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

SALEH, Nurul, AB GHANI, Hadhrami and JILANI, Zairul

Citation:

SALEH, Nurul, AB GHANI, Hadhrami and JILANI, Zairul (2022). Defining factors in hospital admissions during COVID-19 using LSTM-FCA explainable model. Artificial Intelligence in Medicine: 102394. [Article]

Artificial Intelligence In Medicine

Defining Factors in Hospital Admissions during COVID-19 using LSTM-FCA Explainable Model

--Manuscript Draft--

Manuscript Number:	AIIM-D-21-01759
Article Type:	Research Paper
Keywords:	Long Short-Term Memory (LSTM) Formal Concept Analysis (FCA) COVID-19 Hospital Admissions
Corresponding Author:	Nurul Saleh Universiti Malaysia Kelantan Pengkalan Chepa, Kelantan MALAYSIA
First Author:	Nurul Saleh
Order of Authors:	Nurul Saleh
	Hadhrami Ab Ghani
	Zairul Jilani
Manuscript Region of Origin:	Asia Pacific
Abstract:	Recent outbreaks of COVID-19 caused by SARS-CoV-2 infection that started in Wuhan, China, has quickly spread worldwide. The current situation has contributed to a dynamic rate of hospital admissions. A global effort by Artificial Intelligence (AI) and Machine learning (ML) community to develop solutions to assist COVID-19-related research has escalated ever since. Despite overwhelmed efforts from the AI and ML community, many machine learning-based AI systems have been designed as black boxes. In this paper, we propose a framework that utilizes Formal Concept Analysis(FCA) to explain a machine learning technique of Long-short Term Memory (LSTM) upon a dataset of hospital admission due to COVID-19 in the United Kingdom. The novelty of this work lies in increasing the transparency in decision-making using ML to optimize their use in various real-world applications. Both FCA and LSTM 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 can lead to new research to curb the pandemic.

1. **Title:**
   Defining Factors in Hospital Admissions during COVID-19 using LSTM-FCA Explainable Model

2. **Authors and Affiliations:**

   i. Nurul Saleh, Universiti Malaysia Kelantan, Malaysia
   ii. Zairul Jilani, Sheffield Hallam University, United Kingdom

3. **The problem and objective addressed in the paper:**
   LSTM algorithm is used to demonstrate factors affecting hospital admission in time-series prediction. The variety of ML algorithms and the black-box nature of how ML model works makes model interpretability a fundamental element in understanding the model. Hence Formal Concept Analysis (FCA) is used to explain the ML model.

4. **The essence of the approach:**
   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 decision-making, it is essential to describe and understand how the predictions are made while defining and mitigating bias. The mathematical formalism of Formal Concept Analysis (FCA) is used to explain the ML model. The study conducted in this paper was conducted upon a hospital admission due to COVID-19 dataset in the United Kingdom. The novelty of this research lies in increasing the transparency in decision-making to optimize their use in various real-world applications.

5. **Major novel contributions:**
   The LSTM-FCA approach uses the performance of time series prediction in finding contributing factors to hospital admissions. The significant variables are then explained using association rules derived from Formal Concept Analysis (FCA).

6. **Journal or conference papers from authors with significant overlap with this study:**
   i. Simultaneous Modelling and Clustering of Visual Field Data (Jilani, M.Z.M.B., Tucker, A. & Swift, 2016): This paper proposed a novel model-based technique namely Simultaneous Modelling and Clustering (SMC) in finding associated variables in a high-dimensional dataset of Glaucoma.
   ii. An Application of Generalised Simulated Annealing towards the Simultaneous Modelling and Clustering of Glaucoma (Jilani, M.Z.M.B., Tucker, A. &Swift, 2019): The paper extended the model-based (SMC) technique using the continuous optimisation method Generalised Simulated Annealing to improve the search of clusters in glaucoma dataset.
   iii. Hybrid Conceptual Modeling for Simulation: An Ontology Approach During Covid-19 (Saleh.N & Bell.D, 2021). This study used Formal Concept Analysis to derive important elements in domain knowledge to support simulation modeling for COVID-19
   iv. Ontology Derived Conceptual Modeling for Simulation (Saleh. N & Bell. D, 2020). This study deployed Formal Concept Analysis to derive important relationships in Emergency Department for process mining.

