Improving mortality forecasting using a hybrid of Lee–Carter and stacking ensemble model

Samuel Asante Gyamerah[1*], Aaron Akyea Mensah[1], Clement Asare[1] and Nelson Dzupire[2]

Abstract

Background Mortality forecasting is a critical component in various fields, including public health, insurance, and pension planning, where accurate predictions are essential for informed decision-making. This study introduces an innovative hybrid approach that combines the classical Lee–Carter model with advanced machine learning techniques, particularly the stack ensemble model, to enhance the accuracy and efficiency of mortality forecasts.

Results Through an extensive analysis of mortality data from Ghana, the hybrid model's performance is assessed, showcasing its superiority over individual base models. The proposed hybrid Lee–Carter model with a stack ensemble emerges as a powerful tool for mortality forecasting based on the performance metrics utilized. Additionally, the study highlights the impact of incorporating additional base models within the stack ensemble framework to enhance predictive performance.

Conclusion Through this innovative approach, the study provides valuable insights into enhancing mortality prediction accuracy. By bridging classic mortality modeling with advanced machine learning, the hybrid model offers a powerful tool for policymakers, actuaries, and healthcare practitioners to inform decisions and plan for the future. The findings of this research pave the way for further advancements and improvements in mortality forecasting methodologies, thus contributing to the broader understanding and management of mortality risks in various sectors.

Keywords Mortality forecasting, Lee–Carter model, Stack ensemble, Machine learning, Hybrid approach

Background

Accurate predictions of death rates hold significant importance for insurance companies and experts in public health. This accuracy aids in the allocation of resources, planning for future commitments, and evaluating the impacts of health-driven interventions. Forecasting the rate of mortality plays a pivotal role within

economies, as evidenced by the distinct consequences of changes in death rates, as seen during the COVID-19 outbreak in 2020. According to an investigation by the American Council of Insurers, life insurance companies paid out more than 90 billion dollars to beneficiaries as a result of deaths related to COVID-19 during that year (Dore 2023). Nonetheless, mortality rates are susceptible to a range of influences beyond just pandemics. Aspects such as shifts in climate patterns and property damages have vital roles in molding these rates. Additionally, genetic data also play a part in mortality (Angus and van der Poll 2013). The financial strain caused by uncertainties in mortality patterns imposes the necessity of establishing a framework or set of methodologies for forecasting mortality trends in the foreseeable

*Correspondence:
Samuel Asante Gyamerah
asante.gyamerah@knust.edu.gh
[1] Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[2] Mathematical Sciences Department, University of Malawi, Zomba, Malawi

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 2 of 16

future. However, erroneous mortality predictions could carry significant economic and societal consequences, particularly affecting industries like insurance and pensions (Bhardwaj and Agarwal 2022). Inaccurate forecasts can lead to underfunded pension programs, resulting in financial stress for retirees and the government. Several statistical approaches have been devised for predicting mortality rates. Among them, the Lee–Carter model has gained prominence as the most commonly utilized one. Nevertheless, the Lee–Carter model has its own set of drawbacks when it comes to handling mortality data, as highlighted in (Gyamerah et al. 2023). To address these limitations, the applications of machine learning have emerged as a potential remedy. This has been pointed out by researchers such as (Bjerre 2022; Marino et al. 2023; Nigri et al. 2019). Machine learning techniques offer a more adaptable and data-centric approach to modeling mortality trends, capable of capturing complex interactions and nonlinear relationships between variables and mortality outcomes (Berrang-Ford et al. 2021).

Numerous researchers have endeavored to address the limitations of the Lee–Carter model by either extending it or combining it with other robust models, aiming to achieve precise and accurate mortality forecasts. For instance, Li et al. (2011) assessed the assumption of linearity of the mortality index in the Lee–Carter model. They investigated how the model's parameters change over time and whether there are any structural breaks in the mortality index of the Lee–Carter model for mortality forecasting. The study utilized mortality data from the USA, Canada, and the United Kingdom from 1960 to 2006. They proposed a modified version of the Lee–Carter model that considers these structural breaks by using the Zivot-Andrews test to determine if the mortality index in the Lee–Carter model is best explained by the simple random walk models or the broken trend stationary model. Their results demonstrated that this modified and improved model provides more accurate mortality forecasts compared to the original Lee–Carter model. In a study by Rawak (2022), the generalized linear model (GLM) framework of the Lee–Carter model was enhanced by incorporating additional factors influencing mortality. The inclusion of these factors, along with time-factor modulation, significantly improved the model's adequacy for the 14 selected countries. Danesi et al. (2015) addressed the limitations of the Lee–Carter model by comparing different variants of the traditional model to forecast mortality rates in sub-populations in Italy. The outcomes underscored the effectiveness of models that strike a balance between complexity and flexibility. They suggested exploring alternative multi-population datasets and pertinent methods for cross-population forecast comparisons in future research. In

Leng and Peng (2016), the two-step estimation procedure proposed by Lee and Miller (2001) might not effectively capture the underlying dynamics of the mortality index. Doubts were raised about the validity of future mortality projections using the two-step inference procedure for the Lee–Carter model and its extensions. Thus, they recommended adopting an efficient test to ascertain whether the mortality index genuinely follows a nonstationary autoregressive order of 1 [AR (1)] process, enhancing the reliability of the two-step inference procedure. Richman and Wuthrich (2019) extended the Lee–Carter model's applicability using neural networks, automatically selecting the optimal structure to eliminate manual specification. Their approach outperformed the traditional models in out-of-sample forecasting, excelling in learning complex relationships autonomously. Similarly, Hong et al., (2021) proposed a hybrid model combining the Lee–Carter model, artificial neural networks, and random forest models for mortality forecasting. The hybrid model proposed in their work demonstrated superior performance in predicting mortality rates as compared to the traditional models. Highlighting the significance of machine learning in healthcare, Darabi et al., (2018) emphasized the role of machine learning models in predicting mortality risks for intensive care unit (ICU) patients. They applied gradient boosting trees and neural networks to patient data, with gradient boosting trees outperforming neural networks in accuracy. Austin et al. (2012) assessed the benefits of ensemble-based methods for predicting 30-day mortality in patients with cardiovascular conditions, showing improved accuracy compared to conventional regression trees. A recent study by Marino et al. (2023) integrated neural networks into the Lee–Carter framework using recurrent neural networks (RNN) and long short-term memory (LSTM) networks. The integration demonstrated the efficacy of machine learning in enhancing the predictive capability of the Lee–Carter model for reliable long-term mortality projection.

Recent studies have shown that machine learning models have the potential to capture complex time series trends better than traditional statistical models (Gyamerah et al. 2019, 2023; Onyema et al. 2022; Alaje et al. 2022), potentially improving the accuracy and flexibility of mortality forecasts. However, machine learning models can also lack interpretability, making it challenging to understand the underlying mechanisms that drive mortality trends (Gandin et al. 2021). Therefore, there is still the need to consider the potential of the traditional mortality forecasting models when implementing machine learning algorithms for the mortality forecast. While a few studies have made attempts to address the limitations of the Lee–Carter model by combining it with

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 3 of 16

certain machine learning approaches (Marino et al. 2023; Nigri et al. 2019; Richman and Wuthrich 2019), there remains a noticeable gap in the existing literature concerning the integration of an ensemble of machine learning models with the Lee–Carter model. In this study, we seek to develop a new mortality forecasting model that can integrate the strengths of machine learning ensembles and traditional statistical forecasting models (the Lee–Carter model). Hybridizing the Lee–Carter model with an ensemble of machine learning models will provide a balanced approach to mortality forecasting. While the Lee–Carter model has a history of use and interpretability, machine learning will enhance its capability by accommodating structural changes and reducing biases, ultimately yielding a more accurate mortality forecast. The proposed model will combine an ensemble of machine learning and the Lee–Carter model to create an accurate and interpretable model by leveraging the strengths of the Lee–Carter model's ability to model trends in mortality rates over time while incorporating machine learning methods to capture more complex and nuanced mortality trends. Even though existing literature has highlighted the potential of ensemble models, there remains an unresolved question concerning the selection of base models and the optimal number of such models to achieve enhanced accuracy in mortality forecasting. Our study contributes to the existing literature on hybridizing machine learning models with the traditional mortality forecasting models in three main ways: (1) to develop a stack ensemble machine learning algorithm that combines multiple machine learning models to create a more robust model; (2) to examine how the number of base models affects the prediction accuracy; and (3) to assess the performance of each distinct base model in predicting mortality data after hybridizing with the Lee–Carter model. The rest of the study is organized as follows: Chapter two explains the materials and methodology adopted in the study. Chapter three presents the results and a discussion. Then, chapter four concludes the work and presents some recommendations.

