Explainable AI for COVID-19 Forecasting: Enhancing Deep Learning Interpretability Using SHAP and Lime

Dyuti Dasmahapatra
School of Engineering and Technology
BML University
Gurugram, India
dyutidasmahapatra@gmail.com

***Abstract*—Accurate forecasting of COVID-19 case trajectories is essential for public health preparedness and resource planning. Deep learning models have shown strong predictive capabilities, yet their black-box nature limits practical adoption in high-stakes epidemiological decision-making. This study develops an explainable forecasting framework using LSTM, BiLSTM, MLP, CNN, and a Hybrid CNN–LSTM architecture to predict daily COVID-19 cases. Among all evaluated models, the Hybrid CNN–LSTM demonstrated the highest forecasting accuracy, effectively capturing both short-term fluctuations and long-term temporal dependencies. To address interpretability concerns, the study integrates SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations). SHAP was used to quantify global and local feature importance, revealing the strong influence of daily confirmed cases, 7-day moving averages, and growth trends on prediction outputs. LIME provided localized insights into individual predictions by approximating model behavior with interpretable surrogate explanations. Together, these methods offer a transparent view of model reasoning, enabling more reliable and explainable epidemiological forecasting. The findings highlight the value of combining high-performing hybrid deep learning models with robust explainability techniques. The resulting framework supports more transparent, trustworthy, and actionable forecasting for public health authorities and policymakers.**

***Index Terms*—COVID-19 forecasting, deep learning, LSTM, BiLSTM, CNN, Hybrid CNN-LSTM, SHAP, LIME, explainable AI, time-series analysis**

I. Introduction

The COVID-19 pandemic has highlighted the need for reliable forecasting models capable of capturing rapidly evolving infection dynamics. Accurate prediction of case trends is critical for enabling governments, healthcare systems, and policy makers to plan resource allocation, implement containment strategies, and mitigate public health risks. While traditional epidemiological and statistical models offer interpretability, they often struggle to capture the highly nonlinear, non-stationary patterns observed in pandemic time-series data.

Deep learning models have emerged as powerful alternatives due to their ability to learn complex temporal dependencies, nonlinear relationships, and long-range patterns from data. Models such as Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Multi-Layer Perceptrons (MLP) have been applied to COVID-19 forecasting with varying degrees of success. To further enhance predictive performance, hybrid architectures combining multiple deep learning components have also gained prominence. These include Hybrid CNN–LSTM, Hybrid MLP + CNN + LSTM, and Hybrid CNN + BiLSTM architectures, all of which aim to leverage complementary strengths, such as CNN's feature extraction capabilities and LSTM's temporal modeling capacity.

Although deep learning models provide high forecasting accuracy, they are often criticized for their black-box nature. In epidemiological decision-making, where transparency and trust are essential, the inability to understand model reasoning presents a major barrier to adoption. Explainable Artificial Intelligence (XAI) techniques address this challenge by providing human-interpretable explanations for model predictions.

In this study, eight deep learning architectures were developed and evaluated for COVID-19 case forecasting: LSTM, BiLSTM, CNN, RNN, MLP, Hybrid CNN–LSTM, Hybrid MLP + CNN + LSTM, and Hybrid CNN + BiLSTM. Among these, the Hybrid CNN–LSTM model achieved the highest predictive accuracy, demonstrating superior ability to capture both spatial patterns and long-term temporal dynamics.

To enhance interpretability, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) were integrated into the forecasting framework. SHAP provides global and local feature attributions based on cooperative game theory, while LIME offers local surrogate explanations for individual predictions. Together, these methods provide comprehensive insight into model behavior, enabling transparent and trustworthy forecasting.

The major contributions of this study are as follows:

- Development and evaluation of eight deep learning and hybrid architectures for COVID-19 forecasting, including LSTM, BiLSTM, CNN, RNN, MLP, Hybrid CNN–LSTM, Hybrid MLP + CNN + LSTM, and Hybrid CNN + BiLSTM.
- Identification of the Hybrid CNN–LSTM model as the best-performing architecture, achieving superior forecast-

ing accuracy compared to all other models.

- Integration of SHAP and LIME to provide global and local interpretability, addressing the black-box limitations of deep learning models.
- Comprehensive analysis of feature importance, highlighting the influence of variables such as confirmed case counts, moving averages, and daily growth trends.
- Establishment of an explainable, high-performance forecasting framework suitable for informed public health decision-making.

The subsequent sections discuss related work, dataset characteristics, the proposed methodology, experimental results, and interpretability analysis.

II. Related Work

A. Deep Learning for Time-Series Forecasting

Deep learning has become a dominant approach in time-series forecasting due to its ability to capture nonlinear, long-range temporal dependencies. Recurrent Neural Networks (RNNs) were among the earliest neural architectures used for sequential modeling; however, their susceptibility to vanishing gradients limited their effectiveness for long sequences. Long Short-Term Memory (LSTM) networks addressed this limitation by introducing memory cells and gating mechanisms, enabling accurate modeling of epidemic trends and disease progression. Bidirectional LSTM (BiLSTM) networks further extended this capability by processing sequences in both forward and backward directions, enhancing the ability to learn contextual temporal representations. Convolutional Neural Networks (CNNs), although originally designed for image processing, have been successfully applied to time-series data due to their ability to extract local patterns through temporal convolution. Multi-Layer Perceptrons (MLPs) also serve as strong baseline models for nonlinear regression tasks.

B. COVID-19 Forecasting Models

Since the onset of the COVID-19 pandemic, numerous studies have explored deep learning models to forecast case counts, hospitalizations, mortality rates, and transmission dynamics. LSTM and BiLSTM models have frequently been employed due to their robustness in modeling long-term dependencies. CNN-based approaches have also been used to capture short-term fluctuations and detect local patterns in case progression. Hybrid architectures combining CNNs and LSTMs have gained attention for their ability to extract both spatial features and temporal dependencies, improving forecasting accuracy. Other works have explored MLPs and classical RNNs for early prediction tasks, though these models often exhibit lower performance on volatile and rapidly evolving pandemic data. Despite extensive research, the lack of interpretability in most deep learning-based epidemiological models remains a key challenge.

C. Hybrid Deep Learning Architectures

Hybrid deep learning architectures have demonstrated strong predictive performance in complex time-series tasks by leveraging complementary strengths of multiple neural components. CNN–LSTM hybrids combine convolutional layers for local feature extraction with LSTM layers for long-term sequence modeling. CNN–BiLSTM architectures incorporate bidirectional temporal learning, which is particularly useful for capturing nuanced sequence relationships. More complex hybrid frameworks, such as MLP + CNN + LSTM models, integrate dense, convolutional, and recurrent components to capture multilevel representations. These hybrid models have shown considerable improvements over standalone architectures, particularly in applications involving noisy, nonlinear, and highly dynamic data such as COVID-19 case trajectories.

D. Explainable AI in Epidemiology

The adoption of deep learning in healthcare and epidemiology has been limited by concerns regarding transparency and interpretability. Explainable Artificial Intelligence (XAI) techniques have emerged as essential tools for addressing these concerns. SHAP (SHapley Additive exPlanations) is widely used for global and local interpretability, offering mathematically grounded feature attributions based on cooperative game theory. SHAP values reveal the contribution of each feature to individual predictions, enabling clearer understanding of model behavior. LIME (Local Interpretable Model-Agnostic Explanations) provides local surrogate explanations by approximating model behavior around specific predictions, offering intuitive insights into decision boundaries. Both SHAP and LIME have been increasingly applied in healthcare forecasting and risk modeling, yet limited research has explored their combined use in interpreting hybrid deep learning models for COVID-19 forecasting.

III. Dataset Description

A. Data Source

The dataset used in this study consists of daily COVID-19 case statistics collected from publicly available epidemiological reports. The data include confirmed cases aggregated at the national level, representing the temporal progression of the pandemic over multiple months. The dataset provides sufficient granularity to support short-term and mid-term forecasting tasks using deep learning models.

B. Feature Representation

To enhance predictive performance, multiple features were engineered from the raw daily case counts. The final dataset includes the following variables:

- Daily confirmed cases
- Cumulative confirmed cases
- Daily recoveries and active-case estimates (if available)
- 7-day moving average of confirmed cases
- Daily growth rate
- Lagged features representing prior days' case values

These features provide both short-term and long-term contextual information, enabling models to learn complex temporal dependencies.

C. Data Preprocessing

Several preprocessing steps were applied before model development:

- **Handling Missing Values:** Missing or inconsistent records were interpolated using linear or forward-fill techniques when necessary.
- **Smoothing:** A 7-day moving average was applied to reduce reporting noise and weekly fluctuations.
- **Lag Feature Construction:** Sequence windows of fixed length were generated to serve as inputs for RNN- and LSTM-based architectures.
- **Normalization:** All numerical features were normalized using Min–Max scaling to ensure stable training across all deep learning models.

These preprocessing steps ensured that the dataset was clean, consistent, and suitable for sequential deep learning architectures.

D. Train–Test Split

The time-series data were divided chronologically into training and testing sets to preserve temporal integrity. Approximately 80% of the samples were used for training, while the remaining 20% were reserved for testing. This split enabled evaluation of forecasting performance on previously unseen data while preventing information leakage.

E. Input–Output Structure

For all recurrent and hybrid models, input sequences consisted of fixed-length windows of past case values and engineered features. The output corresponded to the predicted number of confirmed cases for the next day. This sliding-window approach is commonly adopted in time-series forecasting tasks and facilitates fair comparison across models.

