Deep Learning via LSTM Models for COVID-19 Infection Forecasting in SAARC

Ajharul Islam
Department of Information and Communication Technology
Comilla University
Comilla-3506, Bangladesh
ajharrimon90@@gmail.com

Md. Sharif Hossen
Department of Information and Communication Technology
Comilla University
Comilla-3506, Bangladesh
sharif.5613@gmail.com

***Abstract*—Coronavirus is a profoundly irresistible infection that has almost incapacitated the worldwide economy. Its capacity to transmit human-to-human information plunges the planet into chaos. In this case, our goal in this work is to forecast the future circumstances of a new Coronavirus in order to reduce its impact, We've got presented a deep learning-based forecast of Covid-19 variants in Bangladesh as well as SAARC. The data sources it is taken into account the number of confirmed and Location of Covid-19. To create the suggested approach and predict and forecast the Covid-19 various variants of covid-19, long short term memory (LSTM) variants based on recurrent neural networks (RNN), such as Stacked LSTM, are used. For all sample of datasets from all SAARC nations, Stacked LSTM forecasting behaviors of variants of Covid-19 with high accuracy and extremely low error. The upward/downward trend of anticipated Covid-19 instances is also visually shown, which can assist researchers and policymakers reduce mortality and morbidity rates by directing Covid-19 in the proper direction with awareness.**

Keywords—- COVID-19, Prediction, Deep learning, RNN, LSTM.

I. INTRODUCTION

Since its discovery in Wuhan, China, in 2019, the new coronavirus has spread around the world. The coronavirus strain is extremely contagious, affecting around 471M people worldwide. Till March 21, 2022, killing 6.08Mlives. This virus can cause mild, moderate, or severe symptoms, depending on the person. Causes respiratory distress syndrome. Covid19 vaccine was first unavailable. The only way to get out of that situation is to it was necessary to detect social distance in order to stop the virus from spreading. Large-scale testing to discover positive cases, and isolation of affected individuals.

Several vaccines were later developed to confer acquired protection to the coronavirus. The immunizations, however, are rendered ineffective due to the covid-19 mutation. Several countries have succeeded in controlling the illness by implementing the appropriate steps and administering the right amount of vaccinations, however, Bangladesh has struggled to manage the problem. The enormous population and insufficient vaccination supplies have made it extremely difficult to return to normalcy. As a result, it's critical to keep the infection from spreading [6]. Due to the virus's severe effects and ease of replication, in the worst affected districts, the national government has imposed lockdown, if not the whole country. It would be altogether more fruitful experiencing the same thing on the off chance that the quantity of affirmed cases could be anticipated somewhat early.

As quickly as possible, apply lockdown should be implemented in all impacted areas. As a result, the spread will narrow and the number of traders will fall. Of instances that have been confirmed. The overall number of affected persons has surpassed 1,951,363 as of March 30, 2022[1]. Bangladesh has a population of 1,309,910 people. Bangladesh has an average of 14,886 people. As of March, daily new cases had reached a peak of 98 percent. 03, 2022 these figures are rapidly increasing and will continue to do so [4]. To have a negative impact on people's lives, hospitals, and the country's currencies [3].

Early aware of Covid19 as they occur is the most notable period for effective disease management. Advances in Artificial Intelligence (AI) research now make it possible to create aware of Covid19. The advantages of deep learning are that it extracts features from The Kind of Covid19, Daily number of Covid19 Case, Percentage of Case, and Total Case in Each Country. For better aware of Covid19 in the research follow forecasting using LSTM .LSTM likes, special kind of its architecture is an artificial recurrent neural network (RNN), which is a common deep learning model.

A. Background and Motivation

The COVID-19 flare-up has brought about countless passing the whole way across the world, and it represents an obscure danger to general wellbeing, food frameworks, and the climate. A great many people are in danger of falling into serious destitution because of the pandemic's business and social burden, and the quantity of individuals who are undernourished, which is at present projected to be around 690 million, could move to 132 million before the finish of the period. So, as a result SAARC (South Asian Association

for Regional Cooperation) as well as our Country directly or indirectly affected by COVID-19[4]. Nowadays, this virus is more concern for every country's economy. By, using modern media we all know that COVID-19 has more variants, some of the variants are really harmful to all of us. So, if we find out these variants and analyze the forecast of specific variants this will be the best option for every country's government.

