computation

MDPI

Article

Smoothing Techniques for Improving COVID-19 Time Series Forecasting Across Countries

Uliana Zbezhkhovska [1] and Dmytro Chumachenko [2,3,4,5,*]

[1] Scientific and Methodical Department for Quality Assurance of Educational Activities and Higher Education, Ivan Kozhedub Kharkiv National Air Force University, 61023 Kharkiv, Ukraine; ulianazbezhkhovska@gmail.com
[2] Mathematical Modelling and Artificial Intelligence Department, National Aerospace University "Kharkiv Aviation Institute", 61072 Kharkiv, Ukraine
[3] Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4] Ubiquitous Health Technology Lab, University of Waterloo, Waterloo, ON N2L 2G5, Canada
[5] Balsillie School of International Affairs, Waterloo, ON N2L 6C2, Canada
* Correspondence: dichumachenko@gmail.com

Abstract: Accurate forecasting of COVID-19 case numbers is critical for timely and effective public health interventions. However, epidemiological data's irregular and noisy nature often undermines the predictive performance. This study examines the influence of four smoothing techniques—the rolling mean, the exponentially weighted moving average, a Kalman filter, and seasonal–trend decomposition using Loess (STL)—on the forecasting accuracy of four models: LSTM, the Temporal Fusion Transformer (TFT), XGBoost, and LightGBM. Weekly case data from Ukraine, Bulgaria, Slovenia, and Greece were used to assess the models' performance over short- (3-month) and medium-term (6-month) horizons. The results demonstrate that smoothing enhanced the models' stability, particularly for neural architectures, and the model selection emerged as the primary driver of predictive accuracy. The LSTM and TFT models, when paired with STL or the rolling mean, outperformed the others in their short-term forecasts, while XGBoost exhibited greater robustness over longer horizons in selected countries. An ANOVA confirmed the statistically significant influence of the model type on the MAPE ($p = 0.008$), whereas the smoothing method alone showed no significant effect. These findings offer practical guidance for designing context-specific forecasting pipelines adapted to epidemic dynamics and variations in data quality.

Keywords: COVID-19 forecasting; time series smoothing; STL decomposition; LightGBM; LSTM; Temporal Fusion Transformer; Kalman filter; epidemic modeling

check for updates

Academic Editor: Simone Brogi

Received: 28 April 2025
Revised: 19 May 2025
Accepted: 21 May 2025
Published: 3 June 2025

Citation: Zbezhkhovska, U.; Chumachenko, D. Smoothing Techniques for Improving COVID-19 Time Series Forecasting Across Countries. *Computation* **2025**, *13*, 136. https://doi.org/10.3390/computation13060136

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Accurate, forward-looking estimates of epidemic activity remain a frontline decision support tool for ministries of health, hospital networks, and multilateral agencies. Almost five years after the first reports of SARS-CoV-2, this virus still produces hundreds of thousands of new laboratory-confirmed infections every month. It continues to strain the intensive care capacity in several regions [1]. At the same time, regular surveillance streams have become more heterogeneous as countries have downscaled testing, shortened reporting lags, or adopted sentinel schemes. These shifts intensify the statistical noise already created by the clearance of backlogs, holiday effects, and sudden behavioral

changes, and they complicate the task of building forecasting systems that can be transferred from one national setting to another.

A well-established way to attenuate this noise is to smooth the raw time series before fitting the model. Classical filters, like moving averages (MAs), Holt–Winters exponential smoothing, and Kalman variants, have long been applied to epidemic curves. Recent reviews list these among the most frequently recommended preprocessing steps for health emergencies [2]. Empirical studies confirm their practical value: centered MAs improved short-term COVID-19 projections by reducing the mean absolute percentage error below 6% [3]. At the same time, seasonal–trend decomposition using Loess (STL), coupled with the seasonal autoregressive integrated moving average (SARIMA), captured extreme surges more faithfully than unfiltered baselines [4]. Adaptive Kalman filtering has also been used successfully to track non-stationary case volatility in real time, outperforming static statistical models during the Omicron wave [5]. On the modeling side, modern machine learning architectures such as long short-term memory (LSTM) networks and the Temporal Fusion Transformer (TFT) increasingly dominate the accuracy rankings for multi-horizon epidemic forecasts. LSTM variants augmented with attention mechanisms recently delivered sub-10% two-week error rates for Japanese prefecture-level predictions [6], and TFT has proved resilient in other pandemic-affected domains that require interpretable time series estimates [7].

Despite these advances, the literature still treats the choice of smoother and the choice of forecasting model as largely independent design decisions. Comparative studies usually fix one dimension (for example, they benchmark several learners on a single smoothed signal or test several smoothers with a single learner) and rarely explore how these dimensions interact when the data quality, epidemic phase, and prediction horizon vary across countries. As a result, practitioners lack systematic evidence on whether the benefits of smoothing depend more on the statistical properties of the filter, on the inductive biases of the learner, or on the forecasting window itself.

This study aims to determine how four common smoothing techniques—the rolling mean, the exponentially weighted moving average (EWMA), a Kalman filter, and STL—affect the short- (3-month) and medium-term (6-month) accuracy of four forecasting models (LSTM, TFT, XGBoost, and LightGBM) when applied to weekly COVID-19 case data from Ukraine, Bulgaria, Slovenia, and Greece.

To achieve this aim, the following tasks were formulated:

1. Assemble and standardize weekly COVID-19 incidence series for the four countries;
2. Apply the four smoothing methods to every series with uniform parameter settings;
3. Train and evaluate each forecasting model on every smoothed dataset for both prediction horizons using the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE);
4. Use a two-way ANOVA to quantify the main and interaction effects of the choice of smoothing and model architecture on the forecast error.

This study is, to our knowledge, the first to jointly evaluate the effects of smoothing techniques and forecasting models across multiple national COVID-19 datasets with varied quality of reporting. The results offer practical, generalizable guidance for public health forecasting pipelines.

The expected contribution of this work is four-fold. Firstly, it offers the first cross-country, full-factorial comparison that jointly varies the preprocessing and model selection in epidemic forecasting. Secondly, it identifies specific smoother–model pairs that minimize the error and reduce the variance under different horizon lengths and data quality profiles. Thirdly, it shows that while the model architecture is the main driver of accuracy, the right smoothing technique can still cut the variance in the forecasts by up to one-third, providing

clear guidance for pipeline design. Fourthly, it offers reproducible code and country-level recommendations that public health agencies can adapt quickly during future outbreaks.

The remainder of this article is organized as follows. Section 2 reviews the recent research on smoothing and machine learning approaches to epidemic forecasting and positions our work within this literature. Section 3 details the data sources, preprocessing steps, smoothing pipelines, and experimental design, including the model architectures and error metrics. Section 4 reports the empirical results for all model–smoother–horizon combinations. Section 5 interprets these findings in light of prior studies, highlights the practical implications for public health agencies, and discusses this study's limitations. Section 6 concludes with the main takeaways and outlines directions for future research.

2. Analysis of the Current Research

Smoothing techniques such as simple MAs, the EWMA, Kalman filters, STL, and exponential smoothing are commonly employed to prepare epidemic data for forecasting models [8–10]. These methods aim to isolate underlying trends and seasonal patterns, thereby improving the performance of both statistical and machine learning models. Traditional smoothing approaches, such as moving averages, have been widely applied to epidemic data [11]. Although these methods are simple and effective for reducing noise, they can also lag behind real-time changes and potentially obscure critical inflection points. Conversely, exponential smoothing techniques provide more responsiveness by assigning greater weight to recent data [12].

The EWMA, a variant of exponential smoothing, has gained prominence in epidemic forecasting due to its ability to detect small shifts in time series data while maintaining computational efficiency. In [13], the authors demonstrated that EWMA charts accurately captured the evolution of COVID-19 mortality in the USA, identifying critical pandemic phases and peak periods across multiple waves. Similarly, Ref. [14] utilized EWMA-based techniques, including robust variants, to enhance the stability and predictive accuracy of epidemic time series from Iraq, particularly in the presence of outliers and irregular fluctuations. The paper [15] further showed that integrating EWMA charts with time series models (e.g., ARIMA and VARMA) enabled the earlier detection of outbreak signals compared to conventional observation-based approaches, highlighting this method's utility for real-time surveillance.

Kalman filters offer a more adaptive approach. In [16], a Kalman filtering framework was employed to estimate the stochastic volatility in daily COVID-19 case rates using time-varying parameters, demonstrating its robustness in capturing volatility clustering, a common feature of epidemic data. In [17], the authors combined Kalman filtering with an ensemble machine learning pipeline, including random forest, to analyze influential environmental and demographic variables. Then, they used the Kalman filter to predict the case trends in the short term. Additionally, Ref. [18] presented a two-stage approach that first classified Indian states based on COVID-19 risk using a naïve Bayes classifier, followed by using Kalman-based smoothing to refine the state-level outbreak predictions. Advanced variants, such as the Cubature Kalman Filter [19], have been used within compartmental models to estimate epidemic states in real time.

Although Kalman filters are highly effective for adaptive smoothing and dynamic estimation, they are often complemented by methods that provide clearer structural decomposition. One such method is STL, a more flexible technique that separates time series into three components: trend, seasonal, and residual. Several studies have successfully applied STL in the context of COVID-19. In [4], STL was combined with the SARIMA to forecast extreme increases in daily COVID-19 cases in Jakarta. The hybrid STL-SARIMA model allowed for accurate estimation of the value-at-risk of infection surges, yielding a

low error and demonstrating STL's ability to enhance the accuracy of the upper-bound predictions under volatile conditions. Similarly, Ref. [20] highlighted the robustness of STL in smoothing time series with pandemic-induced disruptions, showing that STL could be applied without modifications even when the traditional methods (e.g., X13-ARIMA-SEATS) required the treatment of outliers or adjustment to seasonal patterns.

The effectiveness of smoothing methods often depends on the characteristics of the forecasting model. ARIMA models have been widely used to forecast COVID-19 trends across regions, showing a reliable short-term performance [21,22]. Variants like ARIMAX have allowed for the integration of external features such as comorbidities, enhancing hospital resource planning [23]. Hybrid approaches, such as ARIMA-LSTM, combine linear and nonlinear modeling capabilities and have outperformed standalone models in their accuracy [24].