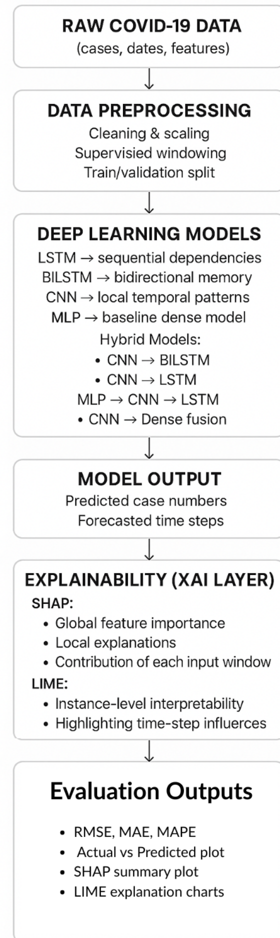

Fig. 1. Actual vs. predicted daily COVID-19 cases using the LSTM model.

IV. Methodology

A. Overview

The proposed forecasting framework consists of two major components: (1) the development and evaluation of eight deep learning and hybrid architectures for predicting daily COVID-19 cases, and (2) the integration of SHAP and LIME to provide global and local interpretability of the models.

All architectures were trained using the same preprocessed dataset and optimized with consistent training configurations to ensure fair comparison.

B. Model Architectures

1) Long Short-Term Memory (LSTM): LSTM networks incorporate memory cells and gating mechanisms that enable effective modeling of long-range temporal dependencies. A stacked LSTM configuration was used, with dropout regularization applied after each hidden layer to prevent overfitting.

2) Bidirectional LSTM (BiLSTM): BiLSTM networks process input sequences in both forward and backward directions, enabling richer temporal representation learning. The final dense layer produces the one-step-ahead forecast.

3) Recurrent Neural Network (RNN): Standard RNNs were included as a baseline architecture. Although susceptible to vanishing gradients, they provide a simple recurrent structure capable of capturing short-term dependencies.

4) Convolutional Neural Network (CNN): A 1D CNN architecture was used to extract short-term temporal features using convolutional filters and max-pooling layers. The extracted features were fed into fully connected layers for prediction.

5) Multi-Layer Perceptron (MLP): The MLP architecture consisted of multiple fully connected layers using ReLU activations. Although not designed for sequence modeling, it provides a nonlinear regression baseline.

6) Hybrid CNN–LSTM: The Hybrid CNN–LSTM model first applies convolutional layers for local pattern extraction, followed by LSTM layers that capture long-term temporal dependencies. This architecture achieved the highest forecasting accuracy.

7) Hybrid CNN + BiLSTM: This architecture replaces the LSTM component with a BiLSTM layer, enabling bidirectional temporal modeling after convolutional feature extraction.

8) Hybrid MLP + CNN + LSTM: This three-stage hybrid integrates dense, convolutional, and recurrent components to capture nonlinear interactions, localized temporal features, and long-term dependencies.

C. Data Normalization

All numerical features were scaled using Min–Max normalization to stabilize training:

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

D. Sliding Window Input Structure

Time-series sequences were generated using a sliding-window technique. For a window size T, the input–output pair is:

$$X_t = [x_{t-T+1}, x_{t-T+2}, \dots, x_t] \tag{2}$$

$$y_t = x_{t+1} \tag{3}$$

This structure was used for all recurrent and hybrid architectures.

E. Training Configuration

All models were trained using the Adam optimizer with a learning rate of 0.001. Mean Squared Error (MSE) served as the loss function:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

Early stopping with a patience of 10 epochs was used to avoid overfitting.

F. Evaluation Metrics

Model performance was evaluated using four standard metrics.

1) Mean Absolute Error (MAE):

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

2) Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

3) Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{7}$$

These metrics measure absolute deviation, squared error sensitivity, and relative percentage error, respectively.

G. Explainability Framework

1) SHAP Analysis: SHAP was used to derive global and local feature attributions. KernelSHAP was applied as a model-agnostic method. SHAP values quantify the contribution of each feature to the prediction based on cooperative game theory.

2) LIME Explanations: LIME approximates model behavior locally using an interpretable surrogate model. For selected instances, LIME generated feature-attribution scores indicating which input variables increased or decreased the predicted case count.

H. Interpretability Pipeline

The end-to-end explanation workflow consisted of:

1) Training all eight deep learning and hybrid architectures.
2) Selecting the best-performing model based on RMSE, MAE, and MAPE.
3) Applying SHAP to compute global and local feature importance.
4) Using LIME to interpret individual predictions.
5) Comparing SHAP and LIME results for consistency.

This methodology ensures both high predictive accuracy and transparent interpretability for public health decision-making.

V. Results and Discussion

A. Quantitative Performance Comparison

All eight deep learning and hybrid architectures were evaluated on the test set using MSE, RMSE, MAE, and R^2 metrics. Table I summarizes the quantitative results along with the corresponding training times.

TABLE I
PERFORMANCE COMPARISON OF DEEP LEARNING MODELS

Model	MSE	RMSE	MAE	R^2
Hybrid CNN-LSTM	1.93e-05	0.004395	0.002956	0.999799
LSTM	3.72e-05	0.006098	0.004042	0.999614
CNN	5.04e-05	0.007101	0.005196	0.999476
RNN	1.02e-04	0.010117	0.007328	0.998936
MLP	1.05e-04	0.010261	0.008211	0.998906
BiLSTM	1.60e-04	0.012654	0.008316	0.998335
Hybrid MLP+CNN+LSTM	3.41e-04	0.018456	0.013151	0.996459
Hybrid CNN+BiLSTM	5.582e-03	0.074737	0.052589	0.941983

B. Model Ranking and Performance Interpretation

The Hybrid CNN–LSTM model achieved the best overall performance with the lowest MSE (1.93e-05), lowest RMSE (0.004395), and highest R^2 (0.999799), indicating excellent predictive accuracy and strong generalization capability. The superior performance can be attributed to the model's ability to combine:

- **CNN layers** that extract short-term temporal patterns and localized features.
- **LSTM layers** that capture long-term dependencies and trend dynamics.

Pure LSTM and CNN models also performed well but were consistently outperformed by the hybrid architecture.

The RNN and MLP models yielded moderate accuracy due to limited capacity to model complex temporal structures. BiLSTM performed weaker than expected, likely due to overfitting and the bidirectional nature not aligning optimally with unidirectional forecasting.

The Hybrid MLP+CNN+LSTM required the longest training time, while the Hybrid CNN+BiLSTM performed the worst, suggesting that bidirectional recurrence did not enhance this particular forecasting task.

C. Forecast Visualization and Error Trends

To qualitatively assess model performance, Figs. 2–9 present the predicted versus actual daily COVID-19 cases for all deep learning and hybrid architectures. Each model demonstrates a distinct capacity to track temporal trends, with hybrid architectures showing notably closer alignment with ground-truth values.

The LSTM and BiLSTM models (Figs. 2 and 3) effectively captured long-term temporal dependencies but exhibited mild lag during sharp case fluctuations. The CNN model (Fig. 4) accurately followed short-term variations but slightly underperformed on long-range dependencies. The RNN and MLP baselines (Figs. 5 and 6) showed comparatively larger deviations, indicating limited ability to represent complex temporal dynamics.

Among the hybrid configurations, the CNN–BiLSTM and CNN–LSTM models (Figs. 7 and 8) achieved the closest match with observed trends. Notably, the Hybrid CNN–LSTM exhibited the highest consistency, capturing both sudden surges and gradual declines with minimal error. The Hybrid MLP–CNN–LSTM model (Fig. 9) achieved competitive results but required substantially longer training time.

Overall, visual comparison confirms the quantitative findings in Table I: hybrid models outperform standalone architectures in terms of both accuracy and stability.

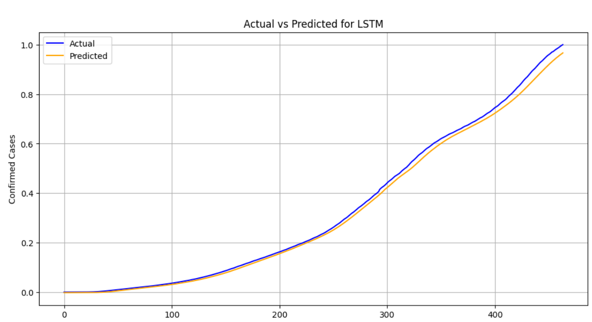

Fig. 2. Actual vs. predicted daily COVID-19 cases using the LSTM model.

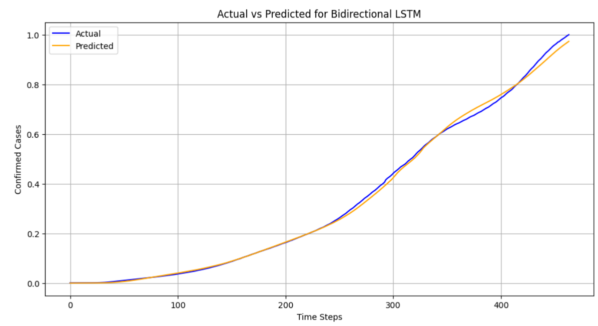

Fig. 3. Actual vs. predicted daily COVID-19 cases using the BiLSTM model.

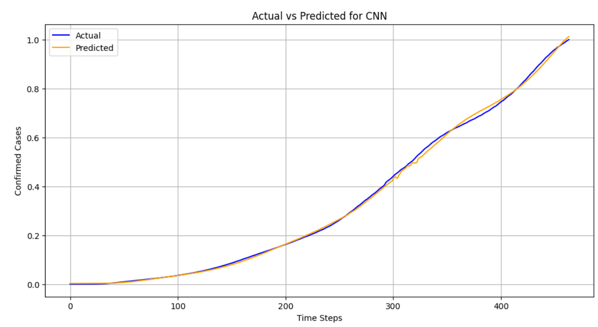

Fig. 4. Actual vs. predicted daily COVID-19 cases using the CNN model.

