Southeast University (SEU)

Department of Computer Science & Engineering (CSE)

CSE4000: Research Methodology

Impact of Temperature to Predict Covid-19 Transmission in Bangladesh: A Comparative Study by Using Deep Learning

A dissertation submitted to the Southeast University in partial fulfillment of the requirement for the degree of
B. Sc. in Computer Science & Engineering

Submitted By:

Md. Rakibul Hassan Rony	Bishajit Chakraborty	Md. Nur Ur Rahman
ID: 2018000000012	ID: 2018000000010	ID: 2015000000083

Supervised by

Md. Mijanur Rahman

Assistant Professor
Department of CSE
Southeast University

January, 2022

Letter of Transmittal

January 04, 2022

The Chairman,
Department of Computer Science & Engineering,
Southeast University,
Banani, Dhaka.

Through: Supervisor, Md. Mijanur Rahman

Subject: Submission of Research Report on the topic **"Impact of Temperature on Covid-19 Transmission in Bangladesh: A Comparative Study Using Deep Learning"**

Respected Sir,
It gives a great pleasure to submit our research report **"Impact of Temperature on Covid-19 Transmission in Bangladesh: A Comparative Study Using Deep Learning"** for the course Research methodology (CSE4000). Working on this topic was a privilege for us. This research was completed by following your instructions and meeting Southeast University's requirements.

This report has been written with the utmost sincerity and effort. As part of our degree requirements, we ask for your approval of this research report.

Thank You so much for your kind cooperation.

Sincerely Yours

Name: Md. Rakibul Hassan Rony
Id: 2018000000012
Batch: 48
Program: CSE

Name: Bishajit Chakraborty
Id: 2018000000010
Batch: 48
Program: CSE

Name: Md. Nur Ur Rahman
Id: 2015000000083
Batch: 39
Program: CSE

Supervisor

Md. Mijanur Rahman
Assistant Professor
Department of CSE
Southeast University

Certificate

This is to certify that the research paper **"Impact of Temperature on Covid-19 Transmission in Bangladesh: A Comparative Study Using Deep Learning"** is the evidence of research work for the partial fulfilment of the requirement for a degree of Bachelor of Science in computer science and engineering from Southeast University.

This paper was completed under my supervision, and the task was completed effectively, according to the records.

Authors

Name: Md. Rakibul Hassan Rony
Id: 2018000000012
Batch: 48
Program: CSE

Name: Bishajit Chakraborty

Supervisor

Md. Mijanur Rahman
Assistant Professor
Department of CSE
Southeast University

Id: 2018000000010
Batch: 48
Program: CSE

Name: Md. Nur Ur Rahman
Id: 2015000000083
Batch: 39
Program: CSE

Abstract

Coronavirus disease 2019 (COVID-19) has already shaken the world by killing and infecting millions of people worldwide due to the lack of effective vaccines and medicines. Researchers around the world are struggling to find more appropriate solutions to stop this pandemic. Different predictions are made based on the trend data like new cases, deaths, recoveries and etc. In Bangladesh, no significant evidence was identified that could explain how temperature affects COVID-19 transmission. So, it can be reevaluated by considering latest datasets with more significant features. This study focuses on how COVID-19 transmission in Bangladesh varies with temperature. Here, we try to predict the upcoming per day new infections using a Deep Learning-based model Long Short Term Memory (LSTM) network for both short and long-term. The model also compared with other country's data, particularly on India, Saudi Arabia, Italy, and Malaysia. The results demonstrate that when using temperature data to predict the short-term, the model's accuracy improves slightly for Bangladesh but differs for other nations. Although using temperature data in long-term prediction, the model's accuracy remains unchanged. This study try to answer the question regarding correlation between

proximity to temperature and the Covid-19 transmission in Bangladesh. Further studies are needed to establish causal relationship and develop preventative measures.

Acknowledgements

First and foremost, we extend thanks to Almighty Allah, whose might has enabled us to successfully conclude our research.

After that, we'd like to express our gratitude to our esteemed supervisor, **Md. Mijanur Rahman,** Assistant Professor at Southeast University's Department of Computer Science Engineering. The research behind it would not have been possible without our supervisor's unwavering support. From our first experience with machine learning to the final copy of this paper, his enthusiasm, competence, and meticulous attention to detail have been an inspiration and kept us on track.

We'd also want to thank **Shahriar Manzoor** sir, Associate Professor and Chairman of Southeast University's Department of Computer Science and Engineering, for appointing Md. Mijanur Rahman as our supervisor.

We'd also want to express our gratitude to the entire faculty at Southeast University's Department of Computer Science and Engineering for their support and encouragement. Finally, we reward our efforts and hard work day after day with right responsibility with pleasure and appreciation.

Finally, we'd want to express our gratitude for all of your kindnesses, which were critical to the completion of our research.

Table of Contents

Abstract………………………………………………………………………… I

Acknowledgement……………………………………………………………… II

Contents……………………………………………………………………….. III

List of Figures…………………………………………………………………. IV

Chapter 1: Introduction……………………………………………………….. **8**

1.1 Statement of the problem……………………………………….……...8

1.2 Significance of the Study.………….………………….……………..8

1.3 Objectives……………………….……………………………………. 10
1.4 Report Outline ………………………………………………………..10
Chapter 2: Literature Review………………………………………………… **11**
Chapter 3: Methodology……………….………………………………………… **12**
3.1 Covid-19 Dataset…………………………………………………...12
3.2 Data Preprocessing ...…………………………………………….....13
3.3 Model Details …………………………………………………………14
3.3.1 Deep learning…...…………………………………………...14
3.3.2 Recurrent Neural Network(RNN)…………………………14
3.3.3 Long-short-term-memory (LSTM)………………….……..14
3.3.4 Training and Testing…………….……………………………16
3.3.5 Prediction Accuracy Measurement……………….……….17
Chapter 4: Implementation..…………………………………………………. 18
4.1 Data Collection…………………………………………………… 18
4.2 Data Preprocessing...…………………………………………… 18
4.2.1 Dataset Snapshot ……………………………………19
4.2.2 Dataset Null Check……………………………………………19
4.2.3 Min Max Scaler……………………………………………..20
4.2.4 Building the Model……………………………………...21
Chapter 5: Results and Discussion.………………………………………. **22**
Chapter 6: Conclusion ………………………………………………………..**30**
6.1 Conclusion …………………………………………………………… 30
6.2 Future Scope ………………………………………………………. 30
Chapter 7: References

List Of Figure

Figure 1: Methodology Diagram

Figure 2: Snapshot of the dataset

Figure 3: Null checking

Figure 4: Min-Max Scaler

Figure 5: Building the model

Figure 6: Actual and predicted COVID-19 new cases(short-term) from 16 April to 26 April considering tempereture data.

Figure 7: Actual and predicted COVID-19 new cases(long-term) from 28 march to 28 April considering tempereture data.

