



Research paper



Defining factors in hospital admissions during COVID-19 using LSTM-FCA explainable model

Nurul Izrin Md Saleh^{a,*}, Hadhrami Ab Ghani^a, Zairul Jilani^b

^a Department of Data Science, Universiti Malaysia Kelantan, 16100 Pengkalan Chepa, Kelantan, Malaysia

^b Department of Computing, Sheffield Hallam University, Sheffield S1 1WB, United Kingdom

ARTICLE INFO

Keywords:

Long Short-Term Memory (LSTM)
Formal Concept Analysis (FCA)
COVID-19
Hospital admissions

ABSTRACT

Outbreaks of the COVID-19 pandemic caused by the SARS-CoV-2 infection that started in Wuhan, China, have quickly spread worldwide. The current situation has contributed to a dynamic rate of hospital admissions. Global efforts by Artificial Intelligence (AI) and Machine Learning (ML) communities to develop solutions to assist COVID-19-related research have escalated ever since. However, despite overwhelming efforts from the AI and ML community, many machine learning-based AI systems have been designed as black boxes. This paper proposes a model that utilizes Formal Concept Analysis (FCA) to explain a machine learning technique called Long-short Term Memory (LSTM) on a dataset of hospital admissions due to COVID-19 in the United Kingdom. This paper intends to increase the transparency of decision-making in the era of ML by using the proposed LSTM-FCA explainable model. Both LSTM and FCA are able to evaluate the data and explain the model to make the results more understandable and interpretable. The results and discussions are helpful and may lead to new research to optimize the use of ML in various real-world applications and to contain the disease.

1. Introduction

The novel coronavirus (COVID-19), an infectious disease that causes severe acute respiratory syndrome was discovered for the first time in Wuhan, China, in November 2019. Ever since the virus has infected and killed millions of people globally. The first UK COVID-19 cases were reported on January 31, 2020. Within three months, daily cases rose sharply to more than 33,000 cases. Other countries that had a spike in daily cases during the early period were the US, Brazil, Italy, Spain, and Iran. The figure terrified the globe, and on March 11, 2020, the World Health Organization (WHO) proclaimed the outbreak a pandemic.

Due to the aggressive number of cases, the entire healthcare system has to respond and make decisions promptly to ensure it does not fail. Preventive measures like social distancing, wearing face coverings, hygienic lifestyle, i.e., hand washing and disinfecting surfaces, and lockdown are enforced by governments worldwide. Moreover, patients with positive COVID-19 have to be admitted into an isolation area with stringent procedures to prevent the disease from spreading. When this work is carried out, vaccines have been released to the public to contain the disease. This situation has created opportunities for researchers to study COVID-19 in any aspect while using COVID-19 datasets that are publicly available.

In this study, LSTM algorithm of Machine Learning is used to demonstrate factors affecting hospital admission in time-series prediction. As machine learning becomes more prevalent in accommodating and accelerating the decision-making process, it is essential to describe and understand how the predictions are made while defining and mitigating bias. These ML-based decisions have led to the development of application that improves people's health, safety, economic well-being, and other aspects of life [1–3]. An understandable and explainable ML model is a difficult topic to address due to varieties of machine learning algorithms and the nature of how machine learning model training works, yet model interpretability has become a fundamental element in making model predictions understandable [4,5]. Hence Formal Concept Analysis (FCA) is used to explain the ML model. This paper is organized into the following sections: Section 1 outlines the review of previous work, motivation, and objectives. Section 2 discusses the method used in this work. The results are presented in Section 3, and discussed in Section 4. Finally, the paper is concluded in Section 5.

1.1. Exploration and review

In fighting the pandemic, data has shown that the first UK mass vaccination programme started in early December 2020. [6] reported

* Corresponding author.

E-mail address: nurulizrin.ms@umk.edu.my (N.Z. Md Saleh).

that when 50% of the adult population has been vaccinated, death tolls are reduced by 95% and hospital admissions are reduced by 80%. [7,8] have reported that a rollout of the Pfizer BioNTech and Oxford AstraZeneca vaccines has led to a substantial fall in severe COVID-19 cases requiring hospital admissions in Scotland. The impact of the lockdown in France and its efficiency in combating COVID-19 has been assessed using a stochastic age-structured transmission model that includes data on age profile and social relationships [9]. The model evaluated the impact of lockdown as well as the best options for dealing with the health crisis after the lockdown was lifted.

A study [10] on hospitalization rates and characteristics of patients hospitalized reported that the COVID-19 associated hospitalization rate in the early period of the pandemic in the US was 4.6 per 100,000 population, and that the rate increased with age. While in Iran, a study on the hospitalization and death rates among patients with multiple sclerosis has discovered that the rate of hospitalization was 25% higher than the general population [11]. School re-opening and hospitalization in the US have not resulted in an increase in COVID-19 hospitalizations and interestingly, the virus spread among school staff, but not among students [12]. Furthermore, a study on sociodemographic, clinical and laboratory factors on admission associated with COVID-19 mortality in hospitalized patients was conducted to identify associations between baseline characteristics on hospital admission and mortality in patients with COVID-19 in Spain [13].

LSTM, which was introduced by [14] in 1997, is a Recurrent Neural Network (RNN) based architecture that is widely used in natural language processing and time series forecasting. LSTM is a neural network version that addresses the issue in a traditional neural network on reasoning from previous events. LSTM has networks with loop to pass information from one step to the next one and support information to persist. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate (Fig. 1).

Many researchers in AI promote their expertise in understanding COVID-19 and finding solutions to reduce the worldwide threat to living society. The focal study of the study's interest is mostly predictive analytics towards COVID-19 cases, infections, and spread. Deep learning seems to be the popular technique used and extensively explored for prediction. Narinder S. et al. [15] in their study used deep learning algorithms in predicting COVID-19 cases as well as understanding the exponential behaviour in a number of cases. Several techniques such as LSTM, Recurrent Neural Network (RNN), and Support Vector Regression (SVR) were used.

