Hybrid Deep Learning and Lee-Carter Model for Mortality Forecasting: A Study of US Adults Aged 35-80

Collins Abaitey[1] , Eric Abakah[1*] , Clement Asare[1] , Emmanuel Oppong[1] , Fareeda Maltiti Baba-Sadiq[1] , Elshaddai Appiah[1]

[1]Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

[*] Corresponding Author: ehrickabakah98@gmail.com

Abstract

This study introduces a hybrid approach that enhances mortality forecasts by integrating machine learning techniques specifically Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Neural Networks (NN) with the traditional Lee-Carter model. After deriving the time index κ_t from the Lee-Carter model, these deep learning models were employed to capture complex temporal patterns, which were then incorporated into the Lee-Carter framework to improve forecasting accuracy. The hybrid models were evaluated using historical mortality data from the United States, covering ages 35 to 80 years from 1975 to 2020. Among the models tested, the LSTM model outperformed all others, demonstrating superior capability in capturing time dependencies and producing more accurate mortality forecasts. The integration of LSTM with the Lee-Carter framework led to significant improvements in predictive accuracy. This research demonstrates that combining traditional statistical approaches with modern deep learning techniques, particularly LSTM, offers a powerful method for enhancing mortality forecasting by effectively modeling time-dependent patterns. These findings provide valuable insights and tools for policymakers, actuaries, and healthcare professionals to improve planning and decision-making.

Keywords: Machine learning, LSTM model, BiLSTM, Neural Networks, Lee-Carter

1. INTRODUCTION

Hybrid mortality modeling has become increasingly relevant as researchers strive to improve the accuracy and reliability of mortality forecasts. In actuarial science, demography, and public health, accurate mortality predictions are crucial for understanding and preparing for future trends in life expectancy and mortality rates. Hybrid approaches, which integrate traditional statistical methods with machine learning techniques, offer enhanced capabilities for capturing complex patterns in mortality data. This integration allows for more precise forecasting, accommodating both short-term fluctuations and long-term trends in mortality rates.

The importance of accurate mortality forecasting is highlighted by its significant economic implications. For example, life insurance companies and pension funds face considerable vulnerability to longevity risk, which arises when policyholders live longer than anticipated, resulting in the need for higher-than-expected disbursements. The global market for longevity risk is substantial, with pension-related liabilities alone estimated to be between USD 60 trillion and USD 80 trillion [2]. According to the World Health Organization, global life expectancy increased significantly from 66.8 years in 2000 to 73.4 years in 2019, representing a gain of more than six years [31]. Although healthy life expectancy increased from 58 years to 64 years during this period, reflecting an 8% rise, this improvement has been mainly attributed to reductions in mortality rates rather than a decrease in years lived with disability. In other words, the overall increase in life expectancy (6.6 years) has outpaced the growth in healthy life expectancy (5.4 years). This decline in mortality rates across all age groups can largely be attributed to significant advancements in nutrition, public health, and medical research aimed at the prevention and treatment of diseases. Furthermore, the COVID-19 pandemic highlighted the urgency of accurate mortality prediction. During the pandemic, the United States alone saw life insurance companies pay over ninety billion dollars to beneficiaries due to COVID-19-related deaths [1]. This demonstrates the critical role of mortality forecasts in managing unexpected health crises and economic risks.

There is a long history of mortality modeling that began with the publication of Gompertz's law of mortality in his seminal paper [9]. Since then, numerous

models have been proposed, although the endeavor to forecast mortality remains relatively recent. Currently, many of the mortality forecasting techniques in use, particularly those employed by European statistical offices, are extrapolative in nature [28]. These methods leverage observed consistency in age patterns and mortality trends over time, as noted in [4] and cited in [14]. Extrapolative techniques are considered more objective and user-friendly, and they are often thought to yield more accurate forecasts compared to expectation methods—based on expert opinions—and explanation methods, which predict mortality by specific causes or explanatory models [4].

In 1992, Lee and Carter [18] introduced a novel approach to long-term mortality forecasting, known as the Lee-Carter (LC) model. For many years, the Lee and Carter [18] methodology has been the standard extrapolative technique for mortality projections. This model decomposes mortality into two components: a time-varying effect that causes all age-specific central mortality rates to exhibit a consistent pattern of stochastic evolution over time, and two age-specific parameters for each age group. However, despite its remarkable capabilities and robust performance, the Lee-Carter model has several limitations as highlighted by several researchers [17, 20, 22, 11, 25]. Notably, its assumption of constant age-specific mortality reduction rates leads to unrealistic estimates and underestimation of mortality improvements among older populations [6].

In recent years, hybrid forecasting models that incorporate deep learning and other machine learning approaches have gained increasing significance across various fields, including actuarial science and computer science. Several studies have highlighted the importance of these models [20, 22, 11]. The key emphasis is about how machine learning techniques offer a more flexible and data-driven method for predicting mortality trends. That is, Machine learning abilities to capture intricate interconnections and nonlinear relationships between covariates and mortality outcomes. Additionally, machine learning approaches can also enhance the evaluation of mortality estimates produced by traditional stochastic models, such as the Lee-Carter and Renshaw-Haberman models [7]. There is a growing body of innovative research focused on employing machine learning methods in both the non-life and life sectors, particularly in the realm of mortality forecasting [26, 7]. However, some previous studies have combined the Lee-Carter model with fixed effect regression models such as random forest and decision trees [17]. Nonetheless, these models often fail to adequately represent time-dependent patterns and nonlinear correlations in mortality

data, resulting in erroneous estimates [11]. In actuarial and demographic applications, neglecting to include temporal dependence in mortality trends can lead to poor risk management and decision-making. Therefore, to enhance the accuracy of mortality forecasts and capture the complex time-dependent patterns in mortality data, this study investigates the use of deep learning models to predict κ_t in the Lee-Carter model.