Highlights

Defining Factors in Hospital Admissions during COVID-19 using LSTM-FCA Explainable Model

Nurul Saleh,Zairul Jilani

- Explainable Model
- Long-Short Term Memory
- Formal Concept Analysis

Defining Factors in Hospital Admissions during COVID-19 using LSTM-FCA Explainable Model

Nurul Saleh[a], Hadhrami Ab Ghani[a] and 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

Recent outbreaks of COVID-19 caused by SARS-CoV-2 infection that started in Wuhan, China, has quickly spread worldwide. The current situation has contributed to a dynamic rate of hospital admissions. A global effort by Artificial Intelligence (AI) and Machine learning (ML) community to develop solutions to assist COVID-19-related research has escalated ever since. Despite overwhelmed efforts from the AI and ML community, many machine learning-based AI systems have been designed as black boxes. In this paper, we propose a framework that utilizes Formal Concept Analysis(FCA) to explain a machine learning technique of Long-short Term Memory (LSTM) upon a dataset of hospital admission due to COVID-19 in the United Kingdom. The novelty of this work lies in increasing the transparency in decision-making using ML to optimize their use in various real-world applications. Both FCA and LSTM 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 can lead to new research to curb the pandemic.

1. Introduction

The novel corona virus (COVID-19) was discovered for the first time in Wuhan, China, in November 2019 and has spread around the world at an accelerated rate. COVID-19 is a contagious disease that causes severe acute respiratory syndrome. The first UK COVID-19 cases were reported on 31 January 2020. Within three months, daily cases rose sharply to more than 33,000 cases. Other countries that had a spike number in daily cases during the early period were the US, Brazil, Italy, Spain, and Iran. The figure terrified the globe, and on 11 March 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, hygiene lifestyle, i.e., hand washing and disinfecting surfaces, and lockdown 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 opened 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 decisions that affect people's health, safety, economic well-being, and other aspects of life, it's essential to describe and understand how the predictions are made while defining and mitigating bias. This is a difficult topic to address because of the variety of machine learning algorithm types and the nature of how machine learning model training works, yet model interpretability has become a fundamental element in making model predictions understandable [20, 10]. Hence Formal Concept Analysis (FCA) is used to explain the ML model. This paper is organised into the following sections: Section 1 outlines the review of previous work , motivation, and objectives. Section 2 discusses the method used in this work. We present out results in Section 3, and discuss the results in Section 4. Finally, we conclude our work in Section 5.

1.1. Exploration and Review

In fighting the pandemic, data has recorded that the first UK mass vaccination program started in early December 2020. [5] reported that when 50% of the adult population has been vaccinated, death tolls are reduced by 95% and hospital admissions are reduced by 80%. [26, 13] 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 lock-down 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[7]. 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 [9] 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, which the rate increased with age. While in Iran, a study on the hospitalization and death rate among patients with multiple sclerosis has discovered that the rate of hospitalization was 25% far more than the general population [21]. School re-opening and hospi-

nurulizrin.ms@umk.edu.my (H.A. Ghani)
ORCID(s):

talization in the US have not resulted an increase in COVID-19 hospitalizations and interestingly, the virus spread among school staffs, but not among students [11]. Furthermore, study on sociodemographic, clinical and laboratory factors on admission associated with COVID-19 mortality in hospitalized patients is conducted to identify associations between baseline characteristics on hospital admission and mortality in patients with COVID-19 in Spain [19].

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 and also to understand the virus and further predict the upcoming issue [16]. Recent studies indicate that elderly and fragile people are affected by COVID-19, but it has also claimed many young lives. [1] 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 RMSE errors of 34.83 and 45.70 for short and long-term prediction [4]. Another study [24] has used various regressors in predicting new cases, deaths, and recoveries of COVID-19 suggests that LSTM is the second best method amongst others. [3] 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 implement mathematical formalisms to the discovery and representation of causal relations to support domain knowledge [23, 22]. Formal Concept Analysis (FCA) is used by [2] to uncover connections between vaccines and other attributes and proposes a rational strategies to design vaccination schemes to curb the COVID-19 pandemic.