Machine learning models like XGBoost and LightGBM can handle complex, nonlinear data structures and benefit substantially from smoothed input features such as rolling averages and lagged statistics [25]. In [26], an ensemble framework incorporating LightGBM, XGBoost, and random forest was developed to forecast daily COVID-19 cases in the USA. This study emphasized the importance of feature engineering, including smoothed time-lagged variables and policy intervention indicators, and found that LightGBM achieved the highest predictive accuracy among the base learners. In [27], LightGBM and LSTM were used to predict daily new cases in Indonesia. Although LSTM yielded a slightly better accuracy, LightGBM performed competitively, especially when smoothed and normalized features were used. Furthermore, Ref. [28] focused on real-time COVID-19 predictions across five cities in Saudi Arabia, comparing the accuracy and computational efficiency of XGBoost and LightGBM. The results demonstrated that while LightGBM offered strong scalability, XGBoost consistently delivered faster computation, making it more suitable for time-critical public health forecasting.

Deep learning models, especially recurrent architectures like LSTM and GRUs [29], have shown strong potential for COVID-19 time series forecasting. These models are well suited to learning long-term temporal dependencies and nonlinear patterns, but their effectiveness depends heavily on the quality of the input data. Several studies have shown that smoothing and normalization significantly enhance the performance of LSTM when predicting daily case numbers. For example, in [30], LSTM-based models were used for multi-step COVID-19 infection forecasting in India, successfully capturing both the first and second pandemic waves and providing two-month-ahead forecasts. In [16], an optimized LSTM model with a modified output gate (popLSTM) outperformed the standard LSTM models across multiple countries by increasing the prediction accuracy by over 4%. Another study [31] proposed an LSTM framework augmented with an attention mechanism and transfer learning for small-area urban time series predictions. The model, trained on smoothed sequences, demonstrated improved robustness in handling both jagged and flat segments of the COVID-19 curve. These findings underscore the critical role of data preprocessing and model enhancements, such as attention and transfer learning, in improving deep-learning-based forecasting.

While LSTM and its variants are now widely used for epidemic time series modeling, the TFT, a state-of-the-art deep learning architecture for interpretable time series forecasting, remains underexplored in COVID-19. Initially proposed in [32], the TFT combines attention mechanisms with gating layers to model short- and long-term temporal patterns while retaining interpretability. Despite its success in other domains, such as electricity load and retail forecasting, the current literature lacks substantial applications of the TFT to COVID-19 case predictions.

The variability in the COVID-19 data across countries, stemming from differences in their testing capacities, reporting standards, and healthcare infrastructure, poses significant challenges for model comparability and forecast reliability. In response, smoothing techniques have been widely used to harmonize epidemic curves and reduce local fluctuations. Studies have demonstrated their effectiveness in enhancing the forecasting performance in diverse national settings. For instance, penalized smoothing splines revealed the mortality patterns across multiple countries [33], while centered moving averages improved the imputation accuracy in datasets with reporting gaps [34]. Exponential smoothing showed sensitivity to parameter tuning in ASEAN countries [35] and proved helpful in capturing real-time changes in trends during rapidly shifting epidemic phases in Thailand [36]. Other work found that smoothing improved the model performance in cases with short-term volatility [37] and across a range of forecasting methods, including Prophet, Holt–Winters, and LSTM techniques [38].

Beyond improving the predictive accuracy, smoothing enhances forecasts' interpretability and practical usability. Smoothed outputs have been shown to clarify key events such as lockdowns and vaccination effects [33,35], improve the attention and weight distribution of machine learning models [36], and reduce bias during imputation [34].

Despite the proven effectiveness of individual smoothing techniques and forecasting models, prior studies have rarely offered systematic comparisons across multiple smoothing methods, model architectures, and national contexts simultaneously. To address this gap, our study evaluates the combined effects of different smoothing strategies and advanced forecasting models across diverse countries with varying data quality and epidemic volatility. This approach aims to identify effective, context-specific preprocessing strategies that enhance epidemic forecasting in real-world public health settings.

3. Materials and Methods

3.1. The Data and Preprocessing

For this research, we used publicly available COVID-19 case data from the World Health Organization https://data.who.int/dashboards/covid19/data (accessed on 2 March 2025), which includes officially reported new infections, cumulative cases, and deaths for all member countries. The dataset spans the pandemic's full course, from January 2020 through to the end of December 2024, allowing for a long-range temporal analysis.

For this study, we focused on four countries: Ukraine, Bulgaria, Slovenia, and Greece. This selection was based on both qualitative and quantitative criteria, including their geographic proximity, comparable pandemic trajectories, and consistent data reporting. In addition, these countries demonstrated a high degree of correlation in their weekly new case counts, with the Pearson's coefficients exceeding 0.85 over the study period. This ensured epidemiological similarity while preserving administrative diversity and variability in the surveillance practices. To reduce reporting noise and align with the goals of medium-term forecasting, we aggregated the daily case counts into weekly totals, using Monday as the start of the week. This also compensated for inconsistencies caused by weekend reporting delays or administrative batching.

We applied quantile-based normalization to stabilize the scale of reported cases and mitigate the impact of extreme fluctuations [39]. This approach transforms the empirical distribution of each country's time series to approximate a standard normal distribution. The transformation promotes a uniform statistical structure across inputs by mapping the original values to a normal distribution, facilitating more stable and efficient learning, particularly in neural network models. This normalization strategy was selected after hyperparameter tuning on Ukrainian data, where it yielded the lowest forecast error and most stable convergence, particularly for the LSTM and TFT models. While opti-

mized on one country, the same method was applied uniformly across all datasets to ensure comparability.

3.2. Smoothing Techniques

To prepare the weekly time series for forecasting, we applied four distinct smoothing methods to each country's case counts: the MA, the EWMA, Kalman filtering, and STL.

The moving average uses a 7-day symmetric window centered at each time point, incorporating three observations before and after the current value. The smoothed value at time t is given by

$$\tilde{y}_t = \frac{1}{7} \sum_{i=-3}^{3} y_{t+i} \tag{1}$$

This method effectively reduces short-term fluctuations and weekly seasonality but may blur sudden structural changes, especially near outbreak peaks. To address this limitation, we also applied the EWMA, defined recursively as

$$\tilde{y}_t = \alpha y_t + (1 - \alpha)\tilde{y}_{t-1} \tag{2}$$

where $\alpha \in (0, 1)$ is the smoothing factor. Our experiments used a span of 7 days, corresponding to

$$\alpha = \frac{2}{\text{span} + 1} = 0.25. \tag{3}$$

This formulation allows the model to adapt to short-term trends while maintaining smoothness.

The Kalman filter provides a probabilistic framework that models the observed series as a linear dynamical system with Gaussian noise. It estimates an unobserved state x_t recursively as new observations v_t become available:

$$x_t = x_{t-1} + \omega_t, \qquad \omega_t \sim \mathcal{N}(0, Q) \tag{4}$$

$$y_t = x_t + v_t, \qquad v_t \sim \mathcal{N}(0, R) \tag{5}$$

We estimated the parameters using the expectation-maximization algorithm over 10 iterations. The Kalman filter captures the hidden temporal structure and smooths the series by balancing model predictions with the observed data.

To isolate long-term patterns, we used STL, which expresses the observed series as the sum of three components:

$$y_t = T_t + S_t + R_t \tag{6}$$

where T_t is the trend, S_t is the seasonal component, and R_t is the residual. The seasonal period was fixed at 7 days to reflect the weekly periodicity observed in the aggregated data. Only the trend component Tt was retained and used as input to the forecasting models.

All smoothing methods were applied independently to each country's weekly time series using consistent parameters. Our goal was not to optimize each technique per country but to compare their relative impact on the forecasting performance under a uniform preprocessing pipeline.

3.3. The Forecasting Models and Experimental Design

To evaluate how different smoothing techniques affected the accuracy and robustness of COVID-19 case forecasts, we designed a comprehensive experimental setup involving four predictive models that spanned both classical machine learning and deep learning

paradigms: LightGBM, XGBoost, LSTM, and the TFT. Each model was trained on four versions of the preprocessed time series across multiple countries, enabling a systematic comparison of the forecasting performance across smoothing strategies, modeling approaches, and time horizons.

For LightGBM and XGBoost, we adopted a feature-based forecasting formulation. The input feature set included fixed lags (1, 2, 7, and 14 days), rolling statistics (mean, standard deviation, min, and max), percentage changes, and STL-derived components (trend, seasonal, and residual). LightGBM was trained with L1 loss, while XGBoost optimized the squared error using a consistent set of hyperparameters across countries to ensure fair evaluation.

We employed two deep learning models, LSTM and the TFT, to explicitly capture the temporal dependencies. Both were trained on smoothed weekly case counts using a 48-week input window and generated predictions for 12-week and 24-week horizons. Both the LSTM and TFT models were trained on the same input structure: smoothed new case counts enriched with engineered time-varying covariates. These included lagged values (up to 90 days), rolling window statistics (mean, std, min, and max), percentage changes (1-day and 7-day), and components extracted via STL. No external covariates such as vaccination coverage, policy interventions, or mobility indicators were used. LSTM was implemented as a deep recurrent network optimized using quantile regression, enabling predictive intervals through quantile loss:

$$L_{\text{quantile}} = \sum_{q \in Q} \max(q(y - \hat{y}),\ (q-1)(y - \hat{y})) \tag{7}$$

with $Q = \{0.1,\ 0.25,\ 0.5,\ 0.75,\ 0.9\}$.

The TFT used the same quantile-based strategy, with four attention heads, a hidden size of 256, and all engineered features provided as time-varying covariates. Standard regularization techniques (early stopping, dropout, and gradient clipping) were applied to both models.

All models were trained on normalized and smoothed weekly case data. Two forecasting scenarios were defined for each dataset, short-term (12 weeks) and medium-term (24 weeks), with the prediction windows beginning in June and September 2024, respectively. To reflect realistic deployment conditions, the models were trained once for each country and forecast horizon and remained fixed throughout the prediction period.

Forecast accuracy was evaluated using the RMSE, MAE, and MAPE, all computed on inverse-transformed predictions. The resulting experimental grid included 128 configurations (4 models × 4 smoothing methods × 2 horizons × 4 countries), enabling a robust comparison across the dimensions of interest.