Methods

In this chapter, we describe and explore our data set and also discuss the general idea behind our model building. We then discuss the traditional Lee–Carter model and its parameter estimation; we also provide brief information on the machine learning models used in our stack ensemble algorithms; and finally, we describe how our proposed hybrid model is developed.

This study employed secondary data on a specific sub-population of Ghana aged 40–83, covering the years 2010–2020. For the analysis, a total of 6,360,292 individuals and 92,062 total deaths were observed. The data includes the variables age, years, and mortality rate. The age variable represents the age groups observed across multiple years. The year variable indicates the specific year for which we aimed to calculate the mortality rate. Lastly, the mortality rate variable represents the log of the rate of mortality for a particular age group across different years. To analyze the mortality patterns, we computed mortality rates for each age by dividing the number of deaths for that age by the total population at risk for that age across the different years. The total log mortality rate for a particular age group was obtained by adding the total number of deaths for that age group across the years and dividing it by the total number of exposures for that age group across the years. Additionally, to calculate the log mortality rate for each year, we added the total number of deaths for that year across the different age groups and divided it by the total number of exposures for that year across the different age groups. This allowed us to examine mortality trends over time in the specified sub-population of interest. By analyzing these variables, we investigated the mortality patterns within specific age groups over time, allowing us to understand how mortality rates varied across different years and age categories.

Exploratory data analysis

Figure 1 depicts the relationship between the age of individuals in Ghana (measured in years) and mortality rates, which clearly reveals a positive correlation. This implies that, on average, mortality rates tend to rise as individuals get older. Additionally, the plot demonstrates that between the ages of 40 and 65, the increase in mortality rates is relatively gradual, but it accelerates beyond age 65. This is a predictable pattern, as individuals over the age of 65 are generally more prone to experiencing higher mortality rates compared to younger individuals. It can also be concluded that the Ghana National Pensions Regulatory Authority can increase the compulsory retirement age from 60 to 65. The scatter plot depicting the mortality rates of individuals in Ghana over time demonstrates a consistent and linear negative correlation. This indicates that mortality rates tend to decrease as the years progress. The trend is logical, as advancements in healthcare and improved living conditions over time contribute to longer life expectancies for individuals of all ages.

Model building framework

We develop a hybrid model that combines the traditional Lee–Carter model with different machine learning algorithms. Our approach for the hybrid model is to combine the Lee–Carter model with a stack of the optimal number of base models instead of combining the individual models with the Lee–Carter model. Specifically, we

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 4 of 16
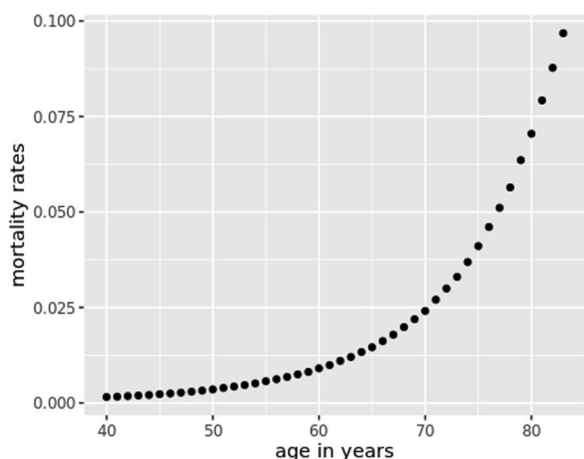
a: Mortality rate against age in years of individuals.
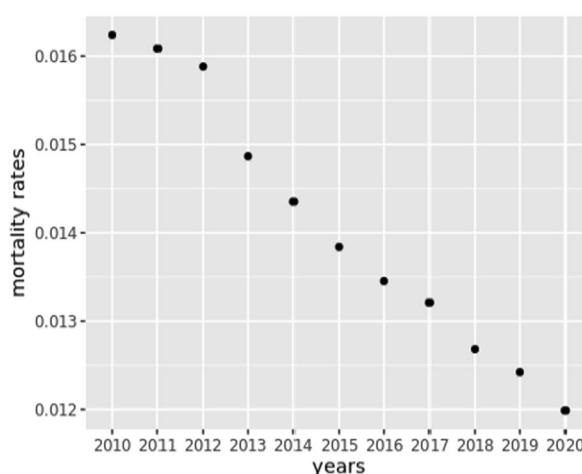
b: Mortality rate against years for individuals.

Fig. 1 Plots depicting mortality rates among individuals

develop a stack ensemble algorithm taking the generalized linear model (GLM), random forest (RF), decision tree (DT), extreme gradient boosting (XGBoost), and the neural network as the base models. The stack ensemble algorithm is then used together with the Lee–Carter model to predict mortality.

The Lee–Carter model

The Lee–Carter model is a statistical approach designed for analyzing time series data. It breaks down the natural logarithm of mortality rates within a given population into two key elements: a temporal trend and an age-based pattern. This model operates under the assumption that the logarithmic value of mortality rates in a specific year and age group can be expressed as a linear equation involving distinct components, namely an age-dependent base-line factor, an age-related slope factor, a chronological reference, and a random error component. The theoretical Lee–Carter model is given by:

$$\log\left(m_{x,t}\right) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \tag{1}$$

where $m_{x,t}$ is the central rate of mortality for age group x at time t. The α parameter represents the average log mortality rates by age over time. It captures the overall level of mortality rates across age groups. The β_x parameter indicates how rapidly or slowly the respective age-specific rates change with respect to κ_t. The κ_t parameter represents the level of mortality at a given time, and $\epsilon_{x,t} \sim \left(0, \sigma^2\right)$. To obtain a unique solution for the Lee–Carter model, Lee and Carter (1992) applied constraints to ensure that the age-specific slope terms sum to 1 and the time index sums to 0. $\sum \beta = 1$ and $\sum \kappa = 0$.

The parameters of the Lee–Carter model are generally estimated using two common approaches (Richman

and Wuthrich 2019). The first approach is applying the singular value decomposition (SVD) method to a matrix of centered log mortality rates (Lee and Carter 1992). A second approach is to treat the Lee–Carter model as a statistical model and assume an appropriate statistical distribution from which we can then fit a nonlinear model and then use a maximum likelihood method to estimate the parameters. In this study, we adopted the approach described by Lee and Carter (1992) in their original paper. The age-specific intercept term is estimated as the average over time of the log central death rates. We use the method of least squares to estimate the age-specific intercept terms.

$$\log\left(m_{x,t}\right) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$
$$\sum_{1}^{T}\log\left(m_{x,t}\right) = T\alpha_x + \sum_{1}^{T}\epsilon_{x,t}$$
$$\sum_{1}^{T}\log\left(m_{x,t}\right) = T\alpha_x \tag{2}$$
$$\alpha_x = \frac{\sum_{1}^{T}\log\left(m_{x,t}\right)}{T}$$

The coefficients $\beta_x \kappa_t$ of the Lee–Carter model are estimated in two stages. First, the singular value decomposition method (SVD) is applied to the matrix of logarithms of rates after subtracting the averages over time of the (log) age-specific rates.

In the second stage, the time index is modeled to minimize errors in the logs of death rates. This re-estimation ensures that the stipulated number of deaths equals the actual number of deaths. Next, we show mathematically how the β_x and κ_t parameters of the Lee–Carter model are estimated using the approach from (Lee and Carter

1992). First, define a matrix A_x as the centered log of death rates by centering the log of the rates at the age-specific death rates.

$$A_x = \log \left(m_{x,t} \right) - \alpha_x$$

Next, we apply the SVD method to the centered log death rates to estimate $\beta_x \kappa_t$.

$$\text{SVD}(A_x) = \text{SVD}\left(\log \left(m_{x,t} \right) - \alpha_x\right) \tag{3}$$

Equation (3) will result in three matrices: U, Σ, and V^T, where U and V are orthogonal matrices and Σ is a diagonal matrix.

$$U = \begin{pmatrix} u_{x_1,t_1} & u_{x_1,t_2} & \cdots & u_{x_1,t_n} \\ u_{x_2,t_1} & u_{x_2,t_2} & \cdots & u_{x_2,t_n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{x_m,t_1} & u_{x_m,t_2} & \cdots & u_{x_m,t_n} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_{x_1,t_1} & \sigma_{x_1,t_2} & \cdots & \sigma_{x_1,t_n} \\ \sigma_{x_2,t_1} & \sigma_{x_2,t_2} & \cdots & \sigma_{x_2,t_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_m,t_1} & \sigma_{x_m,t_2} & \cdots & \sigma_{x_m,t_n} \end{pmatrix}$$

$$V = \begin{pmatrix} v_{x_1,t_1} & v_{x_1,t_2} & \cdots & v_{x_1,t_n} \\ v_{x_2,t_1} & v_{x_2,t_2} & \cdots & v_{x_2,t_n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{x_m,t_1} & v_{x_m,t_2} & \cdots & v_{x_m,t_n} \end{pmatrix}$$