In other study [18] explains a deterministic approach 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 by [8] 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 [20] - domain knowledge is an essential part of explainability.

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, our work aims to demonstrate a more transparent and interpretable machine learning model.

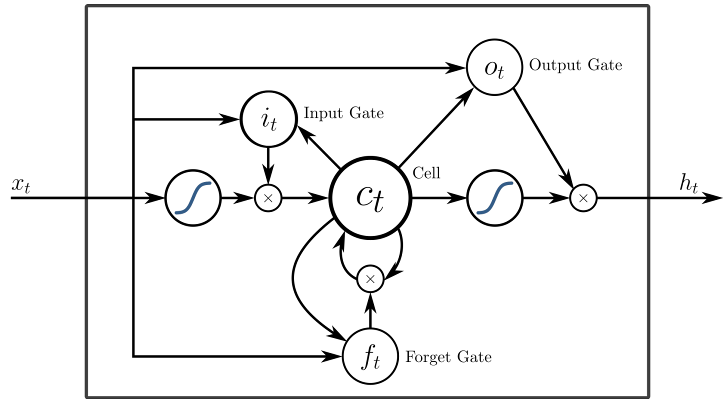


Figure 1: LSTM Components Diagram

1.3. Objectives

The purpose of this study is to find contributing factors of hospital admissions in the UK due to COVID-19 using forecasting method as well as proving the model-based approach. And 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. To explain the model, Formal Concept Analysis (FCA) mathematical model is used in addition to interpret domain knowledge and defining rules associating to the knowledge, further explanation in Section 2.4. LSTM, which was introduced by [12] in 1997, is a Recurrent Neural Network (RNN) based architecture that is widely used in natural language processing and time series forecasting. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate (Figure 1).

Empirical experiments are carried out using LSTM, and FCA is used to support the LSTM results. The hypothesis underlying this research is that the less error rate LSTM prediction (of hospital admissions) with the dataset and independent variables, the more significant are the variables in contributing factor to the 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 [14] [15].

Therefore, a few sets of experiments extensively investigated on the combination of the dependent variables that

