International Journal of Applied Mathematics, Sciences, and Technology for National Defense
Volume 4, Issue 2, 107-120
e_ISSN: 2985-9352, p_ISSN: 2986-0776
https://www.journal.foundae.com/index.php/JAS-ND
DOI : https://doi.org/10.58524/app.sci.def.v4i2.1146

Lee-Carter–ARIMA hybrid approach and machine learning for mortality rate forecasting in the United States: Implications for national defense and population risk assessment

Vita Nuraini
UIN Syarif Hidayatullah Jakarta
INDONESIA

Mahmudi*
UIN Syarif Hidayatullah Jakarta
INDONESIA

Nina Fitriyati
UIN Syarif Hidayatullah Jakarta
INDONESIA

Madona Yunita Wijaya
UIN Syarif Hidayatullah Jakarta
INDONESIA

Irma Fauziah
UIN Syarif Hidayatullah Jakarta
INDONESIA

Article Info

Article history:

Received: March 16, 2025
Revised: May 6, 2026
Accepted: June 12, 2026

Keywords:

ANN
ARIMA
Forecasting
Lee-Carter
Random Forest

Abstract

Background: The Accuracy of mortality rate forecasting plays an important role in various decision-making processes in the life insurance sector, including determining premium amounts. In addition, it also contributes to assessing the readiness of human resources to support national defense, as well as to conducting risk assessments aimed at maintaining demographic stability. The United States mortality data was selected as the study object due to the availability of comprehensive and high-quality.
Aims: This study explores five hybrid approaches that combine stochastic models and machine learning, along with one non-hybrid approach to assess their potential to improve forecasting accuracy.
Method: In this study, the Lee-Carter–ARIMA, Lee-Carter–Random Forest, Lee-Carter–ANN, Lee-Carter–ARIMA–Random Forest, Lee-Carter–ARIMA–ANN, and ANN models were evaluated. These models were applied to mortality rate data from nine divisions in the United States (US), stratified by gender, using training data from 1966 to 2005 and test data from 2006 to 2015. The best model is determined based on the smallest Mean Absolute Percentage Error (MAPE) value while also considering the interpretability of the model.
Result: The study's results show that, across the number of divisions, the Lee-Carter–ARIMA–Random Forest model produces the smallest MAPE values most often. However, in terms of average MAPE, the Lee-Carter–ARIMA–ANN model performs better, with MAPEs of 9.66% for females and 9.28% for males. Furthermore, neither of these models yields a substantial improvement in predictive accuracy compared with the Lee-Carter–ARIMA model.
Conclusion: Considering the relatively small decrease in MAPE and the difficulty of interpreting machine learning models due to their black box nature, the Lee-Carter–ARIMA model demonstrates the best overall performance relative to the other models. Nevertheless, the Lee-Carter–ARIMA–Random Forest and Lee-Carter–ARIMA–ANN models show potential as alternative approaches that merit further investigation and may contribute to national defense planning and support the maintenance of demographic stability.

To cite this article: Nuraini, V., Mahmudi, Fitriyati, N., Wijaya, M. Y., & Fauziah, I. (2026). Lee-Carter–ARIMA hybrid approach and machine learning for mortality rate forecasting in the United States: Implications for national defense and population risk assessment. *International Journal of Applied Mathematics, Sciences, and Technology for National Defense*, 4(2), 107-120.

***Corresponding Author:**
Mahmudi, UIN Syarif Hidayatullah Jakarta, Indonesia, Email: mahmudi@uinjkt.ac.id

INTRODUCTION

Modeling and forecasting future mortality rates are one of the most important issues in life insurance, demography, and other social sciences (Chen & Khaliq, 2022). The high accuracy of the models used for forecasting is an important aspect that is continually improving, enabling predictions to support accurate decision-making, especially for insurance companies (Hong et al., 2021; Kaukuntla, 2021). Beyond the insurance sector, mortality rate forecasting has a significant impact on governments (Raftery & Ševčíková, 2023), particularly in supporting national planning for schools, hospitals, and other public services, as well as in risk management and national defense (Vollset et al., 2020). Therefore, a modelling approach capable of accurately predicting mortality rates is needed.

Various stochastic models have been widely used by insurance companies to model and forecast future mortality rates. One of the most well-known models is the Lee-Carter model (Lee & Carter, 1992). Several studies show that the Lee-Carter model provides fairly accurate and stable results (Lee & Miller, 2001; Zuo et al., 2025), however several studies show that the accuracy performance of the Lee-Carter model declines in certain cases. Therefore, a number of researchers have proposed various modifications and extensions, such as Lee-Carter–ARIMA, or machine learning based variations, to improve accuracy and flexibility in modeling diverse future mortality dynamics (Abdulkarim & Garko, 2015; Atsalakis et al., 2007; Camarda & Basellini, 2021; Hong et al., 2021; Sakr et al., 2017). Nevertheless, significant differences in performance are still observed depending on the evaluation metrics used in machine learning methods. This means that more complex machine learning models don't always produce higher prediction accuracy.

Along with the development of mortality forecasting methods, several studies have also conducted direct comparisons between stochastic models and machine learning methods (Nuraini & Fitriyati, 2025). For example, Bjerre (2022) and Qiao et al. (2023) compared tree-based machine learning models with traditional stochastic mortality models. Although various studies have compared the performance of stochastic and machine learning models, in-depth evaluations of the effectiveness of both in modeling and predicting mortality rates haven't been widely conducted (Nuraini & Fitriyati, 2025), particularly in hybrid model configurations. Therefore, a modeling approach is needed that not only compares the two methods separately but also combines them into a hybrid model. This study focuses on the development and performance evaluation of hybrid models that combine Lee-Carter, ARIMA, RF, and ANN models, as well as a single machine learning model. These models are Lee-Carter–ARIMA, Lee-Carter–Random Forest, Lee-Carter–ANN, Lee-Carter–ARIMA–Random Forest, Lee-Carter–ARIMA–ANN, and ANN developed to model and predict mortality rate data from nine divisions in the United States (US) from 1966 to 2015, for both males and females. The main contributions of this study are (1) comprehensively evaluating the effectiveness of the Lee-Carter–ARIMA hybrid model based on machine learning on multi-division mortality data, and (2) providing a critical analysis of the trade-off between model accuracy and interpretability by examining how hybrid and single models balance predictive performance and transparency in mortality forecasting.