Therefore, the $\text{SVD}(A_x) = \sigma_1 u_{x,1} v_{t,1} \cdots \sigma_k u_{x,k} v_{t,k} = \sum_1^k \sigma_i u_{x,i} v_{t,i}$, where k is the rank of the matrix. The study uses a rank of 1 for the estimation of the parameters β_x and k_x, and β_x is estimated from the first column of the matrix U. To get a unique solution for our model, we apply the restrictions in the original Lee–Carter model to the first column of U by dividing the values in the first column of the matrix U by the sum of the values in the column.

$$\beta_x = \frac{\left(u_{1,1} \quad u_{1,2} \quad \cdots \quad u_{1,t} \right)}{\sum u_{x,1}} \tag{4}$$

κ_t is estimated from the multiplication of the first singular values and the first column of matrix V. Applying the restriction on κ_t,

$$\kappa_t = \sum u_{x,1} \sigma_1 \left(v_{1,1} \quad v_{1,2} \cdots \quad v_{1,t} \right) \tag{5}$$

After estimating the parameters of the Lee–Carter model, we utilize the proposed machine learning models to re-estimate and predict the time index κ_t within the Lee–Carter model. Subsequently, we replace the forecast

values for the time index κ_t along with the already estimated values for β_x and α_x, assumed to be constant and consistent (Hong et al. 2021; Lee and Carter 1992), into the original Lee–Carter model for our mortality prediction. In the subsequent stages of the study, the focus shifts toward delving into the theoretical aspects of the proposed base machine learning models and their application in predicting and forecasting the time index κ_t within the Lee–Carter model. Additionally, the stacking ensemble algorithm is employed to amalgamate these algorithms, thereby enhancing the predictive accuracy beyond what the individual machine learning algorithms can achieve on their own (Breiman 1996).

Stacking ensemble

The stacking algorithm, also known as stacked generalization or stacking ensemble, is an ensemble machine learning algorithm that was introduced by Lee and Miller (2001). Stacking, like any other type of ensemble technique, combines multiple learning algorithms to produce a desired predictive model. The basic idea of the stacking ensemble algorithm is to train a meta-learner on the output of several base models to produce a final model that has higher predictive accuracy compared to individual weak learners. In the first stage, the base learners are fitted to the data, cross-validation is done on the models to help prevent over-fitting, and the predictions from each model are saved. The meta-learner uses the output (predictions) of the weak learners as input and the target variables in the original data as the target variables in the second stage. It attempts to learn how to combine the weak learners to produce an optimal model to make better predictions. In this study, we consider the decision tree, random forest, neural network, and generalized linear model as base learners in the first stage of the stacking ensemble algorithm. We conducted extensive experiments involving various meta-learner models, including the generalized linear model (GLM), random forest (RF), XGBoost, and neural network (NN). In Table 1, we provide the performance metrics, such as RMSE, MAPE, and MAE, for each of these models. Our findings revealed that GLM outperformed the other meta-learner models, displaying the lowest RMSE, MAPE, and MAE values, thereby indicating a higher level of accuracy. We chose GLM as the meta-learner for its outstanding

Table 1 Performance metrics for each of the meta-learners used

	GLM	RF	XGBoost	NN
RMSE	0.039143	0.074639842	0.068451098	0.042021144
MAPE	0.02167	0.044750708	0.042520553	0.02541529
MAE	0.0372259	0.074784325	0.069914282	0.046118753

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 6 of 16

interpretability, computational efficiency, and alignment with our data characteristics, particularly in the context of mortality prediction (Saadatmand et al. 2023). This selection enhances both the accuracy and transparency of our stacking ensemble model, in accordance with the guidance from (Kablan et al. 2023) for a well-informed model selection process.

We adopt the stacking ensemble technique with the aim of achieving accurate estimates and forecast for the time index κ_t of the Lee–Carter model. In the subsequent steps, we describe how weak learners are used to make predictions in the first stage of our stacking ensemble model. We also discuss how the meta-learner utilizes the predictions of the weak learners in the first stage to make the final prediction.

Generalized linear models (GLM)
Generalized linear models (GLMs) are a class of regression models that are used for analyzing the relationships between variables (a predictor variable and a target response). GLMs have the ability to model nonlinear relationships, which makes them more adaptable and able to capture intricate patterns in the data than a typical linear regression model. The use of link functions and the choice of appropriate probability distributions for the response variable contribute to this flexibility. GLMs can be adapted to a variety of data types, including continuous, binary, count, or non-negative continuous data, by carefully choosing an appropriate link function and response distribution. For instance, the logistic link function and a binomial distribution are frequently utilized when dealing with binary outcomes. In the same way, the log link function with a Poisson distribution is widely used for counting data. GLMs can provide accurate predictions and trustworthy estimations for a variety of applications because of their versatility. Mathematically, the GLM can be expressed as:

$$g(E(Y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \qquad (6)$$

where $g(.)$ is a link function, $E(Y)$ is the mean (expected value) of the target variable and $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$ are coefficients associated with the predictors.

Decision tree regression (DT)
Decision trees are popular machine learning models that have shown the potential of handling classification and regression tasks. A decision tree consists of root, internal, and leaf nodes. Branch nodes in the tree correspond to choices between various alternatives, while leaf nodes represent classifications or decisions. The tree structure captures the decision-making process by branching based on different features, leading to specific outcomes

at the leaf nodes. The algorithm produces estimates by calculating the average of the response variable values for data points within the same region identified by the regression tree.

The random forest algorithm (RF)
The random forest algorithm is based on the fundamental idea of combining multiple random decision trees. This concept was first mentioned by (Breiman 2001). The random forest algorithm was defined as a collection of $\hat{h}(., \theta_1), \cdots, \hat{h}(., \theta_n)$, where the $\hat{h}(., \theta)$ are the random tree predictors and $\theta_1, \cdots, \theta_n$ are independent random variables. The random forest predictor $\widehat{h_{\mathrm{RF}}}(x)$ is derived by amalgamating this ensemble of random trees. The random forest can be represented mathematically as.

$$\widehat{h_{\mathrm{RF}}}(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{h}(x, \theta_i) \qquad (7)$$

Extreme gradient boosting (XGBoost)
The extreme gradient boosting algorithm is built on the gradient boosting framework which essentially combines weak learners to form a single model that is robust and has a higher predictive accuracy. In the XGBoost framework, the weak learners learn from the errors of the previous weak learner to improve its accuracy gradually forming a more robust model. Given a dataset with N observations and M features, the XGBoost algorithm aims to learn a prediction function $F(x)$ that maps input feature vectors x to a continuous target variable y. At each boosting iteration, XGBoost trains a new weak learner, typically a decision tree, to capture the residual errors from the previous model. The prediction function for the tth model, denoted as $F_t(x)$, can be expressed as:

$$F_t(x) = F_{t-1}(x) + \sum_{m=1}^{M} \gamma_{t,m} h_m(x) \qquad (8)$$

where $F_{t-1}(x)$ is the prediction function of the previous model at iteration $t-1$, $\gamma_{t,m}$ represents the learning rate (shrinkage parameter) for the tth model and the mth tree, $h_m(x)$ denotes the output of the mth decision tree, which is a function of the input features x. The final prediction function $F(x)$ is the sum of the individual tree outputs over all boosting iterations:

$$F(x) = F_0(x) + \sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{t,m} h_m(x) \qquad (9)$$

where $F_0(x)$ represents the initial prediction usually the mean of the target variable. The model also adopts regularization techniques and has an in-built cross-validation

structure that helps the model to prevent over-fitting even if the data is relatively small.

Neural network (NN)

Neural networks, a group of powerful and flexible machine learning models, draw their inspiration from the structure and operation of the human brain. They excel at learning from complex data patterns, making them ideal for diverse tasks, including time series forecasting. Neural networks consist of interconnected artificial neurons organized in layers, resembling the brain's architecture. Each neuron takes inputs, applies a weighted sum, and introduces nonlinearity through an activation function. During training, the network iteratively adjusts the connection weights and biases to minimize the difference between predicted and actual outcomes. This optimization process, typically achieved through backpropagation and other techniques, enables the neural network to continuously improve its predictive abilities. Interested readers should refer to (Richman and Wuthrich 2019).