Similarly, a comparative case study was carried in [16] out across different countries to learn on the factors causing the COVID-19 besides predicting the future timeline of the pandemic. The study demonstrated forecasting of the confirmed and death cases due to COVID-19. A few of LSTM versions were used in the work and it was found that Convolutional LSTM performed best in the forecasting. At the point of the study, it was forecasted that the number of COVID-19 confirmed and death cases for both countries would increase. The author therefore suggested some immediate actions for both countries to stop the disease from widely spreading.

Meanwhile, a different focus of study by [17] predicts COVID-19 progression in countries. Y Li, W Jia et al. studied on lockdown enforcement as the preventive measure in containing the epidemic COVID-19. They propose a transfer learning approach together with the recurrent neural network (RNN) to measure the significance of the predictor variable (lockdown) in predicting the disease progression. Their promising results show that the lockdown measure significantly improves the prediction performance as compared to other predictive modelling methods without lockdown. Their proposed method achieved the least mean absolute percentage error (MAPE) score (0.005) amongst the other three methods used in the work. The results directly suggest that lockdown measures and extending the period of lockdown are still necessary in containing COVID-19 from spreading.

A study carried out by Rohitash C. et al. [18] used univariate and multivariate time series LSTM for forecasting the spread of COVID-19 infection in India. They employed Bi-directional (BD-LSTM) and Encoder-decoder (ED-LSTM) in their experiments to obtain long-term forecasts (of the progression of COVID-19) for two months. The study found that the univariate LSTM model provides the best performance of forecasting. Their forecasting results suggest low likelihood of another wave infection in October and November 2021.

There was a study on forecasting the dates for containing the COVID-19 disease from spreading using LSTM [19]. The study concerned about the economy and social impact in the countries, thus having predicted dates in hand that the pandemic will subside helps the government, policymakers, entrepreneurs and businesses make appropriate decisions to recover from the disease impact. The LSTM method was validated on the New Zealand dataset between April and May 2021, and reported a correct prediction (dates) in containing COVID-19. However, the predictive model was reported not performing well in some other countries. The authors claimed that the predictive model would be more reliable in those countries if restrictions were not lifted. Having the poor modelling performance due to restrictions is a research opportunity for this work.

In general, ML approach with the setting of COVID-19 pandemic has been used to identify patients at high risk, their death rate, and other abnormalities in order to understand the virus and further predict the upcoming issue [20]. Recent studies indicate that elderly and fragile people are affected by COVID-19, but it has also claimed many young lives. [21] applies machine learning to identify and predict people based on their vulnerability or resistance to possible COVID-19 infection using genetic variants from asymptomatic, mild, and severe COVID-19 patients. The ML model produces useful findings that aid stakeholders in their decision-making.

Using ML approach — LSTM, a study to forecast COVID-19 transmission in Canada, reported that their model has resulted in RMSE errors of 34.83 and 45.70 for short and long-term prediction [22]. Another study [23] has used various regressors in predicting new cases, deaths, and recoveries of COVID-19 suggests that LSTM is the second best method amongst others. [24] in their study of forecasting COVID-19 infection in India with LSTM, Bi-directional LSTM (BD-LSTM), and Encoder-Decoder LSTM (ED-LSTM) claims that LSTM model gives the best performance for more cases compared to BD-LSTM and ED-LSTM. Their best RMSE values on univariate and multivariate LSTM are 1403 and 4572 respectively.

Obtaining causal linkages in COVID-19 data and presenting them in a way that makes them easy to use has also received a lot of attention. Studies have implemented mathematical formalisms of Formal Concept Analysis (FCA) to discover and represent causal relations of domain issues [25,26]. The foundation of FCA is in its ability to construct a set of logical implications from context-specific ideas by applying ordered set theory and domain-specific knowledge lattices. Other study by [27] has also used FCA to uncover relationships between vaccines and other attributes that cause the outbreaks of COVID-19. This study proposes a rational strategy to design vaccination schemes for curbing the COVID-19 pandemic. Hence, the result is exclusively based upon the natural settings of hierarchical ordered data and does not 'learn' nor leverage it to predict future outcomes.

In another study [28], a deterministic approach has been developed using an SEIR-mathematical modelling framework to explore the concept of optimal and robust interventions across a range of different non-pharmaceutical interventions (NPI) scenarios. An epidemiological mathematical model has been proposed in [29] for capturing and predicting the spread of COVID-19 with a simulation model which is performed using the two-step generalized exponential time-differencing method. In general, mathematical formalisms are applied to make the model more explainable, as per described by [4], since domain knowledge is an essential part of explainability. It is found that the main interest from literature was to obtain higher predictive accuracy

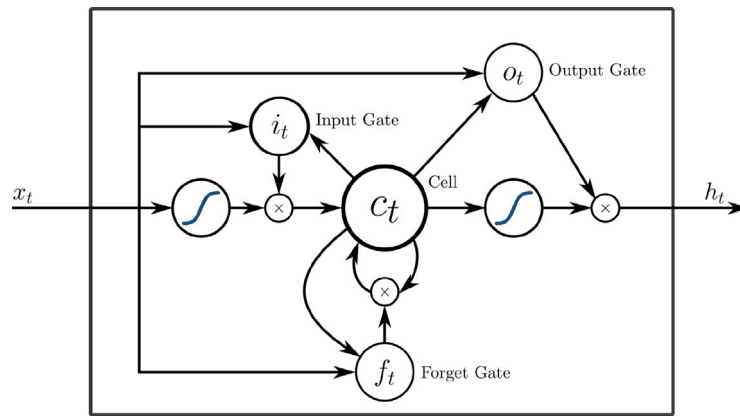


Fig. 1. LSTM components diagram.

to support future decision making. There is a need to identify and explain the contributing factors to high predictive accuracy. Furthermore, getting the variables relationship and its magnitude in forecasting COVID-19 would be able to comprehend the underlying reason of the high predictive accuracy.