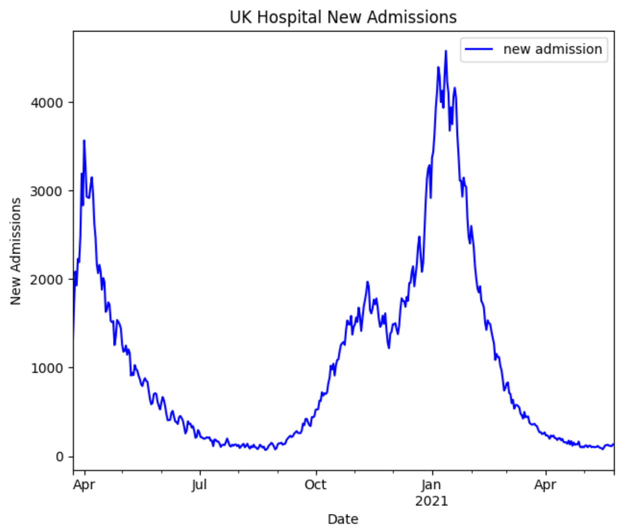


Figure 2: UK Hospital Admissions

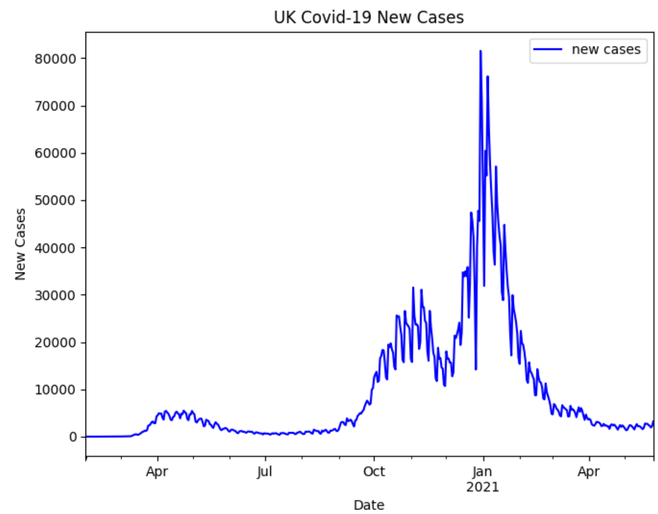


Figure 3: UK COVID-19 New Cases

could predict the target variable with less root means square error (RMSE).

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

We strategized investigation into a few sets of experiments and observed the empirical experiment results before making any judgement toward the hypothesis. Each experiment was run 25 times to obtain consistency 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 it 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 include the UK government and Institute for government UK organisation's. In total, there were 428 records of the dataset used in the experiment. They (total of 7 variables) include new admissions as the predicted variable, total cases, new cases, seasons, national lockdown, number of people have received first dose vaccine, and second vaccine. The variables are coded in Table 1. Figure 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, we come out with statistical values of admissions and new cases by seasons (Table 2 and 3). From figure 4 and 5 it clearly show that high numbers for both admissions and new cases during the winter season.

2.2. LSTM Experiment

LSTM networks are parameters dependent such as optimisers, number of epochs, number of batches, and data partitioning. As finding at the optimum parameters that best predict the dataset is out of the scope of our study, we set the following parameter values in the experiments: Optimiser: Adam, Epoch: 100, Batch size: 35.

Six independent variables are used in this work. We research for the significant variables that highly contribute the number of hospital admission in the UK using LSTM. From these variables (Table 1), there are 63 variables' combina-

Table 1
List of Variables

Variable Code	Variable Name	Variable Type
DV	new admissions	dependent variable
IV1	total cases	
IV2	new cases	independent variable
IV3	seasons	
IV4	national lockdown	
IV5	first dose	
IV6	second dose	

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

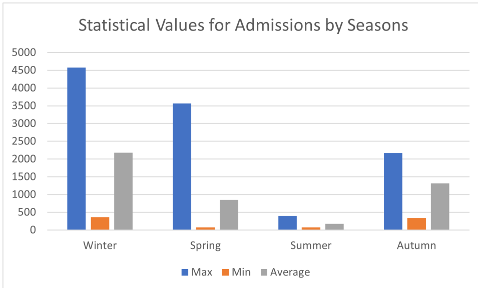


Figure 4: Statistical Values for Admissions by Seasons

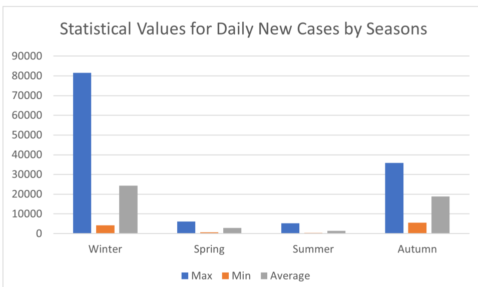


Figure 5: Statistical Values for Daily New Cases by Seasons

Table 4
Univariate LSTM Results

Mean	Min	Max	Std. Dev.
56.06839	54.90627	57.73614	0.851513

tions 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 time series of the data (stationarity and non-stationarity). The result of the test with 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 the input data has unit root. This indicates that the time series of hospital admissions dataset has trend and seasonality effect. Owing to this discovery, we include seasons variable in our investigation.

We also ran preliminary experiments of LSTM for univariate time series forecasting (with 25 samples) to test the method on our dataset. Table 4 and Figure 6 show the results of the univariate LSTM experiments (the best RMSE value - iteration 18th). Based on our initial experiment results, we set out Residual Mean Square Root (RMSE) tolerance to 55. Within our formal multivariate LSTM experiments, we only accept the variables combinations with RMSE values that less than the tolerance value.