Formation of the stacking ensemble algorithm

The data set was split into training and testing data sets. We used the training data set to train the models and the testing data set to determine the predictive performance of the models on the new data set. The data set contained the time index κ_t of the Lee–Carter model. We assumed that the values of κ_t for the first three years could serve as a reliable indicator for predicting the value of κ_t in the fourth year. This approach, adopted by Nigri et al. (2019) has proven to be a convenient method of estimating the time series variable κ_t, as also described by Hong et al. (2021). In this study, we conducted an 80–20% data split to allow for a robust assessment of our models' capabilities, where 80 percent of the data was used for training the models and the remaining 20 percent served as the testing dataset. For the first-level prediction, we utilized various combinations of machine learning algorithms discussed earlier as the base learners, while employing a generalized linear regression model as the meta-learner for all combinations. Each base learner was trained on the training dataset, and a fivefold cross-validation was performed on each learner. This cross-validation allowed us to estimate and evaluate the predictive performance of the models and prevent over-fitting issues. The cross-validation predictions from each base learner were then combined with the original response target variable, forming a new set of input data known as "meta features" for the meta-learner. We subsequently trained the meta-learner on these meta-features. The final stack ensemble algorithm is then combined with the Lee–Carter model by using the stack ensemble to accurately predict the time index for the Lee–Carter model. This allows for

a more accurate mortality rate prediction by the Lee–Carter model. The forecast values are then substituted in the original Lee–Carter model to make mortality rate predictions. The age-specific terms of the Lee–Carter model are assumed to be constant through time (Lee and Carter 1992).

In the study, we carefully considered and selected hyperparameters for each model, as detailed in Table 2. The choice of these hyperparameters was based on a thorough understanding of their impact on model performance. Therefore, we diligently tuned and configured them to ensure optimal results. The selection process involved a combination of empirical experimentation and best practices within the field, and the rationale behind each choice was rooted in the specific requirements of our mortality forecasting task.

To assess the predictive accuracy of our proposed model on new data, we generated predictions using the test dataset for each of the base learners. Subsequently, the predictions from the base learners were fed into the meta- learner to generate ensemble predictions. We then conducted a comparative analysis of the root-mean-squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) of the base learner models and the stacked ensemble model. Our objective was to achieve lower values of RMSE, MAE, and MAPE for our proposed model, as this indicates a higher predictive performance. The error metrics used in this are mathematically defined below:

Root-Mean-Squared Error:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_{\text{pred},i} - y_{\text{actual},i} \right)^2} \qquad (10)$$

Mean Absolute Error:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_{\text{pred},i} - y_{\text{actual},i} \right| \qquad (11)$$

Mean Absolute Percentage Error:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_{\text{pred},i} - y_{\text{actual},i}}{y_{\text{actual},i}} \right| \times 100 \qquad (12)$$

Discussions
Results

In this section, we conduct a graphical analysis of mortality rates across various age groups over time. Additionally, we present visualizations of the parameter estimates of the Lee–Carter model obtained from the singular value decomposition. The visualizations provide insights into the age-specific mortality patterns and

Gyamerah *et al. Bulletin of the National Research Centre* (2023) 47:158

Page 8 of 16

Table 2 Hyperparameter settings for the machine learning models

Model	Parameter	Value
Neural networks (NN)	Hidden layers configuration	(60, 50)
	Cross-validation	nfold = 5
	Fold assignment method	Fold assignment = modulo
	Early stopping	Stopping round = 50
	Early stopping metric	Stopping metric = RMSE
	Tolerance	Tolerance = 0
	Number of epochs	Epochs = 50
	Activation function	Activation = rectifier
Extreme gradient boosting (XGBoost)	Number of trees	ntress = 5000
	Cross-validation	nfold = 5
	Fold assignment method	Fold assignment = modulo
	Early stopping	Stopping round = 50
	Early stopping metric	Stopping metric = RMSE
	Tolerance	Tolerance = 0
	Learning rate	learn_rate = 0.1
	Subsample rate	Sample rate = 0.8
	Maximum tree depth	max_depth = 5
Random forest (RF)	Number of trees	ntrees = 1000
	Number of splits (mtries)	mtries = 2
	Cross-validation	nfold = 5
	Fold assignment method	Fold assignment = modulo
	Early stopping	Stopping round = 50
	Early stopping metric	Stopping metric = RMSE
	Tolerance	Tolerance = 0
	Subsample rate	Sample rate = 0.8
	Maximum tree depth	max_depth = 30
Generalized linear model (GLM)	Alpha	Alpha = 0.1
	Cross-validation	nfolds = 5
	Early stopping	Stopping round = 50
	Early stopping metric	Stopping metric = rmse
	Tolerance	Tolerance = 0

how the model captures the changes in mortality rates over the studied time period. The study uses historical data from Ghana, which comprises the number of deaths for ages 40–83 years and the central exposure at risk for the different ages from 2010 to 2020. The analysis was conducted using the R programming language, and the coding implementation was carried out by the authors.

Figure 2 illustrates how the mortality rates derived from our historical data change with advancement in age. The plot of mortality rates against different age groups over time clearly shows that mortality rates tend to rise with advancing age. The increase in mortality rates is relatively gradual and slow for the age groups between 40 and 60 but accelerates rapidly after the age group of 65. Notably, the steepest and most rapid increase occurs in the oldest age group, from age group 80 to age group 83,

reaching a peak rate of approximately 0.12 at age group 84. This observation is consistent with the data set, which shows a high number of deaths for individuals in the age range of 80–83. Figure 3 shows the mortality trends of individuals in Ghana from age 40–83 from year 2010 to the year 2020. From the plot, we can observe a downward movement of the mortality rates as the years increase. This is generally expected as there has been improvement in healthcare, the discovery of new technologies that increase life expectancy, and more people continuing to be educated.

Tables 3, 4, and Fig. 4 present the estimated κ_t values of the Lee–Carter model using the SVD plotted against years, and a clear decline in the overall level of mortality is evident across the different years. This consistent downward trend equips the Lee–Carter model to
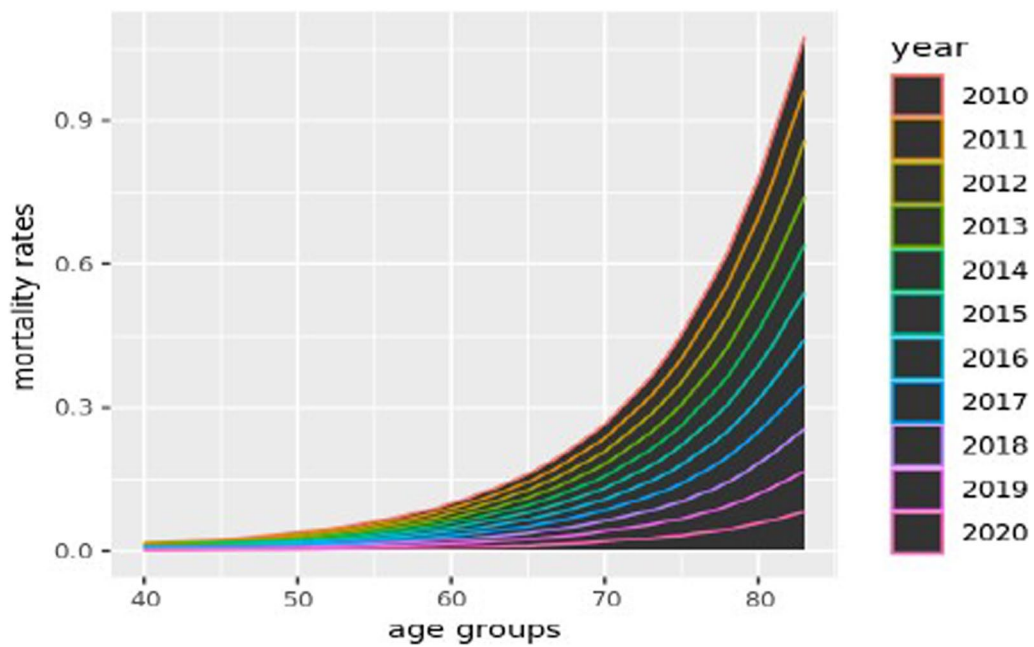
Fig. 2 Mortality rate patterns by age group of individuals in Ghana
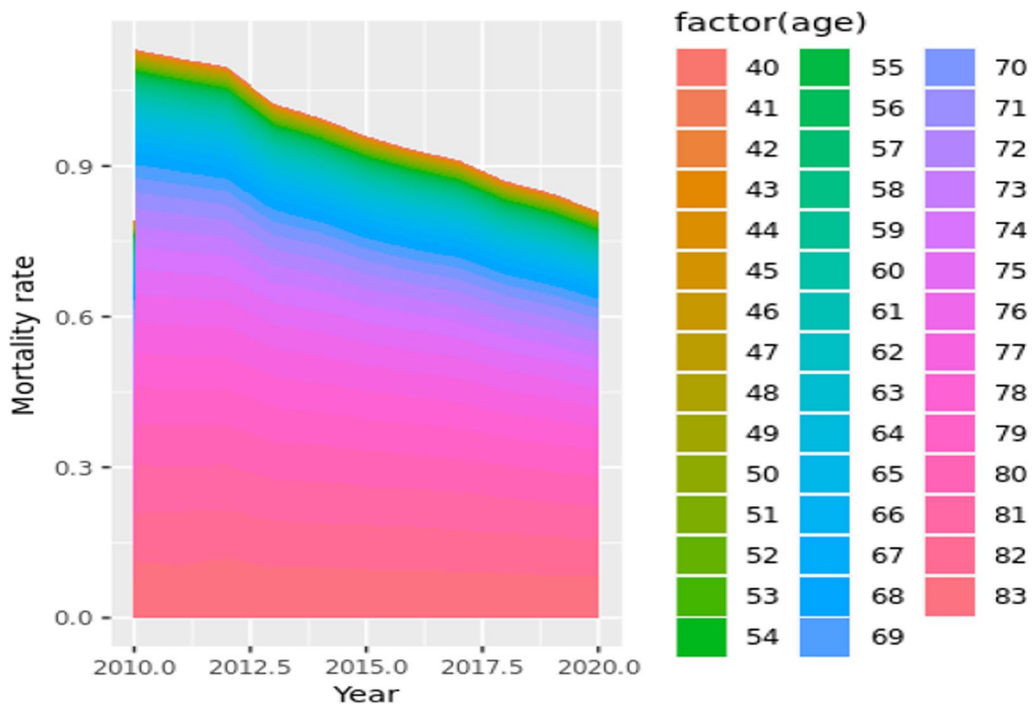

