

# Forecasting Mortality Rates: A Comparative Analysis of LSTM-Based and Lee-Carter Models

Caroliny Rodrigues Nascimento  
Master's Program in Computer Engineering  
University of Pernambuco  
Recife, Brazil  
crn1@ecomppoli.br

Roberta Andrade de Araújo Fagundes  
Master's Program in Computer Engineering  
University of Pernambuco  
Recife, Brazil  
roberta.fagundes@upe.br

**Abstract**— Accurate mortality forecasting is essential for informing long-term planning in areas such as public health, pension systems, and insurance. However, capturing the underlying patterns in mortality data remains challenging due to demographic shifts, regional disparities, and the presence of complex nonlinear trends over time. This study investigates the application of deep learning architectures (Long Short-Term Memory, Bidirectional Long Short-Term Memory, and Convolutional Neural Networks combined with Long Short-Term Memory) for forecasting mortality rates, and compares their predictive performance with the traditional stochastic Lee-Carter model. Mortality data from the Human Mortality Database, covering the United States from 1941 to 2020 and segmented by demographic region and gender, were used to train and evaluate the models. A hyperparameter optimization process using random search was conducted prior to model training. Performance was assessed using mean absolute error and root mean square error. The results demonstrate the superior performance of deep learning models compared to the Lee-Carter model. These findings support the use of neural networks as a robust and flexible alternative for mortality forecasting in demographic studies.

**Keywords**— Mortality rate, Machine Learning, LSTM, Bi-LSTM, CNN-LSTM

## I. INTRODUCTION

The analysis and forecasting of mortality rates constitute a fundamental activity for disciplines such as demography, public health, insurance, and social security. Traditional statistical models, notably the Lee-Carter model [1], have been extensively employed for this purpose over the past decades, offering an efficient stochastic approach to describe mortality patterns and project their temporal evolution. However, such methods are predicated on linear assumptions and simplifications that may not adequately capture the complex and nonlinear dynamics often observed in long-term time series.

With recent advancements in the field of artificial intelligence, particularly in deep learning, novel approaches have been explored for time series modeling. Recurrent neural networks, such as Long Short-Term Memory (LSTM) networks, have proven particularly effective in capturing long-range temporal dependencies, thereby overcoming limitations inherent in conventional architectures for forecasting tasks. Extensions like Bi-LSTM (Bidirectional LSTM) enable the exploitation of contextual information from both directions of the temporal sequence, while hybrid architectures such as

CNN-LSTM integrate the local feature extraction capabilities of convolutional networks (CNNs) with the sequential modeling prowess of LSTMs.

From a technical standpoint, this study distinguishes itself by conducting a controlled and reproducible comparative analysis between the Lee-Carter model and a suite of LSTM-based deep learning architectures, namely standard LSTM, Bidirectional LSTM (Bi-LSTM), and CNN-LSTM, within a unified experimental framework. While previous works have explored both deep learning models and hybrid combinations with statistical approaches, few have implemented a comprehensive comparison across multiple LSTM-based architectures under standardized conditions. The methodology ensures consistency in data preprocessing, temporal segmentation, hyperparameter optimization (via Random Search), and evaluation metrics, enabling a fair assessment of the architectural contributions of each model. An additional technical contribution of this work is the full implementation of the Lee-Carter model in Python, built from scratch using matrix decomposition techniques. This represents a new computational resource, as most prior studies rely on statistical software or closed packages, limiting integration with deep learning pipelines.

Beyond its methodological contributions, this study addresses a critical practical demand: enhancing the precision and adaptability of mortality forecasts in the face of rapidly changing demographic and epidemiological landscapes. Accurate mortality projections are essential for long-term financial planning in pension and retirement systems, enabling policymakers to anticipate population aging and adjust benefits and contributions accordingly. In the insurance sector, precise mortality models support risk-based pricing, solvency assessment, and capital reserve planning. Moreover, in public health, mortality forecasting informs the strategic allocation of healthcare resources, the design of intervention programs, and preparedness for emerging health crises, such as pandemics or climate-related mortality shifts.

The remainder of this paper is organized as follows: Section 2 presents related work on mortality forecasting and machine learning models. Section 3 describes the dataset and the theoretical foundations of the Lee-Carter model and LSTM-based architectures. Section 4 details the proposed methodology, including preprocessing, training, and optimization procedures. Section 5 reports and discusses the results obtained from the experiments. Finally, Section 6 provides concluding

remarks and suggests directions for future research.