2.4. Formal Concept Analysis

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

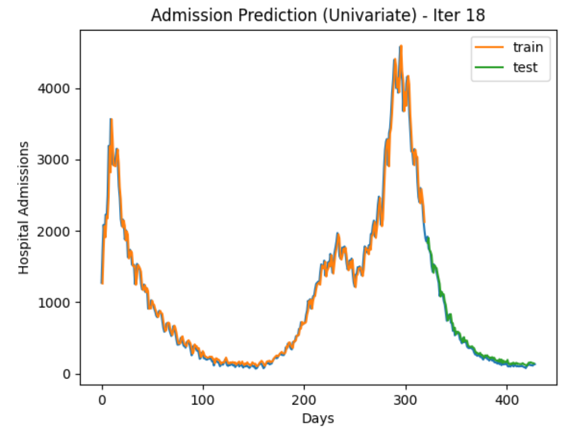


Figure 6: LSTM Univariate Time Series Forecasting

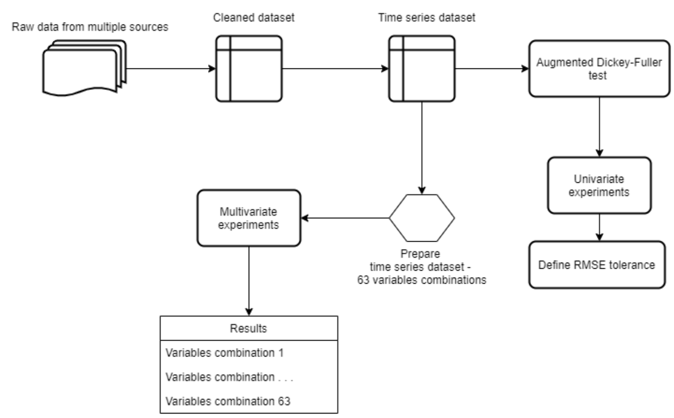


Figure 7: High-level LSTM Experiment

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

developed as a mathematical paradigm for concept formalisation and conceptual reasoning by [27]. FCA examines the relationships between a group of objects and their properties as stated by[6]. 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[17]. 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 Figure 8). The second set of output is the hierarchical relationships of objects exist in the domain - *concept lattice* (See Figure 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:
  Intent (X') = y $\in Y' \mid \forall y \in Y'$, (x,y) $\in R$
  Extent (Y') = x $\in X' \mid \forall x \in X', (x, y) \in R$
  For X' $\subseteq X$, Intent (X') is the attributes owned by all objects of X', and 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 2.2** Formal Concept
  A Formal Concept C in a context is a pair (X', Y') that satisfies Y' = Intent (X') and X' = Extent(Y')
  i.e., C is a Formal Concept $\Leftrightarrow$ for X' $\in CandY' \in C$, Extent(Intent(X')) = X',
  and symmetrically, Intent(Extent(Y')) = Y'.

- **Definition 2.3** 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 admission due to COVID-19. This exploration manages to show the relationships between agents' behavior when dealing with COVID-19 pandemic. Data consist 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 into if there is an increase from a day-to-day basis. In a cross-table, associating an object to the attributes created a concept hierarchy that can be visualized using the concept lattice. Figure 8 shows a cross-table where formal contexts are mapped with ConExp software using the hospital admission data. The cross-table describes formal context existed as per described in **Definition 2.1**. The little circles in Figure 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 mapping of formal context into a table, named Cross-Table. (See Figure 8) is conducted according to the main aim of this study which is to define the factors in hospital admission during COVID-19.

