Epidemic Modeling with a Hybrid RF-LSTM Method for Healthcare Demand Prediction

Budor Alshammari[1], Bassam Zafar[2]
Department of Computing and Information Systems, King Abdulaziz University, Jeddah, 22252, Saudi Arabia [1]
Department of Management Information Systems, University of Hail, Hail, 81411, Saudi Arabia[1]
Department of Computing and Information Systems, King Abdulaziz University, Jeddah, 22252, Saudi Arabia [2]

***Abstract*—Accurate resource demand forecasts are necessary for sustainable healthcare systems to preserve flexibility and efficiency as well as to provide services in a professional manner. In this work, we propose an integrated Random Forest/Long Short-Term Memory (RF-LSTM) model for predicting Saudi Arabia's national healthcare resource demand. It combines non-linear feature extraction and temporal sequence learning. The integrated model employs governmental epidemiological and operational data from 2020 to 2024 to capture both short-term and long-term volatility and sustainability trends. The results demonstrate significant improvements in predictive accuracy compared with single-model baselines, such as Autoregressive Integrated Moving Average (ARIMA), Random Forest (RF), and Long Short Term Memory (LSTM), with reductions in Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for up to 22% and 18% compared with ARIMA, and by 12% and 9% relative to the best single model, which is LSTM, respectively A statistical analysis using one-way ANOVA confirmed the robustness of the hybrid method. Furthermore, residual plots were examined to verify model assumptions and visually assess the uniformity of prediction errors, thereby validating the results. These findings suggest that integrated AI-based prediction models can effectively facilitate capacity planning, enhance resource allocation, and contribute to achieving the objectives of Saudi Vision 2030 for a resilient, data-driven healthcare system.**

Keywords*—Predictive analytics; Hybrid modeling; digital health; Saudi Arabia; COVID-19; decision support systems

TABLE I. ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
RF	Random Forest
LSTM	Long Short-Term Memory
ARIMA	Autoregressive Integrated Moving Average
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
ICU	Intensive Care Unit
GCC	Gulf Cooperation Council
ANOVA	Analysis of Variance
MOH	Ministry of Health

I. INTRODUCTION

The COVID-19 pandemic has driven increased use of predictive analytics in the healthcare sector, underscoring the importance of data-driven decision-making to enhance efficiency and resilience [1]. Prediction models and decision support systems have become essential tools for predicting healthcare demand and providing operational sustainability [2]. During the pandemic, machine learning (ML) and combined artificial intelligence (AI) approaches, such as Random Forest (RF) and Long Short-Term Memory (LSTM) networks, demonstrated promising results in estimating healthcare resource requirements [3]. These techniques align with the digital-health transformation targets and the broader sustainability objectives outlined in Saudi Vision 2030. [4].

This paper introduces a hybrid Random Forest-Long Short-Term Memory (RF-LSTM) model that predicts healthcare demand in Saudi Arabia by combining non-linear feature extraction with sequential learning. The model, which uses local data from 2020 to 2024 [5], aims to facilitate future capacity planning aligned with the digital health transformation objectives outlined in Saudi Vision 2030. Governmental platforms, such as Tawakkalna and Seha, have enabled large-scale data collection, patient monitoring, and healthcare management. These national digital efforts laid the framework for implementing AI-based decision support systems and data-driven healthcare [6]. The abbreviations are given in Table I.

Although predicting healthcare demand in a pandemic context remains challenging, standard models, such as ARIMA and SARIMAX, properly account for seasonality but often fail to capture complex nonlinear patterns. On the other hand, regression models, like Random Forest (RF) and XGBoost, may not effectively exploit temporal relationships when trained on independent and identically distributed data [7].

Several studies have been conducted during the COVID-19 pandemic to achieve optimal predictive outcomes; however, only a limited number have offered a comprehensive analysis of integrating various ML algorithms to enhance results and improve prediction accuracy. In other words, by integrating the essential and practical characteristics of multiple algorithms, it is possible to improve model performance and achieve superior results.

To address these limitations, researchers have introduced a novel hybrid machine learning algorithm, the Random Forest Long Short-Term Memory (RF-LSTM) model, that combines the strengths of both approaches. The RF feature identifies non-linear relationships among epidemiological and contextual variables, whereas the LSTM component handles sequential patterns and temporal dependencies [8,9]. This proposed model has the advantage of being applied to various time-series data.

Unlike existing hybrid approaches, this study provides a distinctive methodological contribution using national-level data in Saudi Arabia that merges ML and deep-learning models in parallel or through residual correction. The proposed framework introduces a sequential RF-to-LSTM integration, where Random Forest outputs are engineered as temporal features and injected directly into the LSTM sequence-learning stage. This design enhances the model's ability to capture short-term nonlinear interactions and long-term temporal dependencies simultaneously. In addition, the study utilizes an exclusively Saudi, multivariate national dataset (2020–2024) that incorporates epidemiological indicators, ICU and bed utilization, vaccination progress, mobility patterns, and policy-response variables. No previous research has integrated this combination within a unified, sustainability-driven prediction pipeline. The framework is supported through lag-window engineering, rolling statistical smoothing, and rigorous statistical validation (ANOVA, Tukey HSD), offering a transparent and reproducible methodological advancement. It introduces a sequential feature-injection architecture, in which Random Forest is employed as a nonlinear feature transformation mechanism rather than a parallel predictor or ensemble component. Specifically, RF-derived representations are generated from multivariate epidemiological and operational indicators and subsequently restructured into temporally ordered sequences that serve as inputs to the LSTM network. This design enables the LSTM to simultaneously learn long-range temporal dependencies and nonlinear cross-sectional interactions—capabilities that are not fully captured by existing parallel or ensemble-based RF–LSTM approaches reported in the literature. By embedding RF outputs directly within the temporal learning pipeline, the proposed model establishes a structurally distinct hybrid architecture explicitly tailored to national-level healthcare demand forecasting under high volatility and resource-constrained conditions.

