Peer review history for:

Azeez A, Osuji G, Mutambayi R, Ndege J. Machine-learning forecasting model of tuberculosis cases among children in South Africa. S Afr J Sci. 2025;121(11/12), Art. #16658.
https://doi.org/10.17159/sajs.2025/16658

HOW TO CITE:
Machine-learning forecasting model of tuberculosis cases among children in South Africa [peer review history]. S Afr J Sci. 2025;121(11/12), Art. #16658. https://doi.org/10.17159/sajs.2025/16658/peerreview

Reviewer 1: Round 1

Not openly accessible under our Publishing peer review reports policy.

Reviewer 2: Round 1
Date completed: 02 July 2024
Recommendation: Accept / **Revisions required** / Resubmit for review / Resubmit elsewhere / Decline / See comments
Conflicts of interest: None

Does the manuscript fall within the scope of SAJS?
Yes/No
Is the manuscript written in a style suitable for a non-specialist and is it of wider interest than to specialists alone?
Yes/No
Does the manuscript contain sufficient novel and significant information to justify publication?
Yes/No
Do the Title and Abstract clearly and accurately reflect the content of the manuscript?
Yes/No
Is the research problem significant and concisely stated?
Yes/No
Are the methods described comprehensively?
Yes/**No**
Is the statistical treatment appropriate?
Yes/No/Not applicable/Not qualified to judge
Are the interpretations and conclusions justified by the research results?
Yes/Partly/No
Please rate the manuscript on overall contribution to the field
Excellent/**Good**/Average/Below average/Poor
Please rate the manuscript on language, grammar and tone
Excellent/Good/**Average**/Below average/Poor
Is the manuscript succinct and free of repetition and redundancies?
Yes/No
Are the results and discussion confined to relevance to the objective(s)?
Yes/No
The number of tables in the manuscript is
Too few/**Adequate**/Too many/Not applicable

The number of figures in the manuscript is

Too few/**Adequate**/Too many/Not applicable

Is the supplementary material relevant and separated appropriately from the main document?

Yes/No/**Not applicable**

Please rate the manuscript on overall quality

Excellent/**Good**/Average/Below average/Poor

Is appropriate and adequate reference made to other work in the field?

Yes/No

Is it stated that ethical approval was granted by an institutional ethics committee for studies involving human subjects and non-human vertebrates?

Yes/No/Not applicable

If accepted, would you recommend that the article receives priority publication?

Yes/No

Are you willing to review a revision of this manuscript?

Yes/No

With regard to our policy on 'Publishing peer review reports', do you give us permission to publish your anonymised peer review report alongside the authors' response, as a supplementary file to the published article? Publication is voluntary and only with permission from both yourself and the author.

Yes/No

Comments to the Author:

This is a very interesting study that has the potential to contribute towards the use of technology to improve the prediction of various health conditions in South Africa.

Please address all 15 comments added to the attached manuscript.

Please also proofread the document to ensure that all grammatical errors are rectified.

[See Appendix 1 for Reviewer 2's comments made directly on the manuscript]

Author response to Reviewer 2: Round 1

1. Need to briefly explain what machine learning algorithms are and how they are able to "predict incidence rates, improve diagnostics and close the treatment gap in childhood TB". Provide a simplified explanation of this.

AUTHOR: Thank you. We've added a simplified explanation of machine learning algorithms, highlighting how they help in predicting TB incidence rates, improving diagnostics, and closing treatment gaps in childhood TB.

2. Explain what these features are, especially to a non-specialist audience.

AUTHOR: Thank you for your review. We have revised the sentence to include a simplified explanation of "chest X-ray features" to ensure clarity for a broader audience. The revised sentence now reads: "A new TB diagnostic algorithm was recently and conditionally approved by the WHO, with a sensitivity of 85% when using chest X-ray features (visible signs of lung conditions), or 30% specificity without them."

3. Provide examples of features that would be included in a TB prediction model.

AUTHOR: Thank you for your review. We have revised the sentence to include the features as suggested.

4. Additionally, explain what time series data is and why this applies to your study.

AUTHOR: Thank you for your comment. We believe that explaining what time series data is, may not be necessary for this context, as the focus is on using ML approach in the TB prediction model. The inclusion of time series data and its relevance to our study is implicit in the forecasting approach we've adopted. However, if further clarification is needed, we would be happy to expand on this point.

5. Although the author highlights a few machine learning algorithms used in forecasting, they have not provided an overview of previous works relating to algorithms used in predicting/forecasting TB. Please provide this information in the background section.

AUTHOR: Thank you for your review. We have revised and added previous studies to the background section as suggested in line 9-13.

6. Please justify the selection of these models in your study.

AUTHOR: Thank you for your review. We have revised and added the justification and inclusion of those models in our study as suggested in line 50-54. Highlighted in yellow.

7. Please indicate whether 120 observations were the total number of available cases, or this is the shortlisted number of cases. If this is not the total available number (form ERT.NET), please explain the process used to select these 120.

AUTHOR: The 120 observations refer to monthly data points spanning from January 2010 to December 2019, covering a 10-year period. These 120 observations represent the total number of available monthly data points for TB incidence over this timeframe, not a shortlisted number of cases.
The data was not selected or filtered from a larger dataset; rather, it includes all the available monthly incidence data for TB from 2010 to 2019, as per the data provided by ERT.NET. Each observation corresponds to the total number of TB cases reported in a given month during this period, which was used as the basis for the time series analysis and model forecasting.

8. Please provide references to support this.

AUTHOR: Thank you for your review. We have added a citation to justify the statement.

9. Would it not be more appropriate for the train and test dataset to comprise of data from the same time period? Please explain the reason for your selection.

AUTHOR: Thank you. Splitting the dataset this way reflects the typical practice in time series forecasting, ensuring the model is tested on unseen future data while learning from a representative historical dataset. This helps to evaluate its real-world predictive capability.

10. Why are these the only metrics used to compare the models? Please provide references to justify this selection.

AUTHOR: Thank you. These metrics are widely used in forecasting studies to provide a well-rounded assessment of model performance. Their selection ensures that we capture different aspects of prediction accuracy, making them suitable for evaluating the models in our study.

11. Create separate graphs with their own legend for each category i.e. single models vs combined models.

AUTHOR: Thank you for your suggestion. We have separated the figure into figure 3a and figure 3b with each having different legend.

12. Ensure the labels on the legend are displayed in full and not cut-off.

AUTHOR: We have attended to it as you suggested. All the labels are displayed in full.

13. Use different colours to represent each model and update the legend to indicate this.

AUTHOR: Thank you for your suggestion. We have updated the legend but stick to the colours.

14. Rephrase this paragraph to improve its clarity.

AUTHOR: Thank you for your suggestion. We have rephrased it accordingly.

15. Compare your findings to other studies that used machine learning models to predict TB cases. This will help to highlight the contribution of this study.

AUTHOR: Thank you for your suggestion. We have included some studies that use ML model to predict TB cases.

Reviewer 3: Round 1

Not openly accessible under our Publishing peer review reports policy.

Reviewer 1: Round 2

Not openly accessible under our Publishing peer review reports policy.

Appendix 1: Reviewer 2's comments on manuscript (Round 1)

Machine Learning Forecasting Model of Tuberculosis Cases Among Children in South Africa.

