“Longevity risk management through Machine Learning: state of the art”

AUTHORS	Susanna Levantesi https://orcid.org/0000-0002-4644-4358 https://publons.com/researcher/3409859/susanna-levantesi/ Andrea Nigri https://orcid.org/0000-0002-2707-3678 https://publons.com/researcher/2987858/andrea-nigri/ Gabriella Piscopo http://orcid.org/0000-0003-1292-5767 https://publons.com/researcher/3840171/gabriella-piscopo/

ARTICLE INFO	Susanna Levantesi, Andrea Nigri and Gabriella Piscopo (2020). Longevity risk management through Machine Learning: state of the art. *Insurance Markets and Companies*, *11*(1), 11-20. doi:10.21511/ins.11(1).2020.02
DOI	http://dx.doi.org/10.21511/ins.11(1).2020.02
RELEASED ON	Wednesday, 25 November 2020
RECEIVED ON	Friday, 16 October 2020
ACCEPTED ON	Monday, 16 November 2020

LICENSE	This work is licensed under a Creative Commons Attribution 4.0 International License

JOURNAL	"Insurance Markets and Companies"
ISSN PRINT	2616-3551
ISSN ONLINE	2522-9591
PUBLISHER	LLC “Consulting Publishing Company “Business Perspectives”
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
44	0	0

© The author(s) 2026. This publication is an open access article.

BUSINESS PERSPECTIVES

LLC "CPC "Business Perspectives"
Hryhorii Skovoroda lane, 10,
Sumy, 40022, Ukraine

www.businessperspectives.org

Susanna Levantesi (Italy), **Andrea Nigri** (Italy), **Gabriella Piscopo** (Italy)

LONGEVITY RISK MANAGEMENT THROUGH MACHINE LEARNING: STATE OF THE ART

Abstract

Longevity risk management is an area of the life insurance business where the use of Artificial Intelligence is still underdeveloped. The paper retraces the main results of the recent actuarial literature on the topic to draw attention to the potential of Machine Learning in predicting mortality and consequently improving the longevity risk quantification and management, with practical implication on the pricing of life products with long-term duration and lifelong guaranteed options embedded in pension contracts or health insurance products. The application of AI methodologies to mortality forecasts improves both fitting and forecasting of the models traditionally used. In particular, the paper presents the Classification and the Regression Tree framework and the Neural Network algorithm applied to mortality data. The literature results are discussed, focusing on the forecasting performance of the Machine Learning techniques concerning the classical model. Finally, a reflection on both the great potentials of using Machine Learning in longevity management and its drawbacks is offered.

Keywords Classification, Regression Tree, longevity risk, Machine Learning, Neural Network

JEL Classification C6, G22, J11

Received on: 16th of October, 2020
Accepted on: 16th of November, 2020
Published on: 25th of November, 2020

© Susanna Levantesi, Andrea Nigri, Gabriella Piscopo, 2020

Susanna Levantesi, Associate Professor, Department of Statistics, Sapienza University of Rome, Italy.

Andrea Nigri, Ph.D., Department of Statistics, Sapienza University of Rome, Italy.

Gabriella Piscopo, Associate Professor, Department of Economics and Statistical Science, University of Naples Federico II, Italy. (Corresponding author)

This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conflict of interest statement:
Author(s) reported no conflict of interest

INTRODUCTION

The availability of large datasets and the advances in Artificial Intelligence (AI) to analyze and extract information from the data represent a challenge for the insurance sector. It is well known that the insurance industry is a data-driven business, so AI can have significant consequences on its processes and decisions. A survey conducted by Deloitte (2018) and the European Financial Management Association shows that the activities in which AI could have the strongest influence vary substantially from one sector to another. In the insurance sector, 56% of the respondents consider the Risk Management the area in which AI has the greatest impact, versus 29% in the banking sector.

Regarding the insurance business, longevity risk management is presented as part of the wider Enterprise Risk Management (ERM) of an insurance company (Pitacco, 2020). The ERM is a comprehensive approach that goes from risk identification to risk assessment (that goes from product design to pricing to natural hedging techniques), capital allocation, and risk monitoring. Certainly, there are areas of risk management where the use of AI is more developed, like the risk monitoring of the underwriting process, and others where the potential benefits of automated predictive modeling have not yet been fully exploited. This is the case of the management of longevity risk in life business, to assess which many insurers still rely on traditional methods and make predictions resorting to classical demographic frameworks based on

extrapolative models. Some stochastic longevity models have been proposed starting from the classical financial literature on interest rates, due to similarities between the process of mortality and interest rate (e.g., Milevsky & Promislow, 2001; Cairns et al., 2006; Biffis, 2004; Dahl, 2004; Schrager, 2006). Other models exploit traditional statistical tools to fit the real data model using the Singular Value Decomposition of the matrix of mortality rate by age and time or the Principal Component Analysis. The Lee-Carter model (Lee & Carter, 1992) is widely recognized as the cornerstone of mortality modeling and forecasting; it introduces a model for central death rates involving both age and time-dependent terms. The Renshaw-Haberman model (Renshaw & Haberman, 2003, 2006), for the first time, introduces a cohort parameter to catch the observed variations in mortality among individuals belonging to different cohorts. The assumption that the error term of the model is described by white noise is violated because the force of mortality at older ages rather shows a higher variability than at younger. For this reason, Brouhns et al. (2002) have extended the basic model by describing the number of deaths according to a Poisson distribution. Other approaches use penalized splines to smooth mortality and derive future mortality patterns (Currie et al., 2004). Generally, all these models have been applied to national data consisting of yearly observations. However, actuaries are more interested in deriving the underlying mortality of given portfolios. An important difference between mortality data aggregated at the national level and specific mortality relates to the range of the observation period: often, a given portfolio data does not contain observations during very long periods as in the national dataset, though it contains many other descriptive variables.