The primary aim of this study is to hybridize the Lee-Carter model with deep learning algorithms, such as Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and neural networks, to enhance the accuracy of predicting time-dependent mortality trends. The objectives of the study are 1) to develop a hybrid mortality forecasting model by integrating the Lee-Carter model with deep learning techniques such as LSTM, Bi-LSTM, and neural networks, and 2) to evaluate the performance of these deep learning models in predicting mortality for individuals aged 35 to 80. Achieving these objectives will answer the following research questions: 1) How can the integration of the Lee-Carter model with deep learning techniques improve the accuracy of mortality predictions for the target age group? 2) How do the predictive performances of the deep learning models LSTM, Bi-LSTM, and neural networks compare in terms of accuracy and reliability? This study is significant as it seeks to improve mortality forecasting accuracy, particularly for the aging population and the associated financial risks of longevity. By integrating traditional statistical methods with advanced deep learning techniques, this research aims to enhance predictive capabilities essential for policymakers, public health officials, and insurance providers. The Focus on individuals aged 35 to 80 years also provides insights into this demographic's unique health challenges. Ultimately, the findings could inform effective mortality management strategies and policy decisions, contributing to a better understanding of contemporary mortality dynamics. The rest of the study is structured as follows: Section 2 presents the materials and methods; Section 3 presents the results with discussions; and Section 4 concludes the study with recommendations for future research.

2. MATERIALS AND METHODS

Dataset

This study analyzed secondary data from the Human Mortality Database for a selected sub-population of the United States aged 35 to 80 from 1975 to 2020. The dataset includes a total of 89,139,773 individuals, comprising 49,202,432 females and 39,937,341 males. Within this population, there were 1,257,755 recorded deaths, with 509,970 females and 747,785 males. The data encompasses several variables: year, age, group (gender), mortality, exposure, and deaths. To analyze mortality trends over time in the United States, we computed the annual log mortality rate using the following formula:

$$\text{Log Mortality Rate} = \frac{\sum_{i=1}^{n} D_i}{E} = \log\left(\frac{D_i}{E}\right)$$

where D_i represents the total number of deaths for age group i in a specific year and E is the total population at risk (exposure) in that age group. Examining these age-specific mortality rates over several years allowed for a better understanding of fluctuations in mortality risk both within and between age groups.

Figure 1 illustrates the relationship between age and mortality rate. The plot reveals a clear upward trend, indicating that individual mortality rates increase with age. This consistent rise in mortality rates underscores the critical role of age in understanding and forecasting mortality trends. Such a pattern is anticipated and aligns with the fundamental premise of many mortality forecasting models, which assert that age is the primary determinant of mortality probability [30, 29]. Additionally, there is a pronounced increase in mortality rates among individuals aged 70 and older. This steep rise highlights the heightened vulnerability of older age groups to mortality, emphasizing the need for targeted interventions and healthcare strategies tailored to this population.

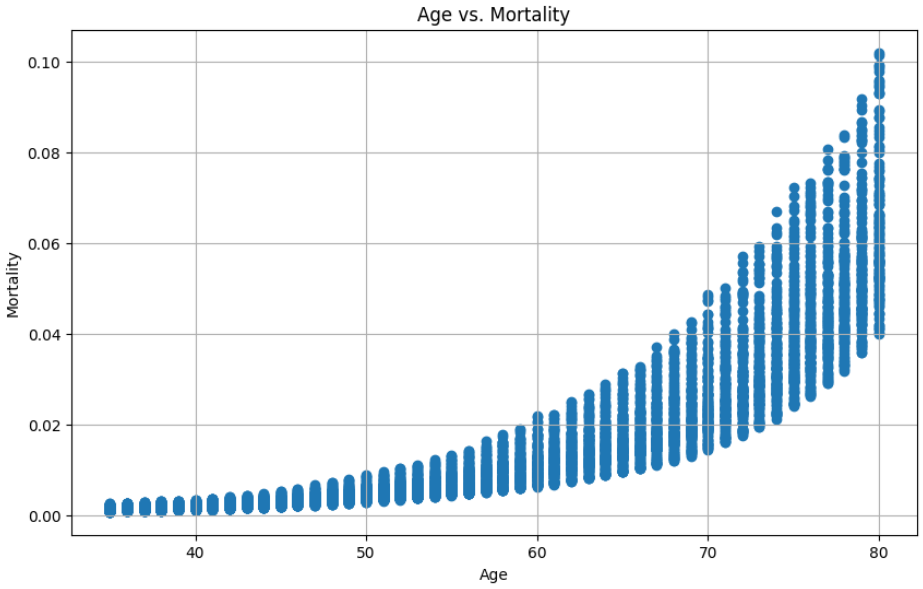

Figure 1: Plot of age against mortality rate

Figure 3 illustrates a general decline in mortality rates for both females and males over the years, reflecting an overall improvement in mortality trends. A notable feature of the graph is the persistent gap between the blue and orange lines, which indicates that males consistently experience higher mortality rates than females. This observation aligns with established patterns across various populations, where males exhibit a 60% greater mortality risk compared to females [32]. However, the gradual reduction of this disparity over time suggests a trend toward convergence, indicating that gender differences in mortality may be diminishing. Interestingly, the most recent data points, clustered around 2020, reveal a marked increase in mortality rates for both sexes. This upward trend can be linked to the significant impact of the COVID-19 pandemic, which has profoundly affected mortality patterns across the globe.

