0,1}	Categorical data
8.	Death	{0,1}	Categorical data
9.	Cough	{0,1}	Categorical data
10.	Fever	{0,1}	Categorical data
11.	Breathing_prob	{0,1}	Categorical data
12.	Pneumonia	{0,1}	Categorical data
13.	RDS	{0,1}	Categorical data
14.	Res_failure	{0,1}	Categorical data
15.	Weakness	{0,1}	Categorical data
16.	Malaise_bodysore	{0,1}	Categorical data
17.	Rhinorhea	{0,1}	Categorical data
18.	Sorethroat	{0,1}	Categorical data
19.	Sputum	{0,1}	Categorical data
20.	Non_respi_sym	{0,1}	Categorical data
21.	Drymouth	{0,1}	Categorical data
22.	Nausea	{0,1}	Categorical data
23.	Anorexia	{0,1}	Categorical data
24.	Diarrhoea	{0,1}	Categorical data
25.	Headache	{0,1}	Categorical data
26.	Conjunctivites	{0,1}	Categorical data
27.	Hypertension	{0,1}	Categorical data
28.	Primarymyelofibrosis	{0,1}	Categorical data
29.	Cardica_arryhmia	{0,1}	Categorical data
30.	Heartfailure	{0,1}	Categorical data
31.	Myocardial_infraction	{0,1}	Categorical data
32.	Renal_dz	{0,1}	Categorical data
33.	SepticShock	{0,1}	Categorical data
34.	Multiple_organ_failure	{0,1}	Categorical data
35.	other_sym	{0,1}	Categorical data
36.	Asymptomatic	{0,1}	Categorical data

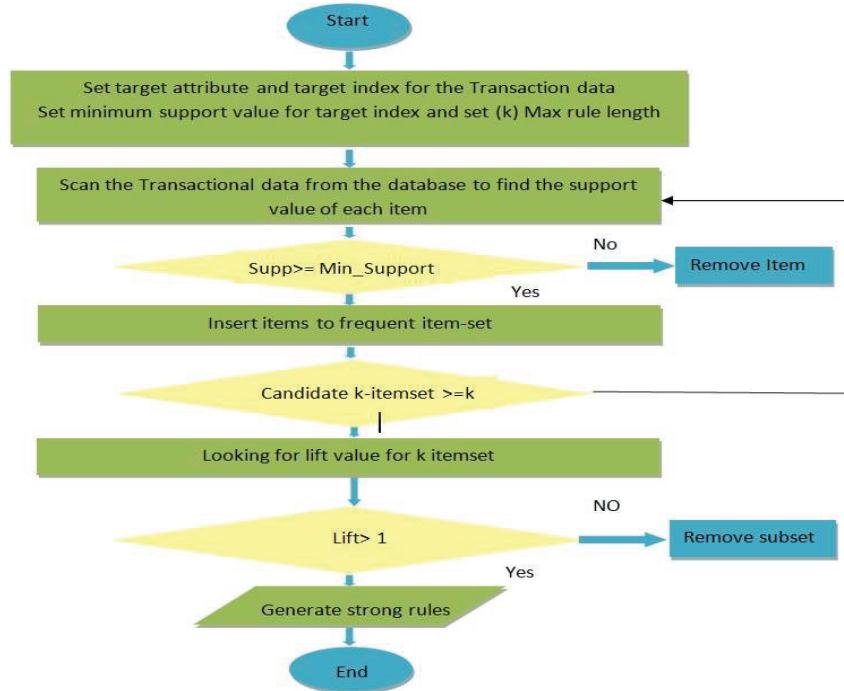


Figure 2 Shows the working steps of the algorithm for class value-based market basket analysis with real/categorical data.

in Table 1). For our experiment, we only used 21 attributes (age, gender, country, hospitalization, ChronicDZ YN, Cough, Fever, Breathing Problem, Pneumonia, RDS, Res failure, Weakness, Malaise, Bodysore, Sputum, Headache, Cardica arryhmia, Heart Failure, Myocardial infraction, Renal dz, SepticShock, Death).

Data were extracted from the GitHub web portal on October 22, 2021, and patients' data up to May 27, 2020 are included in this dataset [34]. The median age of the patients was 52 years, and 57 per cent of them were men. Due to COVID-19, it is claimed that a total of one hundred and twenty-five (125) people lost their lives, accounting for 8% of the total (1560). In this paper, we discussed a new variant that was detected earlier in China than in Singapore in the month of December 2020 as reported in a report published on the internet by the leading newspaper in India. Figure 6 shows that fever was the most common symptom (67 percent), followed by cough (37%), malaise/body aches (11%), and pneumonia (11%). In 1–5%

of patients, migraine, sputum creation, squeamishness, diarrhea, respiratory difficulty problems, and septic shock were completely recorded. Myocardial limited necrosis, dissatisfaction, and renal ailment were found in about 1% of the cases in this study. The recurrence of constant hypertension was 5%, diabetes 4%, and kidney and coronary cardiopathy 1%.