In Section 1, the advancements of the recent literature on longevity risk management with the introduction of Machine Learning techniques applied to mortality forecast are retraced to improve the classical models' predictability. Section 2 presents the family of the generalized age-period cohort models and the main accuracy measures. Section 3 retraces the method presented in the literature on mortality modeling with ML: a brief overview of the Classification and Regression Trees (CART) approach applied to mortality and a global perspective on the NN in mortality modeling is provided. An introduction to the basic idea behind NN architecture is presented, and the benefits of NN in both integration and full replacement of canonical models are explored. Sections 4 and 5 are devoted to describing and discussing the results and the benefits of using ML techniques as complementary to standard mortality models. This way of operating is likely to please the longevity risk managers who are unwilling to use not immediate explainable algorithms. The last section is devoted to the findings.

1. LITERATURE REVIEW

Recently, AI in general and Machine Learning (ML) in particular have appeared in actuarial research and practice. Moreover, while the scientific productions and the applied researches boast an increasing number of contributions in non-life insurance (Wüthrich et al., 2019; Ferrario et al., 2018; Gabrielli et al., 2019; Noll et al., 2020), the life insurance sector is suffering the unwillingness of the demographers to replace the traditional models with a sort of "black boxes", referring to how ML algorithms operate. Nevertheless, the first results of the application of ML techniques to mortality forecasts and longevity risk management have been presented during the last years. At first, some early unsupervised methods for mortality analysis have been introduced with application to different medicine fields; lately, they have been exploited by demographers (Carracedo et al., 2018) and actuaries (e.g., Huang & Browne, 2017; Piscopo et al., 2017). Deprez et al. (2017) introduce the use of supervised ML algorithms to improve the fitting of the log mortality rates. Levantesi and Pizzorusso (2018) extend the model and take advantage of ML to produce better mortality forecasts in the Lee-Carter model. The authors show that mortality modeling can benefit from ML, which better captures patterns that traditional models do not identify. Specifically, Deprez et al. (2017) use decision trees to detect the weaknesses of different mortality models and investigate the cause-of-death mortality. Levantesi and Pizzorusso (2018) use decision trees, random forest, and gradient boosting to calibrate an ML estimator. Finally, Levantesi and Nigri (2019) focus on the random forest al-

http://dx.doi.org/10.21511/ins.11(1).2020.02

gorithm. Hainaut (2018) proposes the application of Neural Network (NN) to model mortality rates. Richman and Wuthrich (2018) implement the Lee-Carter model in the multiple population setting where parameters are estimated using deep Neural Networks. Nigri et al. (2019) extend the original Lee-Carter model to catch the nonlinear mortality trend introducing a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) to improve the coherence of the mortality forecasts. Nigri et al. (2020) provide an autoregressive model leveraging Deep Learning (DL) to forecast, independently and simultaneously, summary longevity measures. Maier et al. (2019) discuss how it is possible to deepen a rich application data set with the survival modeling to create a life score used in an automatic underwriting system. A common component of the underwriting process is the estimation of individual-level mortality risk. Traditionally, this is performed manually using human judgment and point-based systems, subjected to inconsistency, limiting the ability to price the products efficiently. On the contrary, the availability of wide historical data sets represents a challenge for ML to support real-time automated decision-making deepening the available data.

ML techniques permit us to integrate a stochastic model with a data-driven approach. In addition to improving the underwriting process, a more accurate longevity risk quantification can sustain the reinsurance of this "toxic risk" (Blake et al., 2006) and the development of the longevity capital market. The contribution given by the use of hedging assets designed for longevity management has been highlighted in Cocco and Gomes (2012). Different longevity assets have been designed: firstly, mortality and longevity bonds, afterward forwards and swaps. The significant reduction in the forecasting error reached through the application of ML and DL techniques is particularly useful in the pricing of these products. The analysis of the impact of mortality rates provided by these models on the pricing of longevity derivatives has been investigated, for example, in Levantesi and Nigri (2019).

This paper aims to describe the main results of the recent actuarial literature on applying AI methodologies to mortality forecasts to improve both fitting and forecasting of the models traditionally used to stimulate their use in the actuarial practice, with relevant implications in the longevity risk management. Note that all the contributions mentioned here work on aggregate data by country, age, and gender, not on individual features.

