Forecasting All-Cause Mortality: Leveraging Causes-of-Death Data through Neural Networks

Manny Adachi
Master of Health Data Science candidate

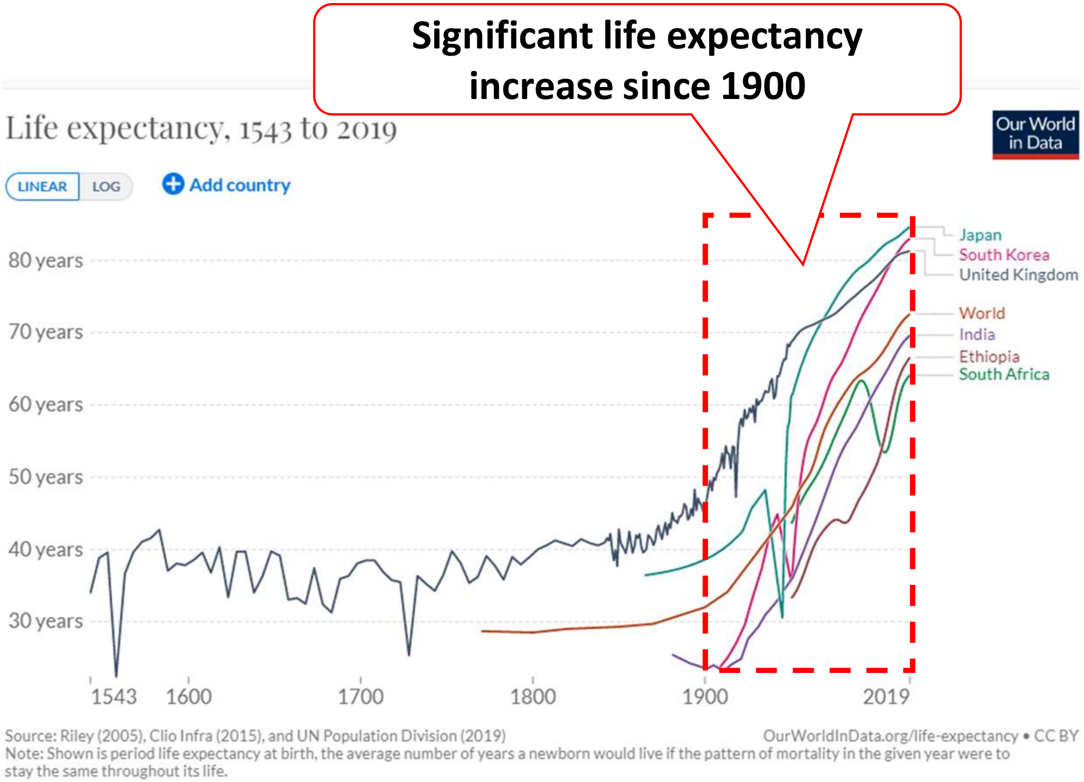
Supervised by: Dr. Andrés Villegas, Dr. Katja Hanewald
School of Risk & Actuarial Studies

Content

- Background
- Overview of past studies
- Project aim
- Methodology
- About data
- Result
- Discussion
- Conclusion

Background

Mortality forecasting – important for decision making in both public and private sectors



Example responses by sectors

Public

- Revisions of the social security plan and policy

Private

- Revisions of insurance product prices and corresponding reserves

Source of image: [Life Expectancy - Our World in Data]

Overview of past studies

Research gaps in all components of forecast models, but “cause-of-death (CoD) + neural network” is likely the prominent gap.

Forecast model components		
Base stochastic models	**Involved feature variables**	**Neural network type**
■ **No single model is better** than others. (Tabeau et al., 1999; Cairns et al., 2009)	■ Age, year (time), and gender are mandatory.	■ The models with especially the **convolutional neural networks (CNN)** have outperformed the conventional models (Perla et al., 2021, and Wang et al., 2021).
■ Various models have been proposed, but **Lee-Carter (LC) model** and its extensions are most frequently referred.	■ **CoD** has sometimes been used, but had **some critical drawbacks** in old studies (e.g., low quality/ limited availability of data) (Lyu et al., 2020).	■ However, **CoD has apparently not been involved** when modelling with the neural networks.

Project aim

Can a CNN-based (LC) model using cause-of-death as a feature variable outperform the non-CNN-based (LC) models?

Reasons of the aim	Compared models
Gap in previous studies ■ As mentioned in previous page **Feasibility** ■ The neural networks can <u>**detect automatically the complex correlations among features**</u>: more efficient and reliable feature engineering ■ The base model will be a LC-type model due to its simplicity.	**New model** ■ (CoD based) LC-type model with CNN **Benchmark models** ■ See the next page

Methodology – Overview

We have compared the new model (LC-CNN using CoD as a feature) to the other variant models to assess the effect by each model component.

