Performance, Pitfalls, and a New Paradigm for Pandemic Preparedness:

A Global Analysis of COVID-19 Forecasting Models : AI Integration and Finetuned

By: Rishu Ranjan Raj

IIT KANPUR, M-TECH DoMS

241140027

COLLAB LINK: https://colab.research.google.com/drive/15THGRGM40I0RdfH_Sbmy8QJOkss-B-57?usp=drive_link

GOOGLE DRIVE LINK: FOR DELIVERABLES

https://drive.google.com/file/d/1atjatN24ao6h-vz6oukKhV0AiFHb59RG/view?usp=drive_link

Executive Summary

The COVID-19 pandemic placed an unprecedented demand on epidemiological forecasting, elevating predictive models from academic tools to critical instruments of public policy and resource allocation. This report provides a comprehensive analysis of the diverse array of forecasting models deployed across various World Health Organization (WHO) regions. It evaluates their performance, dissects the complex reasons for observed disparities, and culminates in a set of strategic recommendations for strengthening future pandemic preparedness.

The analysis reveals that no single forecasting model or methodological class—be it mechanistic, statistical, or machine learning-based—proved universally superior. Performance was profoundly context-dependent, fluctuating based on the specific WHO region, the phase of the epidemic, the quality of available data, and the target variable being predicted (e.g., cases versus deaths). Mechanistic models, while useful for early scenario planning, were often based on unrealistic assumptions of population homogeneity. Statistical models like ARIMA struggled with the non-linear shocks caused by interventions and new variants. Machine learning models showed promise in capturing complex dynamics but were hampered by initial data scarcity and a lack of interpretability.

A central finding is the consistent outperformance of ensemble models, which

aggregate predictions from multiple individual models. This suggests that acknowledging and synthesizing diverse perspectives is a more robust strategy than searching for a single, perfect forecasting tool. However, even the best models often exhibited poor calibration, displaying overconfidence by providing overly narrow prediction intervals. This was a critical failure, as communicating uncertainty is as important as the point forecast itself for sound decision-making.

Performance varied significantly by region. In the Americas, many sophisticated models failed to outperform simple baselines, and forecast errors were linked to social determinants of health. In Europe, human "crowd" forecasts proved superior for predicting cases, highlighting the value of integrating qualitative information, while computational models excelled at predicting deaths. In South-East Asia, models accounting for seasonality (SARIMA) were particularly effective. In Africa, the catastrophic failure of uncontextualized models from the global north underscored the absolute necessity of local data and expertise.

The core determinants of forecast accuracy were identified as: the quality and biases of underlying data; the dynamic nature of the pandemic, with shifting transmission rates and viral evolution; the impact of human behavior and public health interventions, which actively work to invalidate model assumptions; and the unique socio-ecological context of each region.

Based on this exhaustive analysis, this report puts forth a new paradigm for the role of forecasting in public health. The ultimate goal should not be the pursuit of a flawless predictive model, but rather the development of a **resilient and adaptive decision-making system** that can effectively utilize imperfect, uncertain forecasts. Key recommendations include:

1. **Adopting an Adaptive Policy Framework:** Shifting from rigid plans to flexible strategies that use forecasts for scenario analysis and are continuously updated with real-time data.
2. **Implementing Forecast-Driven Resource Management:** Using predictive models to proactively manage medical supply chains and strategic stockpiles, anticipating surges in demand to prevent shortages.
3. **Integrating Forecasts into Ethical Allocation Frameworks:** Using forecasts to inform pre-established, transparent frameworks for the fair allocation of scarce resources *before* a crisis peaks.
4. **Strengthening the Global Data and Modeling Infrastructure:** Investing in equitable, high-resolution data collection; building local modeling capacity in all

regions; standardizing evaluation protocols to include probabilistic scores like the Weighted Interval Score (WIS); and fostering collaborative modeling platforms.

The COVID-19 pandemic was a stress test that revealed critical vulnerabilities in our global health security architecture. By learning from the successes and failures of our forecasting efforts and embedding these lessons into international frameworks like the WHO Pandemic Agreement, the global community can build a more resilient, responsive, and equitable system to face the pandemics of the future.

1. A Taxonomy of COVID-19 Forecasting Models: From Mechanistic to Machine Learning

The global response to the COVID-19 pandemic spurred an unprecedented proliferation of epidemiological forecasting models. These tools, ranging from classical differential equation models to sophisticated artificial intelligence algorithms, became central to public discourse and policy-making.[1] Understanding their underlying principles, core assumptions, and inherent limitations is essential for interpreting their performance and appreciating the challenges of predicting a novel pathogen's trajectory. This section provides a taxonomy of the primary model archetypes used during the pandemic, establishing a conceptual framework for the subsequent analysis of their real-world performance.

1.1 Mechanistic Compartmental Models: The Epidemiological Foundation

Mechanistic models form the traditional bedrock of epidemiological modeling, seeking to capture the fundamental mechanisms of disease transmission within a population.[2] These models are based on systems of differential equations that divide the population into distinct compartments representing different stages of the disease.

The most basic is the **Susceptible-Infected-Recovered (SIR)** model, which categorizes individuals into one of three states.[3] A more refined and widely used variant during the pandemic was the

Susceptible-Exposed-Infected-Recovered (SEIR) model.[4] The addition of the "Exposed" compartment accounts for the latent period of the disease, during which an individual is infected but not yet infectious—a key feature of SARS-CoV-2.[5] These models were further extended to include additional states relevant to COVID-19, such as Quarantined (Q), Deceased (D), and Insusceptible (P), leading to more complex structures like the SEIQRPD model.[6] The primary output of these models is often the estimation of the reproduction number (

R), a critical parameter for public health policy that measures the average number of secondary infections caused by a single infected individual.[1]

Despite their conceptual elegance, these models are built on significant simplifying assumptions. Their most critical and frequently criticized assumption is that of **homogeneous mixing**, which posits that all individuals in a population are equally susceptible and have an equal probability of coming into contact with one another.[8] This assumption, known to be unrealistic even at the time of the models' creation, ignores the vast heterogeneity in social contact rates, population density, and individual susceptibility.[2] This simplification led some early, influential models to predict catastrophic outcomes, such as over 80% of a population being infected in an unmitigated first wave, a figure that did not materialize in most countries.[8]

Furthermore, mechanistic models face challenges with parameter estimation. Key variables, such as the proportion of asymptomatic but infectious carriers, are often not empirically observable and must be estimated through statistical procedures based on observed data like case counts or deaths.[2] Another major limitation is the common use of fixed, time-dependent transmission rates. This approach fails to capture the dynamic reality of a pandemic, where transmission rates change rapidly in response to non-pharmaceutical interventions (NPIs), voluntary changes in public behavior, or the emergence of new viral variants.[2]

1.2 Statistical Time-Series Models: Capturing Trends and Seasonality

In contrast to mechanistic models that focus on underlying disease dynamics, statistical time-series models are data-driven, aiming to identify and extrapolate patterns such as trends and seasonality from historical data.[10] The most prominent family of such models used for COVID-19 forecasting is the Autoregressive Integrated

Moving Average (ARIMA) class.

ARIMA models, also known as the Box-Jenkins method, are composed of three key components [10]:

- **Autoregressive (AR):** This component assumes that the current value in the series can be predicted based on a linear combination of its own past values, or lags. The parameter 'p' denotes the number of lag observations included.[12]
- **Integrated (I):** This component addresses the model's core requirement for the data to be stationary—meaning its statistical properties like mean and variance are constant over time. If the data exhibits trends or other forms of non-stationarity, it is transformed through differencing (subtracting a past value from the current value). The parameter 'd' represents the number of times the data is differenced.[14]
- **Moving Average (MA):** This component models the error term as a linear combination of past forecast errors. It helps account for random shocks in the time series. The parameter 'q' denotes the size of the moving average window.[12]

A critical extension of this model is the **Seasonal ARIMA (SARIMA)**. This model incorporates additional seasonal components, denoted as (P,D,Q)m, to explicitly model repeating patterns or cycles that occur at regular intervals (e.g., weekly or yearly).[16] Given that COVID-19 data often exhibited weekly reporting cycles, SARIMA models proved particularly useful.[11]

The primary strength of ARIMA-family models lies in their ability to capture temporal dependencies in a structured way and their relative ease of interpretation.[10] However, they have significant limitations. Their foundational assumption is that the relationship between past and future values is

linear, and that the underlying data is stationary (or can be rendered stationary).[19] This makes them inherently ill-suited for capturing the abrupt, non-linear shifts caused by events like the imposition of a lockdown, the relaxation of restrictions, or the sudden emergence of a more transmissible variant.[19] The process of identifying the optimal parameters

(p,d,q) and (P,D,Q) often relies on interpreting autocorrelation (ACF) and partial autocorrelation (PACF) plots, a process that can be subjective and cumbersome.[19] Finally, as primarily univariate models, they rely solely on the past values of the target variable itself, implicitly assuming that all the information needed for a forecast is contained within that single data stream.[14]

1.3 Machine Learning and AI Approaches: Modeling Non-Linear Dynamics

The limitations of traditional models led many researchers to turn to machine learning (ML) and artificial intelligence (AI) techniques, which are adept at identifying complex, non-linear patterns in large datasets.[9]

Polynomial Regression is a relatively simple ML technique that extends linear regression by adding polynomial terms (e.g., x2,x3) to the model equation.[24] This allows it to fit curves to the data, making it useful for capturing non-linear trends, such as the characteristic S-shaped logistic growth curve seen in some epidemic waves.[24] The general form of a polynomial regression equation is

y=β0+β1x+β2x2+...+βpxp+ϵ.[24] However, its flexibility is also its weakness. High-degree polynomials are extremely prone to

overfitting, where the model captures noise in the training data rather than the underlying signal, leading to poor generalization on new data.[24] They can also produce wildly erratic and unreliable extrapolations outside the range of the training data.[27]

A more powerful class of models for sequential data is **Recurrent Neural Networks (RNNs)**. Unlike standard neural networks, RNNs have internal loops that allow information to persist, making them suitable for tasks where sequence order is important.[28] However, standard RNNs struggle with the "vanishing gradient problem," which makes it difficult for them to learn dependencies over long sequences.[28] The

Long Short-Term Memory (LSTM) network is a specialized type of RNN designed to overcome this limitation.[2] LSTMs feature a more complex architecture with "memory cells" and three types of "gates" (input, forget, and output) that regulate the flow of information. This structure enables them to effectively learn and remember long-term patterns and dependencies, making them theoretically well-suited for forecasting complex time series like disease outbreaks.[28]

The primary challenge for these advanced models, particularly in the context of a novel pathogen, is their immense **data requirement**.[2] Deep learning models like

LSTMs require substantial amounts of training data to learn effectively. In the early days of the COVID-19 pandemic, the scarcity of data made their direct application difficult and prone to issues like gradient explosion, limiting their utility when it was most needed.[2] Furthermore, these models are often considered "black boxes" because their internal decision-making processes are not easily interpretable, which can be a significant barrier to their acceptance in high-stakes policy environments.[23]

1.4 Hybrid and Ensemble Models: The Power of Synthesis

Recognizing that no single model is perfect, two other approaches gained prominence: hybrid models and ensemble models. **Hybrid models** seek to combine the strengths of different modeling philosophies. For example, a model might use the differential equations from a mechanistic SEIR model but allow a recurrent neural network to learn the time-varying transmission rate, thereby combining the interpretability of the former with the flexibility of the latter.[2]

Ensemble models take a different approach: instead of building one superior model, they aggregate the forecasts from a large number of diverse, independent models.[31] The final ensemble forecast is typically the median or a weighted average of the individual predictions. This approach operates on the principle of "wisdom of the crowd," where the collective judgment is often more accurate and reliable than that of any single expert. The US COVID-19 Forecast Hub, a collaboration led by the Centers for Disease Control and Prevention (CDC), is a prime example of this strategy in action, collecting and combining forecasts from dozens of research groups.[31] As subsequent sections will show, ensemble forecasts have consistently been among the top performers throughout the pandemic, demonstrating that a strategy of diversification and synthesis is highly effective in managing model uncertainty.[31]

The development of these synthetic approaches represents a crucial evolution in thought. It signifies a move away from the search for a single "perfect" model and towards a framework that acknowledges the inherent limitations of all models. The consistent outperformance of ensembles is not merely a statistical artifact; it is a powerful demonstration that the most robust forecasting strategy is one that hedges against the inevitable failure of any individual approach. This understanding has profound implications for how public health agencies should structure their forecasting capacity, suggesting that fostering a diverse ecosystem of competing models and developing robust methods to synthesize their outputs is more valuable than placing a bet on any single methodology. This shift from a monolithic to a pluralistic approach is one of the most important lessons learned from the pandemic's modeling efforts.

The pandemic laid bare a fundamental "Catch-22" in epidemiological forecasting. In the early, chaotic phase of an outbreak, when data is scarce and the disease dynamics are poorly understood, accurate forecasts are most desperately needed to guide urgent policy decisions. However, this is precisely the environment where most sophisticated models perform worst. Mechanistic models must rely on unverified assumptions, and data-driven models lack the data they need to be trained. Conversely, in the later, more stable phases of an epidemic when data is plentiful, models become more accurate, but their marginal utility for decision-making may have decreased. This paradox highlights the need for a modeling strategy that is adaptive across the lifecycle of a pandemic.

