OPEN ACCESS
CC
KJMR
Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 1 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

A DEEP LEARNING APPROACH TO OPTIMIZING COVID-19 MORTALITY PREDICTIONS WITH RECURRENT NEURAL NETWORKS (RNNs)

Fazal Malik*
Department of Computer Science, Iqra National University Peshawar, Khyber Pakhtunkhwa (KPK), Pakistan

Hamid Ullah
Department of Computer Software Engineering, University of Engineering and Technology, Mardan, Khyber Pakhtunkhwa, Pakistan

Ashraf Ullah
Institute of Computer Science and Information Technology, University of Science and Technology Bannu, Khyber Pakhtunkhwa (KPK), Pakistan

Said Khalid Shah
Institute of Computer Science and Information Technology, University of Science and Technology Bannu, Khyber Pakhtunkhwa (KPK), Pakistan

Rahmat Hussain
Institute of Computer Science and Information Technology, University of Science and Technology Bannu, Khyber Pakhtunkhwa (KPK), Pakistan

Muhammad Javed
Institute of Computer Science and Information Technology, University of Science and Technology Bannu, Khyber Pakhtunkhwa (KPK), Pakistan

Corresponding author: fazal.malik@inu.edu.pk
DOI: https://doi.org/10.71146/kjmr243

Article Info

Abstract

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has posed significant challenges to global health since its onset in December 2019. While numerous models have been developed to predict COVID-19 outcomes, many fall short in accurately forecasting mortality rates, focusing instead on clinical severity. This study addresses these gaps by proposing an advanced approach using Recurrent Neural Networks (RNNs) to improve mortality prediction. Although existing models have employed various machine learning (ML) and deep learning (DL) techniques, they often rely on limited data sources and lack a dedicated focus on global mortality forecasting. Our research introduces RNNs as a primary tool for addressing this limitation, enhancing predictive accuracy in the context of COVID-19 deaths. The methodology involves four key phases: (1) Data Acquisition from the "COVID-19 Symptoms Dataset," which includes variables such as symptoms, confirmed cases, and death rates; (2) Data Processing, which encompasses data cleaning, visualization, and feature selection; (3) Supervised Learning using the RNN model for prediction; and (4) Performance Evaluation, where the RNN's performance is assessed in terms of accuracy, precision, recall, and F1-score. Results show that the RNN model significantly outperforms previous approaches, achieving an accuracy of 94%, precision of 92%, recall of 93%, and F1-score of 92%. This study demonstrates the effectiveness of RNNs in enhancing COVID-19 mortality predictions, providing critical insights for better global forecasting and aiding public health efforts in resource allocation and response planning.

Keywords:
COVID-19 Pandemic, Mortality Prediction, Deep Learning (DL), Clinical Severity, Recurrent Neural Networks (RNNs), Supervised Learning, Data Analysis.

Introduction

A novel coronavirus identified in December 2019 in Wuhan, China, was linked to pneumonia cases and rapidly spread, leading to a global pandemic. The World Health Organization (WHO) declared, COVID-19 a public health emergency in January 2020 and later a pandemic in March 2020. COVID-19, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has affected over 81,552 cases in China and more than 1.4 million cases globally. The virus spread faster outside China, notably in the USA, Italy, and Spain. Initially termed 2019-nCoV, the virus was later officially named SARS-CoV-2 [1].

The virus is known as a coronavirus, and the disease it causes is termed COVID-19 (where "co" stands for "corona," "vi" for "virus," and "d" for "disease"). Coronaviruses are a family of viruses ranging from the common cold to more severe illnesses like Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). These viruses typically spread among animals, with some capable of transmission between animals and humans, known as zoonotic events. COVID-19 represents a newly discovered strain that has caused an ongoing outbreak of respiratory disease [2]. For instance, Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) emerged in Mainland China and Hong Kong in 2003, causing significant respiratory issues. Similarly, in 2012, the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) led to an outbreak of MERS in Saudi Arabia, the United Arab Emirates, the Republic of Korea, and other countries [3]. Both SARS-CoV and MERS-CoV were new to science, but through human intervention, these outbreaks were contained. By 2020, coronaviruses, particularly SARS-CoV-2, which causes COVID-19, had become widely recognized globally. While SARS-CoV-2 is part of a larger family of viruses, not all may realize that it is linked to previous infectious outbreaks. Scientists have acquired significant knowledge about coronaviruses, but gaps remain, and the possibility of a coronavirus pandemic was largely unforeseen before the 21st century [4].

The International Committee on Taxonomy of Viruses has identified over 40 coronaviruses, with seven infecting humans, including SARS-CoV1, MERS-CoV, and SARS-CoV-2, noted for high mortality rates [5]. Electronic medical records of Ethiopian COVID-19 patients were analysed using seven machine learning models: J48 decision tree, random forest (RF), k-nearest neighbour (k-NN), multi-layer perceptron (MLP), Naïve Bayes (NB), extreme gradient boosting (XGBoost), and logistic regression (LR). Metrics like sensitivity, specificity, precision, and receiver operating characteristic (ROC) assessed model performance [6]. Coronaviruses, initially zoonotic, undergo genetic modifications enabling human transmission, often causing severe infections. Ongoing reinfections by mild coronaviruses like OC43 and 229E necessitate vigilance [7]. The Kaggle COVID-19 dataset provides details on patients' symptoms, death cases, mortality rates, and confirmed cases. Key symptoms include dry cough, high fever, sore throat, and difficulty breathing, appearing within 2–14 days [8]. Fever involves elevated temperature (irregular, remittent, intermittent, or continuous). Persistent cough may indicate bronchitis or pneumonia, while shortness of breath (dyspnea) manifests as air hunger or chest tightness, worsened by exertion, extreme temperatures, or high altitude [9].

Machine Learning (ML) and Deep Learning (DL) techniques have been extensively applied to predict COVID-19 mortality. A cost-effective tool employing Support Vector Machine (SVM) and Feed forward Neural Network (FNN) models utilized clinical and radiomic features from high-resolution computed tomography (HRCT) scans. The integration of these features achieved the highest area under the receiver operating characteristic (AUC), demonstrating its effectiveness in clinical settings [10]. Robust models incorporating radiomic and neural network features from chest X-rays (CXRs) used certified software and data from 1,816 patients across five hospitals. The pipeline included feature selection and mortality prediction using AdaBoost (ADA), Quadratic Discriminant Analysis (QDA), and Random Forest (RF). Metrics such as accuracy (ACC), AUC, and sensitivity (SENS) validated strong predictive performance across balanced and imbalanced datasets [11]. Recent advancements feature three supervised deep learning models: Convolutional Neural Network (CNN)-based CV-CNN, Long Short-Term Memory (LSTM) combined with CNN in CV-LSTM + CNN, and IMG-CNN using converted clinical dataset images, enhancing COVID-19 mortality prediction [12]. These techniques highlight the importance of ML and DL in addressing pandemic-related challenges.

Advanced deep learning techniques, including Long Short-Term Memory (LSTM), bidirectional LSTM, Convolutional Neural Networks (CNN), hybrid CNN-LSTM, Multilayer Perceptron's (MLP), and Recurrent Neural Networks (RNN), were compared for predicting COVID-19 mortality. Bayesian optimization improved performance using a dataset of case counts, demographics, and socioeconomic factors, revealing model efficacy variations to guide public health strategies [13]. Gradient Boosting outperformed 14 Machine Learning (ML) algorithms using OpenDataSus data, including vaccination information [14].

The global COVID-19 pandemic presents a significant health threat, especially to vulnerable populations like the elderly and neonates. Existing Machine Learning (ML) and Deep Learning (DL) models primarily focus on predicting clinical severity, limiting their broader applicability in forecasting mortality and reducing their generalizability across diverse populations. These models are often constrained by localized data, hindering their effectiveness for global predictions.

