STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–17.

Hybrid Deep Learning Model: LSTM and 2BiGRU for Predicting Coronavirus (COVID-19)

Thanaa Moustafa *, Mohamed A.Makhlouf , Hosam E.Refaat

Department of Information System, Faculty of Computers and Informatics, Suez Canal University, Egypt

Abstract The COVID-19 pandemic has had a major global health impact, highlighting the urgent need for accurate predictive models to forecast the virus's spread. This research explores the use of deep learning techniques to improve the accuracy of COVID-19 case predictions. Traditional machine learning methods often struggle with the complexities of time-series data inherent in pandemic forecasting, which motivates the use of advanced deep learning models. This study employs the LSTM-2BiGRU model, a sophisticated deep learning architecture, to predict new COVID-19 cases using two datasets: historical data from OurWorldInData and medical data with historical disease records. The model was trained to leverage time-dependent factors and achieve high prediction performance. The LSTM-2BiGRU model achieved a significant improvement over traditional machine learning models, with an accuracy of 76% and a Mean Absolute Error (MAE) of 8% for the historical dataset within a 7-day forecast window. When applied to the epidemiology dataset, the model demonstrated even higher accuracy, ranging from 80% to 90% across different prediction periods (1 to 14 days), with a Mean Absolute Percentage Error (MAPE) between 10% and 15%. These findings demonstrate the potential of deep learning models like LSTM-2BiGRU to provide more accurate and timely forecasts for COVID-19, with a substantial reduction in Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) compared to previous studies. This underscores the model's improved performance and supports better-informed public health decisions.

Keywords Machine Learning, Deep Learning, COVID-19, Time Series, LSTM, GRU

DOI: 10.19139/soic-2310-5070-2271

1. Introduction

Since its emergence in December 2019, the coronavirus (COVID-19) has rapidly spread across the globe, [1, 2] causing widespread morbidity and mortality. The transmission of COVID-19 is facilitated by human interaction, making containment measures such as travel restrictions, physical distancing, and frequent hand hygiene critical to limiting its spread. The most common symptoms of COVID-19 include fever and shortness of breath, but in more severe cases, additional symptoms such as chest pain, sputum production, and hoarseness can also occur. In its more advanced stages, the virus can lead to pneumonia, which has a reported fatality rate of 5.8 %. In fact, the mortality rate of COVID-19 has surpassed that of the 1918 Spanish flu pandemic, highlighting the severe impact of the virus. [3, 4, 5] In response to the global health crisis, medical researchers from various disciplines have been actively seeking solutions to control the spread of the virus. Epidemiological models have been employed to track and predict the progression of COVID-19 cases. In certain regions, the confirmed case numbers have reached a point of saturation, where the rate of new infections has plateaued due to a reduction in the number of susceptible individuals. Governments have adopted stringent measures such as nationwide lockdowns to isolate infected individuals and minimize social contact, effectively decreasing the likelihood of further transmission[6, 7, 8, 9, 10]. Machine learning (ML) and deep learning (DL) algorithms have become indispensable tools in epidemiological

---

*Correspondence to: Thanaa Moustafa (Email: sanaamostafa59@gmail.com). Department of Information System, Faculty of Computers and Informatics, Suez Canal University, Egypt.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

analysis and forecasting. These algorithms are particularly effective when working with large-scale epidemiological datasets, enabling researchers to model epidemic patterns and predict future trends. Such predictive models are essential for taking proactive steps to curb the spread of the virus and mitigate its impact. Leveraging real-time data from sources like the Johns Hopkins University dashboard which already gets updated data from OurWorldInData, machine learning and deep learning models have been employed to observe the daily dynamics of COVID-19 and to forecast future case numbers [11, 12, 13, 14]. This research focuses on the application of deep learning models, specifically Long Short-Term Memory (LSTM) networks and Bidirectional Gated Recurrent Units (BiGRU), to improve the prediction accuracy of COVID-19 cases. LSTMs, a specialized form of Recurrent Neural Networks (RNNs), are designed to capture long-term dependencies in sequential data, making them particularly suited for time-series forecasting. Meanwhile, BiGRU models allow for the extraction of features in both forward and backward directions, leading to more robust and meaningful predictions [15, 16, 17, 18, 19].

1.1. Study Contributions

By combining the strengths of LSTM and 2BiGRU architectures, this study aims to enhance the accuracy of COVID-19 case predictions, surpassing the performance of traditional machine learning algorithms, which typically achieve an average accuracy of around 64% for common classification tasks. The workflow of the proposed model, illustrating the step-by-step process, is presented in Figure 1 to provide a visual overview of the methodology. The primary contributions of this study are as follows:

- Two types of datasets were collected and processed. The first dataset consists of real-time COVID-19 data specific to Egypt, which was preprocessed by removing irrelevant columns and values to retain only confirmed cases and Dates. The second dataset is an epidemiological dataset of positive and negative COVID-19 cases from Mexico, sourced from the General Directorate of Epidemiology, Secretariat of Health, and made publicly available on their official website. This dataset was similarly cleaned and feature-engineered, with the top 10 features selected from an initial set of 41.
- A hybrid model combining Long Short-Term Memory (LSTM) and Bidirectional Gated Recurrent Units (BiGRU) was developed, with distinct attributes tailored to suit the scale and characteristics of each dataset.
- The model was trained on the Egypt dataset, focusing on confirmed cases over 7 days, and its performance was compared with previous studies.
- Based on the model's superior performance on historical data demonstrated by its low Root Mean Squared Error (RMSE) the same model was trained on the Mexican epidemiological dataset over 1 to 14 days. The model's performance was then compared to that of traditional machine learning classifiers and a standalone LSTM model, prior to the integration of BiGRU

1.2. Paper Organization

In Sect. 2, a literature review of several recent related research was investigated. Section 3 presents the dataset characteristics. Section 4 illustrates the proposed model phases and design. Section 5 is the main discussion of the output results, efficiency, and limitations. Section 6 presents the conclusion that summarizes the over all work in this study with the future directions.

2. Releated Work

Numerous studies have utilized machine learning and deep learning models to predict and analyze the spread of COVID-19, as well as its associated symptoms and case severity. Reviewing key works in this area highlights significant progress and existing gaps. Torad et al [20] examined the symptoms and transmission of COVID-19, noting that the primary symptoms include fever, dry cough, and myalgia, with an incubation period ranging from 1 to 14 days. The authors emphasized the transmission of the virus through respiratory droplets and surface