METHOD

The entire modeling and forecasting process in this study uses mortality rate data from nine U.S. Census divisions: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific. The data were obtained from the U.S. State Life Tables database, available through the Harvard Dataverse in the U.S. Mortality Database repository (https://dataverse.harvard.edu/dataverse/usfmdb). This database provides central mortality rates for 24 age groups (0-110+) by gender for the period 1941-2022. In this study, the data used was limited to the period from 1941 to 2015 to avoid extraordinary events such as World War II (Elder et al., 2009), the Polio Epidemic (Spooner & Dattani, 2025), the Asian Flu Pandemic (Trotter Y. et al., 1959), the Measles Epidemic (Hinman et al., 1983), the Opioid Epidemic (Ayoo et al., 2020), and the Covid-19 Pandemic (World Health Organization, 2020), so that the estimation results are more stable and representative. All mortality data from 1966 to 2015 for females and males across all divisions were preprocessed before use in modeling and forecasting.

Determining Parameter Values in the Lee-Carter Model

In general, the Lee-Carter model models the logarithm of the mortality rate $m_{x,t}$ as a combination of age-specific constants that describe the general pattern of mortalitybased on age (a_x), sensitivity coefficients that indicate how quickly or slowly the mortality rate decreases at a certain age in response to changes in k_t (b_x), and the overall mortality rate index in year t (k_t). Mathematically, the Lee-Carter model equation is written in Equation (1):

$$\ln\left(m_{x,t}\right) = a_x + b_x k_t \tag{1}$$

Can be rewritten in the form of an Equation (2):

$$m_{x,t} = e^{a_x + b_x k_t} \tag{2}$$

The Lee-Carter model has limitations or constraints on its parameters, as written in Equation (3):

$$\sum_{x=x_1}^{x_p} \widehat{b_x} = 1 \text{ dan } \sum_{t=t_1}^{t_n} \widehat{k_t} = 0 \tag{3}$$

The first step in estimating the parameters of the basic Lee-Carter equation is to calculate tha parameter a_x, as shown in Equation (4):

$$a_x = \frac{1}{t} \sum_{t=t_1}^{t_n} \ln(m_{x,t}) \tag{4}$$

The parameters b_x and k_t are estimated through the Singular Value Decomposition (SVD) process of the matrix $\ln\left(m_{x,t}\right) - a_x$. The matrix $\ln\left(m_{x,t}\right) - a_x$ is assumed to be A, where A is an $m \times n$ matrix that has a factorization in the form of Equation (5):

$$SVD(A_{n\times m}) = U_{n\times n} D_{n\times m} V_{m\times m}^{T} \tag{5}$$

Where, U and V are orthogonal matrices, D has the same dimensions as A and contains singular values, and V^T is the transpose of matrix V.

Next, the parameters b_x and k_t are calculated based on Equation (6) and (7):

$$\widehat{b_x} = \frac{U_{x,1}}{\sum U_{x,1}} \tag{6}$$

$$\widehat{k_t} = V_{t,1} D_{1,1} \sum_{x=x_1}^{x_\omega} U_{x,1} \tag{7}$$

The entire SVD process is performed using the ‘svd’ package in RStudio and the Python programming language with the ‘numpy.linalg.svd()’ function from the ‘NumPy’ library. Next, the estimated results of the parameters a_x, b_x, and k_t are used in the modeling and forecasting process of mortality rates using a hybrid method, as will be explained in the next section and the entire modeling and forecasting process of all models is shown in Figure 1.