To address these limitations, this study proposes an optimized approach using advanced Deep Learning techniques, specifically Recurrent Neural Networks (RNNs), to improve the accuracy of global COVID-19 mortality predictions. The study utilizes the "COVID-19 Symptoms Dataset" from Kaggle, which includes key features such as symptoms, confirmed cases, and death rates. RNNs are employed to capture temporal dependencies and complex data patterns, enhancing the model's ability to predict mortality rates.

The methodology is divided into four phases: Phase-1 (Data Acquisition) involves using the Kaggle dataset, including symptoms like cough, fever, and difficulty breathing. Phase-2 (Data Processing and Analysis) includes pre-processing steps, such as handling missing values and normalization, followed by exploratory analysis to detect trends. Phase-3 (RNN Implementation) involves training the model to predict mortality by using sequential data patterns, with techniques like data augmentation to address imbalanced datasets. Phase-4 (Performance Evaluation) compares the model's performance with existing approaches, using metrics such as precision, recall, F1-score, and accuracy.

This study introduces RNNs as a significant improvement for predicting COVID-19 mortality, overcoming the limitations of prior models focused on clinical severity with limited data. The findings demonstrate that RNNs provide deeper insights, improving global prediction accuracy and facilitating better resource allocation and public health planning.

The subsequent sections will review current methodologies in Section Literature Review, outline the proposed framework in Section Methodology, discuss experimental results in Section Results and Discussion, and provide recommendations for further research in Section Conclusion.

1. Literature Review

This review analyses techniques predicting COVID-19 mortality rates using the COVID-19 Symptoms Dataset, including symptoms, death cases, and confirmed cases. Various techniques are focusing on features like dry cough, fever, sore throat, and difficulty breathing.

COVID-19 mortality risk factors were identified using decision tree and dimension reduction algorithms in a retrospective study of 3,008 Iranian patients (March-November 2020). Key predictors included old age, chest pain, low respiratory rate, oxygen saturation <93%, mechanical ventilator use, and neurological/cardiovascular disorders. Gender, fever, myalgia, dizziness, and gastrointestinal symptoms were not significant. Machine learning models efficiently identified critical factors for early management of high-risk patients. The C4.5 decision tree algorithm classified based on entropy. Dimension reduction techniques, including Principal Component Analysis (PCA), Partial Least Squares (PLS), and t-distributed Stochastic Neighbour Embedding (t-SNE), highlighted key features for multivariate analysis [15]. Predictive models were developed using J48 decision tree, random forest (RF), k-nearest neighbour (k-

NN), multi-layer perceptron (MLP), Naïve Bayes (NB), extreme gradient boosting (XGBoost), and logistic regression (LR) [16].

COVID-19 transmission is significantly affected by environmental factors, with increased infectivity and mortality in colder climates. Seasonal variations, analysed using Ensemble Empirical Mode Decomposition (EEMD), indicate a 59.71% rise in colder Southern Hemisphere countries and a 46.38% decrease in warmer Northern Hemisphere regions. A modified Susceptible, Exposed, Infectious, Recovered (SEIR) model integrating seasonal factors quantifies these effects. Seasonality alone is insufficient to control transmission, highlighting the need for continuous public health measures and readiness for colder season surges [17].

The recent coronavirus disease 2019 (COVID-19) pandemic, impacting countries globally, has significantly increased interest in mathematical models for describing disease dynamics and predicting pandemic progression. Notably, it was observed that even a basic logistic model, utilizing a first-order ordinary differential equation, offered reasonable accuracy in capturing case trends across various nations during the initial wave of the pandemic [18]. The study compares the latest techniques for predicting COVID-19 patient mortality using Machine Learning (ML) and Deep Learning (DL) methods. Techniques such as Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Logistic Regression, Random Forest, and Decision Tree are evaluated for their predictive efficacy. The G power calculation determined sample size requirements with an acceptable error of 0.5 [19].

In the past two decades, outbreaks of animal Beta corona viruses caused severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). Since December 2019, the global spread of SARS-CoV-2 (2019-nCoV), originating in Wuhan, China, has heightened public health concerns. Although SARS-CoV-2 has a lower fatality rate compared to SARS-CoV and MERS-CoV, its rapid transmission has led to severe pneumonia and respiratory illness. Recent advancements in Machine Learning (ML) and Deep Learning (DL), such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), have been utilized to enhance the accuracy of predicting COVID-19 patient mortality [20]. Accurate prediction of COVID-19 patient mortality remains challenging. Recent research has utilized advanced techniques, including Polynomial Regression, Auto-Regressive Integrated Moving Average (ARIMA), and Deep Learning methods such as Recurrent Neural Networks (RNN). Polynomial Regression demonstrated the best fit for India, whereas ARIMA and RNN models effectively predicted trends in South Korea, Italy, the United States, and the United Kingdom [21].

Deep learning techniques, particularly Convolutional Neural Networks (CNNs) with transfer learning, are employed to diagnose and monitor COVID-19 using X-ray images, aiding radiologists and doctors. Challenges like limited and imbalanced datasets, model overfitting, and symptom similarity between COVID-19 and pneumonia are addressed by an automated solution that classifies COVID-19 into two categories using nine advanced CNN architectures [22]. Early detection of COVID-19 patient mortality is enhanced by exploring advanced Machine Learning (ML) and Deep Learning (DL) techniques. A hybrid model integrates ML and DL methods, employing ten deep Convolutional Neural Network (CNN) models for feature extraction from CT images. Features from various CNN layers are classified using five ML classifiers. The dataset comprises 2,481 CT images categorized into COVID-19 and non-COVID-19 groups [23]. Artificial Neural Networks (ANN) and Logistic Regression (LR) models were employed to analyse data from 26,867 PCR-positive COVID-19 patients. Key mortality predictors identified by ANN included decreased consciousness, cough, PO2 level, age, chronic kidney disease, fever, headache, smoking status, chronic blood diseases, and diarrhea [24]. The status of artificial intelligence (AI) applications in clinical settings for COVID-19 highlighted the use of big data and AI techniques, including neural networks, classical Support Vector Machine (SVM), and edge learning [25].

Support Vector Machine (SVM) with hyperparameter tuning via the Taguchi method was employed to analyse 19 variables affecting COVID-19 cases and mortality. The hybrid SVM-Taguchi approach enhanced predictions for confirmed cases and deaths, showing superior statistical performance compared to other artificial intelligence methods [26]. Daily development rates and doubling times of cases were used to establish moderation, control, and containment benchmarks (development rates of <10%, 1%, and 0.1%, respectively) with interventions like isolation and lockdown [27].

COVID-19 mortality rates are crucial for assessing disease severity and predicting death rates. Machine learning techniques, including Regression models, XGBoost, Random Forest, and Support Vector Machine (SVM), analyse risk factors and mortality. These models effectively forecast future mortality in epidemics. The Covid-GAN model, based on an Auxiliary Classifier Generative Adversarial Network (ACGAN), generates synthetic chest X-ray images, enhancing detection when integrated with Convolutional Neural Networks (CNNs) [28]. Artificial Intelligence (AI) aids COVID-19 analysis. The CNN-GRU model, developed using a dataset of 4,692 cases, predicts mortality [29]. A recent approach using chest X-ray images and AI methods, including k-nearest neighbour, Bayesian methods, and SVM, showed notable results with Mobile Net and SVM classifiers [30]. Machine learning models for predicting COVID-19 cases identified Exponential Smoothing (ES) as the most effective, followed by Logistic Regression (LR) and Least Absolute Shrinkage and Selection Operator (LASSO), while SVM showed variable performance [31]. Deep learning (DL) analysed Lung Superman (LUS) images, assessing infection severity at edge, video, and pixel levels. Advanced models, such as Spatial Transformer Networks, were introduced for severity prediction. A depth model benchmarked pixel-level COVID-19 biomarkers, yielding promising results with ongoing research [32]. Recent studies on mortality prediction using Machine Learning (ML) and Deep Learning (DL) found LSTM most effective for death predictions [33].