Table 1: Comparative Overview of Major COVID-19 Forecasting Model Architectures

Model Category	Core Principle	Key Assumptions	Data Requirements	Strengths	Weaknesses	Primary COVID-19 Application
Mechanistic (e.g., SEIR)	Models the underlying biological mechanisms of disease transmission through population compartments.	Homogeneous population mixing; fixed transmission rates; key parameters (e.g., asymptomatic rate) are known or can be	Low (relies on epidemiological parameters rather than large datasets).	Provides policy-relevant insights (e.g., estimating R_0); useful for "what-if" scenario analysis.	Unrealistic assumptions about human behavior; struggles with dynamic changes; can be highly sensitive to parameter	Early-stage pandemic planning and evaluating the potential impact of large-scale interventions.

		estimated.			choices.	
Statistical (e.g., ARIMA/SARIMA)	Extrapolates statistical patterns (trends, seasonality) from historical time-series data.	Linear relationships between past and future values; data is stationary or can be made stationary through differencing.	Moderate (requires a consistent time-series of the target variable).	Relatively simple to implement and interpret; effective for stable, short-term forecasting.	Poor at predicting turning points or handling sudden structural breaks; assumes the future will behave like the past.	Short-term (1-4 week) forecasting of cases and deaths, especially where weekly patterns exist.
Machine Learning (e.g., LSTM)	Learns complex, non-linear patterns and long-term dependencies directly from large datasets.	Assumes that patterns present in the training data will persist in the future; requires sufficient representative data.	High (requires large, often diverse, datasets for effective training).	Can capture complex, non-linear dynamics that other models miss; can integrate multiple data streams.	Often a "black box," lacking interpretability; data-hungry, performing poorly with limited data; prone to overfitting.	Analyzing complex trends and integrating multiple data sources (e.g., mobility, demographics) for prediction.
Ensemble	Aggregates forecasts from multiple diverse, individual	Assumes that the collective "wisdom of the crowd" is more	High (relies on the existence of a diverse ecosyste	Consistently more accurate and robust than most individual	Can obscure the specific errors of constituent models;	Official national forecasts (e.g., by the US CDC) to provide a

	models to produce a single, combined forecast.	accurate than any single model; errors of individual models may cancel out.	m of individual models to draw from).	models; mitigates the risk of relying on a single flawed model.	performance depends on the quality and diversity of the models in the ensemble.	reliable, consensus prediction for public health planning.

---

2. Global Performance Analysis: A Comparative Review Across WHO Regions

While the theoretical underpinnings of forecasting models provide a necessary foundation, their true value is determined by their performance in the real world. The COVID-19 pandemic served as an unprecedented, global-scale test of these tools. This section moves from theory to practice, conducting a comparative analysis of model performance across four key WHO regions: the Americas, Europe, South-East Asia, and Africa. The analysis reveals that there was no universally "best" model; instead, performance was highly contingent on the regional context, the specific metric being evaluated, and the nature of the quantity being forecast.

2.1 Evaluation Metrics Deep Dive: Understanding Forecast Accuracy

To compare models rigorously, it is crucial to understand the metrics used to evaluate them. These metrics are not interchangeable; they measure different aspects of a forecast's quality and have different implications for policymakers.

- **Scale-Dependent Errors:** The **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** are two of the most common metrics.[35] Both measure the average magnitude of the forecast error in the same units as the data (e.g., an error of 100 cases). MAE is the average of the absolute differences between predicted and actual values, making it straightforward to interpret.[36] RMSE is the square root of the average of the squared differences. By squaring the errors, RMSE gives disproportionately more weight to large errors.[36] This makes it a preferred metric when large forecast misses are particularly costly and should be heavily penalized.[38]

- **Percentage Errors:** The **Mean Absolute Percentage Error (MAPE)** expresses the average error as a percentage of the actual value, making it a scale-independent metric.[36] This is its primary advantage, as it allows for the comparison of forecast accuracy across regions with vastly different case numbers (e.g., comparing a forecast for the United States with one for New Zealand).[39] However, MAPE has significant drawbacks. It produces infinite or undefined values when the actual value is zero, and it can be biased, as it penalizes over-predictions more heavily than under-predictions.[40]
- **Probabilistic Forecast Scores:** Point forecasts (a single number) are of limited use to policymakers, who also need to understand the range of plausible outcomes. Probabilistic forecasts provide this by giving a full probability distribution for the future value. The **Weighted Interval Score (WIS)** is a state-of-the-art metric for evaluating such forecasts.[43] It generalizes the absolute error to an entire distribution and can be interpreted as a combination of penalties for over-prediction, under-prediction, and the "spread" or sharpness of the prediction intervals.[31] A lower WIS indicates a better forecast. Crucially, WIS evaluates both the accuracy of the central prediction and the **calibration** of the forecast's uncertainty. A well-calibrated model is one whose stated confidence is justified; for example, its 95% prediction intervals should contain the true value 95% of the time.[44]

The choice of metric is not trivial. A model can be accurate (low MAE) but poorly calibrated (high WIS), presenting a dangerously false sense of certainty to decision-makers. A comprehensive evaluation, therefore, requires looking at both accuracy and calibration to get a full picture of a model's utility.

2.2 The Americas: Heterogeneity in Performance Across Waves and Jurisdictions

The forecasting effort in the United States, coordinated by the CDC and the COVID-

19 Forecast Hub, provides one of the most comprehensive datasets for model evaluation.[31] The results from this region are sobering and reveal deep challenges in predictive modeling.

A striking finding was the widespread underperformance of many complex models when compared to simple, naive baselines. One major analysis found that approximately two-thirds of the models submitted to the CDC failed to produce forecasts more accurate than a simple baseline assuming case counts would remain the same as the previous week. One-third of models failed to outperform even a simple linear trend forecast.[34] This suggests that for many models, the added complexity did not translate into better predictive power.

When comparing broad model categories, **ensemble models** generally demonstrated the best overall performance.[34] However, even their superiority was modest, and they were not found to be "significantly" better than the simple baseline models, which in turn outperformed the average of both machine learning and mechanistic epidemiological models.[34] Furthermore, no single modeling approach—whether epidemiological, ML, or hybrid—was consistently the best performer across the different pandemic waves. A model that performed well during the first wave could perform poorly in the next, highlighting the instability of model performance as the pandemic's dynamics changed.[34]

Performance was also highly dependent on context. Forecast skill was generally higher for larger geographical jurisdictions, such as states compared to counties, likely due to more stable data signals at larger scales.[31] Critically, model performance deteriorated significantly during periods of rapid change. During the growth phases of the winter 2020, Delta, and Omicron waves, the 95% prediction interval coverage of many forecasts dropped below 50%, meaning the models were both inaccurate and dangerously overconfident precisely when reliable guidance was most needed.[31]

Finally, forecast accuracy in the US was not equitable. An analysis of forecast errors across social determinants revealed substantial racial disparities. For instance, forecast errors were found to be 21.6% higher for Hispanic populations compared to the average.[32] This disparity is not necessarily a flaw in the models' algorithms themselves, but rather a reflection of the biased data they were trained on. Under-reporting of cases and deaths in minority communities, coupled with biases in auxiliary data sources like cellphone mobility data, created a distorted picture of reality that the models learned and projected into the future.[32]

2.3 Europe: Contrasting Human Judgment with Computational Forecasts

The German and Polish Forecast Hubs provided a unique natural experiment by collecting and evaluating not only computational model forecasts but also predictions from human experts and laypeople, termed "crowd forecasts".[33] The results challenge the assumption that purely computational approaches are always superior.

For the highly uncertain task of forecasting **COVID-19 cases**, crowd forecasts consistently outperformed all other methods, including sophisticated semi-mechanistic models and even the aggregated Hub ensemble forecast, across all time horizons.[33] This suggests that human forecasters were more adept at integrating diverse, often qualitative, information that is difficult to encode in a mathematical model. For example, humans could anticipate the impact of an announced lockdown policy or understand how changes in testing strategies might affect reported case numbers, allowing them to adapt their predictions more quickly than models that were simply extrapolating past trends.[33]

However, the reverse was true for forecasting **COVID-19 deaths**. For this task, computational models, and particularly the Hub ensemble, were consistently superior to crowd forecasts.[33] The key difference is the availability of a strong leading indicator: case numbers. Models that could effectively use case data to predict the subsequent wave of deaths performed well. This highlights a critical principle: the best forecasting tool depends heavily on the nature of the target variable and the structure of the available information.

A pervasive issue across all forecast types in Europe was **overconfidence**. Both individual models and ensembles systematically produced prediction intervals that were too narrow, failing to capture the true degree of uncertainty.[33] This indicates poor calibration. Interestingly, some analyses found that performance-weighted ensembles, while sometimes more accurate in their point predictions, could be even more poorly calibrated than simple equal-weighted ensembles.[44] This creates a difficult trade-off for policymakers between forecast accuracy and the reliability of its stated uncertainty.

2.4. South-East Asia: The Critical Role of Seasonality in Model Accuracy

Analyses from the South-East Asian region, including countries like Indonesia, the Philippines, and Malaysia, offer a clear lesson on the importance of correctly identifying and modeling underlying data structures, particularly seasonality.[11]

In multiple comparative studies from this region, **SARIMA models consistently and significantly outperformed standard ARIMA models**.[18] The key to this success was the identification of a strong 7-day seasonal pattern in the case reporting data, corresponding to a clear weekly cycle (e.g., lower reporting on weekends). By explicitly modeling this weekly seasonality, SARIMA models achieved lower RMSE, MAE, and MAPE values, confirming that accounting for this feature was critical for accuracy.[18]

The competition between statistical models and machine learning in this region was more nuanced. Some research indicated that deep learning models like LSTM could achieve higher predictive accuracy than SARIMA, particularly when trained on longer datasets (e.g., 20 weeks of data versus 4 weeks).[47] However, the same research found that ARIMA models often demonstrated superior stability and reliability across different time frames.[47] This points to a classic trade-off: the potential for higher accuracy from complex, non-linear ML models versus the robustness and stability of more traditional statistical methods. For public health officials, this choice is not straightforward. A slightly less accurate but more stable and reliable model may be preferable to a more powerful but volatile one, especially when consistency is critical for planning.

2.5. Africa: The Imperative for Context-Specific and Validated Models

The experience of forecasting in Africa provides the starkest warning against a "one-size-fits-all" approach to modeling. In the early stages of the pandemic, influential "flagship" models developed in the global north, trained on data from China and Europe, projected "doomsday" scenarios for the African continent that were grossly inaccurate.[48] These models failed because they were not contextualized to Africa's unique socio-ecological landscape, which includes factors like a significantly younger population demographic, different social and household structures, and varying environmental conditions that may have influenced transmission.[48]

This failure of external models highlights the critical need for local data and local modeling expertise. Studies conducted within Africa revealed a very different picture of model performance. There was no single best model type for the continent; performance varied by country. One analysis of the top 10 most infected African nations found that **cubic and quadratic regression models** provided the best fit for the majority of countries, while others were better described by ARIMA or various exponential smoothing models.[49] Another study focusing on the entire continent also found that a cubic model was superior for predicting both cases and deaths, outperforming a range of other exponential family models.[50] The success of these simpler regression models suggests that either the underlying dynamics were less complex than in other regions, or that data was too sparse and noisy to support more complex model structures.

Beyond forecasting the epidemic's trajectory, a related crisis emerged in the domain of clinical prediction models (CPMs), which aim to predict outcomes for individual patients. A systematic review and external validation of 22 different COVID-19 CPMs found that **all of them performed poorly** when applied to a new patient population.[51] Astonishingly, none of the complex models demonstrated better clinical utility than simple, single-variable predictors like a patient's age or baseline oxygen saturation.[51] This "validation crisis" underscores a fundamental and dangerous gap in the modeling pipeline: models are often developed and published without rigorous external validation, rendering them untrustworthy for real-world clinical or policy implementation.

The collective evidence from these regions leads to an inescapable conclusion: context is paramount. The optimal modeling strategy for one region is not transferable to another without careful adaptation and validation. A global pandemic response strategy must therefore be decentralized, empowering regions to develop, adapt, and validate models that are suited to their specific data landscapes, epidemiological dynamics, and socio-ecological realities. Any attempt to impose a single, centralized modeling approach is destined to fail.

Table 2: Summary of Model Performance Metrics by WHO Region and Model Type

WHO Region	Key Study/Hub	Target Variable	Best Performing Model Type(s)	Key Finding/Insight	Primary Evaluation Metric Used	Source(s)
The Americas	US CDC Forecast Hub	Cases, Deaths	Ensemble, Simple Baselines	Many sophisticated models failed to outperform simple baselines. Performance was unstable across waves and showed racial disparities.	MAPE, WIS	[32]
Europe	German/Polish Forecast Hub	Cases	Crowd-sourced (Human Judgment)	Human forecasters better integrated qualitative information (e.g., policy changes) for uncertain variables like cases.	WIS, Absolute Error	[33]
Europe	German/Polish Forecast	Deaths	Ensemble, Computational	Computational models	WIS, Absolute Error	[33]

	Hub		Models	were superior when a strong leading indicator (cases) was available. Pervasive overconfidence was an issue.		
South-East Asia	Country-specific analyses (e.g., Indonesia, Malaysia)	Cases	SARIMA, LSTM	Identifying and modeling weekly seasonality (via SARIMA) was critical for accuracy. LSTM showed potential but was less stable.	RMSE, MAE, MAPE	18
Africa	Continent-wide and multi-country analyses	Cases, Deaths	Simple Regression (Cubic, Quadratic), various time-series	"Flagship" models from the global north failed due to lack of contextualization. Local models	R^2, various time-series metrics	48

				must be developed and validated.		