Fig. 3 Mortality rate patterns by age group of individuals in Ghana over time

effectively capture the overall changes in mortality patterns. The observed decline suggests that, on average, mortality rates have been decreasing over time. This download trend may be attributed to various factors, such as advancements in healthcare, positive lifestyle changes, and other improvements in public health practices over the years.

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 10 of 16

Table 3 Estimated κ_t (2010–2015)

Year	κ_t
2010	5.4609337
2011	4.8139260
2012	4.1926829
2013	2.0038117
2014	0.9860853
2015	−0.1303990

Table 4 Estimated κ_t (2016–2020)

Year	κ_t
2016	−1.3068399
2017	−1.8863190
2018	−3.5730888
2019	−4.4964154
2020	−6.0643773

Figure 5 showcases the estimated β_x values obtained using the SVD, displayed against age groups measured in years. The graph provides valuable insights into the rate of change in mortality rates across different ages over time. Specifically, it illustrates how the mortality rate of each age group varies as a function of time. Upon analyzing the graph, it becomes apparent that certain age groups exhibit higher changes in mortality rates, as indicated by their corresponding β_x values. Notably, the rate of change is higher for age groups between 45 and 50, followed by a slight decline after age group 50.

Subsequently, it sharply increases after age group 60 and begins to decrease again after age group 75. This comprehensive visualization allows us to grasp the dynamic patterns of mortality rate changes across various age groups throughout the observed period.

Application of the stack ensemble model to the mortality data

In this section, we focus on predicting the κ_t values of the Lee–Carter model using our stack ensemble machine learning model. We utilize the predicted values to assess the predictive accuracy of the machine learning model. We do this by calculating the root-mean-squared error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) for both the stack ensemble model and the individual base models. We have summarized the error measurements in Tables 5, 6 and 7. Initially, we constructed a stacked ensemble model with three base learners: GLM, RF, and XGBoost. However, we noticed that the predictions from these base learners were not accurate enough. Surprisingly, the predictions from the meta-learner, which aggregated the base learners' outputs, demonstrated better performance than the individual base models. We sought to enhance the ensemble's performance by adding another machine learning model as a base learner, thus bringing the total to four base learners. While this addition resulted in a slight improvement in the ensemble's error, the errors of the base models remained unchanged. In our continued pursuit of optimizing the ensemble, we introduced a neural network as an additional base learner, expanding the total number of base learners to five. This decision
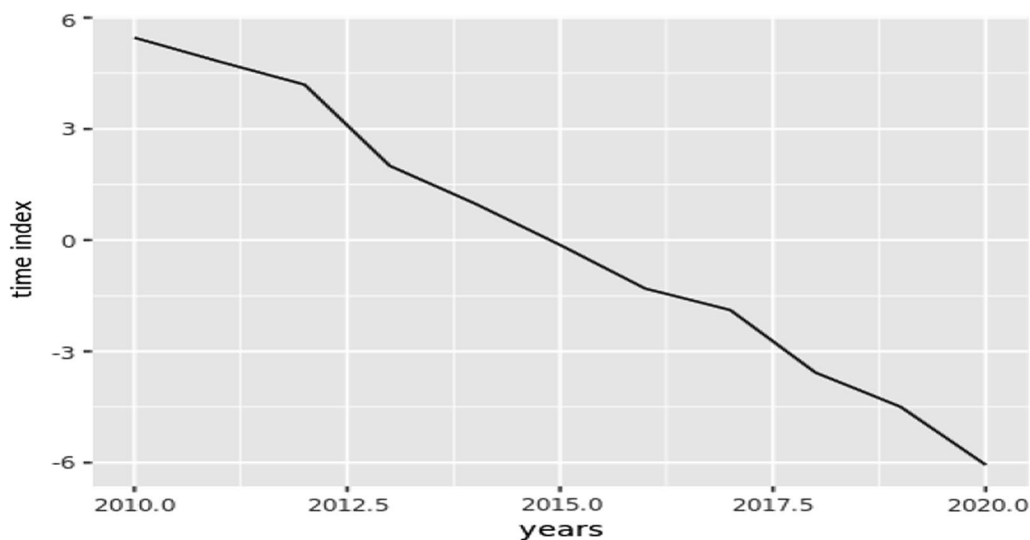

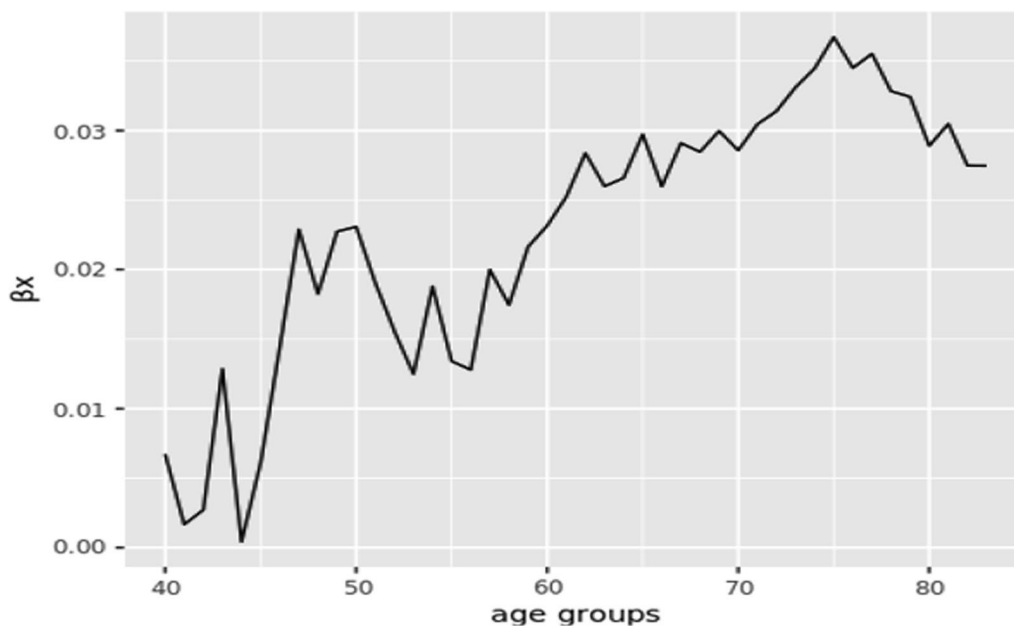**Fig. 4** Time index of the Lee–Carter model plotted against years

Fig. 5 Plot of β_x in the Lee–Carter model as a function of age in years

Table 5 Performance metrics for individual base models in the stack ensemble model and the stack ensemble itself

Machine learning models				
Metric	Stack ensemble	Random forest	XGBoost	GLM
RMSE	0.1307365	1.761899	1.747702	1.758636
MAPE	0.05448399	0.9773364	0.9798964	0.9851048
MAE	0.1271199	1.669135	1.607744	1.623914

The stack ensemble comprises three base models: random forest (RF), XGBoost, and generalized linear model (GLM)

yielded a remarkable drop in the ensemble's error, and surprisingly, the errors associated with the individual base models, including the newly added neural network, remained consistent.