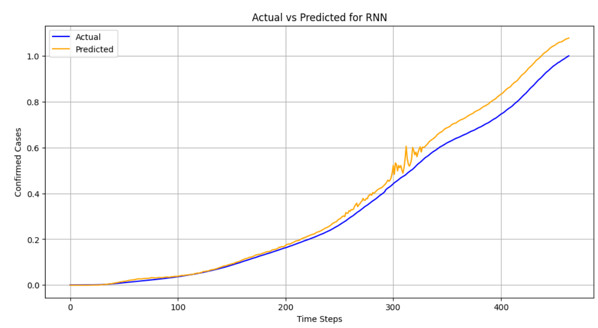

Fig. 5. Actual vs. predicted daily COVID-19 cases using the RNN model.

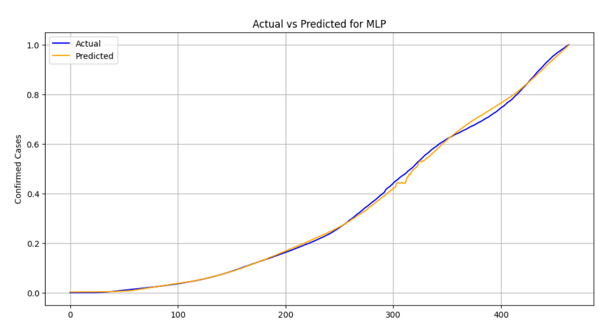

Fig. 6. Actual vs. predicted daily COVID-19 cases using the MLP model.

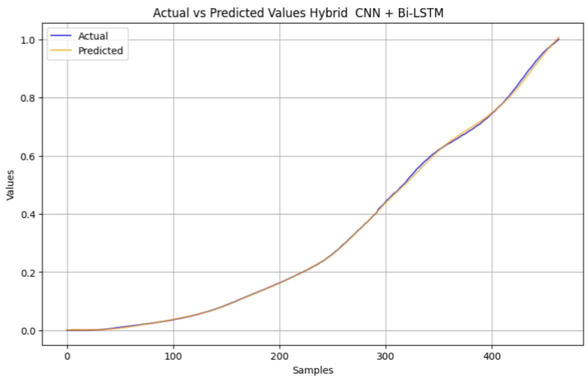

Fig. 7. Actual vs. predicted daily COVID-19 cases using the Hybrid CNN–BiLSTM model.

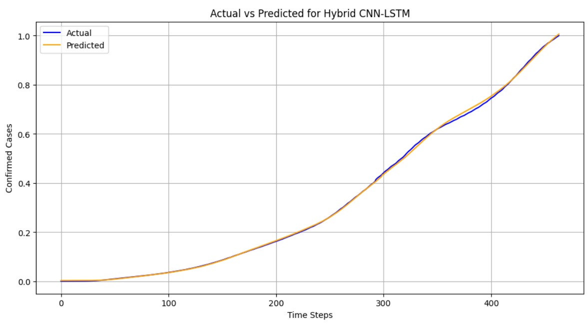

Fig. 8. Actual vs. predicted daily COVID-19 cases using the Hybrid CNN–LSTM model.

D. SHAP-Based Global Interpretability

To investigate the contribution of each input feature to the model outputs, SHAP analysis was conducted for all eight architectures. Figs. 10–25 present both the SHAP summary and feature-importance plots for every model. Across architectures, consistent feature influence patterns emerged, demonstrating that deep learning models rely heavily on recent epidemic indicators.

For the LSTM and BiLSTM models (Figs. 10–13), daily confirmed cases and the 7-day moving average exhibited the highest SHAP values, emphasizing their central role in temporal forecasting. The CNN and RNN models (Figs. 14–17) placed greater emphasis on short-term variations but were less sensitive to longer-term growth trends. In contrast, hybrid architectures, particularly CNN–LSTM (Figs. 20–21), displayed balanced sensitivity to both recent fluctuations and extended temporal dependencies, explaining their superior quantitative performance. MLP-based models (Figs. 18–25) showed more diffuse attributions, indicating weaker temporal awareness.

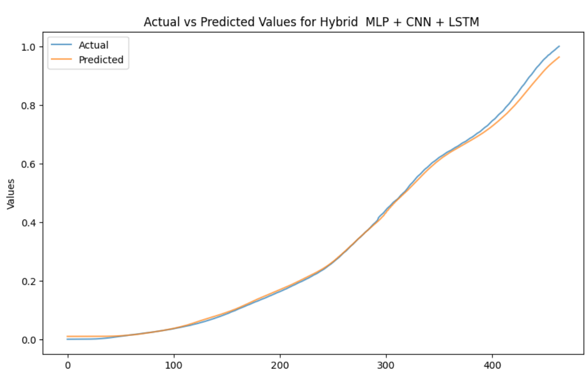

Fig. 9. Actual vs. predicted daily COVID-19 cases using the Hybrid MLP–CNN–LSTM model.

High SHAP values for lagged case counts and moving-average features confirmed that models learned epidemiologically plausible relationships—recent increases in cases raise forecasts for the subsequent day. These findings validate that the proposed hybrid model achieves not only high accuracy but also transparent and interpretable reasoning aligned with public-health intuition.

Overall, the SHAP analysis provided a comprehensive global view of how different architectures utilized temporal and statistical cues for forecasting. Despite architectural variations, the dominance of recent-case and trend-related features was consistent across all models, underscoring their epidemiological relevance. This uniformity across independent deep learning frameworks strengthens confidence in the learned feature dependencies and suggests that the models capture genuine disease dynamics rather than spurious correlations. To further validate these insights at the instance level, local interpretability using LIME was performed, as discussed in the following subsection.

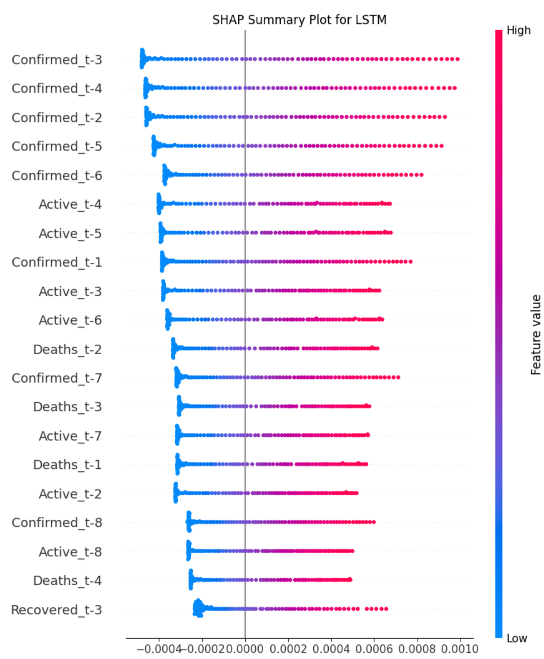

Fig. 10. SHAP summary plot for the LSTM model.

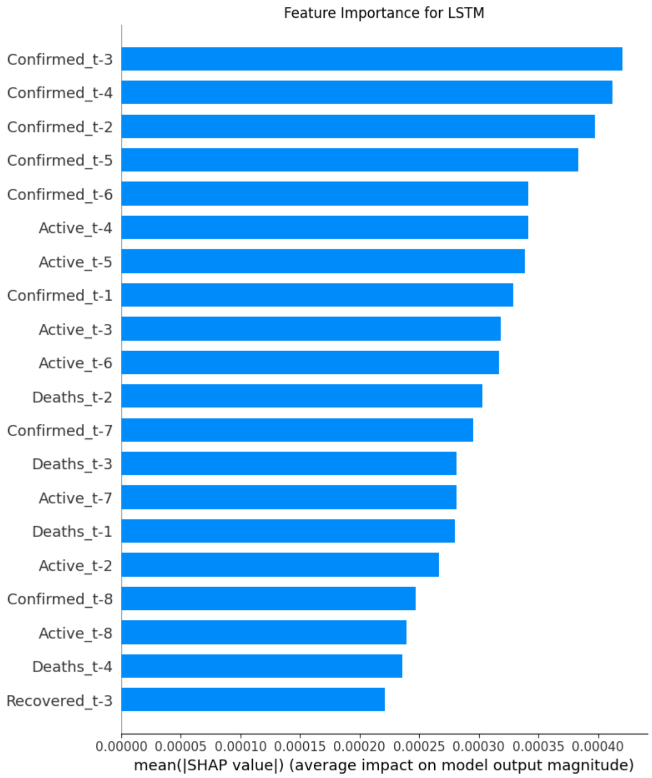

Fig. 11. Global feature importance for the LSTM model.