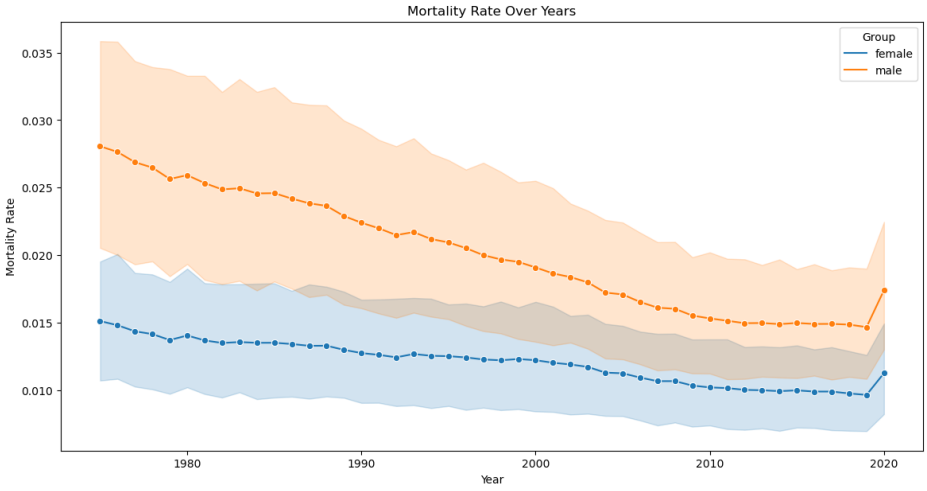

Figure 2: Plot of mortality rates pattern against years.

Methods

We develop a hybrid model that integrates the traditional Lee–Carter model with the three deep learning algorithms (LSTM, BiLSTM, and NN). We seek to hybridize the Lee-Carter model with each of the three deep learning models. This allows for the creation of distinct hybrid models, enabling us to evaluate the performance of each hybrid approach in forecasting mortality rates.

Lee-Carter Model

One popular stochastic model for predicting mortality rates is the Lee-Carter model. In the fields of actuarial sciences and demography, it has grown in importance as a technique for predicting and projecting age-specific death rates. The foundation of the Lee-Carter model is the notion that the overall level of mortality is represented by a time-varying index (t = 1,. . . .., T) and the log of the mortality rate at a specific age (x = 1,. . . .., X) which captures the distinct age pattern of death. The Lee-Carter model can be represented mathematically as follows.

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

where $m_{x,t}$ is the central rate of mortality for age group x at time t. The α_x parameter is the age-specific component of the average age pattern of

mortality which shows how mortality rates vary across different ages. The β_x parameter measures the sensitivity of mortality rates to changes over time for each age group. It shows how changes in the overall mortality level affect mortality rates at different ages concerning the time-dependence. The κ_t parameter is a time-varying parameter that characterizes the overall mortality level in a given year or time. It is interpreted as a measure of the average mortality risk at time t. The $\epsilon_{x,t}$ $(0, \sigma^2)$ is the error term that accounts for the unobserved factors and random variability in mortality rates for each age group x and time t.

To ensure that equation 1 can be identified, then: $\sum \beta = 1$, $\sum \kappa = 0$. To estimate equation 1, we employ a Singular Value Decomposition (SVD) approach described in [18]. We then estimate the age-specific intercept term by finding the average logarithmic transformed death rate for each age over time.

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

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

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

$$\sum_{t=1}^{T} \log(m_{x,t}) = T\alpha_x + \sum_{t=1}^{T} \epsilon_{x,t}$$

$$\sum_{t=1}^{T} \log(m_{x,t}) = T\alpha_x$$

$$\alpha_x = \frac{\sum_1^T log(m_{x,t})}{T} \tag{2}$$

We Subtract these averages from the log-transformed rates to center the data:

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

The matrix A consists of these centered values, where A is an XT matrix (with X being the number of age groups and T being the number of periods). We apply the SVD to the matrix of logarithms of rates after subtracting the log of the age-specific rates. We continue by modeling the time index κ_t to minimize deviations in the logs of the mortality rates.

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

The result of Equation 3 will be three matrices: U, $\sum$, and V^T where U is an orthogonal matrix where the value depends on age, V^T is an orthogonal matrix where its value is time-dependent and $\sum$ is a singular diagonal matrix.

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

$$U^TU = I, \quad V^TV = I$$

$$A = U_1\sigma_1V_1^T + U_2\sigma_2V_2^T + \cdots + U_n\sigma_nV_n^T \tag{4}$$

We used a rank of 1 for the estimation of the parameters β_x and κ_t. β_x is the first column of U and κ_t is the first column of V.

$$\beta_x = \frac{(u_{1,1}u_{1,2}...u_{1,t})}{\sum u_{x,1}}$$

$$\kappa_t = \sum u_{x,1}\sigma_1(v_{1,1}v_{1,t})$$

Long Short-Term Memory Model

LSTM is a form of RNN that was created to address the limitations of standard RNNs in capturing long-term dependencies in sequential data. It was introduced in [12]. The key principle behind how LSTM works is that, rather than using the same feedback loop connection for events that occurred long ago and events that occurred yesterday to forecast what will happen tomorrow, LSTM employs a more sophisticated structure incorporating gates to regulate information flow. Specifically, LSTM networks use a forget gate, an input gate, and an output gate to regulate which information is retained, updated, and sent on to the next time step. This design enables LSTM to efficiently preserve and use both long-term and short-term memories. This allows LSTMs to perform language-related tasks.