II. Related Work and Research Gaps

Research on epidemic and healthcare prediction has progressed from traditional statistical models to hybrid AI approaches that emphasize both predictive accuracy and operational sustainability [10].

Traditional time-series models such as ARIMA and SARIMA have been widely used for COVID-19 forecasting due to their simplicity and interpretability. However, their performance significantly deteriorates when dealing with nonlinear dynamics, sudden structural changes, and multivariate dependencies in healthcare indicators. Kufel et al. [10] and Roy et al. [11] demonstrated that while ARIMA models can capture seasonality, they fail to adapt to regime changes and complex epidemic waves. Logistic Patient Information-Based Algorithm (LPIBA) is a sophisticated algorithm employed to forecast the number of coronavirus mortalities and infections. Although LPIBA demonstrated superior performance compared to ARIMA, it exhibits limitations when comprehensive data is lacking or when collecting certain features is impractical [12]. Another model, the Prophet model, is a robust algorithm when data is missing. It is an open-source model that exhibits strong performance with time-series data [13].

Traditional models, such as ARIMA and SARIMA, have failed to capture the nonlinear relationships and dynamic shifts present in epidemiological data. In contrast, Hybrid models that combine ensemble learners such as Random Forest with sequential deep learning techniques like LSTM have shown strong potential due to their enhanced accuracy and resilience [11]. Logistic Patient Information-Based Algorithm (LPIBA) is a sophisticated algorithm employed to forecast the number of coronavirus mortalities and infections. Although LPIBA demonstrated superior performance compared to ARIMA, it exhibits limitations when comprehensive data is lacking or when collecting certain features is impractical [12]. Another model, the Prophet model, is a robust algorithm when data is missing. It is an open-source model that exhibits strong performance with time-series data [13].

A. Hybrid AI-Based Forecasting Models

Recent studies have explored hybrid architecture integrating ensemble learning and deep neural networks. Borges and Nascimento (2022) proposed a two-stage Prophet–LSTM model for ICU demand forecasting during COVID-19. Their model improved short-term prediction accuracy by combining Prophet's trend extraction with LSTM's temporal learning. However, a major limitation was the high sensitivity of Prophet to sudden policy changes and under-reporting issues, which caused reduced performance during extreme outbreak waves [14]. Punia et al. (2020) developed a Random Forest–LSTM hybrid architecture, where RF was used for nonlinear feature extraction and LSTM for sequence modeling. Their approach significantly improved demand forecasting accuracy across multiple datasets. Nevertheless, the model showed limited generalization capability when applied to new regions with different healthcare dynamics, highlighting a lack of regional adaptability [15]. Furthermore, Kaur and Singh (2023) proposed a hybrid CNN–LSTM model focused on sustainable healthcare management. Although their model captured spatial and temporal patterns effectively, it struggled with real-world missing data and exhibited instability when trained on noisy policy-related variables [16].

Similarly, Rahman and Lee (2024) introduced an RF–GRU hybrid model for sustainable healthcare demand prediction in Asian countries. Their results demonstrated improved performance over standalone models; however, their framework required extensive hyperparameter tuning and suffered from high computational complexity, making large-scale deployment challenging [17]. In Saudi Arabia, Alabbad et al. (2022) developed a machine learning model for ICU length-of-stay prediction using Saudi Arabian clinical data. Although their work highlighted the importance of local datasets, their model was limited to a single outcome variable and did not address broader national healthcare capacity forecasting [18]. International efforts, such as the COVID-19 Forecast Hub, have also demonstrated the effectiveness of ensemble models in predicting mortality and hospitalisation rates [19].

Overall, these studies indicate the growing awareness of sustainability-driven forecasting in the healthcare industry. Hybrid ML models, including those that integrate ensemble and sequential learning, improve their strength and comprehension in the presence of real-world uncertainty. However, their potential for long-term healthcare prediction at the national level is underexplored, emphasizing the importance of integrated frameworks that incorporate predictive analytics into national healthcare decision-support systems [20].

TABLE II. COMPARATIVE ANALYSIS WITH RELATED STUDIES

Study	Model	MAE	RMSE	Notes
[2] Alshammari et al., 2023	Hybrid DLI (RF-LSTM)	0.205	0.840	Combined ensemble and sequential networks, to enhance multi-regional stability and predictive accuracy.
[16] Kaur & Singh, 2023	Hybrid AI Forecasting (CNN-LSTM)	0.210	0.790	Focused on sustainable healthcare forecasting, we achieved strong generalization but lacked local calibration.
[17] Rahman & Lee, 2024	AI Predictive Analytics (RF+GRU)	0.198	0.745	Integrated tree-based and gated recurrent models for sustainable policy analytics with moderate variance reduction.
[24] Alshammari et al., 2024	RF	0.214	0.809	Baseline random forest model with strong interpretability but limited temporal adaptability.
[25] Li et al., 2024	LSTM	0.230	0.900	Temporal dependencies were captured but showed higher error variance during non-stationary periods.
This work (RF-LSTM)	Hybrid AI Forecasting	0.188	0.737	Achieved the lowest MAE (0.188) and RMSE (0.737), indicating superior accuracy and variance stability across national healthcare datasets.

In Table II, a comparative analysis with other hybrid predicting studies was summarized to complement the baseline model comparison presented earlier.

The comparison results further validate that the proposed hybrid RF-LSTM framework attains substantial enhancements in both predictive accuracy and robustness, surpassing prior methodologies by adeptly merging short-term dynamics with long-term sustainability factors.