## II. RELATED WORK

The Lee-Carter model [1] stands as one of the most influential and widely adopted approaches in actuarial mortality projections. Since its inception, numerous extensions have been developed, including multivariate versions, models incorporating cohort effects, and applications across different countries and population strata [2]–[4]. Nevertheless, such models remain reliant on linear assumptions and may exhibit limitations in contexts characterized by more intricate temporal structures.

With the advent of deep learning techniques, neural network-based approaches have been increasingly applied to the modeling of mortality time series. Perla et al. [5] explored the use of LSTM networks for modeling mortality rate series and demonstrated that these models can outperform classical statistical models in terms of predictive accuracy. Lindholm and Palmborg [6] investigated strategies for improving data efficiency in training LSTM networks, including alternative temporal partitioning and ensembling techniques.

Other studies have explored more complex architectures. Zain and Alturki [10] proposed a hybrid CNN-LSTM model, combining local pattern extraction via convolutional layers with the sequential modeling capabilities of LSTMs. Their results showed significant improvements, particularly in series exhibiting seasonality and noise. Lu et al. [11] also demonstrated that CNN-LSTM models can outperform traditional autoregressive models and pure LSTMs in both short- and long-term forecasts. Chen and Khaliq [8], as well as Mathonsi and Van Zyl [12], further emphasize the efficacy of deep learning-based models in mortality forecasting, even in contexts marked by high variability and missing data.

A promising research direction involves the hybridization of classical statistical models with machine learning techniques. For example, Jo et al. [13] proposed a model combining LSTM networks with latent topic modeling to enhance mortality prediction.

Contributions from the Brazilian Congress on Computational Intelligence (CBIC) have also advanced technical development in time series analysis. Barreto and Albuquerque [14] proposed an unsupervised approach using clustering-based local autoencoders for novelty detection in time series. Silva [15] introduced a probabilistic fuzzy methodology with seasonal ensembles for forecasting complex time series. Campos et al. [16] applied neural stochastic processes to monthly hydrological data, showing the capacity of these models to capture patterns with high variability. Leite et al. [17] evaluated the application of LSTM networks in real-world time series forecasting, focusing on robustness and stability under noisy data conditions.

Another promising avenue involves the use of transfer learning for mortality forecasting in countries with limited data. Gautam et al. [18] applied LSTM with transfer learning to forecast COVID-19 cases and deaths in low-resource settings, demonstrating the feasibility of training effective models even in data-scarce environments. This type of approach is particularly

relevant for mortality forecasting in underrepresented or developing populations, where data collection can be inconsistent.

## III. MODELS

### A. Lee-Carter

Lee-Carter model is used to model and project time series of mortality rates. It is a log-linear model that decomposes the observed mortality rate into three main components: an average age pattern, a temporal mortality index, and an age sensitivity factor to temporal variations. The basic formulation of the model is given by:

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t} \quad (1)$$

Where

- $m_{x,t}$  represents the observed central mortality rate at age  $x$  in year  $t$ ;
- $a_x$  is the average log-mortality pattern for each age  $x$ , capturing the general level of mortality by age across the entire observation period;
- $b_x$  is the age-specific sensitivity parameter, which measures how mortality at a given age responds to changes in the overall mortality trend, larger values of  $b_x$  indicate that age group  $x$  is more affected by variations in  $k_t$ ;
- $k_t$  is the time-varying mortality index, a latent factor that summarizes the evolution of mortality over time, usually showing a long-term decline as mortality improves;
- $\epsilon_{x,t}$  represents the error term, accounting for the unexplained variation in mortality rates.

The estimation of  $a_x, b_x$  and  $k_t$  in (1) is performed by applying Singular Value Decomposition (SVD) to the centered log-mortality matrix. Because the decomposition is not unique, identification constraints are imposed (2)(3). These restrictions ensure that the parameters are uniquely defined and interpretable.

$$\sum_t k_t = 0 \quad (2)$$

$$\sum_x b_x = 0 \quad (3)$$

In forecasting applications, the mortality index  $k_t$  is treated as a time series and typically projected using autoregressive integrated moving average (ARIMA) models, while the age parameters  $a_x$  and  $b_x$  remain fixed after estimation. This structure provides a parsimonious but powerful framework, capable of capturing both the general mortality decline and the heterogeneous responses across different ages. Despite its simplicity, the Lee-Carter model has become a benchmark in demography and actuarial science due to its interpretability, statistical rigor, and forecasting performance [1].