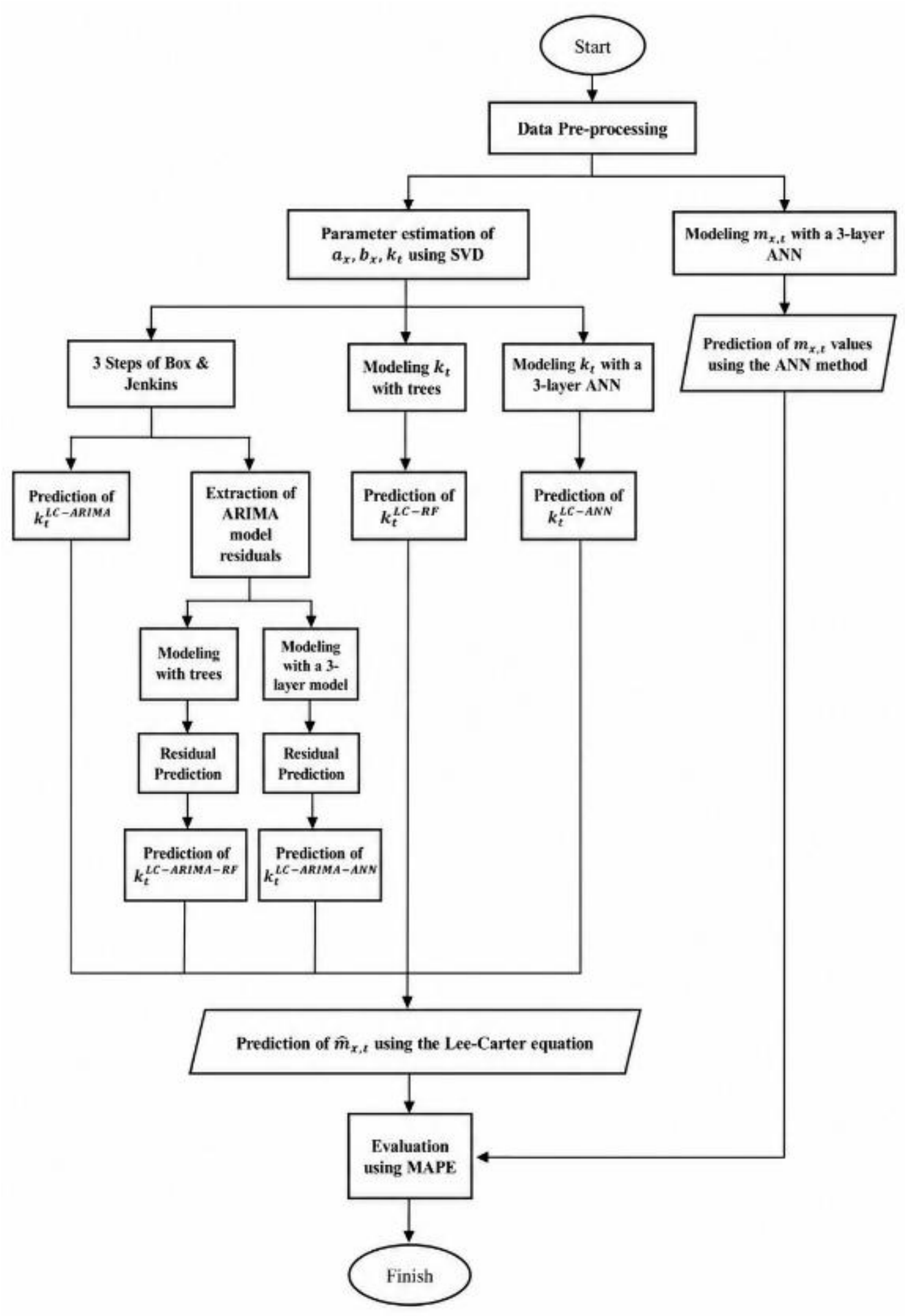

Figure 1. Flowchart of modeling and forecasting for the six models in this study.

Lee-Carter–ARIMA

The Lee-Carter–ARIMA model is a hybrid that combines the Lee-Carter stochastic approach with the Autoregressive Integrated Moving Average (ARIMA). In this model, the value of k_t, which represents the mortality trend over time, is then treated as a time series and analyzed using the ARIMA modeling is carried out in three iterative steps introduced by Box and Jenkins (Box et al., 2008) model identification, model estimation, and model diagnostic (Engle, 1982; Ljung & Box, 1978; Razali & Wah, 2011). Several models obtained through the three stages were then evaluated, and the best model was selected based on the fulfilment of model diagnostic assumptions and/or has the smallest Akaike Information Criterion (AIC) (Akaike, 1974) value will be used to predict k_t for the next few years. The predicted value of $\hat{k}_t$ is then substituted into Equation (1) along with the parameter values a_x and b_x to obtain the value of $\ln(\hat{m}_{x,t})$. This value is the exponentiated to obtain the predicted mortality rate $\hat{m}_{x,t}$.

Lee-Carter–Random Forest

Unlike the previous section, this section uses an alternative machine learning-based approach, namely Random Forest. In this study, a Random Forest regression model is used to predict the mortality trend component. The prediction results in regression cases use the average of all decision tree predictions (Probst & Boulesteix, 2018). Before performing the modeling process with Random Forest, the k_t value are first transformed into a time series data pattern, where the first values are used as input nodes and the fourth value represents the output node as the desired target value (Breiman, 2001). Next, the modeling and forecasting of k_t are carried out using the Python programming language with the help of the 'RandomForestRegressor' library from the 'sklearn.ensemble' module, based on the pattern that has been formed. The number of trees used in modeling varies from 50, 100, 150, to 400, resulting in 8 models. Each model is used to predict the value of k_t. Subsequently, each prediction result is processed further in the same way as the final stage of the Lee-Carter-ARIMA method to obtained the predictions of the mortality rate $\hat{m}_{x,t}$ for the coming years.

Lee-Carter–ANN (Artificial Neural Network)

After parameter k_t modeling was performed using the Random Forest approach, the Artificial Neural Network (ANN) method will be used in this section. The ANN method was chosen because the resulting model can approximate various forms of nonlinearity in the data and is known as a universal approximator (Zhang, 2003). Unlike the process in the Lee-Carter–Random Forest model, in the Lee-Carter–ANN model, the k_t value estimated goes through a normalization stage using the min-max method (Hong et al., 2021). After the normalization process, three processing unit layers need to be determined for ANN modeling. The determination of the input and output layers is carried out in the same way as in the random forest method. For the hidden layer, a single hidden-layer feedforward network is used (Zhang et al., 1998). The number of neurons in the hidden layer is determined based on rules of thumb (Hong et al., 2021). A trial-and-error process was conducted to evaluate the optimal number of neurons in the hidden layer to be used in k_t forecasting. Each model with a different number of neurons was tested for performance using the root mean square error (RMSE). Then, the optimal number of neurons, based on the smallest RMSE value, is used to model the normalized k_t. The ANN model in this study is implemented in Python using the 'MLPRegressor' library from the 'sklearn.neural_network' module. The ANN model obtained was then used to predict the normalized k_t value. The predicted $\hat{k}_t$ result was first denormalized before carrying out the final process as in the Lee-Carter-ARIMA method.

Lee-Carter–ARIMA–Random Forest

The Lee-Carter–ARIMA–Random Forest model is constructed through three main stages, namely mortality parameter estimation using the Lee-Carter model, k_t parameter modeling using ARIMA, and the use of Random Forest to improve prediction results through residual modeling. The residual modeling stage is carried out by defining the residuals from the value predicted by ARIMA ($\hat{k}_t^{ARIMA}$). These residuals are then modeled using the Random Forest with a procedure similar to modeling k_t in the Lee-Carter–Random Forest model. The residual prediction results are then added to the k_t value predicted using the ARIMA model ($\hat{k}_t^{ARIMA}$), resulting in the final k_t prediction that is a combination of the ARIMA and Random Forest methods can be expressed in Equation (8):

$$\hat{k}_t^{LC-ARIMA-RF} = \hat{k}_t^{ARIMA} + \hat{e}_t^{RF} \tag{8}$$

The predicted value $\hat{k}_t^{LC-ARIMA-RF}$ is then processed according to the final stage of the Lee-Carter-ARIMA method.

Lee-Carter–ARIMA–ANN

The Lee-Carter–ARIMA–ANN model was developed as an extension of the Lee-Carter–ARIMA model by adding an ANN to handle nonlinear patterns that the ARIMA model cannot capture. As in the Lee-Carter–ARIMA–Random Forest model, the residuals from the ARIMA modeling of the k_t parameter are used as input for the ANN for further modeling. The residual modeling process is carried out using the same steps as in the k_t value modeling in the Lee-Carter–ANN model. The sum of $\hat{k}_t^{ARIMA}$ and $\hat{\varepsilon}_t^{ANN}$ produce the value $\hat{k}_t^{LC-ARIMA-ANN}$, then processed according to the final stage of the Lee-Carter-ARIMA method.