B. Objectives

- Increasing awareness of Covid19 and advances in Forecasting followed by deep learning.
- Recurrent neural network (RN'N) architecture is often used for sample of data forecasting in this research project.
- This project uses the concept of e behavior of temporal correlations, applying deep learning models such LSTM (Long short term memory).
- These model are used for Forecasting, feature extraction and unlabeled data and finding hidden structures in it.
- Using by analyzing data graph find out most concern variant in SAARC.

C. Problem Definition

- To find out situation of all SAARC Countries how much affected by Covid19 except Afghanistan and Bhutan.
- Find out all variants of covid19 from dataset using data preprocessing.
- To find which variant of Covid-19 widely spread for each country
- Manual detection of harmful variant of Covid-19 using data graphs.
- Hence, it is required to develop computational methods that will make forecasting about harmful variant of Covid-19.

II. LITERATURE REVIEW

A. Comparative Studies

Authors	Goal	Future
[1] Sourabh Shastri, Sachin Kumar, Paramjit Kour, Vibhakar Mansotra, Kuljeet Singh,	The authors of this work employed deep learning models to show Covid19 predictions for India and the United States. Both nations' confirmed Covid19 cases and mortality cases are taken into account. The limited availability of Covid-19 data makes projecting time series data difficult. The predictions are modeled using the extension of the Recurrent Neural Network (RNN) as a Long Short Term Memory (LSTM) cell and its variants, including Stacked, Bi-Directional, and Convolutional LSTM. For all four datasets, With error rates ranging from 2.0 to 3.3 percent, the convolutional LSTM model performs better than the other two models. In our experimental comparison investigation, The Stacked LSTM model has the worst performance. The number of confirmed cases and fatal cases in both countries would rise over the course of the following month, according to the authors' forecasts.	In the future, authors may said they look at the entire economy loss in different sectors after the conclusion of Covid19 and devise a strategy to recover it, which would aid nations in regaining their economic footing. We intend to estimate potential Covid19 instances in other countries, as well as verifying Covid19 aerosol transmission.
[2] Hafsa Binte Kibria, Oishi Jyoti, Abdul Matin	More study is being done to predict the COVID-19, as the outcome of this tragedy has a direct impact on our daily life. Various angle of prediction models have been introduced by many number of researchers. Authors employed a variety of techniques in this project. AR (8, 0), MA (0, 7), ARMA (8, 7), ARIMA (8, 1, 7), and rolling forecast are all options. To anticipate the spread of COVID-19, researchers used ARIMA (8, 1, and 7) models. Bangladesh is experiencing a pandemic. The most accurate forecasting methodology for predicting In Bangladesh, the daily confirmed cases trend was revealed to be ARIMA (8, 1, 7). The number of confirmed cases produced each day was approximated.	More study is needed to determine which forecasting approach or model is the most accurate for various scenarios throughout the world.

	Using this model over the following month.	
[3] Faheem Aslam, Ruqia Khan, Maira Aslam, Yasir Tariq Mohmand Tahir, Mumtaz Awan,	The current analysis is based on COVID-19 verified case predictions for 14 days beginning July 1, 2020. It use the Automatic ARIMA model to anticipate daily COVID-19 case increase using data from the subcontinent from March 15, 2020 to June 30, 2020. According to the study's findings, the number of cases would increase in Pakistan and India while remaining stable in Bangladesh. The findings might help policymakers prepare for future crises caused by coronavirus's most lethal repercussions.	Authors said by case of prediction In the future, governments should devise counter-strategies to cope with economic challenges, health crises, and offer the essential assistance to front-line workers and the impoverished.
[4] Milad Haghani, Michiel C.J. Bliemer, Floris Goerlandt	The bibliometric characteristics of these works, as well as those addressing coronaviruses in general, are discussed by the authors in this succinct review. By doing a scoping analysis of the COVID-19 literature, we also emphasize the key safety-related aspects that these studies have thus far addressed. Our results demonstrate that, in addition to medical and clinical aspects like vaccine and treatment safety, issues like patient transport safety, occupational safety of healthcare professionals, bio-safety of laboratories and facilities, social safety, food safety, and particularly mental/psychological health and domestic safety have received the most attention from the scientific community in relation to the COVID-19.	More scientific studies conducted during and after the present outbreak might better equip human civilizations and assist them in preparing for future global health catastrophes. In this sense, the current effort intends to emphasize the safety factors associated with the COVID-19 pandemic, on which some research has previously been published. The authors believe that this will encourage other academics to discover gaps in the extent of the safety dimensions offered, in order to create a more diversified picture of the current pandemic's safety consequences.