2. METHODS

The class of the generalized age-period cohort (GAPC) models (Villegas et al., 2015) embraces most of the stochastic mortality models proposed in the literature, where the following predictor explains the effects of age a, calendar year t, and cohort c:

$$\eta_{a,t} = \alpha_a + \sum_{i=1}^{n} \beta_a^{(i)} \kappa_t^{(1)} + \beta_a^{(0)} \gamma_{t-a}, \qquad (1)$$

where α_a is the age-specific parameter giving the average age-specific mortality pattern; $\kappa_t^{(i)}$ denotes the time-varying index for the general mortality and $\beta_a^{(i)}$ modifies its effect across ages, so it represents the deviations from the trend of age as $\kappa_t^{(i)}$ varies, their product is the age-period factor of the mortality trends; γ_{t-a} is the cohort parameter and $\beta_a^{(0)}$ modifies its impact across ages ($c = t - a$ is the year of birth), for this $\beta_a^{(0)} \cdot \gamma_{t-a}$ is the cohort term.

According to the GAPC framework, the Lee-Carter model is described as follows:

$$\log\left(m_{a,t}\right) = \alpha_a + \beta_a^{(1)} \kappa_t^{(1)}, \qquad (2)$$

where the predictor $\eta_{a,t}$ is the logarithm of the central death rates, $m_{a,t}$. Throughout the paper, it will refer to the Lee-Carter model as extended by Brouhns et al. (2002), based on the assumption that the observed number of deaths, $D_{a,t}$, in the population are realizations the random variable, which follows a Poisson process.

It is necessary to introduce in the model two constraints to solve the identifiability questions of the parameters:

$$\sum_{t \in \mathcal{T}} \kappa_t^{(1)} = 0, \qquad \sum_{a \in \mathcal{A}} \beta_a^{(1)} = 1.$$

The time trend $\kappa_t^{(1)}$ is described by an autoregressive integrated moving average (ARIMA) process to estimate the future probabilities. An ARIMA(0,1,0) usually provides the best fit to the data.

Mortality forecasting literature does not provide a vast space for ML modeling. It is still unfamiliar to both actuaries and demographers. The main contributions are from Deprez et al. (2017), Levantesi and Pizzorusso (2019), Levantesi and Nigri (2019); the common idea behind all these works is to improve the fitting accuracy of canonical models using ML algorithms. In other words, correct the mortality surface produced by standard stochastic mortality models. All of the proposed methods, adopting CART, calibrate an ML estimator used to adjust (and improve) the mortality rates estimated by the original mortality model. Those authors show that mortality modeling can benefit from ML, which better captures patterns that traditional models do not identify. Specifically, Deprez et al. (2017) imply decision trees to overcome the weaknesses of different mortality models, also investigating the cause-of-death mortality. Levantesi and Pizzorusso (2019) use decision trees, random forest, and gradient boosting to calibrate an ML estimator. Finally, Levantesi and Nigri (2019) focus on the random forest algorithm.

All these contributions work on mortality data downloaded from the Human Mortality Database (HMD). At state of the art, there are still no academic contributions in the field of ML working with mortality data of insurance portfolios.

The regression and decision tree algorithm, known as CART, which stands for Classification and Regression Trees, works on the feature space. The algorithm provides a sub-spaces division of the predictor space, partitioning it into a sequence of binary splits (Hastie et al., 2016). The ML literature has rapidly shown relevant evolution, adopting the ensemble methods (e.g., random forest and gradient boosting machine) to improve CART performances aggregating many trees.

In the literature, the ML implementation of mortality models refers to four categorical variables identifying an individual: gender (*g*), age (*a*), calendar year (*t*), and year of birth (*c*). Therefore, the model assigns each individual the feature $x=(g,a,t,c)\in\mathcal{X}$ with $\mathcal{X}=\mathcal{G}\times\mathcal{A}\times\mathcal{T}\times\mathcal{C}$ the feature space, where:

$$\mathcal{G}=\{males,females\},\mathcal{A}=\{0,\ldots,\omega\},$$
$$\mathcal{T}=\{t_1,\ldots,t_n\},\mathcal{C}=\{c_1,\ldots,c_m\}.$$

The feature space X could be enriched with other information such as the income, the marital status, to be or not to be a smoker. The model requires that the number of deaths D_x satisfies the age independence condition in $x\in\mathcal{X}$ and follows a Poisson distribution, $D_x\sim Pois(m_x\cdot E_x)$ for all $x\in\mathcal{X}$, where m_x is the central death rate and E_x are the exposures. Let us denote the expected number of deaths of the chosen stochastic mortality model *j* as d_x^j and let m_x^j the central death rate. Following the framework in Deprez et al. (2017), the initial condition of the model is $m_x=m_x^j$ and:

$$D_x\sim Pois(\psi_x\cdot d_x^j),$$
$$\text{with}\quad \psi_x\equiv 1,\quad d_x^j=m_x^jE_x.$$

Note that the condition $\psi_x\equiv 1$ is equivalent to state that the mortality model completely fits the crude rates. This is an ideal condition as the model might overestimate $(\psi_x\le 1)$ or underestimate $(\psi_x\ge 1)$ the crude rates in the real world. The aim of the approach used in Deprez et al. (2017), Levantesi and Pizzorusso (2019), and Levantesi and Nigri (2019) is essential to calibrate the parameter ψ_x according to an ML algorithm in order to improve the fitting accuracy of the mortality model. The estimator ψ_x is found as a solution of a tree classification algorithm used on the ratio between the death observations and the death calculated through specified mortality model:

$$\frac{D_x}{d_x^j}\sim gender+age+year+cohort. \qquad (3)$$

Let $\hat{\psi}_x^{j,ML}$ denote the ML estimator that is solution of equation (3), where j is the mortality model and ML the ML technique. To reach a better fit of the observed data, the central death rate of the mortality model, m_x^j, are adjusted through $\hat{\psi}_x^{j,ML}$ as follows:

$$m_x^{j,ML}=\hat{\psi}_x^{j,ML}\cdot m_x^j,\quad \forall x\in\mathcal{X}. \qquad (4)$$