Model component	New model	Variant 1	Variant 2	Variant 3	Variant 4 (for female)	Variant 4 (for male)	Variant 5
Base model	LC-type	LC-type	LC-type	LC-type	LC-type	LC-type	LC-type
Feature reduction	Yes (embedding)	Yes (SVD)	Yes (SVD)	Yes (embedding)	No	No	No
CNN usage	Yes	No	No	Yes	Yes	Yes	Yes
CoD usage	Yes	No	Yes	No	No	No	No

Methodology – Lee-Carter (LC) model

Lee-Carter (LC) model

- It is a model used for **forecasting the mortality rate**, firstly introduced by Professor Ronald Lee and Professor Lawrence Carter in 1992.

- It has been **widely used** as a base model in many actuarial and demographic applications due to its **simplicity**, **interpretability**, and relatively **promising performance**.

- Mathematical expression of the LC model is,

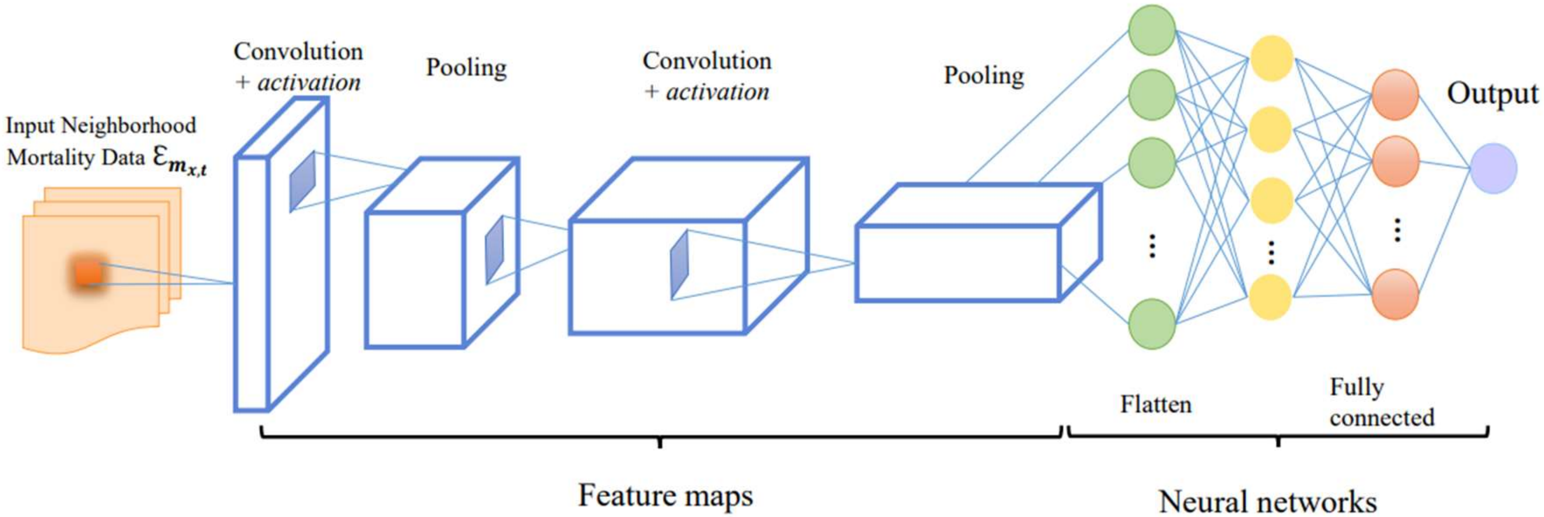
$$\log(\boldsymbol{m}_{x,t}) = \boldsymbol{\alpha}_x + \boldsymbol{\beta}_x \boldsymbol{\kappa}_t + \boldsymbol{\epsilon}_{x,t}$$

  - $\boldsymbol{m}_{x,t}$ is the mortality rate at age $\boldsymbol{x}$ in year $\boldsymbol{t}$.
  - $\boldsymbol{\alpha}_x$ is the average log mortality rate specific to age $\boldsymbol{x}$.
  - $\boldsymbol{\beta}_x$ is the rate of the log mortality change over time at age $\boldsymbol{x}$.
  - $\boldsymbol{\kappa}_t$ is the time index (year-to-year change) of the mortality in year $\boldsymbol{t}$.
  - $\boldsymbol{\epsilon}_{x,t}$ is the error.

Methodology – What are CNNs?

CNN (convolutional neural network)