Recent studies demonstrate the effectiveness of Machine Learning (ML) and Deep Learning (DL) techniques in predicting COVID-19 mortality. Prominent methods include Decision Tree (DT) classifiers, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Logistic Regression (LR) [34]. A study examines the use of machine learning (ML) algorithms to predict mortality in COVID-19 patients. An analysis was conducted using a dataset from Kaggle, evaluating six ML algorithms: KNN, Naive Bayes, SVM, Decision Tree, Random Forest, and Logistic Regression. The analysis identified key predictors of mortality, emphasizing that symptoms are critical for predicting COVID-19-related mortality. The study concludes that the proposed method can be easily updated with new data, aiding frontline doctors in making timely clinical decisions [35].

2. Methodology

The methodology employs a systematic approach to predict COVID-19 mortality rates using the "COVID-19 Symptoms Dataset" from Kaggle. It involves four phases: data acquisition, data processing and analysis with Python's powerful libraries, supervised learning using Recurrent Neural Networks (RNNs) to capture patterns in symptoms and performance evaluation against existing methods. The Long Short-Term Memory (LSTM) model is used to handle time-series data, focusing on key symptoms linked to higher mortality rates, thus improving prediction accuracy. The process is illustrated in Figure 1 and detailed in Algorithm 01.

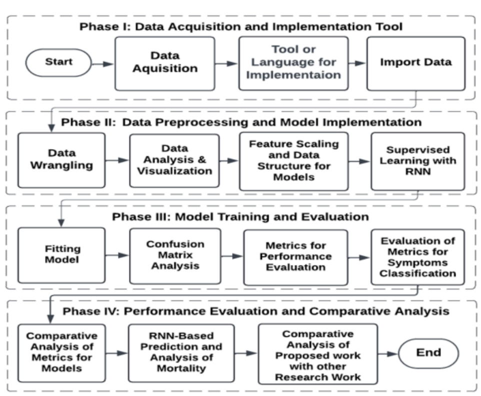

Figure 1. **Proposed work's block diagram**

The proposed research methodology consists of four stages: Data Acquisition: An extensive COVID-19 dataset from Kaggle was utilized, containing patient symptoms, death cases, mortality rates, and confirmed cases. Symptoms include dry cough, high fever, sore throat, and difficulty in breathing, processed using Python and Anaconda Jupiter Notebook for quality analysis. Data Processing and Analysis: Data visualization techniques, such as scatter plots and histograms, along with dimensionality reduction methods, identified key features differentiating symptoms, death cases, mortality rates, and confirmed cases. Supervised Learning: Deep learning with Recurrent Neural Networks (RNNs) enhances COVID-19 death case prediction, capturing temporal dependencies and improving model generalization through optimized architecture and data augmentation. Performance Evaluation: The RNN model's effectiveness was assessed against current approaches using accuracy, precision, recall, and F1-score, providing a foundation for future research in disease prediction and healthcare improvement.

Algorithm 01: Optimized COVID-19 Mortality Prediction

Input: COVID-19 Symptoms Dataset from Kaggle
Output: Model Performance Metrics and Comparative Analysis

Step 1. Data Acquisition and Preprocessing

// Load the dataset containing COVID-19 symptoms, mortality rates, and other relevant data.

1.1. COVID19_dataset ← Load Dataset("Kaggle/COVID19_Symptoms")

// Extract relevant features (e.g., symptoms, confirmed cases, mortality rates).

1.2. features ← Extract Features(COVID19_dataset)

// Clean the dataset by removing null values, duplicates, and irrelevant information.

1.3. clean data ← Clean Data(features)

// Perform normalization or scaling on the dataset for consistency in model training.

1.4. norm data ← Normalize Data(clean data)

// Split the dataset into training and testing sets.

1.5. train data, test data ← Split Dataset(norm data, train size)

Step 2. Data Visualization

// Visualize the distribution of key symptoms to understand the dataset better.

2.1. symptom_distribution_plot ← PlotSymptomDistribution(train data)

// Generate a heatmap to visualize correlations between symptoms and mortality rates.

2.2. correlation heatmap ← PlotCorrelationHeatmap(train_data)

// Plot the time series of daily mortality rates.

2.3. mortality_trend_plot ← PlotMortalityTrend(train_data)

Step 3. Supervised Learning with RNN

// Initialize the Recurrent Neural Network (RNN) model for time-series analysis.

3.1. rnn_model ← InitializeRNN()

// Train the RNN model using the training dataset.

3.2. trained_rnn_model ← TrainRNN(rnn_model, train_data)

Step 4. Model Evaluation

// Evaluate the performance of the model using test data.

4.1.*performance_metrics ← EvaluateModel(trained_rnn_model, test_data)*
// Calculate accuracy, precision, recall, and F1-score.
4.2.*accuracy, precision, recall, f1_score ← CalculateMetrics(performance_metrics)*
// Generate a confusion matrix to assess the classification performance.
4.3. *confusion_matrix ← GenerateConfusionMatrix(performance_metrics)*
Step 5. ***Comparative Analysis and Discussion***
// Compare the proposed model's accuracy with existing models in the literature.
5.1. *accuracy_comparison ← CompareAccuracy(trained_models, existing_models)*
// Identify areas of significant improvement over past methods.
5.2. *improvement_areas ← IdentifyImprovements(trained_models, test_data)*
//Discuss the implications of the findings and potential applications in healthcare.
5.3. *implications ← DiscussImplications(improvement_areas, accuracy_comparison)*
Step 6. ***End***

2.1. Phase I: Data Acquisition and Implementation Tool

2.1.1. Data Acquisition

In this research, we use the "COVID-19 Symptoms Dataset" from Kaggle, which provides comprehensive information on patients' symptoms, death cases, mortality rates, and confirmed cases. This dataset features various symptoms, including dry cough, high fever, sore throat, and difficulty in breathing. It is an extensive open-source repository that addresses real-world challenges related to COVID-19 [36]. The dataset, derived from observations published by the Government of India, records symptoms over multiple days and is open to further improvements. We selected this dataset to identify significant COVID-19 symptoms and enhance prediction accuracy for mortality rates [37].

2.1.2. Tool or Application for Implementation

We selected Python as the programming language for its simplicity, efficiency, readability, and robust libraries like TensorFlow and Keras, which facilitate deep learning tasks such as Recurrent Neural Networks (RNNs) for time-series analysis and sequential data modeling. Python's extensive use in artificial intelligence (AI) and data science enhances its suitability for our research. For development, Jupyter, an open-source, web-based environment, was chosen for its seamless Python integration, real-time code execution, and visualization capabilities, which are essential for iterative model development and evaluation, particularly with RNNs in sequential data tasks.

2.1.3. Importing Required Libraries

The Recurrent Neural Networks (RNNs) algorithm is implemented in Python using the Anaconda Jupyter Notebook environment, with Pandas for managing tabular data, NumPy for numerical operations, and Seaborn and Matplotlib for visualization. These libraries ensure efficient data manipulation, numerical analysis, and visualization, improving result accuracy. RNNs, well-suited for processing sequential data, capture temporal dependencies, enhancing predictive accuracy. Their ability to model time-series data addresses challenges in mortality prediction, providing deeper insights into patient outcomes and improving diagnostic accuracy in medical research.