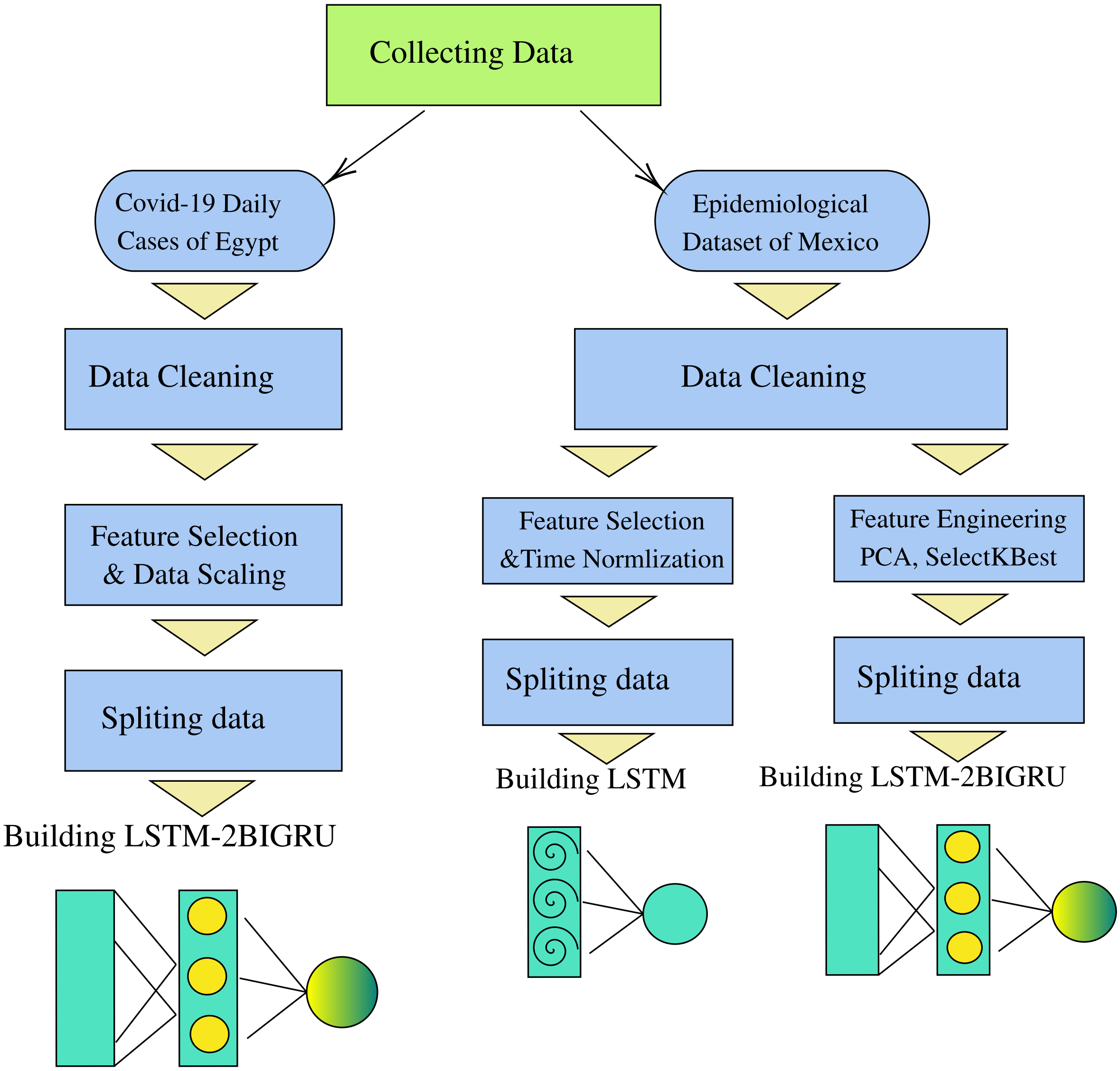

Figure 1. Work-Flow of This Study.

contact. Their research identified a need for further investigation into the clinical and epidemiological features of COVID-19 patients in Egypt, particularly the impact of health status and work stress. Using a Random Forest Classifier, the study achieved a mean accuracy of 74.77% in predicting case severity based on vital signs, providing insight into feature importance rankings.

M. Marzouk et al [18] significant contribution analyzed COVID-19 cases in Egypt by utilizing data on confirmed cases, recoveries, and deaths. The study employed Long Short-Term Memory (LSTM) networks to forecast the spread of the outbreak, showing that LSTM outperformed other models. However, the study also highlighted the complexity of deep learning models, noting that while such models can capture intricate patterns, their complexity can also pose challenges in terms of interpretability and implementation.

Further S. P. Cumbane et al [17] explored the use of a multi-layer Bidirectional Long Short-Term Memory (BiLSTM) model for predicting COVID-19 spread in developing countries. The study demonstrated that BiLSTM models consistently outperformed standard LSTM models in predicting the spread of the virus, achieving superior accuracy across multiple countries and cities. This suggests that the bidirectional nature of BiLSTM allows for a more comprehensive capture of temporal dependencies, leading to more accurate forecasts.

Additionally, N. N. Aung et al [16] leveraged open-source data, including flight data, to perform COVID-19 time series forecasting across 190 countries using machine learning methods. The authors utilized data available until January 31, 2021, to predict the total number of cases over a 14-day period. The study found that while the BiLSTM model showed a median prediction error of 33%, its maximum error was significantly lower than that of traditional models such as ARIMA, indicating that the inclusion of more variables, along with the application of advanced machine learning techniques, can help minimize prediction errors.

D. Fedman et al [15] presented two recurrent neural network-based approaches LSTM and GRU to predict daily confirmed COVID-19 cases, daily positive tests, and total individuals vaccinated. These models achieved remarkable accuracy, with a mean absolute percentage error of less than 1.9%, demonstrating their capability to forecast COVID-19 spread variables effectively.

Author [21] comparative study focused on COVID-19 vaccination prediction using XGBoost and LightGBM (LGBM). While LGBM with all features yielded an accuracy of 80% but high RMSE, a refined feature selection improved the accuracy to 88% with a reduced r2 score when applied to Germany's daily confirmed cases. This highlights the importance of feature engineering in optimizing model performance.

A. Solanki at al [22] utilized three datasets: the Johns Hopkins CSSE repository, Worldometers, and additional sources to develop various predictive models, including SARIMAX, HoltWinters, and polynomial regression. The SARIMAX model achieved a MAPE of 0.236, while the polynomial regression model demonstrated 85% accuracy in predicting the number of affected cases and deaths over a 10-day period. Meanwhile, the LSTM model, optimized using the ADAM optimizer, produced low overall prediction errors, further validating the effectiveness of deep learning methods in COVID-19 forecasting.

N. E. M. Khalifa at al [19] developed a deep learning sequence-to-sequence regression model using population mobility reports to predict the spread of COVID-19 in Egypt. Achieving 96.69% accuracy for a 3-day prediction window, this study diverged from traditional models by focusing on mobility data rather than daily infection rates, achieving highly accurate predictions with low error.

Finally, S. Bouklif [23] investigating COVID-19 detection through Bi-Directional LSTM and early clinical reports achieved testing accuracy of 87.15%, with precision at 91% and recall at 88%. The proposed model outperformed traditional ensemble models, underscoring the potential of Bi-LSTM to enhance the early detection of COVID-19.