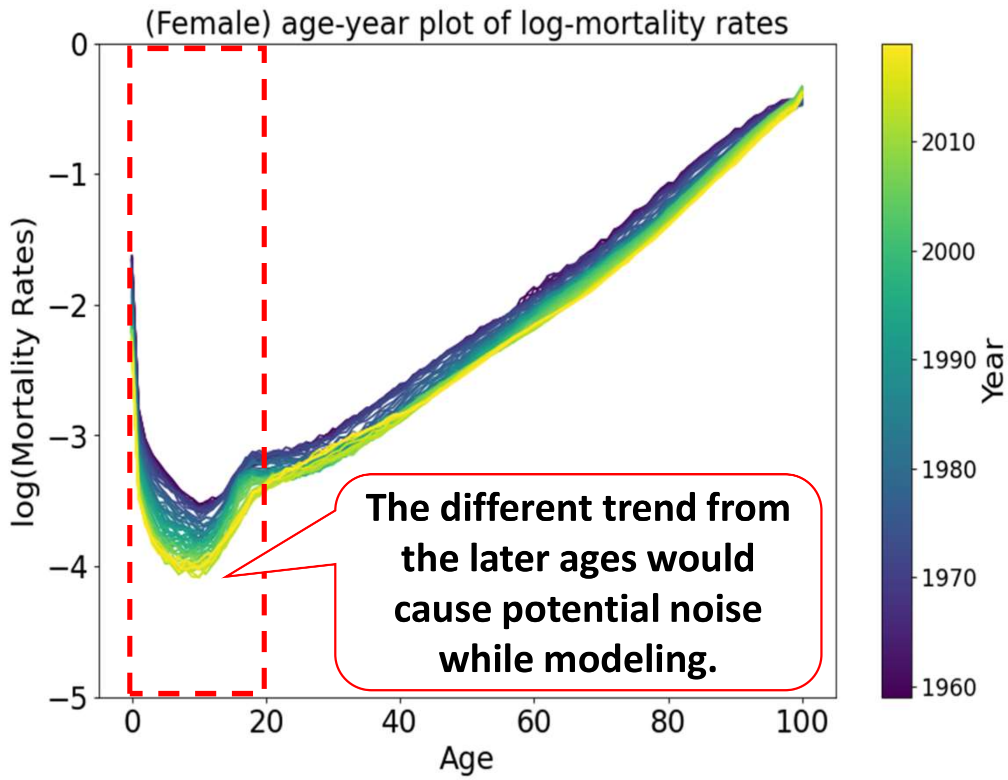
- CNNs are a kind of neural network models, but usually used for image recognition, pattern recognition, and/or computer vision.
- CNNs are often faster and easier to train than the recurrent neural networks (RNN), which are usually used for the time series data.

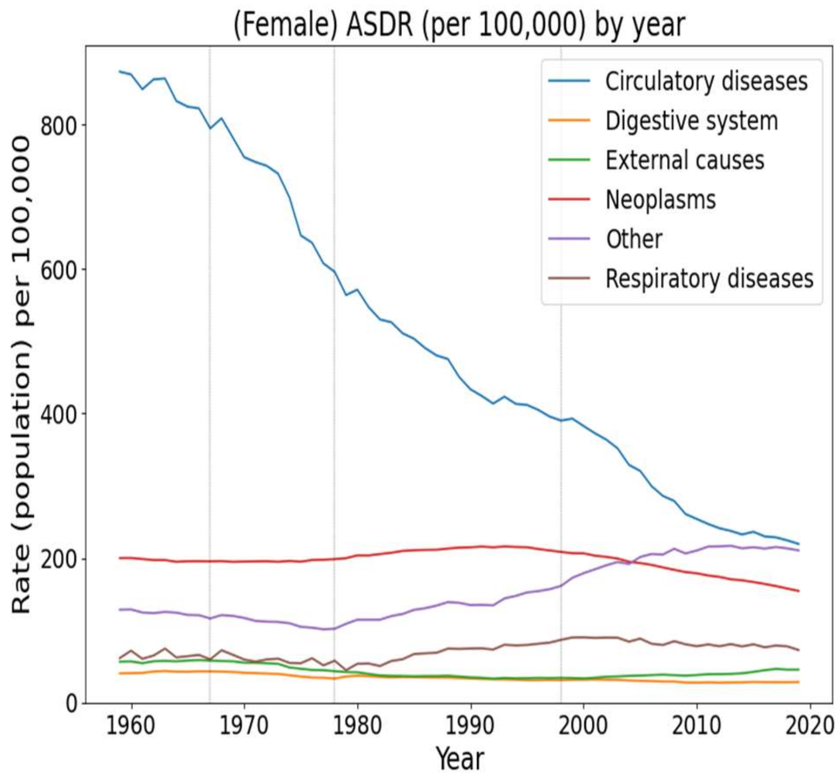


Source: C. Wang et al. (2021), [Neural Networks Part 8: Image Classification with Convolutional Neural Networks - YouTube]

About data

U.S. mortality data of the age 20 and above with cause-of-death and gender breakdown from 1959 to 2019





Source of data: Villegas et al. (2021), National Bureau of Economic Research (n.d.), and Shkolnikov et al. (2021)

Result

The new model partially outperformed the conventional LC models by MSE. The computing time is much longer in the neural network models.

	Model	Computing time (in minute)		MSE total (in 10^{-3}) (Forecast vs. actual)	
		Training	Forecast	Female	Male
CNN	**New model**	215.00	85.83×10^{-3}	53.98	113.90
	Variant 1	2.98×10^{-3}	9.73×10^{-3}	36.18	34.21
	Variant 2	14.18×10^{-3}	35.83×10^{-3}	160.03	133.17
CNN	**Variant 3**	79.12	14.98×10^{-3}	33.78	167.38
CNN	**Variant 4 (for female)**	51.98	18.00×10^{-3}	67.06	Not applicable
CNN	**Variant 4 (for male)**	51.92	18.00×10^{-3}	Not applicable	147.97
CNN	**Variant 5**	52.83	35.55×10^{-3}	527.66	85.62

Discussion

Points

- The new model could **partially outperform the conventional models** (Variant 2 not 1).