Artificial Neural Network (ANN)

This section describes in detail the *Artificial Neural Network* (ANN) model used as a comparison model to the previous hybrid models. The ANN model in this study was built using the ‘MLPRegressor’ function from the ‘scikit-learn’ library in Python. The modeling process was carried out using the same steps as in the ANN method in Lee-Carter–ANN and Lee-Carter–ARIMA–ANN modeling, but the variables modeled were not the k_t parameter or the residuals from the ARIMA modeling process at k_t, but rather directly the $m_{x,t}$, which is the mortality rate at age x and year t. After building the ANN model, the value of $m_{x,t}$ is predicted for the coming years.

After the modeling process using various approaches is complete, the next important step is to evaluate how well each model represents mortality patterns and produces accurate predictions (Ospina et al., 2023). In this study, the Mean Absolute Percentage Error (MAPE) metric is used. Mathematically, MAPE can be defined as Equation (9):

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|\right) \times 100\% \tag{9}$$

Where, y_t represents the actual value in period t, $\hat{y}_t$ represents the model’s predicted value, and $t = 1,2,\dots,n$ represents the time index t with n being the total number of observations.

RESULTS AND DISCUSSION

This section presents the forecasting results obtained from the six model configurations. The comparison is based on the MAPE metric to assess the model's accuracy, while also considering the interpretability of the model.

Overview of Evaluation

This section presents the forecasting results from all six models and compares their performance. The evaluation was conducted using the Mean Absolute Percentage Error (MAPE) evaluation metric. MAPE is intuitive and easy to interpret, facilitating the communication of forecasting results (Makridakis et al., 1984). In addition, the scale-free nature of MAPE (Hyndman & Koehler, 2006) allows comparisons of accuracy levels across age groups or divisions without being affected by differences in data units and scales. However, MAPE is limited when actual values are close to zero, as it can produce very large or undefined error values (Ibrahim et al., 2021). However, in the context of mortality rate data, which generally have positive values and don’t approach zero, MAPE remains an appropriate metric for assessing the accuracy of mortality rate forecasting models. To ensure realistic testing, all models were trained using data from 1966 to 2005 and validated on data from 2006 to 2015.

In the data used in this study, the mortality rate is relatively high for the age group 0, then decreases for the age group 1-4. Subsequently, the mortality rate increases gradually up to the age groups 55-59 or 60-64, before eventually showing a sharp increase for the age groups 65-69 to 110+. In addition, when observed across each age group from 1966 to 2015, a declining pattern with fluctuations over time can be seen. This pattern indicates the presence of a linear component in the form of a long-term trend over time and a nonlinear component in the form of fluctuations (Goswami, 2019; Zhang, 2003), which forms the basis for developing hybrid models as well as the use of single machine learning models in this study.

All Model Performance

Table 1 presents the prediction results for five hybrid models: Lee-Carter–ARIMA, Lee-Carter–Random Forest, Lee-Carter–ANN, Lee-Carter–ARIMA–Random Forest, and Lee-Carter–ARIMA–ANN and a single machine learning model: ANN. The six models were trained for nine divisions, using annual mortality data from 1966 to 2005 and evaluated using data for the period 2006 to 2015 for females and males.

Table 1. MAPE values for Lee-Carter–ARIMA, Lee-Carter–Random Forest, Lee-Carter–ANN, Lee-Carter–ARIMA–Random Forest, and Lee-Carter–ARIMA–ANN by division and gender.

Divisions	Females	Males
New England		
Lee-Carter–ARIMA	10.164%	10.511%
Lee-Carter–Random Forest	13.326%	13.392%
Lee-Carter–ANN	22.481%	17.256%
Lee-Carter–ARIMA–Random Forest	10.144%	10.458%
Lee-Carter–ARIMA–ANN	10.435%	10.517%
ANN	14.275%	14.256%
Middle Atlantic		
Lee-Carter–ARIMA	6.036%	8.345%
Lee-Carter–Random Forest	10.504%	12.922%
Lee-Carter–ANN	17.904%	10.533%
Lee-Carter–ARIMA–Random Forest	6.086%	10.982%
Lee-Carter–ARIMA–ANN	6.036%	8.311%
ANN	11.330%	14.596%
East North Central		
Lee-Carter–ARIMA	9.427%	8.275%
Lee-Carter–Random Forest	11.076%	10.042%
Lee-Carter–ANN	14.258%	13.120%
Lee-Carter–ARIMA–Random Forest	9.488%	8.327%
Lee-Carter–ARIMA–ANN	9.416%	8.273%
ANN	9.788%	8.613%
West North Central		
Lee-Carter–ARIMA	10.685%	9.277%
Lee-Carter–Random Forest	11.065%	10.680%
Lee-Carter–ANN	11.357%	14.433%
Lee-Carter–ARIMA–Random Forest	9.999%	9.241%
Lee-Carter–ARIMA–ANN	10.336%	9.268%
ANN	6.941%	9.835%
South Atlantic		
Lee-Carter–ARIMA	8.432%	8.564%
Lee-Carter–Random Forest	11.198%	12.890%
Lee-Carter–ANN	16.458%	18.041%
Lee-Carter–ARIMA–Random Forest	9.561%	8.399%
Lee-Carter–ARIMA–ANN	8.432%	8.582%
ANN	11.751%	12.062%
East South Central		
Lee-Carter–ARIMA	11.534%	9.953%
Lee-Carter–Random Forest	12.774%	11.830%
Lee-Carter–ANN	13.357%	13.540%
Lee-Carter–ARIMA–Random Forest	11.203%	9.907%
Lee-Carter–ARIMA–ANN	11.390%	9.923%
ANN	9.754%	9.661%
West South Central		
Lee-Carter–ARIMA	9.016%	9.057%
Lee-Carter–Random Forest	10.908%	12.205%

Lee-Carter–ANN	13.290%	12.919%
Lee-Carter–ARIMA–Random Forest	9.514%	9.811%
Lee-Carter–ARIMA–ANN	9.476%	9.029%
ANN	10.503%	12.133%
Mountain		
Lee-Carter–ARIMA	12.348%	9.464%
Lee-Carter–Random Forest	13.442%	12.670%
Lee-Carter–ANN	16.469%	15.099%
Lee-Carter–ARIMA–Random Forest	12.140%	9.159%
Lee-Carter–ARIMA–ANN	12.097%	9.320%
ANN	13.512%	14.069%
Pacific		
Lee-Carter–ARIMA	8.880%	10.434%
Lee-Carter–Random Forest	13.511%	16.088%
Lee-Carter–ANN	19.304%	25.386%
Lee-Carter–ARIMA–Random Forest	8.775%	9.920%
Lee-Carter–ARIMA–ANN	8.855%	10.295%
ANN	14.243%	13.895%