At each time step t of the LSTM Layer Computations, the Long Short-Term Memory (LSTM) layer performs the following operations:

The forget gate determines which information from the previous cell state should be retained and carried forward to the next cell state. The activation of the forget gate f_n is computed as:

$$f_n = \sigma_g(W_f \cdot [h_{n-1}, x_n] + b_f)$$

where σ_g is the sigmoid activation function, W_f is the weight matrix for the forget gate, $[h_{n-1}, x_n]$ represents the concatenation of the previous hidden state h_{n-1} and the current input x_n, and b_f is the bias term. The input gate decides how much of the new information should be added to the current cell state. The activation of the input gate i_n is computed as:

$$i_n = \sigma_g(W_i \cdot [h_{n-1}, x_n] + b_i)$$

Here, W_i is the weight matrix for the input gate, and b_i is the bias term. The temporary cell state $\tilde{C}_n$ represents the new candidate values for the cell state, which are scaled by the input gate before being added to the current cell state. It is computed as:

$$\tilde{C}_n = \sigma_h(W_C \cdot [h_{n-1}, x_n] + b_C)$$

where σ_h is typically the hyperbolic tangent (tanh) activation function, W_C is the weight matrix for the cell state, and b_C is the bias term. The current cell state C_n is updated by combining the previous cell state, scaled by the forget gate, and the new candidate values, scaled by the input gate:

$$C_n = f_n * C_{n-1} + i_n * \tilde{C}_n$$

The forget gate determines how much of the previous cell state C_{n-1} to keep, while the input gate determines how much of the new candidate state $\tilde{C}_n$ to add. The output gate determines the output of the LSTM cell. The activation of the output gate o_n is computed as:

$$o_n = \sigma_g(W_o \cdot [h_{n-1}, x_n] + b_o)$$

where W_o is the weight matrix for the output gate, and b_o is the bias term. The hidden state h_n is the output of the LSTM cell at the current time step. It is computed by applying the output gate to the tanh of the current cell state C_n:

$$h_n = o_n * \sigma_h(C_n) \tag{5}$$

LSTM uses two key activation functions, the sigmoid function transforms any input into a value between 0 and 1. It is defined as:

$$\sigma_g(x) = \frac{1}{1 + e^{-x}}$$

The tanh function transforms any input into a value between -1 and 1. It is defined as:

$$\sigma_h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Bidirectional Long Short-Term Memory Model

Bidirectional LSTMs are a variant of the standard two-long short-term memory (LSTM) that is also designed to handle sequential data. The main idea behind Bi-LSTMs is to analyze the input sequence both forwards and backward in order to allow the model to incorporate information from both past and future states.

The Bi-LSTM architecture is made up of two independent LSTM layers, one of which processes the input sequence forward and the other backward. The outputs of these two LSTM layers are then merged at each time step to get the final result. Unlike standard RNNs, which process input sequences in just one way (either forward or backward), Bi-LSTM processes the sequence in both directions concurrently. The forward LSTM layer is the same as the LSTM model defined in Equation 5. Mathematically, the backward LSTM layer computes the following equations at each time step t:

$$f_t = \sigma(W_f[h_{t+1}, x_t] + b_f) \qquad \text{Forget gate}$$

$$i_t = \sigma(W_i[h_{t+1}, x_t] + b_i) \qquad \text{Input gate}$$

$$c_t = f_t * c_{t+1} + i_t * \tanh(W_c[h_{t+1}, x_t] + b_c) \qquad \text{Cell state}$$

$$o_t = \sigma(W_o[h_{t+1}, x_t] + b_o) \qquad \text{Output gate}$$

$$h_t = o_t * \tanh(c_t) \qquad \text{Hidden state} \tag{6}$$

During the forward pass, the input sequence moves through the LSTM layer from the beginning to the end. At each step, the forward LSTM calculates its state and adjusts its memory cell using the current input along, with the previous hidden state and memory cell. Meanwhile, the input sequence is also processed by the LSTM layer in order starting from the last step and going back to the first. Like in the forward pass the backward LSTM computes its state and updates its memory cell based on the present input and previous hidden state and memory cell.

After both forward and backward passes are finished the hidden states from both LSTM layers are merged at every step. This merging can be as simple as concatenating the hidden states or applying some other transformation. One advantage of Bi-LSTM is its ability to capture not the preceding context, at a given time step (unlike RNNs) but also the subsequent context. By taking into account both future details Bi-LSTM can grasp intricate relationships, within the input sequence.

Neural Network Model

Neural networks are powerful machine learning models inspired by biological neural networks, capable of learning complex patterns and representations from data, [10]. It is composed of three layer: input layer, hidden layers and output layer which are connected by neurons. The neurons within a neural network receive inputs and compute a weighted sum, followed by the application of an activation function that introduces non-linearity. During the training phase, the network undergoes iterative adjustments of connection weights and biases. This process aims to minimize the disparity between predicted and actual outcomes, a technique known as optimization. The neural network is trained using a supervised learning approach, where the model learns to map input data to the desired output by minimizing a loss function, [5].

The activation of a neuron j in layer l is computed as:

$$z_j^l = \sum_{i=1}^{l} w_{ij}^l a_i^{l-1} + b_j^l \tag{7}$$

where w_{ij}^l is the weight, a_i^{l-1} is the activation from the previous layer, and b_j^l is the bias term. In this study, we An activation function σ is then applied to compute the final activation:

$$a_j^l = \sigma(z_j^l) \tag{8}$$

The output layer produces the final prediction, and the performance of the model is measured using a loss function such as the Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \tag{9}$$

where y_i is the actual value and $\bar{y}_i$ is the predicted value.

Neural networks are trained using optimization methods such as Gradient Descent and backpropagation, which modify the weights to reduce the loss function. This allows neural networks to capture complex trends, making them crucial for tasks like credible mortality forecasting.

Formation of the Hybrid Mortality Forecasting Model

We lagged the data before using it to train the LSTM, Bi-LSTM, and neural network models. Lagging the data captures the temporal dependencies inherent in time series data which allows the models to learn from past values and generate accurate future predictions. Lagged data creates a sequence for the model to learn from, allowing it to identify trends, seasonality, and temporal patterns [19, 3]. We assumed that the first three years' κ_t values could serve as a reliable indicator for forecasting the fourth year's value. This approach from [22] has shown to be an excellent method of estimating the time series variable κ_t, which is also described by [13].