B. Scope of the problem

The existing method for analyze Covid-19 variants and forecast more concern variant in SAARC using LSTM methods not find organize research [4]. Here the steps followed for new scope are,

- Proper covid19 variant analysis
- Recognizing healthy nations using data graph
- Identify more concern variant
- Forecasting analysis these methods, as mentioned above, are traditional and time-consuming.

III. SYSTEM MODEL

A. Proposed Work

The project will use Unsupervised Machine Learning approach. Every Machine Learning study start with data collection. Data basically used in Unsupervised Learning are unlabeled. For the proposed system we need to collect from different country which are share by every country's specific organization. In this proposed system we collect data from Kaggle. Once collect the dataset we should apply algorithms in our proposed model LSTM will be best fit forecasting [2]. So, in real world we know that data are basically unorganized most of the case, so we need preprocess, clean and augment this dataset. So, by using Python, TensorFlow library and tools data is being preprocess, clean and augment.

Once we have preprocess dataset for forecast then using LSTM we will find the forecast for most concern variants [2].

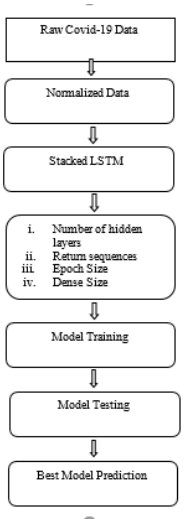

Figure 1: Proposed methodology framework

I. Dataset Acquisition

In order to perform forecasting and data find harmful variant, we need data to train the neural network although proposed model is follow unsupervised learning but proposed model follow self-train methods. For this purpose, a publicly available dataset named "Covid-variants" which present in kaggle. Kaggle is an open source data center. This specific dataset have 100417 data from different countries but, for our proposed research project we take only 5472 sample which data satisfied only SAARC. Now in below some snapshot of dataset in given.

Header	Definition
Date	Time From Begining of Covid
Location	All Countries affected by Covid
Variant	The Kind of Covid
Number Of Sequence	Daily number of Covid Case per kind?
Perc Sequences	Percentage of Case
Probationary	Sometimes the character was given probationary status as an Avenger, this is the date that happened
Number Of Total	Total Case In Each Country

Figure 3: Data definition

II. Data Preprocessing

When we produce new data based on changes to current data, we call it data augmentation. By adding fair changes to data in our training set, we are producing new enhanced data. We know that, Data preprocessing is a step in the data mining and analysis process that transforms raw data into a format that computers and machine learning can comprehend and assess. Real-world data in the form of text, images, video, and other formats is messy. In addition to having errors and inconsistencies, it typically lacks a consistent design and is unfinished. Machines interpret input as 1s and 0s because they desire to handle data in a nice and orderly fashion. Calculating structured data, such as whole integers and percentages, is so straightforward. Unstructured data, such as text and photos, must be cleaned and prepared before being analyzed.