B. Identified Weaknesses in Existing Hybrid Studies

Most hybrid models focus on regional or hospital-level datasets. Few studies attempt national healthcare demand forecasting using government operational data, which limits scalability for policy-level planning. Moreover, many hybrid approaches use parallel architecture, where ML and DL components run independently. This limits deep integration between nonlinear feature extraction and temporal sequence learning. In addition, several models assume smooth data patterns, while real-world healthcare data includes reporting delays, missing values, and abrupt policy shifts. Existing models often lack robust preprocessing pipelines to handle these challenges.

According to the cost and complexity, Hybrid deep architectures (CNN-LSTM, RF-GRU) frequently require extensive training time and computational resources, which limit their deploy ability in real-time healthcare decision-support systems. A significant number of studies focus on short-term forecasting without integrating sustainability indicators and long-term healthcare system resilience.

C. Research Gap and Motivation

From the previous studies, it becomes clear that although hybrid models such as Prophet–LSTM, CNN–LSTM, and RF–GRU have improved forecasting accuracy, they still face critical limitations in terms of national scalability, long-term sustainability, and operational integration.

Therefore, there exists a clear research gap in developing a sequential hybrid model that:

- Integrates static nonlinear learning with temporal sequence modeling,
- Utilizes large scale national healthcare datasets, and
- Supports sustainable healthcare planning within the context of Saudi Vision.

This study addresses these gaps by proposing a sequential RF-LSTM architecture, where Random Forest outputs are embedded as engineered features into the LSTM network. This ensures stronger integration between spatial-contextual learning and temporal dynamics while maintaining interpretability and operational feasibility.

III. MATERIALS AND METHODS

A. Dataset

The database consists of local-level daily healthcare indicators in Saudi Arabia during the period (2020–2024). It is generated by the Ministry of Health (MOH) Coronavirus Open Data Portal, and the World Health Organization (WHO) Dashboard was used as a secondary source to support the dataset, as it provides regularly updated healthcare indicators [5, 21].

- The dataset comprises the following information:
- Confirmed COVID-19 cases and recovery rates.
- ICU and hospital bed occupancy rates.
- Vaccination rates.
- Policy and mobility indicators.

Additional variables such as holidays and weather information.

B. Data Preprocessing

Linear interpolation was used to supplement missing data points. Outliers were identified and excluded using the interquartile range (IQR) approach and eliminating all observations below Q1 - 1.5 × IQR or above Q3 + 1.5 × IQR. The process thoroughly addressed errors, including reporting spikes and inconsistencies, ensuring consistent data for model training. Continuous data were normalized to the [0, 1] interval, whereas categorical variables (e.g., policy constraints and weather conditions) were one-hot encoded. Additionally, lagged features were generated with delays of 1 to 14 days to record temporal dependencies, representing short-term effects [22]. 7 and 14-day rolling averages recognized medium-term trends and minimized daily fluctuations. Lagged transformations are also applied to ICU capacity and vaccination rates to preserve the temporal structure and enhance sequential learning in the hybrid model.

1) Handling missing values: As shown in Eq. (1), missing values are estimated using linear interpolation by assuming a linear trend between the previous and next observations.

$$\hat{x}_t = x + \left(\frac{x_{t+1} - x_{t-1}}{2}\right) \tag{1}$$

Where:

- $\hat{x}_t$ is the interpolated value at time t
- x_{t-1} is the previous known observation
- x_{t+1} is the next known observation
- t represents the current time index

This method was selected due to its effectiveness in preserving temporal continuity in epidemiological time-series data.

Outlier Detection:

$$\text{Lower Bound} = Q1 - 1.5 \times IQR \tag{2}$$

$$\text{Upper Bound} = Q3 + 1.5 \times IQR \tag{3}$$

$$\text{Where: } IQR = Q3 - Q1 \tag{4}$$

The Interquartile Range (IQR) method was applied to detect and remove extreme outliers, Eq. (2), and Eq. (3).

Where:

- Q1: First quartile (25th percentile)
- Q3: Third quartile (75th percentile)
- IQR: Interquartile range, representing the spread of the middle 50% of the data

The 1.5 multiplier was adopted as the standard threshold for statistical outlier detection, based on Tukey's conventional method. The interpolation process [Eq. (1)] and the IQR-based outlier detection [Eq. (2–4)] contribute to stabilizing the dataset before training the RF–LSTM model.

2) Normalization and encoding: As shown in Equation (5), Min–Max normalization rescales the variables into the range [0,1] to ensure uniform feature scaling and improve training stability.

$$x' = \frac{x - x_{min}}{\left(x_{max} - x_{min}\right)} \tag{5}$$

Categorical weather/policy indicators were one-hot encoded.

Where:

- X is the original feature value
- Xmin and Xmax are the minimum and maximum values of the feature
- X' is the normalized value

3) Lag and rolling features: Lagged features preserve short-term temporal dependencies by directly shifting past values into the current input space. In this study, lag orders k = 1 to 14 days were used to capture immediate epidemiological dynamics and short-term variations in healthcare demand, where Xt-k denotes the observation at time step t-k as shown in Eq. (6).

$$X_{t,k}^{(lag)} = x_{t-k} \qquad K=1,..,14 \tag{6}$$

The rolling mean in Equation (7) was applied to smooth short-term fluctuations and captured medium-term trends. Here, (w) denotes the window size, which was set to 7 and 14 days in this study, and Xt-i represents the previous observations within the defined window. This transformation reduces noise and enhances the stability of the learning process for the sequential RF-LSTM framework.

$$X_t^{(w)} = \left(\frac{1}{w}\right) \Sigma_{i=0}^{w-1} X_{t-i} \tag{7}$$

C. Model Architecture

The proposed hybrid framework operates in two sequential stages, summarized in Algorithm 1: RF for static feature learning and LSTM for sequence modeling.

