

How to Cite:

Mann, S., Anand, M., Aggarwal, A., Wadhawan, K., & Jain, H. (2022). Time series forecasting for number of hospitalizations caused due to COVID-19 in the United Kingdom. *International Journal of Health Sciences*, 6(S2), 14113–14124. <https://doi.org/10.53730/ijhs.v6nS2.8706>

Time series forecasting for number of hospitalizations caused due to COVID-19 in the United Kingdom

Dr Suman Mann

Associate Professor, Department of Information Technology, Maharaja Surajmal Institute of Technology, Janakpuri, Janakpuri, New Delhi, 110058, India
Corresponding author email: sumanmann@msit.in

Manan Anand

Department of Information Technology, Maharaja Surajmal Institute of Technology, Janakpuri, Janakpuri, New Delhi, 110058, India
Email: manan.olympian23@gmail.com

Akshit Aggarwal

Department of Information Technology, Maharaja Surajmal Institute of Technology, Janakpuri, Janakpuri, New Delhi, 110058, India
Email: akshitx10@gmail.com

Kartik Wadhawan

Department of Information Technology, Maharaja Surajmal Institute of Technology, Janakpuri, New Delhi, 110058, India
Email: wkartik71@gmail.com

Harshit Jain

Department of Information Technology, Maharaja Surajmal Institute of Technology, Janakpuri, New Delhi, 110058, India
Email: jain2008.harshu@gmail.com

Abstract---Pandemics and epidemics have plagued humanity throughout history. The modern world faced one such devastating disease in 2019 called Coronavirus. As the world is still trying to recover from Coronavirus, an epidemic that might not see its end anytime soon. This study focuses on analyzing and predicting the future hospital admissions that arise due to Covid-19. For this study, the choice of country is the United Kingdom. The data has been procured from reliable internet sources to carry out all necessary experiments. The data set contains daily hospitalisations due to Covid-19 in the United Kingdom. To carry out the time-series forecasting, the predictive model is built using a special Recurrent

Neural Network, also known as Long Short-Term Memory (LSTM) . The final model was built using a Stacked LSTM to predict the number of hospitalisations due to Covid-19 that may arise in the United Kingdom for the next twenty days from the last day of the dataset used. The results of this study show a clear indication that a spike in the number of hospitalisations may arise in the upcoming days.

Keywords--time series forecasting, COVID-19, coronavirus, deep learning, LSTM.

Introduction

Pandemics and epidemics have plagued humanity throughout history. From The Black Death to the Spanish Flu, these diseases have claimed millions of lives over the years. The modern world faced one such devastating disease in 2020 called Coronavirus or Covid-19 caused by the SARS-CoV-2 virus. The origin of this deadly virus is in Wuhan city of Hubei province, China [1]. Covid19 was declared a pandemic on 11 March 2020 by the World Health Organization (WHO). The pandemic brought forth a time of massive unrest and anxiety that was unheard of in the modern world. It brought the world to a standstill, and everyone was forced inside their homes. The virus claimed millions of lives as no vaccination was available in the early stages of the pandemic. The scarcity of hospital beds and equipment around the world was horrific to witness [21]. Countries and cities with high population density like China, India, the United States of America, the United Kingdom and Italy were affected most by the pandemic [17]. A sudden spike in the number of cases in a particular area can completely overwhelm the hospitals in that area. This has been seen worldwide over the past two years in countries, some of which are stated above.

During the pandemic, several factors have had an impact on whether Covid-19 cases will rise or fall in a country. These factors include the effectiveness of vaccines over time, government policies, and different variants of the virus. 2021 has seen many new variants of Covid-19 and the effectiveness of vaccines such as Pfizer, Moderna and Oxford-AstraZeneca against these new variants has been a big question among researchers. Lawmakers around the globe have responded to these sudden surges in cases by stocking up on equipment and resources in advance, many countries around the world also authorized the use of foreign vaccines in the wake of a new spike in Covid-19 cases, but the uncertainty about the actual number of patients that will need professional care in the future undermines any efforts made by the governments. Recent studies have also claimed that SARS-CoV-2 has resulted in many possible mutations, namely Alpha, Beta, Gama, Delta and Omicron.