As shown in Figure 6, among those COVID-19 patients whose age lies between 45 and 84, their death count is high as compared to other age groups. Most patients report symptoms of fever and cough during COVID-19, as shown in Figure 6.

In Table 2, significant rules or views for survival ($N = 1560$) are given. A few rules, like if a patient had malaise_bodysore and hospitalisation (antecedent), then the patient had higher confidence of survival (consequent). In the next rule, if a patient had a malaise_bodysore and an age of 85 years (antecedent), then they had a higher chance of survival (consequent). As shown in Figure 7, if the patient had a fever and was over the age of 85, they had a better chance of survival. There are a few significant rules for the death of patients, like if the patient had a Res_failure and an age in the range of 20–44 years (antecedent), they had a higher chance of death (consequent). In another rule, if the patient had a Res_failure and a Myocardial infraction (antecedent), they had a higher chance of death (consequent). RDS, Pneumonia, and respiratory failure were some of the patient's symptoms (antecedent). If they also had any of these symptoms, there was a higher chance of death (consequent). This can be seen in Figure 8, so this was the case.

Table 3 describes the significant rules for males in different age groups. The few rules for survival are that if a patient was male and aged 45–54 years and had RDS (antecedent), then there was a higher chance of survival (consequent). Patients with chronic diseases have very low chances of survival. If the patient's symptoms consisted of chronic disease along with any one of these symptoms (breathing problem, cough, age ranging between 65 and 84 years old), then there was a high chance of death.

Table 4 describes the significant rules for females in different age groups. A few survival rules for female patients: if a patient had any of the following weaknesses: pneumonia, malaise, body sores, fever, or headache, they had a better chance of survival. If females had an age group of 65–84 years and suffered from chronic disease, then they had a very low chance of survival. If the patient's symptoms consisted of chronic disease along with any one of these symptoms (fever, pneumonia, age in the range of 65–84 years), then there was a high chance of death.

Table 2 Significant rule for survival Vs Death with more than one symptom [No. = 1560]

Rule Number	Antecedents	Consequents	Support	Confidence	Lift
{Rules for survived}					
R1	{Malaise_bodysore = 1, Hospitalization}	{Survived}	0.1	1	1.09
R2	{Malaise_bodysore = 1, Age = >85 yrs}	{Survived}	0.1	1	1.09
R3	{Fever = 1, Age = >85 yrs}	{Survived}	0.2	1	1.09
R4	{Fever = 1, Hospitalization = 0}	{Survived}	0.2, 0.3	1	1.09
{Rules for died}					
R1	{Res_failure = 1, Age = 20-44 yrs}	{Died}	0.01	1	12.48
R2	{Res_failure = 1, Myocardial_infraction = 1}	{Died}	0.02	1	12.48
R3	{Septic Shock = 1, Renal_dz = 1}	{Died}	0.02	1	12.48
R4	{Septic Shock = 1, Cardica_arrhythmia = 1}	{Died}	0.02	1	12.48
R5	{RDS = 1, Cardica_arrhythmia = 1}	{Died}	0.02	1	12.48
R6	{RDS = 1, ChronicDZ_YN = 1, sex = Male}	{Died}	0.02	1	12.48
R7	{RDS = 1, ChronicDZ_YN = 1, Age = 65-84 yrs}	{Died}	0.02	1	12.48
R8	{RDS = 1, ChronicDZ_YN = 1, Age = 45-54 yrs}	{Died}	0.02	1	12.48
R9	{RDS = 1, sex = Male, Age = 20-44 yrs}	{Died}	0.03	1	12.48
R10	{Res_failure = 1, Septic Shock = 0, ChronicDZ_YN = 1}	{Died}	0.1	1	12.48
R11	{RDS = 1, Pneumonia = 1, Septic Shock = 0, sex = Male}	{Died}	0.1	1	12.48
R12	{ChronicDZ_YN = 1, Pneumonia = 1, Age = 45-54 yrs}	{Died}	0.1	1	12.48

(Continued)

Table 2 Continued

Rule Number	Antecedents	Consequents	Support	Confidence	Lift
R13	{ChronicDZ_YN = 1, Pneumonia = 1, RDS = 1, sex = Male}	{Died}	0.1	1	12.48
R14	{ChronicDZ_YN = 1, Pneumonia = 1, Age = 55–64 yrs, sex = Male}	{Died}	0.1	1	12.88
R15	{ChronicDZ_YN = 1, Pneumonia = 1, Res_failure = 1}	{Died}	0.1	0.94	11.79
R16	{ChronicDZ_YN = 1, Pneumonia = 1, RDS = 1}	{Died}	0.1	0.91	11.39
R17	{ChronicDZ_YN = 1, Age = 65–84 yrs, Pneumonia = 1}	{Died}	0.1	0.93	11.59
R18	{ChronicDZ_YN = 1, Pneumonia = 1}	{Died}	0.1, 0.3, 0.4	0.88	10.92

Table 5 describes the significant rules for a chronic condition. If patients had no chronic disease and reported fever, then patients had higher chances of survival. If a male patient’s symptoms consisted of chronic disease along with any one of these symptoms (RDS, pneumonia), then the patient had a higher chance of death.