4. Results

Building on the experimental setup described in the previous section, we evaluated the impact of smoothing techniques on the forecasting accuracy. To support a comparative evaluation, the results are visualized using boxplots (to illustrate the variability in the predictions), heatmaps (to summarize the average performance across countries), and line plots (to compare the predicted versus actual case trajectories).

Figure 1 summarizes the forecasting error distribution across two forecast horizons for each model–smoothing combination.

As expected, shorter-term forecasts generally result in lower absolute errors. However, the model's sensitivity to the choice of smoothing technique remains notable. In the 3-month horizon, the TFT achieves the lowest median RMSE and MAE, particularly when paired with STL or the rolling mean, indicating its strong short-term performance. However,

LightGBM outperforms the other models when considering the MAPE, exhibiting lower variability in the relative error. LSTM, while capable of a competitive median performance under STL, shows significant instability, especially with the other smoothing methods.

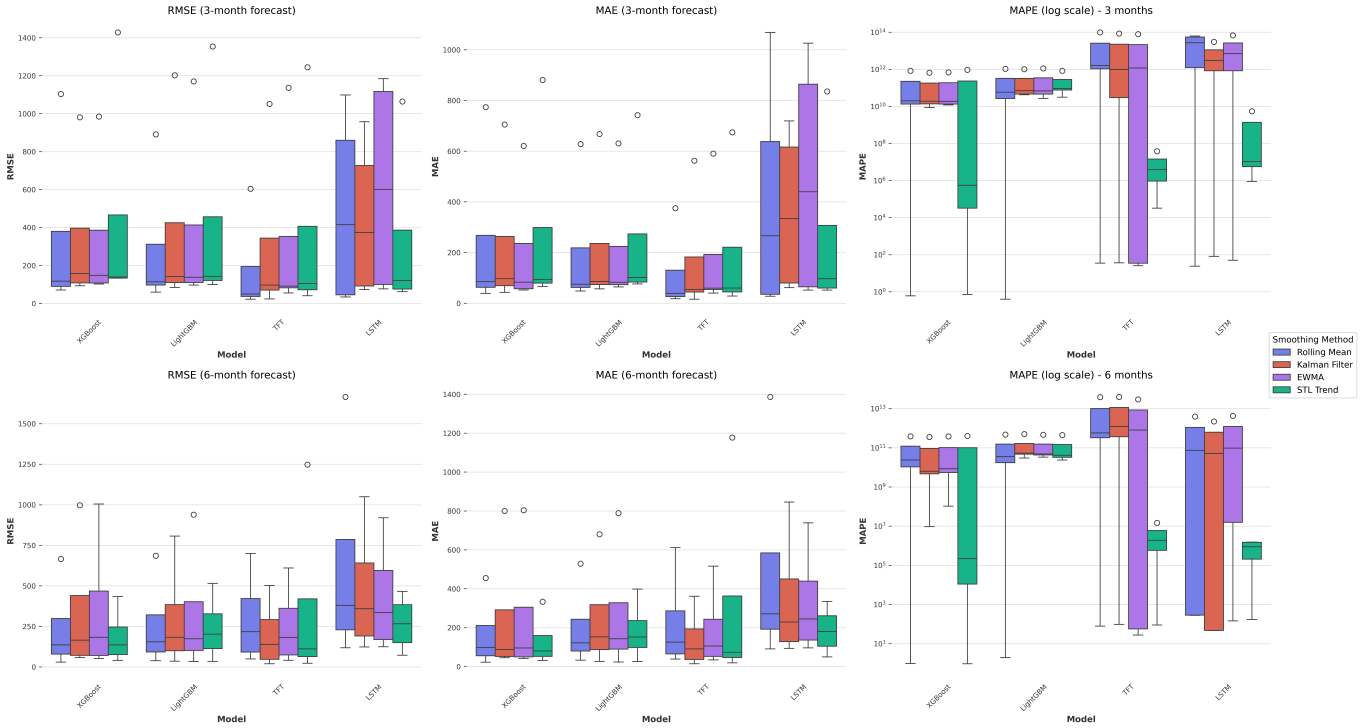


Figure 1. Boxplots for 3- and 6-month forecasting horizons.

For the 6-month horizon, XGBoost consistently outperforms the other models across all three metrics, with the lowest median RMSE, MAE, and MAPE. This suggests that tree-based models are more robust for longer-term forecasts. While neural models like the TFT and LSTM demonstrate potential, they are more prone to error spikes, particularly under less effective smoothing methods such as the Kalman filter or EWMA. The MAPE plotted on a logarithmic scale, due to its high variance, reveals substantial volatility for the neural models in both the 3-month and 6-month forecast periods when paired with the Kalman filter or EWMA smoothing, likely due to small denominators where accurate case counts approach zero. Notably, extraction of the trends using STL consistently reduces this instability, particularly for LSTM, underscoring its effectiveness in stabilizing the relative error across short- and medium-term forecasts.

Figures 2 and 3 present heatmaps of the RMSE, MAE, and MAPE for the 6-month and 3-month forecast horizons.

The patterns shown are supported by the summary of the best-performing combinations in Table 1, which lists the lowest values for each error metric across all settings.

As expected, 3-month forecasts generally yield lower RMSE and MAE values than those of 6-month forecasts, especially for Ukraine and Bulgaria. For instance, in Ukraine, the TFT with STL achieves an RMSE of 144.49 and an MAE of 91.01 for the 6-month horizon, but at 3 months, LSTM with the rolling mean performs better, with a 33.89 RMSE and a 27.6 MAE. In Bulgaria, the TFT paired with the rolling mean achieves an RMSE of 41.68 and an MAE of 28.97 at 3 months, compared to the lowest error for a 6-month forecast for the TFT with the Kalman filter, at a 56.12 RMSE and a 43.76 MAE. These reductions confirm that neural models benefit from shorter horizons, provided that appropriate smoothing is applied.



Figure 2. Heatmaps for 6-month forecasting.

However, the reduction in the error is not uniform. In Slovenia, the TFT with the Kalman filter achieves the lowest RMSE (19.81) and MAE (14.34) for the 6-month horizon, yet the performance for the 3-month horizon is not consistently better across all metrics. The TFT with the rolling mean achieves the best RMSE (22.33), and the TFT with the Kalman filter still yields the lowest MAE (15.94). This suggests that even when shorter-term forecasts are generally more accurate, some combinations at longer horizons can still outperform them depending on the error metric and smoothing method. In Greece, a different pattern emerges. For the 6-month horizon, XGBoost with STL outperforms the neural models in its RMSE (434.03) and MAE (333.09). For 3 months, although the TFT with the rolling mean achieves the lowest RMSE (604.24) and MAE (375.15), the overall errors remain higher than those for the other countries. This indicates that forecasting

challenges in Greece may stem more from the underlying variability in the data or reporting inconsistencies than from the choice of model or horizon.

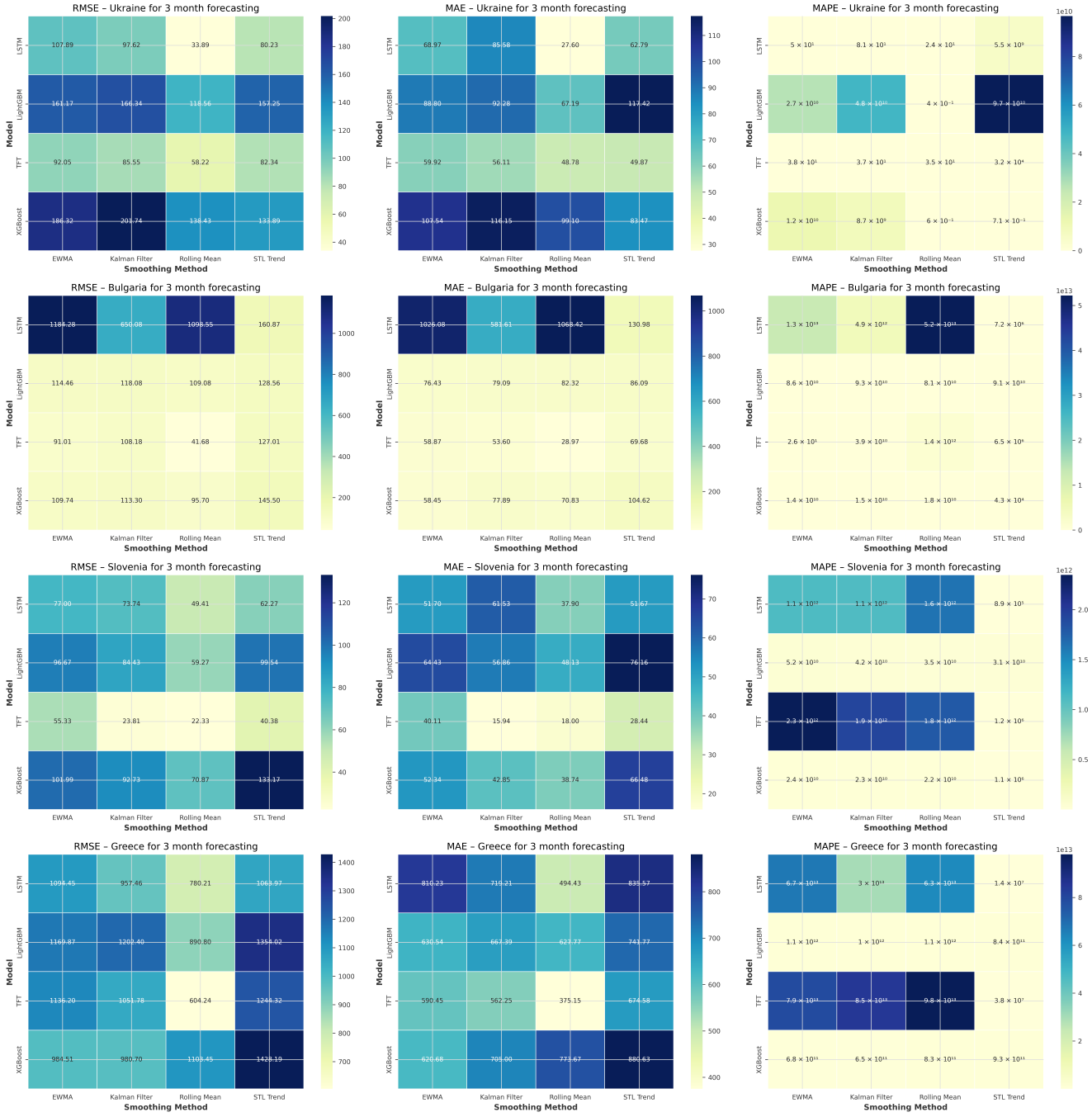


Figure 3. Heatmaps for 3-month forecasting.