Countries with high population density like India suffered most during subsequent waves of Covid-19. The government was not prepared for such a mammoth surge in cases around the country. The strain on hospitals around India, many such countries with underdeveloped health infrastructure and high population density, were unforeseen and people were left without basic healthcare

facilities like - oxygen cylinders, ventilators, basic medicines and numerous lives were lost in this period of time.

Many researchers have tried to predict future spikes in the number of Covid-19 cases, which have been used to estimate the shock and severity of a supposed spike in cases(also termed as Waves) that people and governments could expect. There is another angle with which this could have been looked at, and that is the number of Hospitalizations each day that occurred specifically due to Covid-19. This is similar to predicting the mortality rate of ICU patients [8]. This way, looking at the problem can help analyse the number of Hospitalizations that may arise in the future. In this study, a deep learning-based predictive approach to forecast Covid-19 hospitalizations in the United Kingdom is proposed. The dataset taken is a day-by-day count of hospital admissions due to Covid-19. This is critical for creating a time series (using daily data). RNN based long short term memory (LSTM) variants to design the proposed mode are utilized to build a working model. The variant of LSTM used in this study is Stacked LSTM.

The remnant of the study is organized as follows:

Section 2 puts forward a comprehensive literature review of previously available studies regarding the Corona Virus and its impact in the United Kingdom, Deep Learning for Time-Series Analysis, and why LSTM(Long Short-Term Memory Networks) are better suited for Time-Series data and their advantages over Recurrent Neural Networks. In Section 3, different steps taken to carefully curate a working model are talked about, which includes, but are not limited to, the Dataset, any preprocessing techniques which were used to normalize said data and the proportion in which the dataset is distributed for further evaluation of the model metrics, this section further discusses the architecture of the proposed model which produced the best results. Section 4 discusses the outcomes of the proposed experiment with all possible evaluation metrics. Section 5 concludes this study with discussions over the limitations, possible scope of improvement along with most important learnings which came from this study.

Related Literature

The novel coronavirus outbreak was declared a pandemic by the WHO on 11 March 2020. Ever since it has infected millions of people in 188 countries. The studies of Zhang R. show that airborne transmission especially through the nascent aerosols is highly virulent and represents the dominant route for the transmission of the disease [18]. Multiple patients were left with lasting effects after recovering from Covid-19. The most common aftereffects observed post recovery from Covid-19 were hair fall, lung diseases, joint pain and sleep problems [10].

A series of data that occurs in successive order over a period of time is called time series. Previous studies by Sourabh Shastri show that time series forecasting of Covid-19 data produces encouraging results [15]. The study mentioned above prognosticates the confirmed cases of Covid-19 in India and the United States of America. The focal point of this study was to analyse and predict hospital admissions due to Covid-19, this was facilitated by Recurrent Neural Networks

(RNN) and Long Short Term Memory(LSTM) Networks. RNNs are a class of neural networks that allow previous outputs to be used as inputs while having hidden states, the research by Zachary C. Lipton on Recurrent Neural Networks shows that they are especially useful when working with sequential data like a time series, the reason for this conclusion is that by feeding the previous output as the next input RNNs are able to remember the information of the recent past [9].

In this study[7][20], the author discusses and further shows with experiment how RNN can store a lot of previous information for a short time which makes training more complicated for certain applications. But, since Recurrent Neural Networks suffer from the Vanishing Gradient problem, RNN based models are unable to model long term dependencies which is why LSTM was introduced. Long Short-Term Memory units (LSTM) are a special type of RNN that deal with the vanishing gradient problem encountered by traditional RNNs. Traditional RNNs can only remember recent information but when the interval between a piece of information when it is encountered and when it is needed becomes large RNNs become obsolete. This conclusion was shown in the studies of Steven Elsworth[6].

In their study [2][22][23], A Comparative Study and Analysis of Time Series Forecasting Techniques, the authors discuss how time series forecasting plays an important role in business decisions, even in today's time of machine learning and deep-learning heavy methodologies. The importance of predicting future data based on historical data is expressed, with an assumption that some similarities within data will repeat themselves over time. Time series forecasting is an old problem, and the purpose of this study according to the authors, is to present previously developed methods for different time-series datasets and to compare their results in a collective study. The authors have compared results for the following methods best suited for Time-series forecasting: Regression method, Stochastic approach, Soft computing techniques namely RNNs, and Fuzzy logic forecasting.