We have taken an example with transactional data having 06 instances in Figure 3, and the rule mining method yielded two rules, with a support score of 0.1 and 0.33, confidence of 1, and lift of 1.2 for the antecedent (X) Age = 18–35 and consequent (Y) Live Birth in rule 1. With a confidence level of one, all “Treatment = DI” patients gave birth to live children. Similarly, lift 1.2 indicates that “treatment = DI” and “live birth” are mutually correlated.

In this work, we took data from 45 countries for analysis [34, 41] and found that China and the Philippines had a very high death count in reported patients as compared to other countries. We also found that the mutation present in both countries in December 2020 was dangerous and deadly as compared to other countries, as shown in our study.

In Figure 9 shows the association rules for the target attribute (consequent): Yes, death with a target value. (i) the rule has 22 instances, the rule length is 4, and the minimum value count for segments is 13, implying

Table 3 Significant rules for age (survived/died for sex = male) (N = 1560)

Rules	Antecedents	Consequents	Support	Confidence	Lift
	{Rules for survived}				
R1	{Hospitalization = 1, Country = China, Death = 0, sex = Male, Fever = 1}	{Age = 20-44 yrs}	0.3	0.61	1.94
R2	{RDS = 1, Pneumonia = 0, sex = Male, Death = 0}	{Age = 45-54 yrs}	0.01	1	9.94
	{Rules for died}				
R1	{Death = 1, Breathing_prob = 1, Pneumonia = 0, sex = Male, ChronicDZ.YN = 1}	{Age = 65-84 yrs}	0.03	1	14.44
R2	{Death = 1, ChronicDZ.YN = 1, Country = China, Cough = 0, sex = Male}	{Age = 65-84 yrs}	0.03	0.83	12.04
R3	{Death = 1, Breathing_prob = 1, Pneumonia = 0, sex = Male, Cough = 1}	{Age = 65-84 yrs}	0.03	1	14.44
R4	{Death = 1, ChronicDZ.YN = 1, Hospitalization = 1, Cough = 0, sex = Male}	{Age = 65-84 yrs}	0.04	0.6	8.67
R5	{Death = 1, RDS = 1, Myocardial_infraction = 0, Renal_dz = 0, sex = Male}	{Age = 45-54 yrs}	0.05	0.38	3.79
R6	{Death = 1, Pneumonia = 1, sex = Male}	{Age = 45-54 yrs}	0.05	0.27	2.68
R7	{Death = 1, Country = Philippines, sex = Male, Pneumonia = 1}	{Age = 45-54 yrs}	0.05	0.27	2.64
R8	{Death = 1, Hospitalization = 1, Breathing_prob = 0, Cough = 1, sex = Male}	{Age = 55-64 yrs}	0.05	0.47	4.95
R9	{Death = 1, Res_failure = 1, Myocardial_infraction = 0, sex = Male, SepticShock = 0}	{Age = 65-84 yrs}	0.06	0.5	7.22

Table 4 Significant rules for age (survived/died. for sex = female) (N = 1560)

Rules	Antecedents	Consequents	Support	Confidence	Lift
	{Rules for survived}				
R1	{Weakness = 1, Country = China, Cough = 0, sex = Female, Death = 0}	{Age = 20-44 yrs}	0.02	0.64	1.65
R2	{Hospitalization = 1, Country = China, Death = 0, Pneumonia = 1, sex = Female}	{Age = 20-44 yrs}	0.02	0.61	2.03
R3	{Weakness = 1, Cough = 0, Fever = 1, sex = Female, Death = 0}	{Age = 20-44 yrs}	0.02	0.5	1.6
R4	{Malaise_bodysore = 1, sex = Female, Death = 0, Cough = 0, Fever = 1}	{Age = >85 yrs}	0.03	0.84	1.58
R5	{Headache = 1, Death = 0, sex = Female}	{Age = >85 yrs}	0.03	0.53	2.98
R6	{sex = Female, Hospitalization = 1, Fever = 1, Death = 0, Weakness = 1}	{Age = 45-54}	0.2	0.28	2.74
	{Rules fordied}				
R1	{Death = 1, ChronicDZ_YN = 1, Fever = 1, sex = Female}	{Age = 65-84 yrs}	0.03	0.57	8.25
R2	{Death = 1, sex = Female, ChronicDZ_YN = 1, Hospitalization = 1, Malaise_bodysore = 0}	{Age = 65-84 yrs}	0.05	0.63	9.03
R3	{Death = 1, sex = Female, ChronicDZ_YN = 1, Septic Shock = 0, Pneumonia = 1}	{Age = 65-84 yrs}	0.05	0.54	7.78
R4	{Death = 1, sex = Female, ChronicDZ_YN = 1}	{Age = 65-84 yrs}	0.08	0.5	7.22

support from 10% of the target population (125). (ii) the number of rules is 15, the rule length is 5, and the minimum value count for segments is 100 (80% of the target population of 125). (iii) the number of rules is ten, the rule length is three, and the minimum value count for segments is 113 (90% of the target population of 125).