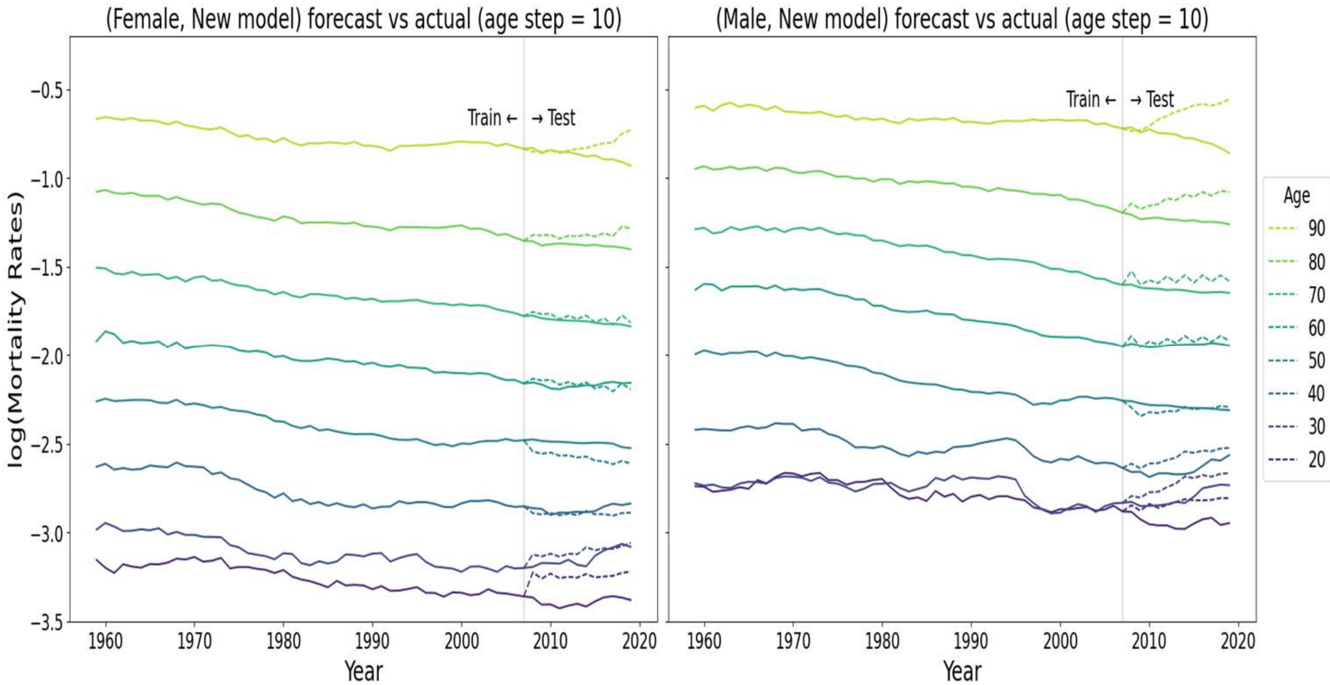
  More investigation with robust combinations of different **hyper-parameters** would be beneficial.

  We could not achieve it due to computational and time cost within the time limitation.

  (e.g., the smaller learning rate while training sometimes improve the model performance, although it often requires more epochs causing **significantly longer training time**.)

- Model development for every population category separately is **reasonable only when each category has absolutely no correlation** to another.

  However, such case is often **unrealistic**.



Conclusion

There is a potential for the model using CNN and CoD completely outperforming the conventional LC models in accuracy. More research is needed.

What we have done (i.e., contribution)

New attempt	Knowing the limitation and possibility
■ Usage of both cause-of-death (CoD) specific data and (convolutional) neural network to forecast all-cause mortality rate	■ Performance of the new model would be enhanced by further investigation. ■ However, always be aware of the computation and time cost

What can be done next, for example?

Hyper-parameters	Data	Feature importance
■ Robust exploration of the optimal hyper-parameter combinations	■ Data of other countries' population for model training and performance assessment	■ It would enhance the understandable explanation of the model to the audience.

Reference

Bertolaccini, L., Solli, P., Pardolesi, A., & Pasini, A. (2017). An overview of the use of artificial neural networks in lung cancer research. *Journal of thoracic disease*, *9*(4), 924–931. https://doi.org/10.21037/jtd.2017.03.157

Brownlee, J. (2018, November 12). *How to Develop Convolutional Neural Network Models for Time Series Forecasting.* Machine Learning Mastery. https://machinelearningmastery.com/how-to-develop-convolutional-neural-network-models-for-time-series-forecasting/

Brunton, S. (2020, Janurary 20). *Singular Value Decomposition (SVD): Overview* [Video]. YouTube. https://www.youtube.com/watch?v=gXbThCXjZFM

Brunton, S. (2020, Janurary 20). *Singular Value Decomposition (SVD): Mathematical Overview* [Video]. YouTube. https://www.youtube.com/watch?v=nbBvuuNVfco

Brunton, S. (2020, Janurary 28). *Principal Component Analysis (PCA)* [Video]. YouTube. https://www.youtube.com/watch?v=fkf4IBRSeEc

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science, 1(2)*, 3–43.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I., 2009. A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States. *North American Actuarial Journal*, 13(1), pp.1-35.