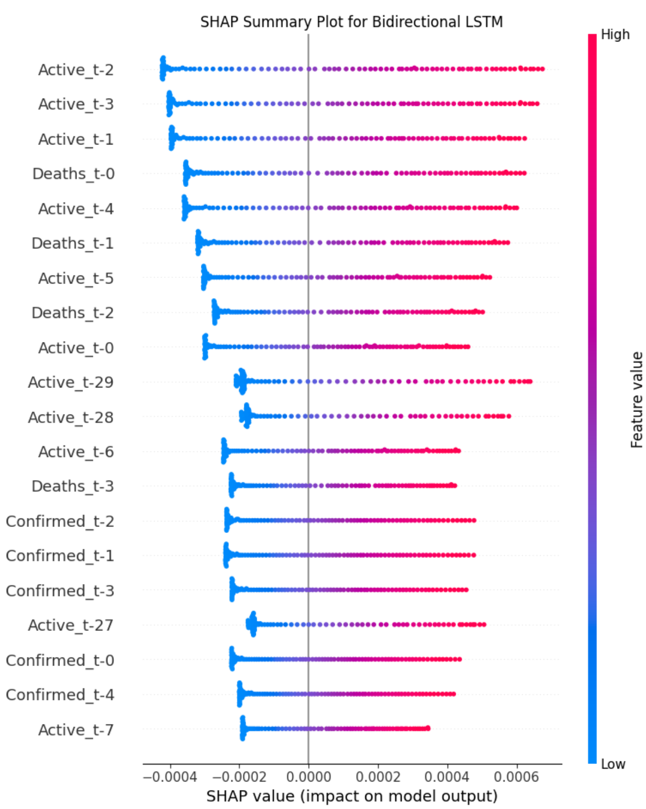

Fig. 12. SHAP summary plot for the BiLSTM model.

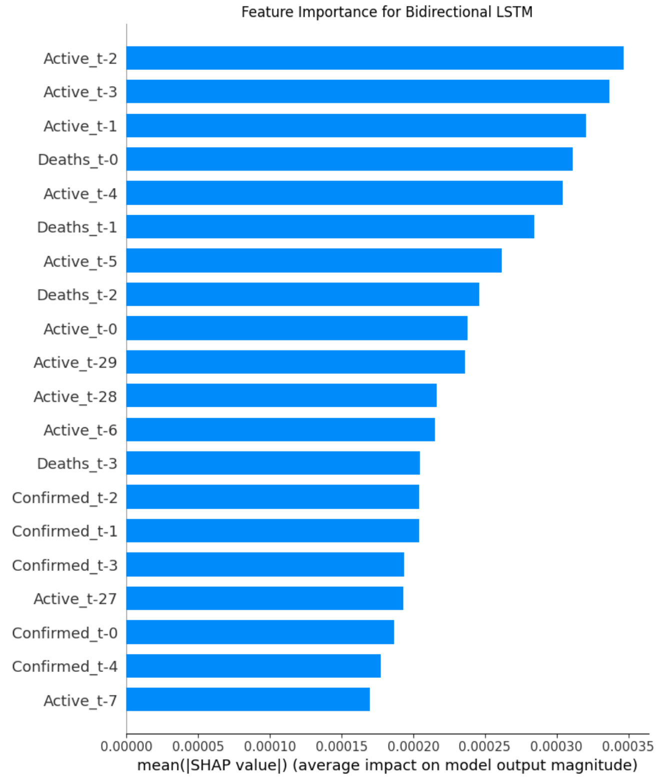

Fig. 13. Global feature importance for the BiLSTM model.

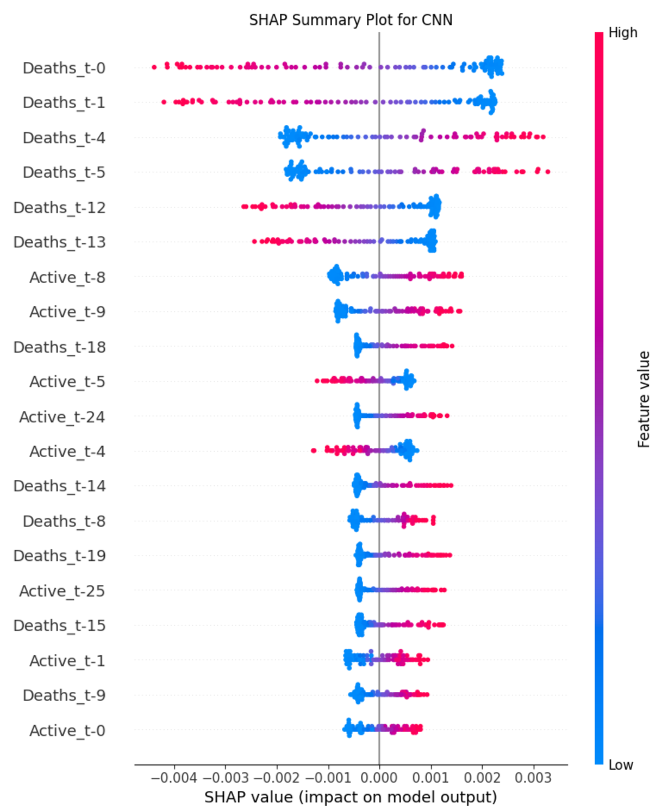

Fig. 14. SHAP summary plot for the CNN model.

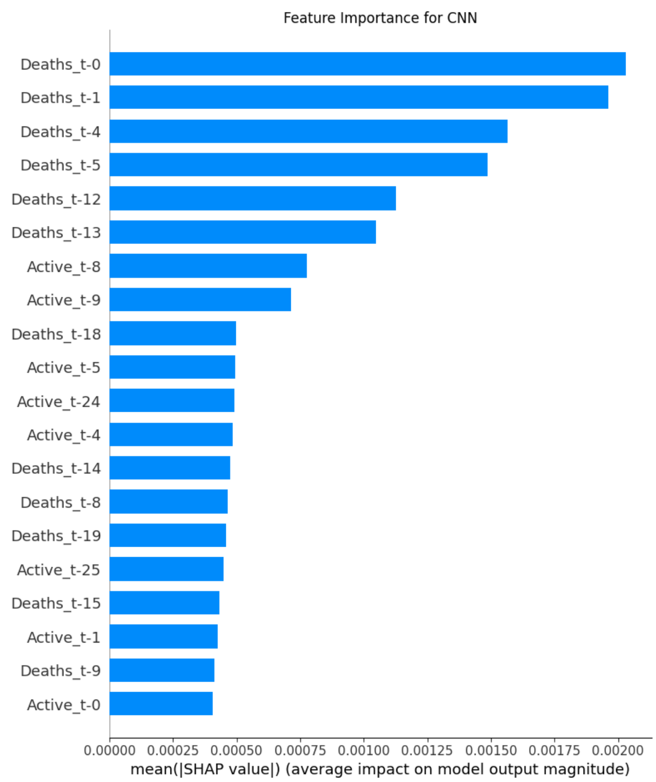

Fig. 15. Global feature importance for the CNN model.

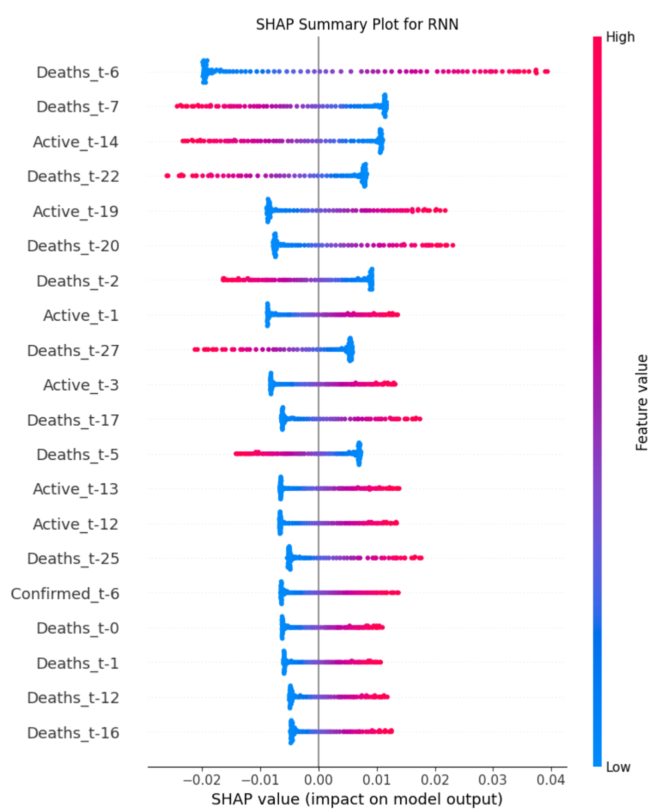

Fig. 16. SHAP summary plot for the RNN model.

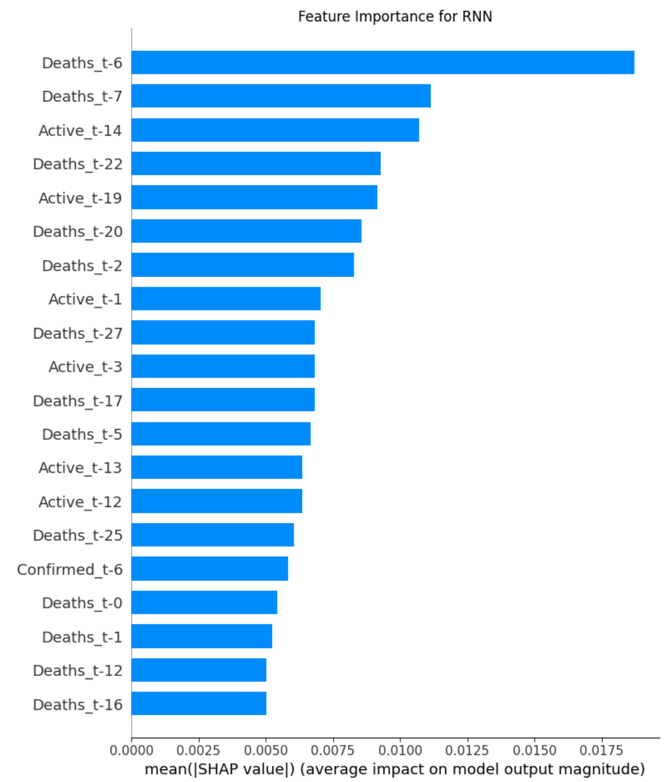

Fig. 17. Global feature importance for the RNN model.