Table 5 Significant rules for survived/died vs.age (N = 1560) for chronic condition

Rules	Antecedents	Consequents	Support	Confidence	Lift
R1	{Death = 0, Fever = 1, Cough = 0, Hospitalization = 1} {With chronic disease}	{Without chronic disease}	0.2,0.3	1	1.1
R1	{RDS = 1, Death = 1, Myocardial_infraction = 0, sex = Male, Hospitalization = 0}	{With chronic disease}	0.01	0.89	9.43
R2	{Pneumonia = 1, sex = Male, Death = 1}	{With chronic disease}	0.3	0.87	9.18

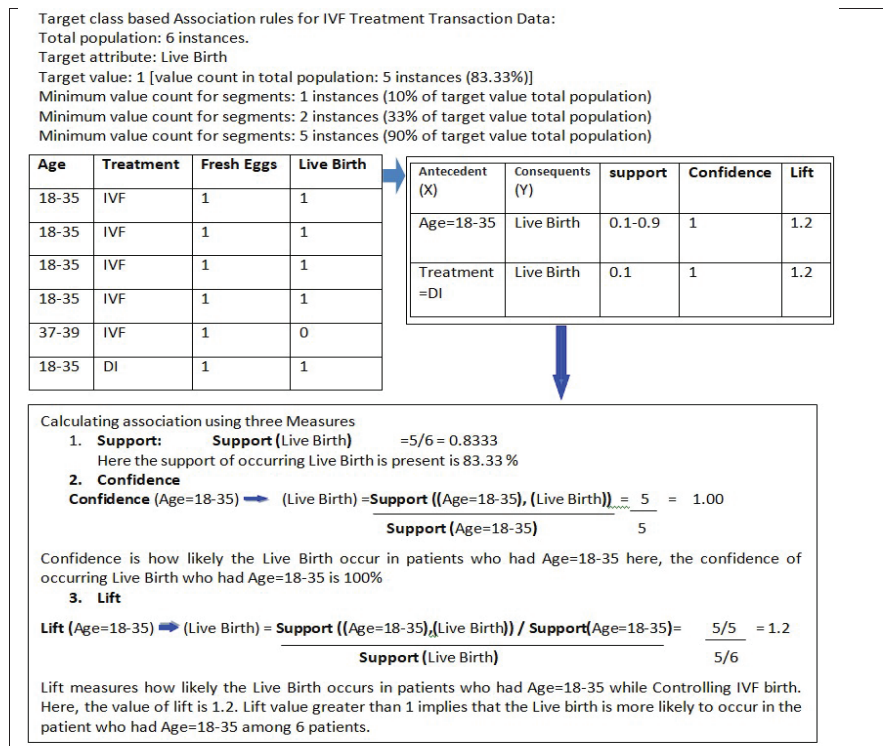


Figure 3 Association Rule Mining examples with transactional data.

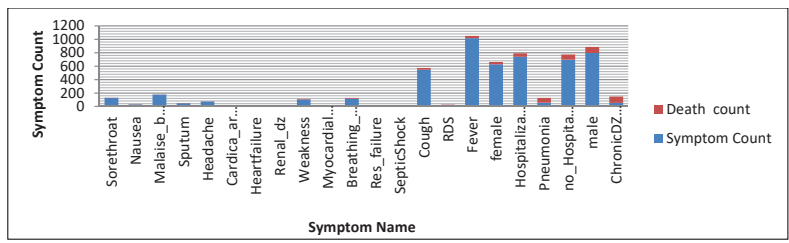


Figure 4 Symptom counts (Death/Survival) vs. symptom name (N = 1560).

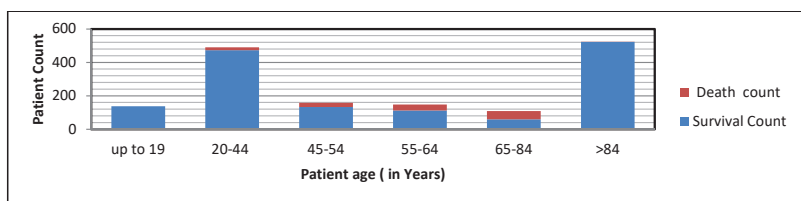


Figure 5 Patient Count (Died/Survived) Vs. Patient age (in years) (N = 1560).