Abstract

Globally, children and young adolescents under 15 years old constitute approximately 11% of all Tuberculosis (TB) cases, with a growing concern over TB infections in children under 5 years old, especially in resource-limited settings. Nonetheless, the true extent of TB burden among children remains inadequately explored in South Africa. The application of Random Forest-Bayesian ARIMA (RF-BARIMA) model for infectious disease prediction has not been previously employed. In this study, we employed the RF-BARIMA model to forecast TB incidences among children under 5 years old in South Africa's Eastern Cape Province, spanning from 2010 to 2019. Comparative analysis demonstrated that the RF-BARIMA model outperformed other models in predictive accuracy and forecast capability. The study's predictions revealed a projected mean of 0.4122 TB cases per month in 2022, with an effective sample size of 4053.488 for TB cases in Eastern Cape Province. These findings indicate a prospective reduction of 1670.85 mean number of TB cases among children under 5 years old in the coming years. The RF-BARIMA model offers enhanced predictive and forecast accuracy in comparison to the single Bayesian ARIMA model. These results provide compelling evidence of significant under-reporting and potentially elevated TB incidence among children under 5 years old in South Africa's Eastern Cape Province, raising important implications for public health policy and intervention strategies.

Keywords: ARIMA model, Bayesian, Random forest, Machine learning, Tuberculosis

Significance:

Childhood tuberculosis (TB) in South Africa is a significant concern, with the majority of cases occurring in children aged 0-4 years. The burden in children mirrors the high burden of the adult epidemic in the country. This research explores the burden of TB among children in South Africa and introduces a novel predictive model that outperforms existing methods to model the good health, well-being and potential to significantly reduce TB cases among children under 5 years old. The result call for urgent public health policy and intervention strategies to address the under-reporting and elevated TB incidence in this vulnerable demographic, further reinforcing the study's global significance.

2. Introduction

In recent years, despite increased global awareness of the prevalence of pediatric tuberculosis (TB), the development of machine learning algorithms to enhance diagnostic methods has been limited. In 2021, approximately 1.2 million children were estimated to contract TB, but only a third of children aged 0 to 5 received proper care and were reported in national TB programs[1,2]. Although collecting and testing many samples may increase diagnostic yield and improve diagnosis, but implementation at the primary care level, where the need for adequate diagnostic tools is most significant, remains challenging [3,4]. Algorithms can be used to predict incidence rates, improve diagnostics and close the treatment gap in childhood TB [5,6]. A new TB diagnostic algorithm was recently and conditionally approved by the WHO, has a sensitivity of 85% with chest X-ray features or 30% specificity without them[7]. However, the algorithm still requires further validation and training for accurate interpretation.

Various forecasting models are employed for infectious diseases, including grey prediction [8,9], exponential smoothing prediction [10], dynamic model [11], Box-Jenkins [12,13], and others. When using the forecasting method to attain accurate prediction, models are developed based on the features of the time series. When the prediction effect of a single model is not optimal, many studies choose the combination model prediction approach [14–17]. The combined model can absorb the advantages of two or more methods in order to achieve greater forecast accuracy.

Bayesian inference is a widely used method for analysing conditional probabilities of events, such as predicting hierarchy time series data [18], seasonality [19,20], and multi-step-ahead prediction [21]. It is also applied in general estimation, prediction task [22], and statistical analysis [23]. In time series forecasting, Bayesian method can be employed to forecast using a Kalman filter and smoothing technique and the Markov chain Monte Carlo (MCMC) method [24]. One study demonstrated the effectiveness of Bayesian networks in accurately predicting clinical parameter time series data [25]. This not only improved computational efficiency but also

optimized the modelling process. Additionally, dynamic Bayesian networks, which combine Kalman filtering models and echoing neural networks, have been utilized to predict multi-step-ahead time series data [21]. Furthermore, researchers like Panagiotelis et al. have proposed the use of Bayesian density techniques, specifically multivariate skewed t-distributions, for forecasting intraday electricity price [26].

Accurately predicting microbiologically confirmed cases of tuberculosis (TB) in young children who are suspected of having the disease is crucial for targeted clinical decision-making and future advancements in diagnostic research initiatives. To achieve this, we developed and evaluated nine machine-learning-based classification models. These models aimed to predict the incidence of microbiologically confirmed TB in children under the age of five in the Eastern Cape, South Africa. The primary objective of this study was to assess whether machine learning algorithms could effectively predict microbiological confirmation in paediatric TB patients. During the model selection process, we carefully analysed various model metrics to evaluate and compare the performance of these machine learning models

3. Methods

The study employs nine machine learning time series prediction models is to perform, evaluate, and compare single and hybrid ML models on the incidence of TB. These models are: Auto-ARIMA models, ARIMA with XGBoost error (Boosted ARIMA) model, Error-Trend-Season (ETS) with exponential smoothing state space model, PROPHET model, Time series linear regression model, Bayesian ARIMA model, Naïve Random Walk (NRW) model, Multivariate Adaptive Regression Spline (EARTH) model. The hybrid models proposed are: RF-Auto-ARIMA models, RF-XGBoost model, RF-ETS model, RF-PROPHET model, RF-LM model, RF-BARIMA model, RF-NRW model, RF-EARTH model.

3.1. Study Area and Data Source

This study focused on tuberculosis (TB) confirmed cases in children under the age of five in the Eastern Cape Province. The researchers collected data on monthly incident cases from January 2010 to December 2019, sourced from the Electronic Tuberculosis Register (ERT.NET) of Eastern Cape's Department of Health. A total of 120 observations were collected over a ten-year period. To ensure reliable and stable results, it's essential to have a minimum of 100 observations when creating predictive models. Therefore, the observations in this study were split into two sets: a training dataset comprising 108 observations (from January 2010 to December 2018) and a testing dataset consisting of 12 observations (from January to December 2019).

3.2. Ethical Approval

This study was approved by the Eastern Cape Department of Health (EC_20210_014) and the Research Ethic Committee, University of [details anonymised by journal administrator] to conduct research in the province.

3.3. Forecast models

1. Auto-ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) has two main forms: non-seasonal and seasonal ARIMA. The non-seasonal is expressed as (p, d, q), where 'p' signifies the autoregressive order, 'd' indicates the differencing order, and 'q' represents the moving average order. Conversely, the seasonal ARIMA model incorporates data seasonality and follows a similar process to non-seasonal ARIMA but considers seasonal patterns. To ensure the ARIMA model's effectiveness, it is crucial that the data exhibits stationarity, maintaining a constant mean and variance throughout the dataset, as described by the relation:

$$Y_i = \mu + \sum_{k=1}^{p} \phi_k \left(Y_{i-k} - \mu\right) + \varepsilon_t + \sum_{i=1}^{q} \theta_i \, \varepsilon_{i-i} \qquad t \in \mathbb{Z} \qquad (1)$$

where μ is the mean process, ε_t is a white noise process with mean zero and variance, $\sigma^2, \phi_p \neq 0$ and $\theta_q \neq 0$. The model is specified with the residual errors [23] as:

$$\varepsilon_i = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + \mu_i - \varphi_1 \mu_{t-1} - \cdots - \varphi_q \mu_{t-q}$$

where $\phi(L)\varepsilon_i = \varphi(L)\mu_i$ for a polynomial with the lag operator ($L^d X_t = X_{t-d}$). L is the lag operator, φ_i is the moving average parameters, p is the order of the lagged observation, d is the degree of difference, and μ_i is the white noise specified by ($\mu_i \sim Normal\ (0, \sigma^2)$). The time series predictors (Y_i) can be predicted by the autoregressive approach given as:

$$Y_i = (1 - L)^d X_t \text{ and}$$

$$\left(1 - \sum_{i=1}^{p} \emptyset_i L^i\right) Y_i = \left(1 + \sum_{i=1}^{q} \varphi_i L^i\right) \varepsilon_i$$

These concepts and equations were used in this study to forecast the values of TB incidence among under-5-year-old children for validation. However, AR=$\phi(z)$ and MA=$\theta(z)$ characteristics polynomials are expressed as: $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ and $\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q$. The difference is taken *d* times until the original series becomes stationary, which is known as "Integrated". In general, a *d*-th order difference can be written as: $Y_t' = (1 - B)^d Y_t$, where B is the backshift operator. However, we created a basic univariate auto-ARIMA machine learning model with a date time feature in the model to generate the ARIMA model.