The MAPE values vary considerably across countries and smoothing methods. For the 3-month horizon, LSTM + STL achieves the lowest MAPE in Slovenia (894,000) and Greece (13,700,000), and it also leads in Greece at the 6-month horizon (1,530,000), highlighting STL's stabilizing effect on neural models in terms of the relative error. However, in low-incidence countries like Slovenia and Greece, the MAPE can become extremely inflated due to division by near-zero actual case counts. In such cases, the RMSE and MAE offer more stable indicators of the performance and should be prioritized when interpreting the results. In contrast, Bulgaria sees its best MAPE value using the TFT with the EWMA at both horizons (27.3 at 6 months and 25.5 at 3 months), while in Ukraine, XGBoost with

STL and LightGBM with the rolling mean perform best for the 6- and 3-month forecasts, respectively. These results suggest that STL can effectively stabilize the relative error in neural models like LSTM, particularly for countries such as Slovenia and Greece. However, no single smoothing method consistently outperforms the others across all countries and models, reinforcing the need for context-specific optimization in epidemic forecasting.

Table 1. Best-performing method per country and forecast horizon.

Country	Forecast Period	Best Method (RMSE)	RMSE	Best Method (MAE)	MAE	Best Method (MAPE)	MAPE
Ukraine	6M	TFT + STL	144.49	TFT + STL	91.01	XGBoost + STL	0.93
	3M	LSTM + RM	33.89	LSTM + RM	27.6	LightGBM + RM	**0.396**
Bulgaria	6M	TFT + Kalman	56.12	TFT + Kalman	43.76	TFT + EWMA	27.3
	3M	TFT + RM	41.68	TFT + RM	28.97	TFT + EWMA	25.5
Slovenia	6M	TFT + Kalman	**19.81**	TFT + Kalman	**14.34**	LSTM + Kalman	46.6
	3M	TFT + RM	22.33	TFT + Kalman	15.94	LSTM + STL	894,000
Greece	6M	XGBoost + STL	434.03	XGBoost + STL	333.09	LSTM + STL	1,530,000
	3M	TFT + RM	604.24	TFT + RM	375.15	LSTM + STL	13,700,000

Note: Bolded values indicate the lowest metric values across countries.

Figures 4–11 visually compare the 6-month and 3-month forecast trajectories across countries and smoothing methods. These plots provide insight into how well the different models track the actual case dynamics, especially during rising or falling trends.

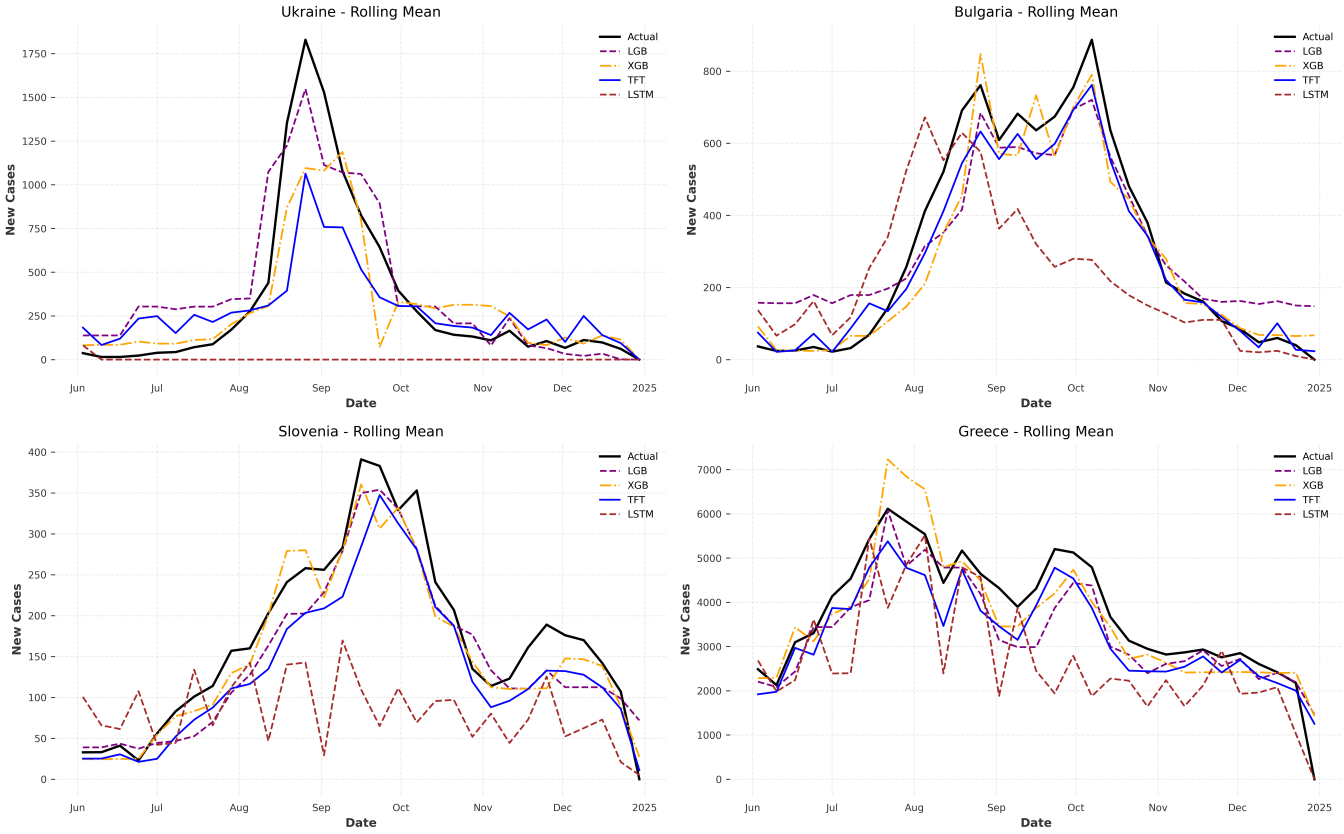


Figure 4. Forecast trajectories for 6-month forecasting and the rolling mean as the smoothing method.

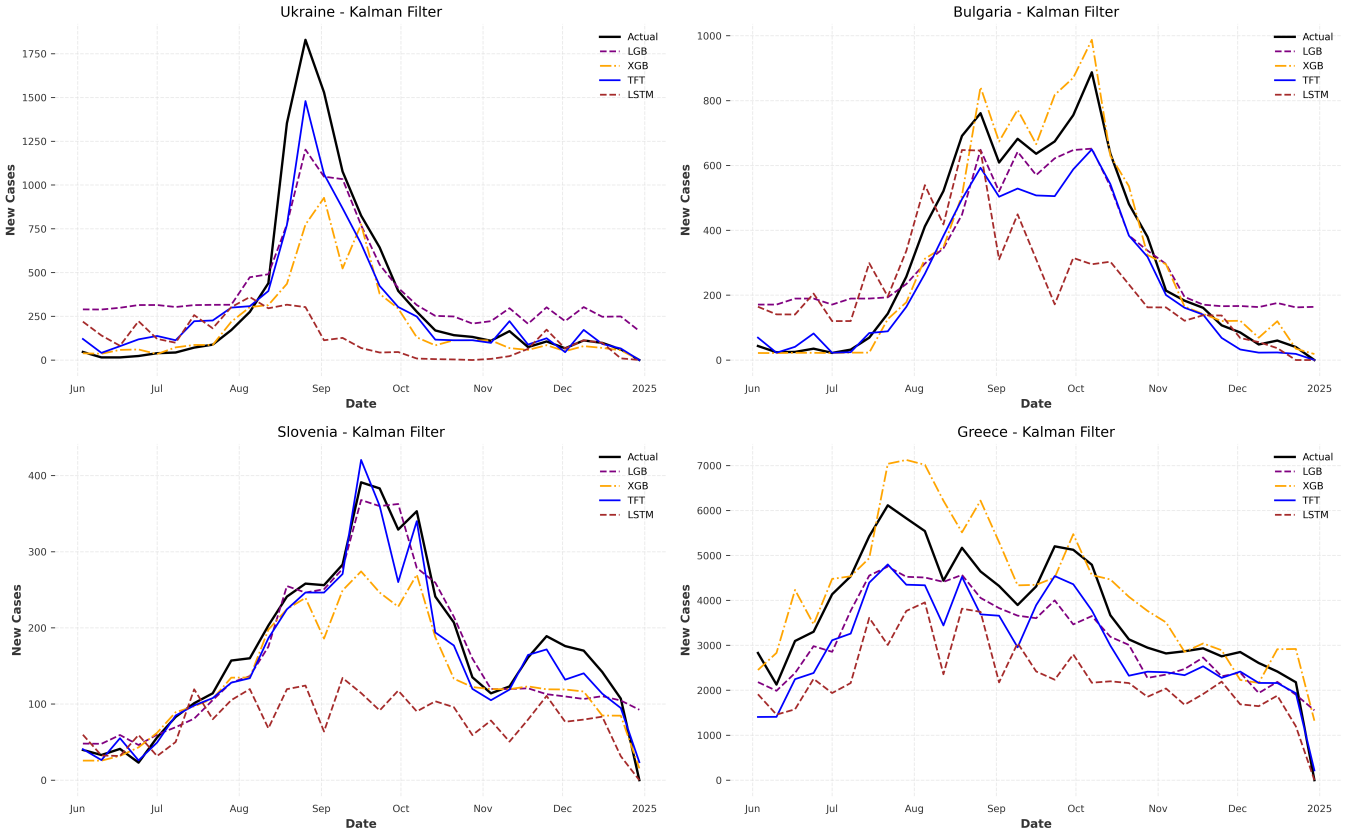


Figure 5. Forecast trajectories for 6-month forecasting and the Kalman filter as the smoothing method.