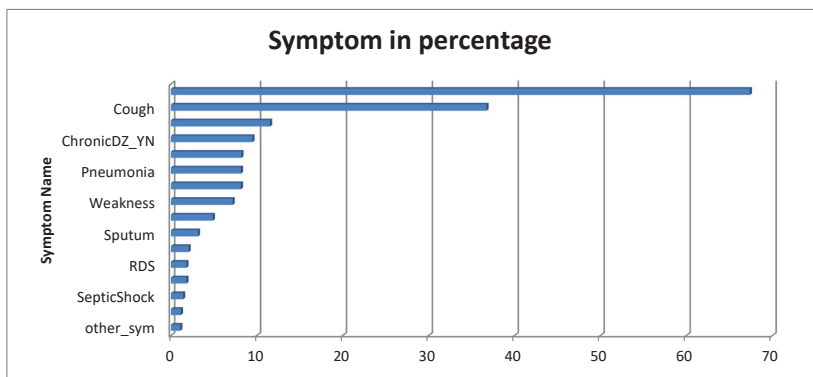


Figure 6 Symptom in percentage Vs. Symptom Name (N = 1560).

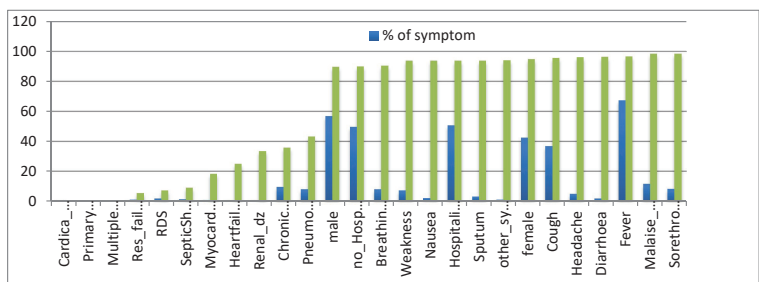


Figure 7 Chance of survival vs. % of symptoms (N = 1560).

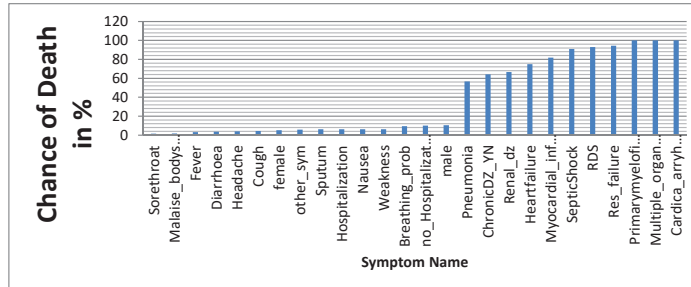


Figure 8 Symptom Name vs. Death Probability for COVID-19 Patients in % (N = 1560).

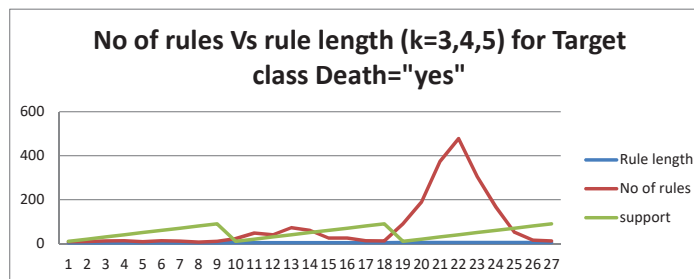


Figure 9 Shows the number of rules for rule length value for support value 0.1 to 0.9.

6 Discussion

The study was undertaken mainly to examine the two most important questions about a patient’s survival or death from COVID-19. In this work, we found a few new rules for patients suffering from COVID-19 who had different symptoms and the presence, absence, and combination of various symptoms was responsible for the survival and death of patients. This work also found the rules for male and female patients, age, and the rules of chronic disease with their survival and death. In this paper, we used the HotSpot Algorithm. A HotSpot is an algorithm that can directly mine association rules from real and categorical data. It can directly mine association rules and dynamically acquire the range of real number intervals without the discretization of real data, thus avoiding the influence of subjective factors in discretization. The processing speed is extremely fast. But like the Apriori algorithm, the HotSpot algorithm has one obvious shortcoming: support selection needs to be set artificially based on experience and cannot be set accurately according to the actual scale of the problem. If the support threshold is set too low, a cumbersome and complicated tree structure may be

generated; if the support threshold is set too high, some associated intervals existing in the rare target attribute values may be ignored. Therefore, in the process of support selection, multiple comparison experiments are needed to determine the optimal support based on the mining results. When it comes to setting up support thresholds for the HotSpot algorithm, it can be very sensitive. This is why it combines intelligent optimization technology with association rule-mining technology [42]. Most of the symptoms rules consisted of Malaise bodysore and Hospitalization, Age = >85 years, Fever, Rest failure Myocardial infarction, RDS, Pneumonia, cough, sex, Breathing problem as antecedents and consequent like survival, death, with chronic disease, without chronic disease and various age ranges.