In general, based on Table 1, the smallest MAPE value in each division was obtained from different models. In addition, differences in the best models across divisions are evident by gender. This shows that mortality data characteristics can vary by division and gender, so model selection for forecasting should consider division and gender factors to achieve more accurate predictions (Murphy et al., 2023; Wu et al., 2021). The MAPE values for the Lee-Carter–ARIMA, Lee-Carter–ARIMA–Random Forest, and Lee-Carter–ARIMA–ANN models don't show significant differences compared to other models. This indicates that adding machine learning components doesn't significantly improve accuracy across all divisions in this study. In other words, the performance of these hybrid models is comparable to the Lee-Carter–ARIMA model in this study. On the other hand, the MAPE values for the Lee-Carter–Random Forest and Lee-Carter–ANN models differ significantly from the other three models. These findings indicate that applying machine learning algorithms to the Lee-Carter component cannot improve performance, especially when the data structure exhibits small fluctuations, as in the mortality rate data used in this study.

Nevertheless, a more detailed analysis based on the number of regions shows a tendency for certain models to dominate. From Table 1, it can be seen that seven out of the 18 datasets have the lowest MAPE values under the Lee-Carter-ARIMA-Random Forest model. Furthermore, the Lee-Carter-ARIMA-ANN model ranks second, with five out of 18 datasets having the lowest MAPE values. The Lee-Carter-ARIMA dan ANN models each account for three out of 18 datasets with the lowest MAPE values. Meanwhile, none of the datasets achieve the lowest MAPE under the Lee-Carter-Random Forest and Lee-Carter-ANN models. These findings indicate that although the addition of machine learning components doesn't always significantly improve accuracy for all regions, the Lee-Carter-ARIMA-Random Forest and Lee-Carter-ARIMA-ANN models still demonstrate relatively strong and dominant performance in several regions.

Comparative Analysis of Models

To consolidate the research findings, Table 2 presents the average MAPE values for all models, split by gender.

Table 2. Average MAPE values for all model based on gender.

Model	Females	Males	Average
Lee-Carter–ARIMA	9.61354%	9.32001%	9.46678%
Lee-Carter–Random Forest	11.97812%	12.52423%	12.25117%
Lee-Carter–ANN	16.09768%	15.59187%	15.84477%
Lee-Carter–ARIMA–Random Forest	9.65672%	9.57824%	9.61748%
Lee-Carter–ARIMA–ANN	9.60821%	9.27986%	9.44403%
ANN	11.34407%	12.12448%	11.73428%

Considering the average MAPE by gender across all divisions, Table 2 shows that the smallest MAPE values were obtained by the Lee-Carter–ARIMA–ANN model, both for females and males, at 9.60821% and 9.27986%, respectively. In addition, the Lee-Carter–Random Forest, Lee-Carter–ANN, and ANN models have high average MAPE values, indicating that these models aren't suitable for modeling and forecasting mortality data in this study. Furthermore, the MAPE values of the Lee-Carter–ARIMA–Random Forest and Lee-Carter–ARIMA–ANN models are not significantly different from the average MAPE of the Lee-Carter–ARIMA model. This finding indicates that adding machine learning models doesn't substantially improve prediction accuracy. To complement the analysis comparing the prediction results of each model with actual data, Figures 2 – 4 are presented as illustrations that provide an overview of the comparison of each model with actual data for females in the 40-44 age groups in the East North Central division and Middle Atlantic, and males in the 40-44 age groups in the Middle Atlantic division.

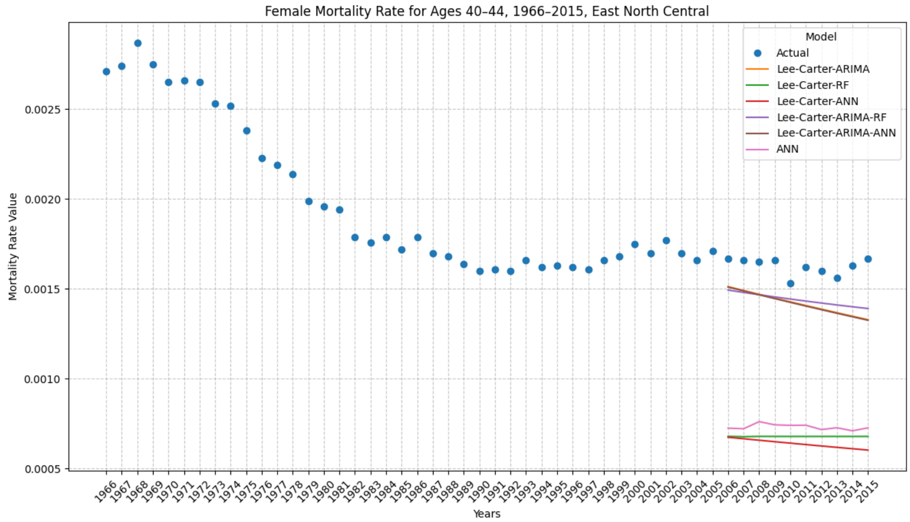

Figure 2. Female mortality rates from actual data and predicted results of all models for the 40-44 age group in the east north central division.

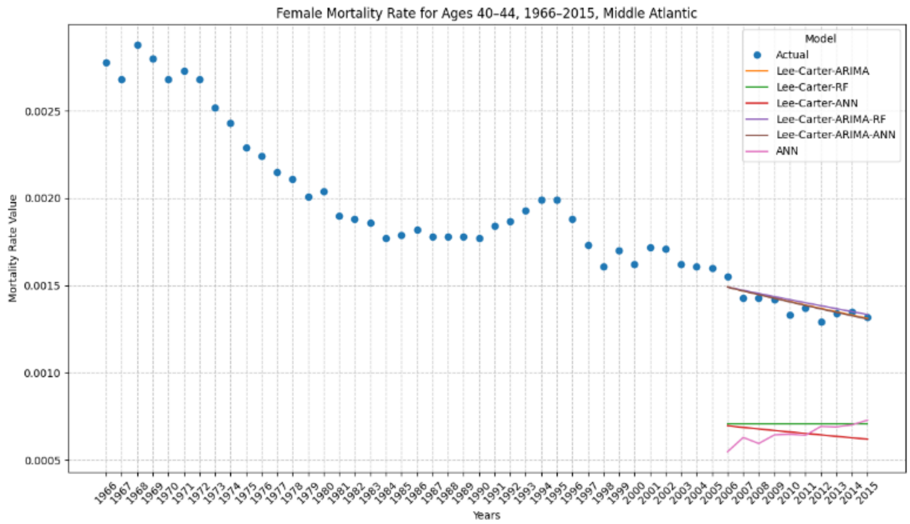

Figure 3. Female mortality rates from actual data and predicted results of all models for the 40-44 age group in the middle Atlantic.