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		
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
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			

Figure 8: The Formal Context of the Hospital Admission

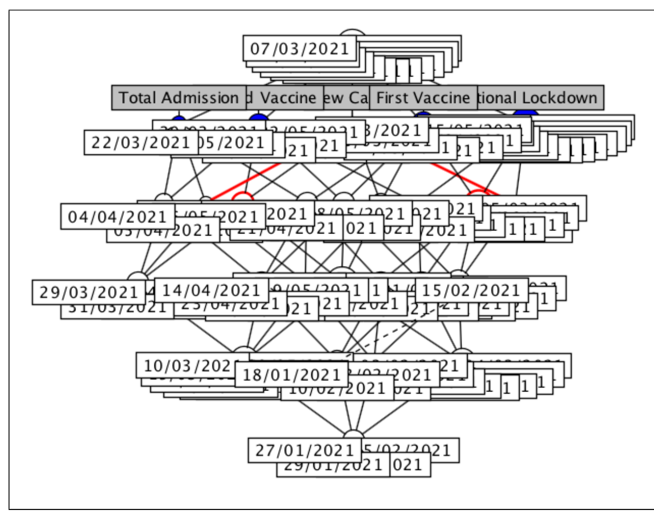


Figure 9: The Concept Lattice of the Hospital Admission

First, the data ranging from January 11th, 2021 to March 6th, 2021 are selected because that was the first time when the vaccination program began. Multiple-valued data, are then transformed into single-valued data. The progress of *New Cases*, *First Vaccine*, *Second Vaccine* and *Total Admission* are compared from 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 rules we seek vary according to the aim of the study and the basic knowledge of the domain data are implicitly gained [25]. Other than that, the *concept lattice* as depicted in Figure 9

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

explained the hierarchical relationship of all the established concepts of the domain.

3. Results

As discussed in the experiment and method, the LSTM experiment was run for each of the variables' combinations (with sampling 25 runs). As we set our tolerance of RMSE to 55, there are 7 variables' combinations with less than the tolerance value presented in our final results (Table 6). Meanwhile, Table 3 shows the results (descending order) for variables combinations with the least significant in predicting hospital admissions (high values of RMSE). Despite not in the lowest list, our all six variables combination experiment also has high RMSE values 463.306, 752.771, and 174.457 for mean, maximum, and minimum respectively with standard deviation 137.981.

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

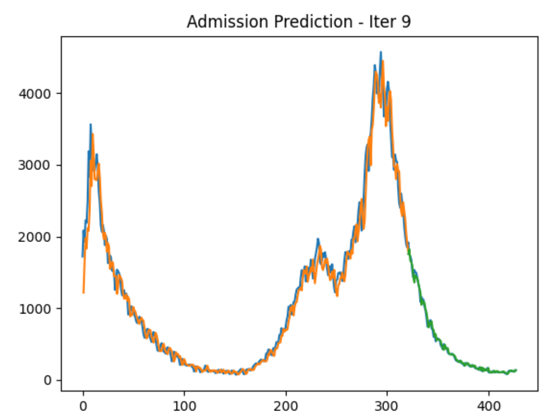


Figure 10: LSTM Multivariate Time Series Forecasting with the Best RMSE

3.1. Association Rules

From conceptual exploration approach conducted, the dependencies between the attributes, i.e. attribute implications or association rules are generated using ConExp. The total of 9 rules are generated to show the relationships between attributes existed 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%)
  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