CHAPTER 1

Introduction

1.1 Statement of the Problem:

Many people in Wuhan, China, became ill and died abruptly around the end of 2019. Chinese authorities found the cause on January 7, 2020, and it was given the temporary moniker "2019-n CoV"[1]. The novel virus was renamed as "Covid-19 virus" following its discovery. Within a short period of time, it infected people and quickly spread to every corner of the globe. Patients with Covid-19 initially had colds, coughs, fevers, and difficulty breathing. Still, some new symptoms have emerged, including diarrhea, headaches, a rash on the skin, darkening of the fingers, and chest pain, and this virus also directly affects the human lungs. . "On 30 January 2020, the World Health Organization (WHO) proclaimed the epidemic a Public Health Emergency of International Concern, and a pandemic on 11 March 2020." [2]. Because coronavirus is a brand-new virus, there was no vaccination available at first. Thus, developing countries have taken actions to combat coronavirus, including vaccine development, lockdown, social isolation, mask use, hand sanitizer use, and enforcing curfews. Simultaneously, specialists are examining a variety of possible explanations of the virus's spread, including lockdown, social isolation, temperature, seasonal influence, and immunization. Some countries have developed the Covid-19 vaccination and are now supplying to other countries. Rest of world is working on developing or importing Covid-19 vaccine. In Bangladesh, the first corona patient was found on 8 March 2020. The infection rate was initially low, but grew with time. When the infection rate started to grow, the government took a number of steps that had been suggested by various authorities and groups.

1.2 Significance of the Study:

Numerous studies have been conducted using machine learning (ML) methods to forecast Covid-19 cases in Bangladesh. The majority of machine learning algorithms require a tremendous quantity of data to train. Due to the fact that the Covid-19 dataset is only one year old, developing credible forecasting models to train on such a little amount of data is extremely difficult [4]. Deep learning (DL) has, on the other hand, proven to be the most successful approach for many predictive tasks in recent years. . DL networks have exploded in popularity among artificial intelligence (AI) methodologies, particularly when compared to traditional machine learning (ML) methods. DL models automate all processes of feature extraction, feature selection, and classification, in contrast to ML techniques [5]. With all of this in mind, it's high time to reevaluate the situation using the most up-to-date datasets and the most promising forecasting models. As far as the author is aware, no research study has been published on forecasting Covid-19 new cases in Bangladesh based on temperature data. We hope that our contribution will spur the government, decision-makers, and scholars in taking additional steps forward. At the moment, many machine learning and deep learning algorithms have been used to forecast Covid-19 transmission. However, since the outbreak began, it has been hard to train machine learning algorithms with appropriate data. Additionally, due to the short dataset, the researchers encountered difficulties determining the influence of temperature on Covid-19 transmission in Bangladesh. While this study has made progress in this area, the objective is not to achieve state-of-the-art performance but rather to investigate the effect of weather on the Covid-19 transmission rate.

This article presents a DL-based Recurrent Neural Network (RNN) architecture LSTM network for estimating the number of patients who may become infected with Covid19 in Bangladesh in the future days. Five impacted countries' statistics are analyzed in this study: Bangladesh, India, Saudi Arabia, Italy, and Malaysia. Furthermore, we have predicted both short-term (10 days) and long-term (30 days) infected cases for both considering temperature and without temperature data. Each layer of the LSTM network is composed of LSTM cells, followed by an activation function and dropout regularization. Each of the LSTM layers is responsible for a distinct set of responsibilities. When it is necessary to remember the syntax of previous data in order to forecast the newest, the LSTM performs well. To enhance the LSTM's learning potential and accuracy, we need to strengthen the connections between the hidden neurons by adding additional hidden layers and neurons. When an RNN is dealing with long sequence data throughout the learning process, vanishing or exploding gradients manifest as a weakness of the RNN [6]. To address this issue, Schmidhuber invented the LSTM, which has gates that enable it to better capture data with long-term dependencies [7]. For very complicated nonlinear and extended data, LSTM parameter optimization is required, and it depends on the data features. The adjustment of parameters is accomplished by adjusting the number of layers and hidden units [8]. The DL model employed in this study was assessed on the publicly accessible Covid-19 dataset, which contains data from March 3, 2020, to April 26, 2021. . Five impacted countries' statistics are analyzed in this study: Bangladesh, India, Saudi Arabia, Italy, and Malaysia. We separated the training and testing data for each country. The data for the selected country is divided into two sections, one for input training and one for output training. Trial and error is used to determine the optimal LSTM parameters and hyper-parameters.

1.3 Objectives:

This study presents a DL-based Recurrent Neural Network (RNN) architecture with a Long short-term memory (LSTM) network for estimating the number of patients who may become infected with Covid-19 in Bangladesh in the future days. Furthermore, we have predicted both short-term (10 days) and long-term (30 days) infected cases for both considering temperature data and without temperature data.

A. The general objective of this research work is to asses and analyze the incidence of COVID-19 against natural temperature variation in Bangladesh.

B. Using a Deep learning based Recurrent Neural Network (RNN) architecture Long short-term memory (LSTM) network for predicting the number of patients who may get infected in the upcoming days with COVID-19 in Bangladesh.

C. Predict both the short-term (10 days) and long-term (30 days) infected cases for both with and without the temperature.

D. Comparison of the prediction among five different countries to over-sure the prediction accuracy.

1.4 Report Outline:

The remainder of the paper is structured as follows:

- **Section 1:** ***Introduction;*** Overall overview, problem statement, research objectives and significance of the study.
- **Section 2:** ***Literature Review;*** The related work of Covid-19 prediction.
- **Section 3:** ***Methodology;*** Covid-19 Dataset Collections, Models Design, Training and Testing Process, and Prediction Accuracy Measurement.
- **Section 4:** ***Result and Discussion;*** The results are presented in section 4 along with a brief discussion.
- **Section 5:** ***Conclusion;*** Summarizes the paper's conclusion

Chapter 2

Literature Reviews

2.1 Analysis:

Time series data has been used to track the COVID-19 epidemic. Researchers are focusing on time series data analysis in most of the predictions. Proper forecasting of the pandemic can raise public awareness about the prevention of COVID-19. LSTM estimation has a distinct advantage over time series data with both short and long sequences. Meanwhile, DL analysis has been placed based on time series analysis. We examined such suggestions in the literature review in the part on time series analysis utilizing COVID-19 data, and simulated their work through the discussions.

N.Yudistira proposed a prediction model for anticipating the upcoming pandemics considering the number of infected rate, death, recovered [4]. G. Pinter et al. proposed a hybrid ML approach for Covid-19 prediction in Hungary [5]. M. Alazab et al. used ML and DL technique to forecast and detect Covid-19 [6]. N. Zheng et al. suggested a hybrid AI model to predict the Covid-19 in China [7]. Most of the algorithms used DL and ML methods to predict any future task based on past incident data. M. Rahman et al. used the SIR model to the impact of control strategies on Covid-19 in Bangladesh [8]. M. Khan et al. proposed the infection trajectory-pathway strategy (ITPS) of Covid-19 Outbreak situations in Bangladesh [9]. A. Chowdhury et al. used ANFIS and LSTM algorithms to predict the COVID -19 upcoming 10 day's newly infected cases in Bangladesh. They considered previous days based on 3 scenarios to predict COVID -19 cases [10]. A. Hridoy et al. presented data-driven estimation methods to predict the possible number of COVID-19 cases in Bangladesh for the upcoming months [11].

In the last few months, we were analyzing and investigating COVID-19 in Bangladesh. COVID-19 predicting a is very challenging because it spreads rapidly. Its symptoms do not appear immediately, resulting in a rapid spread of it. Thus, it is important to check which method gives accurate forecast results and which method does not, during a pandemic in Bangladesh. In this case, the number of new cases, temperature, new tests, new deaths, new recovered, total active are used as features of the LSTM network. In general, LSTM is utilized to solve RNN difficulties, non- linearity, lengthy sequence, and heterogeneous features are some of them.