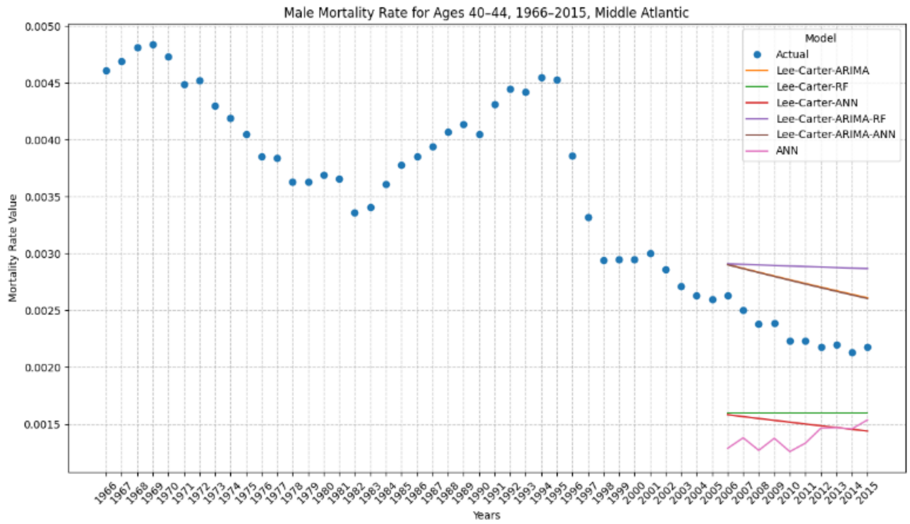

Figure 4. Male mortality rates from actual data and predicted results of all models for the 40-44 age group in the middle Atlantic division.

Furthermore, based on the illustrations in Figures 4 – 6, the prediction results from the Lee-Carter–ARIMA, Lee-Carter–ARIMA–Random Forest, and Lee-Carter–ARIMA–ANN models for both female and male data are quite close to the actual values. In contrast, the prediction results from the Lee-Carter–Random Forest, Lee-Carter–ANN, and ANN models have lower mortality rates than the actual mortality rates. Thus, in this study, the Lee-Carter–ARIMA, Lee-Carter–ARIMA–Random Forest, and Lee-Carter–ARIMA–ANN models are the three that can represent actual values with a relatively high level of accuracy compared to the other three models. However, for some divisions and age groups, there were patterns of underestimation and overestimation in the prediction results using the Lee-Carter–ARIMA, Lee-Carter–ARIMA–Random Forest, and Lee-Carter–ARIMA–ANN models, indicating that each division has different characteristics, so that the mortality prediction patterns formed aren't entirely the same between divisions.

Interpretation, Findings, and Limitations

A comparative evaluation using the MAPE metric confirms that integrating classical statistical models with machine learning architectures via a hybrid framework can yield lower MAPE values than single neural network-based models. However, the MAPE values of the Lee-Carter–Random Forest and Lee-Carter–ANN models are still not good enough to improve the accuracy of mortality data predictions in this study, which means that machine learning modeling on k_t obtained based on Lee-Carter model parameter estimates doesn't have a significant impact. Besides that, these two hybrid models are also not recommended for long-term planning and for use in policy forecasting, particularly in government and defense-related demographic planning and defense (Booth & Tickle, 2008).

In addition, it appears that the MAPE values of the Lee-Carter–ARIMA–Random Forest and Lee-Carter–ARIMA–ANN models aren't significantly different from the average MAPE value of the Lee-Carter–ARIMA model, although they're significantly different from Lee-Carter–Random Forest, Lee-Carter–ANN, and ANN models. This finding indicates that adding of machine learning models doesn't substantially improve prediction accuracy. This is likely due to the residual characteristics of the ARIMA model, which satisfies the assumption of homoscedasticity, so that it's possible that the non-linear pattern in the residual isn't very strong (Engle, 1982). Machine learning models, in this case Random Forest and ANN, generally excel at modeling non-linear patterns (Robinson et al., 2017; Zhang, 2003). Therefore, with residuals with non-linear patterns that aren't very strong, as in this study, their ability to improve accuracy is limited. Thus, despite an average increase in accuracy, the Lee-Carter–ARIMA–Random Forest and Lee-Carter–ARIMA–ANN models still do not improve mortality rate prediction accuracy in this study, indicating that machine learning on ARIMA residuals has little impact. Therefore, in general, given the relatively small decrease in MAPE and the difficulty of interpreting machine learning models due to their black-box nature (Zhang et al., 2022), the Lee-Carter–ARIMA model can be concluded to be the best model in this study.

In accordance with the findings of this study, the Lee-Carter-ARIMA method, as the simplest approach compared to other methods and not involving machine learning techniques, is more suitable for use in planning contexts related to national defense. This is due to its high accuracy in predicting mortality rates, which are directly linked to demography aspects (Cincotta, 2023). Such accuracy is essential in reducing the risk of prediction errors that may affect strategic decision-making, as well as in supporting human resource planning, population resilience, and risk mitigation in facing various future uncertainties (Hyndman & Athanasopoulos, 2018). The necessity of this accuracy stems from the fact that national security and social stability rely on precise mortality data, where even a 1% error margin can translate into thousands of lives lost or billions in defense budget deficits (Shrader et al., 2023). Ultimately, while overestimation leads to budgetary waste and inefficient personnel deployment, underestimation risks paralyzing defense responses, allowing threats to spread without adequate protection.