2. XGBoost model

Extreme Gradient Boosting (XGBoost) is a robust machine learning system that employs gradient boosting decision tree algorithms. It assesses the significance of input features and predicts errors to make final predictions [27–29]. XGBoost has seen continuous refinement and enhancement by various researchers [30]. In practice, it often requires multiple iterations to obtain sufficient accuracy [31]. XGBoost is a powerful gradient boosting machine method [32,33]. The XGBoosting function can be written as:

$$XF^t = \sum_{i=1}^{n} l\left(Y_t \hat{Y}_t^{t-1} + f_t(x_i)\right) + \emptyset(f_t) + c$$

where Y_t is the observed value, $\hat{Y}_t^{t-1}$ represents the predicted value from the last iteration, f_t represents the new function for learning models, x_i is the feature vector, n is the sample size, $\emptyset(f_t)$ is the regularisation term that controls model complexity, and l is the loss function that measures the difference between the label and the prediction in the previous phase to produce the output of the new tree [29,34].

3. ETS model

The Error-Trend-Season (ETS) model uses exponential smoothing within a state space framework as a predictive models for the time series data. The algorithms used in the exponential smoothing techniques generated point forecasts from the state space model expressed as:

$$\text{Forecast equation: } \hat{y}_{t+1|t} = \ell_t$$

$$\text{Smoothing equation: } \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} = \ell_{t-1} + \alpha\varepsilon_t$$

where $\varepsilon_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$ is the residual at time t. The training data errors lead to the adjustment of the estimated level throughout the smoothing process. The model usually has a three-character string identification method using the framework terminology of Hyndman et al. (2002) [35]. The residuals training additive errors, ε_t is assumed to be normally distributed white noise with mean zero and variance σ^2 as:

$$y_t = \ell_{t-1} + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$$

$$\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim \text{NID}(0, \sigma^2)$$

Insert this into the error correction equations for Holt's linear method, we get::

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$

$$b_t = b_{t-1} + \beta\varepsilon_t$$

where β can be set as $\beta = \alpha\beta^*$. Based on the classification, the method is fully automatic and only requires arguments for ETS time series [36,37].

4. PROPHET model

The Prophet model, introduced by Facebook Inc. in 2017 [39], is designed for predicting time series data characterized by significant seasonal patterns and multiple seasons of historical data. [38]. Prophet is known for its resilience in handling missing data and effectively managing outliers [40]. This model breaks down the time series into three primary components.: the seasonal term S_t, the trend term T_t, and the residual term R_t:

$$y_t = S_t + T_t + R_t$$

In addition, to satisfy the needs of the actual scenario, the Prophet model integrates the effect of holidays $h(t)$:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t$$

where $g(t)$ describes a piecewise-linear trend (or "growth term"), $s(t)$ describes the various seasonal patterns, $h(t)$ captures the holiday effects, and ε_t is a white noise error term.

In the Prophet model algorithm used, the first step in this approach is modelling the time-series with specified parameters. The second step was to set the weekly and daily seasonal components to *True* to improve the prediction argument results fitted in the model. Then, the forecast is obtained, and the performance is evaluated.

5. Time Series Linear model

A Time series linear model (TSLM) used a *linear* algorithm function from the *parsnip* workflow to model the trend and seasonality of the data. The function fits the regression model and machine learning predictions using *tidyverse* principles [41]. The linear model assumes a linear relationship that exists between the forecast and predictor variables, expressed as:

$$y_t = X_t\beta + \varepsilon_t$$

This indicates that the errors must have a mean of zero, otherwise, the forecasts are biased. The residuals must not be autocorrelated; otherwise, the forecasts will be inefficient because the data contains more information that can be utilised. The residuals must be independent normal random variables with constant variance in order to create accurate inferences and prediction intervals. The trend is the slope of $y_t = X_t\beta + \varepsilon_t$. The model uses predictions to generate new values for independent features. These features lags are typically used in auto-regressive models.

6. Multivariate Adaptive Regression Spline (MARS) model

The MARS was created by modifying the algorithm process to use a workflow that standardises the pre-processing of the features of MARS machine learning model. The algorithm automatically creates a piecewise linear model that captures the nonlinear relationships in the data by assessing knots similar to step functions [42]. The procedure evaluates each predictor from each of the data points as a knot and creates a linear regression model with the candidate feature(s). Considering non-linear, non-monotonic data where $Y_t = f(X_t)$, the MARS procedure looks for a single point across the range of X_t values where two different linear relationships exist between Y_t and X_t achieve the smallest error (smallest SSE). This process continues until several knots are built to produce highly non-linear predictions, and the knots were used to fit a better relationship with our training data. After the full set of knots were established and identified, we removed those that did not significantly contribute to predictive accuracy.

7. Random Walk (Naive)

The random walk model in this study is based on an algorithm that assumes a variable, Y_t takes a random step away from its previous value in each time period. These steps are independently and identically distributed in size (i.i.d.), represented as $Y_t = Y_{t-1} + w_t$, where w_t is a discrete white noise series. To process the variable, the first difference is calculated, and the model's

mean is applied. Additionally, a backward shift operator B is applied to the random walk as follows:

$$Y_t = \mathrm{B}Y_t + w_t = Y_{t-1} + w_t$$

And stepping back further:

$$Y_{t-1} = \mathrm{B}Y_{t-1} + w_{t-1} = Y_{t-2} + w_{t-1}$$

This process was repeated until the end of the time series to get:

$$Y_t = (1 - \mathrm{B} + \mathrm{B}^2 + \cdots)\, Y_t$$

$$\Longrightarrow\ Y_t = w_t + w_{t-1} + w_{t-2} + \cdots$$

We used a workflow interface for adding pre-processing of the time series data into the algorithm functions of the random walk model.

8. Bayesian_ARIMA model

In this context, an ARIMA model was created similarly to model 1, but with a Bayesian approach. The Bayesian structural regression algorithm was employed to specify a Bayesian structural time series model before fitting to effectively handles unobserved components within time series data and provides more accurate uncertainty estimates. This approach manages uncertainty more effectively by allowing the measurement of posterior uncertainty for individual components, control over component variances, and the incorporation of prior assumptions into the model. The model can be represented as follows:

$$Y_t = \mu_t + x_t\beta + S_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

$$\mu_{t+1} = \mu_t + \upsilon_t, \upsilon_t \sim N(0, \sigma_\varepsilon^2)$$

where μ_t is the unobserved trend, x_t represents a set of regressors, and S_t denotes seasonality. This technique does not rely on differencing, lags, or moving averages. The model was created using the fit function from a stan algorithm of *bsts* package workflow to pre-process the time series data.

3.4. Random Forest model and Bayesian_ARIMA Models

The proposed hybrid model used in this study was based on a combination of the Random forest (RF) model and all the previously mentioned ML models: random forest and ARIMA model (RF-ARIMA), random forest and ARIMA with XGBoost Errors model (RF-XGBoost), random forest and ETS model (RF-ETS), random forest and Prophet model (RF-PROPHET), random forest and Bayesian ARIMA model (RF-BARIMA), random forest and Naïve-Random walk model (RF-NRW), random forest and Multivariate Adaptive Regression Spline model (RF-EARTH), which makes the hybrid model an excellent choice for forecasting. Random forests are a bagged decision tree modification that creates a wide number of de-correlated trees to improve predictive performance. RF uses learning algorithm that achieves strong predictive performance with minimal hyperparameter tuning.