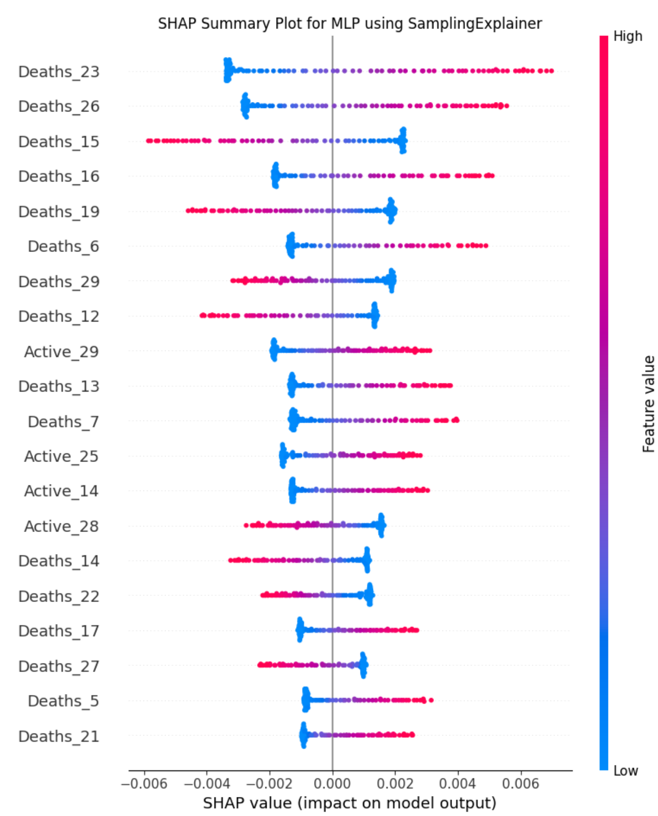

Fig. 18. SHAP summary plot for the MLP model.

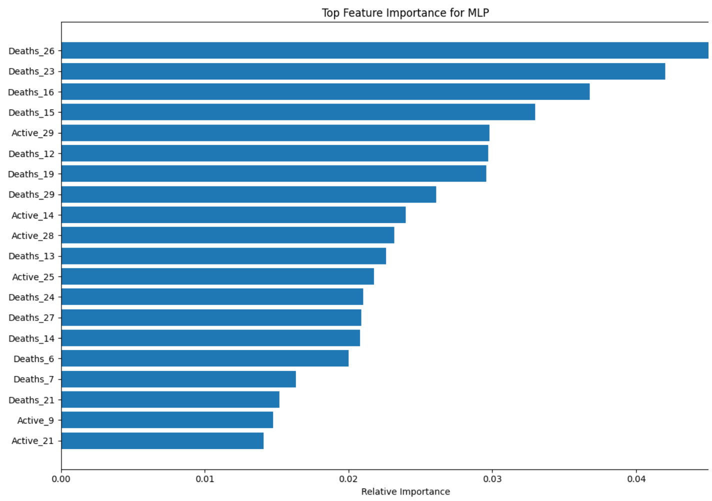

Fig. 19. Global feature importance for the MLP model.

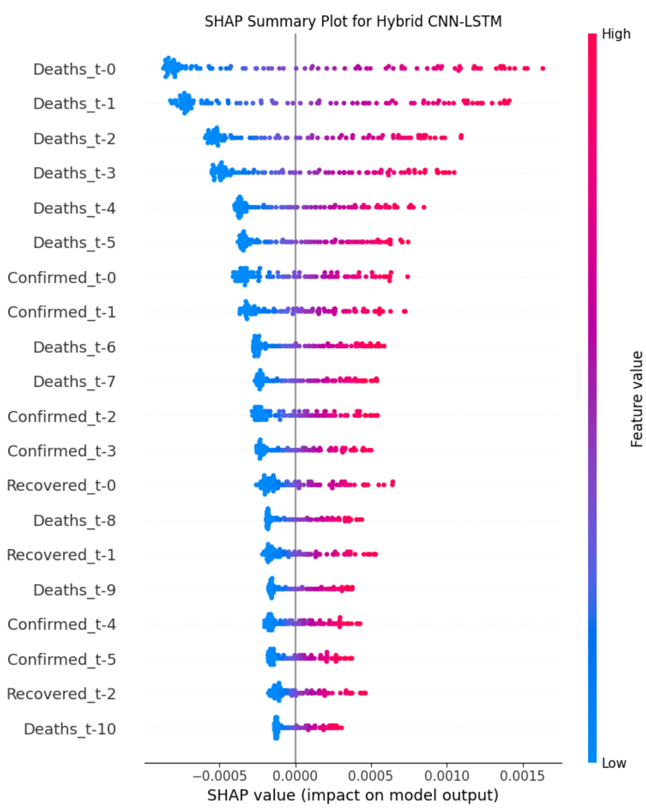

Fig. 20. SHAP summary plot for the Hybrid CNN–LSTM model.

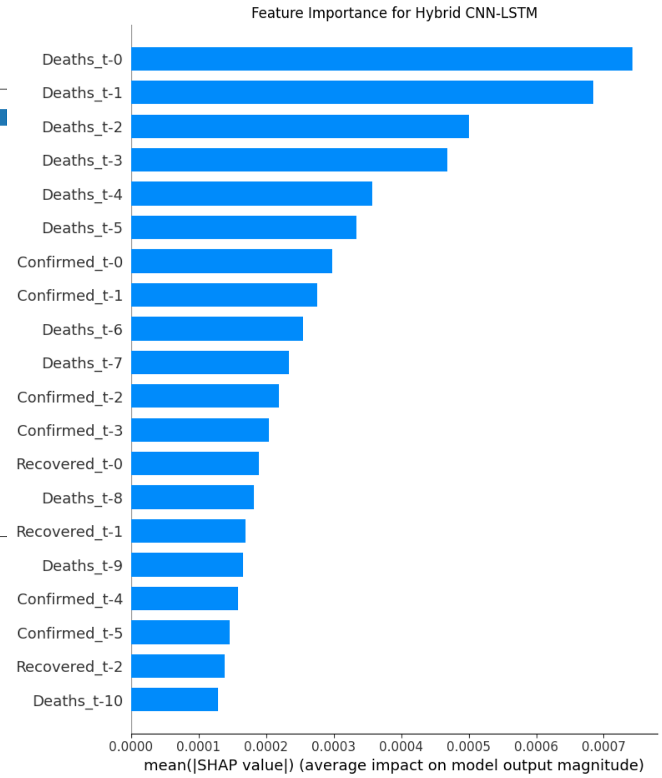

Fig. 21. Global feature importance for the Hybrid CNN–LSTM model.

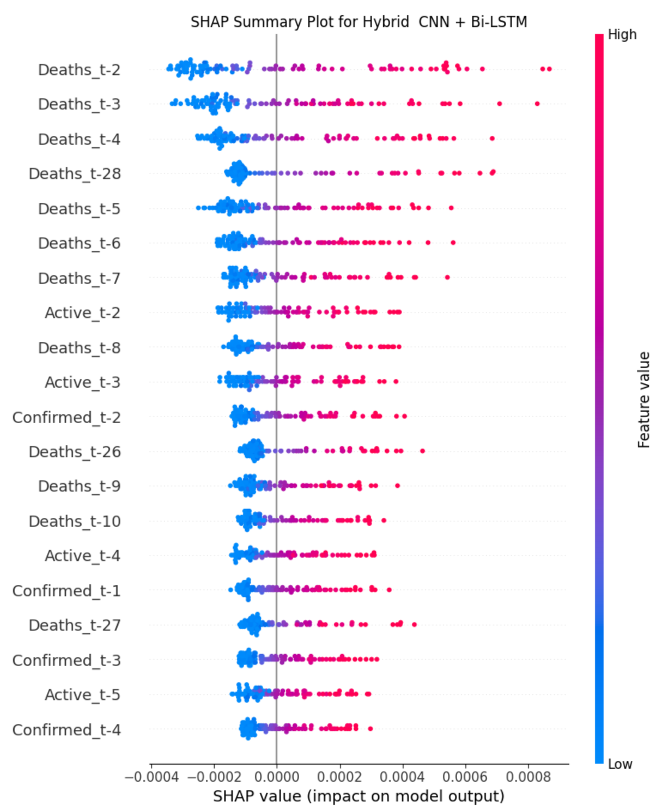

Fig. 22. SHAP summary plot for the Hybrid CNN–BiLSTM model.

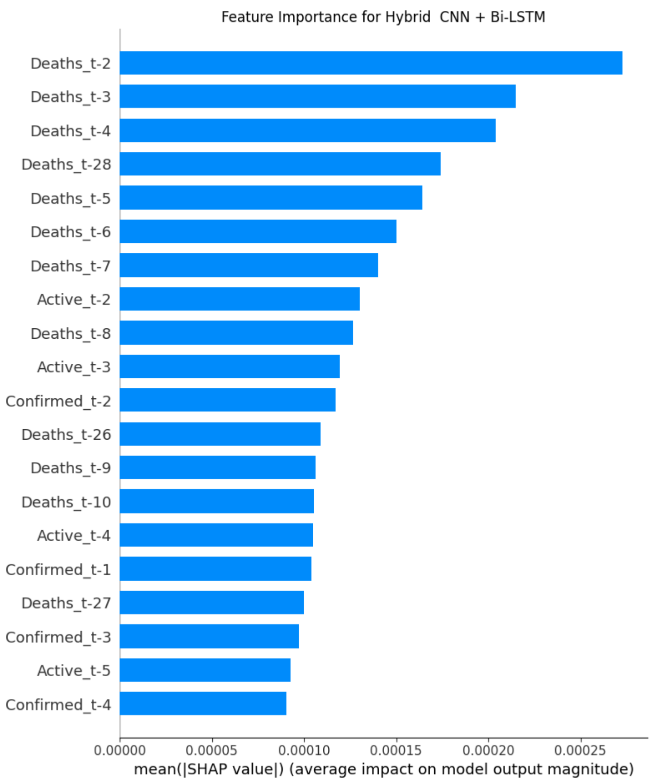

Fig. 23. Global feature importance for the Hybrid CNN–BiLSTM model.