1.2. Motivation

This study is motivated by the urge to address the factors that contribute to hospital admissions to help in making informed decisions and to respond quickly in managing the pandemic. Additionally, this work is also motivated by the availability of massive databases and current breakthroughs in ML approaches. These successful models are frequently used in a black box fashion, with no information provided regarding how they arrive at their conclusions. Lack of transparency feature can be a severe disadvantage, and due to this, the work aims to demonstrate a more transparent and interpretable machine learning model.

1.3. Objectives

The purpose of this study is to find contributing factors to hospital admissions in the UK due to COVID-19 using the forecasting method as well as to prove the model-based approach. And they explain the black-box result of LSTM using Formal Concept Analysis (FCA).

2. Method

This research devised Long-Short Term Memory (LSTM) networks to find associated and significant variables in predicting hospital admissions in the UK. As depicted in Fig. 7, the cleaned dataset is converted into time series before 63 datasets (Table 5) are prepared for LSTM experiments. Multivariate LSTM experiments are carried out in this work which predict hospital admissions using combinations of six variables. Whilst univariate experiments are run to set the error tolerance for the prediction. To explain the model and results obtained by LSTM experiments, Formal Concept Analysis (FCA) mathematical model is employed to interpret associating rules to the domain knowledge, further explanation in Section 2.4.

Empirical experiments are carried out using LSTM, while FCA is used to support the LSTM results. The hypothesis underlying this research is that the lower the error rate of LSTM prediction (of hospital admissions) with the dataset and independent variables, the more significant the variables in contributing to hospital admissions. This approach which namely model-based was tested in a few research work that cluster and classify significant and highly associate variables in predicting Glaucoma disease based on the model performance [30,31].

Table 1
List of variables.

Variable code	Variable name	Variable type
DV	New admissions	Dependent variable
IV1	Total cases	Independent variable
IV2	New cases	
IV3	Seasons	
IV4	National lockdown	
IV5	First dose	
IV6	Second dose	

Therefore, a few sets of experiments extensively investigated on the combination of the dependent variables that could predict the target variable with less root means square error (RMSE).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (observed_i - predicted_i)^2}$$

Experiments in this work are divided into several phases to investigate and observe the empirical experiment results before making any judgement towards the hypothesis. Each experiment was run 25 times to obtain consistent results and statistical values. Later in the study, the method of attribute exploration from FCA is used to explore the relationship between the attributes and whether they will be able to explain the LSTM result.

2.1. Dataset

The dataset of hospital admissions due to COVID-19 used in this work was formed and cleaned from a few sources, including the UK government and Institute for Government UK organizations. In total, there were 428 records in the dataset used in the experiment. They (a total of 7 variables) include new admissions as the predicted variable, total cases, new cases, seasons, national lockdown, number of people who have received the first dose vaccine, and second vaccine. The variables are coded in Table 1. Figs. 2 and 3 show the daily number of hospital admissions due to COVID-19 and new COVID-19 cases from March 2020 to May 2021.

Since seasons variable is used in this study, the statistical values of admissions and new cases are tabulated by seasons (Tables 2 and 3). From Figs. 4 and 5, it is clearly shown that high numbers for both admissions and new cases were recorded during the winter season.

2.2. LSTM experiment

LSTM networks are parameter dependent, such as optimizers, number of epochs, number of batches, and data partitioning. As finding the optimum parameters that best predict the dataset is out of the scope of

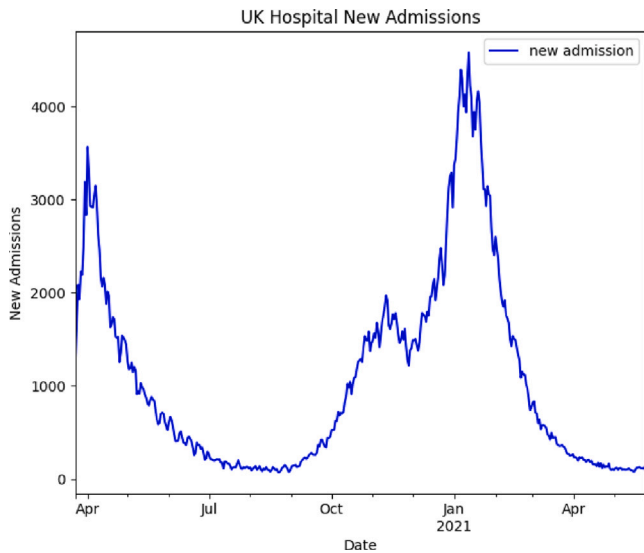


Fig. 2. UK hospital admissions.

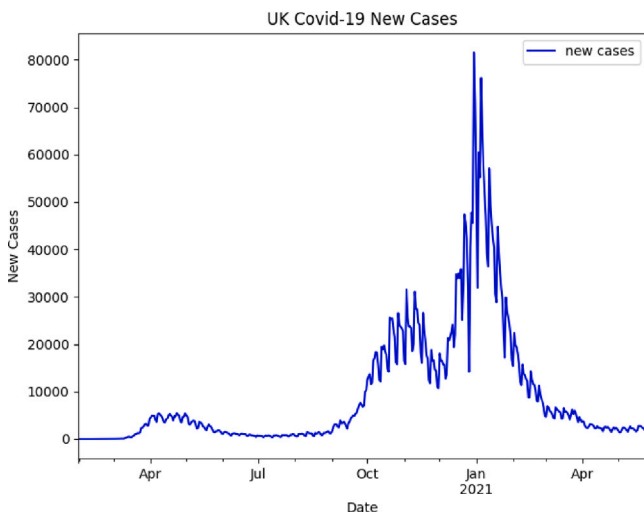


Fig. 3. UK COVID-19 new cases.

Table 2
Statistical values for admissions by seasons.

Season	Max	Min	Average
Winter	4576	364	2180
Spring	3565	78	849
Summer	394	72	174
Autumn	2168	340	1316

Table 3
Statistical values for daily new cases by seasons.