Center for Disease and Control Prevention. (n.d.). *Public-Use Data Files*. https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple

Koehrsen, W. (2018, October 2). *Neural Network Embeddings Explained*. Towards Data Science. https://towardsdatascience.com/neural-network-embeddings-explained-4d028e6f0526

Koissi, M., Shapiro, A.F., & Högnäs, G. (2004). Fitting and Forecasting Mortality Rates for Nordic Countries Using the Lee-Carter method.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659-671.

Lyu, P., de Waegenaere, A., & Melenberg, B. (2020). A Multi-population Approach to Forecasting All-Cause Mortality Using Cause-of-Death Mortality Data. *North American Actuarial Journal*, 25(sup1), S421–S456.

National Bureau of Economic Research. (n.d.). *Mortality Data - Vital Statistics NCHS Multiple Cause of Death Data*. Retrieved April 1, 2022, from https://www.nber.org/research/data/mortality-data-vital-statistics-nchs-multiple-cause-death-data

Natrajan, S., Sathish, G., & Jeyakumar, B. (2018), Data Wrangling and Data Leakage in Machine Learning for Healthcare. *JETIR- International Journal of Emerging Technologies and Innovative Research*, ISSN:2349-5162, Vol.5, Issue 8, page no.553-557.

Reference (continued)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. & others (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825--2830.

Perla, F., Richman, R., Scognamiglio, S., & Wüthrich, M. V. (2021). Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, 2021(7), 572–598.

Richman, R. & Wüthrich, M. V. (2021). A neural network extension of the Lee-Carter model to multiple populations. *Annals of Actuarial Science, 15(2)*, 346–366.

Schweizerische Aktuarvereinigung. (n.d.). *Actuarial Data Science.* https://www.actuarialdatascience.org/ADS-Tutorials/

Shkolnikov, V., Barbieri, M., & Wilmoth J. (2021, March 17). *The United States of America, Exposure to risk (period 1x1)*. The Human Mortality Database. https://www.mortality.org/hmd/USA/STATS/Exposures_1x1.txt

Tabeau, E., Ekamper, P., Huisman, C., & Bosch, A. (1999). Improving overall mortality forecasts by analysing cause-of-death, period and cohort effects in trends. *European Journal of Population, 15(2)*, 153–183.

Tam, A. (2021, November 20). *Using CNN for financial time series prediction*. Machine Learning Mastery. https://machinelearningmastery.com/using-cnn-for-financial-time-series-prediction/

Villegas, A.M., Bajekal, M., Haberman, S., & Zhou, L. (2021, September). Analysis of Historical U.S. Population Mortality Improvement Drivers 1959–2016. Society of Actuaries Research Institute. https://www.soa.org/globalassets/assets/files/resources/research-report/2021/2021-historical-us-population-mortality-improvement-drivers.pdf

Wang, C. W., Zhang, J., & Zhu, W. (2021). Neighbouring prediction for mortality. *ASTIN Bulletin, 51(3)*, 689–718.

Wang, Q., Hanewald, K., & Wang, X. (2021). Multistate health transition modeling using neural networks. *Journal of Risk and Insurance, 89(2)*, 475–504.

Wilmoth, J. R. (1995). Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies, 5(4)*, 293–319.

World Health Organization. (2021, October 4). *Ageing and health*. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

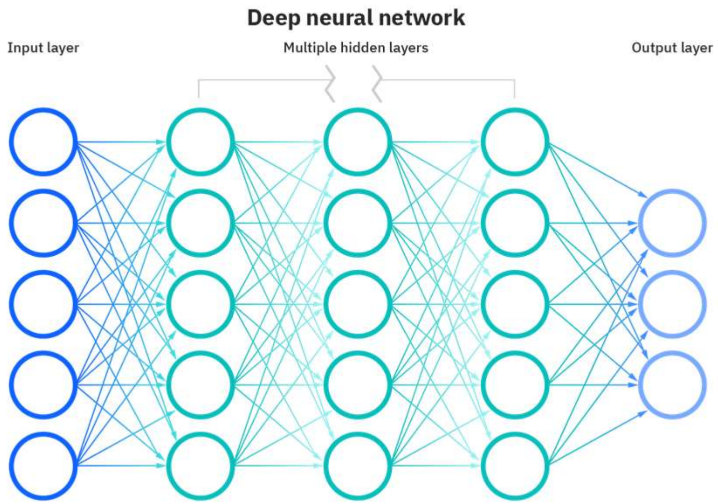
Appendix

Methodology – what are neural networks?