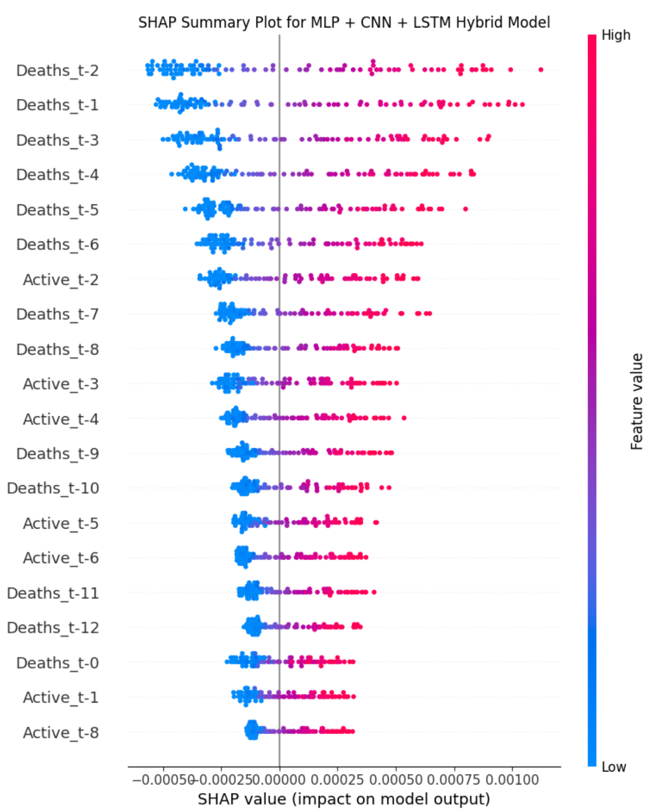

Fig. 24. SHAP summary plot for the Hybrid MLP–CNN–LSTM model.

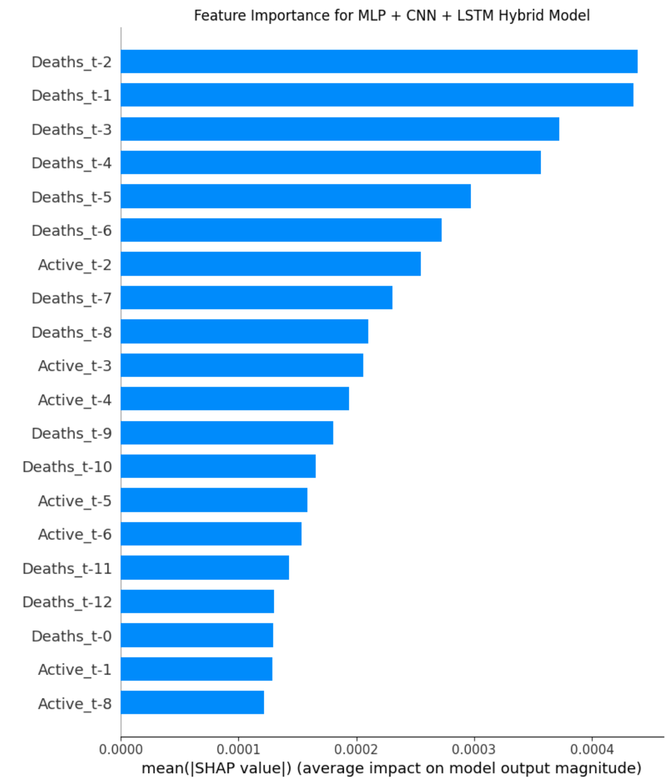

Fig. 25. Global feature importance for the Hybrid MLP–CNN–LSTM model.

E. LIME-Based Local Interpretability

To complement the SHAP-based global analysis, LIME explanations were generated to evaluate the local interpretability of individual predictions across all models. Figs. 26–41 display the local explanation maps and aggregate feature-importance rankings derived using LIME.

Across all architectures, three key predictors consistently influenced localized outcomes:

- Recent lag values,
- Short-term trend indicators, and
- Daily growth-rate fluctuations.

The LSTM and BiLSTM models (Figs. 26–29) emphasized recent lag features as dominant contributors for most predictions. CNN and RNN models (Figs. 30–33) focused more heavily on immediate temporal variations, while the MLP baseline (Figs. 34–35) exhibited less interpretable distributions. Hybrid models, particularly CNN–LSTM and CNN–BiLSTM (Figs. 36–39), produced coherent local explanations, showing strong correspondence between feature influence and model predictions. Notably, the Hybrid CNN–LSTM model displayed consistent positive contributions from recent confirmed cases and 7-day moving averages when case counts rose, mirroring SHAP-based global trends.

The overall consistency between SHAP and LIME analyses validates the robustness of the interpretability framework and confirms that the model's internal reasoning aligns with realistic epidemic dynamics.

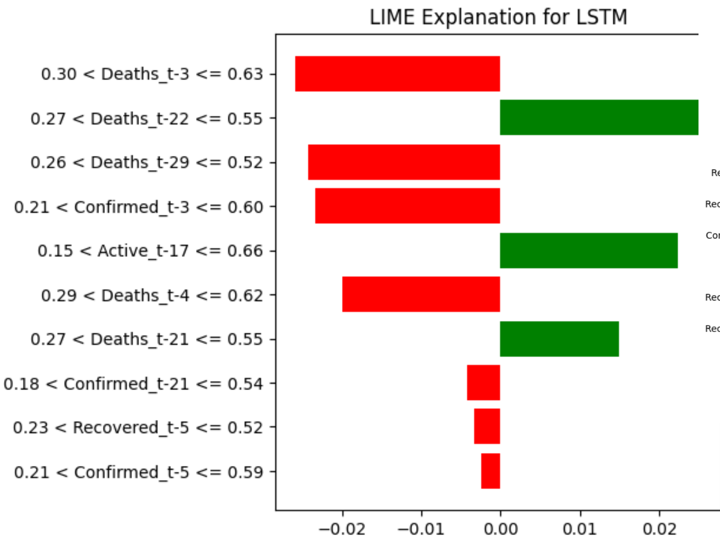

Fig. 26. LIME explanation for an individual prediction using the LSTM model.

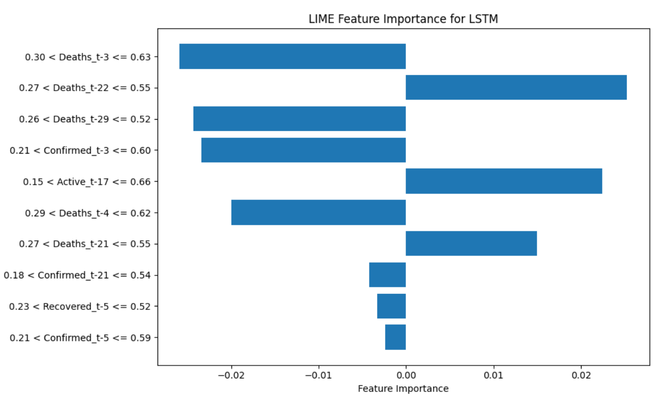

Fig. 27. Aggregate LIME feature importance for the LSTM model.

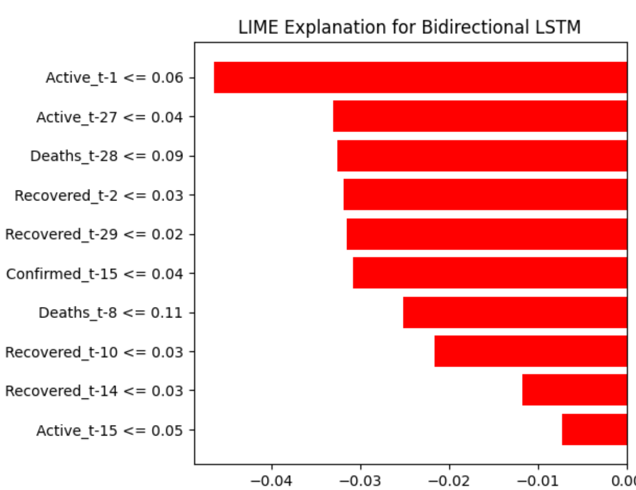

Fig. 28. LIME explanation for an individual prediction using the BiLSTM model.

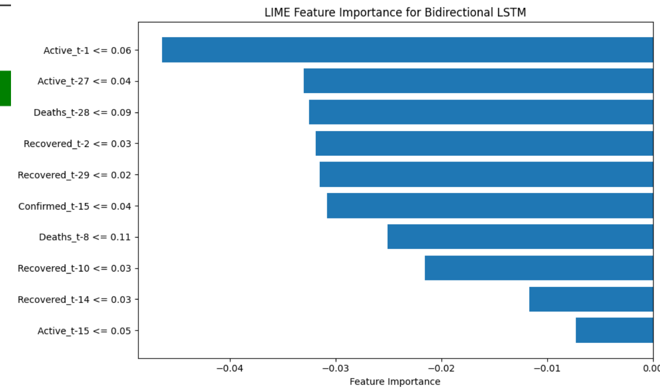

Fig. 29. Aggregate LIME feature importance for the BiLSTM model.