---

3. The Determinants of Forecast Accuracy: Unpacking Regional and Methodological Disparities

The significant variation in forecast performance across different models and regions was not random. It was driven by a confluence of systematic factors, from the quality of the input data to the unpredictable nature of human behavior. Understanding these determinants is crucial for building more robust and reliable forecasting systems for future health emergencies. This section synthesizes the evidence to explain *why* model performance varied so dramatically, moving from observation to a causal analysis of the key drivers of forecast error.

3.1 The "Garbage In, Garbage Out" Principle: Impact of Data Quality and Availability

The most fundamental determinant of any model's accuracy is the quality of the data it is fed—a principle colloquially known as "garbage in, garbage out".[4] Throughout the pandemic, the global data landscape was characterized by extreme heterogeneity, inconsistency, and uncertainty, which formed the weak foundation upon which all forecasts were built.

Data collection methods, testing capacity, and reporting standards varied enormously between countries and even within them.[1] This meant that "confirmed cases" was not a consistent metric; it was a function of a nation's testing strategy and resources. In the early stages, limited testing capacity meant that only the most severe cases were identified, leading to a massive underestimation of true infection counts.[52] This data scarcity was a particular challenge for data-intensive machine learning models, which require large datasets to be trained effectively.[2] In these data-poor environments, parsimonious models—those with fewer parameters, like simple exponential growth or SIR models—were often more practical, even if their assumptions were flawed.[1]

Even when data was available, it was often biased and lagged. Mortality and morbidity statistics were shown to vary significantly by location and time depending on the accuracy of accounting and differences in data collection protocols.[1] For example, studies of US death certificates revealed a systematic under-reporting of COVID-19 as the cause of death for several minority groups, including a 33% undercount for non-Hispanic American Indian or Alaska Natives.[32] Similarly, lower testing rates in rural areas led to a disproportionate under-detection of cases in those communities.[32] When models are trained on this incomplete and biased data, they do not just learn the dynamics of the virus; they also learn the systemic biases of the data collection system itself. Consequently, a forecast error for a specific community is often not just a "model error" but a symptom of a deeper failure in the public health data infrastructure. Improving forecasting, therefore, is inextricably linked to foundational investments in equitable, high-resolution, and timely data collection.

3.2 The Shifting Landscape: How Epidemic Phases and Viral Evolution Defy Static Models

A second major challenge is that a pandemic is not a static phenomenon. It is a dynamic, evolving process, and models that assume a stable underlying structure are bound to fail. The effective reproduction number, R, is not a fixed biological constant but a highly variable parameter that changes over time and location, heavily influenced by public health interventions and population behavior.[1] Models that rely on fixed or simple time-dependent transmission rates cannot keep pace with this reality.[9]

This dynamism explains why forecast skill was not constant throughout the pandemic. Multiple analyses from the US and Europe showed that models performed worst during periods of rapid change—the turning points at the beginning and peak of epidemic waves.[31] During these phases of exponential growth or decline, models struggled to keep up, often lagging reality by several weeks.[34] Their performance was much better during more stable, linear periods of the pandemic. This is a critical weakness, as accurate forecasts are most essential precisely during these volatile turning points to inform urgent policy action.

The evolution of the virus itself presented another layer of complexity. The emergence of new variants like Delta and Omicron, with significantly different characteristics of transmissibility and immune evasion, represented major **structural breaks** in the time-series data. Most forecasting models are not designed to anticipate such fundamental shifts in the underlying data-generating process. As a result, the mean error of many models was observed to worsen with each successive pandemic wave, as the virus continued to evolve away from the patterns the models had been trained on.[34]

3.3 The Human Factor: Modeling the Impact of NPIs, Public Behavior, and Mobility

Epidemics do not unfold in a vacuum; they are shaped by human actions. This creates a reflexive loop that is incredibly difficult to model. Non-pharmaceutical interventions (NPIs) like lockdowns, mask mandates, and social distancing are explicitly designed to alter the course of the pandemic, breaking the very trends that forecasting models seek to extrapolate.[1] This leads to the

paradox of intervention: a forecast of a dire outcome can be successful precisely because it inspires public health actions that prevent the forecast from coming true.[52] For example, an influential early model's prediction of 2.2 million deaths in the US was widely criticized as inaccurate in hindsight, but it was instrumental in catalyzing the stringent lockdown measures that ensured the prediction was not realized.[52] This means that models cannot be evaluated solely on their predictive accuracy in a vacuum; their utility as decision-support tools that can change the future must also be considered.

Beyond government mandates, individuals engage in **adaptive behavior**, voluntarily changing their contact patterns in response to perceived risk.[53] When case counts are high, people may reduce their social activities, creating a natural brake on transmission that is often not explicitly included in simple compartmental models.[9] Agent-based models, which simulate the decisions of individual "agents," are theoretically capable of capturing these complex feedback loops but are difficult to parameterize and computationally expensive.[54]

To account for these behavioral dynamics, many modeling groups incorporated large-scale human mobility data from sources like Google and Apple as a proxy for social contact.[32] While this was shown to improve the accuracy of some models, this data is not a neutral reflection of society. It is derived primarily from smartphone users, creating sampling biases against certain populations, such as the elderly or those in lower socioeconomic strata. These biases can then be baked into the models, perpetuating and amplifying health inequities in the forecasts they produce.[32]

3.4 The Socio-Ecological Context: Why a "One-Size-Fits-All" Approach Fails

Finally, the spread of an infectious disease is not merely a biological process; it is a socio-ecological one. The trajectory of an epidemic is profoundly shaped by a region's unique context, including its population density, age distribution, household structure, cultural norms, public health infrastructure, and even climate.[2]

The most glaring illustration of this principle was the failure of early global models when applied to the African continent. These models, developed and trained on data from China and Europe, failed to account for Africa's distinct socio-ecological factors, most notably its much younger population demographic. The result was a series of "doomsday" predictions that were orders of magnitude higher than the observed reality, undermining the credibility of the models and leading to accusations of "pseudo-science".[48] This experience served as a powerful and necessary corrective, demonstrating that models cannot be naively transported across contexts without careful validation and adaptation.

This imperative for contextualization is a unifying theme across all regions. The success of SARIMA models in South-East Asia was due to their ability to capture a specific local data pattern (weekly reporting cycles).[18] The success of human forecasters in Europe was tied to their ability to process local, qualitative information about policy changes.[33] The success of simpler regression models in some African nations may have reflected the unique data environment and transmission dynamics in those specific settings.[49] Effective modeling is not about finding a universal key, but about crafting the right key for each specific lock. This requires a "whole-of-society" approach to modeling that integrates epidemiological data with demographic, economic, and behavioral information to create a more holistic and accurate picture of pandemic risk.[9]

4. Strategic Public Health Recommendations for Future Pandemic Preparedness and Response

The comprehensive analysis of COVID-19 forecasting models yields a clear, overarching conclusion: the pursuit of a single, perfect predictive model is a futile endeavor. The inherent uncertainties of a novel pathogen, the complexities of human behavior, and the diversity of global contexts make this an impossible goal. The strategic imperative, therefore, must shift. The goal is not to achieve certainty, but to build a public health architecture that is resilient, adaptive, and capable of making effective decisions under conditions of deep uncertainty. This section translates the analytical findings of this report into four pillars of actionable recommendations for national and international health authorities.

4.1 Fostering an Adaptive Policy Framework: Integrating Forecasts into Real-Time Decision-Making

The static, pre-written pandemic plans of the past proved inadequate for the dynamic reality of COVID-19. Future preparedness must be built on a foundation of **adaptive management**, a framework that treats policies not as fixed directives but as hypotheses to be continuously tested, evaluated, and updated in response to new evidence.[56] Forecasts are the engine of such a system.

Recommendation: National and regional health authorities should establish formal adaptive policy frameworks that integrate forecasting directly into the decision-making cycle.

Implementation Steps:

- **Establish Formal Feedback Loops:** Create dedicated crisis management teams that include both policymakers and modelers. This ensures that modelers are working on policy-relevant questions and that policymakers understand the

capabilities and, crucially, the uncertainties of the forecasts they are using.[57] Regular, structured communication is essential.
- **Embrace Scenario Analysis:** Use forecasting models not just to produce a single "most likely" prediction, but to conduct "what-if" scenario analyses.[59] For example, models can be used to estimate the potential impact of relaxing or tightening specific NPIs, providing a quantitative basis for comparing policy options.
- **Develop Tiered, Flexible Response Plans:** Design public health response plans with clear, pre-defined tiers of action linked to forecast-based indicators (e.g., projected hospital admissions). However, these plans must be flexible, allowing leaders to adapt and course-correct as real-time surveillance data and updated forecasts become available.[58] This approach balances the need for proactive planning with the ability to respond to unforeseen developments.

4.2 Proactive Resource Management: A Forecast-Driven Approach to Supply Chain and Stockpile Strategy

The pandemic exposed severe fragilities in the global medical supply chain, leading to critical shortages of personal protective equipment (PPE), diagnostics, and other essential supplies.[61] A reactive approach to resource management is insufficient. Forecasts provide the ability to be proactive, anticipating needs before they become critical.

Recommendation: Public health agencies and healthcare systems must leverage short- and medium-term forecasts to drive a proactive and resilient strategy for resource allocation, supply chain management, and strategic stockpiling.

Implementation Steps:

- **Build Supply Chain Resilience:** Move away from fragile "just-in-time" supply chains towards more resilient models. This involves using forecasts to anticipate demand surges, diversifying suppliers to avoid single points of failure, increasing data transparency across the supply chain with modern, cloud-based Enterprise Resource Planning (ERP) systems, and renegotiating vendor contracts to prioritize reliability over just the lowest cost.[61]
- **Modernize Strategic Stockpiles:** Evolve the concept of the Strategic National

Stockpile (SNS) from a static, one-size-fits-all repository to a dynamic, multi-tiered system informed by forecasts.[64] Long-range forecasts of plausible pandemic scenarios can inform what critical, hard-to-source items should be held in a central government stockpile. Shorter-range forecasts can trigger the release of these assets and guide the use of more flexible mechanisms like vendor-managed inventory for more common supplies.[64]
- **Guide Regional and Local Allocation:** At the regional and hospital level, forecasts of patient loads can enable administrators to make timely decisions about staffing levels, bed capacity, and the distribution of critical resources like ventilators.[67] This proactive internal allocation can help optimize resource use and delay or prevent the need to resort to crisis standards of care.[68]

4.3 An Ethical Framework for Scarcity: Using Forecasts to Guide Resource Allocation

During a severe pandemic, demand for critical resources like ICU beds, ventilators, or novel therapeutics can outstrip supply, forcing horrific life-or-death allocation decisions. These tragic choices must not be made in an ad-hoc manner in the heat of a crisis. They must be guided by a pre-established, transparent, and fair ethical framework.

Recommendation: Health authorities must develop and institutionalize clear ethical frameworks for the allocation of scarce medical resources, using epidemiological forecasts as a key input to activate and guide this process proactively.

Implementation Steps:

- **Plan Proactively:** Use long-range forecasts of potential pandemic severity to anticipate when resource scarcity is a plausible outcome. This anticipation should trigger the activation of a dedicated **Triage or Allocation Committee** *before* the crisis reaches its peak, allowing for considered deliberation rather than panicked reaction.[69]
- **Establish Clear, Value-Based Principles:** The framework must be grounded in core ethical principles, including: the equal value of all persons (prohibiting discrimination based on race, gender, disability, etc.); maximizing benefit (utility, e.g., saving the most lives); prioritizing the worst-off (those in greatest medical

need); and ensuring fairness and consistency.[69] Forecasts can help operationalize the "maximize benefit" principle by allowing committees to model the population-level outcomes of different allocation strategies.

- **Ensure Procedural Fairness:** The process must be transparent, with allocation criteria clearly defined and publicly communicated to the extent possible.[70] To maintain trust and objectivity, the allocation decisions should be made by a dedicated committee, separating the role of the frontline clinician (who advocates for their individual patient) from the role of the allocator (who must consider the needs of the entire population).[69]

4.4 Strengthening the Global Data and Modeling Infrastructure for Enhanced Preparedness

The analysis in this report makes one thing unequivocally clear: the quality of our forecasts is a direct reflection of the quality of our data and modeling infrastructure. A truly global and equitable pandemic response requires a foundational investment in this infrastructure.

Recommendation: The WHO and its Member States should lead a coordinated effort to invest in a global pandemic intelligence infrastructure that prioritizes equitable data systems, local modeling capacity, standardized evaluation, and collaborative platforms.