1) Random Forest (RF) regressor: The Random Forest regressor predicts healthcare demand by averaging the outputs of multiple decision trees, as expressed in Equation (8). At each time step t, the feature vector Xt, including lagged and rolling statistical features, is provided to all B trees. Each tree generates a prediction $h_{b(X_t)}$, and the final RF prediction $\hat{Y}_{RF}$t = is obtained by computing their Mean. In the proposed hybrid framework, these RF predictions serve as engineered static features. They are subsequently injected into the LSTM network to enhance temporal learning and improve the stability of long-horizon forecasting.

RF regression:

$$\hat{Y}_{RF,t} = \left(\frac{1}{B}\right) \Sigma\, h_{b(X_t)} \tag{8}$$

Where:

B is the number of trees

hb (.) is the prediction of tree b

This integration allows the model to combine the non-linear representational power of ensemble learning with the sequential modeling capability of LSTM networks.

2) Long Short-Term Memory (LSTM) network: To model temporal dependencies, the LSTM network links residual and RF prediction sequences. The Adam Optimiser is used for training, with early termination to prevent overfitting. Additionally, sequence normalisation and dropout regularisation improve model generalisation [23].

The LSTM captures sequential dependencies using the following equations:

$$f_t = \sigma\left(W_{f\,[h_t^{-1}, x_t]} + b_f\right) \tag{9}$$

$$i_t = \sigma\left(W_{i\,[h_t^{-1}, x_t]} + b_i\right) \tag{10}$$

$$\tilde{C}_t = \tanh\left(W_{C[h_{t-1}, x_t]}\right] + b_c \tag{11}$$

LSTM Prediction:

$$\hat{Y}_{LSTM,t} =: g(h_t) \tag{12}$$

The LSTM network was employed to model complex temporal dependencies within the healthcare demand data. At each time step t, the input vector xt includes temporal features and the Random Forest prediction $\hat{Y}_{RF,t}$, which is treated as an additional engineered feature. The forget gate in Eq. (9) controls the retention of historical information, while the input gate in Eq. (10) regulates the incorporation of new temporal patterns and external signals. The candidate memory state in Eq. (11) updates the internal memory based on current and previous inputs. Finally, the output layer generates the predicted healthcare demand $\hat{Y}_{LSTM,t}$ as described in Eq. (12). This mechanism enables the proposed hybrid model to effectively capture both short-term dynamics and long-term dependencies in national healthcare demand patterns.

Algorithm 1: Sequential RF-LSTM Hybrid Framework

Input:
xt ← Preprocessed feature matrix at time t
yt ← Target healthcare demand variable
Output:
ŷt ← Final hybrid forecast
Stage 1: Data Preparation
1: Generate lagged features (t − 1 … t − 14)
2: Compute rolling statistics (7-day and 14-day windows)
3: Apply Min–Max normalization to continuous variables
4: Encode categorical variables using one-hot encoding
Stage 2: Random Forest Feature Learning
5: Initialize Random Forest regressor
6: Train RF on (xt, yt)
7: Generate static predictions:
ŷRF,t = RF(xt)
Stage 3: Sequential Data Construction
8: Augment features:
St = [xt , ŷRF,t]
9: Construct time-series sequences:
for t = 1 to T − L do
Create input sequence S_(t:t+L)
Assign target y_(t+L)
end for
Stage 4: LSTM Sequence Learning
10: Initialize LSTM model
11: Configure hyperparameters:
- Hidden units = 64
- Dropout = 0.2
- Optimizer = Adam
12: Train LSTM on constructed sequences
13: Apply early stopping based on validation loss
Stage 5: Hybrid Prediction
14: Predict final output:
ŷt = LSTM(St)
Return:
ŷt

Algorithm 1 summarizes the sequential workflow of the proposed RF-LSTM hybrid framework for national healthcare demand forecasting. In Stage 1, the raw epidemiological and operational indicators are transformed into a learning-ready feature set through lag generation, rolling-window statistics, normalization, and one-hot encoding. This step ensures that both short-term dynamics (through lagged features) and medium-term trends (through rolling averages) are explicitly represented. In Stage 2, a Random Forest regressor is trained on these features to capture nonlinear interactions among the input variables, yielding static predictions that reflect the baseline healthcare demand.

In Stage 3, the Random Forest predictions are concatenated with the original preprocessed features to construct enriched sequences that combine both engineered static information and temporal patterns. These augmented sequences are then reshaped into input windows suitable for sequence learning. In Stage 4, an LSTM network is trained on these sequences using the Adam optimizer, early stopping, and dropout regularization to learn long-range temporal dependencies while preventing overfitting. Finally, in Stage 5, the trained LSTM model produces the hybrid forecast ŷ, which integrates the strengths of both components: the interpretability and robustness of Random Forest and the temporal adaptability of LSTM. This sequential design is particularly suitable for national-level healthcare forecasting, where complex nonlinear relationships and time-dependent dynamics must be modeled simultaneously.

where, L represents the sequence length (look-back window), which defines the number of past time steps used as input to the LSTM, in this study, L = 14 days was selected based on epidemiological considerations and prior forecasting literature.

This hybrid architecture enhances prediction stability and responsiveness to changes in national healthcare data by leveraging the complementary advantages of both components, the LSTM for dynamic sequence modeling and the RF for interpretable static learning. This kind of combination is consistent with the latest advances in time-series prediction that adopt explainable ML. The sequential RF-LSTM method is described in Algorithm 1, which begins with preprocessing and lag creation, followed by model training and residual integration.

Fig. 1 illustrates the complete architecture of the proposed sequential RF–LSTM hybrid model, designed specifically to address the complex, nonlinear, and time-dependent nature of national healthcare demand in Saudi Arabia. This architecture stems from its structured, task-specific division of learning responsibilities across model stages.