The following table illustrates the dataset used for the LSTM, Bi-LSTM, and NN models, including the actual κ_t values and their lagged versions:

Table 1: κ_t values used for training and testing the models

Year	κ_t **values**	**lag1**	**lag2**	**lag3**
1978-01-01	9.605133	10.171352	11.318423	12.040781
1979-01-01	8.279512	9.605133	10.171352	11.318423
1980-01-01	8.768251	8.279512	9.605133	10.171352
1981-01-01	7.834674	8.768251	8.279512	9.605133
1982-01-01	7.133978	7.834674	8.768251	8.279512
⋮	⋮	⋮	⋮	⋮
2016-01-01	-9.022713	-8.882640	-9.100307	-9.005688
2017-01-01	-9.040511	-9.022713	-8.882640	-9.100307
2018-01-01	-9.326591	-9.040511	-9.022713	-8.882640
2019-01-01	-9.794005	-9.326591	-9.040511	-9.022713
2020-01-01	-3.408362	-9.794005	-9.326591	-9.040511

In the study, we employed an 80%-20% data split to allow for a thorough evaluation of our models' capabilities, with 80% of the data used for training and the remaining 20% serving as the testing dataset. We used the lagged data for training, while the actual κ_t values were used for testing.

For the initial forecasts, we used LSTM, Bi-LSTM, and Neural Networks. Each model was trained on the lagged training dataset, and 5-fold cross-validation was used to assess and evaluate the models' prediction performance while avoiding overfitting. Using cross-validation, we were able to successfully integrate the predictions from each model. The hybrid model properly predicts the time index κ_t for the Lee-Carter model. This integration improves the Lee-Carter model's ability to forecast mortality rates. The forecast values are then substituted into the original Lee-Carter model to estimate mortality rates. The age-specific parameters in the Lee-Carter model are expected to remain constant across time.

To assess the predictive accuracy of our proposed model on data, we generated predictions for each individual model, particularly LSTM, Bi-LSTM, and NN. We next compared three widely utilized Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) to assess the performance of each model. Our goal was to obtain lower values of RMSE, MAE, and MAPE for our suggested models, since this suggests improved prediction performance and generalization to the data. The error metrics utilized in this study are mathematically defined as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{11}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{12}$$

where y_i represents the actual values, $\hat{y}_i$ denotes the predicted values, and n indicates the number of observations.

3. RESULTS AND DISCUSSION

Results

Figure 3 shows the relationship between age and mortality rates, with data points color-coded by year and ranging from 1975 to 2020. The red regression line overlays the scatter plot, which emphasizes the overall trend between age and the mortality rate. The color gradient changes from purple (previous years) to yellow (recent years) to add a time component to the analysis. The figure clearly shows a positive correlation between age and mortality rates. As individuals age, their mortality rate steadily rises. The mortality rate increases gradually between the ages of 35 and 50. The scatter points are densely packed and relatively low, which implies a consistent and reduced mortality risk among younger people. This slow move implies that while mortality does increase with age, the danger is quite minor until the age of 50. It then starts to rise significantly between the ages of 50 and 65. The points begin to spread out further, and the slope of the regression line becomes steeper than in the 35-50 age range. After the age of 65, it accelerates rapidly. The points go up the y-axis, and the regression line's steep slope indicates a considerable increase in mortality. The significant increase suggests that people over 65 have a significantly higher chance of death.

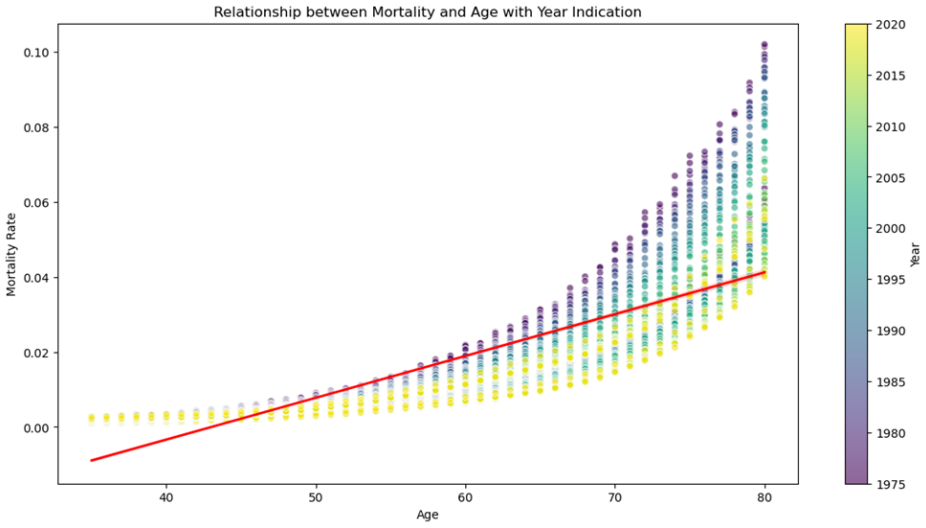

Figure 3: Plot of mortality rates pattern by age group.

Figure 4 shows the mortality trends of individuals in the United States from age 35 to age 80 from the year 1975 to the year 2020. From the plot, we can see an overall downward trend in mortality rates for most age groups from 1980 to about the year 2010. This decreasing trend indicates that advancements in various areas, including healthcare, living conditions, and overall quality of life, have led to a progressive drop in mortality rates over the last few decades. However, this positive trend appears to be disrupted around the year 2020, with a notable increase in mortality rates across practically all age groups. This sudden rise could be attributed to the COVID-19 pandemic, which had a considerable influence on worldwide death rates, particularly among older age groups who are more vulnerable to the virus, [21].