3. MATERIALS and METHODS

The objective of this study is to develop a robust and accurate model for predicting COVID-19 case numbers using deep learning techniques. To achieve this, we utilized two distinct datasets and employed a hybrid LSTM-2BiGRU architecture for time-series forecasting. The approach undertaken in this work aimed to improve predictive performance over traditional machine learning models by leveraging the sequential dependencies present in epidemiological and real-time COVID-19 data. This section outlines the materials, data sources, and methodologies employed throughout the study. The emphasis is placed on effective data preprocessing, feature selection, and model construction to achieve optimal predictive accuracy. By systematically preparing the data and evaluating the model's performance on different datasets, this work contributes to the growing body of research focused on forecasting epidemic trajectories and informing public health interventions.

3.1. Data Collecting and Preprocessing

Two datasets were used for the model development: a real-time dataset consisting of daily confirmed COVID-19 cases from Egypt was obtained from the OurWorldInData website [24, 25], and a detailed epidemiological dataset from Mexico was sourced from the General Directorate of Epidemiology, Secretariat of Health in Mexico [26]. These datasets provided the basis for training and evaluating the predictive model. Preprocessing steps were applied to ensure data integrity, including removing irrelevant columns, handling missing values using mean imputation, where the average value of the respective feature was calculated and substituted for missing entries, and scaling features to prepare the data for modeling.

The first dataset was collected from daily reported confirmed COVID-19 cases, specifically extracting data for Egypt. The dataset spans from January 21, 2020, to March 2023, encompassing 515,759 confirmed cases. The dataset was filtered to retain only three relevant columns: Country, Cases, and Dates. As shown in figure 2, these columns provided the basis for building the time-series forecasting model. The Cases column was scaled using the MinMaxScaler [27] to normalize the data. The MinMaxScaler function transforms features by scaling them within a specified range (0,1), as in fig 3. The equation for the MinMaxScaler 1 is expressed as follows:

$$X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{1}$$

Where:

- X is the original value.
- X_{min} is the minimum value in the dataset.
- X_{max} is the maximum value in the dataset.
- X' is the scaled value after applying the MinMax scaling.

	Country/Region	Date	Cases
23	Egypt	2020-02-14	1
24	Egypt	2020-02-15	1
25	Egypt	2020-02-16	1
26	Egypt	2020-02-17	1
27	Egypt	2020-02-18	1

	Country/Region	Date	Cases
1138	Egypt	2023-03-05	515698
1139	Egypt	2023-03-06	515698
1140	Egypt	2023-03-07	515698
1141	Egypt	2023-03-08	515759
1142	Egypt	2023-03-09	515759

Figure 2. Egypt Confirmed Cases of COVID-19

Scaling the data is particularly important for machine learning models that are sensitive to feature magnitude, such as neural networks, gradient descent-based models, and those that rely on distance metrics, including Support Vector Machines (SVM) [28, 29] and k-Nearest Neighbors (KNN) [30]. The scaling ensures effective convergence during training and helps the model better capture relationships within the data.

The Second dataset was sourced from the Epidemiology Directorate of the Secretariat of Health in Mexico. It contains 41 features related to pre-existing conditions, such as asthma, pneumonia, and diabetes, which may influence the outcome of COVID-19 tests (i.e., positive or negative results) features are written in Spanish. So, the names of the columns were translated into English. After preprocessing, 18 relevant features were selected for the analysis. The outcome variable was scaled to binary form, where 0 represents a negative COVID-19 test result and 1 represents a positive result. Additionally, date features were converted into numerical values to facilitate model processing.

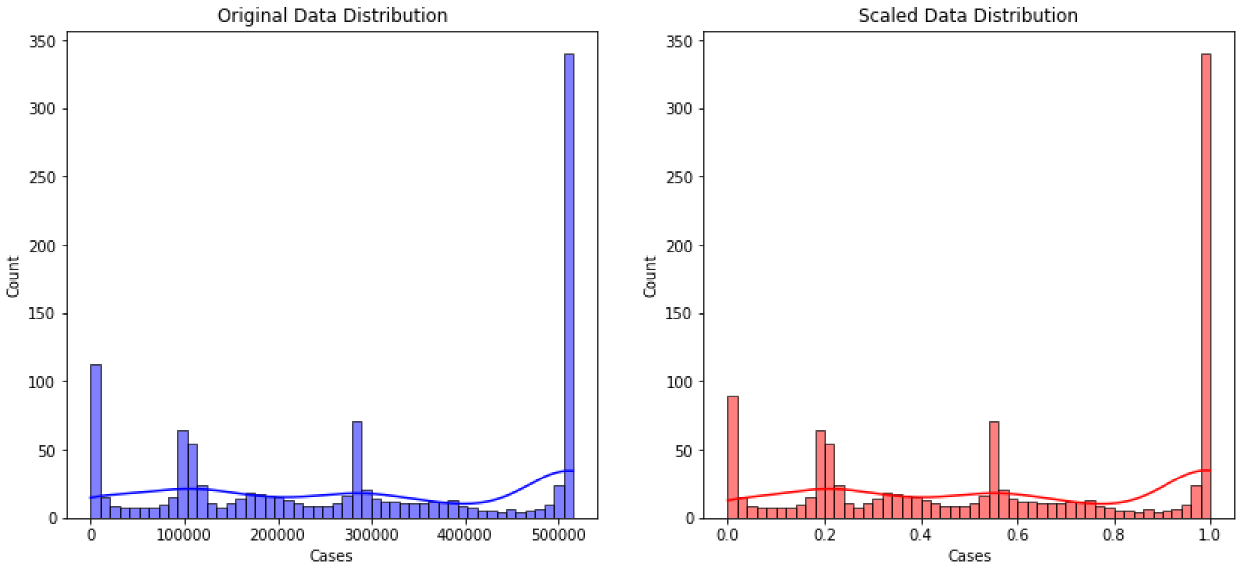

Figure 3. Distribution of Scaled Values VS Original

3.2. Feature Engineering and Dimensionality Reduction

Using the epidemiology dataset provided by the Secretariat of Health in Mexico, we performed feature engineering and applied feature importance analysis using a 'RandomForestClassifier' the output in figure 4 with 'randomState=42' eq2 This method was employed to rank features based on their contribution to prediction outcomes. The results highlighted AGE as the most critical feature, followed by ICU admission, intubation, and pneumonia, which align with known clinical risk factors in COVID-19 outcomes. Subsequently, dimensionality reduction was applied using Principal Component Analysis (PCA) eq4 to identify patterns in the data and reduce noise, ensuring the model could focus on the most relevant information without overfitting, followed by feature selection with the SelectKBest method eq3. The top 10 features were identified for model training, The top-ranked features further validated the importance of clinical factors such as comorbidities (e.g., diabetes, obesity, hypertension). This pipeline reduced the impact of noisy and irrelevant data, focusing the predictive model on the most informative features.

$$\hat{y} = \text{mode}(T_1(X), T_2(X), \ldots, T_m(X)) \tag{2}$$

Where:

- $\hat{y}$ is the predicted class.
- $T_i(X)$ represents the prediction of the i-th decision tree on input X, where $i = 1, 2, \ldots, m$ and m is the total number of trees in the forest.
- $\text{mode}(\cdot)$ denotes the majority vote function, which selects the most frequent class from the predictions of the individual decision trees.