In this analysis, the first measure involved permuting the TB data and recording the prediction error on the *out-of-bag* portion of the data using Mean Squared Error (MSE) for each tree in the regression classification. Mean Squared Error and Variance explained were then calculated using *out-of-Bag-Error* Estimation. The model utilized two-thirds of the training data for training and the remaining portion for validating the trees, in which a total of 500 trees were generated. The results showed a 72.9% increase in MSE for the model's variable importance and an 82.7% variance explained by the model. The analysis included plotting the Error against the number of trees and also the absolute residuals values against the probability distribution of random variables (Figure 1). It was observed that as more trees were added and averaged, there was a decreasing trend in the Error.

The study recorded errors using Mean Squared Error (MSE). Residuals from the Random Forest (RF) model were obtained and compared to actual TB cases, and accuracy measures were calculated. These residual values were then used to fit various hybrid models, including RF-ARIMA, RF-XGBoost, RF-ETS, RF-PROPHET, RF-BARIMA, RF-NRW, and RF-EARTH. Then, hybrid models were employed to forecast TB cases for the years 2020, 2021,

and 2022. Finally, the predictive accuracy of both single machine learning models and the hybrid machine learning model was compared to determine which model performed best in terms of predictive accuracy.

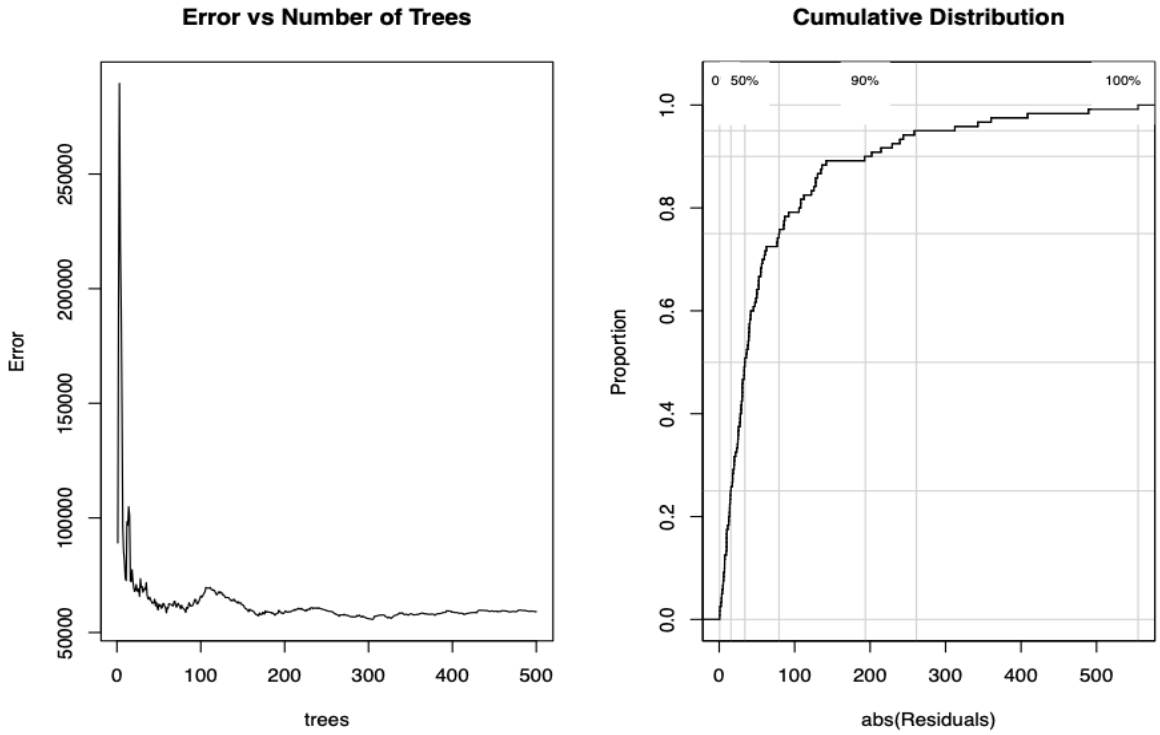

Figure 1: Plot of Error vs number of Trees (Left) and Cumulative Distribution of residuals (right)

3.5. Accuracy metrics and Model Evaluation

In our study, we assessed the true model accuracy by comparing predicted values to actual values, using a range of performance metrics to precisely measure model accuracy. To comprehensively evaluate predictive performance and determine the best model, we employed three specific parameter metrics: Mean Absolute Error (MAE), Mean Absolute Scaled Error (MASE), and Root Mean Square Error (RMSE):

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

$$MASE = \frac{MAE}{MAE_{naive}} = \frac{\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}$$

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

In our assessment, we evaluated the accuracy of the model using a test dataset and then recalibrate to improve its forecasting accuracy across the entire dataset.

3.6. Data analysis

The statistical analyses were conducted utilizing various packages in R software (Version 4.1.0). These packages offered the essential tools to facilitate robust model-building and ensure the most accurate model configuration. Exploring the detailed procedures and codes used are accessible at https://github.com/.

4. Results

The temporal trend of TB incidents (Figure 2) indicates a discernible upswing in reported TB cases between 2016 and 2019. It is noteworthy that the confirmed case count experienced fluctuations across different months, displaying a distinct pattern influenced by both annual seasonality and overarching long-term trends. The data was partitioned for model training and evaluation, resulting into an 80/20 split ratio of training and testing subsets for optimal balance between model exposure and robust performance assessment.

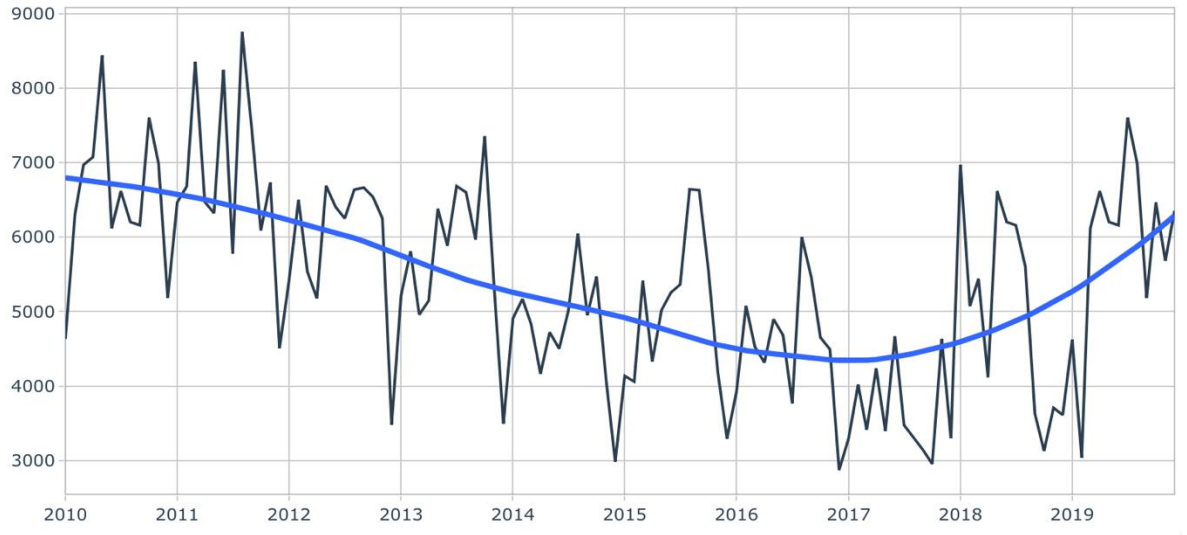

Figure 2: Monthly TB cases among under-5-year-old children between 2010-2019 from Eastern Cape Province, South Africa.