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

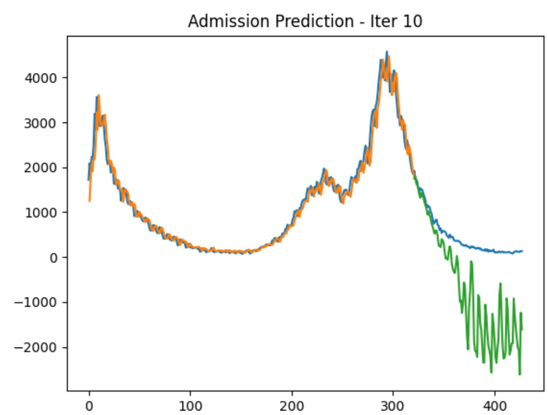


Figure 11: LSTM Multivariate Time Series Forecasting with the Highest RMSE

- **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

Our LSTM experiments resulted 7 best variables combinations in predicting hospital admissions in the UK. From these results, it clearly indicates three variables have strong association towards our target variable -*National lockdown*, *New cases* and *First Vaccine*. The first vaccine variable coincides with [5] [13] where vaccination programme in the UK reduces hospital admissions, and a single dose vaccine is effective in preventing hospital admissions. We 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 our defined tolerance RMSE. Surprisingly, the seasons variable is not listed as one of the variables that best in LSTM prediction. Nevertheless, we found that seasons and first vaccine are the best pairing in our LSTM experiment with RMSE values 65.736, 76.656, and 52.631 for mean, maximum, and minimum respectively (standard deviation: 5.769). On the contrary, second vaccine is the least significant variable in our LSTM prediction as this variable presence in the bottom 5 variables combinations.

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 contributed to new cases and hospital admission in the UK between January 11th, 2021 - March 6th, 2021 as vaccines rollout progresses and lockdown imposes 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 imposes *Lockdown* link to increases in *First Vaccine* rollout. From both rules, we have deduced that *Lockdown* and *First Vaccine* have a strong implication in number of cases and total admission in UK hospitals. The target variables from the LSTM result generated - *National Lockdown*, *New Cases* and *First Vaccine* and rules generated by FCA - *National Lockdown*, *New Cases*, *First Vaccine* and *Second Vaccine* have a strong coorelation with total admission number due to COVID-19 in hospitals in the UK.

5. Conclusion

From the study, the utility of methods explained, both LSTM and FCA are feasible in finding association variables and generating rules or hypothesis in the data. We employed LSTM, a deep learning approach to forecast the factors impacting admission due to COVID-19 and FCA method of attribute exploration to develop rules or relationship between the attributes. The innovative 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. We 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 bring new insights that may result in the development of new research aimed at controlling the pandemic.

Based on the promising RMSE values in our LSTM prediction and FCA discoveries a number of research opportunities can be considered in future work. The LSTM parameter values can be further explored to optimise the prediction. With the optimised prediction, a new set of significant variables or pattern could be found as we are interested to see how seasons in the UK impact the hospitalisations. 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). We noted that our experiments were run on a dataset which vaccines and seasons were observed less than six months. Another interesting future work is that to test our approach on datasets from other countries for the same target variable.

Acknowledgement

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

References

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

[2] Burgos-Salcedo, J., 2021. A comparative analysis of clinical stage 3 covid-19 vaccines using knowledge representation. medRxiv doi:10.1101/2021.03.07.21253082.

[3] Chandra, R., Jain, A., Chauhan, D.S., 2021. Deep learning via lstm models for covid-19 infection forecasting in india. arXiv preprint arXiv:2101.11881 .

[4] Chimmula, V.K.R., Zhang, L., 2020. Time series forecasting of covid-19 transmission in canada using lstm networks. Chaos, Solitons & Fractals 135, 109864.

[5] Cook, T., Roberts, J., 2021. Impact of vaccination by priority group on uk deaths, hospital admissions and intensive care admissions from covid-19. Anaesthesia 76, 608–616.

[6] Davey, B.A., Priestley, H.A., 2002. Introduction to lattices and order. Cambridge university press .

[7] Di Domenico, L., Pullano, G., Sabbatini, C.E., Boëlle, P.Y., Colizza, V., 2020. Impact of lockdown on covid-19 epidemic in île-de-france and possible exit strategies. BMC medicine 18, 1–13.

[8] Furati, K., Sarumi, I., Khaliq, A., 2021. Fractional model for the spread of covid-19 subject to government intervention and public perception. Applied Mathematical Modelling 95, 89–105. doi:https://doi.org/10.1016/j.apm.2021.02.006.

[9] Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., Prill, M., Chai, S.J., Kirley, P.D., Alden, N.B., et al., 2020. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—covid-net, 14 states, march 1–30, 2020. Morbidity and mortality weekly report 69, 458.

[10] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018. A survey of methods for explaining black box models. ACM computing surveys (CSUR) 51, 1–42.

[11] Harris, D.N., Ziedan, E., Hassig, S., 2021. The effects of school reopenings on covid-19 hospitalizations. National Center for Research on Education Access and Choice .