## B. LSTM

Long Short-Term Memory (LSTM) networks are a specialized class of recurrent neural networks (RNNs) developed to address the difficulty of learning long-term dependencies in sequential data. Introduced by Hochreiter and Schmidhuber [19], LSTMs are particularly effective in scenarios where the prediction of future values depends heavily on a long and complex historical context, such as in time series forecasting.

In this context, LSTMs are especially advantageous due to their ability to capture long-range temporal dependencies, which are often critical for achieving accurate forecasts. This is essential in time series data, where the future state is influenced by multiple past values, and the relationships among them may be non-linear and time-variant. The LSTM architecture includes three key components—gates—that regulate the flow of information through the network: the input gate, the forget gate, and the output gate. The input gate controls which new values from the current input should be stored in the memory cell. The forget gate determines which information from the past should be discarded, allowing the model to eliminate irrelevant or outdated data. The output gate decides which information from the memory cell should be output, ensuring that only the most relevant information contributes to the prediction [19].

Throughout both training and inference, LSTMs process time series data in a stepwise fashion. First, they evaluate which parts of the historical data can be forgotten. Next, they identify and incorporate new information deemed relevant for the prediction task. The memory cell is updated by integrating both retained past information and newly acquired data, enabling the model to preserve a dynamic and contextually appropriate internal representation. Finally, the LSTM produces a forecast based on the updated memory state, drawing on the most relevant elements stored in memory to generate an accurate prediction for the next value in the series [20].

## C. BI-LSTM

Bidirectional Long Short-Term Memory (Bi-LSTM) networks represent an extension of LSTM networks specifically designed to capture sequence information from both directions; that is, from past to future and from future to past. This inherent capability renders Bi-LSTMs particularly effective in time series tasks where contextual dependencies can be intricate and non-linear across time.

The primary distinction between a conventional LSTM and a Bi-LSTM lies in the Bi-LSTM's utilization of two LSTM layers operating in opposing directions. One layer processes the input sequence in the forward temporal direction (from beginning to end), while the other layer processes the sequence in the reverse direction (from end to beginning). Consequently, each temporal point within the sequence gains access to information from both its past and future simultaneously, providing a more comprehensive context for generating predictions [20].

## D. CNN-LSTM

The CNN-LSTM architecture embodies a hybrid

approach that synergistically combines Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks. This combination aims to harness the spatial feature extraction capabilities of CNNs alongside the ability of LSTMs to model long-range temporal dependencies. This integrated structure proves particularly advantageous in time series modeling tasks where inputs exhibit both relevant local patterns and, concurrently, a sequential structure over time.

The initial phase of the architecture involves one or more one-dimensional convolutional layers (1D-CNN), which are responsible for applying sliding filters across the input sequences. These layers autonomously learn pertinent local representations by capturing short-range patterns, such as trends, oscillations, or localized anomalies within the time series. Subsequent to convolution, pooling operations can be applied for dimensionality reduction and reinforcement of local invariance.

The outputs of the CNN are then reconfigured as temporal sequences and fed into one or more LSTM layers. These LSTM layers are tasked with capturing long-term temporal dependencies between the patterns extracted by the convolutions. The LSTM maintains updated hidden states throughout the sequence, preserving relevant information and discarding noise via its input, forget, and output gates. The ultimate output of the LSTM can be connected to dense layers for regression or classification objectives.

From a computational perspective, CNN-LSTM enables the model to jointly and in a supervised manner learn both spatial representations and temporal dependencies [10], [11].

## IV. PROPOSED METHOD

The methodological flow adopted in this study, depicted in Figure 1, was conceived with the objective of ensuring a systematic and replicable analysis of mortality time series.