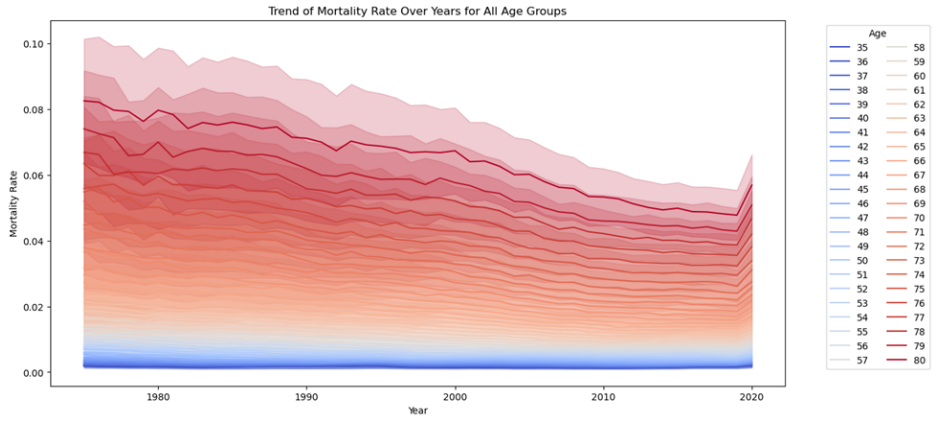

Figure 4: Plot of a trend showing mortality rate against years for all age groups

Tables 2, 3, 4 present the estimated k_t values for males, females, and the total population as derived from the Lee-Carter model using Singular Value Decomposition (SVD). The k_t values serve as indicators of mortality rates, which exhibit a general declining trend over time.

Table 2: Estimated male κ_t

Year	kt_values
1975	14.401631
1976	13.733420
1977	12.610189
1978	12.027303
1979	10.693669
⋮	⋮
2016	-10.676321
2017	-10.693316
2018	-10.777154
2019	-11.297666
2020	-4.767075

Table 3: Estimated female κ_t

Year	kt_values
1975	8.705711
1976	8.011024
1977	6.931560
1978	6.449875
1979	5.236618
⋮	⋮
2016	-7.166915
2017	-7.189877
2018	-7.736503
2019	-8.155179
2020	-2.496647

Table 4: Estimated Total κ_t (1975 - 2020)

1975-1997		1998-2020	
Year	κ_t	Year	κ_t
1975	12.04078149	1998	-0.05833998
1976	11.31842258	1999	-0.20085416
1977	10.17135173	2000	-0.80713630
1978	9.60513269	2001	-1.51180769
1979	8.27951194	2002	-1.98819315
1980	8.76825079	2003	-2.68310507
1981	7.83467391	2004	-4.23448208
1982	7.13397824	2005	-4.47961205
1983	7.18130481	2006	-5.63225902
1984	6.76811536	2007	-6.60085383
1985	6.74531104	2008	-6.71262441
1986	6.25808518	2009	-7.79976763
1987	5.75464300	2010	-8.43229879
1988	5.56548467	2011	-8.70511172
1989	4.51438524	2012	-9.07272625
1990	3.72790227	2013	-9.00568829
1991	3.18708345	2014	-9.10030728
1992	2.49613971	2015	-8.88263957
1993	2.99694232	2016	-9.02271288
1994	2.33911521	2017	-9.04051145
1995	2.03341385	2018	-9.32659108
1996	1.33569908	2019	-9.79400455
1997	0.44426087	2020	-3.40836221

Figure 5 presents the estimated β_x values derived from the Singular Value Decomposition (SVD), plotted against age groups measured in years. The graph illustrates that different age groups exhibit varying degrees of change in mortality rates, as reflected in their corresponding β_x values. Notably, the rate of change is most pronounced for individuals aged 35 to 38, with a slight decline observed after age 38. Subsequently, there is a significant increase around age 40, followed by another decline at age 63. This comprehensive depiction enhances our understanding of the dynamic patterns in mortality

rate variations across various age groups over the observed period.

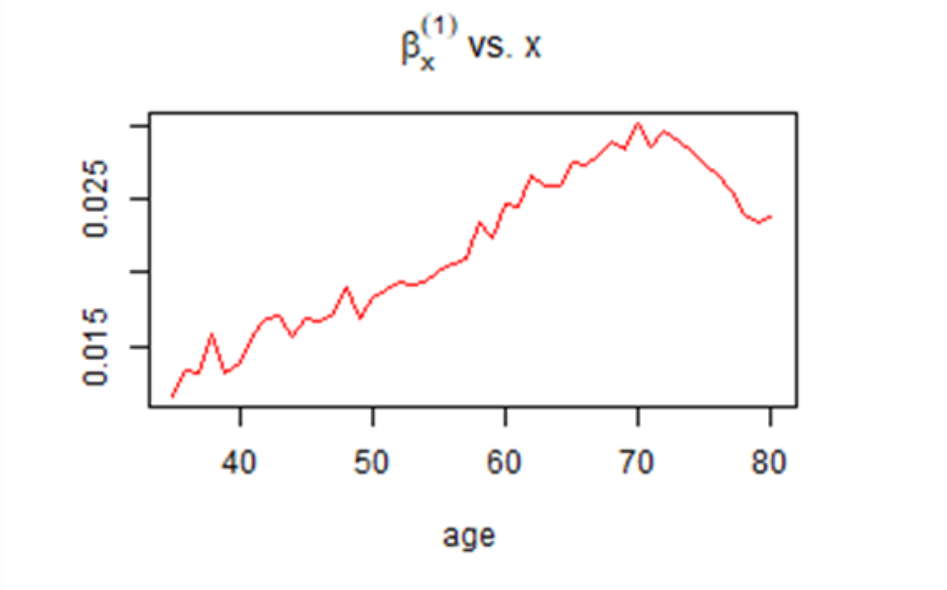

Figure 5: Plot of β_x of the Lee-Carter model against age

Analyzing these plots allows for a clearer understanding of how mortality rates are structured across different ages and how they evolve over time for females. The trends depicted in these parameters facilitate more accurate forecasting of future mortality rates by capturing the underlying patterns and their progression.

Application of the LSTM, Bi-LSTM, and the neural network models to the United States mortality data set.