Basics of neural networks (NN)

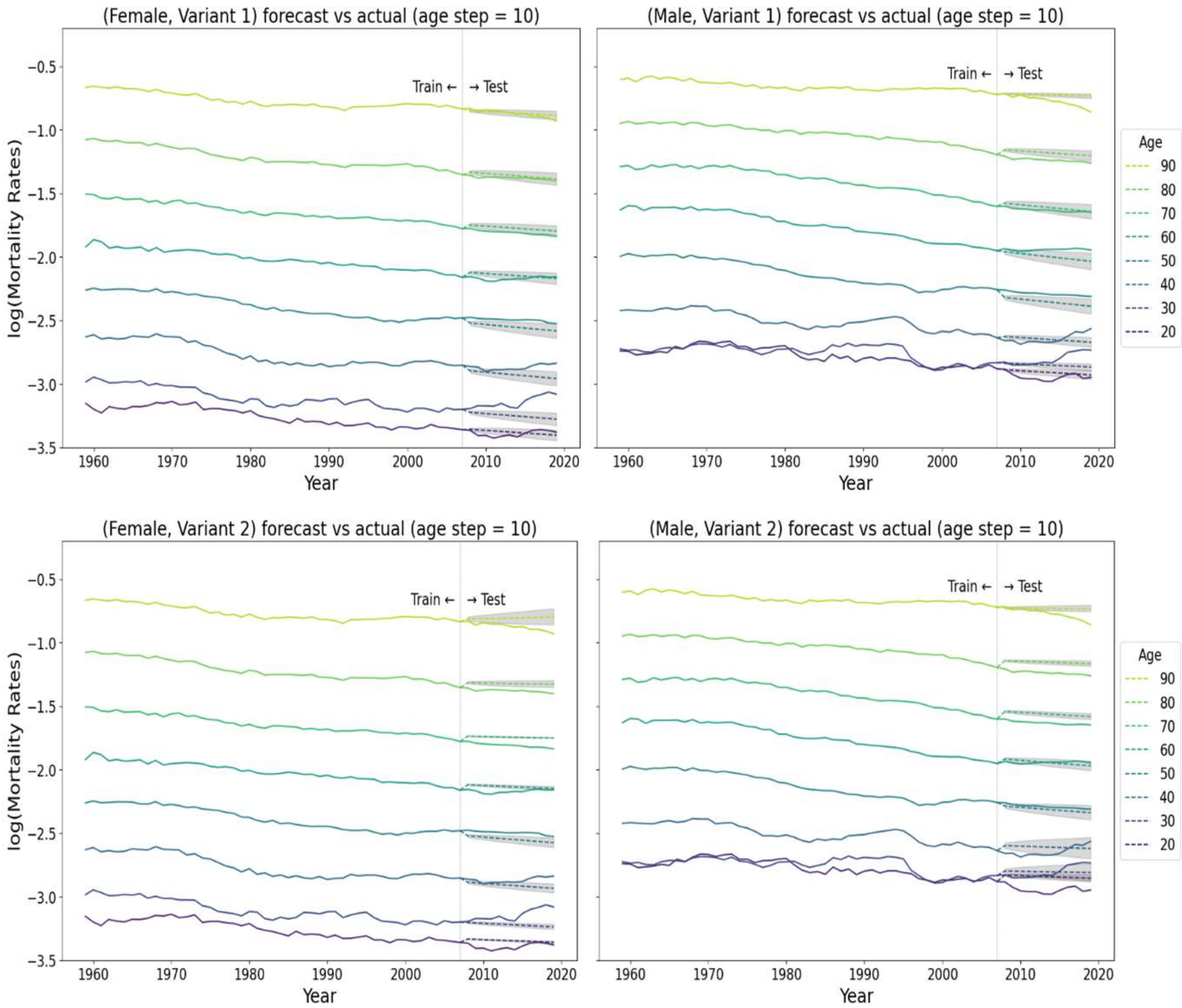
- They are **a subset of machine learning**. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

- **FCN** (feed-forward fully connected neural networks) is the **most basic form of NN**. Each layer in FCN is connected to every part of the previous layer.