Season	Max	Min	Average
Winter	81523	4239	24291
Spring	6196	720	2878
Summer	5318	368	1416
Autumn	35833	5598	18833

the study, the following parameter values are set in the experiments: Optimizer: Adam, Epoch: 100, Batch size: 35.

Six independent variables are used in this work. Furthermore, the significant variables that highly contribute to the number of hospital admissions in the UK are studied in this paper using LSTM. From

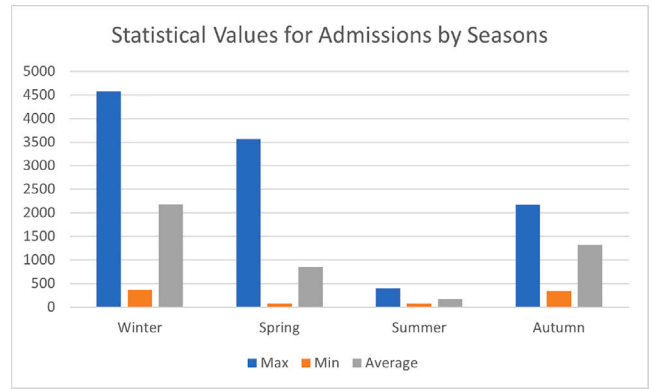


Fig. 4. Statistical values for admissions by seasons.

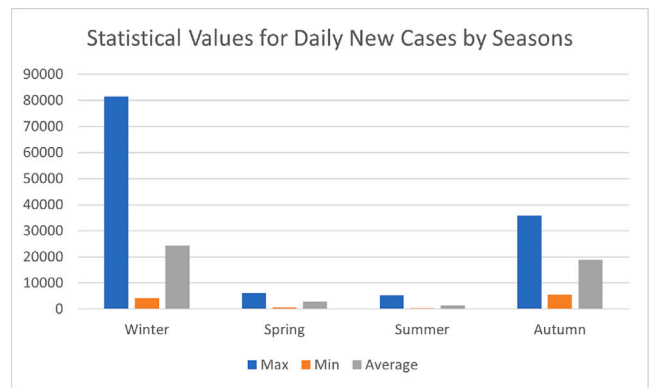


Fig. 5. Statistical values for daily new cases by seasons.

Table 4
Univariate LSTM results.

Mean	Min	Max	Std. Dev.
56.06839	54.90627	57.73614	0.851513

these variables (Table 1), there are 63 variable combinations of LSTM multivariate experiments as tabulated in Table 5.

2.3. Initial observation

Upon dataset conversion into time series, the Augmented Dickey–Fuller test was performed on the dataset to find the nature of the time series (stationarity and non-stationarity). The result of the test with a 0.05 significant level found that the time series dataset in hand is non-stationarity. The p -value is 0.266137 which is greater than 5% or 0.05 times the input data has a unit root. This indicates that the time series of hospital admissions dataset has trend and seasonality effects. Owing to this discovery, the season variable is included in this investigation.

Preliminary experiments of LSTM have also been run for univariate time series forecasting (with 25 samples) to test the method on the dataset. Table 4 and Fig. 6 show the results of the univariate LSTM experiments (the best RMSE value — iteration 18th). Based on the initial experiment results, the Residual Mean Square Root (RMSE) tolerance is set at 55. Within the formal multivariate LSTM experiments, only the variable combinations with RMSE values that are less than the tolerance value are accepted.

2.4. Formal concept analysis

To explain the causality of the LSTM result earlier, Formal Concept Analysis (FCA) is used. FCA was originally developed as a mathematical

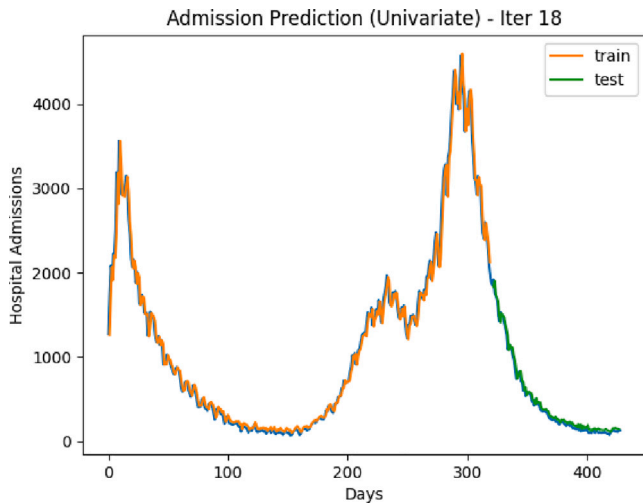


Fig. 6. LSTM univariate time series forecasting.

Table 5
Combination of variables.

Total variables	Variables combinations	Number of sets
6	1	6
	2	15
	3	20
	4	15
	5	6
	6	1
Total experiments		63

paradigm for concept formalization and conceptual reasoning by [32]. FCA examines the relationships between a group of objects and their properties as stated by [33]. The hierarchical property of concept lattices in FCA not only describes the relationships between attributes, but it also serves as a strong foundation for defining the structural property of the application domain [34]. Generally, FCA produces two sets of output. The first set of output is a list of all the interdependencies or rules that exist between the attributes in the attribute set *formal concept* – implications set (See Fig. 8). The second set of output is the hierarchical relationships of objects that exist in the domain – *concept lattice* (See Fig. 9). Followings are the list of definitions of FCA used in this study:

Definition 2.1 (Formal Context).

A succession of three similar things (X, Y, I) where objects X , attributes Y and a binary relation I between X and Y , i.e., $I \subseteq X \times Y$. $(x, y) \in I$ states that the object x has attribute y .

Definition 2.2 (Intent and Extent).

When (X, Y, I) is a context, $X' \subseteq X$ and $Y' \subseteq Y$, the Intent function maps the objects to the attributes, and the Extent function maps the attributes to objects:

$$\text{Intent}(X') = y \in Y' \mid \forall y \in Y', (x, y) \in R$$

$$\text{Extent}(Y') = x \in X' \mid \forall x \in X', (x, y) \in R$$

For $X' \subseteq X$, $\text{Intent}(X')$ is the attributes owned by all objects of X' , and $\text{Extent}(Y')$ is the set of all objects that own the attributes Y' . These two functions show a Galois connection and formal concepts for the domain.