The mortality improvement reached by the ML algorithm is given by the relative changes of central death rates $\Delta m_x^{j,ML}=\hat{\psi}_x^{j,ML}-1$. The values of $\hat{\psi}_x^{j,ML}$ are then obtained by applying an ML algorithm. As shown in Levantesi and Nigri (2019), this approach can also be used to analyze the traditional mortality models' limits.

http://dx.doi.org/10.21511/ins.11(1).2020.02

The ML algorithms used in the application provided by the specific literature on mortality modeling are Decision Tree (DT), Random Forest (RF), and Gradient Boosting Machine (GBM) and are summarized as follows.

Let $\left(X_\tau\right)_{\tau\in\mathcal{T}}$ be the partition of X, where J is the number of distinct and non-overlapping regions. The DT estimator, given a set of variables x, is defined as: $\hat{\psi}^{DT}\left(x\right)=\Sigma_{\tau\in\mathcal{T}}\hat{\psi}_\tau 1_{\{x\in X_\tau\}}$, where $1_{\{x\in X_\tau\}}$, is the indicator function. The regions $\left(X_\tau\right)_{\tau\in\mathcal{T}}$ are found by minimizing the residual sum of squares.

RF aggregates many DTs, obtained by generating bootstrap training samples from the original dataset (Breiman, 2001). This algorithm's main characteristic is to select a random subset of predictors at each split, thus preventing the dominance of strong predictors in the splits of each tree (James et al., 2017). The RF estimator is calculated as

$$\hat{\psi}^{RF}\left(x\right)=\frac{1}{B}\sum_{b=1}^{B}\hat{\psi}^{(DT)}\left(x\,|\,b\right),$$

where B is the number of bootstrap samples and $\hat{\psi}^{(DT)}\left(x\,|\,b\right)$ is the DT estimator on sample b.

GBM considers a sequential approach in which each DT uses the information from the previous one to improve the current fit (Friedman, 2001). Given the current fit $\hat{\psi}\left(x_{i-1}\right)$, at each stage i (for i = 1, 2, ..., N), GBM algorithm provides a new model adding an estimator h to the fit: $\hat{\psi}\left(x_i\right)=\hat{\psi}\left(x_{i-1}\right)+\lambda_i h_i\left(x\right)$, where $h_i\in H$ is the family of differentiable functions and λ is the multiplier derived through the optimization problem.

According to Levantesi and Pizzorusso's (2019) framework, an ML estimator improves mortality forecasts obtained by implementing the Lee-Carter model. The ML estimator is modeled as follows:

$$\log\left(\hat{\psi}_x^{LC,ML}\right)=\alpha_a^{\psi}+\beta_a^{(1,\psi)}\kappa_t^{(1,\psi)},\qquad(5)$$

where the parameters $\alpha_a^{\psi},\beta_a^{(1,\psi)}$, and $\kappa_t^{(1,\psi)}$ correspond to the parameters $\alpha_a,\beta_a^{(1)}$, and $\kappa_t^{(1)}$ in the classical Lee-Carter model. From equations 4 and 5, the following model is obtained, which improved the Lee-Carter through ML:

$$\log\left(m_x^{LC,ML}\right)=\left(\alpha_a^{\psi}+\alpha_a\right)+\\+\beta_a^{(1,\psi)}\kappa_t^{(1,\psi)}+\beta_a^{(1)}\kappa_t^{(1)}.\qquad(6)$$

The authors underline that adopting the original mortality model for both fitting and forecast of ML estimator leads to an improvement in projection and makes it easy to understand the effect of such improvements on the model's parameters.

Among the CART methods, the Lee-Carter model might be enhanced through the following model suggested by Levantesi and Nigri (2019). The authors provide a novel view in which ML estimator's future values are consequences of the extrapolation of $\hat{\psi}_s$ previously smoothed with 2-dimensional P-splines (Eilers & Marx, 1996). As a result, forecasted values are strongly dependent on the smoothing process. Indeed extrapolation considers future observations as missing values estimated by the 2-dimensional P-splines (Currie et al., 2006).

The mean absolute percent error measures the model's accuracy,

$$MAPE=\frac{100}{N}\sum_x\left|\frac{m_x-\hat{m}_x}{m_x}\right|,$$

and the mean absolute error, $MAE=\sum_x\left|m_x-\hat{m}_x\right|/N$, where N is the data dimension, m_x and $\hat{m}_{\mathbf{x}}$ are respectively the observed and the estimated central death rates. While the models accuracy of forecasting is calculated through the root mean squared error,

$$RMSE=\sqrt{\sum_x\left(m_x-\hat{m}_x\right)^2/N},$$

which is used to compare the future mortality rates in an out-of-sample test.