In this section, we delve into the performance evaluation of three machine learning models: Long Short-Term Memory, Bidirectional LSTM, and Neural Network. The evaluation procedure focuses on assessing each model's predicted accuracy, using evaluation metrics that give essential insights into their respective performances. We summarize the error metrics in the Table 5.

Table 5: Error Metrics

Metric	MSE	MAE	MAPE
LSTM	5.3590	**1.2734**	**0.3083**
Bi-LSTM	**5.3298**	1.2973	0.3101
NN	5.4586	1.4879	0.3292

From the Table 5, we can observe that the Bi-LSTM has the lowest MSE indicating that it has the smallest average squared errors. The LSTM tends to have the lowest MAE and MAPE indicating that on average, LSTM's predictions are closer to the actual values. With the lowest MAE and MAPE, the LSTM showed superior in predicting κ_t.

We compare the predicted κ_t values from 2015 to 2019, produced using hybrid models, with the actual κ_t values. The predicted κ_t values are then used to estimate mortality rates by replacing them into the original Lee Carter equation. We assume constant values for α_x and β_x, as proposed by Lee and Carter in 1992. Table 6 shows predicted and actual death rates for the years 2015 to 2019.

Table 6: Actual and Predicted κ_t Values

Year	**Actual κ_t**	**Predicted κ_t (LSTM)**	**Predicted κ_t (Bi-LSTM)**	**Predicted κ_t (NN)**
2015	-8.8826396	-8.803983	-8.768156	-8.543645
2016	-9.0227129	-8.776583	-8.737502	-8.488410
2017	-9.0405115	-8.773879	-8.737778	-8.438983
2018	-9.3265911	-8.780485	-8.749879	-8.477440
2019	-9.7940046	-8.842212	-8.811502	-8.616756

Discussion

Our mortality forecasting study identified persistent patterns in mortality rates, with older individuals experiencing a higher mortality rate compared to younger individuals. This finding aligns with previous research by Pierce et al. [23], which revealed that mortality rates rise exponentially with age. Contributing factors include weakened immune systems, accumulated health risks over time, and increased susceptibility to chronic diseases [8]. These results underscore the necessity for specialized healthcare treatments tailored

to distinct age groups to address their unique vulnerabilities. While preventive measures and early detection programs may be more effective for older populations, it is crucial for younger individuals to address lifestyle and environmental factors to mitigate long-term health risks.

Our analysis of mortality rates in the US revealed a consistent decline across all age groups over time. This positive trend can be attributed to several factors, including improved education, better healthcare services, and enhanced living conditions [33]. Increased health awareness and access to medical facilities have led to earlier disease identification and treatment, significantly reducing mortality rates. Public health initiatives, such as immunization programs and anti-infective disease campaigns, have been instrumental in this decline. Furthermore, improved nutrition and hygiene have yielded broad health benefits by reducing the incidence of malnutrition and waterborne infections. This integrated approach to health has resulted in a more robust population capable of effectively managing diseases.

The time index k_t from the Lee-Carter model demonstrates a strong decreasing trend, indicating a systematic decline in mortality rates. This pattern suggests improvements in population health and lifespan, consistent with the demographic transition hypothesis [24]. This hypothesis explains how a country's birth and death rates shift from high to low as it develops. Factors such as better healthcare, improved living conditions, and public health initiatives contribute to this decline [15]. Advances in medical technology, enhanced access to healthcare services, and effective disease prevention and management programs have all played a role in this upward trend. These enhancements not only increase life expectancy but also improve the quality of life for many individuals.

Initially, the error metrics for our models were significant, reflecting challenges in effectively collecting and modeling historical data. Machine learning models, including LSTM, Bi-LSTM, and neural networks, are particularly sensitive to data quality issues such as noise and outliers [34]. These challenges can lead to overfitting or underfitting, adversely affecting the forecast accuracy of each model. Our findings are consistent with other research on mortality prediction. For instance, Scognamiglio et al. [27] demonstrated that several recurrent neural network models, including LSTM and Bi-LSTM, frequently outperform traditional models like the Lee-Carter model in capturing nonlinear patterns and providing accurate projections. These neural network models excel at

processing time-series data and can learn complex temporal relationships, making them well-suited for mortality forecasting. However, the study also highlighted the challenges of predicting sudden fluctuations in mortality rates, a common issue across all modeling methodologies [16]. In our case, the unexpected occurrence of the COVID-19 pandemic caused a sudden shift in mortality rates, complicating accurate forecasts.

4. CONCLUSION

This study aimed to enhance mortality forecasting by integrating the Lee-Carter model with hybrid forecasting models, including LSTM, Bi-LSTM, and Neural Networks, focusing on accurately capturing temporal dependencies in mortality trends. Utilizing lagged data for training enabled effective modeling of these trends, with the LSTM hybrid model outperforming the Bi-LSTM and Neural Networks due to its capacity to capture long-term dependencies in sequential data. While Bi-LSTM models may lead to overfitting, the LSTM's emphasis on historical data points proved more effective for mortality forecasting. The LSTM-Lee-Carter hybrid model yielded the most accurate mortality rate estimates, illustrating the potential of combining hybrid models with traditional actuarial frameworks to improve prediction accuracy. The findings offer valuable insights for policymakers, actuaries, and healthcare practitioners, enhancing decision-making and preparation for demographic and health-related challenges.

References

[1] American Council of Life Insurers (ACLI). New research from acli shows u.s. mortality rates at historical low, 2021. URL https://www.acli.com/posting/nr21-060#:~:text=New%20research%20from%20ACLI%20shows,since%20the%201918%20Influenza%20Epidemic. Accessed: 2024-08-30.

[2] David Blake and Johnny Li. Longevity risk and capital markets: the 2022–2023 update. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 49(2):229–233, 2024.

[3] Pranav Bodapati. Doing multivariate time series forecasting with recurrent neural networks. https://www.databricks.com/blog/2019/09/10/doing-multivariate-time-series-forecasting-with-recurrent-neural-networks.html, 2019. Accessed: [Insert Access Date].