The SelectKBest method selects the top k features based on their scores. The equation for computing the score of each feature can be represented as:

$$\text{Score}(X_i) = f(X_i, y) \tag{3}$$

Where:

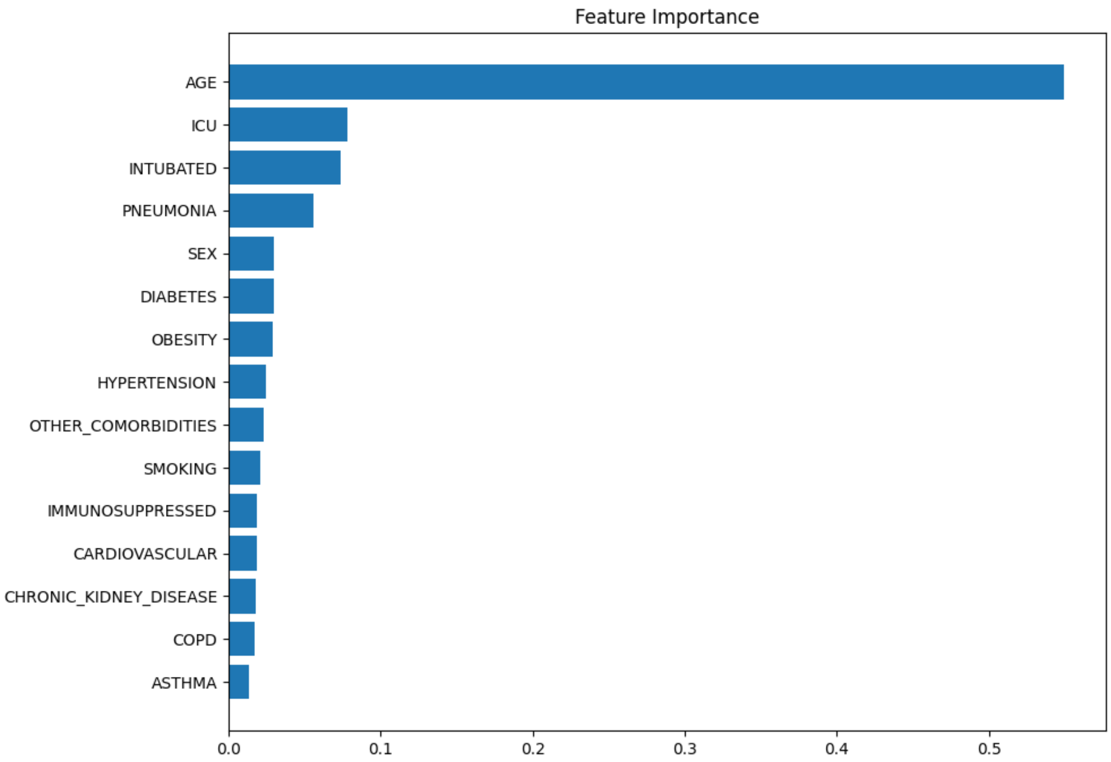

Figure 4. Feature Importance Analysis: RandomForestClassifier

- Score(X_i) is the score for feature X_i.
- $f(X_i, y)$ is a function that evaluates the relationship between the feature X_i and the target variable y. This function can be based on various statistical tests, such as ANOVA, Chi-square, or mutual information.

The PCA aims to transform the data into a new coordinate system where the greatest variance lies on the first coordinates (principal components). The equation for PCA can be written as:

$$Z = XW \tag{4}$$

Where:

- Z is the matrix of transformed features (principal components).
- X is the original data matrix with dimensions $n \times p$ (where n is the number of samples and p is the number of original features).
- W is the matrix of eigenvectors (principal components) derived from the covariance matrix of X.

The principal components are the directions of maximum variance, and the transformation is achieved by projecting the original data onto these components.

4. Model Phases and Design

The datasets were split into training and testing sets, with 80% of the data used for training and 20% reserved for testing. The predictive model utilized a hybrid LSTM-2BiGRU architecture [31, 32, 33, 34], designed to capture both short- and long-term dependencies in the sequential data.

4.1. Applying Model for First Dataset

For Egypt dataset 3.1, The model was trained using 60-time steps, meaning it used $60 days$ of past data to predict the number of cases within the next $7 days$. The model was configured with 100 units per layer, alongside a 30% dropout rate to prevent overfitting as shown in figure 5. The trade-off between the number of units and overfitting was considered: a higher number of units allows the model to learn complex patterns but increases the risk of overfitting. In contrast, fewer units simplify the model but may limit its capacity to identify intricate relationships in the data.

```
Layer (type)              Output Shape            Param #
=================================================================
lstm_4 (LSTM)               (None, 60, 100)         40800
_________________________________________________________________
dropout_12 (Dropout)        (None, 60, 100)          0
_________________________________________________________________
bidirectional_8 (Bidirection (None, 60, 128)          63744
_________________________________________________________________
dropout_13 (Dropout)        (None, 60, 128)          0
_________________________________________________________________
bidirectional_9 (Bidirection (None, 128)              74496
_________________________________________________________________
dropout_14 (Dropout)        (None, 128)              0
_________________________________________________________________
dense_4 (Dense)             (None, 7)               903
=================================================================
Total params: 179,943
Trainable params: 179,943
Non-trainable params: 0
```

Figure 5. LSTM-2BiGRU Model Summary for Egypt Dataset

To assess the performance of the model, several key metrics[35, 36, 37] were calculated and the output of evaluation during training model is shown in figure6:
Mean Absolute Error (MAE): MAE is used to measure the average magnitude of the errors in a set of predictions, without considering their direction. It is calculated as eq5:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

where:

- y_i: Actual value.

- $\hat{y}_i$: Predicted value.
- n: Total number of data points.

Mean Absolute Percentage Error (MAPE): MAPE calculates the average percentage error between predicted and actual values, providing a normalized view of the prediction accuracy. The formula is eq6:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{6}$$

Where:

- y_i: Actual value.
- $\hat{y}_i$: Predicted value.
- n: Total number of data points.

Root Mean Square Error (RMSE): RMSE is the square root of the average squared differences between actual and predicted values. It penalizes larger errors more than smaller ones, making it sensitive to outliers It is computed as eq7

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

Where:

- y_i: Actual value.
- $\hat{y}_i$: Predicted value.
- n: Total number of data points.

The performance of the proposed models was evaluated using precision, recall, accuracy, and F1-Score. These metrics were computed according to the following equations:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

where:

- TP: True positives.
- TN: True negatives.
- FP: False positives.
- FN: False negatives.

The LSTM-2BiGRU model achieved the following metrics: MAE of 8%, MAPE of 12%, RMSE of 10%, Precision of 76%, Recall of 68%, F1 Score of 72%, and Accuracy of 76%. These metrics illustrate the model's ability to predict future COVID-19 cases with a high degree of accuracy and minimal error. The results obtained in this study were compared with previous research, demonstrating the model's efficiency. A detailed comparison is presented in Table 1. The primary objective of this study is to validate the model's performance on the medical dataset from Mexico. We achieved the best results by considering long days for prediction (7 days), with low Error percentages 8%, and higher accuracy compared to the study [19] with the same dataset and days.