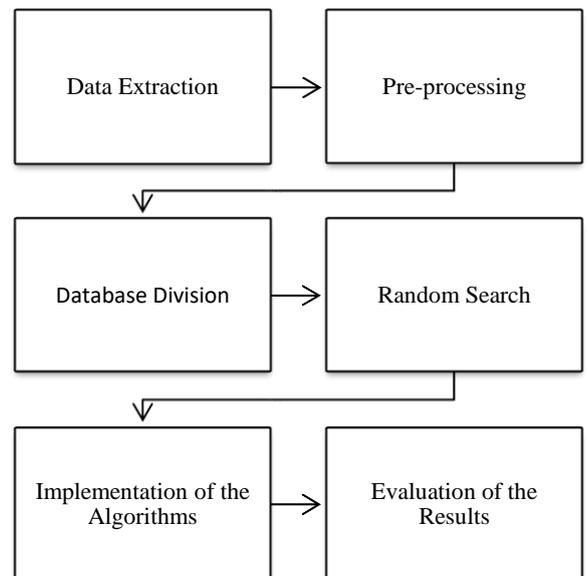


Fig.1. Stages of the applied methodology

### A. Data Extraction

For the experiment, 18 distinct databases were utilized, each comprising 8,881 records, organized by demographic region and gender segmentation.

These datasets were extracted from the Human Mortality Database (HMD) platform, a widely recognized source in actuarial and demographic studies. The records within these databases correspond to age-specific mortality rates, organized on an annual frequency, spanning the period from 1941 to 2020.

### B. Preprocessing

The data preprocessing stage commenced with the consolidation of the 18 individual databases. The columns of interest were extracted: the year vector (Year) and the time series of mortality rates (mx).

To facilitate the application of LSTM, Bi-LSTM, and CNN-LSTM models, the series was reshaped into a one-dimensional format. Subsequently, it was normalized to enhance training stability and optimize the convergence of deep learning algorithm optimization algorithms.

### C. Database Division

Following preprocessing, the data underwent division into training and test sets, adhering to an 80% for training and 20% for testing proportion. Given the sequential nature of temporal data, historical data preceding 2004 were allocated to compose the training set, while subsequent years were designated for the test set, specifically for forecasting mortality rates.

### D. Random Search

With the datasets appropriately partitioned, hyperparameter optimization for the models was conducted utilizing the Random Search technique. This approach was selected due to its efficiency in exploring high-dimensional search spaces, thereby enabling the identification of optimal configurations that minimize errors [15].

In this experiment, search spaces were defined for parameters such as: number of hidden units (ranging between 32 and 256) and dropout rate (between 0.0 and 0.5). Each combination was evaluated based on the validation error. Additionally, the models were configured for a one-step-ahead forecasting approach, utilizing a look-back period of 5 years to predict the mortality rate for the subsequent year.

### E. Application of Algorithms

For the experiment using the Lee-Carter model, an implementation in Python was necessary (Fig. 2). The data were organized into a two-dimensional matrix with ages in rows and years in columns. The mortality rate was transformed using the natural logarithm. The average by age  $a_x$  was then calculated as the average of each row of this log-transformed matrix.

Subsequently, the matrix was centered by age, subtracting  $a_x$  from each row, and singular value decomposition (SVD) was applied to the centered matrix. The first singular vector of the matrix  $Uf$  was used as an estimate of  $b_x$ , while the first vector of  $V^T$  weighted by the

corresponding singularity, provided the values of  $k_t$ . To ensure the identification of the model,  $b_x$  as normalized so that the sum of its elements was equal to 1, adjusting  $k_t$  proportionally.

Finally, with the parameters  $a_x, b_x \in k_t$  estimated, it was possible to reconstruct the adjusted log-mortality matrix.

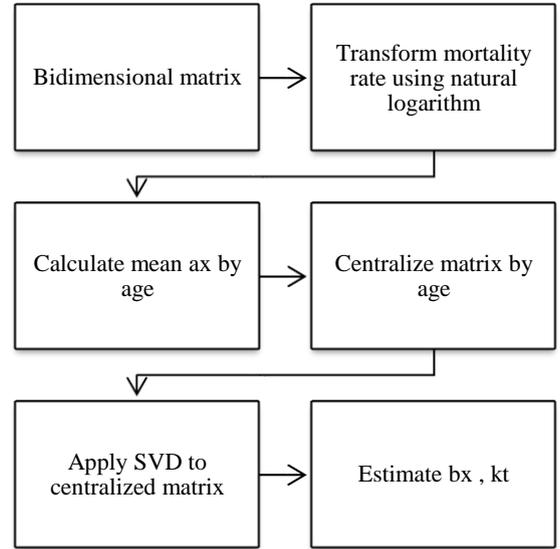


Fig.2. Stages of Lee-Carter Model Construction in Python

The construction of the LSTM, Bi-LSTM, and CNN-LSTM models was performed using the Keras library, integrating the optimized hyperparameters identified in the preceding step.