[4] Heather Booth and Leonie Tickle. Mortality modelling and forecasting: A review of methods. *Annals of actuarial science*, 3(1-2):3–43, 2008.

[5] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers*, pages 177–186. Springer, 2010.

[6] Carlo G Camarda and Ugofilippo Basellini. Smoothing, decomposing and forecasting mortality rates. *European Journal of Population*, 37: 569–602, 2021.

[7] Philippe Deprez, Pavel V Shevchenko, and Mario V Wüthrich. Machine learning techniques for mortality modeling. *European Actuarial Journal*, 7:337–352, 2017.

[8] David Furman, Judith Campisi, Eric Verdin, Pedro Carrera-Bastos, Sasha Targ, Claudio Franceschi, Luigi Ferrucci, Derek W Gilroy, Alessio Fasano, Gary W Miller, et al. Chronic inflammation in the etiology of disease across the life span. *Nature medicine*, 25(12):1822–1832, 2019.

[9] Benjamin Gompertz. Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value

of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115):513–583, 1825.

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[11] Samuel Asante Gyamerah, Aaron Akyea Mensah, Clement Asare, and Nelson Dzupire. Improving mortality forecasting using a hybrid of lee–carter and stacking ensemble model. *Bulletin of the National Research Centre*, 47(1):158, 2023.

[12] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[13] Wei Hong Hong, Jia Hui Yap, Ganeshsree Selvachandran, Pham Huy Thong, and Le Hoang Son. Forecasting mortality rates using hybrid lee–carter model, artificial neural network and random forest. *Complex & Intelligent Systems*, 7:163–189, 2021.

[14] Fanny Janssen. Advances in mortality forecasting: introduction. *Genus*, 74(1):21, 2018.

[15] Hans-Peter Kohler and Jere R Behrman. Benefits and costs of the population and demography targets for the post-2015 development agenda. *Copenhagen Consensus Center Population and Demography Assessment Paper (as of 3 October)*, 2014.

[16] Nikolaos Kourentzes, Devon K Barrow, and Sven F Crone. Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, 41(9):4235–4244, 2014.

[17] Jennifer Lee, Jinjin Cai, Fuhai Li, and Zachary A Vesoulis. Predicting mortality risk for preterm infants using random forest. *Scientific reports*, 11(1):7308, 2021.

[18] Ronald D. Lee and Lawrence R. Carter. Modeling and forecasting u.s. mortality. *Journal of the American Statistical Association*, 87(419): 659–671, 1992.

[19] Cody Maklin. Forecast time series with LSTMs using TensorFlow 2 and Keras in Python. https://www.justintodata.com/forecast-time-series-lstm-with-tensorflow-keras/, n.d. Accessed: [Insert Access Date].

[20] Mario Marino, Susanna Levantesi, and Andrea Nigri. A neural approach to improve the lee-carter mortality density forecasts. *North American Actuarial Journal*, 27(1):148–165, 2023.

[21] Amber L Mueller, Maeve S McNamara, and David A Sinclair. Why does covid-19 disproportionately affect older people? *Aging (albany NY)*, 12 (10):9959, 2020.

[22] Andrea Nigri, Susanna Levantesi, Mario Marino, Salvatore Scognamiglio, and Francesca Perla. A deep learning integrated lee–carter model. *Risks*, 7(1):33, 2019.

[23] Carl A Pierce, Paula Preston-Hurlburt, Yile Dai, Clare Burn Aschner, Natalia Cheshenko, Benjamin Galen, Scott J Garforth, Natalia G Herrera, Rohit K Jangra, Nicholas C Morano, et al. Immune responses to sars-cov-2 infection in hospitalized pediatric and adult patients. *Science translational medicine*, 12(564):eabd5487, 2020.

[24] Samuel H Preston and Andrew Stokes. Sources of population aging in more and less developed countries. *Population and Development Review*, 38(2):221–236, 2012.

[25] Arthur E Renshaw and Steven Haberman. A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3):556–570, 2006.

[26] Ronald Richman and Mario V Wüthrich. A neural network extension of the lee–carter model to multiple populations. *Annals of Actuarial Science*, 15(2):346–366, 2021.

[27] Salvatore Scognamiglio, Francesca Perla, Mario Marino, Andrea Nigri, and Susanna Levantesi. A deep learning integrated lee–carter model. *Risks*, 7:33, 03 2019. doi: 10.3390/risks7010033.

[28] Lenny Stoeldraijer, Coen van Duin, Leo van Wissen, and Fanny Janssen. Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the netherlands. *Demographic Research*, 29:323–354, 2013.

[29] United Nations, Department of Economic and Social Affairs, Population Division. World population ageing 2019: Highlights, 2019. https:

//www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Highlights.pdf.

[30] John R Wilmoth, Kirill Andreev, Dmitri Jdanov, Dana A Glei, C Boe, M Bubenheim, D Philipov, V Shkolnikov, and P Vachon. Methods protocol for the human mortality database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality. org [version 31/05/2007]*, 9:10–11, 2007.

[31] World Health Organization. World health statistics 2022. https://www.who.int/news/item/20-05-2022-world-health-statistics-2022, 2022. Accessed: 2024-10-18.

[32] Yu-Tzu Wu, Albert Sanchez Niubo, Christina Daskalopoulou, Dario Moreno-Agostino, Denes Stefler, Martin Bobak, Sian Oram, Martin Prince, and Matthew Prina. Sex differences in mortality: results from a population-based study of 12 longitudinal cohorts. *Cmaj*, 193(11): E361–E370, 2021.

[33] Anna Zajacova and Elizabeth M. Lawrence. The relationship between education and health: Reducing disparities through a contextual approach. *Annual Review of Public Health*, 39:273–289, 2018. doi: 10.1146/annurev-publhealth-031816-044628.

[34] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.