III. Proposed Stacked LSTM Model

A Recurrent Neural Network (RNN) variant called Long Short Term Memory (LSTM) is utilized to address RNN's drawbacks. As shown in Fig. 3.4, LSTMs are capable of learning long-term dependencies by substituting memory cells for the hidden layers of an RNN. Several gate units, including input gate (it), output gate (ot), and forget gate (ft), are paired with the activation function to model LSTMs and learn the behavior of temporal correlations. In Eqs. (1)— (5) The LSTM cell's operating method is also mathematically stated [2].
Now,

$$i_t = \sigma\left(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i\right) \quad (1)$$

$$f_t = \sigma \cdot W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f \Sigma \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc} x_t + W_{hc} h_{t-1} + b_c\right) \quad (3)$$

$$o_t = \sigma\left(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_t + b_o\right) \quad (4)$$

$$h_t = o_t \tanh(c_t) \quad (5)$$

Where i, f, c, o are input gate, forget gate, memory cell, and output gate, and i, f, c, o are input gate, forget gate, memory cell, and output gate, respectively. From memory cells to gate units, Wxi, f, c, and o are diagonal weight matrices. One LSTM variations are utilized to conduct out testing in this work, and they are detailed in the following sections [2].

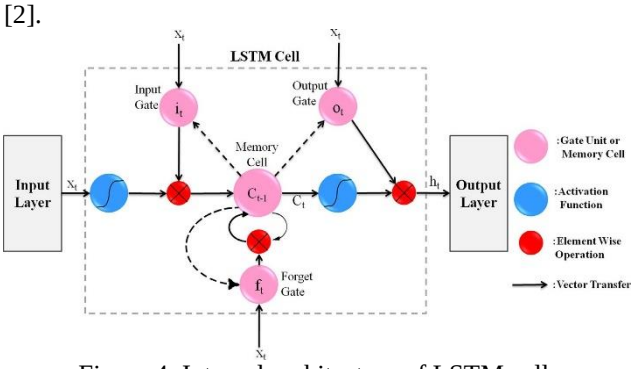

Figure 4: Internal architecture of LSTM cell.

B. Requirement Analysis

Software requirement: Using Google Colaboratory as an experimental environment. Also use built in google colab GPU support which model is Tesla k80.
Hardware requirement: A computer, a strong internet connection.Machine Learning: Machine Learning (ML) is the process of analyzing data, learning from it, and then making predictions about fresh data using algorithms.
Deep Learning: Machine learning has a branch called deep learning. It employs algorithms that are based on the structure and function of neural networks in the brain. Deep learning, like machine learning, is primarily based on algorithms that learn from data. These algorithms or models, on the other hand, are based on the structure and function of neural networks in the brain. As a result, this learning will take place in either supervised or unsupervised settings.
Supervised learning: when a deep learning model learns and infers from previously labeled data, this occurs.
Unsupervised learning: when the model learns and infers from unlabeled data, this happens. The goal of a deep learning model is to determine if a dataset serves better accuracy. We had about 5,472 sample of covid19 variant stored on a disk, and we were to give these sample to our model to learn here, we know that LSTM is follow self-supervised learning. If the model were learning in a supervised form, each sample of data would be forecasting for covid19.

- If the model were learning in a supervised form, the data sample would be labeled.
- If the model were learning in an unsupervised form, the data sample would be labeled.

Because the algorithms are based on the structure of *recurrent* neural networks, LSTM models are referred to as artificial neural networks.

IV. EXPERIMENTS

A. Find out Harmful variant

The graphs below indicate which Covid-19 version is most often used in each nation. Here the graphs about 19 kinds of Covid19 variants.

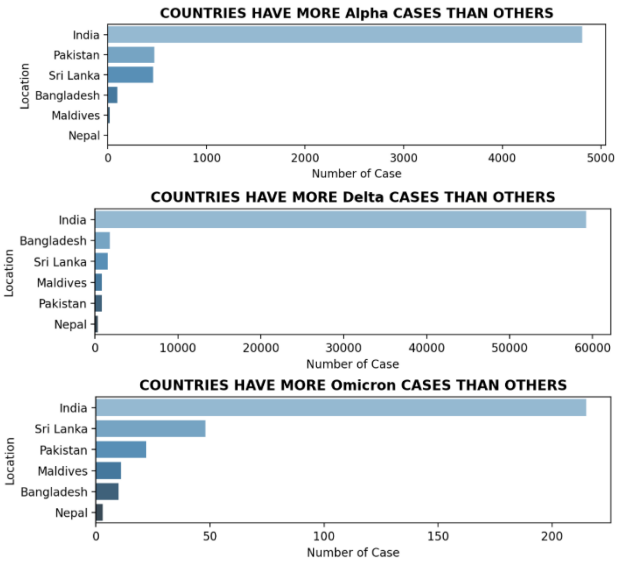

Figure 5: Covid19 case analysis graphs

From, the Figure 5 we can see that Omicron variant of covid19 is very much concern for SRRAC as well as our Country.
Now, the following graph will represent how increasing Omicron variant of covid19.