Implementation Steps:

- **Invest in the Data Foundation:** The single most important step to improve forecasting is to improve the data it relies on. This requires sustained investment in national and local public health surveillance systems to produce timely, high-resolution, and reliable data.[72] Specific efforts must be made to identify and correct for known data gaps and biases, particularly those affecting marginalized and under-served communities.[32]
- **Build and Empower Local Modeling Capacity:** The failure of external models in regions like Africa highlights the urgent need to decolonize disease modeling.[48] International partners must support the development of local modeling expertise in all WHO regions, providing training, resources, and access to data. This is the only way to ensure the development of contextually appropriate models that

reflect local realities.[48]

- **Mandate Standardized, Probabilistic Evaluation:** The global health community should adopt a standardized protocol for forecast submission and evaluation, modeled on the success of the US and European Hubs. This standard should mandate that all publicly shared forecasts be **probabilistic** (providing a range of outcomes, not just a single number) and that they be evaluated on both **accuracy** (e.g., MAE) and **calibration** (using a proper score like WIS).[43] This promotes transparency, enables rigorous comparison, and ensures that a model's stated confidence is properly scrutinized.
- **Support and Expand Collaborative Platforms:** The success of ensemble forecasting demonstrates the immense value of collaborative platforms like the COVID-19 Forecast Hub.[31] These platforms foster a diverse modeling ecosystem, encourage innovation, and produce a more robust and reliable consensus forecast. Supporting these hubs and potentially expanding them into a global network should be a key priority for international health organizations.

Table 3: Framework for Adaptive Public Health Interventions Based on Forecast Confidence

Forecast Signal Level	Indicators	Recommended Policy Stance	Resource Allocation Strategy	Public Communication Focus
Level 1: High Confidence / Low Uncertainty	Tight, well-calibrated prediction intervals (low WIS). Multiple models show strong agreement on trend and magnitude.	**Targeted Interventions:** Implement geographically and demographically focused measures. Confidently ease or tighten restrictions based on forecast trajectory.	**Efficient Deployment:** Allocate resources (e.g., mobile testing, vaccines) precisely to predicted hotspots. Maintain optimized supply chain operations.	**Clear Directives:** Communicate specific actions and expected outcomes. Provide clear guidance to the public with high confidence.

Level 2: Moderate Confidence / Known Uncertainty	Wider prediction intervals, but models agree on the general trend (e.g., cases will increase). Moderate WIS scores.	**Broad-Based Mitigation & Contingency Planning:** Implement wider, less-targeted measures (e.g., regional mask mandates). Prepare for multiple plausible scenarios within the forecast range.	**Scenario-Based Pre-positioning:** Pre-position resources to handle both the median forecast and the upper-bound (worse-case) scenario. Activate secondary suppliers.	**Transparent Uncertainty:** Explain the range of possible outcomes. Prepare the public for potential shifts in policy as the situation clarifies. Emphasize the reasons for caution.
Level 3: Low Confidence / High Uncertainty	Very wide prediction intervals. Models disagree on the direction of the trend (some predict increase, some decrease). High WIS scores.	**Defensive Posture & Surveillance Ramp-up:** Maintain baseline public health measures. Avoid major policy shifts. Dramatically increase testing and genomic surveillance to gain clarity.	**Resource Conservation & Mobilization:** Conserve critical resources. Hold strategic reserves. Place reserve supply chains and mutual aid agreements on standby.	**Honesty About Uncertainty:** Be transparent about what is unknown and why. Focus communication on what individuals can do to reduce personal risk (e.g., vaccination, ventilation).
Level 4: Deep Uncertainty / Conflicting Models	Models fail baseline tests. A known structural break has occurred (e.g., new variant of unknown characteristics).	**Assume Worst-Case & Maximize Adaptability:** Implement robust, "no-regrets" measures that are effective	**Strategic Stockpile Mobilization:** Prepare for the potential need to activate the national stockpile. Alert healthcare	**Crisis Communication & Solidarity:** Acknowledge the crisis and the failure of existing models. Call for collective action

	Forecasts are unreliable.	across a wide range of possibilities. Prioritize actions that preserve future options.	systems to prepare for potential crisis standards of care.	and solidarity. Focus on resilience and mutual support.

Conclusion: Towards a More Resilient and Forecast-Informed Global Health Security Architecture

The COVID-19 pandemic subjected the world to the most intense public health crisis in a century, and in doing so, it stress-tested every component of our global preparedness and response systems. This report's analysis of the vast and varied landscape of forecasting models reveals a critical lesson: the challenge was not merely technical, but systemic. The successes and failures of our models were inextricably linked to the strengths and weaknesses of the data ecosystems they relied on, the policy frameworks they sought to inform, and the complex, ever-changing human contexts in which they operated.

We have learned that no single model can serve as a crystal ball. Mechanistic models provided a valuable language for understanding transmission but rested on fragile assumptions. Statistical models captured trends but were broken by the pandemic's shocks. Machine learning models offered power but demanded data that did not exist when it was most needed. The most reliable forecasts often came from the humble synthesis of many imperfect models, a testament to the power of collaboration and intellectual humility. We learned that a forecast's accuracy is distinct from its calibration; a model that is correct on average but dangerously overconfident in its certainty is a flawed tool for a decision-maker navigating a crisis. Most importantly, we learned that context is king—a model developed in one part of the world cannot be naively applied to another without risking catastrophic error and a profound loss of public trust.

The path forward, therefore, is not a relentless search for a better algorithm alone. It is the construction of a more resilient, intelligent, and equitable **global health**

security architecture. This requires a paradigm shift: from seeking certainty to managing uncertainty. The recommendations outlined in this report—fostering adaptive policy frameworks, driving proactive resource management, embedding ethics into crisis planning, and building a robust global data and modeling infrastructure—are the pillars of such a system. They represent a move away from a reactive posture to a proactive and forecast-informed one.

By investing in high-quality, equitable data collection, we provide the fuel for all good models. By fostering local modeling expertise, we ensure those models are relevant and trusted. By standardizing evaluation and demanding probabilistic forecasts, we promote transparency and a more honest appraisal of uncertainty. And by integrating these improved forecasts into adaptive policy and resource frameworks, we build a system that can bend without breaking, learning and evolving in real-time as a crisis unfolds. The WHO Pandemic Agreement, born from the crucible of this pandemic, provides a generational opportunity to codify these lessons into tangible, binding commitments for the global community.[74] By seizing this opportunity, we can transform the painful lessons of COVID-19 into a more resilient future, one in which we are better prepared to face the next pandemic not with perfect foresight, but with wisdom, adaptability, and a shared commitment to protecting all of humanity.

Works cited

1. The challenges of modeling and forecasting the spread of COVID-19 ..., accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7382213/
2. Enhancing COVID-19 Prevalence Forecasting: A Hybrid Approach ..., accessed on August 6, 2025, https://www.mdpi.com/2673-9909/4/2/22
3. Predictive Models for Forecasting Public Health Scenarios: Practical Experiences Applied during the First Wave of the COVID-19 Pandemic - MDPI, accessed on August 6, 2025, https://www.mdpi.com/1660-4601/19/9/5546
4. COVID-19: Understanding – and misunderstanding – epidemic models - Monash Lens, accessed on August 6, 2025, https://lens.monash.edu/@technology/2020/04/16/1380098/covid-19-understanding-and-misunderstanding-epidemiology-models
5. Two new compartmental epidemiological models and their equilibria - medRxiv, accessed on August 6, 2025, https://www.medrxiv.org/content/10.1101/2021.09.03.21263050v1.full-text
6. A Comparative Analysis of Forecasting Models on COVID-19 | Request PDF - ResearchGate, accessed on August 6, 2025, https://www.researchgate.net/publication/361561202_A_Comparative_Analysis_of_Forecasting_Models_on_COVID-19
7. The challenges of modeling and forecasting the spread of COVID-19 - PNAS,

accessed on August 6, 2025, https://www.pnas.org/doi/10.1073/pnas.2006520117
8. Limitations of models for guiding policy in the COVID-19 pandemic | medRxiv, accessed on August 6, 2025, https://www.medrxiv.org/content/10.1101/2022.06.30.22277091v1.full-text
9. Mathematical models for COVID-19: applications, limitations, and ..., accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7386428/
10. Using the SARIMA Model to Forecast the Fourth Global Wave of Cumulative Deaths from COVID-19: Evidence from 12 Hard-Hit Big Countries - MDPI, accessed on August 6, 2025, https://www.mdpi.com/2225-1146/10/2/18
11. Predictive analytics of COVID-19 cases and tourist arrivals in ASEAN based on covid-19 cases - PMC, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9546420/
12. Master ARIMA: Your Guide to Time Series Forecasting - Investopedia, accessed on August 6, 2025, https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp
13. What are ARIMA Models? | IBM, accessed on August 6, 2025, https://www.ibm.com/think/topics/arima-model
14. What Is ARIMA Modeling? | Master's in Data Science, accessed on August 6, 2025, https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arima-modeling/
15. ARIMA for Time Series Forecasting: A Complete Guide - DataCamp, accessed on August 6, 2025, https://www.datacamp.com/tutorial/arima
16. Time Series forecasting using SARIMA in Python | by Tirtha Mutha | Medium, accessed on August 6, 2025, https://medium.com/@tirthamutha/time-series-forecasting-using-sarima-in-python-8b75cd3366f2
17. ARIMA & SARIMA: Real-World Time Series Forecasting - neptune.ai, accessed on August 6, 2025, https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide
18. (PDF) Forecasting COVID-19 cases for Top-3 countries of Southeast ..., accessed on August 6, 2025, https://www.researchgate.net/publication/368974578_Forecasting_COVID-19_cases_for_Top-3_countries_of_Southeast_Asian_Nation
19. What are the limitations of ARIMA models? - Milvus, accessed on August 6, 2025, https://milvus.io/ai-quick-reference/what-are-the-limitations-of-arima-models
20. Understanding the Limitations of ARIMA Forecasting | Towards Data Science, accessed on August 6, 2025, https://towardsdatascience.com/understanding-the-limitations-of-arima-forecasting-899f8d8e5cf3/
21. Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions - PMC, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7986567/
22. What are the limitations of the time series models (ARIMA, etc) for Forecasting? - Reddit, accessed on August 6, 2025, https://www.reddit.com/r/quant/comments/1826m0x/what_are_the_limitations_o

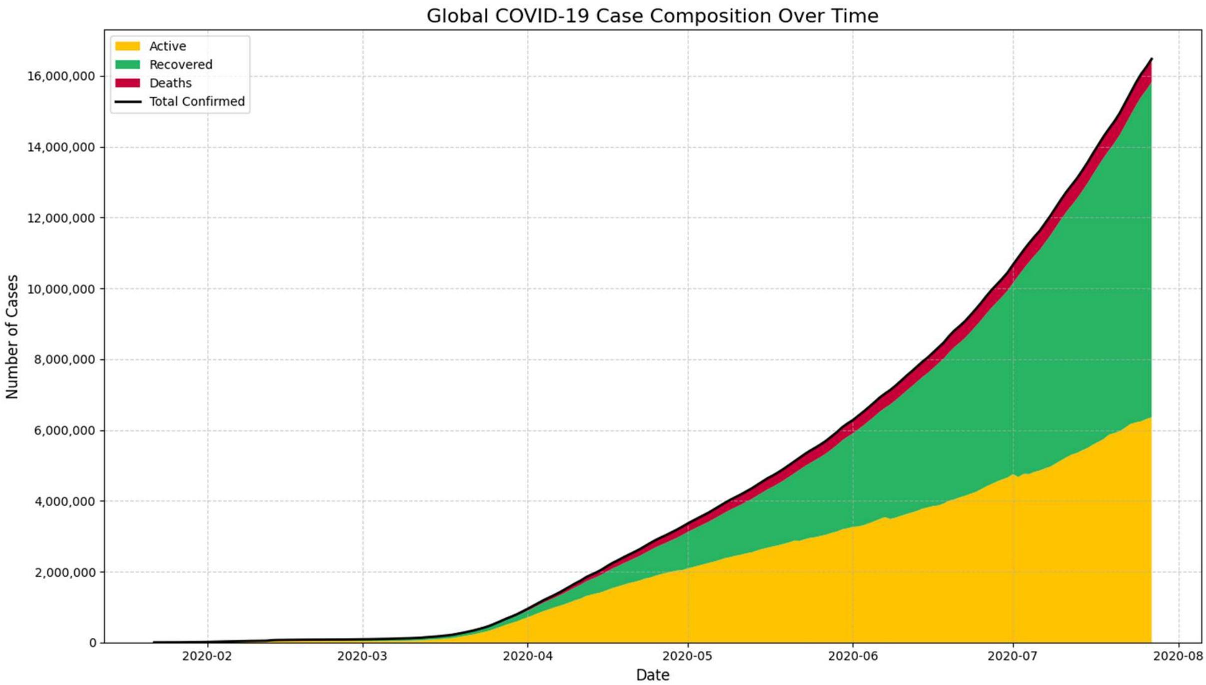
f_the_time_series/
23. A Comparative Study on Deep Learning Models for COVID-19 Forecast - MDPI, accessed on August 6, 2025, https://www.mdpi.com/2227-9032/11/17/2400
24. Polynomial Regression | Forecasting Class Notes - Fiveable, accessed on August 6, 2025, https://library.fiveable.me/forecasting/unit-4/polynomial-regression/study-guide/rRZvQN97tcKqI5Xv
25. Implementation of Polynomial Regression - GeeksforGeeks, accessed on August 6, 2025, https://www.geeksforgeeks.org/machine-learning/python-implementation-of-polynomial-regression/
26. Understanding Polynomial Regression: A Powerful Tool for Complex Relationships | by Data Overload | Medium, accessed on August 6, 2025, https://medium.com/@data-overload/understanding-polynomial-regression-a-powerful-tool-for-complex-relationships-d2394a898fd6
27. [D] Why do "polynomials" (e.g. polynomial regression) have a bad reputation of overfitting? : r/statistics - Reddit, accessed on August 6, 2025, https://www.reddit.com/r/statistics/comments/mv1o8f/d_why_do_polynomials_eg_polynomial_regression/
28. Introduction to Time Series Forecasting with LSTMs | CodeSignal Learn, accessed on August 6, 2025, https://codesignal.com/learn/courses/time-series-forecasting-with-lstms/lessons/introduction-to-time-series-forecasting-with-lstms
29. Time Series Stock Prediction with LSTM | by Gabe Nosek - Medium, accessed on August 6, 2025, https://gabenosek.medium.com/time-series-stock-prediction-with-lstm-eb04f2224c22
30. Comparative analysis of time series prediction models for visceral ..., accessed on August 6, 2025, https://digitalcommons.aaru.edu.jo/cgi/viewcontent.cgi?article=3455&context=amis
31. Challenges of COVID-19 Case Forecasting in the US, 2020–2021 (Journal Article) - OSTI, accessed on August 6, 2025, https://www.osti.gov/biblio/2350971
32. Auditing the fairness of the US COVID-19 forecast hub's case prediction models | PLOS One, accessed on August 6, 2025, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0319383
33. Comparing human and model-based forecasts of COVID-19 in ..., accessed on August 6, 2025, https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010405
34. Accuracy of US CDC COVID-19 forecasting models - PMC, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11233691/
35. RMSE in Epidemiology: A Comprehensive Guide - Number Analytics, accessed on August 6, 2025, https://www.numberanalytics.com/blog/ultimate-guide-to-rmse-in-epidemiology
36. Evaluating forecast accuracy (MAE, RMSE, MAPE) | Intro to Time ..., accessed on August 6, 2025, https://library.fiveable.me/intro-time-series/unit-8/evaluating-