The research article by Zhou and Kan [21], examines the impact of the year-long measures due to Covid-19 in the U.K. on different Socio-economical groups of people living in the U.K. during the pandemic. This study focuses on the impact created by the three lockdowns imposed by the government of the U.K., and how different measures such as social-distancing and work from home policies had a different scale of impact on people from various social groups.

Proposed Model

Data distribution

In this paper, the dataset of confirmed hospital admissions due to SARS-CoV-2, This way, looking at the problem which is commonly known as the novel coronavirus or Covid-19, in the United Kingdom is taken into consideration. The Dataset of the United Kingdom is taken from Our World in Data Org. Our World in Data is an online forum that focuses on tackling problems the world has faced over centuries and working towards solving them collaboratively. Dataset considered in this study consists of data from 28-03-2020 to 14-10-2021 in

chronological order with corresponding new hospital admission on that particular day are shown in Table 1.

Table 1. Hospital bed occupancy on a Daily-basis in the UK.

Date	Beds Occupied
28-03-2020	8278
29-03-2020	9525
30-03-2020	11093
31-03-2020	12597
01-04-2021	13637
02-04-2020	13875
03-04-2020	15546
04-04-2020	17502
05-04-2020	18855
06-04-2020	19493
07-04-2020	20234

The data shows hospital admissions in the United Kingdom due to Covid-19. The admissions being considered are new for new patients and do not include the previous days' patients. The study of new hospital admissions is crucial in today's world as a sudden spike in Covid-19 related cases can put a country's entire healthcare system in chaos. Thus, providing a way for predicting the number of hospital beds required in the future can be very useful. It opens the door for people to be better prepared for situations like these and can also be vital for procuring essential pieces of equipment beforehand.

The graphical representation of confirmed hospital admissions in the United Kingdom is given below. This representation is a necessary piece of information to visualize the ups and downs, and the in-general trends of hospital-bed occupations in the United Kingdom, which are almost synonymous to the various spikes in covid-19 cases which were observed throughout the year. As observed from the graphical representation of the hospitalisations due to Covid-19 in the United Kingdom, the trend of hospitalisations is random and has no standard order or pattern. But despite the randomness of the data, it is clear from its representation the rise and fall in the number of hospitalisations due to Covid-19 in the UK. This anomaly poses a special challenge as it gives very little information for the model to learn from. The most efficient way to analyze this data is a time series forecasting technique [2].

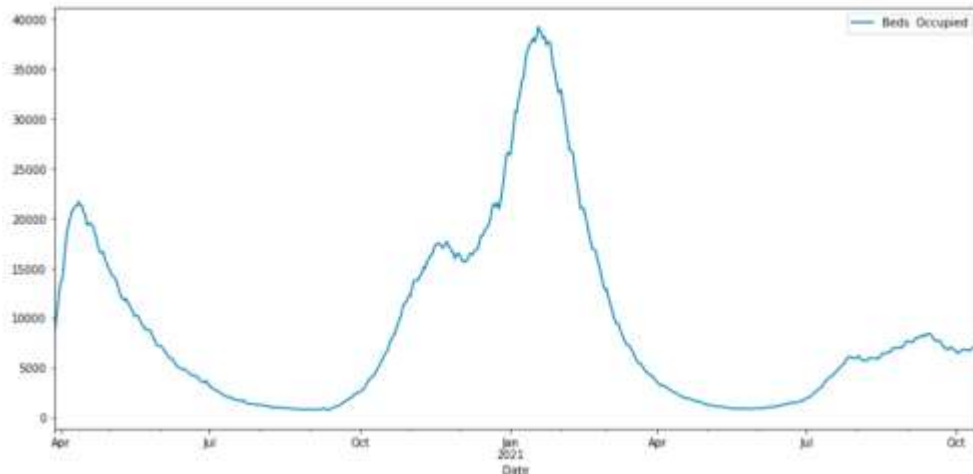


Figure 2. Graphical representation for the Hospital-bed occupancy due to Covid-19 in the UK

Data preprocessing

Data Preprocessing is a technique used to transform raw data into useful information that the machine learning model can understand. The advantage of Data Preprocessing is that it improves the generalization of the model. Data preprocessing generally includes the following steps[14]:

- (1) Data cleaning: Data cleaning involves dealing with missing data and noisy data.
- (2) Data Transformation: Data Transformation is carried out to modify data in a suitable format.