Chapter 3

Methodology

This methodology section is divided into some subsections to get a better overview of the whole work. After discussing the data sources and data preprocessing, the following subsections discuss the model and its variations, and later subsection describes the training testing mechanism of the dataset. Then some mathematical terms are shown, which are used for the model's accuracy measurement.

3.1 Covid-19 Dataset

The COVID-19 dataset for this research is collected from the publicly available dataset provided by "Our World In Data (OWID)" [17] and "NASA Prediction Of Worldwide Energy Resources" [18], which are updating every day. These datasets are available in the format of time series decorated in date, month, and year. And this research is considered the datasets from April 7, 2020, to April 26, 2021 total of 384 days. We collected each day's number of new cases, new deaths, new tests from "OWID" [17], and the recovered cases from "Datahub" [19], and the temperature data from "NASA" [18]. Dataset also includes total active cases, which are calculated as follows:

Total active cases = Total cases – Total deaths – Total recovered (1)

This approach will predict the next 10 days for short-term predictions and the next 30 days for long-term predictions. The model was designed considering the previous 5 timestamps to predict the 6th day's new cases. The table below illustrates the input dataset of this study.

Table 1. Input Framework of the Model

Date	New Cases	Temperature (Celcius)	New Tests	New Deaths	New Recovered	Total Active
2020-04-07	41.0	37.18	981.0	5.0	0.0	114.0
2020-04-08	54.0	37.64	1097.0	3.0	0.0	165.0
2020-04-09	112.0	35.99	1184.0	1.0	0.0	276.0
2020-04-10	94.0	37.25	649.0	6.0	0.0	364.0
2020-04-11	58.0	36.79	1340.0	3.0	3.0	416.0

3.2. Data Preprocessing

Preprocessing data is an essential aspect of any data analysis or prediction strategy. The data has been filtered to exclude null and empty values, as the data may contain null or empty values and irrelevant data. When it comes to data analysis, there are various phases involved. An illustrated figure of the whole methodology is shown in Figure 1.

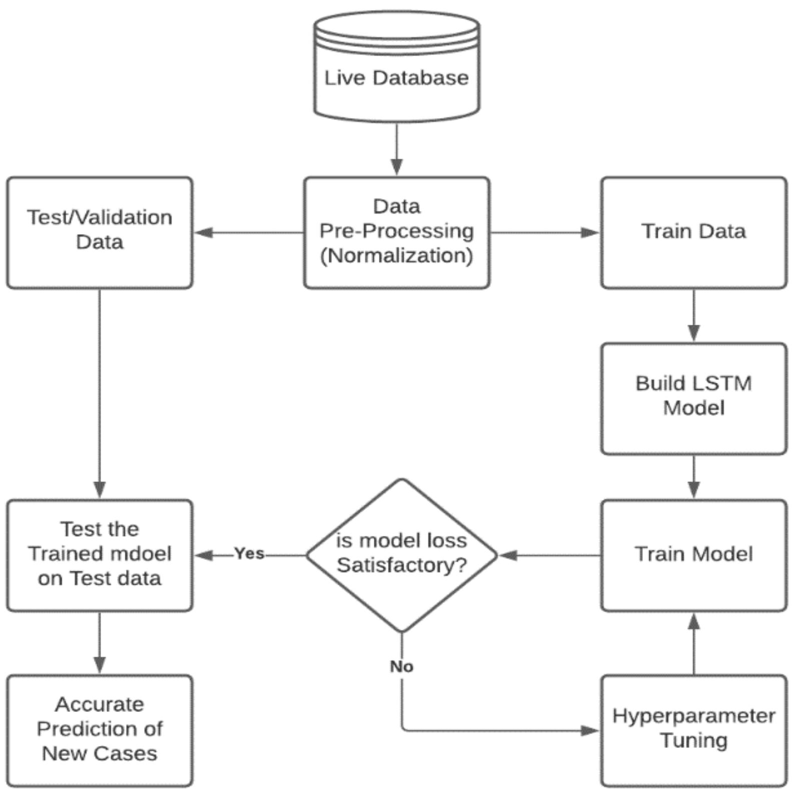


Figure 1. Methodology Diagram

In the case of time series data, normalization is very important for LSTM networks. So, for the input features/variables normalization here, MinMaxScaler is used. MinMaxScaler transforms all the features into the range [0,1] means that the minimum value will be 0, and the maximum value will be 1 for each feature/variable. It is essential to transform the training and testing data into the same amount of scale for avoiding scaling problems during the train and test of the model. The MinMaxScaler is defined as follows,

$$X_{scaled} = (X - X_{min}) / (X_{max} - X_{min}) \quad (2)$$

Where X is the input training dataset, and Xscaled is the output of the normalized training and testing dataset.

3.3. Model Details

3.3.1. Deep learning

Recently, DL has been the most successful approach for the prediction task. This section briefly discusses the core principle of two DL models RNN and LSTM, which will be used later for the COVID-19 forecasting in Bangladesh.

3.3.2. Recurrent Neural Network(RNN)

In the cases of classification and regression problems, a Deep sequential model is more well structured than traditional models [20]. The special feature of RNN is its hidden states which store information of the past sequence. In most cases, RNN is used for prediction because it can analyze sequential data [21]. The critical drawbacks of RNN are the disappearing or expanding gradient problems. RNNs can't solve this problem because they only bind the activation functions of hidden layers from the last time stage. [7].

3.3.3. Long-short-term-memory (LSTM)

LSTM is an RNN addition that handles long-sequence inputs by using a forgetting mechanism [9]. When remembering the syntax of the past data is very important for predicting the latest result, then LSTM works the best. Vanishing gradient and exploding gradient are the most disadvantages of RNN. LSTM can overcome RNN problems. LSTM used prediction, handwriting recognition, speech recognition, and natural language processing (NLP). LSTMs are one of the most effective prediction jobs since they estimate future forecasts based on numerous highlighted aspects in the dataset [22]. LSTM was successful in forecasting because of its ability to handle time-dependent datasets. [23]. The main components of LSTM architecture are the cell state and three regulators. These regulators are called gates: forget gate, input gate, and output gate. The cell states carry various information via gates. Which information is allowed to enter the cell state that decides these gates. Every gate has its weight. Which information should be kept or forgotten during training is decided by these gates using the RNN learning process. Each gate has its own decision, and these decisions are different from each other. The forget gate determines which information from the previous phase should be retained. The input gate decides which data is required for the current state update. The output gate selects which data is required for the following phase. These gates are represented via the following equations:

Input gates: It = sigmoid(Wi[Ht-1,Ct]+bi)
(3)

Forget gate: Ft =sigmoid(Wf[Ht-1,Ct]+bf) (4)

Output gate: Ot =sigmoid(Wo[Ht-1,Ct]+bo) (5)

From the above equations, Wi = Wf = Wo refer respectively to the weight parameters. Ht-1 represents the previous timestamp. Ct represents the current timestamp, and bi= bf=bo refers respectively to the BIOS parameters.

The LSTM model used for this research is constructed by two LSTM layers and a Dense layer. The First LSTM layer has 40, and the second LSTM has 136 hidden neurons, respectively. And the activation function for both layers is Rectified Linear Unit (RELU), which maps all the values in a range from 0 to any positive numbers. Adam optimizer used, and the loss function (LF) used here is Mean Squared Error (MSE). The batch size of 32 and the epoch number is 1000 with an Earlystop callback function with patience 500 is applied to build the model.