2.2. Phase II: Data Processing and Model Implementation

2.2.1. Data Wrangling, and Pre-Processing

The data wrangling process involves cleaning and pre-processing the dataset to eliminate irrelevant information, such as unnecessary values and null entries, while addressing missing values, handling outliers, and transforming the data into a suitable format. This ensures the dataset is refined, error-free, and ready for analysis. In our study, Recurrent Neural Networks (RNNs) were employed as the primary modeling technique, requiring comprehensive data pre-processing to manage redundancies and missing values. Figure 2 shows a heatmap of missing values, emphasizing the importance of meticulous data preparation for improving RNN-based model accuracy and reliability.

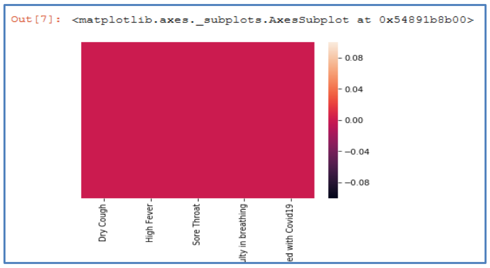

Figure 2. Heatmap of Missing Values in COVID-19 Symptom Data

Figure 2 displays a heatmap showing missing values in the dataset, where red areas indicate data presence and colour absence indicates missing values.

The heatmap is entirely red, signifying no missing values across features such as Dry Cough, High Fever, and COVID-19 infection status. This confirms the dataset's readiness for analysis. The absence of missing values validates the effectiveness of our data cleaning, ensuring a high-quality dataset for modeling.

Figure 3 illustrates the success of these efforts, showing a clean dataset with no missing or duplicated values, ready for further analysis.

```
Out[4]: Dry Cough                 0
        High Fever                0
        Sore Throat               0
        Difficulty in breathing   0
        Infected with Covid19     0
        dtype: int64
```

Figure 3. Heatmap illustrating the Absence of Missing Values in the Preprocessed Symptom Dataset

2.2.2. Data Analysis

This step explores data using techniques like descriptive statistics, visualization, and correlation analysis. Tables 1 and 2 shows the dataset including confirmed cases, deaths and mortality rates by country. RNNs are applied to improve COVID-19 death prediction. This model identifies patterns in symptoms, enhancing mortality forecasts. RNNs are employed to enhance the prediction of COVID-19 death cases (Tables 1 and 2).

Table 1. Symptom Frequencies and COVID-19 Infection Status

Dry Cough	High Fever	Sore Throat	Difficulty in Breathing	Infected with COVID-19
0	2	3	0	No
15	15	20	16	Yes
4	5	0	0	No
4	7	9	10	No
0	0	1	0	No

6	**0**	**6**	**0**	**No**
16	**17**	**18**	**16**	**Yes**

Table 2 provides an overview of COVID-19 statistics across various USA states, including confirmed cases, deaths, recoveries, mortality ratio, and geographical coordinates. For instance, California reports 796,436 confirmed cases with a mortality ratio of 1.92%, while Alaska has the lowest values. We propose using deep learning techniques, such as Recurrent Neural Networks (RNNs), to improve mortality prediction accuracy based on symptoms and other features. These models can help identify patterns and enhance global prediction efforts.

Table 2. **COVID-19 Mortality Ratio and Statistics by U.S. State**

State	Confirmed Cases	Deaths	Recovered	Mortality Ratio	Latitude	Longitude	USA State Code
Alabama	147153	2488	0	1.69	32.318231	-86.902298	AL
Alaska	7004	45	0	0.64	63.588753	-154.493062	AK
Arizona	215284	5525	0	2.57	34.048928	-111.093731	AZ
Arkansas	77963	1229	0	1.58	35.20105	-91.831833	AR
California	796436	15291	0	1.92	36.778261	-119.417932	CA

California reports the highest confirmed cases at 796,436 with a mortality ratio of 1.92%, while Alaska reports the lowest. We propose using deep learning (RNNs) techniques to enhance COVID-19 mortality prediction accuracy. The dataset aids these models in identifying patterns to improve global prediction accuracy and pandemic management.

2.2.3. Data Visualization

After data cleaning and wrangling, we focus on data visualization to identify patterns, trends, and anomalies, enhancing our understanding of relationships between variables. Using various techniques from the plotting library, we present the dataset graphically with Matplotlib, which facilitates clearer interpretation and aids in selecting appropriate algorithms.

Figure 4 illustrates COVID-19 mortality cases, highlighting fluctuations over 50 days. The data shows peaks and troughs, with a general downward trend in mortality cases.

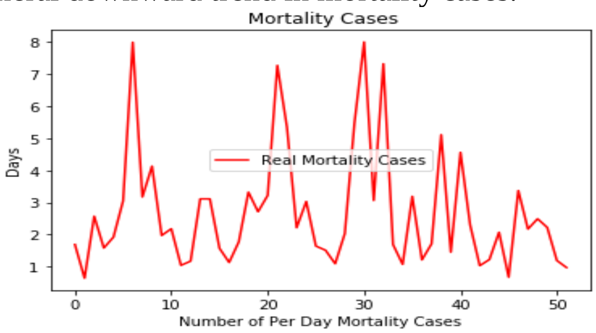

Figure 4. **Daily Mortality Cases Over 50 Days**

This trend may result from effective public health measures, but occasional spikes suggest inconsistencies, possibly due to new variants or lapses in interventions. Data visualization aids in revealing trends that are not immediately apparent from raw data, supporting decision-making. Next, we will apply deep learning regression techniques to predict mortality cases and evaluate model accuracy by comparing predicted and actual outcomes.

2.2.4. Feature Scaling and Data Structure for RNN

After pre-processing the COVID-19 dataset from Kaggle, feature scaling and data structuring are applied for RNNs.

- ***Feature Scaling for RNN***

Feature scaling, a pre-processing technique in machine learning, normalizes or standardizes data to ensure equal contribution of each feature to model predictions, preventing domination by features with larger value ranges.

For RNNs, feature scaling is essential, especially with time-series data like COVID-19 symptom progression. Normalization (scaling between 0 and 1) ensures input features, such as symptoms and mortality rates, have similar ranges, preventing disproportionate importance on larger values. Proper scaling stabilizes gradients during back propagation, facilitating effective learning and capturing temporal relationships for accurate predictions.

- ***Data Structure for RNN***

Data structure organizes and formats data for machine learning models, defining storage, access, and processing for effective model training and operation.

For RNNs, the data is structured to capture the sequential nature of time-series data, reflecting symptom progression for each patient. Sequences representing symptoms over specific periods are fed into the RNN, enabling it to predict future outcomes like severe symptoms or mortality. The proposed technique incorporates feature scaling and data structuring for RNNs, improving COVID-19 death prediction and global accuracy. This structured approach enhances model performance, ensuring reliable outcomes in predicting COVID-19 trends.

2.2.5. Supervised Learning with RNN Model

Supervised learning is fundamental in Artificial Intelligence (AI) and Machine Learning (ML), relying on labelled datasets to train models for predicting outputs on unseen data, particularly in classification and regression tasks. Within this paradigm, RNNs are widely used techniques [38].

RNNs, designed for sequential data, maintain hidden states to capture temporal dependencies, making them ideal for time series and Natural Language Processing (NLP) tasks. Advanced RNN variants, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), address vanishing gradients, ensuring robustness. Together, these techniques exemplify supervised learning's versatility in solving diverse challenges [39].