Definition 2.3 (Formal Concept).

A Formal Concept C in a context is a pair (X', Y') that satisfies $Y' = \text{Intent}(X')$ and $X' = \text{Extent}(Y')$
i.e., C is a Formal Concept \Leftrightarrow for $X' \in \text{Cand}Y' \in C$, $\text{Extent}(\text{Intent}(X')) = X'$,

and symmetrically, $\text{Intent}(\text{Extent}(Y')) = Y'$.

Definition 2.4 (Implications).

An implication $A \Rightarrow B$ holds in (X, Y, I) if and only if $B \subseteq A''$, which is equivalent to $A' \subseteq B'$. The implications hold the set of all concept intents

The attribute exploration method from FCA is conducted using data of hospital admissions due to COVID-19. This exploration manages to show the relationships between agents' behaviour when dealing with COVID-19 pandemic. The data consists of 137 objects (dates range from January 11th, 2021–March 6th, 2021) and 5 attributes (*New Cases*, *National Lockdown*, *First Vaccine*, *Second Vaccine* and *Total Admission*) mapped as 'X' value into the cross table as formal context. 'X' is mapped in if there is an increase on a day-to-day basis. In a cross-table, associating an object to the attributes creates a concept hierarchy that can be visualized using the concept lattice. Fig. 8 shows a cross-table where formal contexts are mapped with ConExp software using the hospital admission data. The cross-table describes the formal context that existed as per the description in Definition 2.1. The little circles in Fig. 9 represent the 11 concepts of the context, and the ascending paths of line segments represent the subconcept–superconcept-relations. The definition of concepts is explained in Definition 2.3.

2.5. Attribute exploration

For the purpose of this study, the formal context is mapped into a table, named Cross-Table. (See Fig. 8) is conducted according to the main aim of this study, which is to define the factors in hospital admission during COVID-19. First, the data ranging from January 11th, 2021 to March 6th, 2021 are selected because that was the first time when the vaccination programme began. Multiple-valued data is then transformed into single-valued data. The progress of *New Cases*, *First Vaccine*, *Second Vaccine* and *Total Admission* are compared on a day-to-day basis. Whenever there is a decrease in *New Cases* and *Total Admission*, the X value will be mapped accordingly. And at the same time, when there is an increase in *First Vaccine*, *Second Vaccine* will also be mapped accordingly. Here, the criteria for the sought rules vary according to the aim of the study, and the basic knowledge of the domain data is implicitly gained [35]. Other than that, the *concept lattice* as depicted in Fig. 9 explains the hierarchical relationship of all the established concepts in the domain.

3. Results

As discussed in the experiment and method, the LSTM experiment was run for each of the variables' combinations (with a sample of 25 runs). As the tolerance of RMSE is set to 55, there are 7 variables' combinations with less than the tolerance value presented in the final results (Table 6).

Meanwhile, Table 3 shows the results (descending order) for variable combinations with the least significant in predicting hospital admissions (high values of RMSE). Despite not being in the lowest list, all six variable combination experiments also have high RMSE values of 463.306, 752.771, and 174.457 for mean, maximum, and minimum, respectively, with a standard deviation of 137.981.

Fig. 10 shows the best result (from iteration 9) with the least RMSE (Mean: 46.568). The amber and green lines from the figures are the predictions of admissions from training and test datasets, respectively. This result indicates “national lockdown” is the most significant variable in predicting hospital admissions. Whilst Fig. 11 exhibits the least best (from iteration 10) LSTM experiment results (“total cases” and “second vaccine”) (see Table 7).

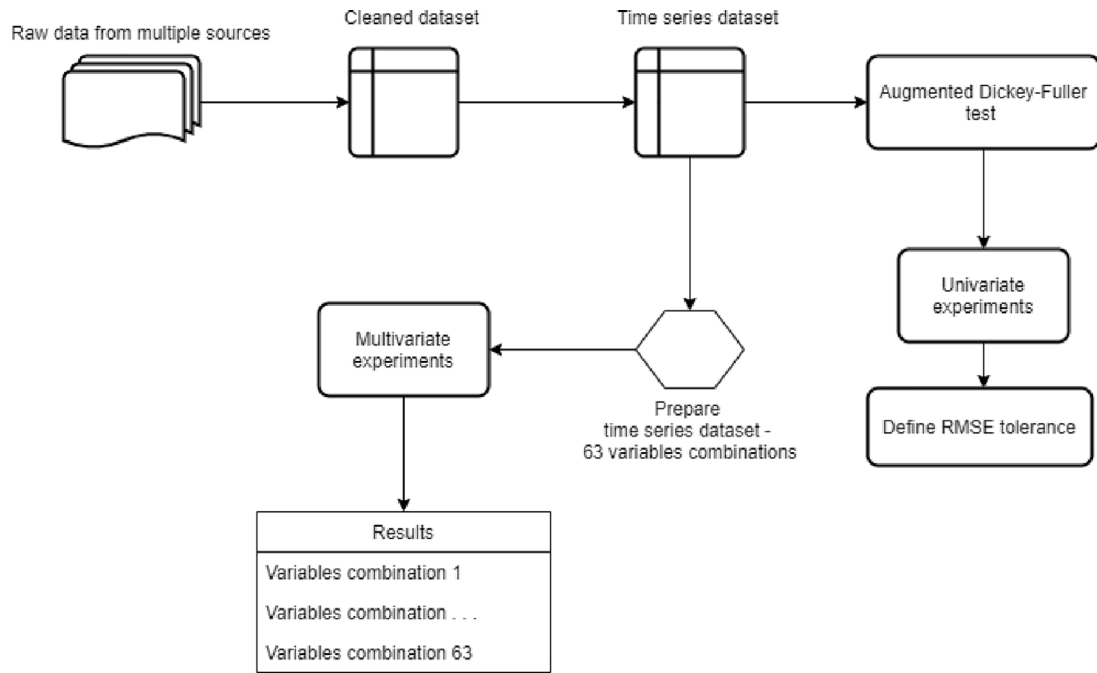


Fig. 7. High-level LSTM experiment.