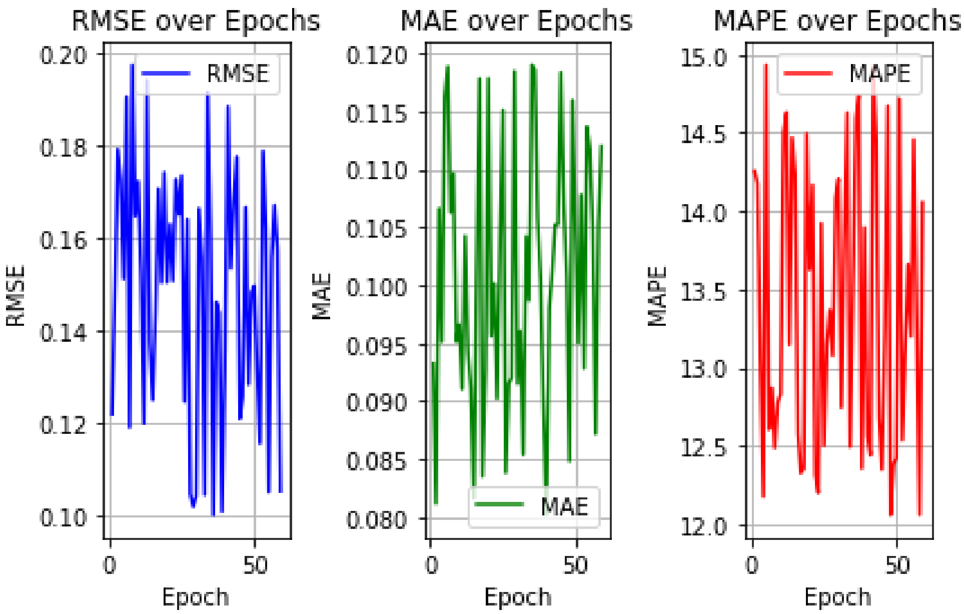

Figure 6. Model Evaluation: RMSE, MAE & MAPE over Epochs

Table 1. Comparison With Other Studies

Study Model	**Days**	**RSME, MAE, MAPE**	**Accuracy**
BILSTM [16]	14 days	MAE 33%	Not provided
SARIMAX India [22]	10 days	MAPE 23%	Not provided
Egypt LSTM [19]	7days	RSME 60%	72.71%
LSTM+GRU [15]	daily	MAE 1.9% RSME 9%	Not provided
This Study LSTM 2BIGRU	7 days	MAE 7.986% RSMS 9.669% MAPE 12.589%	76%

4.2. Applying Model for Second Dataset

For Mexico dataset 3.1 The 2 layers of LSTM model was developed with the following parameters: optimizer = 'adam', loss function ='`mean_squared_error`', and evaluation metrics = ['mae']. Callbacks were implemented to enable early stopping and model checkpointing based on validation loss, with the best model saved as 'best_model.h5' The model's performance was assessed using key metrics, as follows: Precision eq8: 0.625, Recall eq9: 0.462, F1-Score eq10: 0.531, AUC: 0.714 Overall accuracy eq11 was 68% fig9, with a macro average of 64% for both precision and recall. The weighted average precision, recall, and F1-score were 0.67, 0.68, and 0.67, respectively. The test mean absolute error (MAE) fig8, was reported as 0.415. Compared to traditional machine learning methods[38] fig7 , the LSTM and deep learning models demonstrated superior prediction accuracy when utilizing standard model parameters.

For same dataset the target column for results was binarized, with 0 representing negative cases and 1 for positive cases. The data was split into training and testing sets. A time-series framework was adopted by defining time steps of 60 days and a prediction horizon of 1 to 14 days ahead. The LSTM-2BiGRU model was configured with the