The LSTM model was structured with a single LSTM layer, where the number of neurons varied between 32 and 256, defined as a search hyperparameter. Dropout was incorporated after the LSTM layer, with a discard rate ranging from 0 to 0.5. The network's final layer consists of a dense neuron, suitable for univariate time series forecasting.

In the case of the Bi-LSTM model, the salient distinction from the standard LSTM is the deployment of a bidirectional layer, which enables the network to process the input sequence in both directions (forward and backward in time), thereby facilitating the capture of more intricate temporal relationships. The remaining configurations, such as dropout and compilation parameters, adhered to the pattern established for the previous model.

The CNN-LSTM model, conversely, adopts a hybrid approach, integrating convolutions with temporal memory. Initially, a one-dimensional convolutional layer is applied to the input, with an adjustable number of filters between 16 and 64. Subsequently, the data traverse a max pooling layer and a flatten layer, before being processed by a dense layer with 32 to 128 units. Following dropout, the network culminates with an output neuron.

### F. Evaluation of Results

In The performance of the models was evaluated using two standard error metrics in forecasting: Mean

Absolute Error (MAE) and Root Mean Squared Error (RMSE). MAE measures the average magnitude of prediction errors, calculated as the mean of the absolute differences between predicted and observed values. Its key advantage lies in its interpretability, as it expresses prediction errors in the same units as the original data. Unlike RMSE, MAE does not disproportionately penalize large errors, making it more robust in the presence of outliers [7]. In mortality forecasting, MAE serves as a reliable indicator of the average deviation across forecasts, offering a practical assessment of model precision [8].

RMSE, defined as the square root of the mean of the squared differences between predicted and observed values, is more sensitive to large errors due to the squaring operation. By penalizing larger deviations more heavily, RMSE is particularly sensitive to high-variance errors, thus highlighting the presence of substantial discrepancies between predicted and actual values. In the context of LSTM-based forecasting, RMSE complements MAE by quantifying the variability and consistency of model predictions. Recent studies in time series modeling have demonstrated the value of using both MAE and RMSE jointly for a more comprehensive evaluation of model performance [8], [9].

## V. RESULTS

Hyperparameter optimization was conducted utilizing the Random Search technique, with the objective of identifying the optimal configurations for the employed deep learning models (LSTM, Bi-LSTM, and CNN-LSTM). Table I summarizes the hyperparameters ascertained for each architecture, including the number of neurons, batch size, number of epochs, and dropout rate. This phase proved pivotal in ensuring that the models were trained effectively, thereby maximizing their predictive performance and mitigating overfitting.

TABLE I.

COMBINATION OF HYPERPARAMETERS FOUND USING RANDOM SEARCH

Models	Neurons	Batch Size	Epochs	Dropout
LSTM	64	64	100	0.1
BI-LSTM	96	64	100	0.1
CNN-LSTM	64	64	100	0.1

The results obtained from the application of the models (Table II) unequivocally demonstrate the superior performance of deep learning-based models over the traditional stochastic Lee-Carter model in the task of forecasting mortality rates. As presented, the Lee-Carter model yielded a MAE of 0.2671 and an RMSE of 0.3613, values substantially higher than those achieved by the LSTM, Bi-LSTM, and CNN-LSTM models.

Among the proposed models, the three recurrent neural network algorithms exhibited remarkably similar performances, with particular emphasis on the LSTM model, which attained the lowest Mean Absolute Error (MAE = 0.0105), while the Bi-LSTM achieved the lowest (RMSE = 0.0494). The CNN-LSTM model, although

presenting a numerically slightly inferior performance to the others (MAE = 0.0112; RMSE = 0.0499), still demonstrates considerable efficacy.

TABLE II.

RESULTS OF THE EVALUATION METRICS

Model	MAE	RMSE
Lee-Carter	0.2671	0.3613
LSTM	0.0105	0.0497
Bi-LSTM	0.0109	0.0494
CNN-LSTM	0.0112	0.0499

These results indicate that the adoption of deep learning-based architectures, especially when coupled with appropriate hyperparameter optimization strategies such as Random Search, can yield highly effective models for time series forecasting within a demographic context. This is further substantiated by the visual analysis presented in Figure 3.