forecast-accuracy-mae-rmse-mape/study-guide/ijqkb0CAqRaHLBFi
37. Assessment of Accuracy Metrics for Time Series Forecasting | by Satyajit Chaudhuri | Analytics Vidhya | Medium, accessed on August 6, 2025, https://medium.com/analytics-vidhya/assessment-of-accuracy-metrics-for-time-series-forecasting-bc115b655705
38. sMAPE vs MAPE vs RMSE, when to use which regression metric - ML EXPLAINED, accessed on August 6, 2025, https://mlexplained.blog/2023/01/11/smape-vs-mape-vs-rmse-when-to-use-which-regression-metric/
39. 3.4 Evaluating forecast accuracy | Forecasting: Principles and Practice (2nd ed) - OTexts, accessed on August 6, 2025, https://otexts.com/fpp2/accuracy.html
40. Metrics Evaluation: MSE, RMSE, MAE and MAPE | by Jonatasv - Medium, accessed on August 6, 2025, https://medium.com/@jonatasv/metrics-evaluation-mse-rmse-mae-and-mape-317cab85a26b
41. Appendix 3: Error Metrics - OneStream Documentation, accessed on August 6, 2025, https://documentation.onestream.com/1375907/Content/SML/user-guide/error-metrics.html
42. Interpretation of typical MAPE values | Download Table - ResearchGate, accessed on August 6, 2025, https://www.researchgate.net/figure/nterpretation-of-typical-MAPE-values_tbl1_257812432
43. Forecast Evaluation Dashboard - CMU Delphi - Carnegie Mellon University, accessed on August 6, 2025, https://delphi.cmu.edu/forecast-eval/
44. Assessing the Performance of COVID-19 Forecasting ... - medRxiv, accessed on August 6, 2025, https://www.medrxiv.org/content/10.1101/2020.12.09.20246157v3.full.pdf
45. Challenges of COVID-19 Case Forecasting in the US, 2020-2021 - ResearchGate, accessed on August 6, 2025, https://www.researchgate.net/publication/384444452_Challenges_of_COVID-19_Case_Forecasting_in_the_US_2020-2021
46. A retrospective assessment of COVID-19 model performance in the USA - Journals, accessed on August 6, 2025, https://royalsocietypublishing.org/doi/10.1098/rsos.220021
47. Predictive Modeling for Pandemic Forecasting: A COVID-19 Study in New Zealand and Partner Countries - MDPI, accessed on August 6, 2025, https://www.mdpi.com/1660-4601/22/4/562
48. The need to return to the basics of predictive modelling for disease outbreak response: lessons from COVID-19 - PMC, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7664134/
49. (PDF) Modeling and Forecasting of COVID-19 New Cases in Top 10 Infected African Countries - ResearchGate, accessed on August 6, 2025, https://www.researchgate.net/publication/346194475_Modeling_and_Forecasting_of_COVID-19_New_Cases_in_Top_10_Infected_African_Countries
50. Predictive models on COVID 19: What Africans should do? - PMC, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7670905/

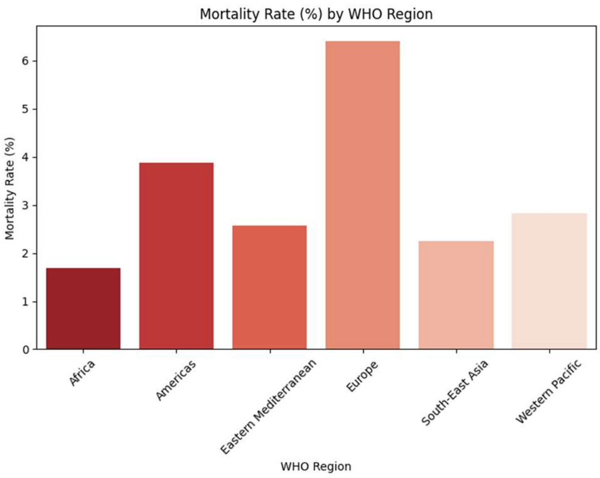
51. Performance of prediction models for COVID-19: the Caudine Forks of the external validation - ERS Publications, accessed on August 6, 2025, https://publications.ersnet.org/content/erj/56/6/2003728
52. Evaluating the Comparative Accuracy of COVID-19 Mortality Forecasts: An Analysis of the First-Wave Mortality Forecasts in the United States - MDPI, accessed on August 6, 2025, https://www.mdpi.com/2571-9394/4/4/44
53. Adaptive human behavior in epidemiological models - PMC, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC3076845/
54. A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology - JASSS, accessed on August 6, 2025, https://www.jasss.org/20/3/2.html
55. Pandemic influenza preparedness and response : a WHO guidance document, accessed on August 6, 2025, https://www.loc.gov/item/2021763175/
56. Adaptive Management and the Value of Information: Learning Via Intervention in Epidemiology | PLOS Biology, accessed on August 6, 2025, https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001970
57. Adaptive Strategies in Healthcare Crisis Management: Navigating Unforeseen Challenges - Syndicate of Hospitals in Lebanon, accessed on August 6, 2025, https://syndicateofhospitals.org.lb/Content/uploads/SyndicateMagazinePdfs/5228_12-13.pdf
58. Public Health Crisis Management Playbook, accessed on August 6, 2025, https://www.oclc.org/realm/public-health-crisis-playbook.html
59. Forecasting post COVID-19 - DiVA portal, accessed on August 6, 2025, http://www.diva-portal.org/smash/get/diva2:1836842/FULLTEXT01.pdf
60. Towards Resilient Healthcare Systems: A Framework for Crisis Management - MDPI, accessed on August 6, 2025, https://www.mdpi.com/1660-4601/21/3/286
61. Health care sector braces for supply chain uncertainty with changing tariff policies - AAMC, accessed on August 6, 2025, https://www.aamc.org/news/health-care-sector-braces-supply-chain-uncertainty-changing-tariff-policies
62. GHX Outlines Five Best Practices to Fortify the Healthcare Supply Chain, accessed on August 6, 2025, https://www.ghx.com/news-releases/2021/ghx-outlines-five-best-practices-to-fortify-the-healthcare-supply-chain/
63. What is Healthcare Supply Chain Resilience - Ryder, accessed on August 6, 2025, https://www.ryder.com/en-us/insights/blogs/logistics/healthcare-supply-chain-resilience
64. Frequently asked questions about the Strategic National Stockpile (SNS) | Mass.gov, accessed on August 6, 2025, https://www.mass.gov/info-details/frequently-asked-questions-about-the-strategic-national-stockpile-sns
65. Mass Counter Measure Response + Strategic National Stockpile Plan - NACCHO, accessed on August 6, 2025, https://www.naccho.org/resource-hub-articles/mass-counter-measure-response-strategic-national-stockpile-plan
66. Receiving, distributing, and dispensing Strategic National Stockpile assets : guide to preparedness, Version 11 - CDC stacks, accessed on August 6, 2025,

https://stacks.cdc.gov/view/cdc/77036
67. Allocation of Scarce Resources During Pandemic: - National Alliance for Care at Home, accessed on August 6, 2025, https://allianceforcareathome.org/wp-content/uploads/Resource_Allocation_Statement_June-2021.pdf
68. Template - Hospital Crisis Standards of Care Resource Allocation Annex - HHS.gov, accessed on August 6, 2025, https://files.asprtracie.hhs.gov/documents/template-hospital-csc-resource-allocation-annex.pdf
69. An Ethics Framework for Making Resource Allocation Decisions Within Clinical Care: Responding to COVID-19 - PMC, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7445717/
70. Whom Should Be Saved? A Proposed Ethical Framework for Allocating Scarce Medical Resources to COVID-19 Patients Using Fuzzy Logic - Frontiers, accessed on August 6, 2025, https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2021.600415/full
71. Ethics and COVID-19: resource allocation and priority-setting - World Health Organization (WHO), accessed on August 6, 2025, https://www.who.int/docs/default-source/blue-print/ethics-and-covid-19-resource-allocation-and-priority-setting.pdf?sfvrsn=4c14e95c_1
72. Improving and Expanding the Evidence Base for Public Health Emergency Preparedness and Response - NCBI, accessed on August 6, 2025, https://www.ncbi.nlm.nih.gov/books/NBK564000/
73. described policy- relevant evidence base for policymaking: an evidence map of COVID- 19 literature | BMJ Public Health, accessed on August 6, 2025, https://bmjpublichealth.bmj.com/content/bmjph/2/2/e000694.full.pdf
74. Pandemic prevention, preparedness and response agreement - World Health Organization (WHO), accessed on August 6, 2025, https://www.who.int/news-room/questions-and-answers/item/pandemic-prevention--preparedness-and-response-accord

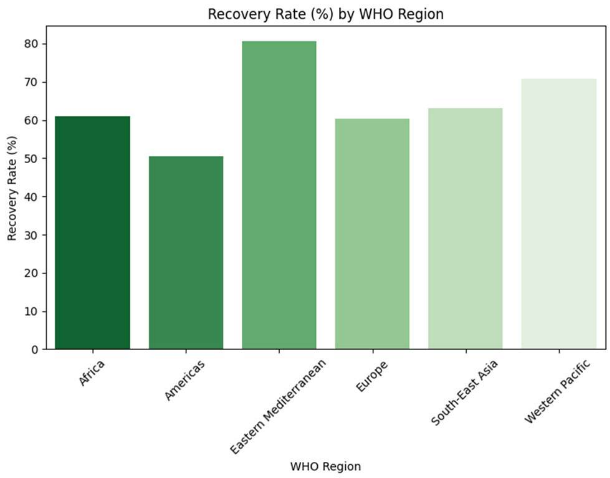
Visualisations: In EDA- Exploratory Data Analysis