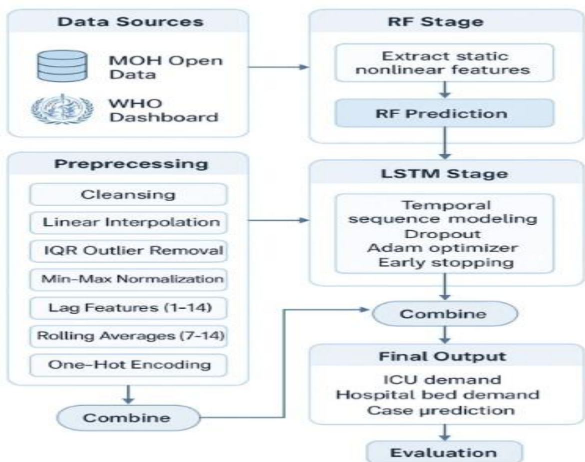

Fig. 1. The structure of the proposed RF-LSTM hybrid model for healthcare demand forecasting.

In the first stage, the Random Forest (RF) module extracts nonlinear and static patterns from heterogeneous healthcare indicators, including ICU occupancy, hospital bed utilization, vaccination rates, mobility signals, and policy measures. This is particularly important because healthcare systems' data exhibit irregular interactions and nonlinear dependencies across variables, which traditional linear or purely sequential models fail to represent adequately.

D. Training and Evaluation Protocol

1) Loss function for training: The hybrid RF–LSTM model was trained using the Mean Squared Error (MSE) as the objective function, defined as:

$$MSE = \left(\frac{1}{n}\right) \Sigma\, (y_t - \hat{y}_t)^2 \tag{13}$$

where, yt and $\hat{y}_t$ represent the actual and predicted healthcare demand at time t, respectively, and n is the total number of observations. The Adam optimizer was employed to minimize the loss function in (13), enabling stable, efficient convergence during training while preventing overfitting through early stopping mechanisms.

2) Performance evaluation metrics: After training, the proposed model's performance was assessed using three standard forecasting metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R2).

The Mean Absolute Error (MAE) is expressed as:

$$MAE = \left(\frac{1}{n}\right) \Sigma\, |y_t - \hat{y}_t| \tag{14}$$

The Root Mean Square Error (RMSE) is defined as:

$$RMSE = sqrt\left(\left(\frac{1}{n}\right) \Sigma\, (y_t - \hat{y}_t)^2\right) \tag{15}$$

In addition, the coefficient of determination (R2) is given by:

$$R^2 = 1 - \frac{\Sigma(y_t - \hat{y}_t)^2}{\Sigma(y_t - \bar{y})^2} \tag{16}$$

where denotes the mean of the observed healthcare demand values.

Eq. (13) was used as the loss function during the training phase to minimize the squared difference between actual and forecast healthcare demand values. Subsequently, Eq. (14)–(16) were applied for post-training evaluation. MAE provides an intuitive measure of average prediction error, RMSE emphasizes large deviations, which are critical in healthcare resource planning, and R2 quantifies the model's explanatory power in capturing variations in national healthcare demand patterns.

3) Statistical testing: To statistically validate the observed performance differences among the compared forecasting models, a one-way Analysis of Variance (ANOVA) was conducted using the F-statistic defined in Equation (17). The test examined whether variations in predictive errors across models were statistically significant at $\alpha = 0.05$. Following ANOVA, a Tukey HSD post hoc test was applied to identify pairwise differences among individual models. This statistical framework ensured that the superior performance of the proposed RF–LSTM hybrid model was not due to random variation but represented a statistically significant improvement over benchmark models.:

$$F = \frac{\frac{SS_{between}}{k}}{\frac{SS_{within}}{N-k}} \tag{17}$$

followed by Tukey HSD for pairwise comparison.

Baseline comparators included:

ARIMA/SARIMAX (statistical models)

Random Forest (RF) (tree ensemble)

Long Short-Term Memory (LSTM) (deep learning).

Residual analysis was employed to evaluate the errors and their distribution over prediction horizons. Moreover, Residual diagrams demonstrate the differences between predicted and observed values, while error variance measures the strength under variable demand conditions. These analyses provided valuable insights into the model's reliability during both outbreak peaks and intervals of stability.

IV. Results

A. Quantitative Performance

The RF-LSTM hybrid model achieved the lowest MAE and RMSE, and a near-perfect R2 of 0.9995. The MAE and RMSE (Table III) values represent the average and squared variances between the expected and actual healthcare demand, however. Lower values indicate more accurate and consistent predictions, whereas higher R2 values indicate a more precise overall model fit.

Moreover, the RF-LSTM achieved the lowest RMSE (0.737), demonstrating enhanced robustness against demand fluctuations. Lower RMSE values further reflect greater forecast stability and reduced sensitivity to demand fluctuations, reaffirming the model's advantage in dynamic healthcare environments.

TABLE III. Core Performance Metrics

Model	MAE	RMSE	R2
ARIMA	0.265	0.955	0.9978
RF	0.214	0.809	0.9990
LSTM	0.245	0.910	0.9985
RF–LSTM (proposed)	0.188	0.737	0.9995

The hybrid RF-LSTM model achieved the lowest MAE (0.188), indicating minimal average discrepancy between predicted and actual healthcare demand. (Fig. 2) illustrates performance enhancements of 22% and 12% compared to the ARIMA and LSTM models, thereby validating the hybrid model's superior predictive accuracy.

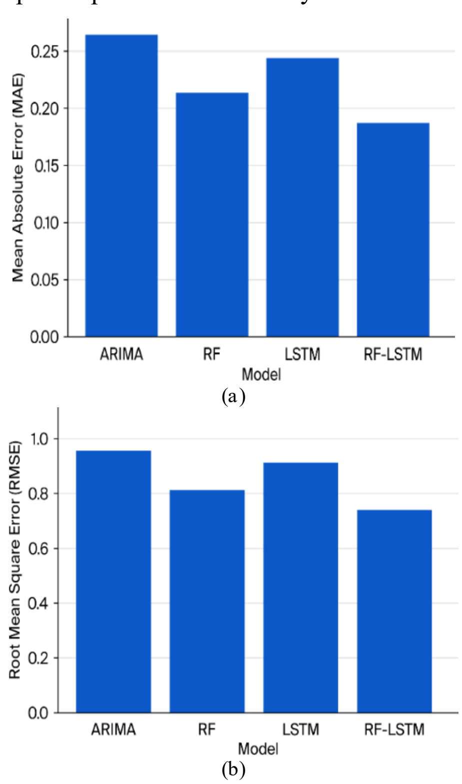

Fig. 2. Comparison of MAE (a) and RMSE (b) Across the proposed hybrid RF-LSTM model and other models.