- ***Recurrent Neural Network***

Recurrent Neural Networks (RNNs) are specialized neural networks designed for sequential data, excelling in tasks like time series analysis, natural language processing (NLP), and sequence-based predictions. For predicting COVID-19 mortality rates, RNNs effectively capture temporal dependencies between symptoms and outcomes, critical for the Kaggle COVID-19 dataset analysis. The dataset includes features like dry cough, high fever, sore throat, and difficulty breathing. RNNs maintain a hidden state to remember past inputs, enabling pattern recognition vital for mortality prediction. The RNN architecture consists of an input layer receiving sequential symptoms (X1, X in Figure 5), hidden layers extracting features and passing outputs to subsequent layers, and an output layer generating final predictions (Y in Figure 5).

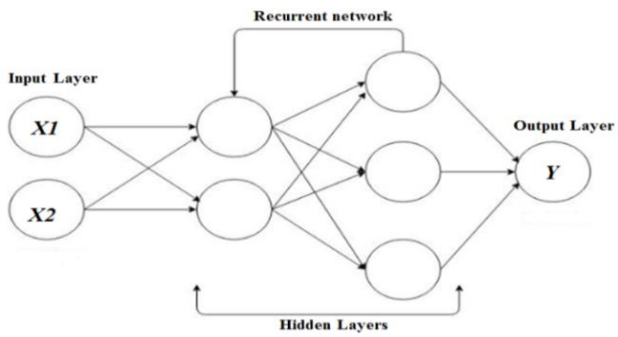

Figure 5. **Recurrent Neural Network (RNN)**

Traditional Recurrent Neural Networks (RNNs) face vanishing or exploding gradient challenges. Advanced architectures like Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) address these, improving long-term sequence learning [40].

3. Results and Discussion

The Results and Discussion section validates RNN models using confusion matrix analysis and performance metrics, highlighting superior accuracy over prior research, enabling reliable predictions, and improving healthcare resource allocation and decision-making.

3.1. Phase III: Model Training and Evaluation

3.1.1. Fitting Model

The Kaggle COVID-19 dataset provides detailed insights into patient symptoms, mortality rates, and confirmed cases. Key features include dry cough, high fever, sore throat, and difficulty breathing, essential for predicting death cases. The dataset is split into training and test sets to ensure real-world representativeness. An optimization method iteratively updates model parameters based on error signals, enhancing classification accuracy and reducing errors. The deep learning using RNN, exploits the strengths of these techniques. Performance is assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis, offering a comprehensive evaluation of the model's capability in predicting COVID-19 death cases.

3.1.2. Confusion Matrix Analysis

A confusion matrix evaluates the validity and error types of a classification model. It identifies false positives (FP) and false negatives (FN), providing insights into misclassifications for clinical interpretation and model refinement. The matrix aids in enhancing diagnostic accuracy. In machine learning, it helps assess model performance by visualizing the types of errors and computing accuracy. The matrix compares actual versus predicted labels, facilitating model optimization. Figure 6 illustrates the confusion matrix for our model, showing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

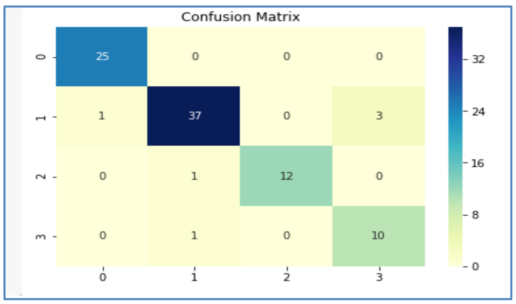

Figure 6. **Recurrent Neural Network (CNN)-Generated Confusing Matrix corresponding to our model**

The elements include True Positives (TP), correctly predicted positive instances; True Negatives (TN), correctly predicted negative instances; False Positives (FP), incorrectly predicted as positive; and False Negatives (FN), incorrectly predicted as negative.

In the multi-class confusion matrix (Figure 6), definitions for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are extended for each class. Macro-averaged values are calculated by averaging these metrics across all classes: Macro-Averaged TP = 21 (average correct positive predictions), Macro-Averaged TN = 63 (average correct negative predictions), Macro-Averaged FP = 1.5 (average incorrect positive predictions), and Macro-Averaged FN = 1.5 (average incorrect negative predictions). Model accuracy and performance metrics, including recall, accuracy, and precision, are computed from these elements for a thorough performance assessment.

3.1.3. Key Metrics for Performance Evaluation of COVID-19 Prediction Models

The proposed research evaluates the performance of Recurrent Neural Networks (RNNs, DL) using a Kaggle COVID-19 dataset containing patient symptoms (dry cough, high fever, sore throat, difficulty in breathing), mortality rates, and confirmed cases. These features are essential for identifying key symptoms and predicting COVID-19 case severity.

Performance is assessed using key metrics—accuracy, precision, recall, and F1 score.
Accuracy (AC) measures the proportion of correctly classified cases, calculated as

$$AC = \frac{TP + TN}{TP + FN + TN + FN} \tag{1}$$

Precision (PR) calculates the proportion of true positives among all positive predictions,

$$PR = \frac{TP}{TP + FP} \tag{2}$$

Recall (RE) measures the percentage of actual positives correctly identified,

$$RE = \frac{TP}{TP + FN} \tag{3}$$

The F1 score balances precision and recall,

$$F1 - Score = 2 \times \frac{PR \times RE}{PR + RE} \tag{4}$$

These metrics, derived from a confusion matrix, provide insights into the models' classification performance. Advanced data processing techniques such as augmentation and dimensionality reduction are used to improve prediction accuracy for COVID-19 death cases. This evaluation emphasizes the models' effectiveness in enhancing predictions and their potential to contribute to healthcare outcomes. By using RNN model, the research aims to improve the reliability of COVID-19 mortality predictions and provide valuable insights for managing the pandemic.

3.1.4. Evaluation of Metrics for COVID-19 Symptoms Classification

This section evaluates proposed models' performance in classifying COVID-19 symptoms (Table 3) using precision, recall, F1-Score, and accuracy for dry cough, fever, sore throat, and breathing difficulty.

Table 3. Performance Metrics of Proposed Models for COVID-19 Symptoms Classification

Features	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Dry Cough	96	100	98	98
High Fever	95	90	92	94
Sore Throat	100	92	96	99
Difficulty in Breathing	77	91	83	95

The model demonstrates high precision and recall for predicting dry cough symptoms, with 96% precision, 100% recall, 98% F1-Score, and 98% accuracy, ensuring minimal false negatives and accurate predictions. For high fever, precision is 95%, and recall is 90%, yielding a 92% F1-Score and 94% accuracy, though some cases may be missed. Sore throat symptoms are detected with 100% precision, 92% recall, 96% F1-Score, and 99% accuracy, showing strong performance in identifying these symptoms. Difficulty in breathing shows lower precision (77%) but high recall (91%), with an F1-Score of 83% and accuracy of 95%, indicating occasional misclassification. The RNN model shows solid performance in predicting COVID-19 symptoms, though performance for high fever and difficulty in breathing reveals room for improvement. The Kaggle dataset, used for evaluation, contains patient symptom data, including dry cough, high fever, sore throat, and difficulty in breathing, contributing to the objective of enhancing COVID-19 death prediction accuracy through RNN technique. Despite strong performance, variability in results suggests potential improvements in future model refinements.

3.2. **Phase IV: Performance Evaluation and Comparative Analysis**

The proposed models' performance is evaluated using precision, recall, F1-Score, and accuracy derived from the confusion matrix, essential for assessing COVID-19 mortality prediction based on patient symptoms.

3.2.1. Comparative Analysis of Metrics for Models in Predicting COVID-19 Mortality

Table 4 presents a performance comparison of RNN model for predicting COVID-19 mortality, focusing on precision, recall, F1 Score, and accuracy metrics.