Comparison of model performance

The MAE, MASE, and RMSE were used for the comparison of models' performance. Among the single models, the Bayesian ARIMA model emerged as the best with lowest MAE (1606.95), MASE (1.45), and RMSE (1833.05) values compared to other models. In the hybrid models, the RF-BARIMA model was the best with lowest MAE (986.89), MASE (0.55), and RMSE (1205.45) values (Table 1). Comparing the accuracy performance of the single and hybrid models, the later showed a better performance. Specifically, the hybrid models showed a better parameter values: μ=0.4122±0.0497, σ=1284.3384±1.4460, ma=-0.7836±0.0009, and sma=-0.0509±0.0012 (Table S1a&b). This substantiates the ability of the hybrid approach in refining parameter estimations and ultimately bolstering forecasting precision.

Table 1: Comparison of Model Performance and Estimated Accuracy Metrics

	Single ML models				Hybrid ML Models			
Id	**Model description**	**MAE**	**MASE**	**RMSE**	**Model description**	**MAE**	**MASE**	**RMSE**
1	Auto-ARIMA	1777.48	1.60	1899.90	RF-Auto-ARIMA	998.04	0.55	1206.60
2	XGBOOST	1760.54	1.58	1993.35	RF-XGBOOST	1321.54	0.73	1484.09
3	ETS (A,N,A)	1921.11	1.73	2064.63	RF-ETS (A,N,A)	1315.27	0.73	1496.79
4	PROPHET	2290.01	2.06	2495.06	RF-PROPHET	1279.29	0.71	1441.30
5	LM	2304.57	2.07	2510.46	RF-LM	1285.67	0.71	1450.99
6	ARIMA	1831.37	1.65	2036.34	RF-ARIMA	1285.67	0.71	1450.99
7	BAYESIAN ARIMA	1606.95	1.45	1833.05	**RF-BAYESIAN ARIMA**	**986.89**	**0.55**	**1205.45**
8	NAÏVE-RANDOM-WALK	1692.81	1.52	1941.07	RF-NAÏVE-RANDOM WORK	1308.65	0.73	1488.40
9	EARTH	3086.13	2.78	3390.73	RF-EARTH	1215.37	0.67	1459.92

Figure 3 showed the forecast plot for the single models. The plot illustrates a consistent upward trajectory in TB incidence cases from the initial phase of the study's forecasting period in 2020 to the end of 2022. Remarkably, Model 7 emerged as the best performing model, boasting the most tightly bounded 80% confidence interval, which attests to the model's exceptional forecasting precision. Model 2 also demonstrated commendable performance, largely attributed to the judiciously specified parameters of its XGBoost components. Nevertheless,

the potential for heightened accuracy remains untapped, as a more refined calibration of parameters (tuning) could yield even better results.

Comparatively, Model 3 closely rivalled the efficacy of Model 2, albeit with a marginally broader test error confidence interval. Contrastingly, Models 4, 5, and 9 exhibited a tendency to overshoot the local trend due to the inherent linear trend components that failed to account for change points. Similarly, Models 1, 6, and 8 grappled with over-fitting the local trend, primarily stemming from the underexplored adjustment of the number of change points. These observations, succinctly depicted in Figure 3 (top), underscore the nuanced interplay of model components and parameters in shaping the accuracy and precision of TB incidence forecasts among children under five years old.

The hybrid models in Figure 3 (bottom) shows a subtle upward trend in TB incidence cases at the beginning of the forecasting period in 2020, succeeded by slight declines in TB cases during 2021 and 2022. Notably, when comparing the performance of the models in predicting the trajectory of TB incidence among children under five, minimal differences are observed.

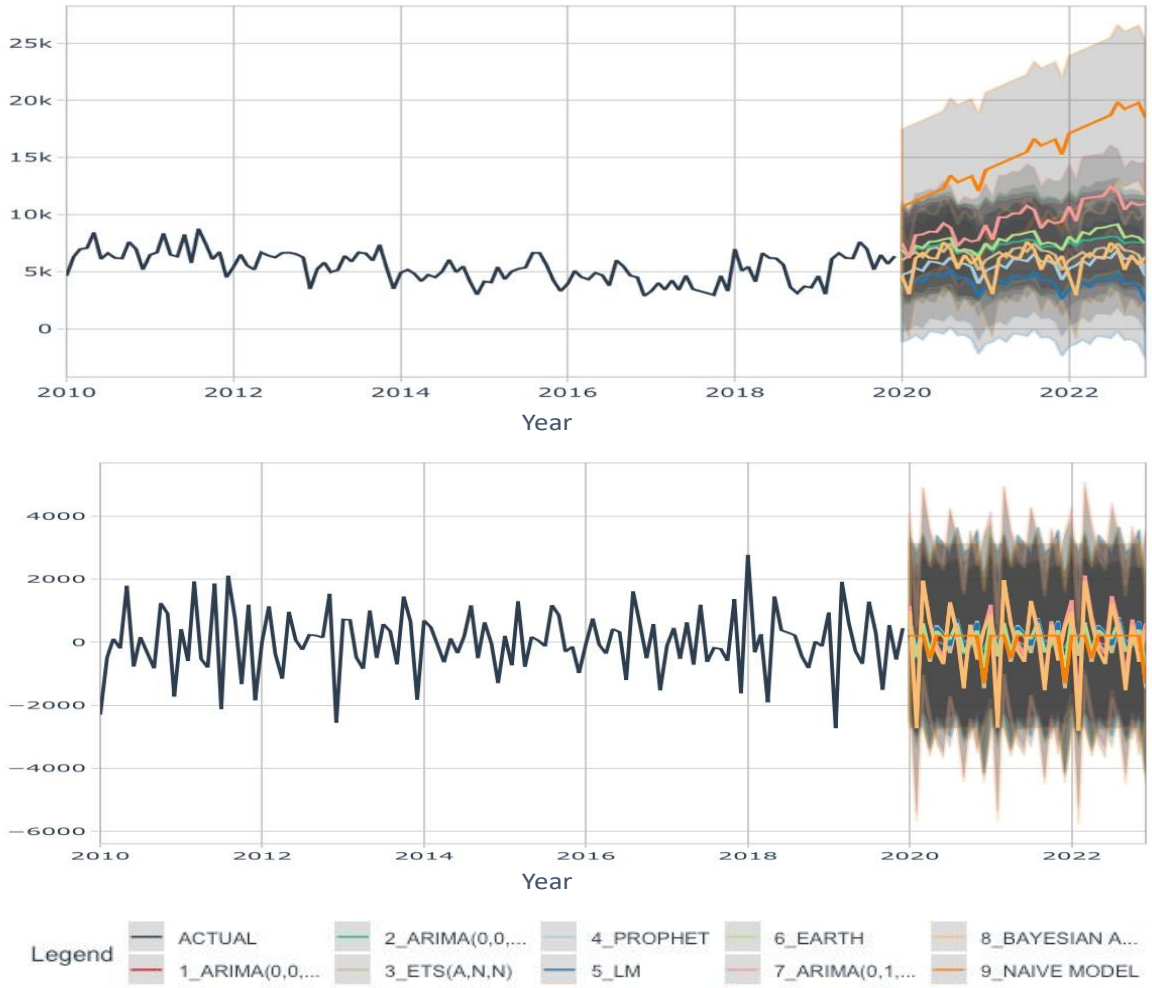

Figure 3: Comparison of all the model performance accuracy between the single models (up) and combined models (bottom) to forecast TB cases for the years 2020, 2021, and 2022.

The posterior predictive performance of the RF-BARIMA model outperformed that of the Bayesian ARIMA model, which was employed to fit the data and validate the compatibility of the fitted model with actual observations (Figure 4). The plot distinctly demonstrates the consistency and adequacy of the model's compatibility in describing the observed data across multiple years. The black dot signifies the distribution of observed outcomes, denoted as "y", while the array of blue lines represents the residual estimates derived from the posterior predictive distribution, labelled as "$\hat{y}$". Notably, the grey area delineates the expected 95% credible interval of the observations to fall within the predicted 95% credible intervals, provided that the model is aptly suited to depict the dataset's characteristics.