A	B	C	D	E	F
	New Cases	National Lockdown	First Vaccine	Second Vaccine	Total Admission
11/01/2...		X			
12/01/2...		X	X		X
13/01/2...	X	X	X		
14/01/2...	X	X	X		
15/01/2...	X	X	X		
16/01/2...		X			
17/01/2...		X			
18/01/2...		X	X	X	X
19/01/2...		X	X		X
20/01/2...	X	X	X		
21/01/2...	X	X	X		
22/01/2...	X	X	X		
23/01/2...		X	X		
24/01/2...		X	X		
25/01/2...		X	X	X	
26/01/2...		X	X		X
27/01/2...	X	X	X	X	X
28/01/2...	X	X	X	X	X
29/01/2...	X	X	X	X	X
30/01/2...		X	X	X	
31/01/2...		X			
01/02/2...		X	X		
02/02/2...		X	X		
03/02/2...	X	X	X	X	
04/02/2...	X	X	X	X	
05/02/2...		X	X	X	
06/02/2...		X	X		
07/02/2...		X			
08/02/2...		X	X	X	
09/02/2...		X	X		
10/02/2...		X	X	X	X
11/02/2...	X	X	X	X	X
12/02/2...	X	X	X		
13/02/2...		X			

Fig. 8. The formal context of the hospital admission.

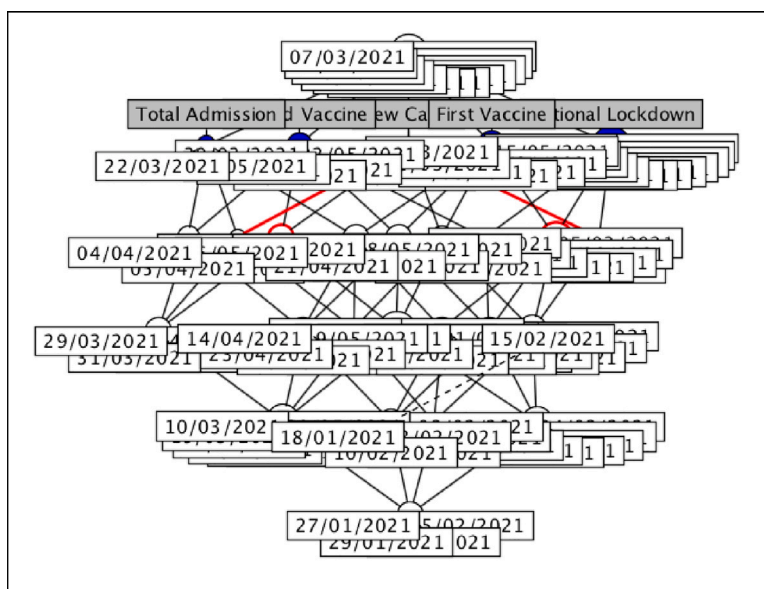


Fig. 9. The concept lattice of the hospital admission.

Table 6
List of variables combinations with least RMSE values.

Variables combination	RMSE value			
	Mean	Maximum	Minimum	Standard deviation
national lockdown	46.568	49.432	44.732	1.721
new cases	46.918	49.470	44.691	1.828
first vaccine	48.934	59.408	42.651	4.554
first vaccine new cases	49.332	58.869	43.121	4.458
new cases national lockdown	50.183	55.214	46.189	2.226
national lockdown first vaccine	53.398	67.625	45.713	5.261
first vaccine national lockdown new cases	54.983	70.512	45.623	6.938

Table 7
List of variables combinations with high RMSE values.

Variables combination	RMSE value			
	Mean	Maximum	Minimum	Standard deviation
total cases second vaccine	1215.418	1410.922	1058.166	102.854
national lockdown second vaccine new cases	1211.848	1403.473	935.851	123.488
second vaccine national lockdown	925.632	1127.558	729.070	106.644
second vaccine	873.184	1135.473	640.760	124.075
second vaccine total cases national lockdown first vaccine	700.847	1002.166	438.400	135.524

3.1. Association rules

From the conceptual exploration approach conducted, the dependencies between the attributes, i.e., attribute implications or association rules, are generated using ConExp. A total of 9 rules are generated to show the relationships between the attributes that exist from the data.

- **Rule 1** (100%)
Decreases in New Cases, imposes of National Lockdown, increases Second Vaccine implies increases in First Vaccine
- **Rule 2** (100%)
Imposes in National Lockdown and decreases in Total Admission implies increases in First Vaccine
- **Rule 3** (95%)

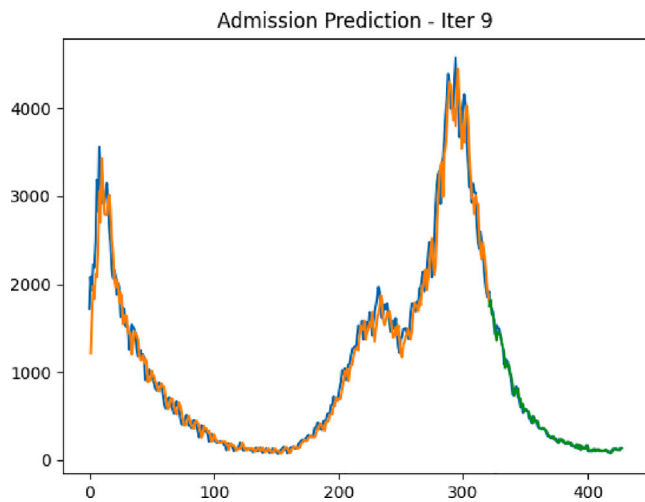


Fig. 10. LSTM multivariate time series forecasting with the best RMSE.

