STATISTICAL SCIENCE

UNIVERSITY COLLEGE LONDON

Fine-grained mortality forecasting with deep learning

Insurance data science conference, London, 19th June 2025

Huiling Zheng*, Hai Wang, Rui Zhu, Jing-Hao Xue

* huiling.zheng.16@ucl.ac.uk

- Mortality rate $q_{x,t}$ is the probability of death at a given age, in a particular year.
- Mortality predictions are crucial assumptions for actuaries.

Annual National Mortality Forecasting, Autoregressive

Traditional forecasting, e.g.,

- Lee-Carter model (Lee and Carter (1992))
- Cairns-Blake-Dowd (CBD) (Cairns et al. (2006))

Deep learning forecasting, e.g.,

- RNN/LSTM (Nigri et al. (2019), Chen and Khaliq (2023))
- CNN (Perla et al. (2021), Wang et al. (2021))
- Transformers (Roshani et al. (2022), Wang et al. (2024))

Year

Age

$q_{18,2000}$	$q_{18,2001}$	$q_{18,2002}$
$q_{19,2000}$		

by Country, Gender

Fine-grained mortality forecasting: exogenous predictors, weekly, regional level data (Robben et al. (2025))

Robben et al. (2025): fine-grained mortality modelling, using XGBoost to link exogenous data (weather and pollution) to weekly death deviations from a Serfling-type Baseline.

Why Fine-grained?

- Climate extremes + ageing population demand weekly, regional forecasts.
- Traditional annual models miss short-term & local spikes.
- Regulators urge climate-aware risk assessment.

We propose **MortFCNet**: forecast weekly death rates from region-specific weather inputs using deep neural networks, to learning these complex relationships, without manual feature engineering.

Data sources

1. Weekly Death Counts
2. Weekly Exposure to Risk
3. Daily Weather Data
4. Geographical Coordinates

Weather Variables	Descriptions
Tmax	Highest air temperature observed each day (°C) recorded 2 meters above ground level.
Tavg	Daily mean air temperature (°C) recorded 2 meters above the ground level.
Tmin	Lowest air temperature reached during the day (°C) recorded 2 meters above the ground level.
Hum	Daily average relative humidity, measured at 2 meters above the ground level. Relative humidity is defined as the percentage of actual humidity relative to saturation humidity.
Rain	Total daily precipitation for the day (mm), reported as liquid-water equivalent and encompassing rain, snow, and hail per square meter.
Wind	Daily mean wind speed in ($m\ s^{-1}$), measured at 10 meters above ground level.

Table 1: Weather factors and descriptions ((ECMWF), 2023)).

Aggregation and feature engineering

1. Population-weighted gridded weather → region-week
2. Weekly averages (Tmax, Tmin, Tavg, RH, rain, wind)
3. Anomalies & extreme indices (e.g., 95th / 5th percentile)
4. Seasonal tag (Spring…Winter)
5. 1-week lags

Baseline and XGBoost

$\boldsymbol{\mu_{t,v}^{g}}$ - fine-grained mortality rate for year t , week v, and region g

Serfling-type Baseline

Poisson GLM: $\log \mu_{t,v}^{(g)} = \theta_0^{(g)} + \theta_1^{(g)} t + \theta_2^{(g)} \sin\left(\frac{2\pi v}{52.18}\right) + \theta_3^{(g)} \cos\left(\frac{2\pi v}{52.18}\right) + \theta_4^{(g)} \sin\left(\frac{2\pi v}{26.09}\right) + \theta_5^{(g)} \cos\left(\frac{2\pi v}{26.09}\right)$

- Calibrated with spatial smoothness via quadratic penalty on neighbour coefficients
- Captures broad seasonality but fixed shape, no exogenous drivers

XGBoost Calibration

- Learns multiplier φ to adjust baseline deaths
- Inputs: engineered weather, spatial & seasonality
- Good in-sample fit; prone to overfit & manual feature burden

Experimental Design

- Train 2013-2018, Test 2019. Cross-validated hyper-parameters. Over 200 NUT-3** regions from Switzerland (CH), France (FR) and Italy (IT), age 65+ only, unisex.