In this experiment, the normalization is done using MinMaxScaler, a library of sci-kit-learn in Python. The MinMaxScaler performs a mathematical operation on each of the values in the dataset[11]. For each value in the dataset, it subtracts the minimum value and then divides it by the range (i.e. the difference between the maximum and the minimum value in the data set).

The formula gives the transformation: $X_{std} = \frac{(X - X_{min})}{(X_{max} - X_{min})}$

After transformation, the dataset is split into a training set and test set. The training and test set are in the ratio 80: 20, respectively. This split is crucial as we need two different sets to train the model and test the results. The ratio of 80:20 is used because the larger the size of the training set the better the performance of the model will be.

Model Description

Recurrent Neural Networks, in short known as RNNs, are an extremely important class of the Deep-learning methodology that is generally used to find correlations

in a time series. Recurrent neural networks work best when the present state depends on previous information[7]. It is a special type of neural network where the output of the previous state is fed as an input to the next state. Recurrent Neural Networks are most commonly developed and used for the tasks that fall under:

- (1) Sequence Generation
- (2) Sequence Labelling
- (3) Sequence Classification

RNN works very well with sequential data such as a time series because of its internal memory [4]. The internal memory allows it to remember important things from the information it receives as input [12]. This allows it to make accurate and precise predictions of the future [19]. These features allow RNN to form a deep understanding of sequential data and its context [8]. The Recurrent Neural Networks have a feedback-loop, the representation of the same, along with the interpretation of an uncoiled RNN can be viewed through Fig 3. and Fig 4. respectively.