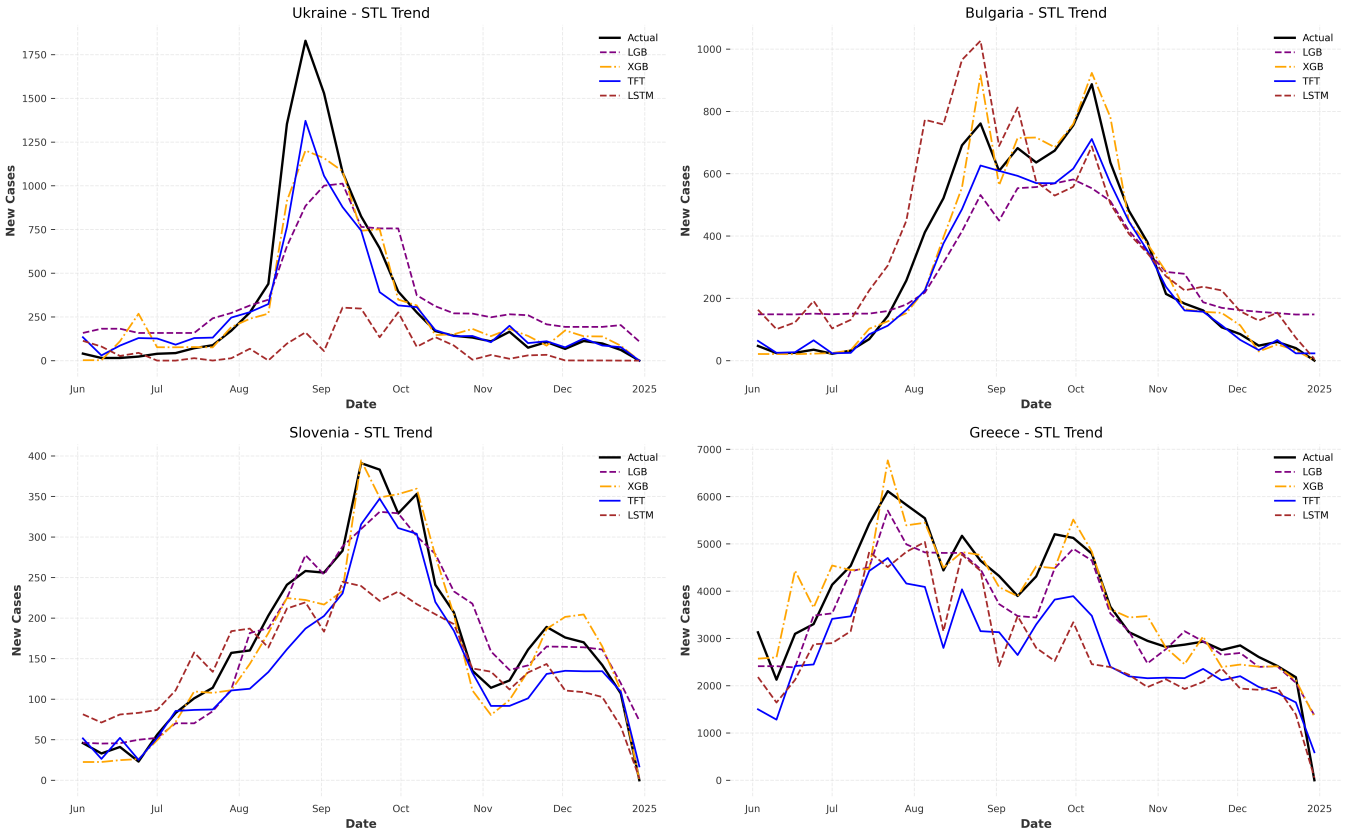


Figure 6. Forecast trajectories for 6-month forecasting and STL as the smoothing method.

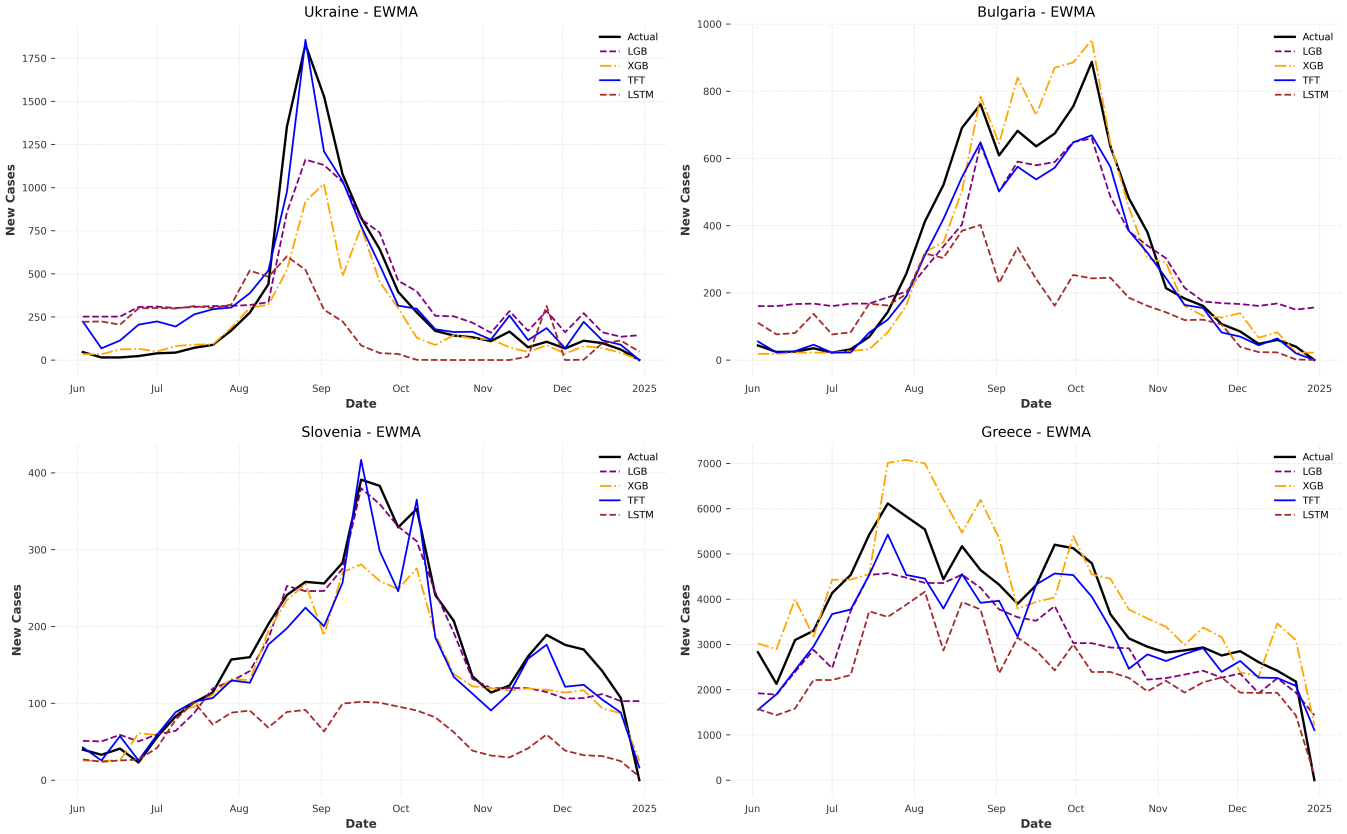


Figure 7. Forecast trajectories for 6-month forecasting and the EMWA as the smoothing method.

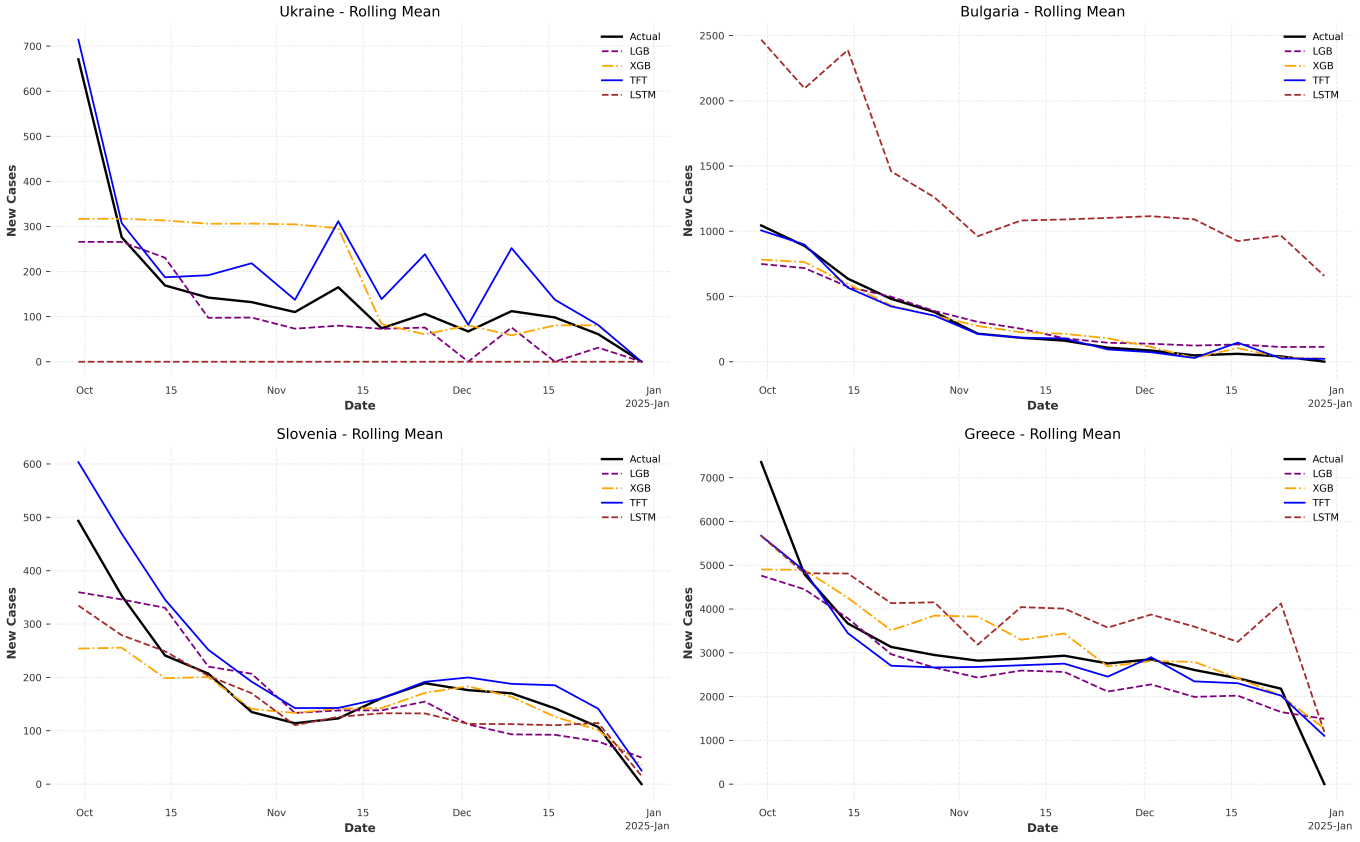


Figure 8. Forecast trajectories for 3-month forecasting and the rolling mean as the smoothing method.