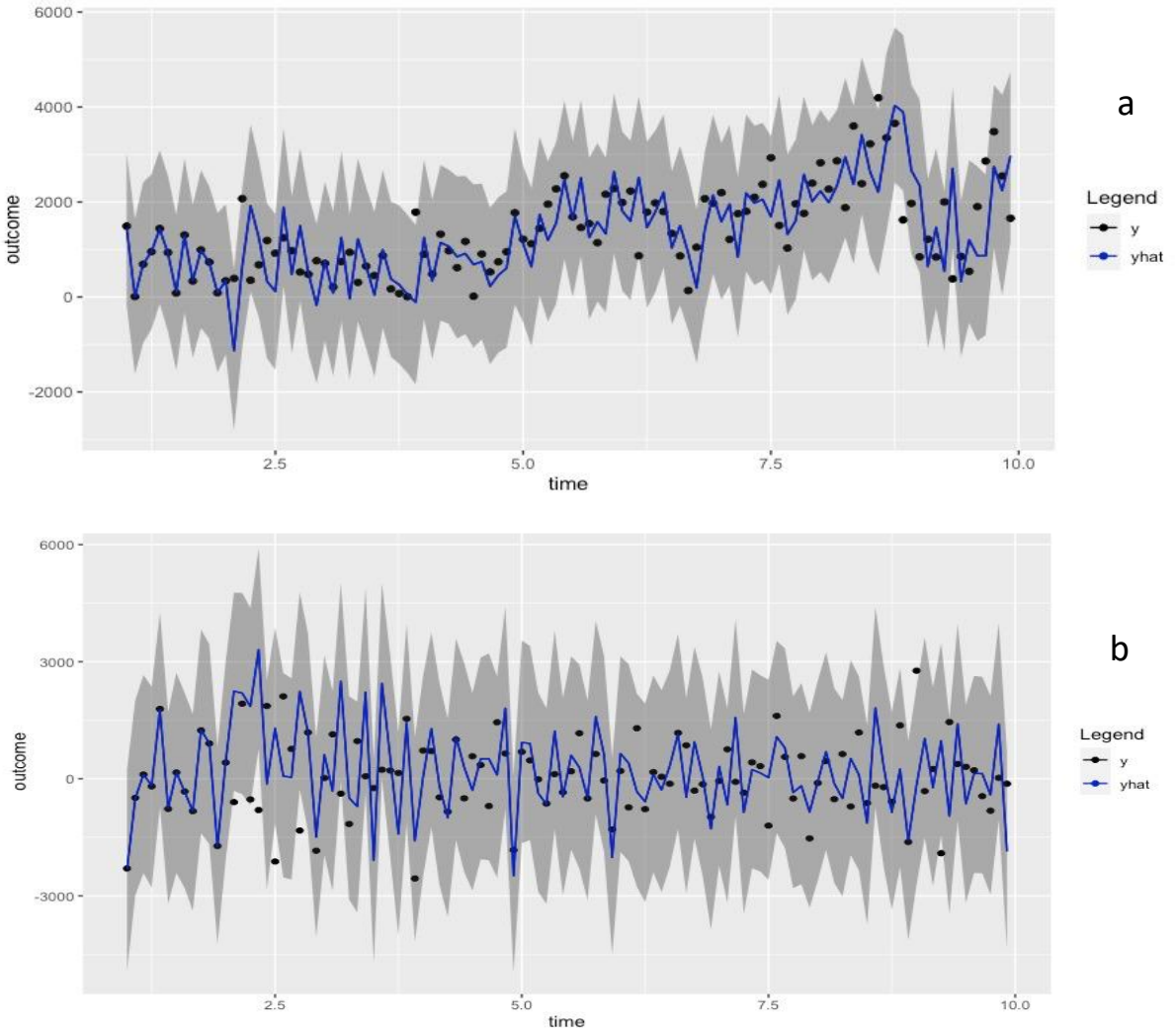

Figure 4: A posterior predictive check for the comparison between the fitted model predicts and the actual observed data for: (a)Bayesian ARIMA model and (b) Random forest-Bayesian ARIMA model.

In comparison of hybrid and single ML models, the RF-BARIMA model exhibited best accuracy in forecasting performance (Figure 5). The plot shows a decreasing temporal pattern in TB incidence cases at the outset of the 2020 forecast period, followed by an increase around mid-2021, and another decline towards the end of the forecasting year.

Figure 5: Comparison of Bayesian ARIMA model (top) and Random forest-Bayesian ARIMA model (bottom) performance accuracy to forecast TB cases for the years 2020, 2021, and 2022.

5. Discussion

The study revealed that, prior to the COVID-19 pandemic period, TB cases among under-5-year-old children had an uptrend and downtrend movement pattern from 2017 to 2019, using a trendline for pattern identification, which revealed a connecting series of lower highs (downtrend) or higher lows (uptrend) forming a level of resistance for future TB case pattern. The TB temporal analysis showed high reported rates of TB incidence among under-5-year-old children in this study, which can be explained by the prevalence of infectious TB in adults,

social interactions between adults and children, and the ventilation conditions in domestic settings [43–45].

The study found that while all models accurately predicted TB cases in children under five, the RF-BARIMA model performed the best. These results align with other research using hybrid models for forecasting infectious diseases, showing that hybrid models generally offer better predictive performance [14,46–48]. Infectious disease data often exhibits various linear and nonlinear features, making single models inadequate for modelling such data. Hybrid models have consistently proven to be the most suitable choice for estimating such complex data [49].

The estimated TB cases in Eastern Cape were 328 per 100,000 among children below 15 years old in 2020, and 7% were among the under-5-year-old children [50,51]. Furthermore, children are the most vulnerable, having a greater risk of getting infected with TB [52]. The study's forecast for 2022 predicts an average of 41.22% TB cases per month, based on a sample size of 4054 TB cases in Eastern Cape Province. This forecast indicates a future decrease of 1671 TB cases among children under 5 years old. These findings align with a similar study conducted in two Kenyan counties that projected TB cases among children under 15 years old, where the estimated TB cases were slightly lower than the general population in Kenya [46].

The findings of this study show a slight declining trend in TB incidence cases at the start of the forecasting period in 2020, followed by slight increases in TB cases in 2021 and 2022. These trends align with a WHO newsletter, which suggested that the number of TB cases could rise in 2021 and 2022, largely due to the impact of the COVID-19 pandemic [50]. This prediction encompasses not only the general population but also includes cases among children below 15 years of age.

The study reveals that TB incidence cases among children under the age of 5 follow a seasonal pattern, with peaks occurring in June, July, and August, followed by a decrease till in the year. This observation is consistent with some previous studies that have also found seasonal

variations contributing to TB infections [53,54]. However, it contrasts with other studies that did not find any association between TB infections and seasonal trends [55]. The study's findings suggesting a connection between TB infections in children and seasons can be explained by the impact of seasonal patterns on TB transmission. Children tend to spend more time indoors during the winter season, which is conducive to TB growth due to factors like overcrowding, increased humidity, and reduced air circulation.

Under-5-year-old children, especially during mid-winter, have been reported to have low vitamin D levels [56]. This deficiency could make these children more susceptible to TB infections during the winter season, potentially contributing to the observed seasonal pattern of TB infections. While children can contract TB at any age, it is most prevalent among those aged 1-4 years, likely due to their underdeveloped immune systems. This aligns with findings in South Africa's Western Cape Province, where the highest TB notification rate among children was reported in the 0-2 years age group [57]. Hospital studies indicate that extrapulmonary tuberculosis is more common in children than pulmonary tuberculosis, and misdiagnosis is a recurring issue [58].