With research results showing the potential for developing Lee-Carter–ARIMA–Random Forest and Lee-Carter–ARIMA–ANN hybrid models to improve the accuracy of mortality rate data forecasting, this study provides practical value for the development of large-scale forecasting systems. Therefore, these models can be utilized to support evidence-based policymaking, including healthcare planning, disaster preparedness, and social security systems, all of which are closely related to national resilience and security. However, the interpretation of the findings still needs to consider a number of methodological and empirical limitations. The main limitation of this study is the use of models that focus solely on the Lee-Carter model and its development via a hybrid approach with ARIMA, Random Forest, and Artificial Neural Network (ANN), without incorporating cohort effects or modifications to other Lee-Carter models. Furthermore, another limitation lies in the data used, which are mortality rate data from nine divisions in the United States for the period 1966 – 2015, selected to avoid the influence of extraordinary events, thereby making the estimation results more stable and representative. Finally, modeling in this study focused on the time index k_t for the Lee-Carter–ARIMA, Lee-Carter–Random Forest, and Lee-Carter–ANN, the ARIMA model residuals against k_t for the Lee-Carter–ARIMA–Random Forest and Lee-Carter–ARIMA–ANN models, and the mortality rate for the ANN model with performance evaluation based on the MAPE values while also considering the interpretability of the model.

CONCLUSION

In this study, five hybrid and one non-hybrid artificial neural network models were proposed to model and predict mortality rates in nine U.S. divisions from 1966 to 2005, for both females and males. The analysis based on the number of divisions with the lowest MAPE values in the model shows the dominance of the Lee-Carter–ARIMA–Random Forest model (7 out of 18). In addition, different results were observed when viewed by the smallest average MAPE by gender. Among the models applied in this study, the Lee-Carter–ARIMA–ANN model had the lowest average MAPE, 9.65672% for females and 9.57821% for males. However, Lee-Carter–ARIMA–ANN and Lee-Carter–ARIMA–Random Forest models show MAPE values that aren't significantly different from the MAPE values of the Lee-Carter–ARIMA model. These findings indicate that adding machine learning doesn't significantly improve prediction accuracy, especially for data with relatively small fluctuations and ARIMA model residuals that meet the assumption of homoscedasticity, as is the case with the data used in this study. Therefore, given that the forecasts' accuracy didn't increase significantly and that interpreting machine learning models is difficult due to their black-box nature, the Lee-Carter–ARIMA model was the best-performing model in this study.

However, Lee-Carter–ARIMA–ANN and Lee-Carter–ARIMA–Random Forest models can be considered alternatives for predicting mortality rates in data with more complex non-linear patterns from various other countries, allowing the effectiveness of the models to be tested across different data characteristics. It is also important to develop machine learning models with a stronger pattern-capturing ability for mortality rate data and to modify the Lee-Carter model by adding cohort effects. Furthermore, the results of this study have implications for actuarial practitioners, suggesting the use of hybrid models to improve the accuracy of mortality projections in support of appropriate decision-making, especially in determining premium values. In addition, this study may contribute to national defense planning and the maintenance of demographic stability, as well as support state stability planning and evidence-based policymaking, particularly in areas closely related to national resilience and security. However, this study also confirms that increased model complexity doesn't always correlate directly with increased mortality forecasting accuracy.

AUTHOR CONTRIBUTIONS

V.N.: Conceptualization, Numerical Simulation, Visualization, Writing – Original Draft. M.: Data Curation, Methodology, Validation, Writing – Review and Editing. N.F.: Formal Analysis, Validation, Supervision, Writing - review & editing. M.Y.W.: Investigation, Formal Analysis. I.F.: Investigation, Formal Analysis. All authors discussed the results and contributed to the final manuscript.

CONFLICT OF INTEREST

The authors declare no competing interest.

REFERENCES

Abdulkarim, S., & Garko, A. (2015). Forecasting maternal mortality rate using particle swarm optimization based artificial neural network. *Dutse Journal of Pure and Applied Sciences*, 1(1), 55–59.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Atsalakis, G., Nezis, D., Matalliotakis, G., Ioana, C., & Skiadas, C. (2007). *forecasting mortality rate using a neural network with fuzzy inference system*. Working Papers 0806, University of Crete, Department of Economics. https://ideas.repec.org/p/crt/wpaper/0806.html

Ayoo, K., Mikhaeil, J. S., Huang, A., & Wąsowicz, M. (2020). The opioid crisis in North America: Facts and future lessons for Europe. *Anaesthesiology Intensive Therapy*, *52*(2), 139–147. https://doi.org/10.5114/AIT.2020.94756

Bjerre, D. S. (2022). Tree-based machine learning methods for modeling and forecasting mortality. *ASTIN Bulletin*, *52*(3), 765–787. https://doi.org/10.1017/asb.2022.11

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*, 3–43. https://doi.org/10.1017/S1748499500000440

Box, G. E. P., Sons., J. W. &, Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis : Forecasting and control*. John Wiley & Sons, Inc. http://www.123library.org/book_details/?id=30721

Breiman, L. (2001). Random forests. *Machine Learning*, 45(2001), 5-32. https://doi.org/10.1023/A:1010933404324

Camarda, C. G., & Basellini, U. (2021). Smoothing, decomposing and forecasting mortality rates. *European Journal of Population*, *37*(3), 569–602. https://doi.org/10.1007/s10680-021-09582-4

Chen, Y., & Khaliq, A. Q. M. (2022). Comparative study of mortality rate prediction using data-driven recurrent neural networks and the lee–carter model. *Big Data and Cognitive Computing*, *6*(4), 134. https://doi.org/10.3390/bdcc6040134

Cincotta, R. (2023). Population age structure and the vulnerability of states to coups d'État. *Statistics, Politics and Policy*, 14(3), 331–355. https://doi.org/10.1515/spp-2023-0029

Elder, G. H., Clipp, E. C., Brown, J. S., Martin, L. R., & Friedman, H. S. (2009). The lifelong mortality risks of World War II experiences. *Research on Aging*, 31(4), 391–412. https://doi.org/10.1177/0164027509333447

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom Inflation. *Econometrica*, 50(4), 987–1007. https://doi.org/10.2307/1912773

Goswami, B. (2019). A brief introduction to nonlinear time series analysis and recurrence plots. *Vibration,* 2(4), 332–368. https://doi.org/10.3390/vibration2040021

Hinman, A. R., Orenstein, W. A., Bloch, A. B., Bart, K. J., Eddins, D. L., Amler, R. W., & Kirby, C. D. (1983). Impact of measles in the United States. *Reviews of Infectious Diseases*, *5*(3), 439–444. http://www.jstor.org/stable/4453053