Table 2., Table 3. and Table 4. give a glance at the model parameters and hyperparameters.

Table 2.

LSTM Layers	Hidden Neurons	Activation Function	Return Sequence
LSTM Layer 1	40	Relu	True
LSTM Layer 2	136	Relu	False

Table 3.

Optimizer	Loss Function	Batch Size	Epochs
Adam	MSE	32	1000

Table 4.

Callback	Monitor	Restore best weights	Patience
EarlyStop	Value loss	True	500

3.3.4. Training and Testing

While predicting the short-term, our test data contains COVID-19 cases from April 17, 2021, to April 26, 2021; we used the rest of the data, which is the date from April 07, 2020, to April 16, 2021 (370 days) to train the model. And we used the test data to validate the model in every epoch.

Similarly, for predicting the long-term from the date March 28, 2021, to April 26, 2021, we used the rest of the data from April 07, 2020, to March 27, 2021 (350 days) to train the model and for validation, we used the test data. This approach is the same for all the countries, including Bangladesh also.

In the cases for experiment with temperature, we used temperature as a feature for the input dataset of the model to see the effect of temperature in predicting new infections of the countries.

3.3.5. Prediction Accuracy Measurement

Root Mean Squared Error (RMSE) is applied in order to measure the LF of the trained model. And for measuring the prediction performance, Mean Absolute Percentage Error (MAPE) is employed.

RMSE is the measurement of the difference between the actual cases and prediction. And the MAPE is a measure of how accurate a forecast system is. It measures this accuracy as a percentage.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}_i\right)^2} \qquad (6)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i-\overline{x}_i}{x_i}\right| \times 100\% \qquad (7)$$

Here, xi is the original values, are the corresponding estimation, and n is the total number of samples.

Chapter 4

Implementation

4.1 Data Collections

The COVID-19 dataset for this study was compiled using publicly available datasets from "Our World In Data (OWID)" [17] and "NASA Prediction Of Worldwide Energy Resources" [18], both of which are updated on a daily basis. These datasets are in the form of time series including day, month, and year information. And the datasets for this study are from April 7, 2020, to April 26, 2021, a total of 384 days. The number of new cases, fatalities, and tests from "OWID" [17], the recovered cases from "Datahub" [19], and the temperature data from "NASA" [18] were all collected on a daily basis.

4.2 Data Preprocessing

For this research, we used Google Collaboratory, which is a Google Research tool. Colab is a web-based Python editor that allows anyone to write and run arbitrary Python code. It's notably useful for machine learning, data analysis, and education.

One of the most difficult aspects of research is data preprocessing. The null check was performed after seeing the snapshot of data and establishing its shape.

4.2.1 Dataset Snapshot

	date	new_cases	temp	new_tests	new_deaths	Recovered	Active
0	2020-04-07	41.0	37.18	981.0	5.0	0.0	114.0
1	2020-04-08	54.0	37.64	1097.0	3.0	0.0	165.0
2	2020-04-09	112.0	35.99	1184.0	1.0	0.0	276.0
3	2020-04-10	94.0	37.25	649.0	6.0	0.0	364.0
4	2020-04-11	58.0	36.79	1340.0	3.0	3.0	416.0
...	...	...	...	...	...	...	...
380	2021-04-22	4014.0	39.68	27429.0	98.0	7266.0	82844.0
381	2021-04-23	3629.0	38.75	25896.0	88.0	5225.0	81160.0
382	2021-04-24	2697.0	40.80	20571.0	83.0	5477.0	78297.0
383	2021-04-25	2922.0	41.76	21922.0	101.0	4301.0	76817.0
384	2021-04-26	3306.0	41.75	25786.0	97.0	4241.0	75785.0

Figure 2: Snapshot of the dataset

4.2.2 Dataset Null Check

```
[ ] # dataset_new.info()
    dataset_new.isna().sum()

    date          0
    new_cases     0
    temp          0
    new_tests     0
    new_deaths    0
    Recovered     0
    Active        0
    dtype: int64
```

Figure 3: Null checking

4.2.3 Min Max Scaler

```
[ ] scaler = MinMaxScaler()
    data_training = scaler.fit_transform(data_training_df)
    print(data_training.shape)
    print(type(data_training))
    # data_training = np.array(data_training)
    data_training

    (374, 6)
    <class 'numpy.ndarray'>
    array([[0.00000000e+00, 8.22425409e-01, 9.67394155e-03, 4.00000000e-02,
            0.00000000e+00, 0.00000000e+00],
           [1.71390903e-03, 8.44562079e-01, 1.30539934e-02, 2.00000000e-02,
            0.00000000e+00, 4.40928544e-04],
           [9.36058009e-03, 7.65158807e-01, 1.55890323e-02, 0.00000000e+00,
            0.00000000e+00, 1.40059655e-03],
           ...,
           [6.78180620e-01, 9.78825794e-01, 7.04449430e-01, 9.50000000e-01,
            3.04169281e-01, 8.79868586e-01],
           [5.47264338e-01, 8.80173244e-01, 5.62662082e-01, 9.30000000e-01,
            3.37363828e-01, 8.64159426e-01],
           [5.76928148e-01, 7.47834456e-01, 5.31979370e-01, 1.00000000e+00,
            3.24759026e-01, 8.52245710e-01]])
```

Figure 4: Min-Max Scaler

4.2.4 Building the Model

```python
def create_model():
  regressor = Sequential()

  regressor.add(LSTM(units = 40, activation = 'relu', return_sequences = True,
                     input_shape = (X_train.shape[1], 6)))

  regressor.add(LSTM(units = 136, activation = 'relu'))

  regressor.add(Dense(units = 1))
  regressor.summary()
  return regressor

regressor = create_model()
```

```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_8 (LSTM)                (None, 5, 40)             7520
_________________________________________________________________
lstm_9 (LSTM)                (None, 136)               96288
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 137
=================================================================
Total params: 103,945
Trainable params: 103,945
Non-trainable params: 0
_________________________________________________________________
```

Figure 5: Building the model

Chapter 5

RESULT AND DISCUSSION

The approach taken in this study is different from previous studies. The risk from the virus was reduced significantly for warm places and countries [24]. Although only temperature data alone can not be a potential factor of COVID-19 transmission rate [25]. Temperature is used as a feature in our model for the input dataset to get an overview of the effect of temperature on the accurate prediction of COVID-19 transmission. This approach is divided into four parts; for the first part, we predicted the short-term with temperature and second, we predicted the short-term without temperature. In the third approach, we predicted the long-term with temperature and lastly predicted the long-term without temperature. Our approaches will help the government and researchers to get an idea to be prepared for the next action that needs to be taken. Hence it will help to slow down the infection rate of the country.

Table 5. and Table 6. show the result of the comparison of prediction including and excluding the temperature data; in this case, only temperature data has been considered. And the outcome shows that the temperature doesn't really have a good impact on the COVID-19 transmission rate. It only differs from 2 to 3% when adding or removing temperature data from the input dataset.