Limitations

This study relied on data collected and reported through the Eastern Cape ERT.NET system, and therefore, the study had no control over the data's quality and accuracy. However, it was assumed that since the data had been submitted to the system, the health facilities in the Province had followed all relevant protocols to ensure data quality. Deep learning and machine learning algorithms typically require a large volume of data to learn effectively. Consequently, the provided dataset may not have been sufficient to allow the algorithm to learn more efficiently.

Conclusion

405 The RF-BARIMA model outperforms other models in predicting TB cases among children
406 under five. This study's findings highlight the persistent under-reporting of TB cases in this age
407 group, suggesting that the actual incidence might be higher than previously estimated. To
408 address this issue, there is a need to re-evaluate the TB monitoring framework data to identify
409 existing gaps and allocate resources to the National TB program urgently.
410 Additionally, the study reveals that TB infections among children under five are influenced by
411 seasonal patterns. This calls for increased investment in TB surveillance, screening, and
412 diagnosis efforts during specific months of the year to curb the spread of infection during peak
413 seasons.
414

References

1. Global Tuberculosis Report 2022. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022.

2. Marais, B. J. et al. Paediatric tuberculosis - new advances to close persistent gaps. Int. J. Infect. Dis. 113, S63-S67 (2021). https://doi.org/10.1016/j.ijid.2021.02.003

3. Walters, E. et al. Specimen Pooling as a Diagnostic Strategy for Microbiologic Confirmation in Children with Intrathoracic Tuberculosis. Pediatr. Infect. Dis. J. 38, e128-e131 (2019). https://doi.org/10.1097/INF.0000000000002240

4. Ioos, V., Cordel, H. & Bonnet, M. Alternative sputum collection methods for diagnosis of childhood intrathoracic tuberculosis: a systematic literature review. Arch. Dis. Child. 104, 629-635 (2019). https://doi.org/10.1136/archdischild-2018-315453

5. Marcy, O. et al. A Treatment-Decision Score for HIV-Infected Children With Suspected Tuberculosis. Pediatrics 144, e20182065 (2019). https://doi.org/10.1542/peds.2018-2065

6. Gunasekera, K. S. et al. Development of a Treatment-decision Algorithm for Human Immunodeficiency Virus-uninfected Children Evaluated for Pulmonary Tuberculosis. Clin. Infect. Dis. 73, e904-e912 (2021). https://doi.org/10.1093/cid/ciab018

7. World Health Organization. WHO operational handbook on tuberculosis: module 5: management of tuberculosis in children and adolescents. (2022).

8. Zhao, Y.-F., Shou, M.-H. & Wang, Z.-X. Prediction of the Number of Patients Infected with COVID-19 Based on Rolling Grey Verhulst Models. Int. J. Environ. Res. Public. Health 17, 4582 (2020).
https://doi.org/10.3390/ijerph17124582

9. Guan, P., Wu, W. & Huang, D. Trends of reported human brucellosis cases in mainland China from 2007 to 2017: an exponential smoothing time series analysis. Environ. Health Prev. Med. 23, 23 (2018).
https://doi.org/10.1186/s12199-018-0712-5

10. Zhang, Y.-Q. et al. Analysis and predication of tuberculosis registration rates in Henan Province, China: an exponential smoothing model study. Infect. Dis. Poverty 9, 123 (2020).
https://doi.org/10.1186/s40249-020-00742-y

11. Martínez-Bello, D. A., López-Quílez, A. & Torres-Prieto, A. Bayesian dynamic modeling of time series of dengue disease case counts. PLoS Negl. Trop. Dis. 11, e0005696 (2017).
https://doi.org/10.1371/journal.pntd.0005696

12. Wang, Y. et al. Temporal trends analysis of human brucellosis incidence in mainland China from 2004 to 2018. Sci. Rep. 8, 15901 (2018).
https://doi.org/10.1038/s41598-018-33165-9

13. Wang, Y. et al. An Advanced Data-Driven Hybrid Model of SARIMA-NNNAR for Tuberculosis Incidence Time Series Forecasting in Qinghai Province, China. Infect. Drug Resist. Volume 13, 867-880 (2020).
https://doi.org/10.2147/IDR.S232854

14. Azeez, A., Obaromi, D., Odeyemi, A., Ndege, J. & Muntabayi, R. Seasonality and Trend Forecasting of Tuberculosis Prevalence Data in Eastern Cape, South Africa, Using a Hybrid Model. IJERPH 13, 1-12 (2016).
https://doi.org/10.3390/ijerph13080757

15. Li, Z. et al. Application of a hybrid model in predicting the incidence of tuberculosis in a Chinese population. Infect. Drug Resist. Volume 12, 1011-1020 (2019). https://doi.org/10.2147/IDR.S190418

16. Wang, Y. et al. Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model. PLOS ONE 13, e0208404 (2018). https://doi.org/10.1371/journal.pone.0208404

17. Wang, Y. et al. Temporal trends analysis of tuberculosis morbidity in mainland China from 1997 to 2025 using a new SARIMA-NARNNX hybrid model. BMJ Open 9, e024409 (2019). https://doi.org/10.1136/bmjopen-2018-024409

18. Novak, J., McGarvie, S. & Garcia, B. E. A Bayesian Model for Forecasting Hierarchically Structured Time Series. (2017) doi:10.48550/ARXIV.1711.04738.

19. Zeng, Z. & Li, M. Bayesian median autoregression for robust time series forecasting. Int. J. Forecast. 37, 1000-1010 (2021). https://doi.org/10.1016/j.ijforecast.2020.11.002

20. Vosseler, A. & Weber, E. Forecasting seasonal time series data: a Bayesian model averaging approach. Comput. Stat. 33, 1733-1765 (2018). https://doi.org/10.1007/s00180-018-0801-3

21. Xiao, Q., Chaoqin, C. & Li, Z. Time series prediction using dynamic Bayesian network. Optik 135, 98-103 (2017). https://doi.org/10.1016/j.ijleo.2017.01.073

22. Rodriguez, A. & Puggioni, G. Mixed frequency models: Bayesian approaches to estimation and prediction. Int. J. Forecast. 26, 293-311 (2010). https://doi.org/10.1016/j.ijforecast.2010.01.009

507 23. Ganics, G. & Odendahl, F. Bayesian VAR forecasts, survey information, and structural
508 change in the euro area. Int. J. Forecast. 37, 971-999 (2021).
509 https://doi.org/10.1016/j.ijforecast.2020.11.001
510
511 24. Zhang, A. Y., Lu, M., Kong, D. & Yang, J. Bayesian Time Series Forecasting with Change
512 Point and Anomaly Detection. (2018).
513
514 25. Van Der Heijden, M., Velikova, M. & Lucas, P. J. F. Learning Bayesian networks for clinical
515 time series analysis. J. Biomed. Inform. 48, 94-105 (2014).
516 https://doi.org/10.1016/j.jbi.2013.12.007
517
518 26. Panagiotelis, A. & Smith, M. Bayesian density forecasting of intraday electricity prices
519 using multivariate skew t distributions. Int. J. Forecast. 24, 710-727 (2008).
520 https://doi.org/10.1016/j.ijforecast.2008.08.009
521
522 27. Zheng, Y. et al. A Learning-Based Model to Evaluate Hospitalization Priority in COVID-19
523 Pandemics. Patterns 1, 100092 (2020).
524 https://doi.org/10.1016/j.patter.2020.100092
525
526 28. Hu, C.-A. et al. Using a machine learning approach to predict mortality in critically ill
527 influenza patients: a cross-sectional retrospective multicentre study in Taiwan. BMJ Open 10,
528 e033898 (2020).
529 https://doi.org/10.1136/bmjopen-2019-033898
530
531 29. Lv, C.-X., An, S.-Y., Qiao, B.-J. & Wu, W. Time series analysis of hemorrhagic fever with
532 renal syndrome in mainland China by using an XGBoost forecasting model. BMC Infect. Dis.
533 21, 839 (2021).
534 https://doi.org/10.1186/s12879-021-06503-y
535
536 30. Li, W., Yin, Y., Quan, X. & Zhang, H. Gene Expression Value Prediction Based on XGBoost
537 Algorithm. Front. Genet. 10, 1077 (2019).
538 https://doi.org/10.3389/fgene.2019.01077