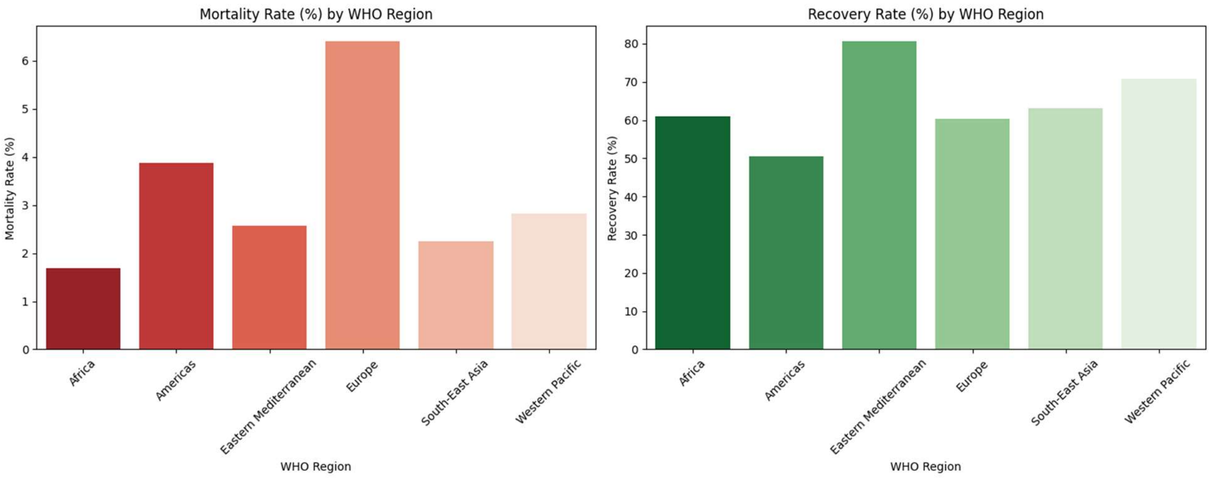
Global COVID-19 Case Composition Over Time
Active
Recovered
Deaths
Total Confirmed
16,000,000
14,000,000
12,000,000
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
Number of Cases
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
Date

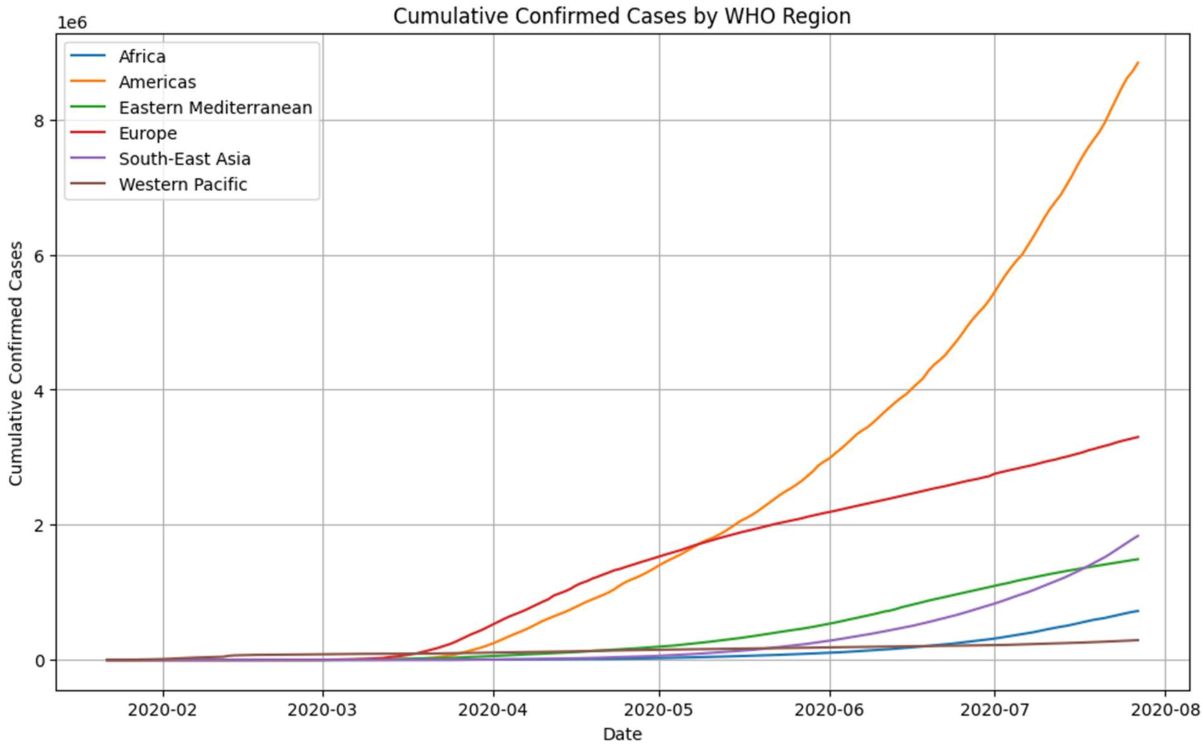


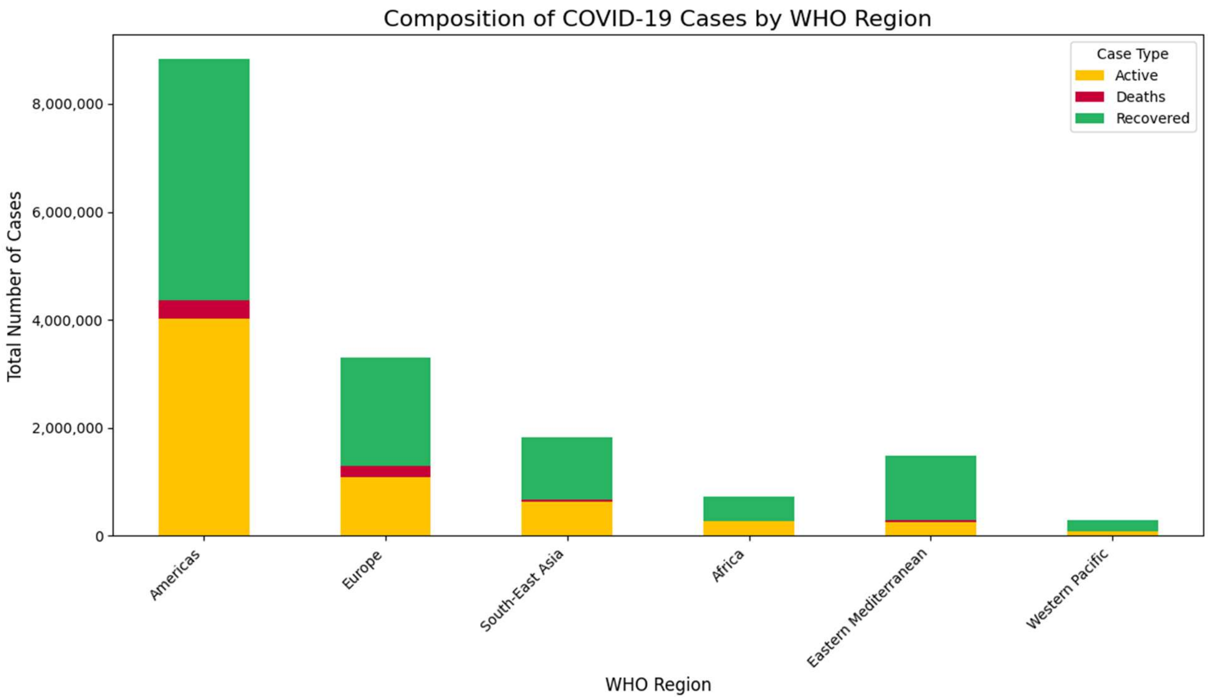
Mortality Rate (%) by WHO Region
Mortality Rate (%)
6
5
4
3
2
1
0
Africa
Americas
Eastern Mediterranean
Europe
South-East Asia
Western Pacific
WHO Region



Recovery Rate (%) by WHO Region
Recovery Rate (%)
80
70
60
50
40
30
20
10
0
Africa
Americas
Eastern Mediterranean
Europe
South-East Asia
Western Pacific
WHO Region

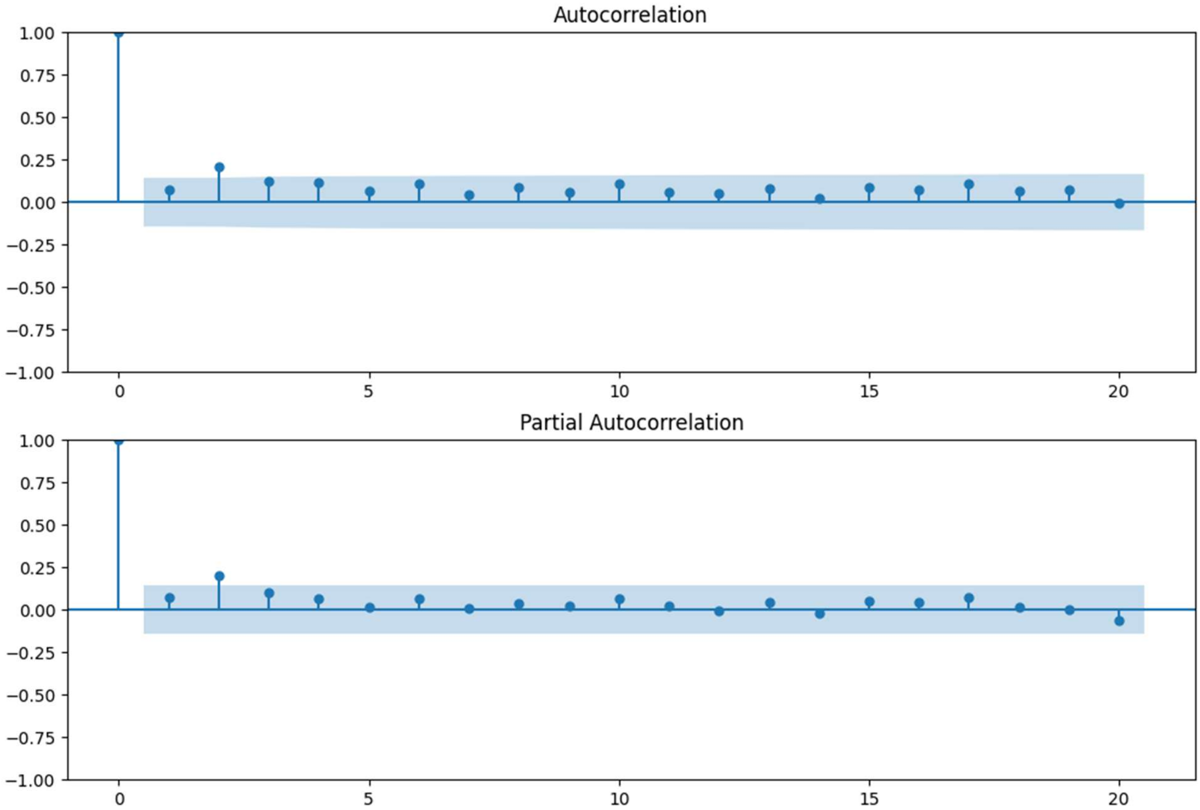

Mortality Rate (%) by WHO Region
Mortality Rate (%)
0
1
2
3
4
5
6
Africa
Americas
Eastern Mediterranean
Europe
South-East Asia
Western Pacific
WHO Region
Recovery Rate (%) by WHO Region
Recovery Rate (%)
0
10
20
30
40
50
60
70
80
Africa
Americas
Eastern Mediterranean
Europe
South-East Asia
Western Pacific
WHO Region

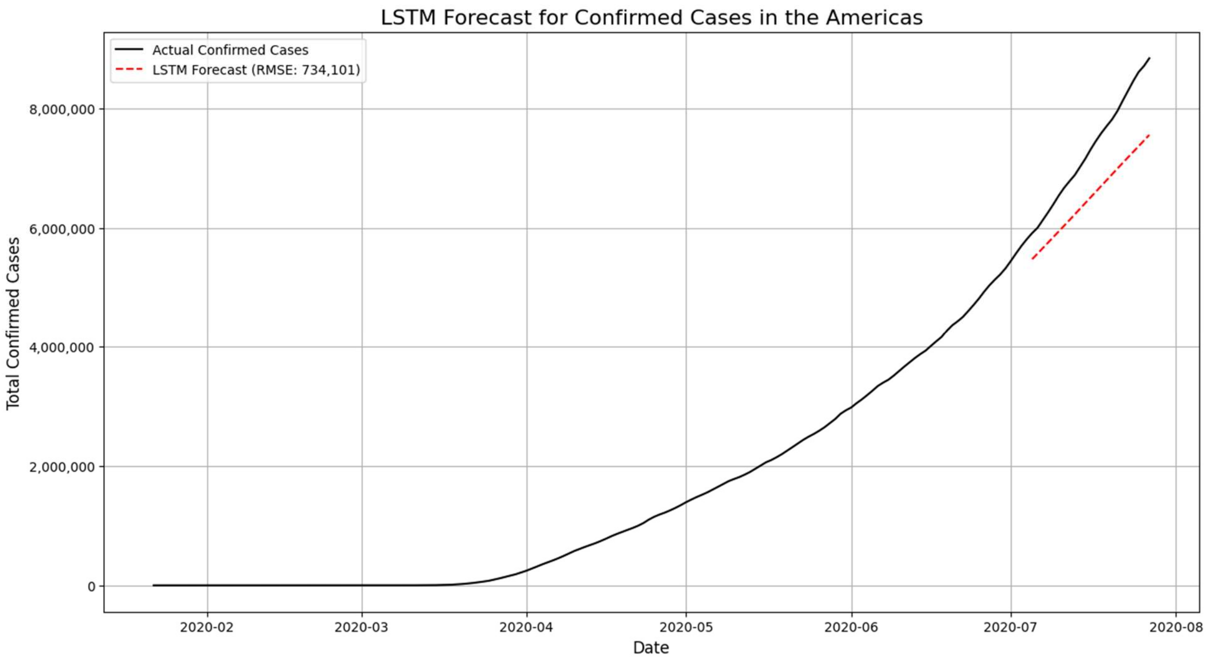

Cumulative Confirmed Cases by WHO Region
1e6
Africa
Americas
Eastern Mediterranean
Europe
South-East Asia
Western Pacific
Cumulative Confirmed Cases
0
2
4
6
8
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
Date