7 Conclusion and Future Work

This framework can take a real number and categorical data as input and can find rules for targeted class value; it means clinical attributes with real no directly used to find symptom rules. The most incessant manifestations in our review include Fever, Cough, disquietude/body sore, Pneumonia, Sore throat Weakness, migraine, Sputum, and Nausea were the most ordinarily revealed side effects in COVID-19 patients and respiratory disappointment addressed one to two percent of indications among COVID-19 patients. Patients suffering from Diseases like Cardiac arrhythmia, multiple organ failure, Primary myelofibrosis, respiratory failure, respiratory distress syndrome, septic shock, myocardial infarction, heart failure, renal disease, chronic disease, and Pneumonia shows higher death rate (order of death rate is left to right for diseases) is discussed in the study. ARM strategies recognized essentially unique indication rules for COVID-19 among more youthful and more seasoned patients, male and female patients, patients with and without constant conditions, and individuals who have persistent conditions. It has also come in the study that patients who have suffered from sore throat, malaise body sore, fever, diarrhea, headache, cough, female, other symptoms, sputum, nausea, weakness, breathing problem their chance of survival are very high (order of survival rate is left to right for diseases) is discussed in the study.

This paradigm can be used to create rules for targeted outcomes or focused therapy. The targeted association rule mining idea, which is based on market basket analysis, can be used to acquire patient behavior with regard to treatment, medical behavior with regard to patient sickness and clinical features, and medicine/therapy success patterns. This research will

aid in the discovery of novel and practical rules that will aid doctors and patients in early detection and medication. The rules have various severe and mild ones responsible for chronic conditions and the survival and death of COVID-19 patients indicate the importance of various symptoms detection and efficient management of COVID-19 patients. We have conducted this study by taking more no of attributes as compared to other studies. This work may be expanded to learn about the effects of corona-variants on different blood groups, as well as the effects of different corona variants on COVID-19 vaccinated persons.

Credit Origin Commitment Proclamation

Anju Singh: Concept, information curation, Divakar Singh: Clinical conversation, amendment and settling the composition. Kamal Upreti: Support in the utilization of affiliation rule mining and settling the composition. Vaibhav Sharma: Support in the utilization of affiliation rule mining, settling the composition. Bhawani singh rathore: Concept, information cleaning. Jagdish Raikwal: Drafting composition and finish.

Presentation of Contending Interest

The creators proclaim that they have no known contending monetary interests or individual connections that might have seemed to impact the work announced in this paper.

References

- [1] WHO, Coronavirus Disease (COVID-19) Pandemic, May 2020. Accessed: 2020-07-27.
- [2] Worldometer, COVID-19 Coronavirus Pandemic Reported Cases and Deaths by Country, Territory, or Conveyance, 2020. Accessed: 2020-07-26.
- [3] Recovery Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, Staplin N, Brightling C, Ustianowski A, Elmahi E, Prudon B, Green C, Felton T, Chadwick D, Rege K, Fegan C, Chappell LC, Faust SN, Jaki T, Jeffery K, Montgomery A, Rowan K, Juszczak E, Baillie JK, Haynes R, Landray MJ., Dexamethasone in

- hospitalized patients with covid-19-preliminary report, *N. Engl. J. Med.* Jul 17:NEJMoa2021436 (2020).
- [4] WHO, Solidarity Clinical Trial for COVID-19 Treatment, 2020. Accessed: 2020-07-27.
 - [5] Jonathan Grein, Norio Ohmagari, Daniel Shin, Diaz George, Erika Asperges, Antonella Castagna, Torsten Feldt, Gary Green, Margaret L. Green, François-Xavier Lescure, et al., Compassionate use of remdesivir for patients with severe covid-19, *N. Engl. J. Med.* 382 (24) (2020) 2327–2336.
 - [6] Bauchner Howard, Phil B. Fontanarosa, randomized clinical trials and covid-19: managing expectations, *JAMA* 323 (22) (2020) 2262–2263.
 - [7] Jeff Craven, COVID-19 Therapeutic Tracker, 2020. Accessed: 2020-12-20.
 - [8] K. Pazhanikumar, S. Arumugaperumal, Association rule mining and medical application: a detailed survey, *Int. J. Comput. Appl.* 80 (17) (2013).
 - [9] Kamran Shaukat, Sana Zaheer, Iqra Nawaz, Association rule mining: an application perspective, *Int. J. Contr. Syst. Instrum.* (1) (2015) 29–38, 2015.
 - [10] Sui-Lee Wee Carol Zimmer, Jonathan Corum, Coronavirus Vaccine Tracker, Updated on December 18, 2020, 2020. Accessed: 2020-12-20.
 - [11] Bloomberg. More, than 1.6 million People Have Been Vaccinated-Covid-19 Tracker, Updated December 19, 2020, 2020. Accessed: 2020-12-20.
 - [12] Kristine A. Moore, Marc Lipsitch, John M. Barry, Michael T. Osterholm, Part 1: the Future of the Covid-19 Pandemic: Lessons Learned from Pandemic Influenza. COVID-19: the CIDRAP Viewpoint, Center for Infectious Disease Research and Policy, 2020.
 - [13] Gangqiang Guo, Lele Ye, Kan Pan, Yu Chen, Xing Dong, Kejing Yan, Zhiyuan Chen, Ning Ding, Wenshu Li, Hong Huang, et al., new insights of emerging sars-cov-2: epidemiology, etiology, clinical features, clinical treatment, and prevention, *Front. Cell Dev. Biol.* 8 (2020) 410.
 - [14] Andreas, Oskar Eriksson, Martin Nordberg, Analysis of scientific publications during the early phase of the covid-19 pandemic: topic modeling study, *J. Med. Internet Res.* 22 (11) (2020), e21559.
 - [15] A. Sayed, Y. Acharya, K.C.V. Long, L. Lynam, M. Tandan, Estimation of clinical comorbidities in covid-19 patients: a systematic re-view and meta-analysis, *Ann. Microbiol. Res.* 4 (1) (2020) 105–111.