539
540 31. Wang, T. et al. Prevalence of hemorrhagic fever with renal syndrome in Yiyuan County,
541 China, 2005-2014. BMC Infect. Dis. 16, 69 (2015).
542 https://doi.org/10.1186/s12879-016-1404-7
543
544 32. Shrivastav, L. K. & Jha, S. K. A gradient boosting machine learning approach in modeling
545 the impact of temperature and humidity on the transmission rate of COVID-19 in India. Appl.
546 Intell. 51, 2727-2739 (2021).
547 https://doi.org/10.1007/s10489-020-01997-6
548
549 33. Babajide Mustapha, I. & Saeed, F. Bioactive Molecule Prediction Using Extreme Gradient
550 Boosting. Molecules 21, 983 (2016).
551 https://doi.org/10.3390/molecules21080983
552
553 34. Nishio, M. et al. Computer-aided diagnosis of lung nodule using gradient tree boosting
554 and Bayesian optimization. PLOS ONE 13, e0195875 (2018).
555 https://doi.org/10.1371/journal.pone.0195875
556
557 35. Hyndman, R. J., Koehler, A. B., Snyder, R. D. & Grose, S. A state space framework for
558 automatic forecasting using exponential smoothing methods. Int. J. Forecast. 18, 439-454
559 (2002).
560 https://doi.org/10.1016/S0169-2070(01)00110-8
561
562 36. Hyndman, R. J., Akram, M. & Archibald, B. C. The admissible parameter space for
563 exponential smoothing models. Ann. Inst. Stat. Math. 60, 407-426 (2008).
564 https://doi.org/10.1007/s10463-006-0109-x
565
566 37. Hyndman, R., Koehler, A., Ord, K. & Snyder, R. Forecasting with Exponential Smoothing.
567 (Springer, 2008). doi:10.1007/978-3-540-71918-2.
568 https://doi.org/10.1007/978-3-540-71918-2
569

570 38. Xie, C. et al. Trend analysis and forecast of daily reported incidence of hand, foot and
571 mouth disease in Hubei, China by Prophet model. Sci. Rep. 11, 1445 (2021).
572 https://doi.org/10.1038/s41598-021-81100-2
573
574 39. Taylor, S. J. & Letham, B. Forecasting at Scale. Am. Stat. 72, 37-45 (2018).
575 https://doi.org/10.1080/00031305.2017.1380080
576
577 40. Rostami-Tabar, B. & Rendon-Sanchez, J. F. Forecasting COVID-19 daily cases using phone
578 call data. Appl. Soft Comput. 100, 106932 (2021).
579 https://doi.org/10.1016/j.asoc.2020.106932
580
581 41. Kuhn M & Vaughan D. parsnip: A Common API to Modeling and Analysis Functions.
582 https://parsnip.tidymodels.org/ (2023).
583
584 42. Boehmke, B. & Greenwell, B. M. Hands-On Machine Learning with R. (Chapman and
585 Hall/CRC, 2019).
586 https://doi.org/10.1201/9780367816377
587
588 43. Shanaube, K. et al. Annual Risk of Tuberculous Infection Using Different Methods in
589 Communities with a High Prevalence of TB and HIV in Zambia and South Africa. PLoS ONE 4,
590 e7749 (2009).
591 https://doi.org/10.1371/journal.pone.0007749
592
593 44. Wood, R. et al. Changing prevalence of tuberculosis infection with increasing age in high-
594 burden townships in South Africa. Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis.
595 14, 406-412 (2010).
596
597 45. Middelkoop, K., Bekker, L.-G., Morrow, C., Zwane, E. & Wood, R. Childhood tuberculosis
598 infection and disease: a spatial and temporal transmission analysis in a South African
599 township. South Afr. Med. J. Suid-Afr. Tydskr. Vir Geneeskd. 99, 738-743 (2009).
600

601 46. Siamba, S., Otieno, A. & Koech, J. Application of ARIMA, and hybrid ARIMA Models in
602 predicting and forecasting tuberculosis incidences among children in Homa Bay and Turkana
603 Counties, Kenya. PLOS Digit. Health 2, e0000084 (2023).
604 https://doi.org/10.1371/journal.pdig.0000084
605
606 47. Cao, S. et al. A hybrid seasonal prediction model for tuberculosis incidence in China. BMC
607 Med. Inform. Decis. Mak. 13, 56 (2013).
608 https://doi.org/10.1186/1472-6947-13-56
609
610 48. Zhou, L. et al. Using a Hybrid Model to Forecast the Prevalence of Schistosomiasis in
611 Humans. Int. J. Environ. Res. Public. Health 13, 355 (2016).
612 https://doi.org/10.3390/ijerph13040355
613
614 49. Chakraborty, T., Chakraborty, A. K., Biswas, M., Banerjee, S. & Bhattacharya, S.
615 Unemployment Rate Forecasting: A Hybrid Approach. Comput. Econ. 57, 183-201 (2021).
616 https://doi.org/10.1007/s10614-020-10040-2
617
618 50. World Health Organization. Global Tuberculosis Report 2022.
619 https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-
620 tuberculosis-report-2022 (2022).
621
622 51. Global Tuberculosis Report 2021. https://www.who.int/teams/global-tuberculosis-
623 programme/tb-reports/global-tuberculosis-report-2021.
624
625 52. Negin, J., Abimbola, S. & Marais, B. J. Tuberculosis among older adults - time to take
626 notice. Int. J. Infect. Dis. 32, 135-137 (2015).
627 https://doi.org/10.1016/j.ijid.2014.11.018
628
629 53. Kirolos, A. et al. Tuberculosis case notifications in Malawi have strong seasonal and
630 weather-related trends. Sci. Rep. 11, 4621 (2021).
631 https://doi.org/10.1038/s41598-021-84124-w
632

633 54. Bodena, D., Ataro, Z. & Tesfa, T. Trend Analysis And Seasonality Of Tuberculosis Among
634 Patients At The Hiwot Fana Specialized University Hospital, Eastern Ethiopia: A Retrospective
635 Study. Risk Manag. Healthc. Policy Volume 12, 297-305 (2019).
636 https://doi.org/10.2147/RMHP.S228659
637
638 55. Jaganath, D. et al. Seasonality of childhood tuberculosis cases in Kampala, Uganda, 2010-
639 2015. PLOS ONE 14, e0214555 (2019).
640 https://doi.org/10.1371/journal.pone.0214555
641
642 56. Graham, D., Kira, G., Conaglen, J., McLennan, S. & Rush, E. Vitamin D status of Year 3
643 children and supplementation through schools with fortified milk. Public Health Nutr. 12,
644 2329-2334 (2009).
645 https://doi.org/10.1017/S1368980008004357
646
647 57. Fares, A. Seasonality of tuberculosis. J. Glob. Infect. Dis. 3, 46 (2011).
648 https://doi.org/10.4103/0974-777X.77296
649
650 58. Datta, M. & Swaminathan, S. Global aspects of tuberculosis in children. Paediatr. Respir.
651 Rev. 2, 91-96 (2001).
652 https://doi.org/10.1053/prrv.2000.0115