Table 4. Overall Performance Comparison of Models for Predicting COVID-19 Mortality

Algorithm	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
RNN	92	93	92	94

The RNN model achieves a precision of 92%, recall of 93%, F1 score of 92%, and accuracy of 94%. Its superior performance indicates that RNN excels in identifying and capturing positive cases, making it the most reliable model for predicting COVID-19 mortality.

The results in Table 4 show that RNN model effectively predict COVID-19 mortality, with RNN outperforming in all evaluated metrics. RNN's superior precision and recall highlight its ability to identify true positives, crucial for accurate mortality prediction. This performance is driven by RNN's ability to capture temporal dependencies in sequential data. The outcomes emphasize the importance of selecting models based on specific goals, such as maximizing recall in mortality predictions. The RNN uses the strengths of each technique to improve prediction accuracy and healthcare outcomes.

3.2.2. RNN-Based Prediction and Analysis of COVID-19 Mortality: Predicted vs. Real Cases

This section discusses using Recurrent Neural Networks (RNNs) for predicting COVID-19 mortality. RNNs, a key deep learning technique, are vital in forecasting trends in complex datasets, with predicted results shown in Figure 7.

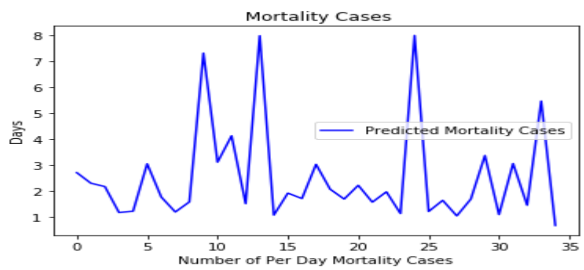

Figure 7. Predicted Mortality Cases

- ***Analyzing Predicted Mortality Cases:***

In Figure 7, the x-axis represents the prediction period in days, while the y-axis shows the predicted mortality cases, with the blue line illustrating the predicted values over time. The predicted mortality cases exhibit a cyclical pattern, indicating possible seasonal variations influenced by disease prevalence, healthcare responses, or public health interventions. Fluctuations in predicted mortality could stem from variations in disease spread, healthcare access, or external factors, suggesting that the model may not fully capture mortality trends and could be influenced by dynamic factors.

Outliers in the data, deviating from the general trend, might indicate anomalies such as sudden outbreaks or data entry errors, which could affect the model's accuracy. These fluctuations highlight the need for refining the model to enhance prediction stability and accuracy, potentially through architecture modifications or incorporating additional features.

Model performance is also sensitive to the quality of training data. Biases or inaccuracies in the data can negatively impact predictions, underscoring the importance of rigorous data cleaning and validation. External factors such as public health policies or socio-economic conditions may further influence mortality rates, warranting further investigation to improve model reliability and predictive power. Overall, while the RNN provides valuable insights, further refinement is necessary to achieve more accurate predictions.

- ***Analyzing Real and Predicted Mortality Cases***

Figure 8 compares real and predicted mortality cases over time, assessing model performance. The x-axis represents days, while the y-axis shows daily mortality cases. The red line indicates actual mortality, and the blue line shows predicted mortality, highlighting how well the model aligns with real-world data. This comparison is key for evaluating prediction accuracy.

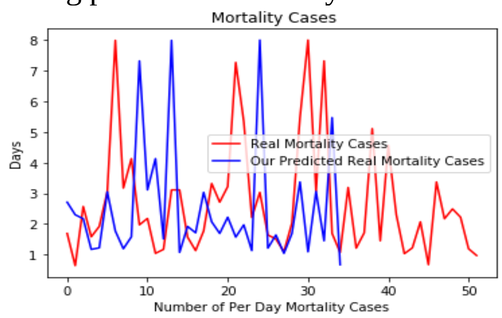

Figure 8. Real and Predicted Mortality Cases.

Both real and predicted mortality cases demonstrate similar trends, showing the model captures general mortality patterns over time. When actual mortality rates increase or decrease, the model's predictions follow the same trend, indicating the model effectively learns underlying patterns essential for predicting future cases.

Fluctuations in both real and predicted mortality rates reflect short-term variations due to factors such as infection rates, healthcare interventions, or changes in testing protocols. The model’s ability to track these fluctuations shows responsiveness, crucial for making realistic predictions in a dynamic pandemic. The moderate correlation between real and predicted cases suggests the model generally follows the trend but with some variability in matching exact values. While effective, the model’s predictions may not always perfectly align with actual cases, highlighting areas for improvement. Understanding discrepancies is essential for refining the model.

Discrepancies between predicted and real mortality cases may arise from factors like uncaptured variables, such as sudden healthcare changes or policy interventions, and data quality issues, including inaccuracies or biases. Identifying these causes is key to enhancing model performance, whether through better feature inclusion or data improvements. Improving the model could involve introducing relevant features such as demographic data (age, gender, comorbidities), healthcare access (ICU beds, ventilators), and environmental conditions (air quality, temperature). These features can provide more contexts, leading to nuanced predictions. Experimenting with architectures like Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU) and tuning hyperparameters (e.g., learning rate, batch size, and layers) may optimize performance, improving temporal pattern recognition.

Ensuring data quality through cleaning, dealing with missing values, and ensuring representativeness is critical for accurate predictions. By addressing these limitations, a more reliable model for predicting mortality cases can be developed, aiding public health planning and timely interventions.

3.2.3. Comparative Analysis of Proposed Work with Other Research Work

The comparative analysis of our proposed work with existing research highlights the performance improvements achieved using advanced machine learning techniques. The analysis is based on the accuracy scores reported in Table 5, which includes research by Husnul Khuluq, et al. [35] and Xiangao Jiang, et al. [5] and our proposed work using Recurrent Neural Networks (RNNs).

Table 5. Comparative Analysis of the Proposed Work in Relation to Other Research Work

Research Work	Data Set	Algorithm	Accuracy Score (%)
Husnul Khuluq, et al. [35]	Covid-19	KNN	75
		Naive Bayes	77
		Logistic Regression	81
		Random Forest	81
		SVM	81
		Decision Tree	79
Xiangao Jiang, et al. [5]	Covid-19	Logistic Regression	50
		KNN (k=5)	80
		Decision Tree (based on Gain Ratio)	70
		Decision Tree (based on Gini Index)	70
		Random Forests	70
		Support Vector Machine	80
Our Proposed Work	Covid-19	RNN	94

Khuluq, Husnul et al., examine the application of machine learning (ML) algorithms for predicting mortality in COVID-19 patients, using a dataset from Kaggle. Six ML algorithms were evaluated: K-Nearest Neighbours (KNN), Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, and Logistic Regression. Key predictors of mortality were identified, with a focus on the importance of symptoms in predicting COVID-19-related mortality. The study concludes that the proposed method, which can be easily updated with new data, supports frontline doctors in making timely clinical decisions. They assessed multiple algorithms on a COVID-19 dataset. Their results demonstrated that Logistic Regression, Random Forest, and SVM achieved accuracies of 81%. In contrast, KNN and Decision Tree models achieved lower accuracies of 75% and 79%, respectively. These findings underscore the effectiveness of ensemble methods like Random Forest and SVM, which outperformed simpler models such as KNN and Decision Trees.