Despite the widespread skepticism among demographers and actuaries, it is possible to record a steady increase of scientific contributions to mortality modeling through DL. Albeit the proposed literature shows prominent approaches, the applications on global panorama seem too uneven, jeopardizing the practitioner's interpretation and most likely pushing down the DL and NN attractiveness. In light of that, a clear recap and exposition of NN in mortality are required.

The term NN refers to a computational model inspired by the human brain, where multiple processing layers are implied to learn information from data represented by multiple levels of abstraction, solving even the most complex problems. Its shape is based on neurons, links between them, and learning algorithms (backpropagation). NN works as a weighted regression in which each unit suggests "weighted" information thanks to the activated links, and the activation function transforms the weighted sum of input signals into output. Albeit DL embodies a wide variety of networks differing from each other by the architectural structure, the basic scheme of NN is presented, the so-called feedforward structure.

Let n_j be the number of neurons in layer j and $\phi_{i,j}$ a generic activation function. The output of the i-th neuron in layer j, denoted by $y_{i,j}$, is computed as follows:

$$y_{i,j} = \phi_{i,j}\left(\sum_{k=1}^{n_{j-1}} \omega_{i,k}^{j} y_{k,j-1}\right). \tag{7}$$

The contributions on this topic are from Hainaut (2018), Richman and Wüthrich (2018), and Nigri et al. (2019, 2020). All of them apply NNs as universal function approximators, aiming to solve different tasks, thus deserving separate discussions. The framework used in Hainaut (2018) aims to replace singular value decomposition (SVD) into the Lee-Carter model. Specifically, the author proposes two stages of estimation: firstly, through the autoencoder, the mortality dataset is reduced in a small number of latent factors. Thus, during the second step, an econometric model is applied to forecast the latent variables. The neural analyzer is adjusted to different countries using as input the centered log-forces of mortality, denoted by:

$$X(t) := \begin{pmatrix} \ln \mu(t, x_{\min}) - \alpha_{x_{\min}} \\ \vdots \\ \ln \mu(t, x) - \alpha_x \\ \vdots \\ \ln \mu(t, x_{\max}) - \alpha_{x_{\max}} \end{pmatrix} t = t_{\min}, \ldots, t_{\max}, \tag{8}$$

where $X(t)$ is a vector of dimensions $n_x = x_{min} - x_{max}$ and α_x is the vector of average log-mortality rates according to the Lee-Carter formulation.

The basic idea is to determine two functions: an encoding and a decoding function denoted by $f^{enc}: R^{n_x} \to R^d$ and $f^{dec}: R^d \to R^{n_x}$. Concerning the classical Lee-Carter model, the factor $\beta_x \kappa_t$ is substituted with a nonlinear function, and log-mortality forces are calculated through the following relation:

$$\begin{aligned} &\ln \mu(t, x) = \alpha_x + f^{dec}\left(x, \kappa_t^{nn}\right) = \\ &\alpha_x + f^{dec}\left(x, f^{enc}\left(X(t)\right)\right). \end{aligned} \tag{9}$$

The encoding and decoding functions are calibrated to minimize the sum of squared residuals between observer and generated mortality curves:

$$\begin{aligned} &\left(f^{enc}, f^{dec}\right) = \\ &= \arg\min \sum_{t=t_{\min}}^{t_{\max}} \| X(t) - \hat{X}(t) \|_2^2 . \end{aligned} \tag{10}$$

This approach aims to overcome the parameter estimation related to SVD, which mainly concerns the lack of nonlinear components, which affects the mortality surface estimation. It is worth noting that the forecasting methods remain the same so that the extrapolation over time is related to canonical models with the classical limitation that are pointed out in Nigri et al. (2019), where a DL integration into Lee-Carter model is suggested. They underline the key role of κ_t parameter to depict the future nonlinear mortality behavior. More precisely, Nigri et al. (2019) propose alternative processes applying an RNN with an LSTM architecture to describe the evolution of κ_t over time. Indeed, taking into account both long and short-term dynamics, the LSTM can treat the time series noise, reproducing it into the forecasted trend. The authors propose an LSTM model that approximates the function f linking κ_t to its delayed values:

$$LSTM: \quad \kappa_t = f\left(\kappa_{t-1}, \kappa_{t-2}, \ldots, \kappa_{t-1}\right) + \varepsilon_t , \tag{11}$$

where ε is a homoscedastic error term. To implement the LSTM algorithm, the dataset is split into the training test, where supervised learning is instructed, and the testing set.

During the training, the network learns the input-output relationship; the function that describes such a link can predict the interested variables, in

http://dx.doi.org/10.21511/ins.11(1).2020.02

this case, κ_t, starting only from the input. In the dataset, the input is a $(n \times J)$ matrix of the time lags of κ_t, and the output is the $(n \times 1)$ vector of its current values, where $n \in N$ is the number of instances. The forecasted values of κ_t are derived recursively and are calculated using only the predicted values $\hat{\kappa}_t$ and not those observed. Despite the remarkable predictive ability of NNs to provide high precision point forecasting, the lack of interval estimation remains the main drawback. In this regard, Marino and Levantesi (2020) extend the approach proposed in Nigri et al. (2019), employing the bootstrap technique to perform a point forecasting of the Lee-Carter time-index through LSTM and produce a confidence interval for the LSTM's parameter uncertainty. It needs to consider that the approaches pursued so far work into a stable framework, aiming to integrate the Lee-Carter model. The pure DL based-approach proposed by Richman and Wüthrich (2018) is introduced, extending the Lee-Carter model to multiple populations in a NN framework. The authors define the NN model in two steps: firstly, they model all mortality rates, considering a feature space that is the year of death, age of last birthday before death, region, and gender. Thereby, though the embedding layer, they model region and gender as categorical features; then, it maps each category in a low dimensional vector, whose parameters are learned during the fitting.