Here the prediction accuracy is measured by RMSE and MAPE for all countries. For the short-term given in Table 5., the accuracy for all the countries is more than 96% to 98%, also Figure 2. shows the accurate and the predicted results. And for the long-term given in Table 6., the accuracy is more than 93% to 97% and the Figure 3. Illustrates the accurate and predicted result of the COVID-19 cases. However, for Bangladesh, short-term prediction from the day April 17, 2021, to April 26, 2021, with temperature data gives us a (MAPE=4.5) and the (RMSE=236). And without the temperature data, our model gives (MAPE=5) and (RMSE=218). As the new case's increasing rate is very high, and there is no specific sequence or pattern to recognize and predict the next pattern of the COVID-19 infection. So, it becomes very hard to predict the future COVID-19 new cases. As we may not get the number of cases predicted as same as the actual cases, this model gives a satisfactory result than the previous studies. And also, for a country like Bangladesh, an (MAPE=4.5) or (MAPE=5) is very much satisfactory for this kind of complex and non-linear datasets.

As said before, we have also compared our model with foreign countries, specifically India, Saudi Arabia, Italy, and Malaysia. And the prediction accuracy of those countries is also very much satisfied because their prediction is so close to the real values. These countries have a 2-3% accurate prediction when the temperature was considered in the input dataset.

In Table 5., RMSE and MAPE are calculated for short-term prediction. The prediction of five countries Bangladesh, India, Saudi Arabia, Italy, and Malaysia, and their accuracy measured in terms of RMSE and MAPE calculated in two ways short-term with temperature, and short-term without temperature.

The result of this table shows that with temperature the MAPE is in the range of 1.9% to 4.5% for all the countries, and without temperature MAPE range is 2.8% to 5%. The result also shows that the values of RMSE and MAPE remain low when the temperature is used except in India.

Table 5. Prediction Accuracy in Terms of RMSE and MAPE of the Model on Different Countries Based on Short Term(10 Days)

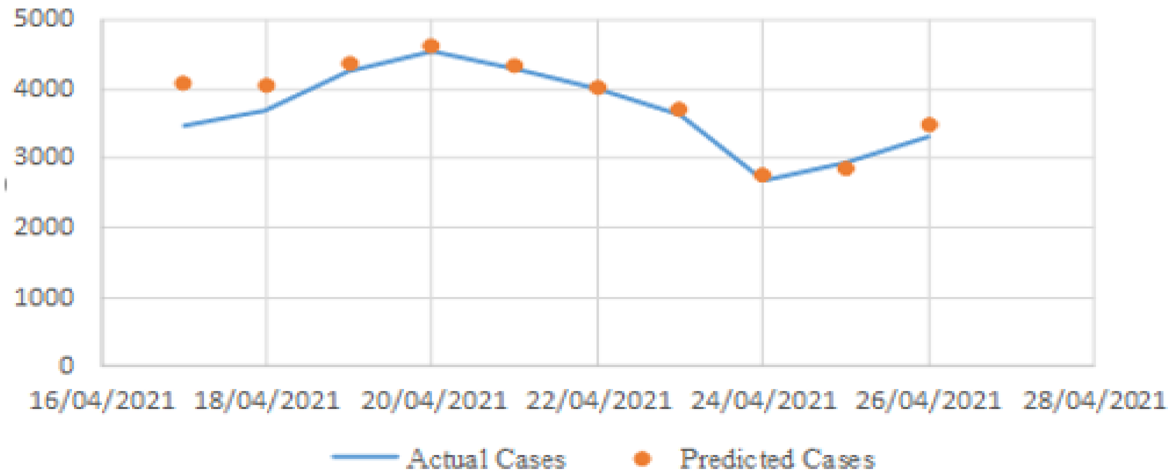
Experimented Countries List	Short-term with Temperature		Short-term without Temperature	
	RMSE	MAPE	RMSE	MAPE
Bangladesh	236	4.5	218	5
India	15122	3.7	12999	3.4
Saudi Arabia	24.4	1.9	42	3.6
Italy	353	2.1	506	2.8
Malaysia	113	3.3	173	6

Table 6. Prediction Accuracy in Terms of RMSE and MAPE of the Model on Different Countries Based on Long Term(30 Days).

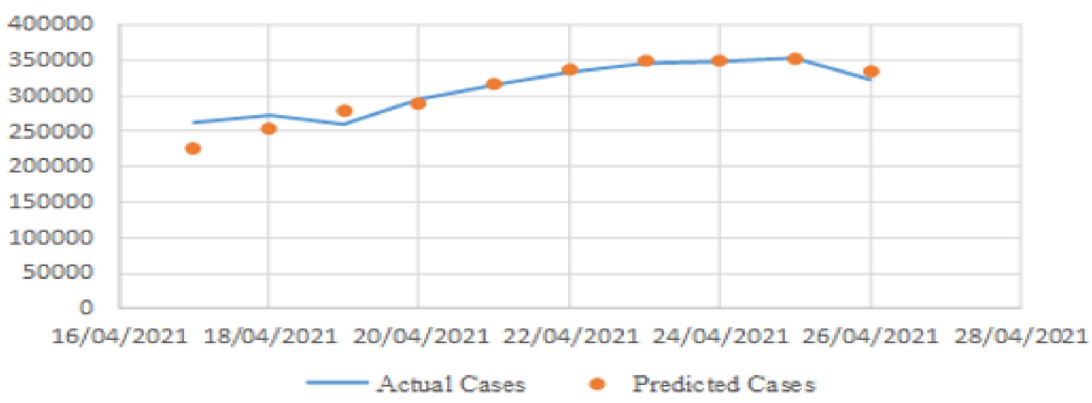
Experimented Countries List	30 days with Temperature		30 days without Temperature	
	RMSE	MAPE	RMSE	MAPE
Bangladesh	651	9.8	680	11.6
India	10983	6.4	11774	6.5
Saudi Arabia	50	4.6	50	4.5
Italy	1836	9.6	1959	10.0
Malaysia	291	14.4	276	14.4

Similarly for Table 6., here, RMSE and MAPE calculated for long-term prediction. The prediction of five countries Bangladesh, India, Saudi Arabia, Italy, Malaysia, and their accuracy measured in terms of RMSE and MAPE calculated in two ways long-term with temperature and long-term without temperature.

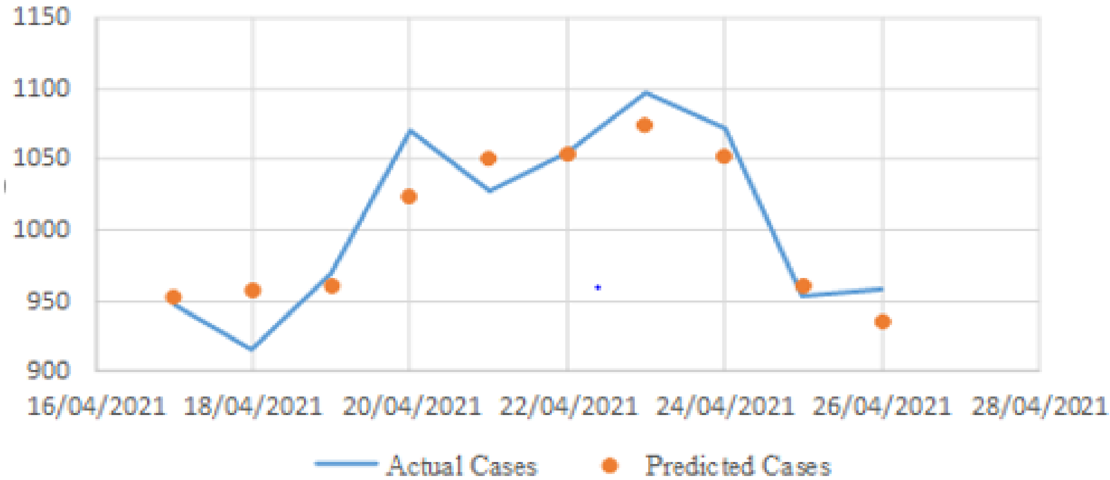
This table results that with temperature the MAPE is in the range of 4.6% to 14.4%, and without temperature, the MAPE range is in 4.5% to 14.4%. Here the MAPE of Bangladesh has changed, but the MAPE of other countries has not changed. In the long-term model, using temperature is almost the same as not using temperature.