Xiangao Jiang et al. evaluated several algorithms on a COVID-19 dataset, including traditional machine learning models such as Logistic Regression, K-Nearest Neighbours (KNN), Decision Trees, Random Forests, and Support Vector Machine (SVM). Logistic Regression achieved an accuracy of 50%. This lower accuracy reflects its limited capacity to handle complex, non-linear relationships in the data, highlighting its ineffectiveness compared to more advanced models. K-Nearest Neighbours (k=5) reached an accuracy of 80%, indicating its strength in managing non-linear data. However, its simplicity also raises the potential for overfitting, making it less robust than more sophisticated models. Decision Trees (both Gain Ratio and Gini Index) achieved an accuracy of 70%. While Decision Trees offer interpretable results, their performance suggests they may not effectively capture the complexities of the dataset. The consistency in their accuracy scores reflects stable but limited effectiveness. Random Forests also achieved an accuracy of 70%, showing an improvement over single Decision Trees. However, this method, though effective, still falls short compared to more advanced techniques. Support Vector Machine (SVM) achieved an accuracy of 80%, demonstrating its capability in handling high-dimensional data. Nevertheless, its performance is still below that of more modern algorithms.

- ***Contributions of the Proposed Work***

In contrast, our proposed work demonstrates significant improvements in predictive accuracy using advanced techniques.

Recurrent Neural Networks (RNNs) achieved the highest accuracy of 94%. This exceptional performance underscores the effectiveness of deep learning techniques, particularly RNNs, in capturing the nuances of the COVID-19 dataset. RNNs excel in handling sequential data and time-series forecasting, making them particularly well-suited for predicting trends over time. The RNN model's accuracy highlights its capability to capture temporal patterns and dependencies with high precision.

- ***Implications and Contributions***

Our research demonstrates significant advancements over traditional machine learning methods. The RNN model achieves notably higher accuracy than traditional algorithms. The RNN model, with its 94% accuracy, shows substantial improvement in prediction capabilities, providing more reliable outcomes compared to existing methods.

Utilizing RNNs represents a considerable advancement over simpler methods. RNNs' proficiency in sequential data analysis significantly enhances performance.

Our research employs a comprehensive COVID-19 dataset, integrating various symptoms and outcome variables. This extensive use of data improves model accuracy by considering a wide range of features and their interactions.

The proposed models are optimized to address the limitations of traditional approaches. The high accuracy achieved reflects the effectiveness of these advanced techniques in providing more reliable predictions for COVID-19 mortality cases.

Improved accuracy in predictions has practical implications for healthcare and disease management. More accurate predictions assist in better resource allocation, timely interventions, and informed decision-making, contributing to more effective management of COVID-19.

The comparative analysis of our proposed work relative to existing research reveals notable advancements in predictive accuracy for COVID-19 mortality cases. Khuluq, Husnul et al. utilized various classical machine learning algorithms to predict COVID-19 outcomes. Their study achieved decent accuracy with Logistic Regression, Random Forest, and SVM. However, these methods have limitations in capturing complex, non-linear patterns in the data. Xiangao Jiang et al. employed several algorithms, with KNN and SVM achieving the highest accuracy. However, Logistic Regression exhibited notably lower performance, indicating that simpler models may struggle with complex patterns in COVID-19 data. In contrast, our proposed approach uses advanced technique Recurrent Neural Networks (RNNs). The results demonstrate a significant improvement in accuracy, with RNN reaching 94%. This enhanced performance is accredited to RNNs' strength in capturing temporal dependencies and patterns over time. Overall, our proposed work makes substantial advancements in COVID-19 mortality prediction by employing advanced deep learning technique. These results in higher accuracy and a better handling of complex data patterns compared to existing research

4. Conclusion and Future Work

This study introduces an optimized approach for predicting COVID-19 mortality rates using advanced deep learning (DL) techniques. Using the Kaggle "COVID-19 Symptoms Dataset," which includes patient symptoms, mortality rates, and confirmed cases, we employ Recurrent Neural Networks (RNNs) to improve prediction accuracy. Unlike existing models focusing primarily on clinical severity, our approach addresses temporal dependencies crucial for accurate predictions. The methodology involves data acquisition, pre-processing, and supervised learning. Key results reveal that the RNN model achieved 94% accuracy, outperforming methods such as K-Nearest Neighbours (KNN), Naïve Bayes, and Logistic Regression, which had accuracies between 50% and 81%. The RNN effectively identifies critical symptoms, such as sore throat and dry cough, capturing the time-sensitive progression of COVID-19. The superior performance of RNNs highlights their capability to analyse sequential data, making them highly effective for mortality prediction. This study underscores the potential of DL in improving prediction reliability and suggests further exploration of temporal models for even greater precision.

Future work should include additional features like socioeconomic factors, comorbidities, and vaccination status to enhance the understanding of mortality risks. Advanced pre-processing and analysis of cyclical or seasonal data patterns could further refine predictions. Real-world testing and clinical validation remain essential to confirm the model's applicability in healthcare settings and support decision-making in COVID-19 treatment and resource allocation.

References

1. Özdemir, Öner. "Coronavirus disease 2019 (COVID-19): diagnosis and management." Journal of Clinical Practice and Research 42, no. 3 (2020): 242.
2. Perc, Matjaž, Nina Gorišek Miksić, Mitja Slavinec, and Andraž Stožer. "Forecasting covid-19." Frontiers in physics 8 (2020): 127.
3. Liu, Zhihua, Pierre Magal, Ousmane Seydi, and Glenn Webb. "Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data." arXiv preprint arXiv:2002.12298 (2020).
4. Jia, Lin, Kewen Li, Yu Jiang, and Xin Guo. "Prediction and analysis of coronavirus disease 2019." arXiv preprint arXiv:2003.05447 (2020).
5. Jiang, Xiangao, Megan Coffee, Anasse Bari, Junzhang Wang, Xinyue Jiang, Jianping Huang, Jichan Shi et al. "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity." Computers, Materials & Continua 63, no. 1 (2020): 537-551.
6. Alie, Melsew Setegn, Yilkal Negesse, Kassa Kindie, and Dereje Senay Merawi. "Machine learning algorithms for predicting COVID-19 mortality in Ethiopia." BMC Public Health 24, no. 1 (2024): 1728.
7. Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao et al. "A novel coronavirus from patients with pneumonia in China, 2019." New England journal of medicine 382, no. 8 (2020): 727-733.
8. García-Ordás, María Teresa, Natalia Arias, Carmen Benavides, Oscar García-Olalla, and José Alberto Benítez-Andrades. "Evaluation of country dietary habits using machine learning techniques in relation to deaths from COVID-19." In Healthcare, vol. 8, no. 4, p. 371. MDPI, 2020.
9. Cao, Jinming, Xia Jiang, and Bin Zhao. "Mathematical modeling and epidemic prediction of COVID-19 and its significance to epidemic prevention and control measures." Journal of Biomedical Research & Innovation 1, no. 1 (2020): 1-19.
10. Verzellesi, Laura, Andrea Botti, Marco Bertolini, Valeria Trojani, Gianluca Carlini, Andrea Nitrosi, Filippo Monelli et al. "Machine and deep learning algorithms for COVID-19 mortality prediction using clinical and radiomic features." Electronics 12, no. 18 (2023): 3878.
11. Iori, Mauro, Carlo Di Castelnuovo, Laura Verzellesi, Greta Meglioli, Davide Giosuè Lippolis, Andrea Nitrosi, Filippo Monelli et al. "Mortality prediction of COVID-19 patients using radiomic and neural network features extracted from a wide chest X-ray sample size: A robust approach for different medical imbalanced scenarios." Applied Sciences 12, no. 8 (2022): 3903.
12. Elshennawy, Nada M., Dina M. Ibrahim, Amany M. Sarhan, and Mohamed Arafa. "Deep-Risk: Deep Learning-Based Mortality Risk Predictive Models for COVID-19." Diagnostics 12, no. 8 (2022): 1847.
13. Tariq, Muhammad Usman, and Shuhaida Binti Ismail. "Deep learning in public health: Comparative predictive models for COVID-19 case forecasting." Plos one 19, no. 3 (2024): e0294289.
14. de Holanda, Wallace Duarte, Lenardo Chaves e Silva, and Álvaro Alvares de Carvalho César Sobrinho. "Machine learning models for predicting hospitalization and mortality risks of COVID-19 patients." Expert Systems with Applications 240 (2024): 122670.
15. Eskandarian, Rahimeh, Roohallah Alizadehsani, Mohaddeseh Behjati, Mehrdad Zahmatkesh, Zahra Alizadeh Sani, Azadeh Haddadi, Kourosh Kakhi et al. "Identification of clinical features associated with mortality in COVID-19 patients." In Operations Research Forum, vol. 4, no. 1, p. 16. Cham: Springer International Publishing, 2023.
16. Moulaei, Khadijeh, Mostafa Shanbehzadeh, Zahra Mohammadi-Taghiabad, and Hadi Kazemi-Arpanahi. "Comparing machine learning algorithms for predicting COVID-19 mortality." BMC medical informatics and decision making 22, no. 1 (2022): 2.