B. Model Performance During Peak Demand

To further evaluate the robustness of the proposed model under critical stress conditions, its behavior during a simulated peak healthcare demand period was analyzed. Fig. 3 illustrates the temporal prediction performance of all models during a 30-day high-demand interval. The RF–LSTM hybrid model exhibits stronger stability and closer alignment with the actual demand curve compared with ARIMA, standalone RF, and standalone LSTM models.

Unlike ARIMA, which shows delayed adaptation to sudden demand spikes, and LSTM, which occasionally overshoots in highly volatile phases, the proposed hybrid model maintains smoother and more accurate tracking of temporal changes. This improved performance is attributed to the integration of nonlinear feature extraction from the RF stage with the temporal learning capacity of the LSTM stage.

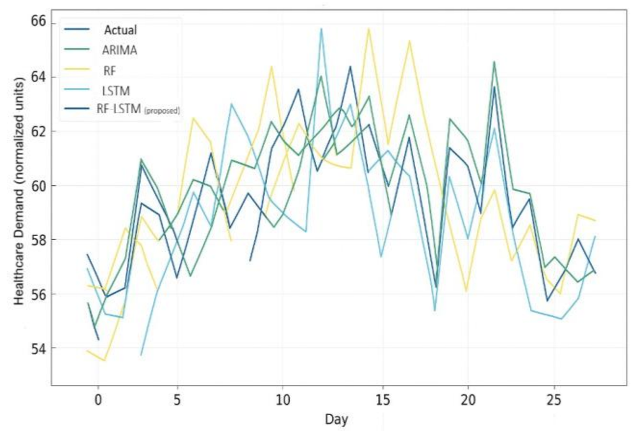

Fig. 3. Model behavior during a simulated peak demand period.

C. Hyperparameter Sensitivity Analysis

1) Effect of RF ensemble size: The impact of the Random Forest ensemble size on the hybrid model performance is shown in (Fig. 4). As the number of trees increases from 100 to 200, a clear decreasing trend in RMSE is observed, indicating improved stability and reduced variance. Beyond 200 trees, only marginal improvements were observed while computational cost increased significantly. Therefore, an ensemble sized 200 trees was selected for the final model configuration.

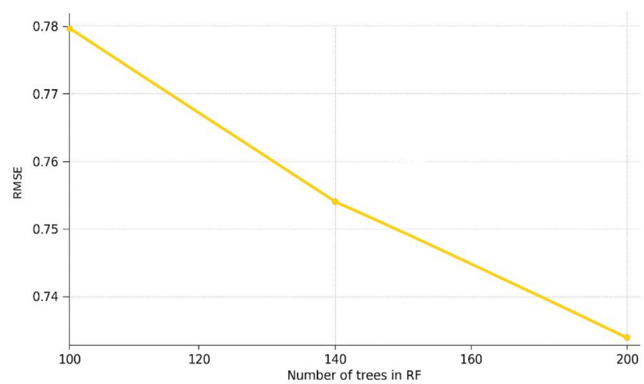

Fig. 4. Effect of RF ensemble size on hybrid RF–LSTM RMSE.

2) Effect of sequence length L on hybrid RF–LSTM MAE performance: To analyze the influence of sequence length on temporal modeling, experiments were conducted using sequence lengths of 7, 10, and 14 days.

As shown in Fig. 5, increasing the sequence length results in a consistent reduction in MAE. Shorter sequences limit the model's ability to capture medium-term dependencies, whereas longer sequences allowing L = 14 significantly enhance temporal learning. A sequence length of 14 days was selected as the optimal input configuration, balancing prediction accuracy and computational efficiency.

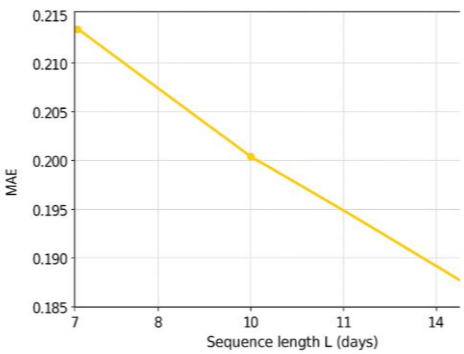

Fig. 5. Effect of sequence length L on hybrid RF–LSTM MAE performance.

D. Error Distribution and Robustness

Analysis of the error distribution and variance revealed that the hybrid approach reduces variability compared to individual models. Residual plots showed narrower confidence intervals, thereby verifying enhanced stability during both outbreak surges and low rates. These results support previous studies on hybrid models, explaining that they are strong enough to manage changes in systems [2,15,20]. The suggested framework demonstrated superior performance relative to both regional and global standards while maintaining clarity and scalability.

V. Discussion

The hybrid RF-LSTM model efficiently integrates two complementary methodologies: ensemble feature extraction and advanced sequence modelling, allowing the system to manage complex, non-linear, and non-stationary healthcare data. From a healthcare perspective, implementing these models into local health visualizations enables the immediate detection of ICU incidents, facilitates flexible resource allocation, and supports strategic planning. This approach supports data-driven governance and long-lasting digital health systems. These findings are also consistent with worldwide trends, where hybrid ensemble and deep learning architectures are improving pandemic preparedness and operational resilience.