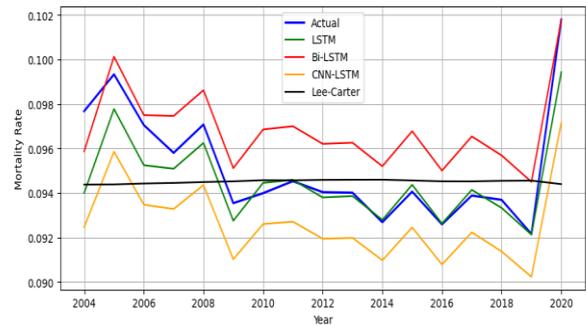


Fig.3. Mortality rate forecast

Figure 3 illustrates the performance of various models in forecasting mortality rates across the years. It is observed that the Lee-Carter model, represented by the black line, while satisfactorily capturing the general level of mortality over time, exhibits a tendency to underestimate the more pronounced annual variations. The inherent smoothness of its forecast curve indicates a limitation of the model in precisely capturing the temporal volatility of the series, particularly in the most recent years, such as 2019 and 2020, where a more discernible divergence between predicted and observed values occurs.

In stark contrast, the deep learning-based models, namely LSTM (green line), Bi-LSTM (red line), and CNN-LSTM (orange line), demonstrate a heightened sensitivity to short-term fluctuations in mortality rates, thereby more accurately tracking the oscillations of the time series. The "Actual" line (blue), representing the true observed data, maintains a significantly closer proximity to the predictions generated by the LSTM and Bi-LSTM models across the majority of the temporal interval. This unequivocally evidences an enhancement in generalization capability when compared to the Lee-Carter model.

Specifically, the LSTM model exhibits a robust capacity for trend following, while the Bi-LSTM, by leveraging bidirectionality to consider both the past and future context of each data point during training, contributes to an even more refined estimate closely aligned with the actual data. Nevertheless, it is notable that in certain years, the predictions from these models display slight deviations, which may suggest potential overfitting or excessive sensitivity to noise present in the data. The CNN-LSTM model, despite also capturing temporal variations, demonstrates greater instability in intermediate years and tends to exhibit a larger divergence from observed values when compared to LSTM and Bi-LSTM. This suggests that the integration of convolutional and recurrent layers, while powerful in other applications, might necessitate a more refined configuration or a more extensive dataset to achieve optimal performance within this specific context.

## VI. CONCLUSION

This study underscores the effectiveness of deep learning architectures, specifically LSTM, Bi-LSTM, and CNN-LSTM, in the task of forecasting mortality rates, especially when compared to the traditional Lee-Carter model. Through rigorous hyperparameter optimization using the Random Search technique, optimal configurations were identified for each model, resulting in robust training performance and reduced risk of overfitting.

The empirical results leave little doubt as to the superior predictive power of deep learning models. All three neural network architectures significantly outperformed the Lee-Carter model across both MAE and RMSE metrics, with LSTM achieving the lowest MAE and Bi-LSTM achieving the lowest RMSE. Although the CNN-LSTM model showed slightly higher error values, it still delivered a strong overall performance, affirming its viability in demographic time series forecasting.

Furthermore, visual inspection of the forecast outputs reinforces the quantitative findings. The deep learning models more closely followed the actual mortality trends, particularly in years marked by greater variability, such as 2019 and 2020. The Lee-Carter model, by contrast, failed to capture these fluctuations with precision, revealing its structural limitations in handling volatile time series.

Notably, the bidirectional structure of Bi-LSTM allowed for an enriched contextual understanding of temporal dynamics, whereas the CNN-LSTM model, while promising, exhibited signs of instability in mid-series predictions, possibly due to architectural complexity or limited training data.

In conclusion, this research highlights the clear advantages of adopting deep learning techniques for mortality forecasting, especially in scenarios involving non-linear trends and complex dependencies. When paired with appropriate hyperparameter tuning strategies, these models offer not only improved accuracy but also enhanced responsiveness to demographic shifts. Future work may explore hybrid approaches and data augmentation techniques to further stabilize performance and extend applicability across diverse population groups.