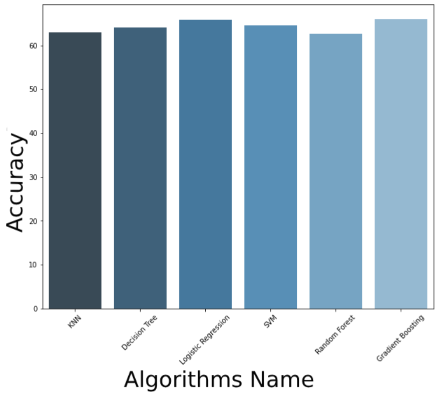

Figure 7. Traditional ML Methods Accuracy with 64%

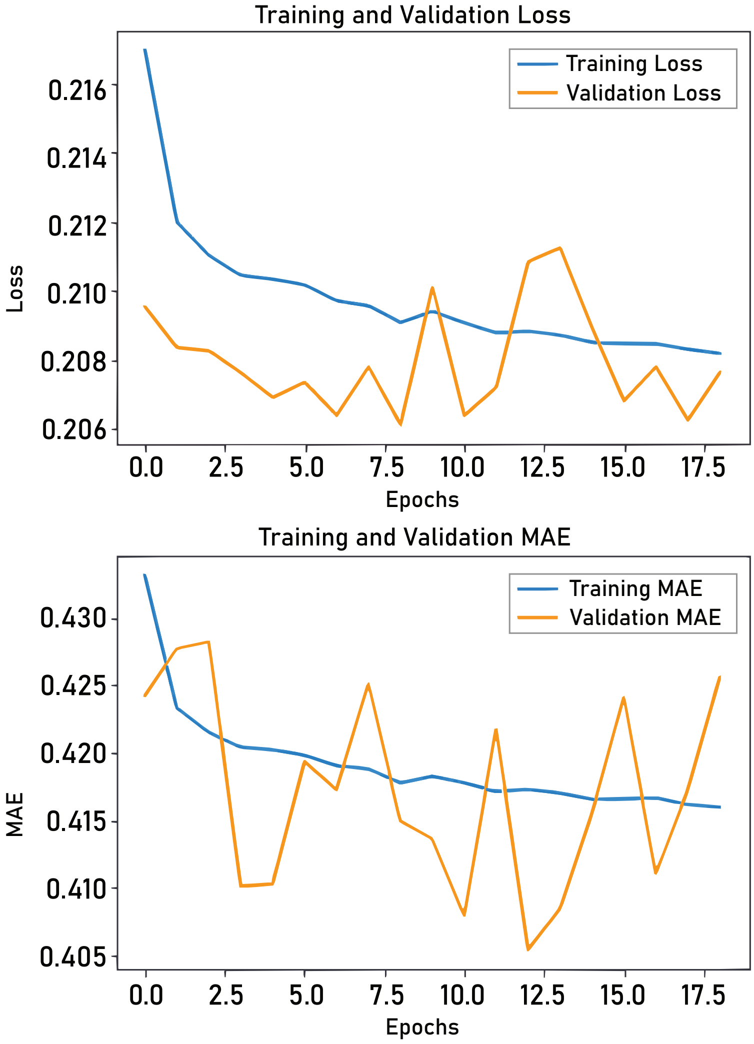

Figure 8. Test Mean Absolute Error & Loss

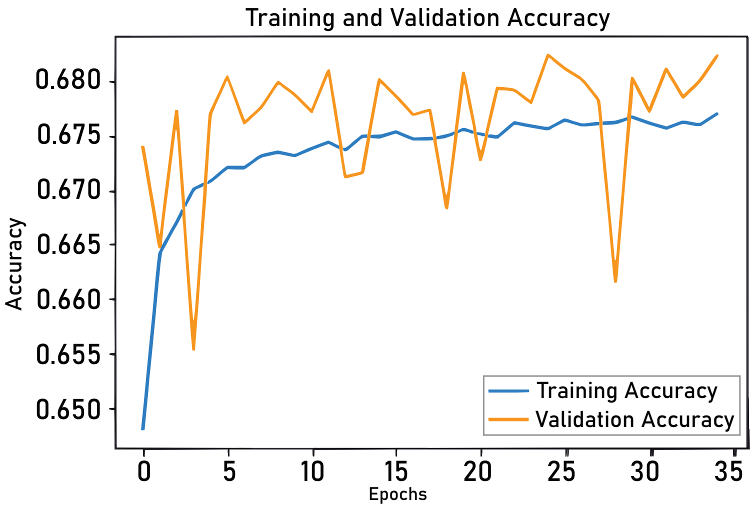

Figure 9. Accuracy of Prediction LSTM

```
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 60, 200)           170400
_________________________________________________________________
dropout_3 (Dropout)          (None, 60, 200)           0
_________________________________________________________________
bidirectional_2 (Bidirection (None, 60, 400)           482400
_________________________________________________________________
dropout_4 (Dropout)          (None, 60, 400)           0
_________________________________________________________________
bidirectional_3 (Bidirection (None, 400)               722400
_________________________________________________________________
dropout_5 (Dropout)          (None, 400)               0
_________________________________________________________________
dense_1 (Dense)              (None, 14)                5614
=================================================================
Total params: 1,380,814
Trainable params: 1,380,814
Non-trainable params: 0
_________________________________________________________________
```

Figure 10. LSTM-2BiGRU Model Summary for Mexico Dataset

Adam optimizerand a *sigmoid* eq12 activation function to accommodate the binary nature of the target variable. We utilized callbacks for early stopping and model checkpointing, and the model was trained over 500 epochs with a batch size of 128 and a dropout rate of 30% as shown in fig10 and below is the Algorithm Pseudo-code 1 represent step-by-step to build and evaluate the proposed model.

The sigmoid activation function [39] is widely used for binary classification tasks as it outputs a value between 0 and 1. This is particularly useful in determining the probability of a binary outcome.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{12}$$

Algorithm 1 Pseudocode for the Proposed Model

```
Input: Dataset $D$
Output: Prediction $\hat{y}$
Import Keras library for creating a deep learning model $M$
Initialize the model with parameters by setting timesteps=60, daysAhead=14, batchSize=128, epochs= 500
Preprocess the dataset $D$ with Features Engineering
for each training epoch do
    for each batch of data do
        Forward propagate the input through the model
        Compute the loss using the objective function
        Backpropagate the error
        Update the model parameters
    end for
end for
Evaluate the model on validation data by MAE, RMSE, MAPE
Output the final prediction $\hat{y}$
```

where:

- x is the input value,
  e is Euler's number (approximately 2.718)

The sigmoid function squashes the input xxx to a range of 0 to 1, making it suitable for binary classification problems, as the output can be interpreted as a probability. The model's performance was assessed using multiple metrics. The Mean Absolute Error (MAE) eq5 was calculated to be 0.118, while the Root Mean Square Error (RMSE) eq7 ranged between 0.130 and 0.140. The precision ranged from 0.88 to 0.92, recall from 0.85 to 0.90, and the F1-score between 0.86 and 0.91. The Area Under the Curve (AUC) ranged from 0.88 to 0.92, and the Mean Absolute Percentage Error (MAPE) eq6 was estimated at 10% to 15%. The overall Accuracy of the model averaged 86%, falling within a range of 80% to 90%.

4.3. Hyperparameter Tuning

To optimize the performance of our hybrid LSTM-BiGRU model, we conducted systematic hyperparameter tuning as shown in table:2. The following parameters were varied during the tuning process:

- **Number of LSTM/GRU Units**: Configurations with 100, 150, 200, and 250 units were tested to determine the optimal number of neurons in each recurrent layer.
- **Dropout Rates**: Values of 0.1, 0.2, and 0.3 were explored to balance model regularization and the prevention of overfitting.
- **Learning Rates**: Three learning rates (0.001, 0.005, and 0.01) were evaluated using the Adam optimizer.
- **Batch Sizes**: Batch sizes of 32, 64, and 128 were tested to balance training speed and convergence stability.
- **Prediction Horizon (*days_ahead*)**: The model was designed to predict between 1 and 14 days ahead based on historical data, enabling flexibility in short- and long-term forecasting.

Training Callbacks: The training process incorporated the following callbacks:

1. *Model Checkpointing*: To save the best model based on validation loss during training.

Table 2. Hyperparameters used in the Model

Hyperparameter	**Values**	**Brief**
Timesteps	60	This value determines the number of past time steps used as input for the model to make predictions
Prediction Horizon	day 1 to 14	enable the model to provide both short-term (1-day ahead) and medium-term (up to 14-days ahead) forecasts.
LSTM Units	200	The first LSTM layer contains 200 units, chosen to balance model complexity and computational efficiency
Bidirectional GRU Units	200	Two Bidirectional GRU layers are included, each with 200 units, to capture both forward and backward temporal dependencies in the data
Dropout Rate	0.3	Applied after each layer to prevent overfitting by randomly dropping 30% of neurons during training
Dense Layer Units	1 to 14	The final dense layer outputs predictions for the next 1-14 days.
Optimizer	Adam	Selected for its adaptive learning capabilities, allowing efficient optimization across a wide range of tasks
Loss Function	Mean Squared Error (MSE)	Used for continuous variable prediction to minimize the error between actual and predicted values.
Metrics	Mean Absolute Error (MAE)	Monitored during training to assess model performance.

2. *Early Stopping*: To halt training if the validation loss did not improve for a specified number of epochs, thereby avoiding unnecessary computation.

For each combination of hyperparameters, the model was trained for up to 500 epochs with a batch size of 128, and its performance was monitored using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) on the validation set. The best-performing hyperparameters were selected based on the lowest validation error.

5. Results and Discussions

The results obtained from the application of the LSTM-2BiGRU model on the epidemiological dataset, along with a comparative analysis between the performance of the LSTM-only model and the LSTM-2BiGRU hybrid model. The main objective is to demonstrate the effectiveness of the LSTM-2BiGRU model in improving prediction accuracy and performance in predicting COVID-19 outcomes.

5.1. Performance on the Epidemiological Dataset

The epidemiological dataset contained historical data from the General Directorate of Epidemiology, Secretariat of Health in Mexico, consisting of 41 features related to comorbidities such as asthma, pneumonia, and diabetes, which potentially affect the COVID-19 test results. After performing feature engineering and selection techniques, including RandomForestClassifier for feature importance analysis and dimensionality reduction using PCA, the most relevant 10 features were selected for the final model. The LSTM-2BiGRU model was built to predict positive and negative COVID-19 cases. Using the Adam optimizer and Sigmoid activation function, the model was trained for 500 epochs with a batch size of 128 and a dropout rate of 0.3. Callbacks for early stopping and model checkpointing were employed to optimize the model training process.