From the performance metrics above, it is evident that the individual base models performed poorly in predicting κ_t values, with high RMSE, MAPE, and MAE

values. The stack ensemble algorithm, on the other hand achieved the lowest RMSE, MAPE, and MAE values, indicating its superior performance in leveraging the strengths of the individual base models for predicting κ_t. Additionally, as the number of base models increased, the errors associated with the stack ensemble decreased. This suggests that the stack ensemble becomes more effective with a greater number of base models.

Actual versus predicted mortality rates
We compare the predicted κ_t values for the years 2016 and 2019 as shown in Table 6, obtained using the stack ensemble algorithms with the actual κ_t values. These predicted κ_t values are then used to make mortality rate predictions by substituting them into the original Lee–Carter model equation. For this purpose, we assume that the α_x and β_x values remain constant, following the approach proposed by Lee and Carter in 1992. Table 8

Table 6 Error measurements of the individual base models in our stack ensemble model and the error measurement of our stack ensemble

Machine learning models					
Metric	Stack ensemble	Random forest	XGBoost	GLM	DT
RMSE	0.1306422	1.761899	1.747702	1.758636	1.747861
MAPE	0.0544645	0.9773364	0.9798964	0.9851048	0.9799522
MAE	0.1270404	1.669135	1.607744	1.623914	1.608126

The stack ensemble model comprises four base models: RF, XGBoost, GLM, and DT

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 12 of 16

Table 7 Error measurements for individual base models and the stack ensemble model

Machine learning models						
Metric	Stack ensemble	Random forest (RF)	XGBoost	Generalized linear model (GLM)	Decision tree (DT)	Neural network (NN)
RMSE	0.039143	1.761899	1.747702	1.758636	1.74786	1.803414
MAPE	0.02167	0.9773364	0.9798964	0.9851048	0.9799522	1.001618
MAE	0.0372259	1.669135	1.607744	1.623914	1.608126	1.704286

The stack ensemble consists of five base models: random forest (RF), XGBoost, generalized linear model (GLM), decision tree (DT), and neural network (NN)

Table 8 Comparison of predicted and actual values using stack ensembles with different numbers of base models

Year	2016	2016	2019	2019
Ensemble	Actual	Predicted	Actual	Predicted
Stack-3	− 1.30684	− 1.210258	− 4.496415	− 4.654073
Stack-4	− 1.30684	− 1.210264	− 4.496415	− 4.653921
Stack-5	− 1.30684	− 1.257513	− 4.496415	− 4.521541

Increasing the number of base models leads to improved prediction accuracy

presents the predicted and actual mortality rate values for the years 2016 and 2019.

Subsequently, we employ suitable graphical representations shown in Figs. 6a, b, 7a, b,8a, b to depict the outcomes. These representations serve the purpose of evaluating the precision of mortality rate forecasts derived from the stack ensemble model's κ_t estimations, shedding light on any disparities between projected and actual figures. This method of examination facilitates the derivation of meaningful insights regarding the stack ensemble algorithm's efficacy in forecasting κ_t values and the resultant implications for mortality rate predictions. Through these visual depictions, we achieve a lucid and

intuitive comprehension of the model's accuracy and its competence in capturing fluctuations in mortality rates over time. Specifically, our graphical analysis involves a comparison between the genuine log mortality rates and those predicted by the stack ensemble model for the years 2016 (represented as 'a') and 2019 (represented as 'b'), focusing on age groups spanning from 40 to 83 years. This specific time frame is chosen due to its application as the evaluation dataset for our stack ensemble, as well as its role in computing error metrics.

Based on the depicted plots, a distinct pattern becomes evident, illustrating a consistent rise in mortality rates with advancing age. This trend holds true for both the observed periods, 2016 and 2019, across various ensemble stack models. Notably, the actual and predicted mortality rates closely adhere to this discernible pattern across all plots. Upon comparing the actual and predicted mortality rates for each respective year (2016 and 2019), a striking overlap is observable. This overlap signifies the remarkable accuracy of our stack ensemble algorithm in forecasting mortality rates that closely mirror those derived directly from historical data. Notably, as we assess models with a greater number of base models, a more favorable fit becomes apparent. This
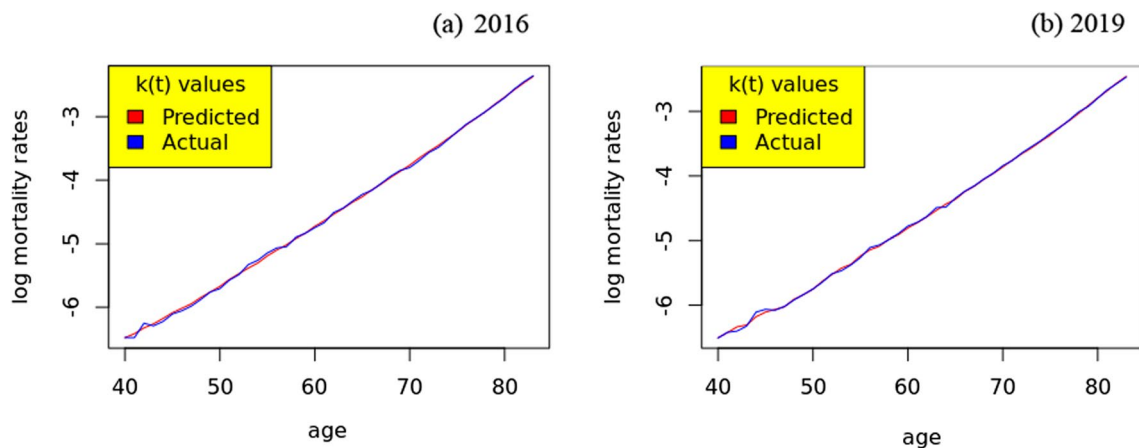

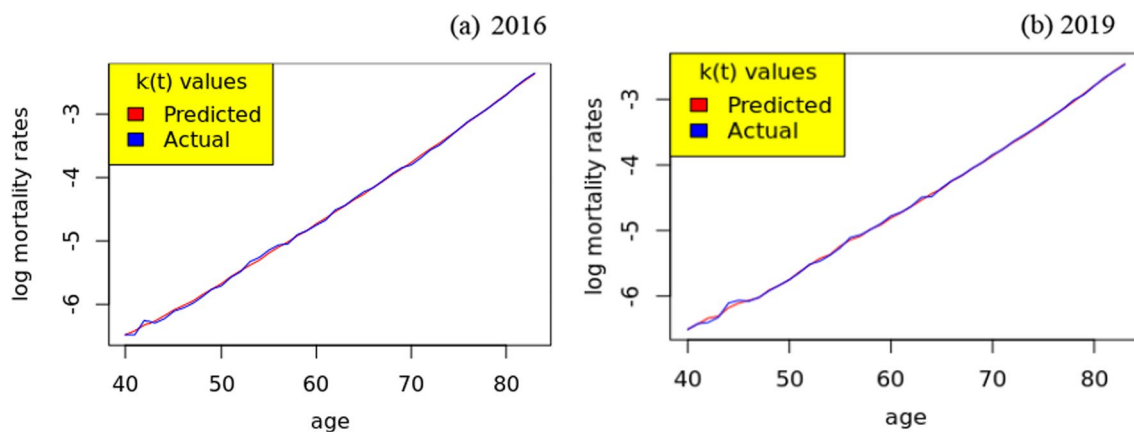**Fig. 6** Actual and predicted log mortality rate of the Stack-3 model