[12] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780.

[13] Iacobucci, G., 2021. Covid-19: Single dose of pfizer and oxford vaccines cuts risk of hospital admission by 80% in over 80s, data suggest.

[14] Jilani, M.Z.M.B., Tucker, A., Swift, S., 2016. Simultaneous modelling and clustering of visual field data, in: 2016 IEEE 29th international symposium on computer-based medical systems (CBMS), IEEE. pp. 213–218.

[15] Jilani, M.Z.M.B., Tucker, A., Swift, S., 2019. An application of generalised simulated annealing towards the simultaneous modelling and clustering of glaucoma. Journal of Heuristics 25, 933–957.

[16] Kushwaha, S., Bahl, S., Bagha, A.K., Parmar, K.S., Javaid, M., Haleem, A., Singh, R.P., 2020. Significant applications of machine learning for covid-19 pandemic. Journal of Industrial Integration and Management 5.

[17] Kuznetsov, S.O., Poelmans, J., 2013. Knowledge representation and processing with formal concept analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 3, 200–215. doi:10.1002/widm.1088.

[18] Morgan, A.L., Woolhouse, M.E., Medley, G.F., Van Bunnik, B.A., 2021. Optimizing time-limited non-pharmaceutical interventions for covid-19 outbreak control. Philosophical Transactions of the Royal Society B 376, 20200282.

[19] Rivera-Izquierdo, M., del Carmen Valero-Ubierna, M., R-delAmo, J.L., Fernandez-Garcia, M.A., Martinez-Diz, S., Tahery-Mahmoud, A., Rodriguez-Camacho, M., Gamiz-Molina, A.B., Barba-Gyengo, N., Gamez-Baeza, P., et al., 2020. Sociodemographic, clinical and laboratory factors on admission associated with covid-19 mortality in hospitalized patients: A retrospective observational study. PLoS One 15, e0235107.

[20] Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable machine learning for scientific insights and discoveries. Ieee Access 8, 42200–42216.

[21] Sahraian, M.A., Azimi, A., Navardi, S., Ala, S., Moghadasi, A.N., 2020. Evaluation of the rate of covid-19 infection, hospitalization and death among iranian patients with multiple sclerosis. Multiple sclerosis and related disorders 46, 102472.

[22] Saleh, N., Bell, D., 2020. Ontology derived conceptual modeling for simulation, in: Operational Research Society Simulation Workshop.

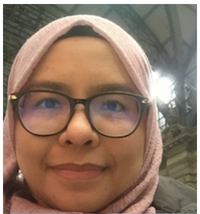
[23] Saleh, N., Bell, D., 2021. Hybrid conceptual modeling for simulation: An ontology approach during covid-19, in: Winter Simulation.

[24] Shahid, F., Zameer, A., Muneeb, M., 2020. Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. Chaos, Solitons & Fractals 140, 110212.

[25] Terabe, M., Washio, T., Motoda, H., Katai, O., Sawaragi, T., 2002. Attribute generation based on association rules. Knowledge and information systems 4, 329–349.

[26] Torjesen, I., 2021. Covid-19: First doses of vaccines in scotland led to a substantial fall in hospital admissions.

[27] Wille, R., 2009. Restructuring lattice theory: an approach based on hierarchies of concepts, in: International conference on formal concept analysis, Springer. pp. 314–339.



Nurul Saleh is a Lecturer at the Department of Data Science at Universiti Malaysia Kelantan, Malaysia. Her research areas are Data Science, Software Engineering, Ontology and Modeling & Simulation.

Mohd Zairul Jilani is a Lecturer in Computer Science at Sheffield Hallam University, United Kingdom. His research areas are mainly in machine learning i.e clustering, classification, optimisation high-dimensional datasets.

Hadhrami Ab Ghani is a Lecturer at the Department of Data Science at Universiti Malaysia Kelantan, Malaysia. His research areas are mainly in Machine Vision, Data Science, Artificial Intelligence and Advanced Communications.

Declaration of interests

☐The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

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