5.2. Comparative Analysis LSTM vs LSTM-2BiGRU

To assess the impact of using the LSTM-2BiGRU hybrid model, we compared its performance with a traditional LSTM-only model. The LSTM-only model achieved an accuracy of 68%, with moderate performance metrics including precision, recall, and F1-score. The application of the LSTM-2BiGRU model on the same dataset, however, significantly improved the results across all metrics. The LSTM-2BiGRU model achieved an average accuracy of 86%, with accuracy falling within a range of 80% to 90%. This represents a substantial improvement over the LSTM-only model. Other performance metrics further, demonstrate the advantages of using the hybrid model:

- Precision: 0.88 to 0.92
- Recall: 0.85 to 0.90
- F1-Score: 0.86 to 0.91
- AUC: 0.88 to 0.92
- Mean Absolute Error (MAE): 0.118
- Root Mean Squared Error (RMSE): 0.130 to 0.140
- Mean Absolute Percentage Error (MAPE): 10% to 15%

These results indicate that the LSTM-2BiGRU model outperforms the traditional LSTM model in terms of both predictive accuracy and error minimization

5.3. Impact of Feature Selection and Scaling

Feature engineering and selection played a critical role in the model's performance. By selecting only the most relevant features and scaling the target variable (COVID-19 test results), the model was better able to capture the underlying patterns in the data. The application of SelectKBest for feature selection and MinMaxScaler for scaling the data improved the convergence of the model and minimized prediction errors. The scaling of the dataset to a range of (0, 1) ensured that the model efficiently handled the input data, allowing for more effective learning. Additionally, the dimensionality reduction through PCA helped in reducing the complexity of the dataset while retaining the most informative features. This contributed to the overall improvement in model performance.

5.4. Ethical Considerations

This study utilized publicly available COVID-19 datasets, including epidemiological data from Mexico, which do not contain any personally identifiable information such as patient names or other private data. The data is aggregated and anonymized, ensuring compliance with ethical standards for research involving human-related datasets. We are committed to upholding principles of fairness and accountability in the development and application of machine learning models. While the datasets used in this study were unbiased in their structure and collection, we acknowledge the potential for algorithmic bias to emerge due to unequal representation of certain demographic or comorbidity-related features. To address this, we conducted feature importance analysis and model validation to ensure that our predictions remain fair and interpretable.

5.5. Discussion of Findings

The results demonstrate that the LSTM-2BiGRU model is highly effective in handling time-series data with complex, multivariate features. The hybrid deep learning architecture successfully captures both short-term and long-term dependencies within the data, leading to enhanced accuracy and robustness in predictions. The bidirectional nature of the BiGRU layers significantly contributes to the model's ability to learn from both past and future time steps, a feature particularly critical in epidemiological forecasting where data patterns are often nonlinear and multidirectional. The model's performance improvement over the LSTM-only architecture highlights the benefits of combining complementary deep learning techniques in hybrid models. This capability is especially advantageous for public health applications, where timely and accurate predictions can aid in decision-making during pandemics. Moreover, the reduction in error metrics and improved predictive accuracy compared to traditional machine learning models further validates the utility of the LSTM-2BiGRU framework for complex, multivariate time-series tasks. However, to enhance the model's real-world applicability, future work should prioritize optimizing computational efficiency. Techniques such as model pruning and quantization could be explored to reduce the model's size and latency, making it more suitable for real-time applications. These improvements could enable the deployment of the model in resource-constrained environments, such as mobile or edge devices, facilitating broader accessibility in public health contexts. In summary, the LSTM-2BiGRU model demonstrates a strong potential for predicting COVID-19 outcomes with high accuracy, reduced error metrics, and robust performance. The study underscores the significant advantages of hybrid deep learning models in addressing complex, multivariate time-series prediction challenges. Future research could also explore integrating decision-making components through Deep Reinforcement Learning (DRL) to further enhance predictive accuracy and applicability in real-world scenarios.

6. CONCLUSION

This study developed and evaluated a hybrid deep learning model, LSTM-2BiGRU, for predicting COVID-19 case trends using two distinct datasets from Egypt and Mexico. The results demonstrate that the proposed model significantly outperforms traditional machine learning models and standalone deep learning architectures. The LSTM-2BiGRU model achieved an accuracy of 76% with a Mean Absolute Error (MAE) of 8% for the Egypt dataset, and 80% to 90% accuracy for the Mexico dataset across various prediction periods. These findings highlight the model's ability to accurately capture the dynamic patterns of COVID-19 transmission and provide reliable forecasts for public health decision-making.

The key contribution of this study is the successful integration of LSTM and BiGRU architectures into a hybrid model, which enhances the ability to capture both short- and long-term dependencies in time-series data. This advancement is particularly important for epidemiological forecasting, where traditional models often fail to handle the complexities of sequential data. The LSTM-2BiGRU model presents a promising tool for predicting future COVID-19 trends and informing timely public health interventions. However, the model's complexity presents certain challenges, particularly in terms of computational demands and the need for high-quality data. Future work could focus on optimizing the model's efficiency, making it more suitable for real-time applications, and incorporating methods to handle missing or noisy data. Additionally, integrating real-time data streams, such as vaccination rates and population mobility, could further enhance the model's predictive power. Overall, this study demonstrates the potential of deep learning models like LSTM-2BiGRU to significantly improve COVID-19 case prediction accuracy, providing a valuable tool for health authorities in managing and mitigating the impact of pandemics.