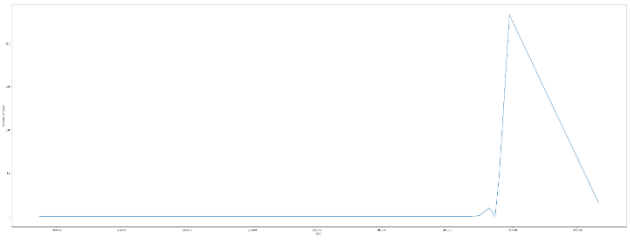

Figure 6: Increasing Omicron variant analysis graph

Here, Figure 6 indicated that, Omicron variant is increasing during December, 2021 to January, 2022. So now, we should give concern Omicron variant in SARRC as well as Bangladesh.

B. Experiments

Tensorflow, Pandas, Numpy, Keras, and Python 3.0 are some of the open source libraries used in the testing, which are run using Google Collaborator. The experimental setup is built on a working environment with a 64-bit version of Windows 10 pro, an Intel(R) Core(TM) i3-5 CPU operating at 2.00GHz, and 8 GB of RAM. One type of recurrent neural network (RNN) utilized to mimic time series forecasting of Covid-19 datasets is the stacked LSTM. This approach is used to anticipate future Covid-19 case values by revealing hidden patterns in time series data. Models are provided historical datasets based on Covid-19 number of cases and location, as shown in Fig 4. I. LSTM models' hyper-parameters are carefully picked and fine-tuned. The workings on one variations, as well as parameter selection, are explained more detail m subsequent sections. Figure 3.1 depicts the suggested methodology's framework. For all SARRC nations, train models of all three RNN versions are utilized to anticipate one month future sequence values of Covid- 19 cases.

I. Stacked LSTM model:
In terms of model depth and complexity, stacked LSTM differs from simple RNN models. We create input acceptable for the model by examining three lag structures and a number of properties of one before working with the designed model. We used MinMaxScaler to normalize the input data, dividing it into a 75 percent training set and a 25% testing set. The stacked LSTM model is used to forecasting Covid-19 instances utilizing dataset [2].

The model was built using a two-layer stacked LSTM structure with 100 hidden neurons. To guarantee that each time step of input data gets an LSTM output, the model must employ return sequences. Additionally, a substantial layer is added to the model to link every neuron to the one after it, creating a fully linked network. The model is evaluated using RMSprops as an optimizer and mean square error as a loss function. Our model is trained for 100 epochs with a validation split of 0.25 and a verbose of 1. The findings of the stacked LSTM on all SAARC nations are presented individually in the results section [2].
Stacked LSTM model is a sequential model. It constructs with as many layers which is given below.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 100)               40800