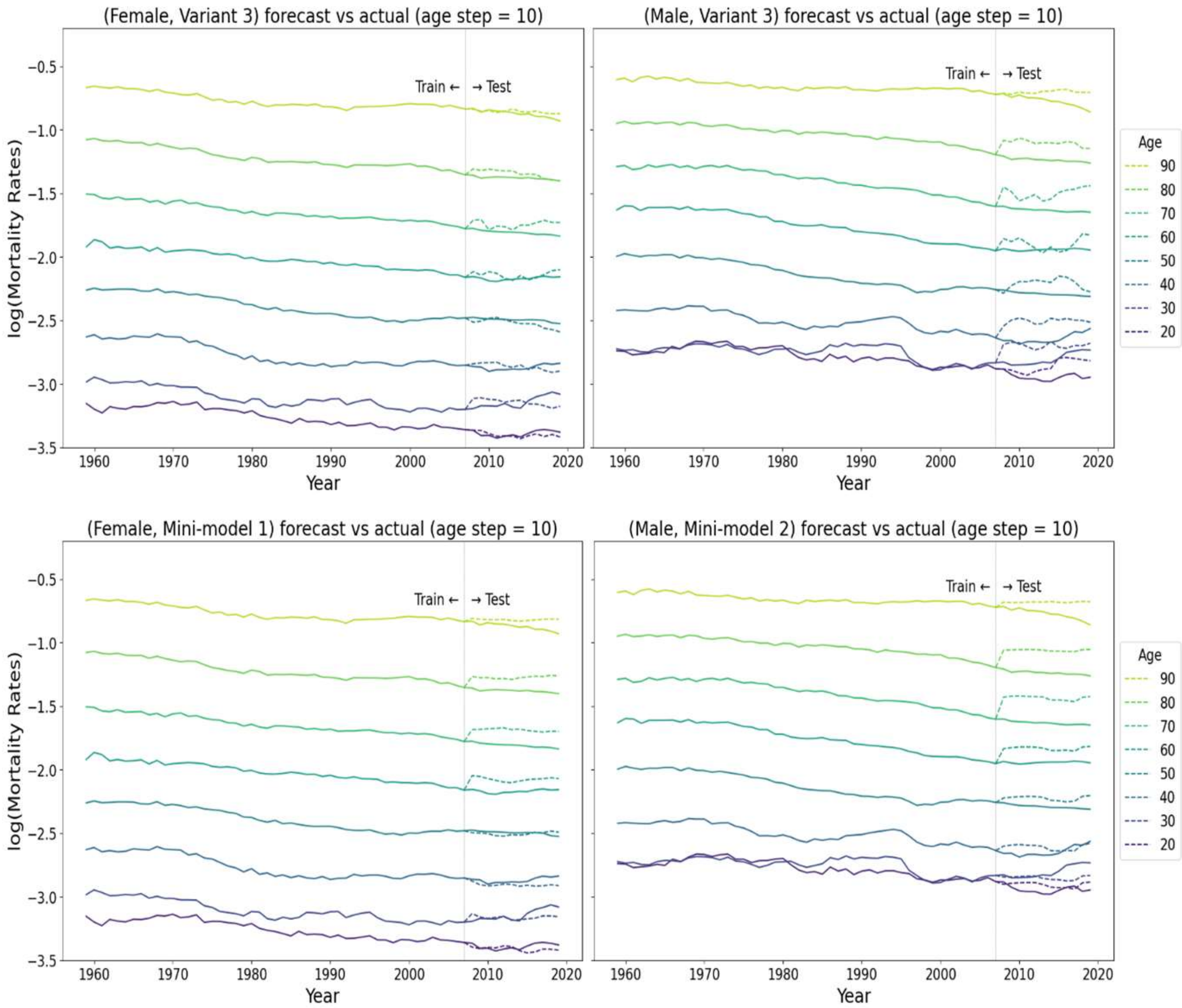


Source: Perla et al. (2021), What are Neural Networks? | IBM

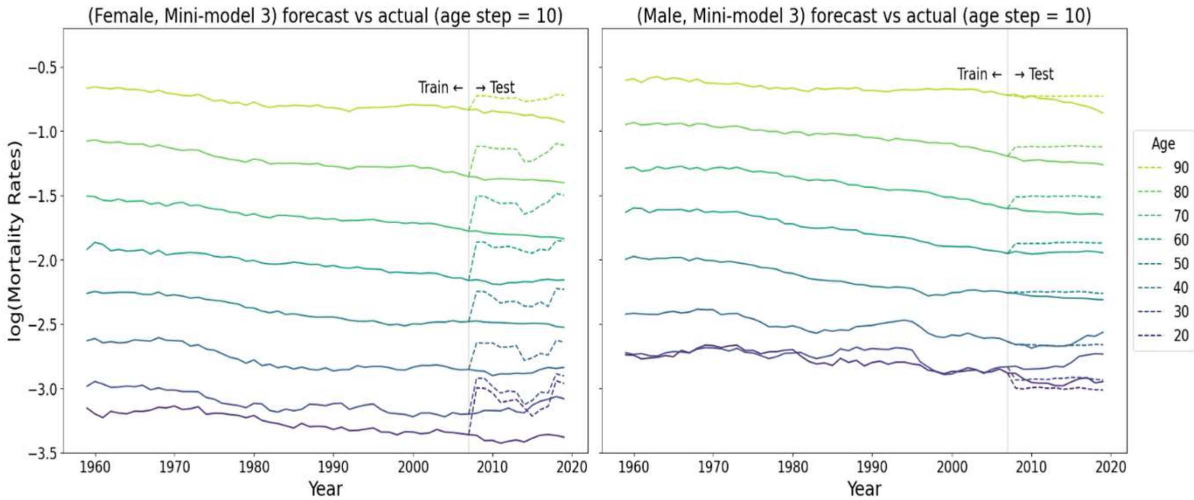
Figures – Forecast vs. actual (Variant 1 and 2)