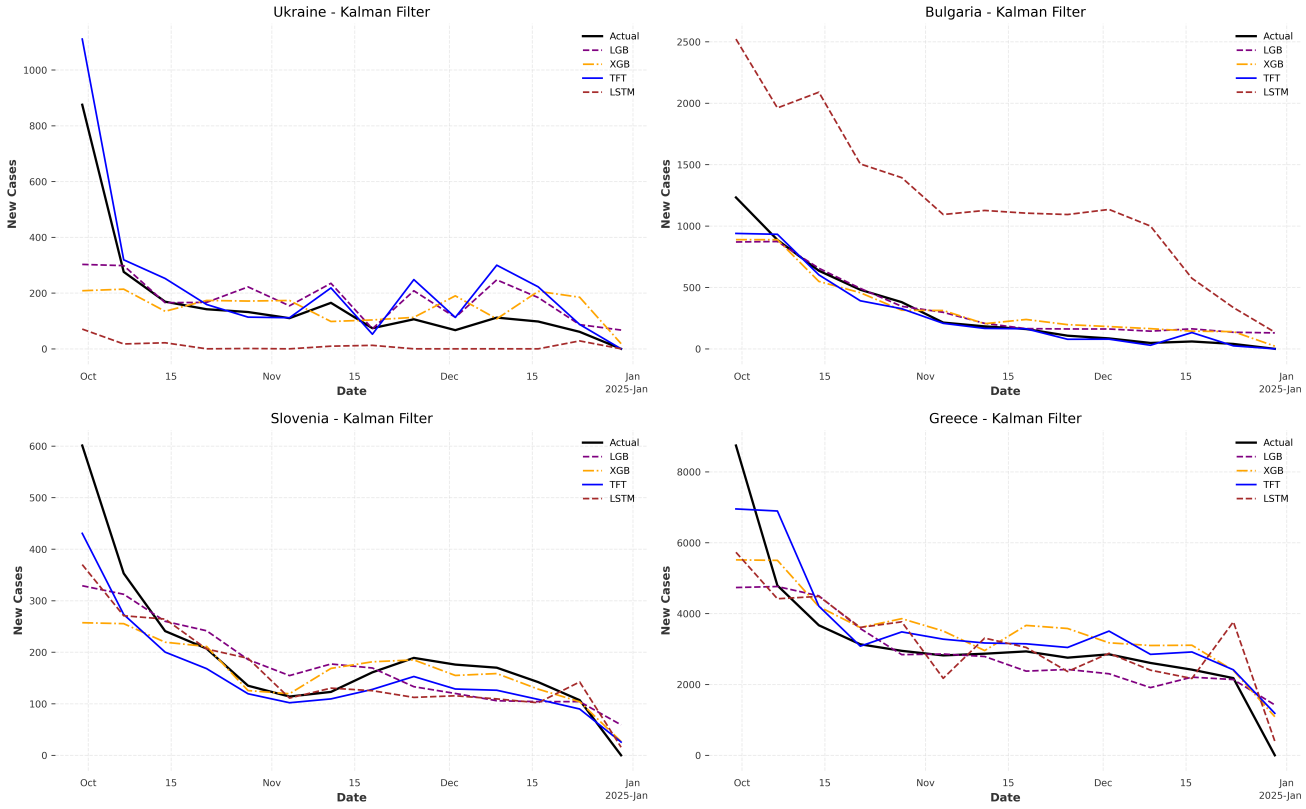


Figure 9. Forecast trajectories for 3-month forecasting and the Kalman filter as the smoothing method.

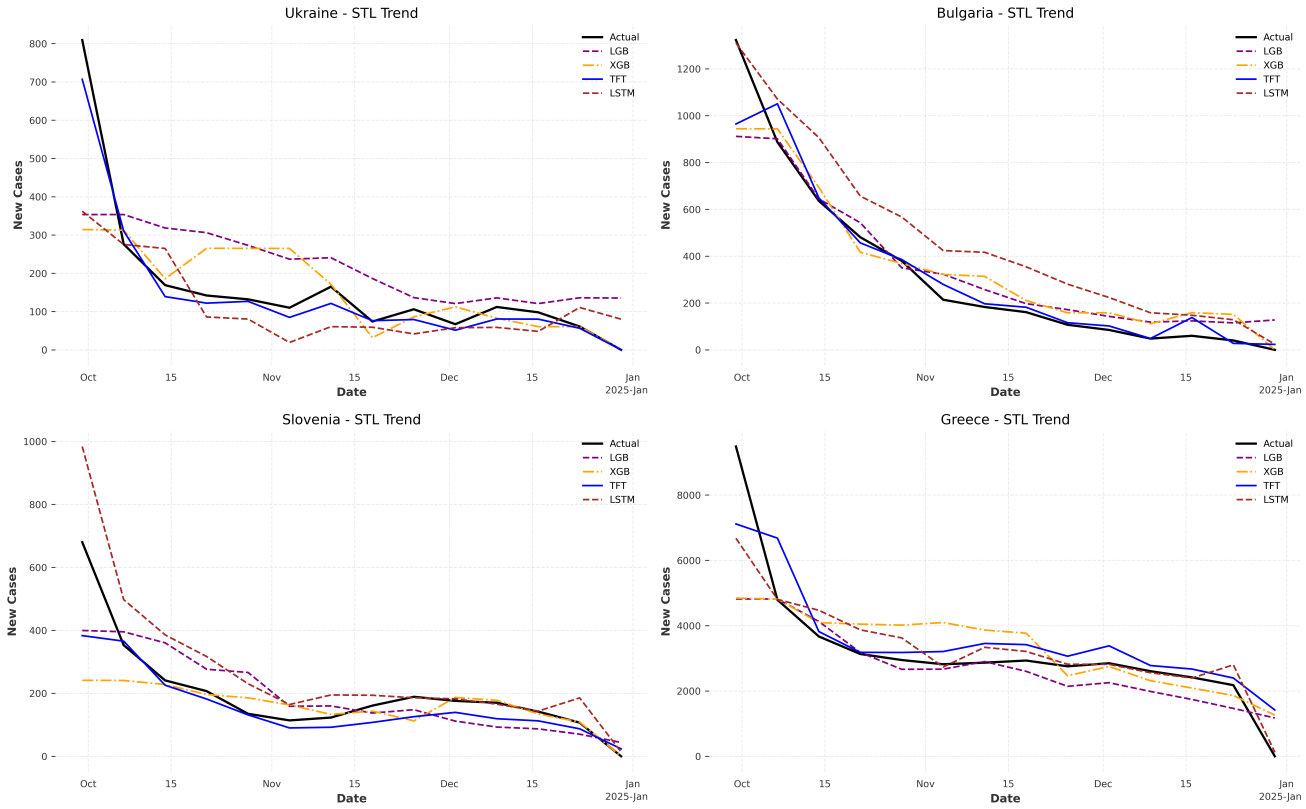


Figure 10. Forecast trajectories for 3-month forecasting and STL as the smoothing method.

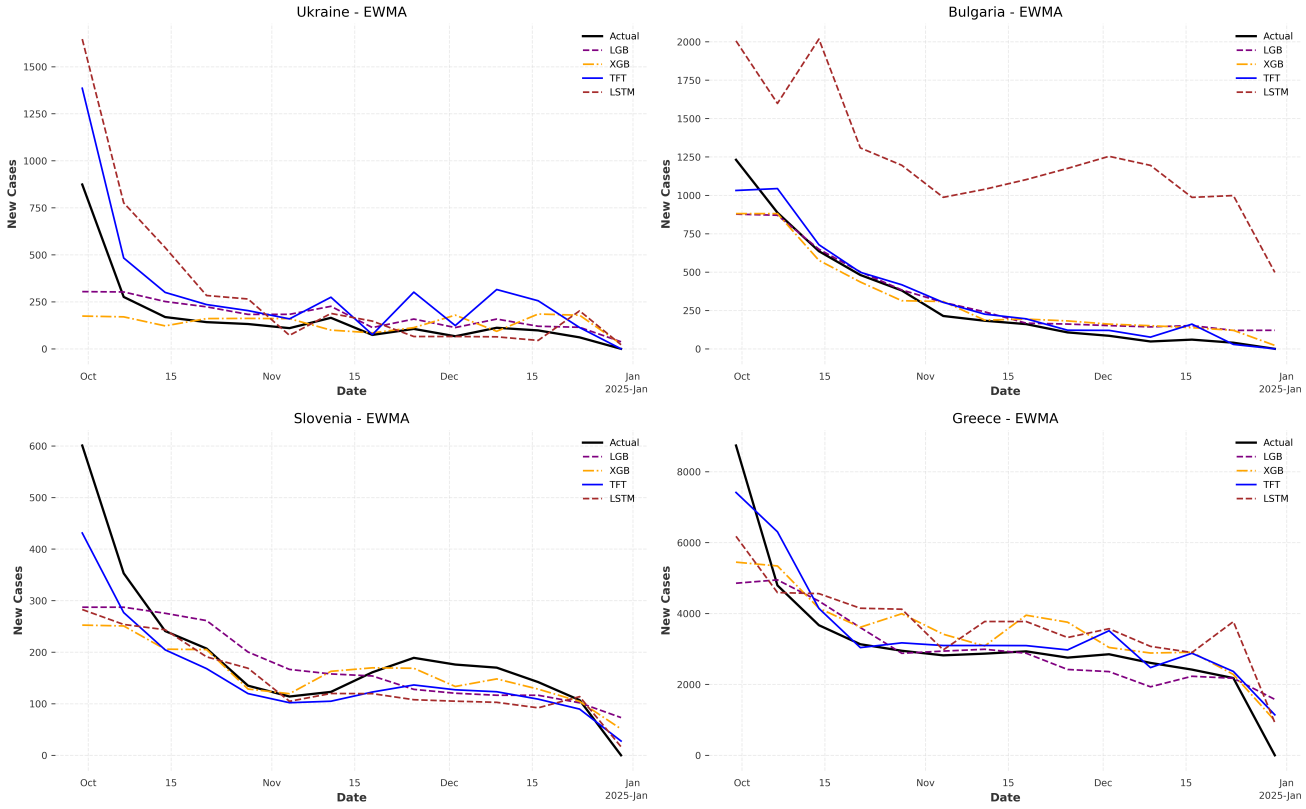


Figure 11. Forecast trajectories for 3-month forecasting and the EMWA as the smoothing method.