 dense (Dense)               (None, 1)                 101

=================================================================
Total params: 40,901
Trainable params: 40,901
Non-trainable params: 0
```

Figure 7: Model Summery

Here, my LSTM model has low loss and validation loss which are 0.394 and 0.0028.
Now graphical representation would be likes,

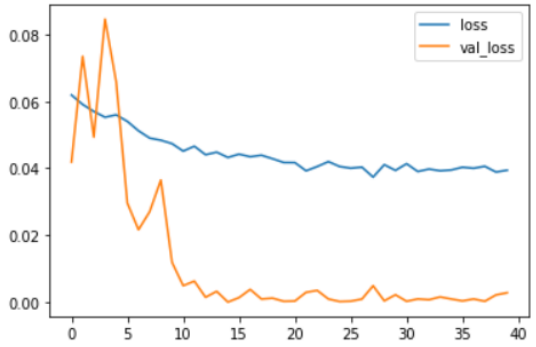

Figure 8: Comparisons of loss and validation loss

C. *Analysis and Forecast*

In this this research project my main goal is analysis about harmful variant of covid19 and forecast about one harmful variant by follow any single person, any single nation from SAARC as well as Bangladesh will be aware of Covid19 harmful variant and people get healthy life [4].Now the forecasting graph will be likes,

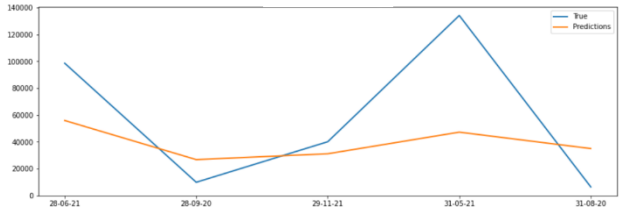

Figure 9: Forecast harmful variant graph

This graph indicates that his difference between actual result and predictions about forecasting, here prediction is more likely to actual result so, it can be consider my forecasting is will be considerable for SAARC nations.

V. CONCLUSION AND FUTURE WORKS

The new coronavirus (Covid-19) has nearly brought the world to a standstill due to its catastrophic pandemic characteristics. Because this outbreak might linger for months, people all around the world must learn to cope with it and incorporate Covid-19 protections into their regular routines. All vital and non-essential goods were kept up for a long period. In our research, we used deep learning models to demonstrate this. Except for Afghanistan and Bhutan, Covid-19 forecasting is available for all SAARC countries. The confirmed Covid-19 cases and mortality cases from all nations are considered 12). Projecting time series data is problematic due to the restricted availability of Covid-19 data.

To mimic the predictions, the Recurrent Neural Network (RNN) is expanded as a Long Short Term Memory (LSTM) cell. Stacked LSTM is one of its variants. The best model is chosen based on the prediction error rate, which is calculated using Mean Absolute Percentage Error (MAPE). In our experimental comparison research, the Stacked LSTM model had the greatest performance.

My projections show that in the next month, there was a rise in the number of cases in SAARC nations. Statistical information and techniques are used to demonstrate the validity of this investigation. When employing real-time Covid-19 data, our best-proposed model's rate of dependability—which yields cutting-edge findings—is sufficient. To summarize, this is the first comparative case study of all SAARC nations to shed light on factors impacting Covid- 19 and project future timeframes for all SAARC nations. Our results will help all countries take the essential safeguards to avoid falling to the Covid-19 outbreak. After the end of Covid-19, we may be able to look at the overall economy loss in different sectors and design a strategy to recover it, which would benefit nations in restoring their economic footing. We plan to determine the number of Covid-19 cases in additional countries and confirm Covid-19 aerosol transmission.

REFERENCES

[1] "Bangladesh: WHO Coronavirus disease (COVID-19) dashboard with vaccination data," *Who.int*. [Online]. Available: https://covid19.who.int/region/searo/country/bd/. [Accessed: 30-Mar-2022].

[2] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study," *Chaos Solitons Fractals*, vol. 140, no. 110227, p. 110227, 2020.

[3] H. B. Kibria, O. Jyoti, and A. Matin, "Forecasting the spread of the third wave of COVID-19 pandemic using time series analysis in Bangladesh," *Inform. Med. Unlocked*, vol. 28, no. 100815, p. 100815, 2022.

[4] F. Aslam, T. M. Awan, R. Khan, M. Aslam, and Y. Tariq Mohmand, "Prediction of COVID-19 confirmed cases in Indo-Pak sub-continent," *J. Infect. Dev. Ctries.*, vol. 15, no. 3, pp. 382–388, 2021.

[5] "Bangladesh COVID - Coronavirus statistics - worldwide," *Worldometers.info*. [Online]. Available: https://www.worldometers.info/coronavirus/country/bangladesh/. [Accessed: 30-Mar-2022].

[6] M. Haghani, M. C. J. Bliemer, F. Goerlandt, and J. Li, "The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review," *Saf. Sci.*, vol. 129, no. 104806, p. 104806, 2020.