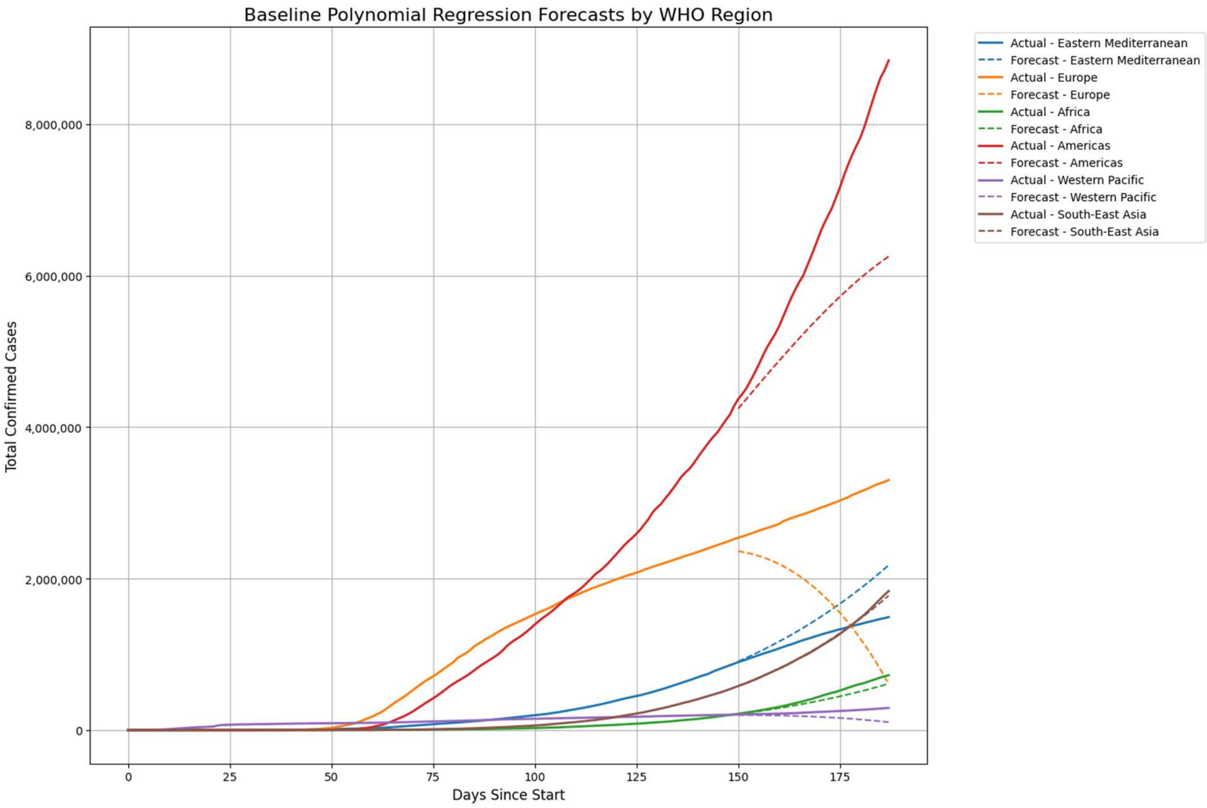

Composition of COVID-19 Cases by WHO Region
Case Type
Active
Deaths
Recovered
Total Number of Cases
8,000,000
6,000,000
4,000,000
2,000,000
0
Americas
Europe
South-East Asia
Africa
Eastern Mediterranean
Western Pacific
WHO Region


CANADA
RUSSIA
AUSTRALIA

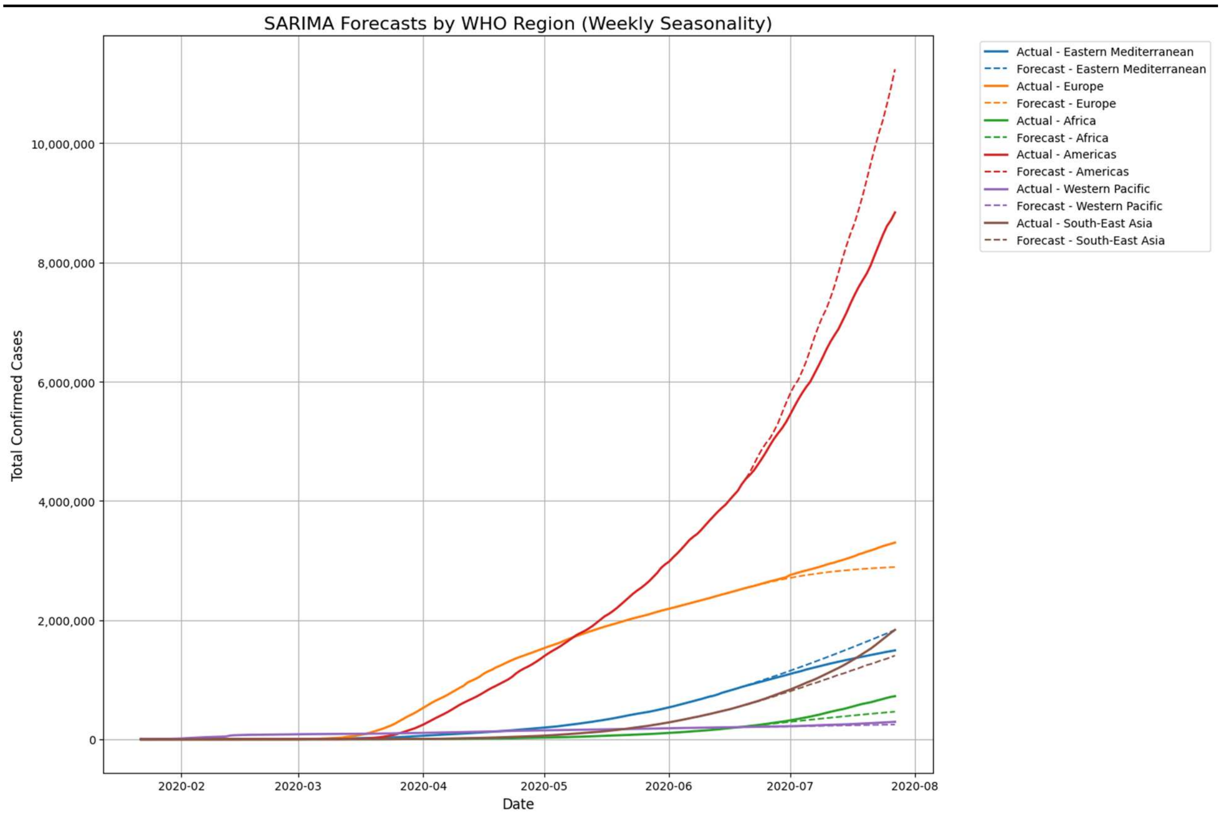
Visualisations: In Modelling & Comparisions



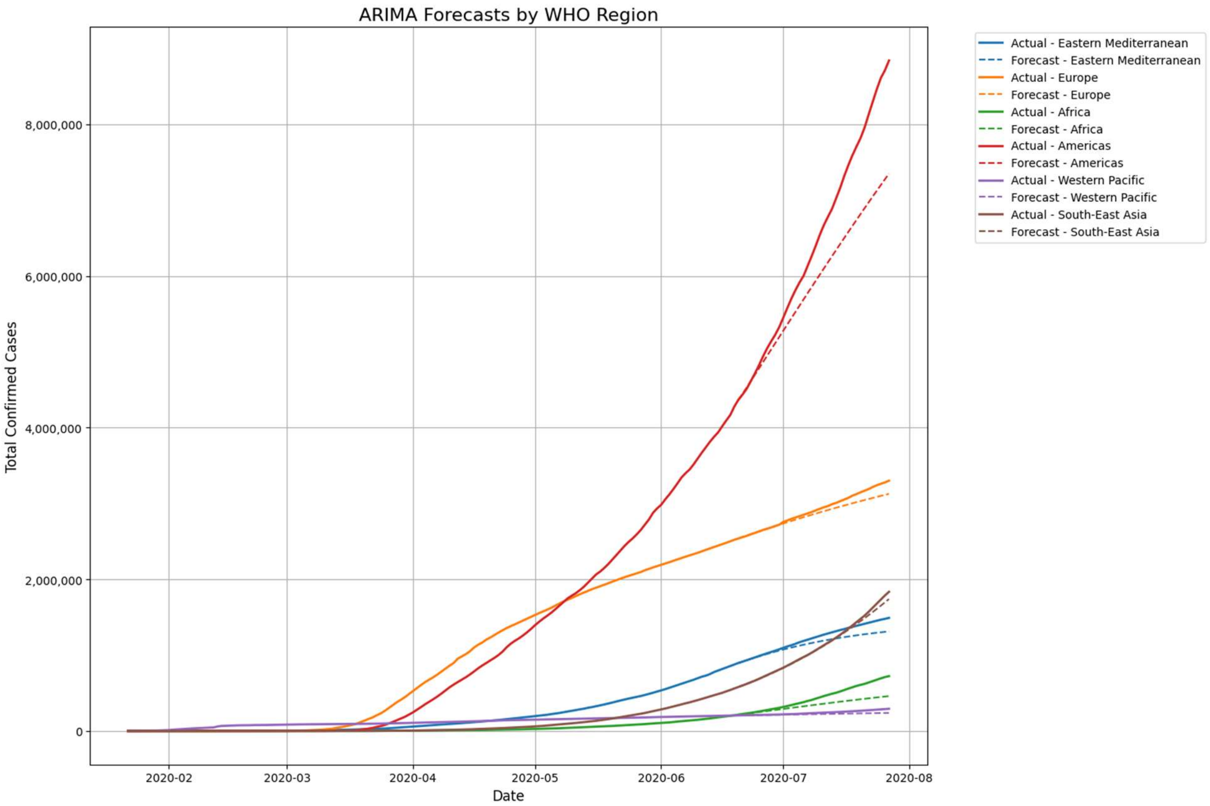




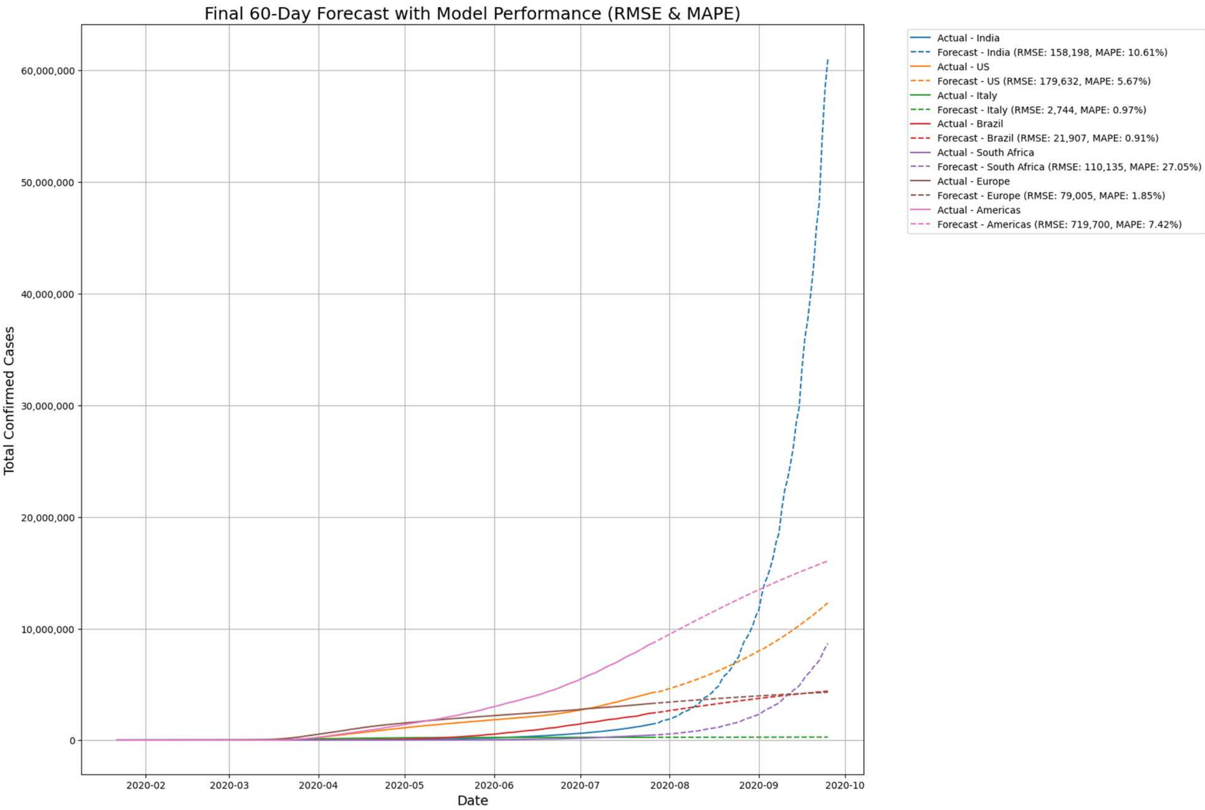
Baseline Polynomial Regression Forecasts by WHO Region
Total Confirmed Cases
Days Since Start
0
25
50
75
100
125
150
175
2,000,000
4,000,000
6,000,000
8,000,000
Actual - Eastern Mediterranean
Forecast - Eastern Mediterranean
Actual - Europe
Forecast - Europe
Actual - Africa
Forecast - Africa
Actual - Americas
Forecast - Americas
Actual - Western Pacific
Forecast - Western Pacific
Actual - South-East Asia
Forecast - South-East Asia