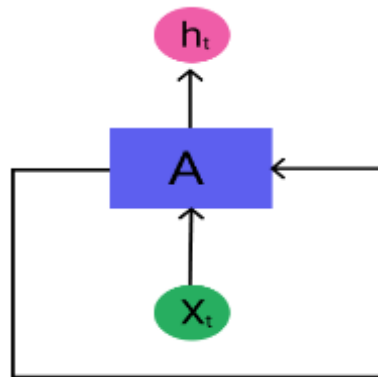


Figure 3. Graphical interpretation of a RNN cell with its feedback loop

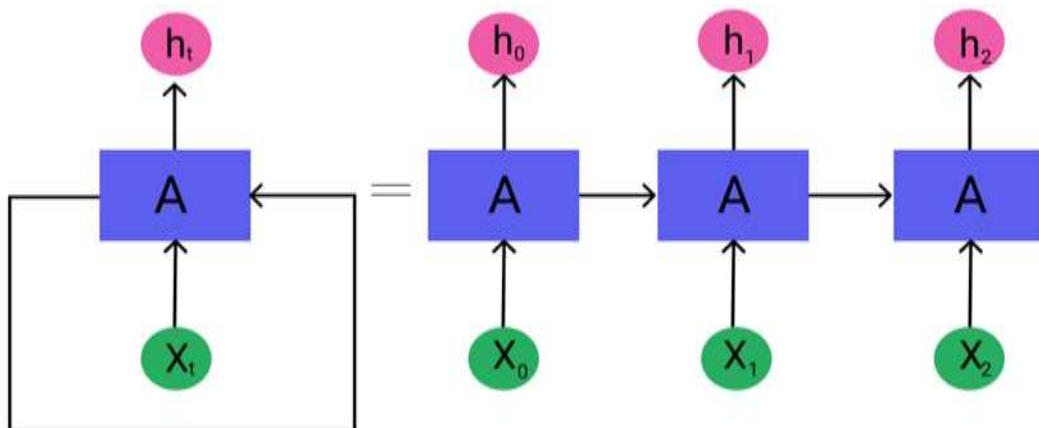


Fig 4. Graphical interpretation of an uncoiled Recurrent Neural Network

LSTM stands for Long Short Term Memory, and it is a special type of recurrent neural network. It is capable of learning order dependence in sequential data[13]. The problem with RNN is that as the gap between a piece of information and the place it becomes relevant or is needed to make a future prediction grows, it is unable to learn to connect the information. This is called the long-term dependency problem[16]. The solution to this problem is LSTM. The default behavior of LSTM is to remember information for a long period of time[9]. The internal architecture of an LSTM cell can be viewed in Fig 5. The LSTM architecture has many mutations, but for the purpose of this experiment, the choice of stacked LSTM gave the best results. A stacked LSTM is a result of vertically stacking multiple layers of LSTM to increase the model complexity[5]. And it is observed that deep RNNs perform better than shallower ones. A graphical representation of the stacked LSTM architecture can be viewed through Fig 6.

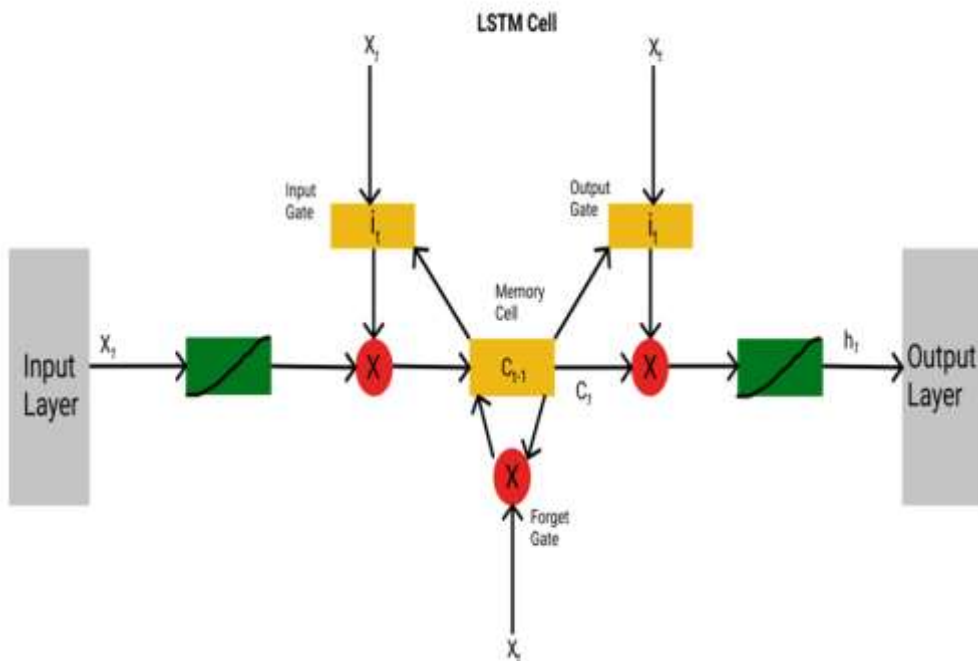


Figure 5. Graphical representation of the internal architecture of a single LSTM cell.

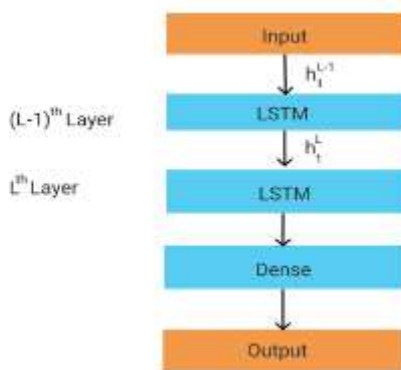


Figure 6. Graphical representation of the Stacked LSTM Architecture

Experiment and Results

The experiments were conducted on Google Colaboratory using python 3.0 with Python libraries like Pandas, Numpy and Keras. The experimental setup is based on a working environment having Intel(R) Core (TM) i5-9400 CPU @ 3.00GHz with 8 GB RAM under 64-bit Windows 10 Operating system. Time series forecasting of Covid-19 hospitalisations is modelled using Stacked LSTM. These models are used to recognise hidden patterns in time series data. During each epoch, the model first analyses the data for the first twenty days in the data frame and generates a prediction for the twenty-first day. It then considers this predicted output for the day and its actual value and skips the first day in the twenty-day window for further steps in the algorithm. Eventually, it creates a new twenty-day window. The model then repeats the same algorithm, generating output for the next twenty-first day (the twenty-second day from the start of the data frame) until it reaches the end of the data frame. The model was trained on 100 epochs, and the loss per epoch graph is given below. The model was trained on 100 epochs, and the loss per epoch graph can be viewed in Fig 8.