17. Liu, Xiaoyue, Jianping Huang, Changyu Li, Yingjie Zhao, Danfeng Wang, Zhongwei Huang, and Kehu Yang. "The role of seasonality in the spread of COVID-19 pandemic." Environmental research 195 (2021): 110874..
18. Pelinovsky, E., M. Kokoulina, A. Epifanova, A. Kurkin, O. Kurkina, M. Tang, E. Macau, and M. Kirillin. "Gompertz model in COVID-19 spreading simulation." Chaos, Solitons & Fractals 154 (2022): 111699.
19. Dinesh, Paidipati, A. S. Vickram, and P. Kalyanasundaram. "Medical image prediction for diagnosis of breast cancer disease comparing the machine learning algorithms: SVM, KNN, logistic regression, random forest and decision tree to measure accuracy." In AIP Conference Proceedings, vol. 2853, no. 1. AIP Publishing, 2024.
20. Hozhabri, Hossein, Francesca Piceci Sparascio, Hamidreza Sohrabi, Leila Mousavifar, René Roy, Daniela Scribano, Alessandro De Luca, Cecilia Ambrosi, and Meysam Sarshar. "The global emergency of novel coronavirus (SARS-CoV-2): An update of the current status and forecasting." International journal of environmental research and public health 17, no. 16 (2020): 5648.
21. Juneja, Mamta, Sumindar Kaur Saini, Harleen Kaur, and Prashant Jindal. "Statistical machine and deep learning methods for forecasting of Covid-19." Wireless Personal Communications (2024): 1-28.
22. Khero, Kainat, Muhammad Usman, and Alvis Fong. "Deep learning framework for early detection of COVID-19 using X-ray images." Multimedia Tools and Applications 83, no. 3 (2024): 6883-6908.
23. Salama, Gerges M., Asmaa Mohamed, and Mahmoud Khaled Abd-Ellah. "COVID-19 classification based on a deep learning and machine learning fusion technique using chest CT images." Neural Computing and Applications 36, no. 10 (2024): 5347-5365.
24. Talkhi, Nasrin, Nooshin Akbari Sharak, Razieh Yousefi, Maryam Salari, Seyed Masoud Sadati, and Mohammad Taghi Shakeri. "Predicting COVID-19 Mortality and Identifying Clinical Symptom Patterns in Hospitalized Patients: A Machine-learning Study." Iranian Journal of Health Sciences 12, no. 1 (2024): 39-48.
25. Hussain, Adedoyin Ahmed, Ouns Bouachir, Fadi Al-Turjman, and Moayad Aloqaily. "Notice of retraction: AI techniques for COVID-19." IEEE access 8 (2020): 128776-128795.
26. Gökler, Seda Hatice. "Prediction of Covid-19 confirmed cases and deaths using hybrid support vector machine-Taguchi method." Computers & Industrial Engineering 191 (2024): 110103.
27. Tellis, Gerard J., Ashish Sood, and Nitish Sood. "How long should social distancing last? Predicting time to moderation, control, and containment of COVID-19." Predicting Time to Moderation, Control, and Containment of COVID-19 (March 28, 2020). USC Marshall School of Business Research Paper (2020).
28. Waheed, Abdul, Muskan Goyal, Deepak Gupta, Ashish Khanna, Fadi Al-Turjman, and Plácido Rogerio Pinheiro. "Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection." Ieee Access 8 (2020): 91916-91923.
29. Tarek, Zahraa, Mahmoud Y. Shams, S. K. Towfek, Hend K. Alkahtani, Abdelhameed Ibrahim, Abdelaziz A. Abdelhamid, Marwa M. Eid et al. "An optimized model based on deep learning and gated recurrent unit for COVID-19 death prediction." Biomimetics 8, no. 7 (2023): 552.
30. Ohata, Elene Firmeza, Gabriel Maia Bezerra, Joao Victor Souza das Chagas, Aloísio Vieira Lira Neto, Adriano Bessa Albuquerque, Victor Hugo C. De Albuquerque, and Pedro Pedrosa Reboucas Filho. "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning." IEEE/CAA Journal of Automatica Sinica 8, no. 1 (2020): 239-248.
31. Rustam, Furqan, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byung-Won On, Waqar Aslam, and Gyu Sang Choi. "COVID-19 future forecasting using supervised machine learning models." IEEE access 8 (2020): 101489-101499.

32. Roy, Subhankar, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben et al. "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound." IEEE transactions on medical imaging 39, no. 8 (2020): 2676-2687.
33. Al-Rashedi, Afrah, and Mohammed Abdullah Al-Hagery. "Deep learning algorithms for forecasting COVID-19 cases in Saudi Arabia." Applied Sciences 13, no. 3 (2023): 1816.
34. Ali, Ahmed M., and Shimaa A. Esmail. "Prediction of COVID-19 Patients using Machine Learning Algorithms." SciNexuses 1 (2024): 58-69.
35. Khuluq, Husnul, Prasandhya Astagiri Yusuf, Dyah Aryani Perwitasari, and Abdul Fadhil. "Multivariate Analysis and Machine Learning: Mortality Predictions In COVID-19 Patients from Comorbidity, Demographic and Laboratory Findings." Journal for ReAttach Therapy and Developmental Diversities 6, no. 10s (2023): 1130-1141.
36. Kaggle. “COVID-19 Symptoms Dataset,” March 27, 2020. https://www.kaggle.com/datasets/prakharsrivastava01/covid19-symptoms-dataset?resource=download.
37. Cohen, Joseph Paul, Paul Morrison, Lan Dao, Karsten Roth, Tim Q. Duong, and Marzyeh Ghassemi. "Covid-19 image data collection: Prospective predictions are the future." arXiv preprint arXiv:2006.11988 (2020).
38. Wynants, Laure, Ben Van Calster, Marc MJ Bonten, Gary S. Collins, Thomas PA Debray, Maarten De Vos, Maria C. Haller et al. "Systematic review and critical appraisal of prediction models for diagnosis and prognosis of COVID-19 infection." MedRxiv (2020): 2020-03.
39. Soni, Kartik M., Amisha Gupta, and Tarun Jain. "Supervised machine learning approaches for breast cancer classification and a high performance recurrent neural network." In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1-7. IEEE, 2021.
40. Analytics Vidhya. (2022, March 10). A brief overview of Recurrent Neural Networks (RNN). Analytics Vidhya. Available Online: https://www.analyticsvidhya.com/blog/2022/03/a-brief-overview-of-recurrent-neural-networks-rnn/. . [Accessed on: 26-08-2024].