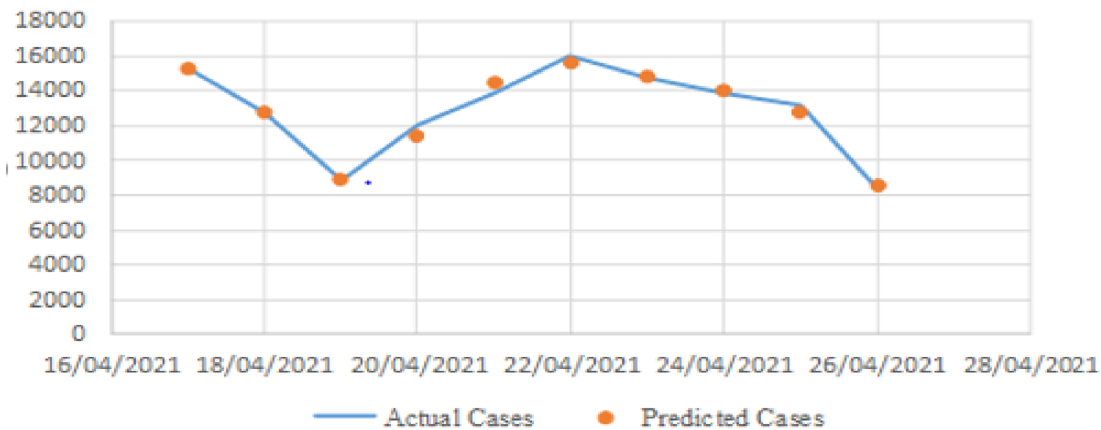
a. COVID-19 new cases short-term prediction in Bangladesh with temperature



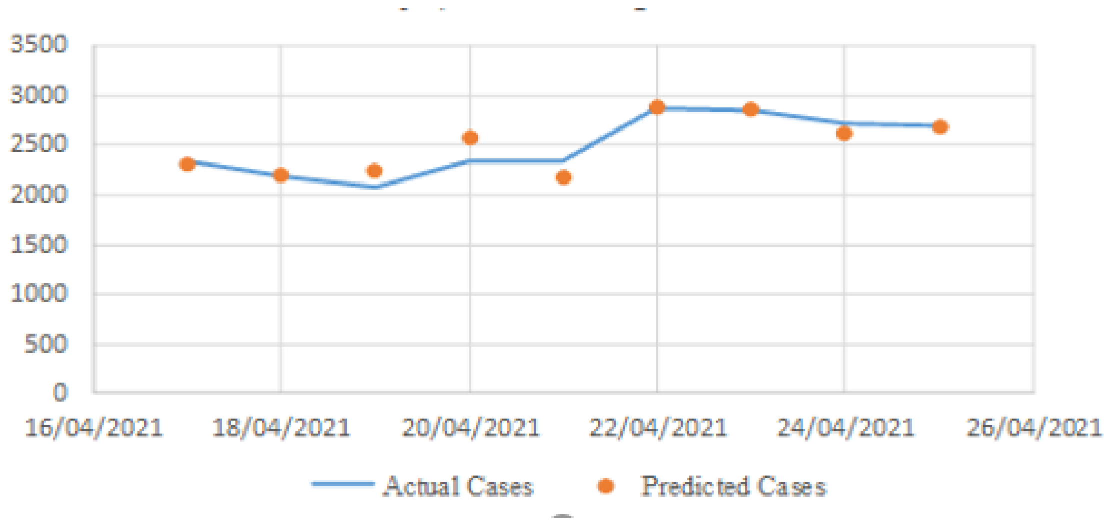
b. COVID-19 new cases short-term prediction in India with temperature



c. COVID-19 new cases short-term prediction in Saudi Arabia with temperature



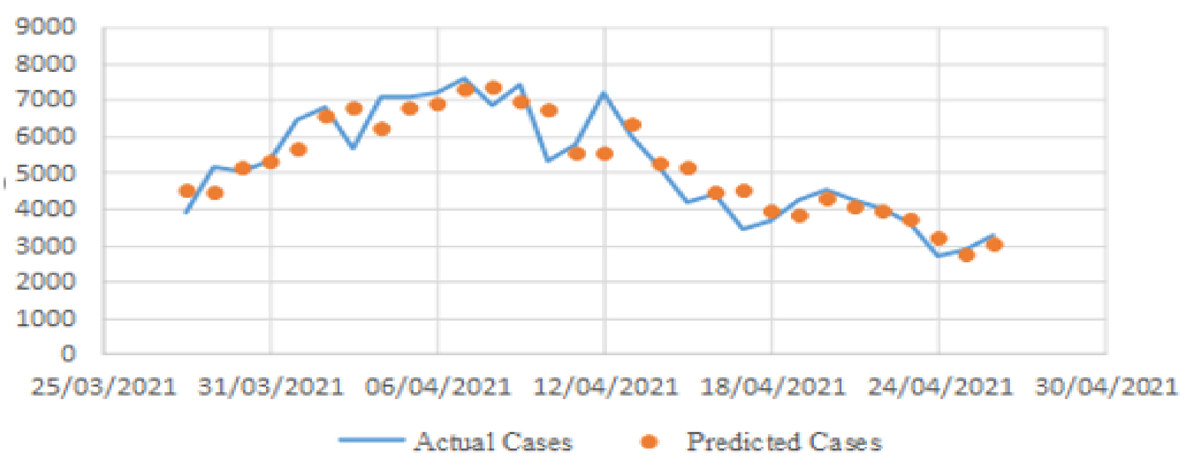
d. COVID-19 new cases short-term prediction in Italy with temperature



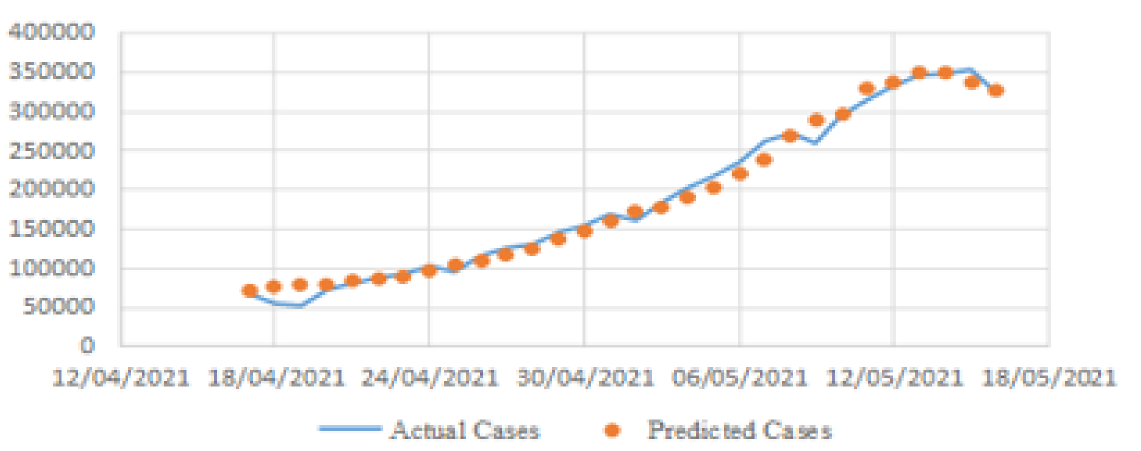
e. COVID-19 new cases short-term prediction in Malaysia with temperature

Figure 6. Actual and predicted COVID-19 new cases(short-term) from 16 April to 26 April considering tempereture data.