**The Nomenclature of Territorial Units for Statistics (NUTS), NUTS-3 represents the smallest regional level.

MortFCNet Architecture

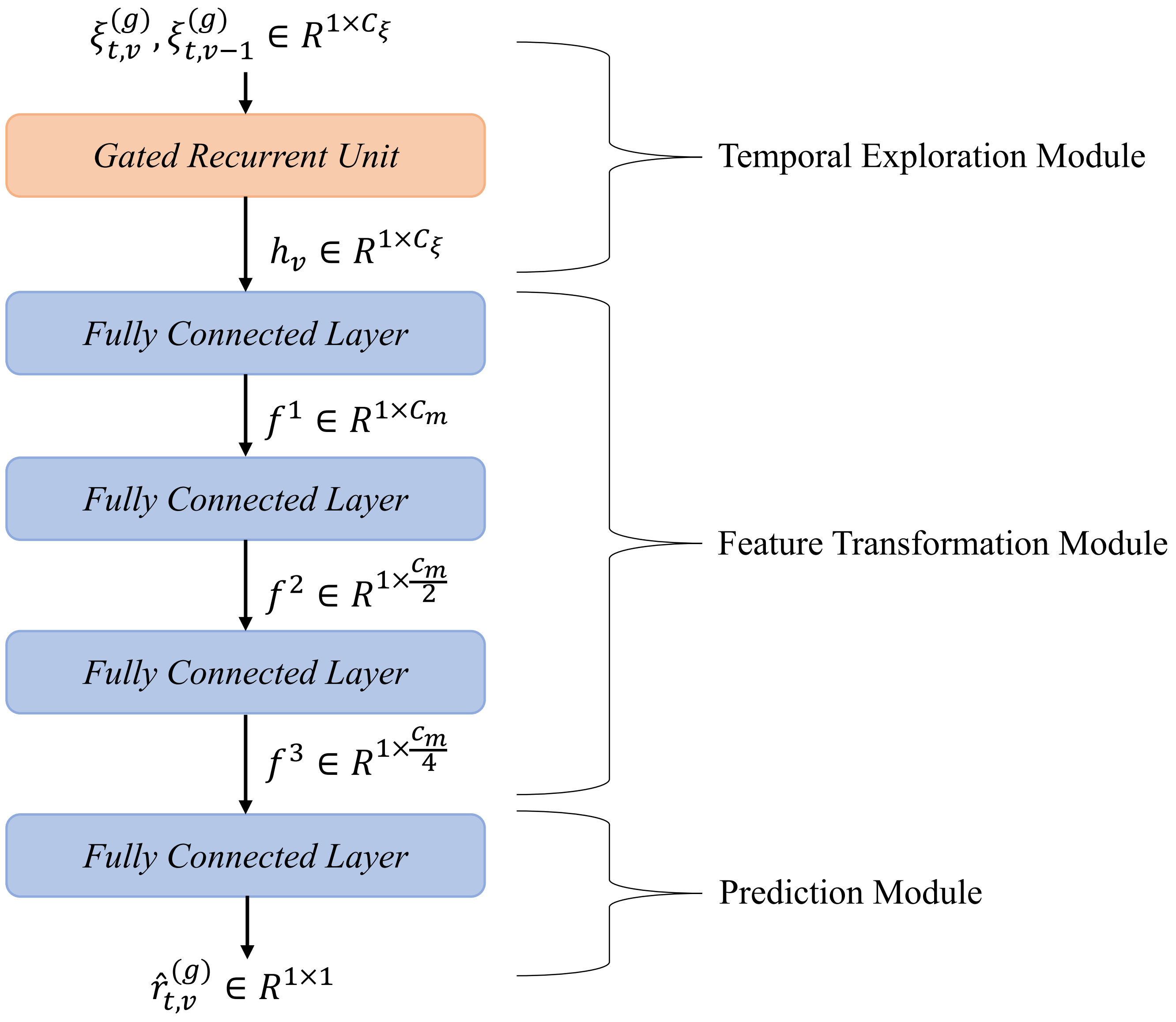