3. RESULTS

Deprez et al. (2017) implement the decision trees algorithm to verify the goodness of mortality estimates obtained by implementing the Lee-Carter model and Renshaw-Haberman model. They backtest a mortality model to highlight the model's ability to explain mortality for each age, year of birth, and gender. The case study is based on the mortality data of Switzerland and the following set of variables $\mathcal{A} = \{0,\ldots,97\}, \mathcal{T} = \{1876,\ldots,2014\}$ for both genders. The results show improvements in the accuracy of the Lee-Carter and Renshaw-Haberman models' mortality fitting because there are factors that are not well caught by these models.

Levantesi and Pizzorusso (2019) enhance the Lee-Carter, Renshaw-Haberman, and Plat mortality projections through DT, RF, and GBM. These algorithms work on the ratio between observed and estimated deaths from a given model. The results show an increase in the accuracy of the projections of all three models. The case study is developed on the Italian mortality dataset with variables $\mathcal{A} = \{0,\ldots,100\}, \mathcal{T} = \{1915,\ldots,2014\}$, and $\mathcal{C} = \{1815,\ldots,2014\}$ for both the genders. They obtain better results for RF that is more effective than DT and GBM. Both fit and forecasting appear improved.

Following the same research line, Levantesi and Nigri (2019) propose to extrapolate the ML estimator on the whole mortality surface using the two-dimensional P-splines and obtain more accurate mortality projections. Moreover, the authors develop a sensitivity analysis of the model to predictors, aiming to investigate if the results are reasonable in a demographic perspective, and a sensitivity analysis on the age interval to analyze if there are some improvements on a reduced dataset. The numerical application is based on the mortality dataset of Australia, France, Italy, Spain, the UK, and the USA with variables $\mathcal{A} = \{20,\ldots,100\}, \mathcal{T} = \{1947,\ldots,2014\}$, and $\mathcal{C} = \{1847,\ldots,1994\}$ of both genders. The authors obtained significant improvements in the mortality projection of the Lee-Carter model by applying the RF algorithm. These results hold for all the analyzed countries.

The innovation through DL models is driven by constant mortality improvement, addressing the need to better understand future mortality dynamics changes. In this scenario, the canonical Lee-Carter model and its evolutions remain the gold standard for comparing future models' performance. Hainaut, 2018 using the data for French, UK, and US mortality rates; the training dataset comprises the years 1946–2000 and the test set by the years 2001–2014. The results show that the NN approach has strongly over perform the Lee-Carter model with and without cohort effects. Since this method acts only as an estimation procedure, the authors state that the empirical evidence supports that a simple random walk might be a plausible method to forecast the latent factors. On the other hand, Nigri et al. (2019) point out the relevant implications of exploiting an extension of the Lee-Carter model, based on an RNN with LSTM architecture to forecast the latent factor κ_t.

The authors investigate six countries, showing better accuracy levels of the mortality trend forecast than the classical approach. Finally, Richman and Wüthrich (2018), pursuing a pure NN approach, project multi-population mortality rates exploiting the main advantages of the multi-population approach, i.e., the robustness compared to single-population models, due to common factors such as population health, economics, and technology. The authors use all countries in the Human Mortality Database, considering data from 1950 to 1999 as the training set and data from 1991 to 1999 as the testing set. Therefore, they assess the forecasting performance on the validation set from 2000 onwards.

4. DISCUSSION

The diffusion of ML techniques in the insurance sector to deal with large available datasets and extract important information from them is increasing competitiveness, bringing to more effective risk management and efficient pricing policies. This work has retraced the advances in recent literature on the application of ML techniques to longevity risk. While it is now a widespread practice in many areas of the insurance business to use modern predictive techniques, in practice, the management of longevity risk is still entrusted to traditional approaches. In this context, the first advantage AI could generate is reducing the information asymmetry between insurer and policyholder: a better understanding and quantification of each policyholder's specific risk can be reached by simultaneously exploiting the information present in national and international datasets and those relating to individual portfolios. This is useful in pricing life products with long-term duration and lifelong guaranteed options embedded in pension contracts. Another AI opportunity is the improvement of longevity risk quantification for health insurance and long term care products thanks to the exploitation of patterns in available medical and socio-economic data since traditional risk models require a very long time to process. AI favors the advances of the microeconomic and structural models, and the risk management is shifting from statistical methods to supervised tree algorithms that automatically select variables of interest and better identify the data connections. Available data could be integrated by those offered by smart sensor technologies for policyholders' health monitoring.