Fig. 7 Actual and predicted log mortality rate of the Stack-4 model
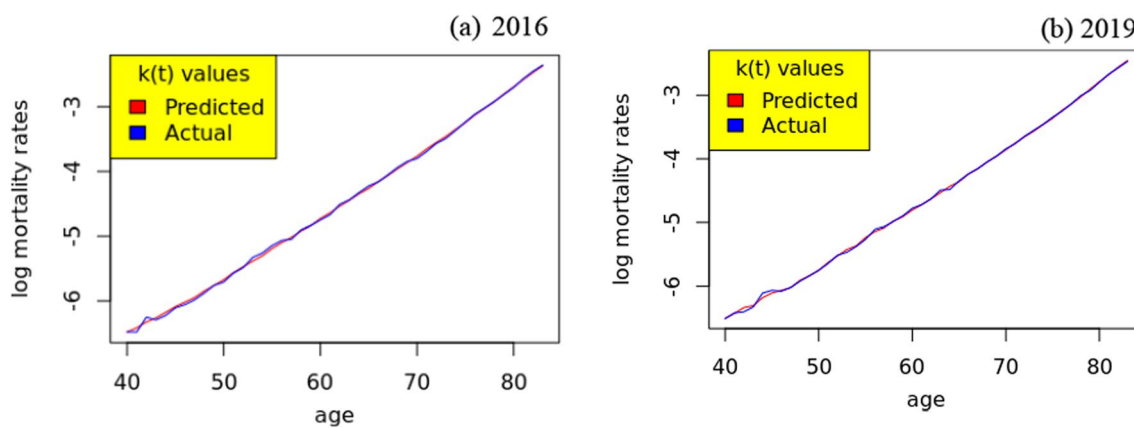

Fig. 8 Actual and predicted log mortality rate of the Stack-5 model

phenomenon aligns seamlessly with our analysis of measurement errors; the errors of the stack ensemble decrease in correspondence with the augmentation of base models. This robust alignment substantiates the reliability and efficacy of our model in predicting mortality rates across diverse age groups. More precisely, our proposed approach for mortality forecasting in this study involves a stack ensemble methodology. This entails utilizing generalized linear models, decision trees, random forests, extreme gradient boosting, and neural networks as base models, with a generalized linear model serving as the meta-learner.

Discussion of results

In this section of the study, we try to establish connections between our approach for mortality forecasting with the approach and findings existing in literature and also reference existing literature to support our findings and recommendations. Through this comprehensive analysis, we aim to provide valuable insights into the implications of our proposed model for mortality forecasting and their relevance in practical scenarios. In our study on mortality rates and mortality forecasting, we observed a consistent trend of mortality levels. The mortality levels among the aged population were higher, on average, than compared to younger age groups. The result is consistent with findings from research conducted in the literature of mortality forecasting (Krasowski et al. 2022). Through our comprehensive data analysis, it is evident that mortality rates at younger ages demonstrate a gradual increase, while mortality rates for older ages exhibit an exponential rise. This pattern is often attributed to factors such as a weakened immune system, prolonged exposure to harmful substances, and accumulated risks over time of the aged population. The observed mortality trends highlight the significance of understanding age-specific vulnerabilities and underscore the importance of implementing targeted interventions and healthcare measures to address the needs of different age cohorts effectively. When analyzing the mortality rate of individuals in Ghana over the

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 14 of 16

years, we observed a consistent and gradual decrease in mortality rates across all age groups as the years progressed. This positive trend can be attributed to the notable improvements in social factors and medical services in the country. Over time, Ghana has experienced significant progress in various areas that have contributed to the decline in mortality rates. One key factor behind this improvement is the rise in the educational level of the population. With increasing access to education, people have become more informed about health practices, leading to better health awareness and healthier lifestyle choices. Additionally, advancements in medical services and healthcare facilities have played a crucial role. People now have better access to healthcare, leading to early detection and treatment of diseases, thus reducing mortality rates (Alaje and Olayiwola 2023; Olayiwola et al. 2023; Schöley et al. 2022). Furthermore, improvements in nutrition and living conditions have positively impacted the overall health of the population. People have better access to nutritious food, resulting in improved overall health and immunity against diseases. The focus on sanitation and hygiene has also contributed to creating a healthier living environment, further reducing the risk of diseases and infections. These positive trends reflect the significant progress made by the country in its pursuit of a healthier and more prosperous society.

The time index κ_t of the Lee–Carter model is a crucial element that reflects the overall level of mortality across time and enables the model to predict mortality rates into the future. We identified a significant downward trend in the time index during the Lee–Carter model data analysis, indicating a systematic decline in mortality rates over the years covered by the historical data. This is a notable result with important implications for population dynamics and mortality forecasts. According to Schöley et al. (2022), observed drop in mortality rates suggests an overall improvement in population health and longevity over time. This pattern conforms with the idea of the demographic transition, a phenomenon in which countries progress from having high rates of births and deaths to having lower rates as a result of economic and social growth (Walaszek and Wilk 2022). Factors such as improvements in healthcare practices, better access to healthcare services, and public health initiatives could be responsible for the decline in death rates (Hao et al. 2020). Additionally, improvements in living conditions, nutrition, sanitation, and education can also result in an increase in life expectancy and reduced mortality rates. This steady decline has significant implications for policymakers and healthcare planners. For accurate projection of future mortality rates and understanding variations in population, it is essential that one understands this trend. When extending the observed trend into the distant future, it is crucial to proceed with caution because a variety of intricate factors affect demographic patterns, and unforeseen occurrences like pandemics, recessions, or changes in lifestyle can have unpredictable effects on mortality rates (Brenner 2021).

The performance metrics obtained from the base models were significantly high, indicating that each model, when used individually, struggled to effectively capture, and model our historical data. As these machine learning models are inherently data-driven, any inconsistencies or inaccuracies in the data set could lead to undesirable results and compromise predictive performance. The high error rates underscore the importance of understanding the limitations and potential pitfalls of using machine learning algorithms (Onyema et al. 2022). These models heavily rely on the quality and representativeness of the data they are trained on. If the data used for training contains noise, outliers, or missing information, the model's ability to make accurate predictions is compromised (Angwin et al. 2022). It is therefore important to pay particular attention to and give careful thought to the base models in the stack ensemble that have significant measurement errors. These errors may propagate into the forecasts of the stack ensemble and have an impact on the performance of the ensemble in general should they remain present in the base models. Recognizing that the quality of the stack ensemble model is affected by the accuracy of the individual base models is a crucial step in building a robust stack ensemble model. If the base models have substantial measurement errors, the ensemble model's ability to fix these errors may be limited, which could result in substandard forecasts from the stack ensemble model. The presence of persistently large measurement errors presents a challenge in achieving optimized forecast accuracy, despite the ensemble's advantage in integrating various models and minimizing the influence of individual model errors. The stack ensemble model is always better than the best single predictor, but a set of good base predictors gives the stack ensemble higher predictive accuracy (Breiman 1996; Gyamerah et al. 2019).

The stack ensemble model for the different combination of base models all showed a relatively lower measurement error as compared to the individual base models. This result validates the ensemble approach's effectiveness in maximizing the positive effects of each of its constituent models' strengths while minimizing each of their limitations (Breiman 1996). The stack ensemble creates a synergistic effect that increases forecasting accuracy by integrating several predictions from the underlying models. The potential of the ensemble to eliminate noise and inconsistencies present in the individual base model predictions helps the ensemble model to achieve the lowest

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 15 of 16

measurement error. The stack ensemble model from the different combination of base models showed that as we increase the number of base models in a stack ensemble, the predictive accuracy increases. We observed from the analysis that the stack ensemble with three base models showed a good predictive performance compared to the base models in the stack ensemble but as we increased the number of base models to four, the predictive accuracy became better. A significant observation was when we increased the number of base models to five, we saw a significant decrease in the errors associated with the stack ensemble. This was an indication that, for a stack model, the predictive performance increases with increasing number of base models. There was no change in the errors associated with the base models since the inclusion of another model to the stack does not affect the predictive performance of the existing base model. The graphical comparison of the predicted and actual κ_t values also supported the observation made from the error measurements. The significant overlap between the predicted and actual values of our proposed ensemble model provides additional evidence of the exceptional predictive performance and reliability of our stack ensemble model in forecasting mortality rates. It is worth noting that if the ensemble is extremely complicated or the underlying models are over-fitted to the training data, the ensemble may not generalize well to unfamiliar, independent data sets, resulting in poor performance.

Conclusions

This research aimed to enhance mortality forecasting accuracy by integrating a hybrid Lee–Carter model with a stack ensemble approach. Combining the strengths of the Lee–Carter model and the versatility of the stack ensemble, this framework demonstrated improved forecasting results. The study's experiments, conducted on real-world mortality data, showcased the stack ensemble's superiority over individual base models. We also found that when building a stack ensemble algorithm, the predictive performance of the model can be improved by increasing the number of base models. This study contributes to the relatively scarce research on mortality forecasting by introducing a hybrid model that combines the classical Lee–Carter model with machine learning models, particularly the stack ensemble. Through this innovative approach, the study provides valuable insights into enhancing mortality prediction accuracy. By bridging classic mortality modeling with advanced machine learning, the hybrid model offers a powerful tool for policymakers, actuaries, and healthcare practitioners to inform decisions and plan for the future. The findings of this research pave the way for further advancements and improvements in mortality forecasting methodologies, thus contributing to the broader understanding and management of mortality risks in various sectors.