Hong, W. H., Yap, J. H., Selvachandran, G., Thong, P. H., & Son, L. H. (2021). Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest. *Complex and Intelligent Systems*, *7*(1), 163–189. https://doi.org/10.1007/s40747-020-00185-w

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.org/fpp/.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/https://doi.org/10.1016/j.ijforecast.2006.03.001

Ibrahim, N., Md Lazam, N., & Shair, S. (2021). Forecasting Malaysian mortality rates using the Lee-Carter model with fitting period variants. *Journal of Physics: Conference Series*, *1988*, 12103. https://doi.org/10.1088/1742-6596/1988/1/012103

Kaukuntla, P. R. (2021). Advancing life insurance pricing accuracy through mortality forecasting: A time-series and survival analysis approach. *International Journal of Multidisciplinary Research and Growth Evaluation*, *2*(1), 729–734. https://doi.org/10.54660/.ijmrge.2021.2.1.729-734

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, *87*(419), 659–671. https://doi.org/10.1080/01621459.1992.10475265

Lee, R., & Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography,* 38, 537-549. https://doi.org/10.1353/dem.2001.0036

Ljung, G., & Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(20), 297-303. https://doi.org/10.1093/biomet/65.2.297

Makridakis, S., Wheelwright, S., & Hyndman, R. (1984). Forecasting: Methods and applications. *The Journal of the Operational Research Society*, 35, 79. https://doi.org/10.2307/2581936

Murphy, S. L., Kochanek, K. D., Xu, J., & Arias, E. (2024). Mortality in the United States, 2023. *NCHS Data Brief,* 2024 Dec(521), CS356116. https://doi.org/10.15620/cdc/170564

Nuraini, V., & Fitriyati, N. (2025). Transformation of traditional models to AI: SLR on the application of machine learning in mortality prediction. *CAUCHY: Jurnal Matematika Murni Dan Aplikasi*, *10*(2), 998–1014. https://doi.org/10.18860/cauchy.v10i2.35972

Ospina, R., Gondim, J. A. M., Leiva, V., & Castro, C. (2023). An overview of forecast analysis with arima models during the covid-19 pandemic: Methodology and case study in Brazil. *Mathematics,* 11(14), 3069. https://doi.org/10.3390/math11143069

Probst, P., & Boulesteix, A.-L. (2018). To Tune or Not to Tune the Number of Trees in Random Forest. *Journal of Machine Learning Research,* 18, 1-18. https://jmlr.org/papers/volume18/17-269/17-269.pdf

Qiao, Y., Wang, C.-W., & Zhu, W. (2023). Machine Learning in long-term mortality forecasting. *The Geneva Papers on Risk and Insurance-Issues and Prectice*, 49, 340-362. https://doi.org/10.1057/s41288-024-00320-5

Raftery, A. E., & Ševčíková, H. (2023). Probabilistic population forecasting: Short to very long-term. *International Journal of Forecasting*, *39*(1), 73–97. https://doi.org/10.1016/j.ijforecast.2021.09.001

Razali, N. M., & Wah, Y. B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.

Robinson, R. L. M., Palczewska, A., Palczewski, J., & Kidley, N. (2017). Comparison of the predictive performance and interpretability of random forest and linear models on benchmark data sets. *Journal of Chemical Information and Modeling*, *57*(8), 1773–1792. https://doi.org/10.1021/acs.jcim.6b00753

Sakr, S., Elshawi, R., Ahmed, A. M., Qureshi, W. T., Brawner, C. A., Keteyian, S. J., Blaha, M. J., & Al-Mallah, M. H. (2017). Comparison of machine learning techniques to predict all-cause mortality using fitness data: the Henry ford exercIse testing (FIT) project. *BMC Medical Informatics and Decision Making*, *17*(1), 174. https://doi.org/10.1186/s12911-017-0566-6

Shrader, J. G., Bakkensen, L., & Lemoine, D. (2023). *Fatal errors: the mortality value of accurate weather forecasts*. (NBER Working Paper Series). National Bureau of Economic Research. http://www.nber.org/data-appendix/w31361

Spooner, F. & Dattani, S. (2025, April 7). *Vaccination eliminated polio from the United States - Our World in Data*. https://ourworldindata.org/data-insights/vaccination-eliminated-polio-from-the-united-states?

Trotter Y., Jr., Dunn, F. L., Drachman, R. H., Henderson, D. A., Pizzi, M., & Langmuir, A. D. (1959). Asian Influenza in the United States, 1957-1958. *American Journal of Epidemiology, 70*(1), 34–50. https://doi.org/10.1093/oxfordjournals.aje.a120063

Vollset, S. E., Goren, E., Yuan, C.-W., Cao, J., Smith, A. E., Hsiao, T., Bisignano, C., Azhar, G. S., Castro, E., Chalek, J., Dolgert, A. J., Frank, T., Fukutaki, K., Hay, S. I., Lozano, R., Mokdad, A. H., Nandakumar, V., Pierce, M., Pletcher, M., … Murray, C. J. L. (2020). Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *The Lancet*, *396*(10258), 1285–1306. https://doi.org/https://doi.org/10.1016/S0140-6736(20)30677-2

World Health Organization. (2020, April 27). *Archived: WHO Timeline - COVID-19*. https://www.who.int/news/item/27-04-2020-who-timeline---covid-19?

Wu, Y. T., Niubo, A. S., Daskalopoulou, C., Moreno-Agostino, D., Stefler, D., Bobak, M., Oram, S., Prince, M., & Prina, M. (2021). Sex differences in mortality: Results from a population-based study of 12 longitudinal cohorts. *CMAJ*, *193*(11), E361–E370. https://doi.org/10.1503/cmaj.200484

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. In *Neurocomputing*, 50, 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhang, Y., Zhang, X., Razbek, J., Li, D., Xia, W., Bao, L., Mao, H., Daken, M., & Cao, M. (2022). Opening the black box: interpretable machine learning for predictor finding of metabolic syndrome. *BMC Endocrine Disorders*, *22*(1), 214. https://doi.org/10.1186/s12902-022-01121-4

Zuo, W., Damle, A., & Tuljapurkar, S. (2025). Sensitivity and uncertainty in the Lee–Carter mortality model. *International Journal of Forecasting*, *41*(2), 781–797. https://doi.org/https://doi.org/10.1016/j.ijforecast.2024.06.010