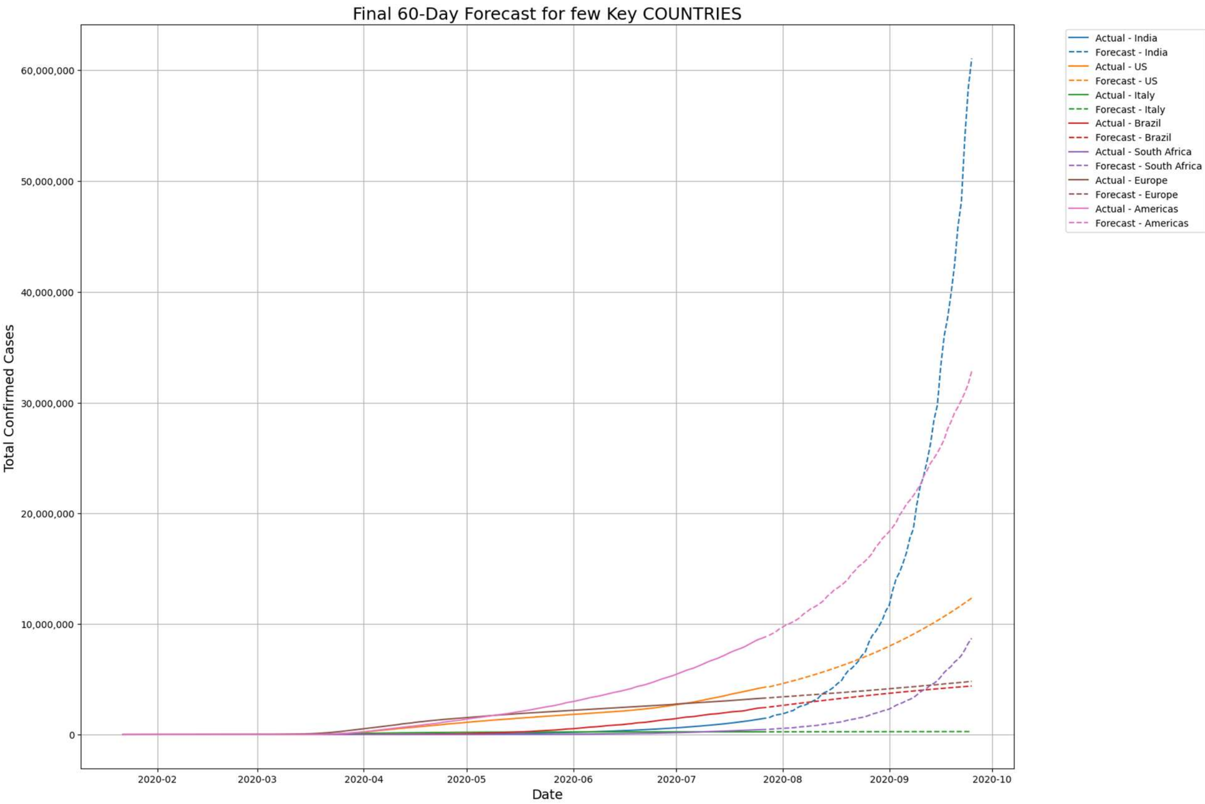


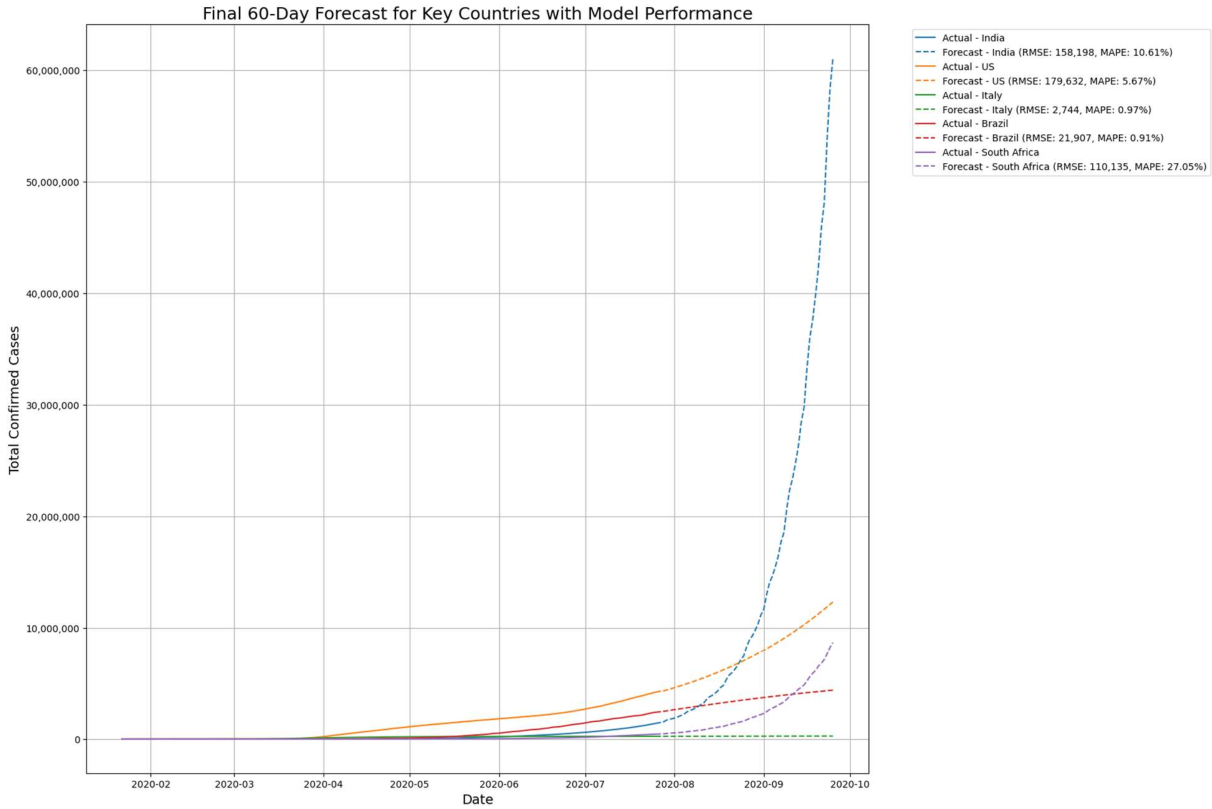
SARIMA Forecasts by WHO Region (Weekly Seasonality)
Total Confirmed Cases
Date
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
0
2,000,000
4,000,000
6,000,000
8,000,000
10,000,000
Actual - Eastern Mediterranean
Forecast - Eastern Mediterranean
Actual - Europe
Forecast - Europe
Actual - Africa
Forecast - Africa
Actual - Americas
Forecast - Americas
Actual - Western Pacific
Forecast - Western Pacific
Actual - South-East Asia
Forecast - South-East Asia



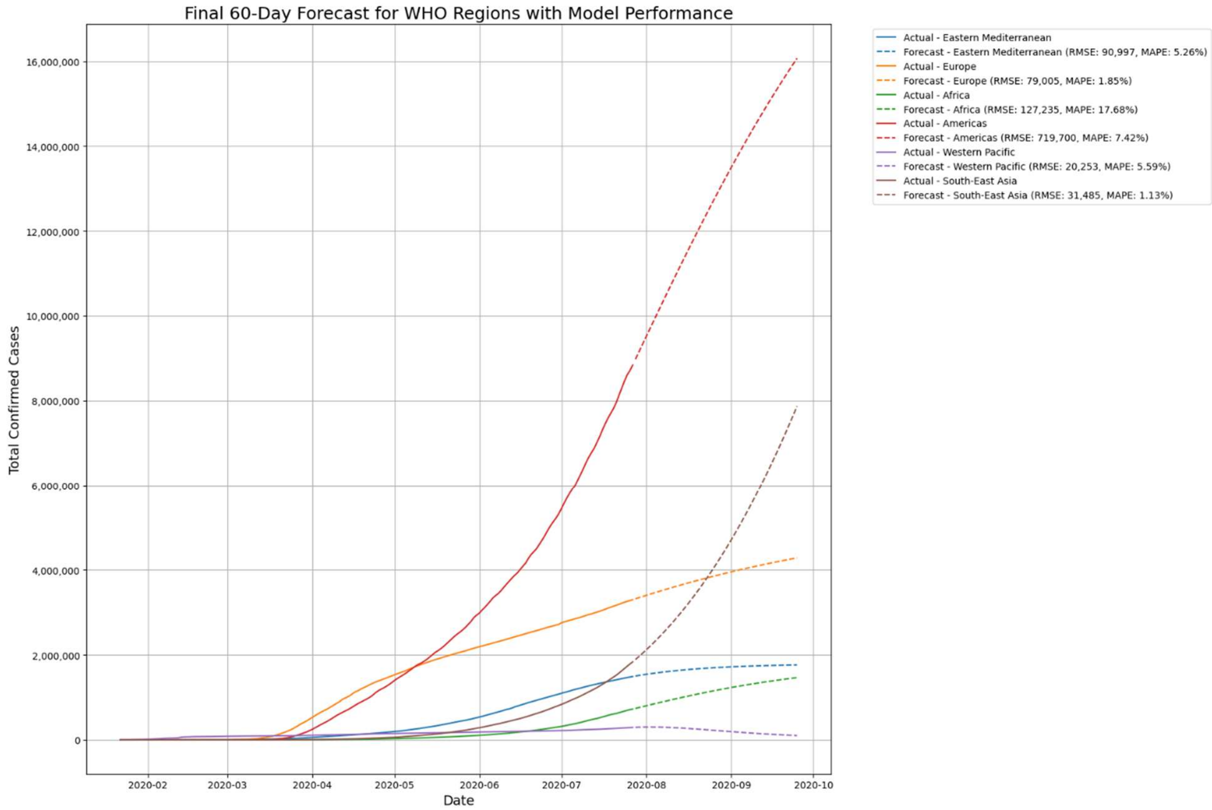
Visualisations: In Forecasting Results


Final 60-Day Forecast with Model Performance (RMSE & MAPE)
Total Confirmed Cases
Date
Actual - India
Forecast - India (RMSE: 158,198, MAPE: 10.61%)
Actual - US
Forecast - US (RMSE: 179,632, MAPE: 5.67%)
Actual - Italy
Forecast - Italy (RMSE: 2,744, MAPE: 0.97%)
Actual - Brazil
Forecast - Brazil (RMSE: 21,907, MAPE: 0.91%)
Actual - South Africa
Forecast - South Africa (RMSE: 110,135, MAPE: 27.05%)
Actual - Europe
Forecast - Europe (RMSE: 79,005, MAPE: 1.85%)
Actual - Americas
Forecast - Americas (RMSE: 719,700, MAPE: 7.42%)


Final 60-Day Forecast for few Key COUNTRIES
Total Confirmed Cases
Date
Actual - India
Forecast - India
Actual - US
Forecast - US
Actual - Italy
Forecast - Italy
Actual - Brazil
Forecast - Brazil
Actual - South Africa
Forecast - South Africa
Actual - Europe
Forecast - Europe
Actual - Americas
Forecast - Americas



Final 60-Day Forecast for Key Countries with Model Performance
Total Confirmed Cases
Date
Actual - India
Forecast - India (RMSE: 158,198, MAPE: 10.61%)
Actual - US
Forecast - US (RMSE: 179,632, MAPE: 5.67%)
Actual - Italy
Forecast - Italy (RMSE: 2,744, MAPE: 0.97%)
Actual - Brazil
Forecast - Brazil (RMSE: 21,907, MAPE: 0.91%)
Actual - South Africa
Forecast - South Africa (RMSE: 110,135, MAPE: 27.05%)



Final 60-Day Forecast for WHO Regions with Model Performance
Total Confirmed Cases
Date
Actual - Eastern Mediterranean
Forecast - Eastern Mediterranean (RMSE: 90,997, MAPE: 5.26%)
Actual - Europe
Forecast - Europe (RMSE: 79,005, MAPE: 1.85%)
Actual - Africa
Forecast - Africa (RMSE: 127,235, MAPE: 17.68%)
Actual - Americas
Forecast - Americas (RMSE: 719,700, MAPE: 7.42%)
Actual - Western Pacific
Forecast - Western Pacific (RMSE: 20,253, MAPE: 5.59%)
Actual - South-East Asia
Forecast - South-East Asia (RMSE: 31,485, MAPE: 1.13%)