- [16] Jean-Marc Adamo, *Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms*, Springer Science & Business Media, 2001.
- [17] Shreshth Tuli, Shikhar Tuli, Gurleen Wander, Praneet Wander, Sukhpal Singh Gill, Schahram Dustdar, Rizos Sakellariou, Omer Rana, Next generation technologies for smart healthcare: challenges, vision, model, trends and future directions, *Internet Technol. Lett.* 3 (2) (2020) e145.
- [18] Adrien Depeursinge, Ann N Leung Anne S Chin, Donato Terrone, Michael Bristow, Glenn Rosen, Daniel L. Rubin, Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution ct, *Invest. Radiol.* 50 (4) (2015) 261.
- [19] Maxwell W. Libbrecht, William Stafford Noble, Machine learning applications in genetics and genomics, *Nat. Rev. Genet.* 16 (6) (2015) 321–332.
- [20] Shreshth Tuli, Shikhar Tuli, Rakesh Tuli, Sukhpal Singh Gill, Predicting the Growth and Trend of Covid-19 Pandemic Using Machine Learning and Cloud Computing, *Internet of Things*, 2020, p. 100222.
- [21] Mohammad Jamshidi, Lalbakhsh Ali, Jakub Talla, Zdeněk Peroutka, Farimah Hadjilooei, Pedram Lalbakhsh, Morteza Jamshidi, Luigi La Spada, Mirhamed Mirmozafari, Mojgan Dehghani, et al., Artificial intelligence and covid- 19: deep learning approaches for diagnosis and treatment, *IEEE Access* 8 (2020) 109581–109595.
- [22] Talha Burak Alakus, Ibrahim Turkoglu, Comparison of deep learning approaches to predict covid-19 infection, *Chaos, Solit. Fractals* 140 (2020) 110120.
- [23] Chunming Xu, Scott A. Jackson, *Machine Learning and Complex Biological Data*, 2019.
- [24] Wolfgang Huber, Vincent J. Carey, Robert Gentleman, Simon Anders, Marc Carlson, Benilton S. Carvalho, Hector Corrada Bravo, Sean Davis, Laurent Gatto, Thomas Girke, et al., Orchestrating high-throughput genomic analysis with bioconductor, *Nat. Methods* 12 (2) (2015) 115–121.
- [25] Rosanna Upstill-Goddard, Diana Eccles, Joerg Fliege, Andrew Collins, Machine learning approaches for the discovery of gene-gene interactions in disease data, *Briefings Bioinf.* 14 (2) (2013) 251–260.
- [26] Pokharel Suresh, Zhenkun Shi, Zucco Guido, Li Yu, Discriminative features generation for mortality prediction in icu, in: *International Conference on Advanced Data Mining and Applications*, Springer, 2020.

- [27] Chandra Prasetyo Utomo, Hanna Kurniawati, Xue Li, Suresh Pokharel, Personalised medicine in critical care using bayesian reinforcement learning, in: International Conference on Advanced Data Mining and Applications, Springer, 2019, pp. 648–657.
- [28] Jessica Vamathevan, Dominic Clark, Czodrowski Paul, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, et al., Applications of machine learning in drug discovery and development, *Nat. Rev. Drug Discov.* 18 (6) (2019) 463–477.
- [29] Timilsina Mohan, Meera Tandan, Mathieu d’Aquin, Haixuan Yang, discovering links between side effects and drugs using a diffusion-based method, *Sci. Rep.* 9 (1) (2019) 1–10.
- [30] Pokharel Suresh, Zuccon Guido, Li Xue, Chandra Prasetyo Utomo, Li Yu, Temporal Tree Representation for Similarity Computation between Medical Patients. *Artificial Intelligence in Medicine*, 2020.
- [31] Jean-Baptiste Lamy, Boomadevi Sekar, Gilles Guezennec, Jacques Bouaud, Brigitte Séroussi, Explainable artificial intelligence for breast cancer: a visual case-based reasoning approach, *Artif. Intell. Med.* 94 (2019) 42–53.
- [32] Rakesh Agarwal, Ramakrishnan Srikant, et al., Fast algorithms for mining association rules, in: *Proc. Of the 20th VLDB Conference*, 1994, pp. 487–499.
- [33] Jiawei Han, Jian Pei, Mining frequent patterns by pattern-growth: methodology and implications, *ACM SIGKDD Explor. Newslett.* 2 (2) (2000) 14–20.
- [34] https://github.com/Mtandan/COVID_ARM.git. Accessed: 2021-22-10.
- [35] Becker, U. and Fahrmeier, L. (2001) Bump hunting for risk: a new data mining tool and its applications. *Computational Statistics* 16 (3) 373–386.
- [36] Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen, Association rule mining to detect factors which contribute to heart disease in males and females, *Expert Syst. Appl.* 40 (4) (2013) 1086–1093.
- [37] Stefan Mutter, Mark Hall, Eibe Frank, using classification to evaluate the output of confidence-based association rule mining, in: *Australasian Joint Conference on Artificial Intelligence*, Springer, 2004, pp. 538–549.
- [38] Tyler McCormick, Cynthia Rudin, David Madigan, A Hierarchical Model for Association Rule Mining of Sequential Events: An Approach to Automated Medical Symptom Prediction, 2011.