Consistent with the earlier findings, LSTM combined with STL produces the most visually accurate forecasts in Slovenia and Greece, closely tracking the actual trends for 3-month forecasting. In contrast, in Ukraine, LSTM with the Kalman filter or the EWMA for 6-month forecasts and LSTM with the rolling mean, the Kalman filter, or the EMWA for 3-month forecasts diverge significantly, where the predictions fluctuate or are flattened unrealistically.

In Bulgaria, the TFT with the EWMA, which shows the best MAPE, also closely follows the real case trajectory. In Ukraine, XGBoost with STL for 6-month forecasting and LightGBM with the rolling mean for 3-month forecasting align better with the actual values, confirming their strong MAPE results. Across most of the countries, STL smoothing improves the shape and stability of the forecasts for the neural models, while the tree-based models offer more consistent but sometimes overly smoothed predictions.

These findings are supported not only visually but also statistically. We conducted a two-way ANOVA on the MAPE, MAE, and RMSE values to validate the observed differences (Table 2).

Table 2. Two-way ANOVA: summary of effects on the MAPE, MAE, and RMSE.

Factor	F-Value for MAPE	*p*-Value for MAPE	F-Value for MAE	*p*-Value for MAE	F-Value for RMSE	*p*-Value for RMSE
Smoothing Method	1.49	0.222	0.095	0.96	0.082	0.97
Model Name	4.13	**0.008**	1.98	0.12	1.066	0.37

The MAPE was the primary focus due to its scale independence and higher variability across conditions. As shown in Table 2, model architecture had a statistically significant effect on the MAPE ($F = 4.13$, $p = 0.008$), while its impact on the MAE and RMSE was in-

significant. The smoothing methods showed no significant effect on any metric, reinforcing that the choice of model is the primary driver of accuracy.

Table 3 lists the best-performing model and smoothing combinations by country and forecast horizon to summarize the dominant configurations observed across all evaluations. These combinations were selected based on the lowest error metrics and were consistent with the abovementioned patterns, such as LSTM with the rolling mean in Ukraine for short-term forecasts and the TFT with the Kalman filter in Bulgaria for longer horizons. In most cases, the strong showing of the neural models further reinforces the statistical findings.

Table 3. Summary of best model–smoothing combinations.

Country	Horizon	Best Model	Best Smoothing
Ukraine	3M	LSTM	Rolling mean
Ukraine	6M	TFT	STL
Bulgaria	3M	TFT	Rolling mean
Bulgaria	6M	TFT	Kalman filter
Slovenia	3M	TFT	None
Slovenia	6M	TFT	Kalman filter
Greece	3M	TFT	Rolling mean
Greece	6M	XGBoost	STL

Beyond identifying the optimal model–smoothing pairs (Table 3), several deeper insights emerged. First, neural models like LSTM and the TFT showed greater sensitivity to volatility in the data, especially for longer horizons, reflecting the patterns seen in epidemic forecasting studies where the performance drops during phase shifts. Second, although our analysis evaluates the models independently, prior research suggests that combining predictions from multiple models (i.e., using ensemble methods) can reduce the variance in forecasts and improve robustness.

Third, smoothing methods like the rolling mean and STL improve the performance for neural models but also help stabilize the predictions from tree-based models, particularly in countries with high case variability. These elements highlight the interplay between the model type, smoother, and forecast horizon in shaping accuracy and reliability.

5. Discussion

Our analysis shows that smoothing is helpful when matched to both the learning algorithm and the forecast horizon. A two-way ANOVA confirmed that the model architecture explained the bulk of the variation in the MAPE ($F = 4.13$, $p = 0.008$), whereas the main effect of smoothing was not significant. Even so, the descriptive results reveal clear interaction patterns.

For the 3-month forecasts, LSTM and the TFT fed with either a centered seven-point rolling mean or the trend component from STL reduced the RMSE by more than 60% in three of the four countries studied. Comparable behavior was reported in [40], which used an STL-SARIMA hybrid to track the daily incidence in Jakarta and attained a MAPE of 0.15. This improvement stems from the removal of the dominant weekly cycle, which otherwise drives gradient instability during backpropagation. Similar advantages for smoothed inputs were observed in a study using multi-step LSTM in Indian states [30].

Kalman smoothing lowered both the RMSE and MAE for the TFT in Bulgaria and Slovenia, echoing the results from a study tracking the volatility in daily case rates [16]. In Greece, however, the same filter suppressed meaningful peaks and inflated the relative error, an outcome previously noted when the variance in the observations was misspecified in epidemic state-space models [40]. These mixed results underline that adaptive filters demand careful tuning to the local reporting noise.

When the horizon widened to 6 months, gradient boosting trees became the most reliable option. XGBoost achieved the lowest median RMSE in half of the long-range runs and showed low sensitivity to the smoother. In [28], a similar conclusion was reached for five Saudi cities: XGBoost offered faster training and more stable long-horizon errors than LightGBM, even though both models shared most engineered features.

Because different model–smoother pairs excel in different settings, an ensemble often outperforms any single contributor. The U.S. Forecast Hub showed that a simple median-weighted ensemble was the most accurate forecaster of mortality for 18 months of the pandemic [41], and similar gains were documented for European case forecasts [14]. Although the present work evaluated the models independently, the diversity of the error profiles indicates that combining neural and tree predictions would further reduce the variance.

This study has several limitations. First, only four European countries were analyzed; extending this experiment to settings with sparser surveillance may change the balance between the cost and benefit of smoothing. Second, the smoothing parameters were fixed across countries to isolate the relative effect of each method. However, future work could explore country-specific tuning to enhance the accuracy further and better reflect national reporting characteristics. Third, weekly aggregation improved the robustness by removing weekday noise and administrative artifacts. Data were used as reported by the WHO. Weekly aggregation minimized the fluctuations, though missing data remain a limitation. However, it may have obscured short-term surges or rapid outbreak dynamics. Exploring daily or biweekly aggregation could capture such patterns better. Operational dashboards that run on daily data should pair light smoothing with anomaly detectors. Additionally, future work may apply mixed-effects models to capture interaction terms and country-specific effects better. Temporal cross-validation or ensemble averaging could help reduce the variance and is worth exploring in follow-up work. Real-time deployment requires balancing the accuracy and computational cost. TFT training requires GPU support, which may limit its accessibility for some agencies. Training the TFT and LSTM was 3–4 times slower than this process for XGBoost or LightGBM, making tree-based models more accessible in resource-limited contexts. Advanced architectures like N-BEATS or diffusion models may offer improvements and are planned for future research. Robust STL and outlier detection techniques may prevent the peak suppression seen with Kalman filtering. This will be tested in subsequent experiments.

For horizons up to 12 weeks, LSTM or the TFT preceded by STL or a short rolling mean would likely give the lowest point error. For 13–26 weeks, a tree ensemble offers greater robustness with a modest computing cost. Both smoothing steps run in linear time, so they add little overhead to real-time pipelines. Updating the model–smoother pairing as new data accrue could help public health agencies keep the predictions reliable under changing surveillance conditions. A real-time system may integrate anomaly detection, residual monitoring, and scheduled retraining to maintain accuracy during regime shifts.

6. Conclusions

This study assessed the effect of smoothing techniques and the model selection on forecasting the COVID-19 case counts across multiple countries and two forecast horizons. Our results indicate that shorter-term forecasts (3 months) consistently reduce the absolute error metrics (the RMSE and MAE), especially when neural models like LSTM and the TFT use smoothing methods such as STL or the rolling mean.

The tree-based models (XGBoost and LightGBM) demonstrated a more robust performance for longer-term forecasting, particularly for countries with more irregular data patterns. Among the smoothing methods, STL was most effective at stabilizing the MAPE

across the models, with statistical tests confirming that model choice significantly impacted the relative accuracy of the forecasts ($F = 4.13$, $p = 0.008$).

While no single combination performed best across all countries and metrics, these results emphasize the importance of aligning the forecast horizon, data volatility, and model architecture with an appropriate smoother. These insights can support the development of more reliable epidemic forecasting pipelines for diverse national contexts.

The proposed forecasting pipeline, combining smoothing techniques with robust machine learning models, can be adapted for real-time deployment by public health agencies. By regularly updating the input data and leveraging model–smoother configurations tailored to the epidemiological profile of each country, health authorities can improve their short- and medium-term outbreak predictions. This, in turn, can support early intervention planning, healthcare resource allocation, and the timely implementation of containment measures during future pandemic waves or similar public health emergencies.

Future work may explore using adaptive training schemes or ensemble techniques to improve the generalizability and real-time applicability. Additionally, incorporating anomaly detection or reporting correction steps before smoothing may help mitigate the risk of obscuring critical short-term shifts in the case dynamics, especially in countries with irregular reporting practices.

Author Contributions: U.Z. and D.C. contributed equally to the conceptualization and writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the National Research Foundation of Ukraine in the framework of the research project 2023.03/0197 on the topic "Multidisciplinary study of the impact of emergency situations on the infectious diseases spreading to support management decision making in the field of population biosafety".