Figure 1: MortFCNet Architecture

Results - Overall

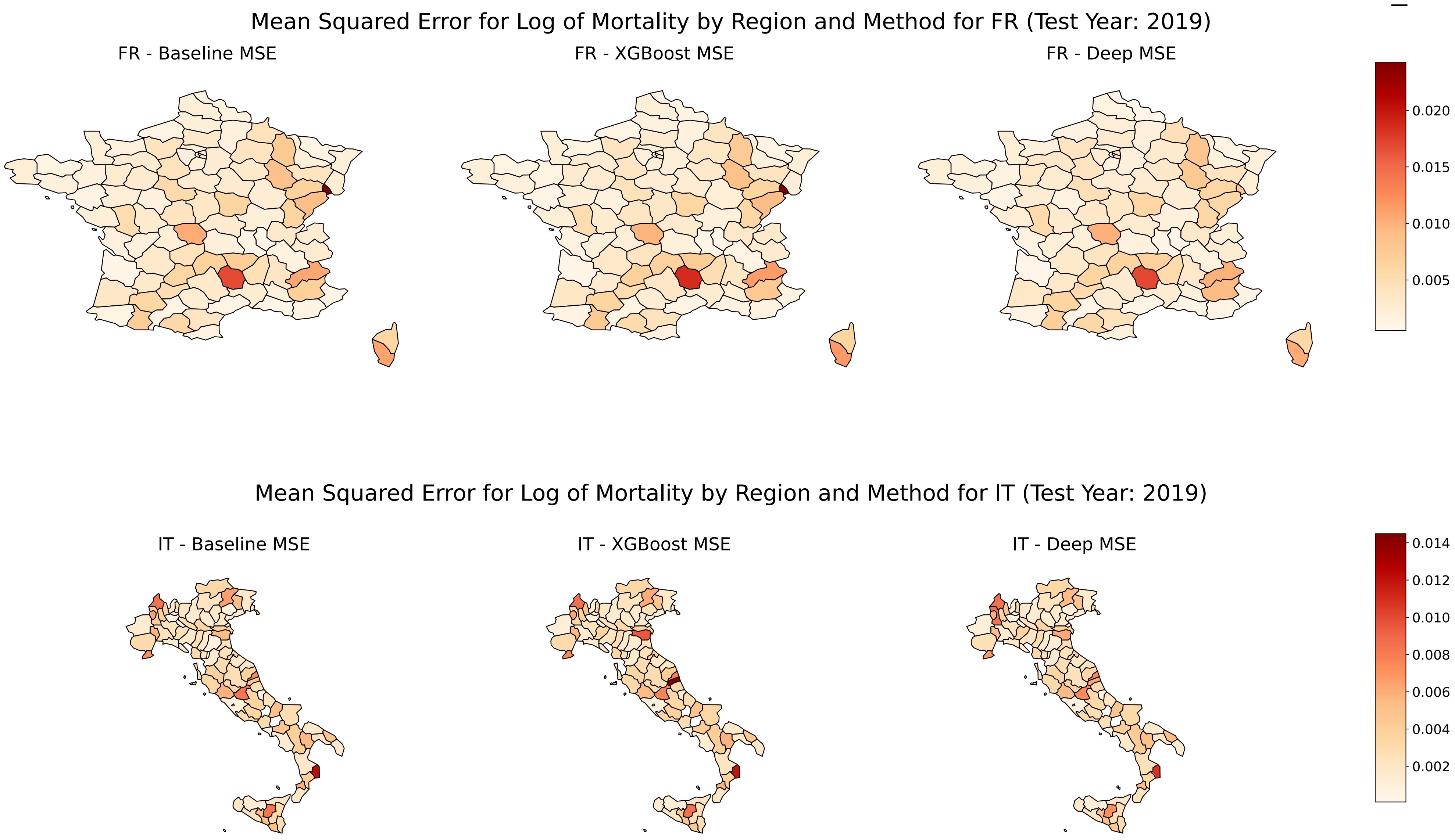


Figure 2: MSE heatmap for a) France, b) Italy, and c) Switzerland. Benchmarking Baseline, XGBoost and MortFCNet.