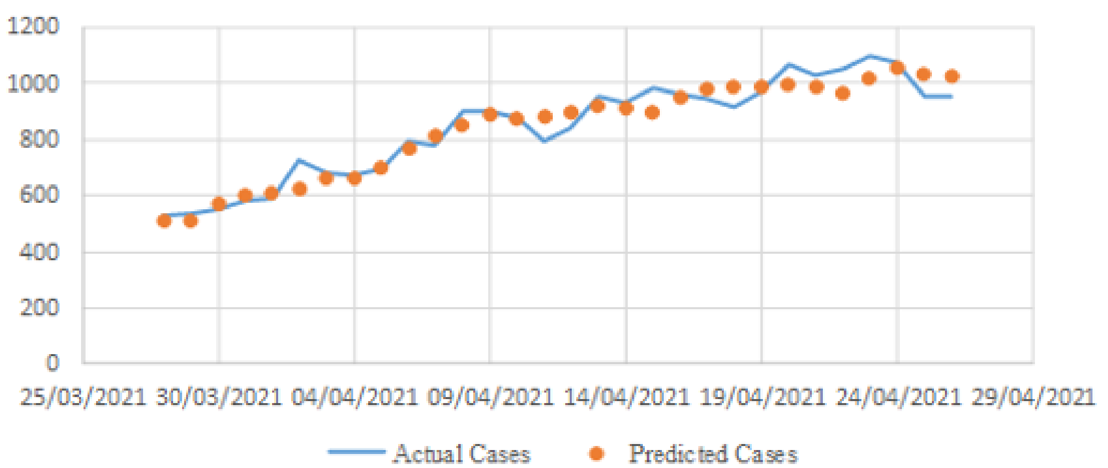
In Figure 6. the short-term prediction with temperature and the actual cases of COVID-19 are shown. Here, the predicted curves are well fitted on the actual curve. The graph of Bangladesh, Saudi Arabia, Italy, and Malaysia shows that after 24 April this country's COVID-19 transmission is going downwards. Whereas, in the case of India, after 25 April the graph is going down. And the predicted curve is also reflecting well the actual curve for all five countries.



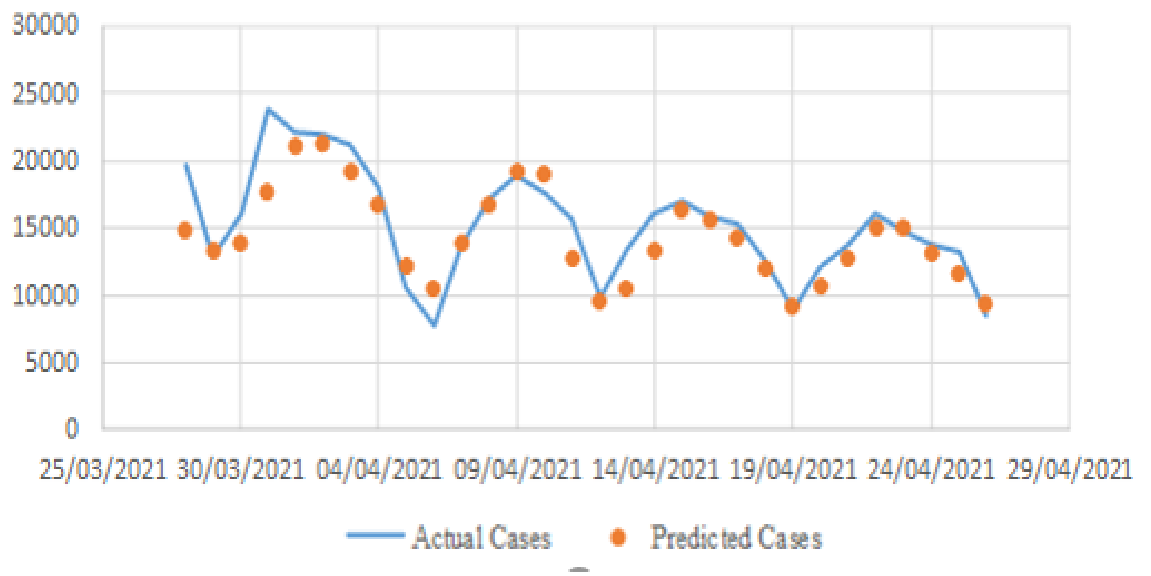
a. COVID-19 new cases long-term prediction in Bangladesh with temperature



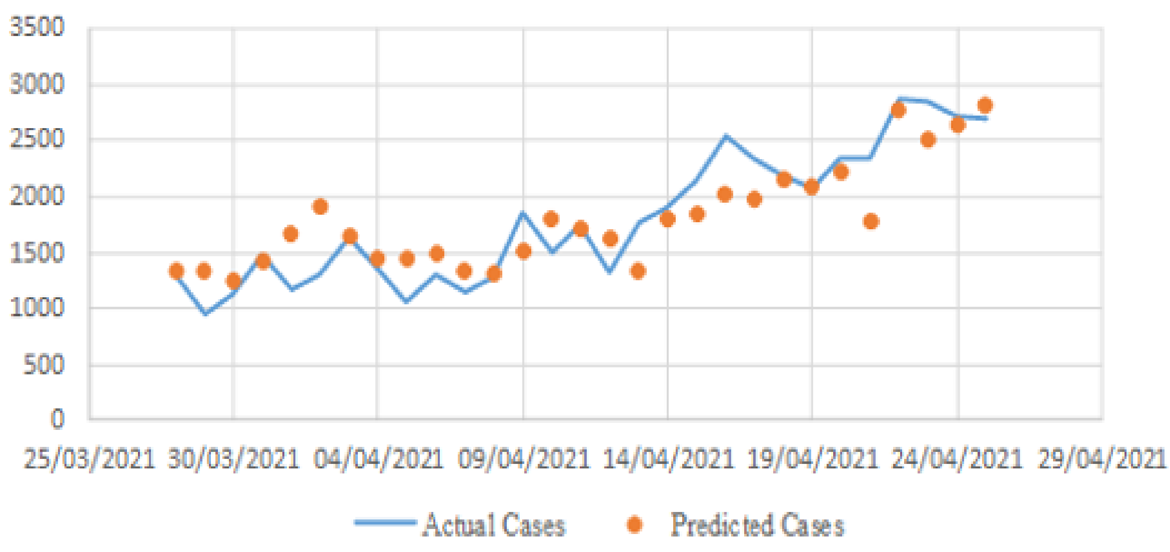
b. COVID-19 new cases long-term prediction in India with temperature



c. COVID-19 new cases long-term prediction in Saudi Arabia with temperature



d. COVID-19 new cases long-term prediction in Italy with temperature



e. COVID-19 new cases long-term prediction in Malaysia with temperature

Figure 7. Actual and predicted COVID-19 new cases(long-term) from 28 march to 28 April considering tempereture data.