## REFERENCES

- [1] R. D. Lee and L. R. Carter, "Modeling and forecasting U.S. mortality," *Journal of the American Statistical Association*, vol. 87, no. 419, pp. 659–671, 1992.
- [2] H. Booth, R. Hyndman, L. Tickle, and P. de Jong, "Lee-Carter mortality forecasting: A multi-country comparison," *Population Studies*, vol. 56, no. 3, pp. 325–348, 2002.
- [3] I. L. Danesi, S. Haberman, and P. Millossovich, "Forecasting mortality in subpopulations using Lee-Carter type models: A comparison," *Insurance: Mathematics and Economics*, vol. 62, pp. 151–161, 2015.
- [4] J. Brouhns, M. Denuit, and J. Vermunt, "A Poisson log-bilinear regression approach to the construction of projected lifetables," *Insurance: Mathematics and Economics*, vol. 31, no. 3, pp. 373–393, 2002.
- [5] F. Perla, R. Richman, S. Scognamiglio, and M. V. Wüthrich, "Time-series forecasting of mortality rates using deep learning," *Scandinavian Actuarial Journal*, vol. 2021, no. 8, pp. 713–733, 2021.
- [6] M. Lindholm and L. Palmborg, "Efficient use of data for LSTM mortality forecasting," *European Actuarial Journal*, vol. 12, pp. 443–469, 2022.
- [7] J. C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.
- [8] Y. Chen and A. Q. M. Khaliq, "Comparative study of mortality rate prediction using data-driven recurrent neural networks and the Lee-Carter model," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 134, 2022.
- [9] G. Li et al., "Forecasting and analyzing influenza activity in Hebei Province, China, using a CNN-LSTM hybrid model," *BMC Public Health*, vol. 24, no. 1, p. 2171, 2024.
- [10] Z. M. Zain and N. M. Alturki, "COVID-19 pandemic forecasting using CNN-LSTM: a hybrid approach," *Journal of Control Science and Engineering*, vol. 2021, no. 1, p. 8785636, 2021.
- [11] Y. Lu et al., "A hybrid CNN-RNN approach for survival analysis in a Lung Cancer Screening study," *Heliyon*, vol. 9, no. 8, 2023.
- [12] T. Mathonsi and T. L. van Zyl, "A statistics and deep learning hybrid method for multivariate time series forecasting and mortality modeling," *Forecasting*, vol. 4, no. 1, pp. 1–25, 2021.
- [13] Y. Jo, L. Lee, and S. Palaskar, "Combining LSTM and latent topic modeling for mortality prediction," *arXiv preprint arXiv:1709.02842*, 2017.
- [14] G. A. Barreto and R. F. Albuquerque, "Unsupervised Time Series Novelty Detection Using Clustering-based Local Autoencoders," in *Anais do XVI Congresso Brasileiro de Inteligência Computacional (CBIC)*, Salvador, Brasil, 2023.
- [15] P. C. Silva, "Probabilistic Forecasting with Seasonal Ensemble Fuzzy Time-Series," in *Anais do XIII Congresso Brasileiro de Inteligência Computacional (CBIC)*, Curitiba, Brasil, 2017.
- [16] L. C. D. Campos, M. M. B. R. Vellasco, and J. G. Lazo, "Processo Estocástico Neural Aplicado em Séries de Afluências Mensais," in *Anais do X Congresso Brasileiro de Inteligência Computacional (CBIC)*, Fortaleza, Brasil, 2011.
- [17] S. Leite, R. Limão, and L. Campos, "Predição de Séries Temporais da COVID-19: Uma Avaliação de Redes Neurais com Células LSTM," in *Anais do XV Congresso Brasileiro de Inteligência Computacional (CBIC)*, Joinville, Brasil, 2021.
- [18] A. Gautam, M. Menon, and S. Sethi, "Forecasting COVID-19 deaths in low-resource settings using transfer learning and LSTM," *PLoS ONE*, vol. 16, no. 6, e0253272, 2021.
- [19] Hochreiter, Sepp and Schmidhuber, Jürgen, "Long short-term memory", *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997.
- [20] X. Chen and P. Moraga, "Assessing dengue forecasting methods: a comparative study of statistical models and machine learning techniques in Rio de Janeiro, Brazil," *Tropical Medicine and Health*, vol. 53, no. 1, p. 52, 2025.