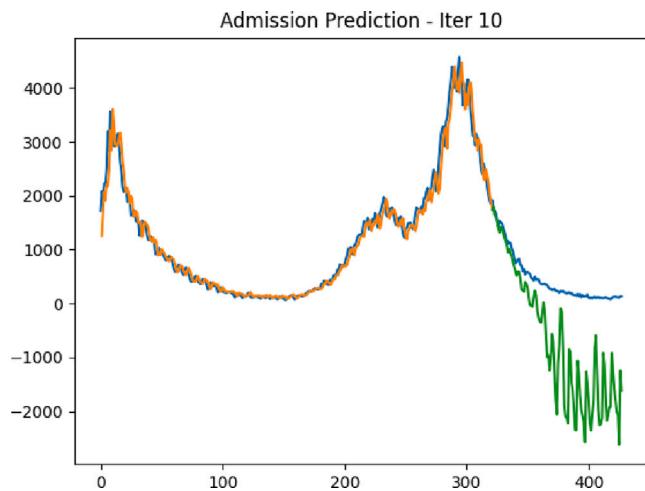


Fig. 11. LSTM multivariate time series forecasting with the highest RMSE.

Imposes of National Lockdown and increases in Second Vaccine implies increases in First Vaccine

- **Rule 4** (95%)
Increases in New Cases and imposes of National Lockdown implies increases in First Vaccine
- **Rule 5** (88%)
Increases in New Cases, increases in First Vaccine, decreases in Total Admission implies increases in Second Vaccine
- **Rule 6** (86%)
Decreases in New Cases, decreases in Total Admission implies increases in Second Vaccine
- **Rule 7** (82%)
Decreases in New Cases, increases in Second Vaccine implies increases in First Vaccine
- **Rule 8** (81%)
Decreases in New Cases implies increases in First Vaccine
- **Rule 9** (80%)
Decreases in New Cases, imposes of National Lockdown, increases in First Vaccine, decreases in Total Admission implies increases in Second Vaccine

4. Discussion

The first vaccine variable coincides with [6,8] where vaccination programme in the UK reduce hospital admissions, and a single dose vaccine is effective in preventing hospital admissions. It is also discovered that predicting hospital admissions using LSTM is best with a single independent variable (top three from the best results). Whilst pairing these three variables in LSTM prediction also presents promising results (below than 55 RMSE). Combining the three variables, however, is still within the defined tolerance of RMSE. Surprisingly, the season variable is not listed as one of the variables for LSTM prediction. Nevertheless, it is found that seasons and the first vaccine are the best pairing in the LSTM experiment, with RMSE values of 65.736, 76.656, and 52.631 for mean, maximum, and minimum, respectively (standard deviation: 5.769). On the contrary, the second vaccine is the least significant variable in the LSTM prediction as this variable’s presence in the bottom 5 variable combinations is quite significant.

Through Formal Concept Analysis (FCA) approach, from 9 association rules generated, 2 rules with clear implications between the attributes in the formal context (with confidence of 100%) are selected, which are **Rule 1** and **Rule 2**. The implication rules depict factors that contributed to new cases and hospital admissions in the UK between January 11th, 2021–March 6th, 2021 as vaccine rollout progresses and lockdown is imposed by the government. **Rule 1** implies that decreases in *New Cases*, when *Lockdown* imposes and increases in *Second Vaccine* rollout there is link to *First Vaccine* rollout. and **Rule 2** implies that decreases in *New Cases* and government imposing textitLockdown link to increases in *First Vaccine* rollout. From both rules, it has been deduced that *Lockdown* and *First Vaccine* have a strong implication in the number of cases and total admissions in UK hospitals.

The target variables from the LSTM result generated — *National Lockdown*, *New Cases* and *First Vaccine* and the rules generated by FCA — *National Lockdown*, *New Cases*, *First Vaccine* and *Second Vaccine* have a strong correlation with the total admission number due to COVID-19 in hospitals in the UK. As has been mentioned earlier, LSTM has demonstrated factors affecting hospital admission due to COVID-19 in time-series manner, whereas the natural setting of hierarchical concept of FCA has pointed out its causal relations. It is important to emphasize that using FCA to explain LSTM is a preliminary approach to an understandable and explainable AI.

5. Conclusion

From the study, the utility of methods explained, both LSTM and FCA are feasible in finding association variables and generating rules or hypotheses in the data. LSTM, a deep learning approach, has been employed in this paper to forecast the factors impacting admission due to COVID-19 and the FCA method of attribute exploration to develop rules or relationships between the attributes. The novelty aspect of this study is shown through the implementation of FCA to support the LSTM results, where the results from FCA have outlined domain knowledge for the explainability of the model. It has been discovered that this study is capable of evaluating data and explaining the model in order to ensure that the outcomes are understandable and interpretable. The findings and discussions may provide new insights that may result in the development of new research aimed at controlling the pandemic.

For future works, based on the promising RMSE values in the LSTM prediction and FCA discoveries, a number of research opportunities can be considered. The LSTM parameter values can be further explored to optimize the prediction. With the optimized prediction, a new set of significant variables or patterns could be found in order to see how seasons in the UK impact hospitalizations. In addition to the seasons variable, another empirical experiment can be carried out on a dataset that stretches a longer period of observations (2 years period that has covered vaccines and seasons). It should be noted that the experiments were run on a dataset on which vaccines and seasons were observed for less than six months. Another interesting future work is testing the approach on datasets from other countries for the same target variable.

Declaration of competing interest

Nurul Saleh reports financial support was provided by University of Malaysia Kelantan.

Acknowledgements

This work is funded by UMK-FUND research grant from Universiti Malaysia Kelantan (UMK), Malaysia.