Results - Overall

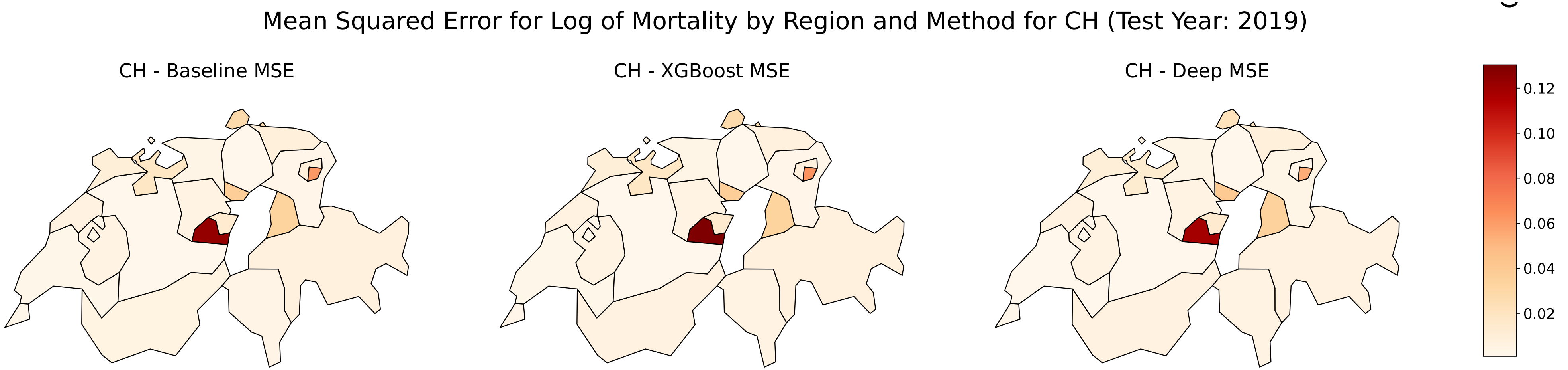
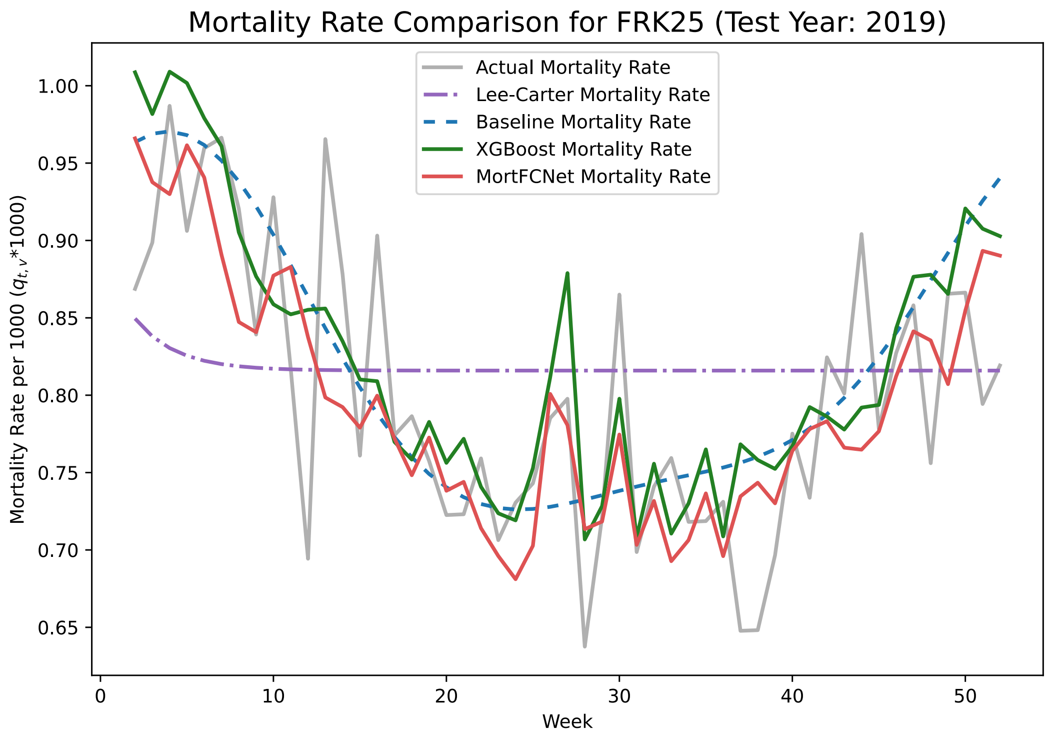


Figure 2: MSE heatmap for a) France, b) Italy, and c) Switzerland. Benchmarking Baseline, XGBoost and MortFCNet.

		MSE ($\times 10^6$)	
		Training (2013-2018)	**Test (2019)**
Overall	Lee-Carter	0.0181	0.0201
Overall	Baseline	0.0150	0.0134
Overall	XGBoost	0.0131	0.0138
Overall	MortFCNet	0.0123	0.0125

Table 2: overall train and test MSE performance across all three countries. Benchmarking Lee-Carter, Baseline, XGBoost and MortFCNet