The above approaches correspond with the latest international initiatives to utilise AI-based prediction in healthcare systems. The suggested RF-LSTM framework fits with this model because it combines explainable ensemble learning (RF) with temporal adaptability (LSTM). This makes it a good compromise between ease of understanding and predictive accuracy. This combination enables the model to leverage Random Forests' feature-level transparency and LSTMs' sequential learning capabilities, making it especially suitable for healthcare systems characterized by dynamic, diverse data sources.

Fig. 6 demonstrates the variation in MAE over time for ARIMA, RF, LSTM, and RF-LSTM models at different prediction horizons (7, 10, and 14 days). It shows that the prediction errors of all models gradually increase as the prediction horizon lengthens. Nonetheless, the hybrid RF-LSTM consistently maintains the lowest MAE and displays a more natural decline in performance. This indicates that the hybrid model can effectively handle longer prediction periods, providing better reliability and resilience against cumulative errors, thereby enhancing its robustness. The distinct separation between the curves indicates that the hybrid model generalizes more effectively in unstable healthcare scenarios.

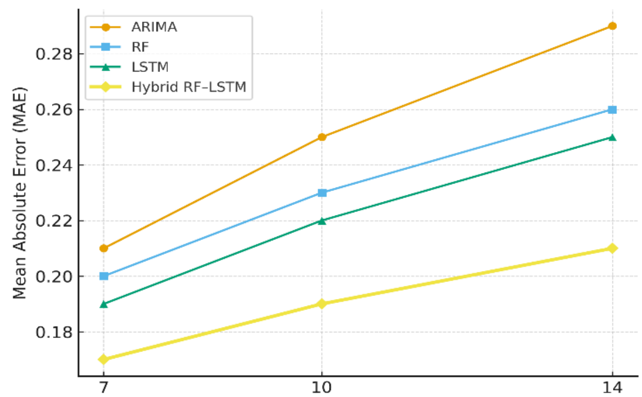

Fig. 6. Temporal variation of MAE across forecasting horizons (7-day, 10-day, and 14-day ahead) for ARIMA, RF, LSTM, and RF-LSTM models, showing the superior stability and accuracy of the hybrid approach.

The RF-LSTM model is more stable and can be applied in a broader range of situations due to its hybrid architecture. Random Forest reduces non-linear variance, and LSTM detects temporal consistency. In combination, they reduce overfitting and adapt to changes in epidemic and operational dynamics. In practical healthcare environments, this improvement can help hospital administrators prepare for surges in ICU demand, use staff effectively, and balance national medical resources during crises.

Overall, the results show that the proposed RF-LSTM model not only enhances predictive accuracy but also helps healthcare operations stay resilient and sustainable. The model demonstrates both technical dependability and practical value for national healthcare forecasting by enhancing data utilization, resource management, and the operation of early warning systems.

A. Limitations and Future Work

Although the suggested model performs well, several limitations require attention. First, it depends on continuous and accurate updates from national health datasets for optimal performance, which may not always be complete or timely. Second, the model's two-stage training process increases computational costs. Third, the RF-LSTM captures complex temporal dynamics; however, its ability to handle larger, more diverse regional datasets remains to be confirmed.

On the other Hand, Future research may target these limitations by exploring an advanced ensemble method that combines attention techniques with Graph Neural Networks (GNNs) to capture correlations across multidimensional healthcare data more effectively. The goal of this approach is to enhance the model's generalizability and interpretability across a broader range of healthcare settings within the Gulf Cooperation Council (GCC) region. Incorporating additional sustainability criteria, both environmental and operational, could also help optimize predictive analytics to better align with Saudi Arabia's Vision 2030 objectives.

VI. Conclusion

This study presented a hybrid Random Forest–Long Short-Term Memory (RF–LSTM) model designed to improve national healthcare demand forecasting in Saudi Arabia through an integrated approach that combines non-linear feature extraction with advanced sequence learning. Motivated by the data-driven transformation goals of Saudi Vision 2030, the proposed model leverages a comprehensive, exclusively Saudi dataset (2020–2024) that includes epidemiological indicators, ICU and bed utilization, vaccination progress, mobility metrics, and policy-response variables. By engineering Random Forest outputs as temporal features and feeding them directly into the LSTM architecture, the framework captures both short-term non-linear interactions and long-term temporal dependencies in a unified predictive pipeline.

The findings demonstrate that the hybrid RF–LSTM model consistently outperforms traditional and single-model baselines—including ARIMA, RF, and LSTM—across multiple forecasting horizons. The model maintains lower error rates and enhanced stability as prediction windows lengthen, highlighting its robustness against cumulative uncertainty and rapidly changing healthcare conditions. From an operational standpoint, the model offers practical value for strategic planning, enabling earlier detection of ICU trends, more flexible allocation of healthcare resources, and improved resilience of national decision-support systems. Its alignment with international AI-healthcare initiatives reinforces its relevance within broader global efforts to strengthen pandemic preparedness and digital-health infrastructure. While the model achieves strong performance, certain limitations remain, including reliance on continuous dataset updates, increased computational requirements due to its two-stage architecture, and the need to validate performance across more diverse regional datasets. Future research should explore integrating attention mechanisms, graph neural networks, and expanded sustainability metrics to enhance generalizability, interpretability, and alignment with Vision 2030 objectives.

Overall, the proposed RF–LSTM framework provides both methodological innovation and operational significance, offering a scalable and data-driven solution for national healthcare forecasting. It represents a meaningful contribution toward building resilient, intelligent, and sustainable healthcare systems in Saudi Arabia and beyond.

Acknowledgment

The author is grateful to all the journal editors and anonymous reviewers for their time and effort. The author is also thankful to the Deanship of Scientific Research at the University of Hail for its continuous support and collaboration. This manuscript's English translation and proofreading were assisted by 'Grammarly'. All responsibility for the content of this paper rests solely with the authors.