- [39] Laszlo Szathmary, Petko Valtchev, Amedeo Napoli, Generating Rare Association Rules Using the Minimal Rare Itemsets Family, 2010.
- [40] Coronavirus disease 2019 (COVID-19): A literature review Journal of Infection and Public Health 13(5) Harapan Harapan, Naoya Itoh, Amanda Yufika, Wira Winardi, Synat Keam, Haypheng Te, Dewi Megawati, Zinatul Hayati, Abram L. Wagner, Mudatsir Mudatsir, April 2020.
- [41] Tandan, M., Acharya, Y., Pokharel, S., & Timilsina, M. Discovering symptom patterns of COVID-19 patients using association rule mining. Computers in Biology and Medicine, 131, 104249, 2021. <https://doi.org/10.1016/j.compbiomed.2021.104249>.
- [42] Rui Pang, Yue Chang, An Improved HotSpot Algorithm and Its Application to Sandstorm Data in Inner Mongolia, Apr 2020.

Biographies



Anju Singh is working as an Associate Professor, Sage University, Bhopal Madhya Pradesh, India. Till now 60 M. Tech dissertations guided by her and 57+ Research papers were published. Her Google research scholar total citations are 202, h-index is 09 and i-index is 08. She has filed one international patent. She has more than 15 years of experience in both teaching and research. She was honored by the Session Chair/Technical Committee Member on National and International conferences. She is a life member of ISTE and CSI and a fellow IETE. Her area of interest includes Deep learning, image processing, nature-inspired algorithms and soft computing.



Divakar Singh is working as an Assistant Professor at University Institute of Technology, Barkatullah University, Bhopal, and Madhya Pradesh, India. He got ranked in the Top 50 head of the department CSE/IT award in all India level for session 2019–20 by uLektz Wall Of Fame. Till now 91 M. Tech and 1 PhD, dissertation guided by him and 130+ Research papers were published. His Google research scholar total citations are 420, h-index is 11 and i-index is 15. He has more than twenty years of experience in both teaching and research. He has served as a Head of the department, in Computer Science and Engineering since 2006 in UIT, Barkatullah University. He has worked as a member of board of studies and as a chairman of, board of studies in the subject CSE/IT/Electronics in the Faculty of Engineering. His area of interest includes Deep learning, image processing, nature-inspired algorithms and soft computing.



Kamal Upreti is currently working as an Associate Professor in the Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology & Management, Delhi. He completed is B. Tech (Hons), M. Tech (Gold Medalist), PGDM(Executive) and PhD in Computer Science & Engineering.

He has published 50+ patents, 32+ books, 15+ magazine issues and 30+ research papers in various reputed international Journals and Conferences.

His areas of research interest include Machine Learning, Wireless Networking, Embedded System and Cloud Computing. He has been chaired many sessions in National and International Conferences across the globe.



Vaibhav Sharma is working as an Assistant Professor in School of Computer Application & Information Technology at SGRR University Dehradun, India since 2013. He has total 12+ years of teaching experience in various Institutes and Universities. He has qualified UGC NET and USET. He is pursuing his PhD in Computer Science. He has published Three Patents, Two Books, and Five Research Papers in national, international journals and conferences and achieved Five best Teacher and Research Excellence awards. He has also participated in many national and international Conferences, Workshops, Webinars, FDP, STC etc. His areas of interest include Programming in C, C# with .Net Framework, RDBMS, Design and Analysis of Algorithms, Python, IOT etc.



Bhawani Singh Rathore is working as a Guest Faculty at University Institute of Technology, Barkatullah University, Bhopal, and Madhya Pradesh, India. He is having 5 years of teaching experience. He has worked as a Teacher

Guardian at UIT, Barkatullah University. He has delivered an Expert lecture in AWS and Google Cloud Chatbot. He has attended FDP at NIT Kurukshetra and NITTTR Chennai. His area of interest includes Google Clouds, Data Mining, and Web APIs.



Jagdish Raikwal is Assistant Professor in the Department of Information Technology at Institute of Engineering & Technology, Devi Ahilya University, Indore. He is having more than 12 years of teaching experience. He has more than 26 research publications in various pre-reviewed Journal and Conferences.