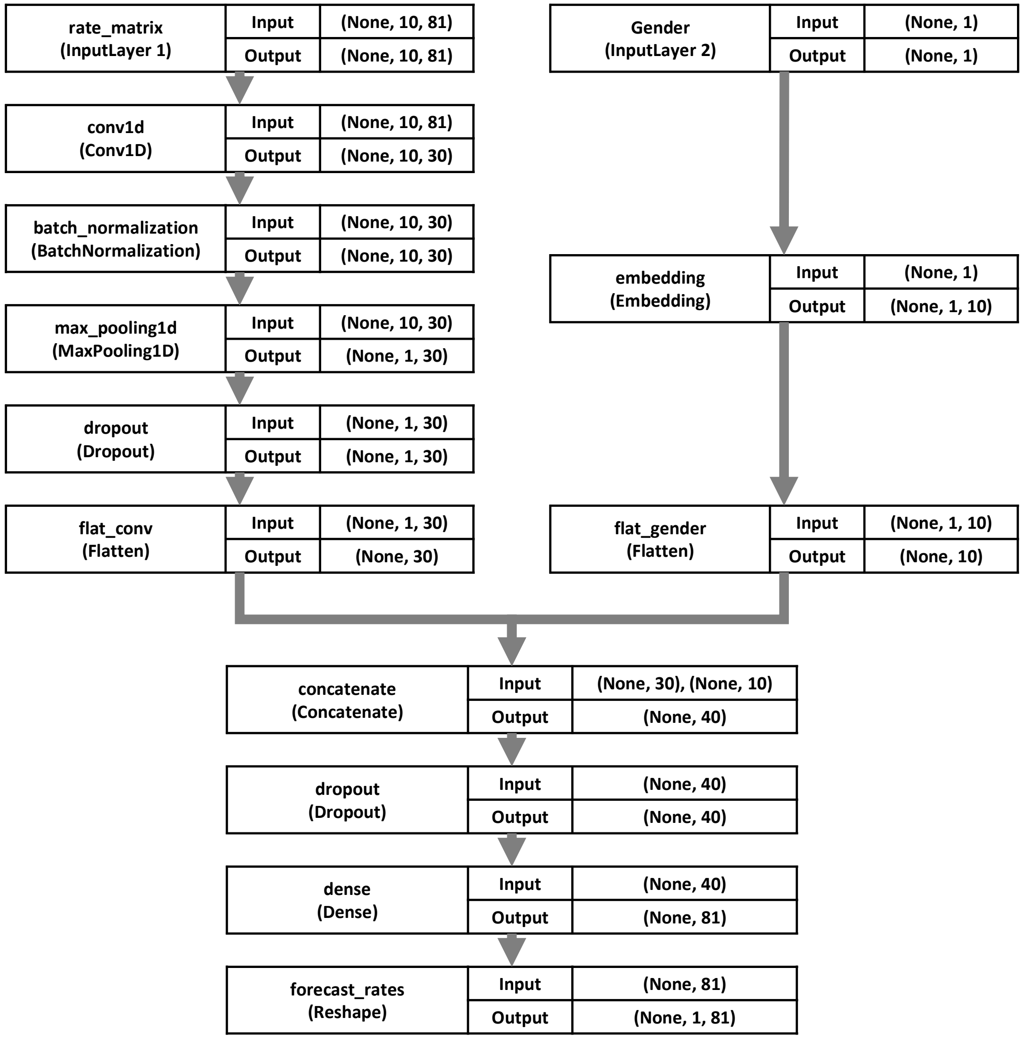
Figures – Forecast vs. actual (Variant 3 and 4)



Figures – Forecast vs. actual (Variant 5)



Model architecture (Variant 3)



Model architecture (Variant 4 and 5)

Female

Layer		Shape
rate_matrix (InputLayer)	Input	(None, 10, 81)
	Output	(None, 10, 81)
conv1d (Conv1D)	Input	(None, 10, 81)
	Output	(None, 10, 50)
batch_normalization (BatchNormalization)	Input	(None, 10, 50)
	Output	(None, 10, 50)
max_pooling1d (MaxPooling1D)	Input	(None, 10, 50)
	Output	(None, 1, 50)
dropout (Dropout)	Input	(None, 1, 50)
	Output	(None, 1, 50)
flat_conv (Flatten)	Input	(None, 1, 50)
	Output	(None, 50)
dropout (Dropout)	Input	(None, 50)
	Output	(None, 50)
dense (Dense)	Input	(None, 50)
	Output	(None, 81)
forecast_rates (Reshape)	Input	(None, 81)
	Output	(None, 1, 81)

Male

Layer		Shape
rate_matrix (InputLayer)	Input	(None, 10, 81)
	Output	(None, 10, 81)
conv1d (Conv1D)	Input	(None, 10, 81)
	Output	(None, 10, 10)
batch_normalization (BatchNormalization)	Input	(None, 10, 10)
	Output	(None, 10, 10)
max_pooling1d (MaxPooling1D)	Input	(None, 10, 10)
	Output	(None, 1, 10)
dropout (Dropout)	Input	(None, 1, 10)
	Output	(None, 1, 10)
flat_conv (Flatten)	Input	(None, 1, 10)
	Output	(None, 10)
dropout (Dropout)	Input	(None, 10)
	Output	(None, 10)
dense (Dense)	Input	(None, 10)
	Output	(None, 81)
forecast_rates (Reshape)	Input	(None, 81)
	Output	(None, 1, 81)

Model architecture (new model)