CONCLUSION

Detailed considerations on the regulatory and ethical aspects deriving from the introduction of ML techniques in the insurance sector are beyond the scope of this work, which instead aims to stimulate reflection on the advantages that these techniques implemented in practice could bring to the quantification of longevity risk. However, besides the great potential of ML, there are also some drawbacks. First of all, one issue concerns the forecasts' bias, which can be significant for mortality datasets. As it is well known in data science, the bias may result from how the data was collected: if the algorithm is trained on a non-representative data set of the real distribution of the population, the forecasts could be biased. This can happen if the algorithms used to estimate a specific insurer's portfolio's longevity risk are calibrated on national data. Simultaneously, the problem is partially solved if internal databases built on the historical experience of the individual company are available. On closer inspection, it falls back on the classic question of the possibility of incurring the so-called basic risk, of which not even the traditional models are free. Another widely known limit of algorithms trained on big data is the possibility of a correlation between variables in this large dataset. On the other hand, through ML, the output from internal risk models can be validated and improved continuously. One thing is certain: the introduction of AI requires companies to change the business culture. The crucial point is to develop processes and tools to give stakeholders the possibility to understand how risks can be identified and managed within limits set by the firm's risk culture. With this awareness, it is important to assess whether the results of the application of ML are in line with the results produced by classical models and to find the explanation for any variance.

http://dx.doi.org/10.21511/ins.11(1).2020.02

In light of the above, it can be argued that ML can offer insurance companies and pension fund managers new tools and methods supporting actuaries in classifying longevity risks, offering accurate predictive pricing models, and reducing losses. In addition to the areas in which ML methodologies are already used, the transition from the traditional modeling of longevity risk to an innovative one could represent a challenge for the life insurance sector.

AUTHOR CONTRIBUTIONS

Conceptualization: Susanna Levantesi, Gabriella Piscopo.
Investigation: Andrea Nigri.
Methodology: Susanna Levantesi, Andrea Nigri.
Supervision: Susanna Levantesi, Gabriella Piscopo.
Visualization: Andrea Nigri.
Writing – original draft: Gabriella Piscopo, Andrea Nigri.
Writing – review & editing: Susanna Levantesi, Gabriella Piscopo.

REFERENCES

1. Alpaydin, E. (2010). *Introduction to Machine Learning* (2nd ed.). Cambridge: Massachusetts Institute of Technology Press. Retrieved from https://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf
2. Biffis, E. (2004). Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics, 37*(3), 443-468. https://doi.org/10.1016/j.insmatheco.2005.05.003
3. Biffis, E., Blake, D., Pitotti, L., & Sun, A. (2014). The cost of counterparty risk and collateralization in longevity swaps. *The Journal of Risk and Insurance, 83*(2), 1-33. https://doi.org/10.1111/jori.12055
4. Blake, D., Cairns, A. J. G., Dowd, K., & MacMinn, R. (2006). Longevity bonds: Financial engineering, valuation and hedging. *Journal of Risk and Insurance, 73*(4), 647-672. https://doi.org/10.1111/j.1539-6975.2006.00193.x
5. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324
6. Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics, 31*(3), 373-393. https://doi.org/10.1016/S0167-6687(02)00185-3
7. Cairns, A. J. G., Blake, D., & Dowd, K. (2006). Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk. *ASTIN Bulletin, 36*(1), 79-120. Retrieved from https://ideas.repec.org/a/cup/astinb/v36y2006i01p79-120_01.html
8. Carracedo, P., Debon, A., Iftimi, A., & Montes, F. (2018). Detecting spatio-temporal mortality clusters of European countries by sex and age. *International Journal for Equity in Health, 17,* 38. https://doi.org/10.1186/s12939-018-0750-z
9. Cocco, J. F., & Gomes, F. J. (2012). Longevity risk, retirement savings, and financial innovation. *Journal of Financial Economics, 103*(3), 507-529. https://doi.org/10.1016/j.jfineco.2011.10.002
10. CRO Forum. (2010). *Longevity* (CRO Briefing Emerging Risks Initiative Position Paper). Retrieved from https://www.thecroforum.org/wp-content/uploads/2012/10/Longevity-Risk-1.pdf
11. Currie, I. D., Durban, M., & Eilers ,P. H. C. (2004). Smoothing and Forecasting Mortality Rates. *Statistical Modelling, 4*(4), 279-298. https://doi.org/10.1191%2F1471082X04st080oa
12. Currie I. D., Durban M., & Eilers P. H. C. (2006). Generalized Linear Array Models with Applications to Multidimensional Smoothing. *Journal of the Royal Statistical Society, 68*(2), 259-280. https://doi.org/10.1111/j.1467-9868.2006.00543.x
13. Dahl, M. (2004). Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts. *Insurance: Mathematics and Economics, 35*(1), 113-136. https://doi.org/10.1016/j.insmatheco.2004.05.003
14. Deloitte. (2018). *AI and you. Perceptions of Artificial Intelligence from the EMEA financial services industry.* Retrieved from https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/technology/deloitte-cn-tech-ai-and-you-en-170801.pdf
15. Deprez, P., Shevchenko, P. V., & Wüthrich, M., (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal, 7*(2), 337-352. https://doi.org/10.1007/s13385-017-0152-4
16. Eilers, P. H. C., & Marx, B. D. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science, 11*(2), 89-102. Retrieved from https://projecteuclid.org/euclid.ss/1038425655
17. Ferrario, A., Noll, A., & Wüthrich, M. V. (2020). *Insights from Inside Neural Networks.* http://dx.doi.org/10.2139/ssrn.3226852