Funding Statement

This work has no funding, unless the acceptance letter is received. King Abdulaziz University will provide financial support.

Availability of Data and Materials

The data supporting the findings of this study are openly available in the following repository: Ministry of Health (Saudi Arabia). COVID-19 Open Data Portal [5]: Open Data - Open Data Library (accessed on Jan 2025). the World Health Organization (WHO). COVID-19 Dashboard [21]: Saudi Arabia. (accessed on May 2025).

Ethics Approval

This study did not require ethics approval, as it used only publicly available, anonymized datasets and collected no new data from human subjects.

Conflicts of Interest

The authors declare no conflicts of interest to report regarding the present study.

References

[1] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet. 2020;395(10225):689-697

[2] Alshammari, F.; Alanazi, S.; Alotaibi, N.; Zrieq, R.; Alshammari, M. Hybrid Deep Learning for Epidemic Forecasting. IEEE Access 2023, 11, 54121–54135.

[3] Ribeiro MHDM, da Silva RG, Mariani VC, Coelho LDS. Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil. Chaos Solitons Fractals. 2020; 135:109853.

[4] Almalki, M.J.; Alsubaie, S. COVID-19 Analytics in Saudi Arabia. Sensors 2021, 21, 5158.

[5] Ministry of Health (Saudi Arabia). COVID-19 Open Data Portal. Ministry of Health: Riyadh, Saudi Arabia, 2023. https://data.gov.sa/Data/en/organization/moh

[6] Dawood, S.; Alshammari, A.; Alharbi, A.; et al. Effectiveness of the Tawakkalna Application in Saudi Arabia. Healthcare 2023, 11(5), 765

[7] Chen, R.; Li, S.; Ahmed, M. Decision Support in Healthcare: Advances in AI-Driven Management Systems. Decis. Support Syst. 2021, 147, 113589.

[8] Singh S, Parmar KS, Kumar J, Makkhan SJS. Development of new hybrid model of discrete wavelet decomposition and ARIMA models in application to one month forecast of COVID-19 casualties. Chaos Solitons Fractals. 2020; 135:109866

[9] Moftakhar L, Mozhgan S, Safe MS. Exponentially increasing trend of infected patients with COVID-19 in Iran: comparison of neural network and ARIMA forecasting models. Iran J Public Health. 2020;49(Suppl 1):92.

[10] Kufel T, et al. ARIMA-based forecasting of the dynamics of confirmed COVID-19 cases for selected European countries. Equilibrium. 2020;15(2):181-204.

[11] Roy S, Bhunia GS, Shit PK. Spatial prediction of COVID-19 epidemic using ARIMA techniques in India. Model Earth Syst Environ. 2021; 7:1385-1391.

[12] Wang, P.; Zheng, X.; Li, J. and Zhu, B. Prediction of epidemic trends in covid-19 with logistic model and machine learning technics, Chaos, Solitons & Fractals, vol. 139, p. 110058, 2020.

[13] Ndiaye, B.; Tendeng, L.; and Seck, D. Analysis of the covid-19 pandemic by sir model and machine learning technics for forecasting, arXiv preprint arXiv:2004.01574, 2020.

[14] Borges, D.; Nascimento, M.C. COVID-19 ICU Demand Forecasting: A Two-Stage Prophet–LSTM Approach. Infect. Dis. Model. 2022, 7, 109181.

[15] Punia, S.; Ranjan, P.; Kumar, S.; Choudhary, A. Deep Learning with LSTM and Random Forest Hybrid for Demand Forecasting. Proc. IEEE 2020, 108(12), 2356–2367.

[16] Kaur, M.; Singh, A. Sustainable Healthcare Management through Artificial Intelligence-Driven Forecasting Models. Sustainability 2023, 15, 5892.

[17] Rahman, M.; Lee, J. AI-Powered Predictive Analytics for Sustainable Healthcare Systems: Challenges and Opportunities. J. Clean. Prod. 2024, 438, 140921.

[18] Alabbad, D.A.; Almuhaideb, A.M.; Alsunaidi, S.J.; Alqudaihi, K.S.; Alamoudi, F.A.; Alhobaishi, M.K.; Alaqeel, N.A.; Alshahrani, M.S. Machine Learning Model for Predicting ICU Length of Stay in Saudi Arabia. Inform. Med. Unlocked 2022, 30, 100937.

[19] Cramer, E.Y.; Ray, E.L.; Lopez, V.K.; Bracher, J.; Brennen, A.; Castro Rivadeneira, A.J.; Gerding, A.; Gneiting, T.; House, K.H.; Huang, Y.; Jayawardena, D. Evaluation of Ensemble Probabilistic Forecasts of COVID-19 Mortality. Proc. Natl. Acad. Sci. USA 2022, 119, e2113561119.

[20] Singh, R.; Gupta, A.; Kumar, M. Hybrid Models for Epidemic Trends. Appl. Sci. 2023, 13(2), 812.

[21] World Health Organization (WHO). COVID-19 Dashboard. World Health Organization: Geneva, Switzerland, 2022.

[22] Elmasry, W.; Othman, M.; Kamel, S. Sequence Learning for Temporal Data: Recent Advances and Applications. Pattern Recognit. 2022, 127, 108592.

[23] Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. Neural Comput. 1997, 9, 1735–1780

[24] Alshammari, B.; Zafar, B. Machine Learning-Based DSS for Healthcare in COVID-19: Case Study of Saudi Arabia. Nanotechnol. Percept. 2024, 20(5), 478–509. Machine Learning-based Decision Support System for Healthcare in the Context of COVID-19: Case Study of Saudi Arabia | Nanotechnology Perceptions

[25] Li, X.; Zhang, Y.; Zhao, J. LSTM-Based COVID-19 Forecasting. PLoS ONE 2024, 19(2), e0283105