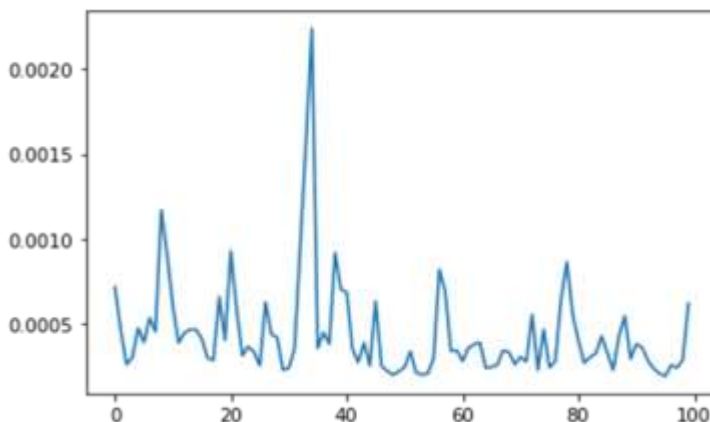


Figure 7. Loss per epoch plot

On running our model for future predictions, we found that the trend for hospital admissions in the United Kingdom due to Covid19 will rise in the next 20 days. The results of which can be viewed in the plot present as Fig 8.

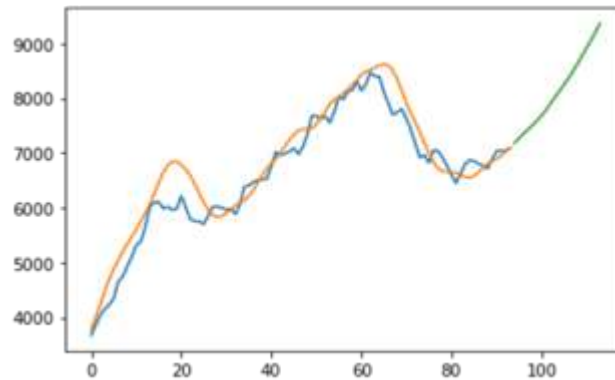


Figure 8. Model predictions for the next 20 days

Conclusion

The aim of this study was to build a Time-Series Forecasting model based on Deep Learning approach, namely the Long Short Term Memory Networks to predict the expected number of hospitalizations that may arise due to Covid-19 and its possible mutations. As can be seen from the experiment results, the hospitalisations due to covid-19 will rise for the next 20 days or so in the United Kingdom. Necessary steps must be taken in order to avoid any unpleasant situation. According to the forecasted findings, the number of hospitalisations in the United Kingdom due to Covid-19 could reach over 9,000 by the next 20 days. The United Kingdom has done a commendable job of vaccinating its population, but as seen in some other countries, Covid-19 cases have previously seen a rise in number[3]. The number may not be as alarming as seen earlier this year or last year, but it unquestionably cannot be ignored.

References

- [1] Andersen, K.G., Rambaut, A., Lipkin, W.I. et al, "The proximal origin of SARS-CoV-2". *Nat Med* 26; 450–452 (2020)
- [2] Athiyarath, S., Paul, M., Krishnaswamy, S (2020). "A Comparative Study and Analysis of Time Series Forecasting Techniques". *SN COMPUTER SCIENCE*. 1, 175.
- [3] Bajema, K. L., Dahl, R. M., Prill, M. M., et al (2021). "Effectiveness of COVID-19 mRNA Vaccines Against COVID-19-Associated Hospitalization - Five Veterans Affairs Medical Centers, United States", February 1-August 6, 2021. *MMWR. Morbidity and mortality weekly report*, 70(37), 1294–1299.
- [4] Donkers, T., Loepp, B., Ziegler, J., (2017). "Sequential User-based Recurrent Neural Network Recommendations". In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 152–160.