18. Friedman, J. H. (2001). Greedy function approximation: A Gradient Boosting Machine. *The Annals of Statistics, 29*(5), 1189-1232. Retrieved from https://projecteuclid.org/euclid.aos/1013203451

19. Gabrielli, A., Richman, R., & Wüthrich, M. (2019). Neural Network Embedding of the Over-Dispersed Poisson Reserving Model. *Scandinavian Actuarial Journal, 1,* 1-29. https://doi.org/10.1080/03461238.2019.1633394

20. Haberman, S., & Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics, 48*(1), 35-55. https://doi.org/10.1016/j.insmatheco.2010.09.003

21. Hainaut, D. (2018). A neural-network analyzer for mortality forecast. *ASTIN Bulletin, 48*(2), 481-508. https://doi.org/10.1017/asb.2017.45

22. Hastie, T., Tibshirani, R., & Friedman, J. (2016). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer. Retrieved from https://web.stanford.edu/~hastie/Papers/ESLII.pdf

23. Huang, F., & Browne, B. (2017). Mortality forecasting using a modified Continuous Mortality Investigation Mortality Projections Model for China I: Methodology and country-level results. *Annals of Actuarial Science, 11*(1), 20-45. https://doi.org/10.1017/S1748499516000142

24. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning: With Applications in R.* New York: Springer. Retrieved from https://www.academia.edu/36691506/An_Introduction_to_Statistical_Learning_Springer_Texts_in_Statistics_An_Introduction_to_Statistical_Learning

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

26. Levantesi, S., & Nigri, A. (2019). A random forest algorithm to improve the Lee-Carter mortality forecasting: impact on q-forward. *Soft Computing, 12*. https://doi.org/10.1007/s00500-019-04427-z

27. Levantesi, S., & Pizzorusso, V. (2019). Application of Machine Learning to Mortality Modeling and Forecasting. *Risks, 7*(1), 26. https://doi.org/10.3390/risks7010026

28. Loeys, J., Panigirtzoglou, N., & Ribeiro, R. (2007). Longevity: A Market in the Making. In *Global Market Strategy*. London: J. P. Morgan Securities Ltd. Retrieved from https://nanopdf.com/download/longevity-june-27-07pmd_pdf

29. Maier, M., Carlotto, H., Sanchez, F., Balogun, S., & Merritt, S. (2019). Transforming Underwriting in the Life Insurance Industry. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(01), 9373-9380. https://doi.org/10.1609/aaai.v33i01.33019373

30. Marino, M., & Levantesi, S. (2020). Measuring longevity risk through a Neural Network Lee-Carter Model. *SSRN*. https://doi.org/10.2139/ssrn.3599821

31. Milevsky, M. A., & Promislow, S. D. (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics, 29*(3), 299-318. https://doi.org/10.1016/S0167-6687(01)00093-2

32. Nigri, A., Levantesi, S., & Marino, M., (2020). Forecasting life expectancy and lifespan disparity: a long short-term memory approach. *Scandinavian Actuarial Journal*. https://doi.org/10.1080/03461238.2020.1814855

33. 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

34. Noll, A., Salzmann, R., & Wüthrich, M. V. (2020). Case Study: French Motor Third-Party Liability Claims. *SSRN*. http://dx.doi.org/10.2139/ssrn.3164764

35. Piscopo, G. & Resta, M. (2017). Applying spectral biclustering to mortality data. *Risks, 5*(2), 24. https://doi.org/10.3390/risks5020024

36. Pitacco, E. (2020). Enterprise Risk Management (ERM) and Quantitative Risk Management (QRM). In *ERM and QRM in Life Insurance. Springer Actuarial*. Springer, Cham. https://doi.org/10.1007/978-3-030-49852-8_2

37. Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics, 45*, 393-404. https://doi.org/10.1016/j.insmatheco.2009.08.006

38. Renshaw, A. E., & Haberman, S. (2003). On the Forecasting of Mortality Reduction Factors. *Insurance Mathematics and Economics, 32*(3), 379-401. https://doi.org/10.1016/S0167-6687(03)00118-5

39. Renshaw, A. E., & Haberman, S. (2006). A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance: Mathematics and Economics, 38,* 556-570. Retrieved from https://bendixcarstensen.com/APC/Renshaw.2006.pdf

40. Richman, R., & Wüthrich, M. (2018). A Neural Network Extension of the Lee-Carter Model to Multiple Populations. *SSRN*. https://doi.org/10.2139/ssrn.3270877

41. Schrager, D. F. (2006). Affine stochastic mortality. *Insurance: Mathematics and Economics, 38*(1), 81-97. Retrieved from http://pages.stern.nyu.edu/~dbackus/BCH/demography/Schrager_affinemortality_IME_06.pdf

42. The Life & Longevity Markets Association. (2010). *Technical note: the q-forward.*

43. Villegas, A. M., Kaishev, V. K., & Millossovich, P. (2015). *StMoMo: An R package for stochastic mortality modelling.* Retrieved from https://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf

44. Wüthrich, M. V., & Christoph, B. (2019). *Data Analytics for Non-Life Insurance Pricing* (Swiss Finance Institute Research Paper No. 16-68). http://dx.doi.org/10.2139/ssrn.2870308

http://dx.doi.org/10.21511/ins.11(1).2020.02