In Figure 7. COVID-19 new cases prediction and the actual cases of short-term with temperature of five countries Bangladesh, India, Saudi Arabia, Italy, and Malaysia are shown. And the graph of India shows that the cases are increasing exponentially. In Bangladesh, the cases are very complex and nonlinear and the case count is decreasing at the end of April.

From the results shown above, it is seen that the model's accuracy of prediction for both including and excluding the temperature data is the closest possible value achieved in comparison with the real COVID-19 cases. The prediction accuracy is almost the same for both with and without temperature data for Bangladesh. However, for other countries, MAPE varies 2 to 3%. Our proposed model not only justifies the effectiveness of temperature but also can predict the COVID-19 cases very accurately. This model can predict the short-term and long-term. The experiment with the five countries Bangladesh, India, Saudi Arabia, Italy, and Malaysia got satisfactory results for both the short-term and long-term. So, it is clear that the effect of temperature is negligible in forecasting the upcoming per day's COVID-19 cases.

Chapter 6

Conclusion

6.1 Conclusion

We already know, Covid-19 has been considered a great threat to the whole world. In this paper, we have proposed deep learning-based models to utilize how effective temperature is in Covid-19 predictions. Researchers are trying to find proper solutions to overcome the Covid-19 problem. It still did not possible to find out proper solutions to overcomes this problem. Like other countries, Bangladesh suffers many problems regarding Covid-19. Accurate forecasting can help to overcome any problems in a short time. In this study, a Deep Learning based model, which is LSTM, an architecture of RNN, is used to determine the effect of temperature in accurately forecasting the upcoming confirmed cases of five different countries, including Bangladesh. The deep learning model is used because of its capability in processing highly complex and nonlinear data. The prediction is made including and excluding the temperature as data for both the short-term and long-term for five countries, namely Bangladesh, India, Saudi Arabia, Italy, and Malaysia. The outcome shows that the effect of temperature in forecasting the newly confirmed cases is negligible.

6.2 Future Scope

This study can be further improved by including other factors like vaccination rate, mask, and sanitizer using rate as features of the model if proper datasets can be achieved.

At this moment, we expect that this effort would be of benefit to the Bangladesh government in making future judgments. This model can be used in predicting Coronavirus disease outbreaks accurately.

REFERENCES.

[1]. M. Johnson, "Wuhan 2019 Novel Coronavirus - 2019-nCoV," Mater. Methods, vol.10,no. January, pp. 1–5, 2020, doi: 10.13070/mm.en.10.2867.

[2]. "Severe acute respiratory syndrome coronavirus 2",Wikipedia,Wikimedia Foundation,12May2021,https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome_coronavirus_2

[3]. "COVID-19 vaccination in Bangladesh", Wikipedia, Wikimedia Foundation, 9 May 2021, https://en.wikipedia.org/wiki/COVID-19_vaccination_in_Bangladesh

[4]. R. Pal, A. A. Sekh, S. Kar, and D. K. Prasad, "Neural network based country wise risk prediction of COVID-19," Appl. Sci., vol. 10, no. 18, 2020, doi: 10.3390/APP10186448.

[5]. A. Shoeibi *et al.*, "Automated Detection and Forecasting of COVID-19 using Deep Learning Techniques: A Review," 2020, [Online]. Available: http://arxiv.org/abs/2007.10785.

[6]. J. Rosindell and Y. Wong, "Biodiversity, the tree of life, and science communication," Phylogenetic Divers. Appl. Challenges Biodivers. Sci., no. 2, pp. 41–71, 2018, doi: 10.1007/978-3-319-93145-6_3.

[7]. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[8]. M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning for time series classification," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, 2014, [Online]. Available:

https://www.m-culture.go.th/mculture_th/download/king9/Glossary_about_HM_King_Bhumibol_Adulyadej's_Funeral.pdf.

[9]. N. Yudistira, "COVID-19 Growth Prediction using Multivariate Long Short Term Memory," *IAENG Int. J. Comput. Sci.*, vol. 47, no. 4, pp. 1–9, 2020.

[10]. G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, "COVID-19 pandemic prediction for hungary; A hybrid machine learning approach," *medRxiv*, 2020, doi: 10.1101/2020.05.02.20088427.

[11]. M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, and S. Alhyari, "COVID-19 prediction and detection using deep learning," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 12, no. May, pp. 168–181, 2020.

[12]. N. Zheng *et al.*, "Predicting COVID-19 in China Using Hybrid AI Model," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 2891–2904, 2020, doi: 10.1109/TCYB.2020.2990162.

[13]. M. M. Rahman, A. Ahmed, K. M. Hossain, T. Haque, and M. A. Hussain, "Impact of control strategies on COVID-19 pandemic and the SIR model based forecasting in Bangladesh," *medRxiv*, 2020, doi: 10.1101/2020.04.19.20071415.

[14]. M. H. R. Khan and A. Hossain, "COVID-19 Outbreak situations in Bangladesh: An Empirical Analysis," medRxiv, pp. 1–15, 2020, doi: 10.1101/2020.04.16.20068312.

[15]. A. A. Chowdhury, K. T. Hasan, and K. K. S. Hoque, "Analysis and Prediction of COVID-19 Pandemic in Bangladesh by Using ANFIS and LSTM Network," *Cognit. Comput.*, pp. 761–770, 2021, doi: 10.1007/s12559-021-09859-0.

[16]. E. E. Hridoy *et al.*, "Forecasting COVID-19 dynamics and endpoint in Bangladesh: A data-driven approach," *medRxiv*, pp. 0–3, 2020, doi: 10.1101/2020.06.26.20140905.

[17]. "Coronavirus(COVID-19) Deaths - Statistics and Research -Our World in Data", May 17, 2021. Accessed on: May 19, 2021[online]. Available:https://ourworldindata.org/covid-deaths

[18]. "NASA Prediction Of Worldwide Energy Resources", Feb 25, 2020.Accessed on: April 17, 2021[online].Available: https://power.larc.nasa.gov/

[19]. "*Novel Coronavirus 2019*", May 5, 2021. Accessed on: May 5,2021[online]. Available on: https://datahub.io/core/covid-19

[20]. Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–27, 2009, doi: 10.1561/2200000006.

[21]. A. Graves, “Generating Sequences With Recurrent Neural Networks,” pp. 1–43, 2013, [Online]. Available: http://arxiv.org/abs/1308.0850.

[22]. P. Arora, H. Kumar, and B. K. Panigrahi, “Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India,” Chaos, Solitons and Fractals, vol. 139, p. 110017, 2020, doi: 10.1016/j.chaos.2020.110017.

[23]. A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, “Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study,” *Chaos, Solitons and Fractals*, vol. 140, p. 110121, 2020, doi: 10.1016/j.chaos.2020.110121.

[24]. Roy, “The role of temperature on the global spread of COVID-19 and urgent solutions,” *Int. J. Environ. Sci. Technol.*, no. November, 2020, doi: 10.1007/s13762-020-02991-8.

[25]. P. Mecenas, R. T. da Rosa Moreira Bastos, A. C. Rosário Vallinoto, and D. Normando, “Effects of temperature and humidity on the spread of COVID-19: A systematic review,” *PLoS One*, vol. 15, no. 9 September, pp. 1–21, 2020, doi: 10.1371/journal.pone.02383