- [5] Du, X., Zhang, H., Van Nguyen, H. and Han, Z., 2017, September. "Stacked LSTM deep learning model for traffic prediction in vehicle-to-vehicle communication". In 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall) (pp. 1-5). IEEE.
- [6] Elsworth, S. and Güttel, S., 2020. "Time series forecasting using LSTM networks: a symbolic approach". arXiv preprint arXiv:2003.05672.
- [7] Hochreiter, S., 1998. "The vanishing gradient problem during learning recurrent neural nets and problem solutions". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02), pp.107-116.
- [8] Hooda, S. , Mann, S., 2020. "A Focus on the ICU's Mortality Prediction Using a CNN-LSTM Model". *International Journal of Psychosocial Rehabilitation*, 24(6), pp.8045-8050.
- [9] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). "A critical review of recurrent neural networks for sequence learning". arXiv preprint arXiv:1506.00019.
- [10] Pan, F., Ye, T., Sun, P., Gui, S., Liang, B., Li, L., Zheng, D., Wang, J., Hesketh, R.L., Yang, L. and Zheng, C., 2020. "Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19)". *Radiology*, 295(3), pp.715-721.
- [11] Patro, S. and Sahu, K.K., 2015. "Normalization: A preprocessing stage". arXiv preprint arXiv:1503.06462.
- [12] Rodrigues, P, Wiles J., Elman J.L., (1999) "A Recurrent Neural Network that Learns to Count, *Connection Science*", 11:1, 5-40, DOI: 10.1080/095400999116340
- [13] S. Hochreiter, J. Schmidhuber, "Long short-term memory, *Neural Computation* 9" (1997) 1735–1780.
- [14] Shalabi, L.A., Shaaban, Z., & Kasasbeh, B. (2006). "Data Mining: A Preprocessing Engine". *Journal of Computer Science*, 2, 735-739.
- [15] Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). "Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study". *Chaos, solitons, and fractals*, 140, 110227.
- [16] Sherstinsky A. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network". *Phys D*. 2020 doi: 10.1016/j.physd.2019.132306.
- [17] Wong DWS, Li Y (2020) "Spreading of COVID-19: Density matters". *PLoS ONE* 15(12): e0242398.
- [18] Zhang, R., Li, Y., Zhang, A. L., et al (2020) : "Identifying airborne transmission as the dominant route for the spread of COVID-19". *National Academy of Sciences*, 117, 14857-14863.
- [19] Zhao, Z., Nehil-Puleoa, K., Zhao, Y. "How well can we forecast the COVID-19 pandemic with curve fitting and recurrent neural networks". *medRxiv preprint 2020* 10.1101/2020.05.14.20102541.
- [20] Anish Batra, Guneet Singh Sethi, Suman Mann," Personalized Automation of Electrical and Electronic Devices Using Sensors and Artificial Intelligence—the Intelligizer System" *Computational Intelligence: Theories, Applications and Future Directions - Volume I, 2019, Volume 798*
- [21] Zhou, M., Kan, M. Y. (2021). "The varying impacts of COVID-19 and its related measures in the UK: A year in review". *PloS one*, 16(9), e0257286.
- [22] Dhruv Rathee, Suman Mann, "Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning" *International journal of computer application*, 2022, Vol 183, no 47, Doi: 10.5120/ijca2018918026.

- [23] Sakshi Hooda, Suman Mann “Examining the Effectiveness of Machine Learning Algorithms as Classifiers for Predicting Disease Severity in Data Warehouse Environments” *Revista Argentina de Clínica Psicológica* ,SCIE indexed,233-251(2020), ISSN0327-6716 DOI: 10.24205/03276716.2020.824
- [24] Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). The COVID-19 pandemic. *International Journal of Health Sciences*, 5(2), vi-ix. <https://doi.org/10.53730/ijhs.v5n2.2937>
- [25] Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2022). Post-pandemic health and its sustainability: Educational situation. *International Journal of Health Sciences*, 6(1), i-v. <https://doi.org/10.53730/ijhs.v6n1.5949>