REFERENCES

1. R. T. Eastman, J. S. Roth, J. S. Roth, K. R. Brimacombe, A. Simeonov, M. Shen, S. Patnaik, and M. D. Hall, “Remdesivir: A review of its discovery and development leading to emergency use authorization for treatment of covid-19,” *ACS central science*, vol. 6, no. 5, pp. 672–683, 2020.
2. WHO, “coronaviruses (covid-19).” https://www.who.int/emergencies/diseases/novel-coronavirus-2019/, 2020.
3. M. Zajenkowski, P. K. Jonason, M. Leniarska, and Z. Kozakiewicz, “Who complies with the restrictions to reduce the spread of covid-19: Personality and perceptions of the covid-19 situation,” *Personality and Individual Differences*, vol. 166, p. 110199, 2020.
4. N. Pearce, J. P. Vandenbroucke, T. J. VanderWeele, and S. Greenland, “Accurate statistics on covid-19 are essential for policy guidance and decisions,” *American Journal of Public Health*, vol. 110, no. 7, pp. 949–951, 2020. PMID: 32324422.
5. A. Nelson, “how the global pandemics compare including death toll, number of cases and symptoms.” , 2020.
6. N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, *et al.*, *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*, vol. 16. Imperial College London London, 2020.
7. Y. Zhang, B. Jiang, J. Yuan, and Y. Tao, “The impact of social distancing and epicenter lockdown on the covid-19 epidemic in mainland china: A data-driven seiqr model study,” *MedRxiv*, pp. 2020–03, 2020.
8. D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, H. J. Schünemann, A. El-Harakeh, A. Bognanni, T. Lotfi, M. Loeb, *et al.*, “Physical distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis,” *The lancet*, vol. 395, no. 10242, pp. 1973–1987, 2020.
9. Y. Zhou, R. Xu, D. Hu, Y. Yue, Q. Li, and J. Xia, “Effects of human mobility restrictions on the spread of covid-19 in shenzhen, china: a modelling study using mobile phone data,” *The Lancet Digital Health*, vol. 2, no. 8, pp. e417–e424, 2020.
10. S. K. Kumaravel, R. K. Subramani, T. K. Jayaraj Sivakumar, R. Madurai Elavarasan, A. Manavalanagar Vetrichelvan, A. Annam, and U. Subramaniam, “Investigation on the impacts of covid-19 quarantine on society and environment: Preventive measures and supportive technologies,” *3 Biotech*, vol. 10, pp. 1–24, 2020.
11. A. Solanki and S. Pandey, “Music instrument recognition using deep convolutional neural networks,” *International Journal of Information Technology*, vol. 14, no. 3, pp. 1659–1668, 2022.
12. R. Ahuja, A. Solanki, and A. Nayyar, “Movie recommender system using k-means clustering and k-nearest neighbor,” in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 263–268, IEEE, 2019.
13. A. Tayal, U. Kose, A. Solanki, A. Nayyar, and J. A. M. Saucedo, “Efficiency analysis for stochastic dynamic facility layout problem using meta-heuristic, data envelopment analysis and machine learning,” *Computational Intelligence*, vol. 36, no. 1, pp. 172–202, 2020.
14. A. Tayal, A. Solanki, and S. P. Singh, “Integrated frame work for identifying sustainable manufacturing layouts based on big data, machine learning, meta-heuristic and data envelopment analysis,” *Sustainable Cities and Society*, vol. 62, p. 102383, 2020.
15. D. Fedman, “Deep learning sequence models for forecasting covid-19 spread and vaccinations,” pp. 357–377, Lecture notes in networks and systems, 2023.
16. N. N. Aung, J. Pang, M. C. H. Chua, and H. X. Tan, “A novel bidirectional lstm deep learning approach for covid-19 forecasting.,” *Dental science reports*, vol. 13, no. 1, pp. 17953–17953, 2023.
17. S. P. Cumbane and G. Gidófalvi, “Deep learning-based approach for covid-19 spread prediction,” *International journal of data science and analytics*, 2024.
18. M. Marzouk, N. Elshaboury, A. Abdel-Latif, and S. Azab, “Deep learning model for forecasting covid-19 outbreak in egypt,” *Process Safety and Environmental Protection*, vol. 153, pp. 363–375, 2021.
19. N. E. M. Khalifa, A. A. Mawgoud, A. Abu-Talleb, M. H. N. Taha, and Y.-D. Zhang, “A covid-19 infection prediction model in egypt based on deep learning using population mobility reports,” *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, 2023.
20. A. A. Torad, F. Shamy, A. M. Kadry, and Z. S. Ahmed, “Using machine learning models to investigate the relationship between corporeal workload and clinical and epidemiological features of patients infected with covid-19 in egypt.,” *JPMA. The Journal of the Pakistan Medical Association*, vol. 73, no. 4, pp. S242–S246, 2023.
21. faegheh ghofrani, “Covid-19 vaccination prediction: Xgboost vs lgbm.” https://www.kaggle.com/code/faeghehghofrani/covid-19-vaccination-prediction-xgboost-vs-lgbm, 2024.
22. A. Solanki and T. Singh, “Covid-19 epidemic analysis and prediction using machine learning algorithms,” vol. 324, pp. 57–78, Nature Publishing Group, 2021.
23. S. Bouktif, A. M. U. D. Khanday, and A. Ouni, “Bi-directional lstm-based covid-19 detection using clinical reports,” *The eurasia proceedings of science, technology, engineering mathematics*, 2023.
24. OurWorldInData, “Ourworldindata covid-19 egypt.” https://ourworldindata.org/coronavirus/country/egypt, 2023.
25. E. Dong, H. Du, and L. Gardner, “An interactive web-based dashboard to track covid-19 in real time,” *The Lancet infectious diseases*, vol. 20, no. 5, pp. 533–534, 2020.
26. Kaggle, “Epidemiology data for positive and negative covid-19 cases in mexico.” https://www.kaggle.com/datasets/marianarfranklin/mexico-covid19-clinical-data, 2020.
27. B. Deepa and K. Ramesh, “Epileptic seizure detection using deep learning through min max scaler normalization,” *Int. J. Health Sci*, vol. 6, pp. 10981–10996, 2022.
28. K. Jain, P. Ghosh, and S. Gupta, “A hybrid model for sentiment analysis based on movie review datasets,” *International Journal on Recent and Innovation Trends in Computing and Communication*, 2023.
29. A. M. ElMassry, A. Alshamsi, A. F. Abdulhameed, N. Zaki, and A. N. Belkacem, “Machine learning approaches for sentiment analysis on balanced and unbalanced datasets,” pp. 18–23, 2024.
30. R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, “Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications,” *Journal of Big Data*, vol. 11, no. 1, p. 113, 2024.
31. Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: Lstm cells and network architectures,” *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

32. P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, “Forecasting cryptocurrency prices using lstm, gru, and bi-directional lstm: a deep learning approach,” *Fractal and Fractional*, vol. 7, no. 2, p. 203, 2023.
33. H. M. Lynn, S. B. Pan, and P. Kim, “A deep bidirectional gru network model for biometric electrocardiogram classification based on recurrent neural networks,” *IEEE Access*, vol. 7, pp. 145395–145405, 2019.
34. F. M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed, “A comprehensive overview and comparative analysis on deep learning models: Cnn, rnn, lstm, gru,” *arXiv preprint arXiv:2305.17473*, 2023.
35. A. Botchkarev, “Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology,” *arXiv preprint arXiv:1809.03006*, 2018.
36. P. Singla, M. Duhan, and S. Saroha, “Review of different error metrics: A case of solar forecasting,” *AIUB Journal of Science and Engineering (AJSE)*, vol. 20, no. 4, pp. 158–165, 2021.
37. O. Rainio, J. Teuho, and R. Klén, “Evaluation metrics and statistical tests for machine learning,” *Scientific Reports*, vol. 14, no. 1, p. 6086, 2024.
38. Kaggle, “Covid-19 prediction using classification models in mexico.” https://www.kaggle.com/code/ritikmeena/covid-19-prediction, 2020.
39. H. Pratiwi, A. P. Windarto, S. Susliansyah, R. R. Aria, S. Susilowati, L. K. Rahayu, Y. Fitriani, A. Merdekawati, and I. R. Rahadjeng, “Sigmoid activation function in selecting the best model of artificial neural networks,” in *Journal of Physics: Conference Series*, vol. 1471, p. 012010, IOP Publishing, 2020.