In addition to the significant progress achieved in this study, there are several fascinating avenues worth exploring to enhance the hybrid Lee–Carter model for predicting mortality. Exploring new ensemble techniques, such as expanding the stack ensemble with additional uncorrelated models, could also yield valuable insights for enhancing model performance. Secondly, future research should broaden the hybrid approach's application to various demographic groups and timeframes to ensure its utility in diverse contexts. Strengthening the model's effectiveness and applicability can be achieved by confirming its robustness across different demographic scenarios and historical periods. Assessing its performance across varied geographic regions, socioeconomic segments, or diverse populations can provide a comprehensive understanding of the model's strengths and limitations. Ultimately, this investigation lays a strong groundwork for future progress in mortality prediction, contributing to more precise and impactful forecasts in healthcare, actuarial science, and public policy domains.

Gyamerah *et al. Bulletin of the National Research Centre*   (2023) 47:158

Page 16 of 16

Declarations

Ethics approval and consent to participate.
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors have no competing interests to disclose in the publication of this paper.

References

Alaje AI, Olayiwola MO (2023) A fractional-order mathematical model for examining the spatiotemporal spread of COVID-19 in the presence of vaccine distribution. Healthc Anal 4:100230. https://doi.org/10.1016/j.health.2023.100230

Alaje AI, Olayiwola MO, Adedokun KA, Adedeji JA, Oladapo AO (2022) Modified homotopy perturbation method and its application to analytical solitons of fractional-order Korteweg–de Vries equation. Beni-Suef Univ J Basic Appl Sci 11(1):139. https://doi.org/10.1186/s43088-022-00317-w

Angus DC, van der Poll T (2013) Severe sepsis and septic shock. N Engl J Med 369(9):840–851. https://doi.org/10.1056/NEJMra1208623

Angwin AJ, Armstrong SR, Fisher C, Escudero P (2022) Acquisition of novel word meaning via cross situational word learning: an event-related potential study. Brain Lang 229:105111. https://doi.org/10.1016/j.bandl.2022.105111

Austin PC, Lee DS, Steyerberg EW, Tu JV (2012) Regression trees for predicting mortality in patients with cardiovascular disease: what improvement is achieved by using ensemble-based methods? Biom J 54(5):657–673. https://doi.org/10.1002/bimj.201100251

Berrang-Ford L, Sietsma AJ, Callaghan M, Minx JC, Scheelbeek PFD, Haddaway NR, Haines A, Dangour AD (2021) Systematic mapping of global research on climate and health: a machine learning review. Lancet Planet Health 5(8):e514–e525. https://doi.org/10.1016/S2542-5196(21)00179-0

Bhardwaj M, Agarwal S (2022) Decision-making optimisation in insurance market using big data analytics survey. In: Big data analytics in the insurance market. Emerald Publishing Limited, pp 57–80. https://doi.org/10.1108/978-1-80262-637-720221004

Bjerre DS (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

Breiman L (1996) Stacked regressions. Mach Learn 24:49–64

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Brenner MH (2021) Unemployment, bankruptcies, and deaths from multiple causes in the COVID-19 recession compared with the 2000–2018 great recession impact. Am J Public Health 111(11):1950–1959. https://doi.org/10.2105/AJPH.2021.306490

Danesi IL, Haberman S, Millossovich P (2015) Forecasting mortality in sub-populations using Lee-Carter type models: A comparison. Insur Math Econ 62:151–161. https://doi.org/10.1016/j.insmatheco.2015.03.010

Darabi HR, Tsinis D, Zecchini K, Whitcomb WF, Liss A (2018) Forecasting mortality risk for patients admitted to intensive care units using machine learning. Proc Comput Sci 140:306–313. https://doi.org/10.1016/j.procs.2018.10.313

Dore K (2023) It's a work in progress. How covid has changed the life insurance marketplace. Accessed 13 Dec 2022 from www.cnbc.com/2022/12/01/how-covid-has-changed-life-insurance-underwriting.html

Gandin I, Scagnetto A, Romani S, Barbati G (2021) Interpretability of time-series deep learning models: a study in cardiovascular patients admitted to Intensive care unit. J Biomed Inform 121:103876. https://doi.org/10.1016/j.jbi.2021.103876

Gyamerah SA, Ngare P, Ikpe D (2019) On stock market movement prediction via stacking ensemble learning method. IEEE Conf Comput Intell Financ Eng Econ 2019:1–8. https://doi.org/10.1109/CIFEr.2019.8759062

Gyamerah SA, Arthur J, Akuamoah SW, Sithole Y (2023) Measurement and impact of longevity risk in portfolios of pension annuity: the case in sub Saharan Africa. FinTech 2(1):48–67. https://doi.org/10.3390/fintech2010004

Hao L, Xu X, Dupre ME, Guo A, Zhang X, Qiu L, Zhao Y, Gu D (2020) Adequate access to healthcare and added life expectancy among older adults in China. BMC Geriatr 20(1):129. https://doi.org/10.1186/s12877-020-01524-9

Hong WH, Yap JH, Selvachandran G, Thong PH, Son LH (2021) Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest. Compl Intell Syst 7(1):163–189. https://doi.org/10.1007/s40747-020-00185-w

Kablan R, Miller HA, Suliman S, Frieboes HB (2023) Evaluation of stacked ensemble model performance to predict clinical outcomes: a COVID-19 study. Int J Med Inf 175:105090. https://doi.org/10.1016/j.ijmedinf.2023.105090

Krasowski A, Krois J, Kuhlmey A, Meyer-Lueckel H, Schwendicke F (2022) Predicting mortality in the very old: a machine learning analysis on claims data. Sci Rep 12(1):17464. https://doi.org/10.1038/s41598-022-21373-3

Lee RD, Carter LR (1992) Modeling and forecasting U.S. mortality. J Am Stat Assoc 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(4):537–549. https://doi.org/10.1353/dem.2001.0036

Leng X, Peng L (2016) Inference pitfalls in Lee-Carter model for forecasting mortality. Insur Math Econ 70:58–65. https://doi.org/10.1016/j.insmatheco.2016.05.016

Li JS-H, Chan W-S, Cheung S-H (2011) Structural changes in the Lee–Carter mortality indexes. N Am Act J 15(1):13–31. https://doi.org/10.1080/10920277.2011.10597607

Marino M, Levantesi S, Nigri A (2023) A neural approach to improve the Lee–Carter mortality density forecasts. N Am Act J 27(1):148–165. https://doi.org/10.1080/10920277.2022.2050260

Nigri A, Levantesi S, Marino M, Scognamiglio S, Perla F (2019) A deep learning integrated Lee–Carter model. Risks 7(1):33. https://doi.org/10.3390/risks7010033

Olayiwola MO, Alaje AI, Olarewaju AY, Adedokun KA (2023) A caputo fractional order epidemic model for evaluating the effectiveness of high-risk quarantine and vaccination strategies on the spread of COVID-19. Healthc Anal 3:100179. https://doi.org/10.1016/j.health.2023.100179

Onyema EM, Almuzaini KK, Onu FU, Verma D, Gregory US, Puttaramaiah M, Afriyie RK (2022) Prospects and challenges of using machine learning for academic forecasting. Comput Intell Neurosci 2022:1–7. https://doi.org/10.1155/2022/5624475

Rawak RH et al (2022) Extending the GLM framework of the Lee–Carter model with random forest recursive feature elimination based determinants of mortality. Sains Malaysiana 51(7):2237–2247

Richman R, Wuthrich MV (2019) Lee and Carter go machine learning: recurrent neural networks. SSRN Electr J. https://doi.org/10.2139/ssrn.3441030

Saadatmand S, Salimifard K, Mohammadi R, Kuiper A, Marzban M, Farhadi A (2023) Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients. Ann Oper Res 328(1):1043–1071. https://doi.org/10.1007/s10479-022-04984-x

Schöley J, Aburto JM, Kashnitsky I, Kniffka MS, Zhang L, Jaadla H, Dowd JB, Kashyap R (2022) Life expectancy changes since COVID-19. Nat Hum Behav 6(12):1649–1659. https://doi.org/10.1038/s41562-022-01450-3

Walaszek M, Wilk J (2022) Population changes during the demographic transition, pp 219–257. https://doi.org/10.1007/978-3-031-06108-0_10

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.