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of Variance
ARIMA	AutoRegressive Integrated Moving Average
ARIMAX	AutoRegressive Integrated Moving Average with Exogenous Variables
CatBoost	Categorical Boosting
CNN	Convolutional Neural Network
COVID-19	Coronavirus Disease 2019
EWMA	Exponentially Weighted Moving Average
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
SARIMA	Seasonal AutoRegressive Integrated Moving Average
STL	Seasonal–Trend Decomposition using Loess
TFT	Temporal Fusion Transformer
VARMA	Vector AutoRegressive Moving Average
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting

References

1. World Health Organization. *COVID-19 Epidemiological Update—Edition 177*; World Health Organization: Geneva, Switzerland, 2024. Available online: https://www.who.int/publications/m/item/covid-19-epidemiological-update-edition-177 (accessed on 2 March 2025).
2. Babanejaddehaki, G.; An, A.; Papagelis, M. Disease outbreak detection and forecasting: A review of methods and data sources. *ACM Trans. Comput. Healthc.* **2025**, *6*, 1–40. [CrossRef]
3. Chhabra, A.; Singh, S.K.; Sharma, A.; Kumar, S.; Gupta, B.B.; Arya, V.; Chui, K.T. Sustainable and intelligent time-series models for epidemic disease forecasting and analysis. *Sustain. Technol. Entrep.* **2024**, *3*, 100064. [CrossRef]
4. Zahrani, A.; Rohmawati, A.A.; Sa'adah, S. STL decomposition and SARIMA model: The case for estimating value-at-risk of COVID-19 increment rate in DKI Jakarta. *Int. J. Inf. Commun. Technol. (IJoICT)* **2021**, *7*, 1–10. [CrossRef]
5. Ghani, M.; Rantini, D.; Maryamah. Kalman filter based on a fractional discrete-time stochastic augmented COVID-19 model. *J. Biosaf. Biosecur.* **2024**, *6*, 76–98. [CrossRef]
6. Jiao, S.; Wang, Y.; Ye, X.; Nagahara, L.; Sakurai, T. Spatio-temporal epidemic forecasting using mobility data with LSTM networks and attention mechanism. *Sci. Rep.* **2025**, *15*, 9603. [CrossRef]
7. Jenko, J.; Costa, J.P.; Vladušič, D.; Bavčar, U.; Čabarkapa, R. Learning from the COVID-19 pandemic to improve critical infrastructure resilience using temporal fusion transformers. In Proceedings of the Computer Science and Information Systems (FedCSIS), Belgrade, Serbia, 8–11 September 2024. [CrossRef]
8. Lotysh, V.; Gumeniuk, L.; Humeniuk, P. Comparison of the effectiveness of time series analysis methods: SMA, WMA, EMA, EWMA, and Kalman filter for data analysis. *Inform. Autom. Pomiary Gospod. Ochr. Środowiska* **2023**, *13*, 71–74. [CrossRef]
9. Kotu, V.; Deshpande, B. Time series forecasting. In *Data Science*, 2nd ed.; Kotu, V., Deshpande, B., Eds.; Morgan Kaufmann: Cambridge, MA, USA, 2019; pp. 395–445. [CrossRef]
10. Ouyang, Z.; Ravier, P.; Jabloun, M. STL decomposition of time series can benefit forecasting done by statistical methods but not by machine learning ones. *Eng. Proc.* **2021**, *5*, 42. [CrossRef]
11. Khikmah, K.N.; Sofro, A. Autoregressive moving average and generalized autoregressive moving average in COVID-19 confirmed cases in Indonesia. *arXiv* **2022**, arXiv:2202.11794.
12. Abidi, F.A.; Radiy, Z.H. Using exponential smoothing methods to analysis COVID-19 time series. In Proceedings of the 2nd International Conference on Engineering and Science to Achieve the Sustainable Development Goals, AIP Conference Proceedings, Tabriz, Iran, 9–10 July 2023; Volume 3092, p. 070014. [CrossRef]
13. Waqas, M.; Xu, S.H.; Anwar, S.M.; Rasheed, Z.; Shabbir, J. The optimal control chart selection for monitoring COVID-19 phases: A case study of daily deaths in the USA. *Int. J. Qual. Health Care* **2023**, *35*, mzad058. [CrossRef]
14. Sherratt, K.; Gruson, H.; Grah, R.; Johnson, H.; Niehus, R.; Prasse, B.; Sandmann, F.; Deuschel, J.; Wolffram, D.; Abbott, S.; et al. Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. *eLife* **2023**, *12*, e81916. [CrossRef]
15. Hsu, C.-R.; Wang, H. EWMA control chart integrated with time series models for COVID-19 surveillance. *Mathematics* **2025**, *13*, 115. [CrossRef]
16. Masum Bhuiyan, M.A.; Mahmud, S.; Islam, M.R.; Tasnim, N. Volatility estimation for COVID-19 daily rates using Kalman filtering technique. *Results Phys.* **2021**, *26*, 104291. [CrossRef] [PubMed]
17. Singh, K.K.; Kumar, S.; Dixit, P.; Bajpai, M.K. Kalman filter based short term prediction model for COVID-19 spread. *Appl. Intell.* **2021**, *51*, 2714–2726. [CrossRef] [PubMed]
18. Rao, V.; Devi, B.; Pratapagiri, S.; Srinivas, C.; Venkatramulu, S.; Raghavakumari, D. Prediction of COVID-19 using Kalman filter algorithm. *Proc. AIP Conf. Proc.* **2022**, *2418*, 030067. [CrossRef]
19. Nanda, S.K.; Kumar, G.; Bhatia, V.; Singh, A.K. Kalman-based compartmental estimation for COVID-19 pandemic using advanced epidemic model. *Biomed. Signal Process. Control* **2023**, *84*, 104727. [CrossRef]
20. Abeln, B.; Jacobs, J.P.A.M. COVID-19 and seasonal adjustment. *J. Bus. Cycle Res.* **2022**, *18*, 159–169. [CrossRef]
21. Singh, R.K.; Rani, M.; Bhagavathula, A.S.; Sah, R.; Rodriguez-Morales, A.J.; Kalita, H.; Nanda, C.; Sharma, S.; Sharma, Y.D.; Rabaan, A.A.; et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. *JMIR Public Health Surveill.* **2020**, *6*, e19115. [CrossRef]
22. Claris, S.; Peter, N. ARIMA model in predicting of COVID-19 epidemic for the Southern Africa region. *Afr. J. Infect. Dis.* **2022**, *17*, 1–9. [CrossRef]
23. Somyanonthanakul, R.; Warin, K.; Amasiri, W.; Mairiang, K.; Mingmalairak, C.; Panichkitkosolkul, W.; Silanun, K.; Theeramunkong, T.; Nitikraipot, S.; Suebnukarn, S. Forecasting COVID-19 cases using time series modeling and association rule mining. *BMC Med. Res. Methodol.* **2022**, *22*, 281. [CrossRef]
24. Jin, Y.; Wang, R.; Zhuang, X.; Wang, K.; Wang, H.; Wang, C.; Wang, X. Prediction of COVID-19 data using an ARIMA-LSTM hybrid forecast model. *Mathematics* **2022**, *10*, 4001. [CrossRef]

25. Ahn, J.M.; Kim, J.; Kim, K. Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting. *Toxins* **2023**, *15*, 608. [CrossRef] [PubMed]
26. Zheng, H.-L.; An, S.-Y.; Qiao, B.-J.; Guan, P.; Huang, D.-S.; Wu, W. A data-driven interpretable ensemble framework based on tree models for forecasting the occurrence of COVID-19 in the USA. *Environ. Sci. Pollut. Res.* **2023**, *30*, 13648–13659. [CrossRef] [PubMed]
27. Dharayani, R.; Hasmawati, H.; Khotijah, S. COVID-19 prediction using LightGBM and LSTM. *J. Sains Komput. Inform.* **2022**, *6*, 838–847. [CrossRef]
28. Sadig, H.E.; Kamal, M.; Rehman, M.U.; Habadi, M.I.; Alnagar, D.K.; Yusuf, M.; Mohammed, M.O.M.; Alqasem, O.A.; Meraou, M.A. Advanced time complexity analysis for real-time COVID-19 prediction in Saudi Arabia using LightGBM and XGBoost. *J. Radiat. Res. Appl. Sci.* **2025**, *18*, 101364. [CrossRef]
29. Luo, J.; Zhang, Z.; Fu, Y.; Rao, F. Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. *Results Phys.* **2021**, *27*, 104462. [CrossRef]
30. Chandra, R.; Jain, A.; Singh Chauhan, D. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLoS ONE* **2022**, *17*, e0262708. [CrossRef]
31. Hu, B.; Han, Y.; Zhang, W.; Zhang, Q.; Gu, W.; Bi, J.; Chen, B.; Xiao, L. A prediction approach to COVID-19 time series with LSTM integrated attention mechanism and transfer learning. *BMC Med. Res. Methodol.* **2024**, *24*, 323. [CrossRef]
32. Lim, B.; Arık, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]
33. Nalmpatian, A.; Heumann, C.; Pilz, S. Forecasting mortality trends: Advanced techniques and the impact of COVID-19. *Stats* **2024**, *7*, 1172–1188. [CrossRef]
34. Elimam, R.; Sutton-Charani, N.; Perrey, S.; Montmain, J. Uncertain imputation for time-series forecasting: Application to COVID-19 daily mortality prediction. *PLoS Digit. Health* **2022**, *1*, e0000115. [CrossRef]
35. Sharif, O.; Hasan, M.Z.; Rahman, A. Determining an effective short term COVID-19 prediction model in ASEAN countries. *Sci. Rep.* **2022**, *12*, 5083. [CrossRef] [PubMed]
36. Thammawijaya, P.; Suphanchaimat, R. Accuracy of COVID-19 prediction modeling techniques. *Outbreak Surveill. Investig. Response J.* **2024**, *17*, 146–154. [CrossRef]
37. Bahuguna, A.; Uniyal, A.; Sharma, N.; Semwal, J. Comparison of exponential smoothing and ARIMA time series models for forecasting COVID-19 cases: A secondary data analysis. *Int. J. Res. Med. Sci.* **2023**, *11*, 1727–1734. [CrossRef]
38. Tuli, S.; Tuli, S.; Tuli, R.; Gill, S.S. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet Things* **2020**, *11*, 100222. [CrossRef]
39. Zhao, Y.; Wong, L.; Goh, W.W.B. How to do quantile normalization correctly for gene expression data analyses. *Sci. Rep.* **2020**, *10*, 15534. [CrossRef]
40. Satrisna, R.D.; Rohmawati, A.A.; Sa'adah, S. Forecasting the COVID-19 increment rate in DKI Jakarta using non-robust STL decomposition and SARIMA model. *Int. J. Inf. Commun. Technol.* **2021**, *7*, 21–30. [CrossRef]
41. 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.; et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2113561119. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.