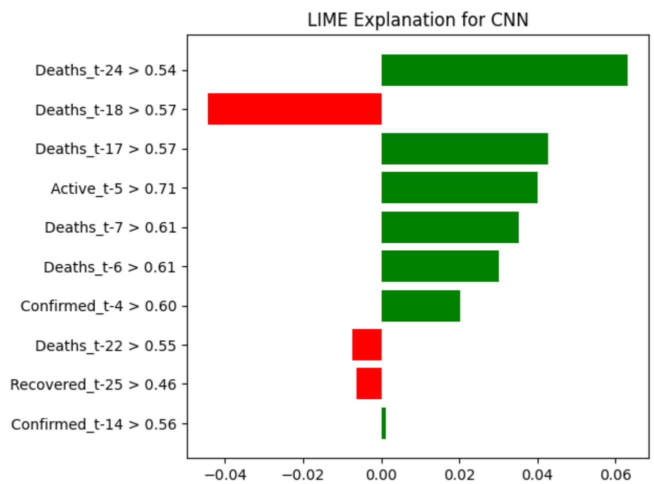

Fig. 30. LIME explanation for an individual prediction using the CNN model.

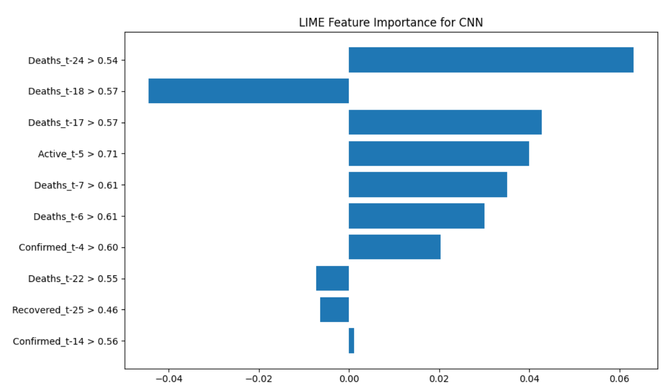

Fig. 31. Aggregate LIME feature importance for the CNN model.

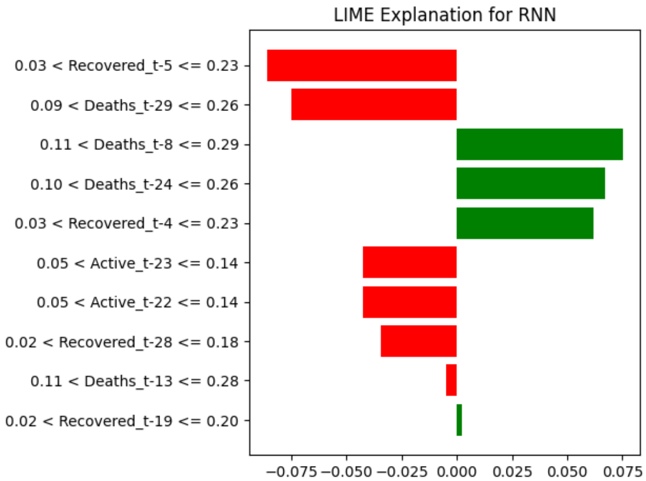

Fig. 32. LIME explanation for an individual prediction using the RNN model.

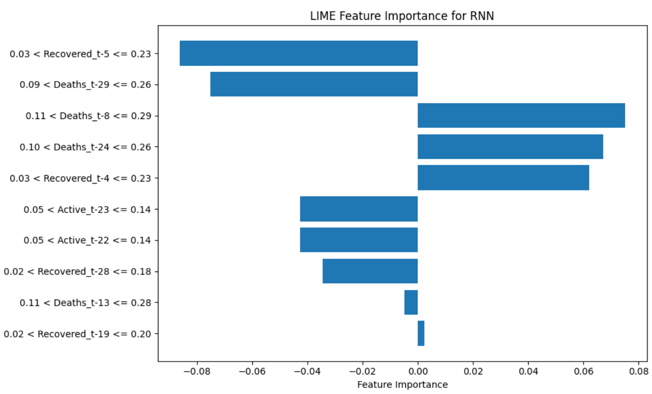

Fig. 33. Aggregate LIME feature importance for the RNN model.

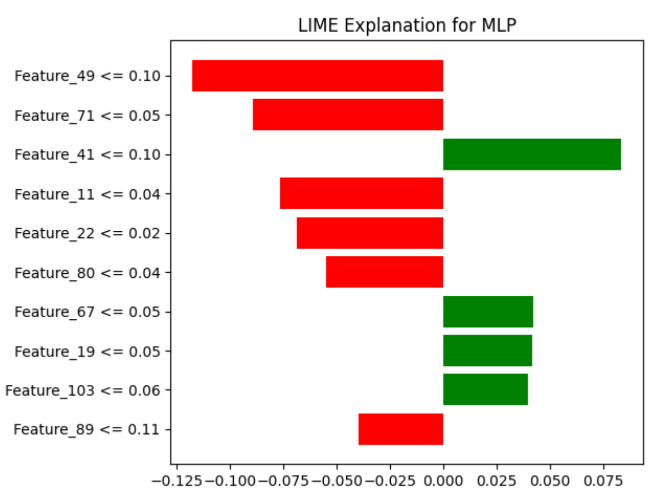

Fig. 34. LIME explanation for an individual prediction using the MLP model.

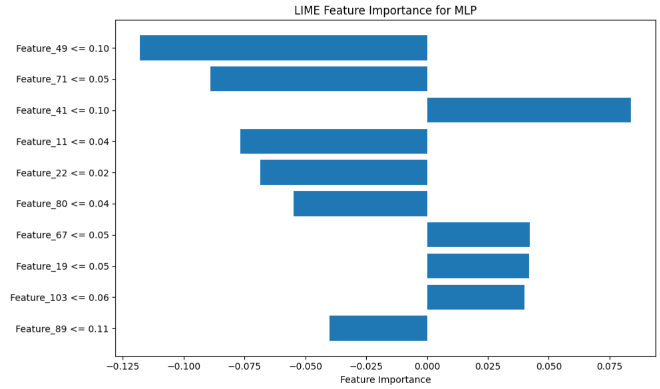

Fig. 35. Aggregate LIME feature importance for the MLP model.

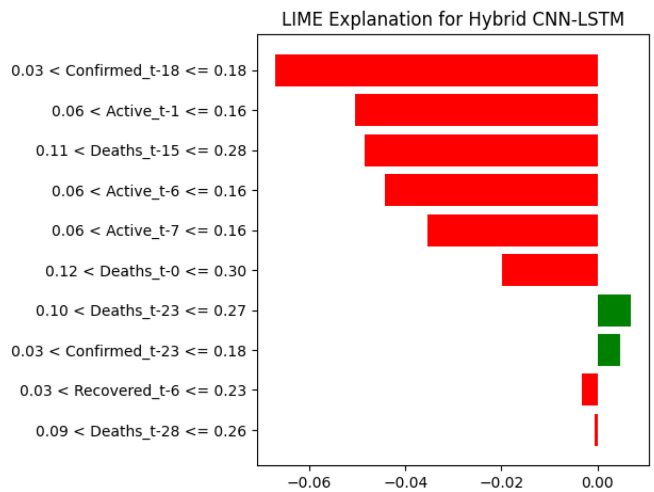

Fig. 36. LIME explanation for an individual prediction using the Hybrid CNN–LSTM model.

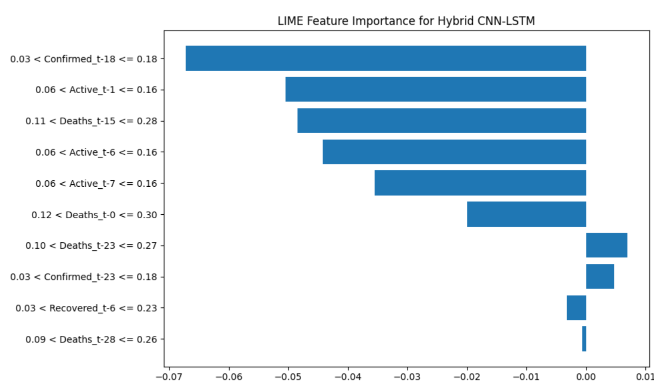

Fig. 37. Aggregate LIME feature importance for the Hybrid CNN–LSTM model.

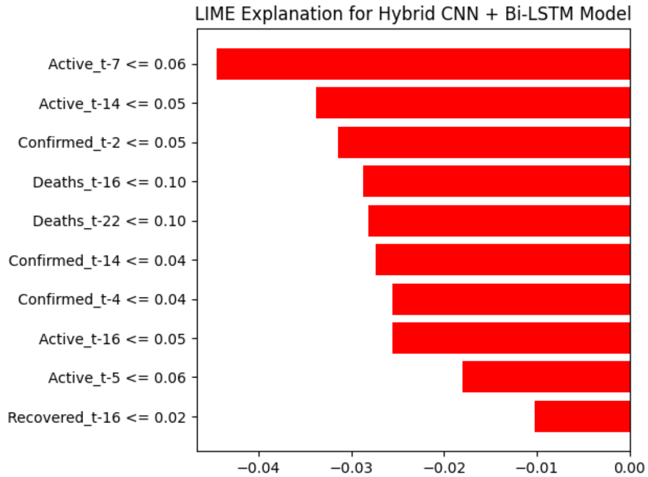

Fig. 38. LIME explanation for an individual prediction using the Hybrid CNN–BiLSTM model.