Results - Regional Case Studies



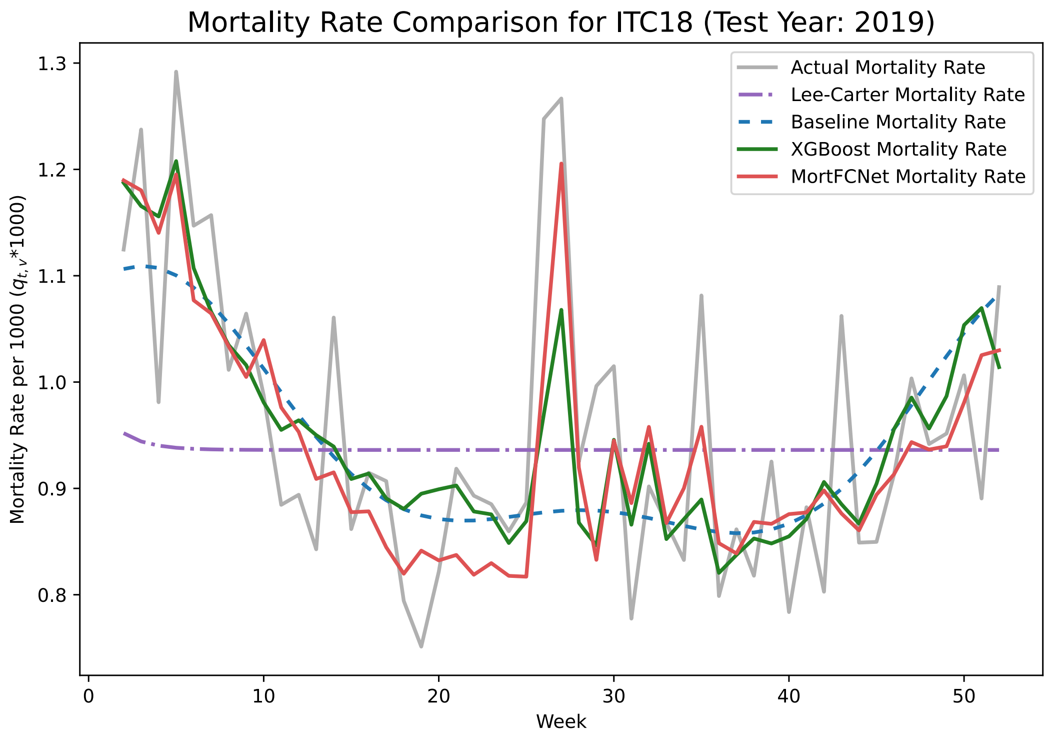


Figure 3: Test MSE for FRK25 and ITC18 regions. Benchmarking Lee-Carter, Baseline, XGBoost and MortFCNet.

Regions	MSE ($\times 10^6$)			
	Lee-Carter	Baseline	XGBoost	MortFCNet
ITC18	0.0177	0.0127	0.0084	0.0068
FRK25	0.0077	0.0044	0.0041	0.0039

Table 3: test MSE performance for regional case studies. Benchmarking Lee-Carter, Baseline, XGBoost and MortFCNet.

Remove anomalies & extremes; keep raw weekly averages

- XGBoost: test MSE slightly increased → needs engineered vars
- MortFCNet: test MSE decreased → automatic feature learning

Model	Training/Test Years	MSE($\times 10^6$)
XGBoost with Feature Engineering	2013–2018	0.0131
	2019	0.0138
XGBoost without Feature Engineering	2013–2018	0.0130
	2019	0.0139
MortFCNet with Feature Engineering	2013-2018	0.0123
	2019	0.0125
MortFCNet without Feature Engineering	2013–2018	0.0128
	2019	0.0123

Table 4: overall train and test MSE performance across all three countries. Benchmarking XGBoost and MortFCNet, with and without feature engineering.

Thank you

Reference

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

Cairns, A. J., Blake, D., & Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty. Journal of Risk and Insurance, 73(4), 687–718.

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., & Perla, F. (2019). A deep learning integrated Lee–Carter model. Risks, 7(1), 33.

Chen, Y., & Khaliq, A. Q. M. (2023). Mortality rates forecasting with data driven LSTM, Bi-LSTM and GRU: The United States case study. Actuarial Research Clearing House, 2023.1.

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.

Wang, C. -W., Zhang, J., & Zhu, W. (2021). Neighbouring prediction for mortality. ASTIN Bulletin: The Journal of the IAA, 51(3), 689–718.

Roshani, A., Izadi, M., & Khaledi, B. (2022). Transformer self-attention network for forecasting mortality rates. Journal of the Iranian Statistical Society, 21(1), 81–103.

Wang, J., Wen, L., Xiao, L., & Wang, C. (2024). Time-series forecasting of mortality rates using Transformer. Scandinavian Actuarial Journal, 2024(2), 109–123.

European Centre for Medium-Range Weather Forecasts (ECMWF). (2023). Copernicus climate data store (cds) [Accessed: 2024-10-26]. https://cds.climate.copernicus.eu

Robben, J., Antonio, K., & Kleinow, T. (2025). The short-term association between environmental variables and mortality: Evidence from Europe. Journal of the Royal Statistical Society Series A: Statistics in Society, 2025; qnaf052.