References

- [1] Buyukarikan B, Ulker E. Classification of physiological disorders in apples fruit using a hybrid model based on convolutional neural network and machine learning methods. *Neural Comput Appl* 2022.
- [2] Bilogrevic I, Huguenin K, Agir B, Jadhwal M, Gazaki M, andHubaux JP. A machine-learning based approach to privacy-aware information-sharing in mobile social networks. *Pervasive Mob Comput* 2016;25:125–42.
- [3] Nieto Y, García-Díaz V, Montenegro C, Crespo RG. Supporting academic decision making at higher educational institutions using machine learning-based algorithms. *Soft Comput* 2019;23:4145–53.
- [4] Roscher R, Bohn B, Duarte MF, Garcke J. Explainable machine learning for scientific insights and discoveries. *Ieee Access* 2020;8:42200–16.
- [5] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv* 2018;51(5):1–42.
- [6] Cook T, Roberts J. Impact of vaccination by priority group on UK deaths, hospital admissions and intensive care admissions from COVID-19. *Anaesthesia* 2021;76(5):608–16.
- [7] Torjesen I. COVID-19: First doses of vaccines in Scotland led to a substantial fall in hospital admissions. *British Med. J. Publ. Group* 2021.
- [8] Iacobucci G. COVID-19: Single dose of pfizer and oxford vaccines cuts risk of hospital admission by 80% in over 80s, data suggest. 2021.
- [9] Di Domenico L, Pullano G, Sabbatini CE, Boëlle P-Y, Colizza V. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Med* 2020;18(1):1–13.
- [10] Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—COVID-NET, 14 states, March 1–30, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69(15):458.
- [11] Sahraian MA, Azimi A, Navardi S, Ala S, Moghadasi AN. Evaluation of the rate of COVID-19 infection, hospitalization and death among Iranian patients with multiple sclerosis. *Multiple Scler Relat Disord.* 2020;46:102472.
- [12] Harris DN, Ziedan E, Hassig S. The effects of school reopenings on COVID-19 hospitalizations. 2021, National Center for Research on Education Access and Choice.
- [13] Rivera-Izquierdo M, del Carmen Valero-Ubierna M, R-delAmo JL, Fernandez-Garcia MA, Martinez-Diz S, Tahery-Mahmoud A, et al. Sociodemographic, clinical and laboratory factors on admission associated with COVID-19 mortality in hospitalized patients: A retrospective observational study. *PLoS One* 2020;15(6):e0235107.
- [14] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80.
- [15] Punn NS, Sonbhadra SK, Agarwal S. COVID-19 epidemic analysis using machine learning and deep learning algorithms. 2020, MedRxiv, Cold Spring Harbor Laboratory Press.
- [16] Shastri S, Singh K, Kumar S, Kour P, Mansotra V. Time series forecasting of COVID-19 using deep learning models: India-USA comparative case study. *Chaos Solitons Fractals* 2020;140:110227.
- [17] Li Y, Jia W, Wang J, Guo J, Liu Q, Li X, et al. Alert-COVID: Attentive lockdown-aware transfer learning for predicting COVID-19 pandemics in different countries. *J Healthc Inform Res* 2021;5(1):98–113.
- [18] Chandra R, Jain A, Singh Chauhan D. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLoS One* 2022;17(1):e0262708.
- [19] Kumar S, Sharma R, Tsunoda T, Kumarevel T, Sharma A. Forecasting the spread of COVID-19 using LSTM network. *BMC Bioinformatics* 2021;22(6):1–9.
- [20] Kushwaha S, Bahl S, Bagha AK, Parmar KS, Javaid M, Haleem A, et al. Significant applications of machine learning for COVID-19 pandemic. *J Ind Integr Manag* 2020;5(4).
- [21] Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. *American Physiological Society Bethesda, MD*; 2020.
- [22] Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* 2020;135:109864.
- [23] Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and bi-LSTM. *Chaos Solitons Fractals* 2020;140:110212.
- [24] Chandra R, Jain A, Chauhan DS. Deep learning via LSTM models for COVID-19 infection forecasting in India. 2021, arXiv preprint arXiv:2101.11881.
- [25] Saleh N, Bell D, Sulaiman Z. Hybrid conceptual modeling for simulation: An ontology approach during COVID-19. In: *Winter simulation conference. IEEE*; 2021.
- [26] Saleh N, Bell D. Ontology derived conceptual modeling for simulation. In: *Operational research society simulation workshop.* 2020.
- [27] Burgos-Salcedo J. A comparative analysis of clinical stage 3 COVID-19 vaccines using knowledge representation. 2021, <http://dx.doi.org/10.1101/2021.03.07.21253082>, MedRxiv, Cold Spring Harbor Laboratory Press.
- [28] Morgan AL, Woolhouse ME, Medley GF, Van Bunnik BA. Optimizing time-limited non-pharmaceutical interventions for COVID-19 outbreak control. *Philos Trans Royal Soc B* 2021;376(1829):20200282.
- [29] Furati K, Sarumi I, Khaliq A. Fractional model for the spread of COVID-19 subject to government intervention and public perception. *Appl Math Model* 2021;95:89–105. <http://dx.doi.org/10.1016/j.apm.2021.02.006>.
- [30] Jilani MZMB, Tucker A, Swift S. Simultaneous modelling and clustering of visual field data. In: *2016 IEEE 29th international symposium on computer-based medical systems. IEEE*; 2016, p. 213–8.
- [31] Jilani MZMB, Tucker A, Swift S. An application of generalised simulated annealing towards the simultaneous modelling and clustering of glaucoma. *J Heuristics* 2019;25(6):933–57.
- [32] Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts. In: *International conference on formal concept analysis. Springer*; 2009, p. 314–39.
- [33] Davey BA, Priestley HA. *Introduction to lattices and order.* Cambridge University Press; 2002.
- [34] Kuznetsov SO, Poelmans J. Knowledge representation and processing with formal concept analysis. *Wiley Interdiscip Rev: Data Min Knowl Discov* 2013;3(3):200–15. <http://dx.doi.org/10.1002/widm.1088>.
- [35] Terabe M, Washio T, Motoda H, Katai O, Sawaragi T. Attribute generation based on association rules. *Knowl Inf Syst* 2002;4(3):329–49.