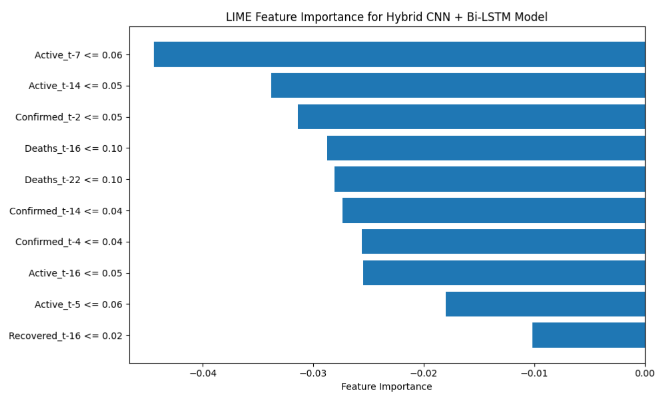

Fig. 39. Aggregate LIME feature importance for the Hybrid CNN–BiLSTM model.

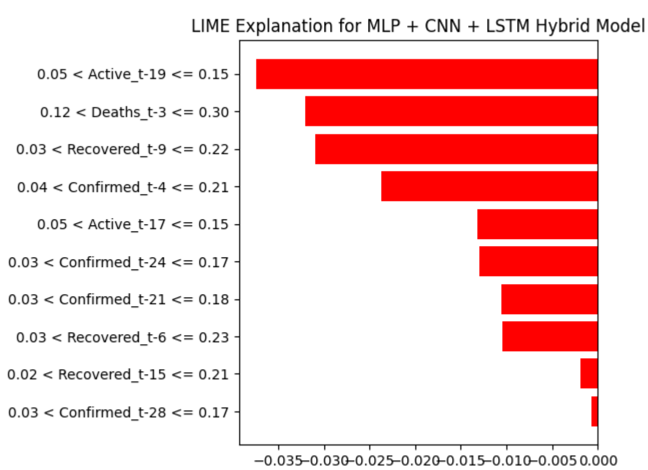

Fig. 40. LIME explanation for an individual prediction using the Hybrid MLP–CNN–LSTM model.

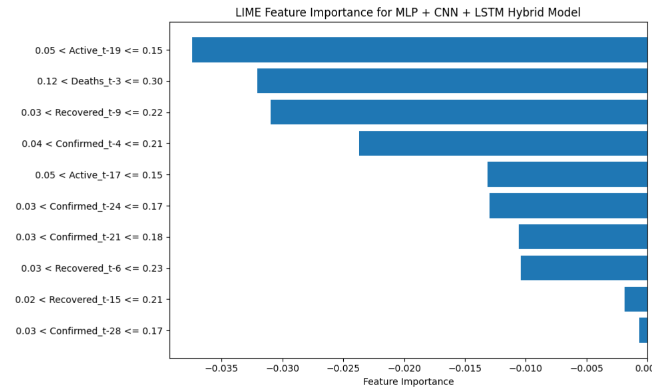

Fig. 41. Aggregate LIME feature importance for the Hybrid MLP–CNN–LSTM model.

The complementary perspectives offered by SHAP and LIME collectively enhance the interpretability of the proposed forecasting framework. While SHAP provides a global understanding of feature contributions across the entire dataset, LIME elucidates localized decision behavior for individual predictions. The convergence of both explanation techniques—highlighting the dominance of recent confirmed cases, moving averages, and short-term growth trends—demonstrates the model's stability and logical consistency. This alignment between quantitative performance and qualitative interpretability underscores the reliability of the Hybrid CNN–LSTM architecture and provides a strong foundation for informed epidemiological decision-making, as discussed in the following section.

F. Discussion

The comparative evaluation of multiple deep learning architectures demonstrates that hybrid models outperform their standalone counterparts for COVID-19 case forecasting. Among all configurations, the Hybrid CNN–LSTM model achieved the best overall performance, attaining the lowest MSE and MAE values with an R^2 exceeding 0.999. Its superior performance stems from the synergy between convolutional layers, which efficiently extract short-term local trends, and recurrent LSTM layers, which capture long-term temporal dependencies in infection trajectories.

Interpretability analysis further reinforced the reliability of this hybrid framework. SHAP-based global explanations revealed that recent confirmed cases, 7-day moving averages, and daily growth rates were the dominant predictors influencing model output. These findings align well with epidemiological intuition, where short-term case surges directly affect near-future transmission patterns. LIME-based local interpretations complemented these insights by providing instance-level transparency—clarifying how each feature contributed positively or negatively to individual predictions. The consistent agreement between SHAP and LIME interpretations confirms that the Hybrid CNN–LSTM model's decision-making process is both stable and logically coherent.

From a practical perspective, these results underscore the importance of coupling predictive accuracy with interpretability in high-stakes forecasting applications. In epidemiological contexts, explainability not only enhances trust among public health officials but also enables data-driven policy interventions supported by transparent evidence. Furthermore, the proposed explainable deep learning pipeline provides a reproducible foundation for extending to other infectious diseases or dynamic systems requiring interpretable temporal modeling.

Overall, the study establishes that high-performing deep learning models, when paired with robust explainability frameworks, can bridge the gap between predictive performance and actionable public health insights.

VI. Conclusion

This study proposed an explainable deep learning framework for COVID-19 case forecasting that combines predictive accuracy with interpretability. Eight neural architectures—including LSTM, BiLSTM, CNN, RNN, MLP, and three hybrid variants—were implemented and compared using standardized evaluation metrics. Among these, the Hybrid CNN–LSTM model consistently achieved the best performance, demonstrating its ability to capture both short-term fluctuations and long-term temporal dependencies inherent in pandemic dynamics.

To address the black-box limitation of deep learning models, the framework integrated SHAP and LIME for model interpretability. SHAP analysis provided global insights into dominant features such as recent confirmed cases, 7-day moving averages, and growth rates, while LIME offered instance-level explanations clarifying localized prediction behavior. The strong alignment between both interpretability techniques validated the transparency and reliability of the proposed model.

By coupling high predictive power with robust explainability, this research contributes to the development of trustworthy AI systems for epidemiological forecasting. The findings underscore that interpretable hybrid models can support more transparent, data-driven decision-making—enhancing preparedness and policy formulation during public health crises. Future work will focus on extending this approach to multi-country datasets, integrating additional contextual variables such as vaccination rates and mobility data, and evaluating real-time deployment scenarios for adaptive forecasting.

VII. Future Work

Future extensions of this study will focus on enhancing both the scalability and generalizability of the proposed framework. First, the model can be expanded to incorporate additional epidemiological and socio-behavioral variables such as vaccination rates, population density, testing frequency, and mobility patterns, enabling richer context-aware forecasting. Second, cross-country and regional transfer learning experiments will be explored to assess the adaptability of the Hybrid CNN–LSTM architecture under diverse outbreak dynamics. Third, integrating attention mechanisms or transformer-based temporal encoders may further improve long-range dependency modeling while preserving interpretability. Finally, a real-time deployment prototype leveraging cloud-based infrastructure and interactive dashboards will be developed to support dynamic monitoring, public health planning, and early-warning decision systems. These directions aim to advance explainable AI for epidemiological forecasting toward practical, high-impact, and trustworthy real-world applications.

References

[1] M. U. Tariq and S. B. Ismail, "Deep learning in public health: Comparative predictive models for COVID-19 case forecasting," *Journal of Public Health Research*, vol. 13, no. 1, pp. 45–59, 2024. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0294289

[2] A. Gupta and S. Gupta, "Deep learning for COVID-19 detection and diagnosis: A review," *Artificial Intelligence in Medicine*, vol. 115, p. 101016, 2021. doi: 10.1016/j.artmed.2021.101016.

[3] R. Sharma and A. Kumar, "Predicting COVID-19 cases using machine learning: A case study of India," *International Journal of Healthcare Management*, vol. 15, no. 1, pp. 12–23, 2022. doi: 10.1080/20479700.2021.1907692.

[4] Y. Liu and X. Wang, "COVID-19 pandemic forecasting with AI: A survey of reinforcement learning and other AI models," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3081–3104, 2023. doi: 10.1007/s10462-022-10100-2.

[5] S. H. Chin and Y. J. Lee, "A hybrid deep learning model for COVID-19 time-series forecasting," *Applied Sciences*, vol. 12, no. 10, p. 4970, 2022. doi: 10.3390/app12104970.

[6] R. Ranjan, V. Kumar, and P. Singh, "Deep learning models for COVID-19 prediction: A review," *IEEE Access*, vol. 9, pp. 125180–125195, 2021. doi: 10.1109/ACCESS.2021.3105145.

[7] B. A. Mohammed and H. F. Ali, "Predictive modeling of COVID-19 spread using machine learning techniques: A case study in the Middle East," *Health Information Science and Systems*, vol. 9, no. 1, p. 12, 2021. doi: 10.1007/s13755-021-00301-y.

[8] A. Bansal and D. Rathi, "Impact of deep learning techniques on COVID-19 forecasting: A comparative study," *Computational and Mathematical Methods in Medicine*, vol. 2022, p. 8791428, 2022. doi: 10.1155/2022/8791428.

[9] Johns Hopkins University, "Novel coronavirus (COVID-19) dataset," 2020. [Online]. Available: https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset

[10] D. Dasmahapatra, "COVID-19 forecasting using explainable AI," 2024. [Online]. Available: https://github.com/dyra-12/Covid-19-Forecasting-using-XAI.git