A Temporal Convolutional Network Approach to Learning From Longitudinal Data

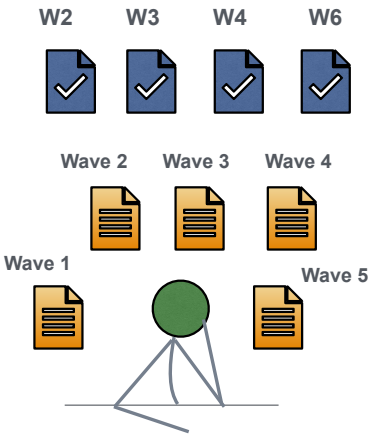
Marjan Qazvini

Email: marjan.qazvini@gmail.com

Longevity 19 Conference, Amsterdam, 2024

Introduction

- 5 files for the life survey, which record information across 5 waves.
- 4 files for the end-of-life survey, which record the participants' status in waves 2, 3, 4 and 6.
- New cohorts were added in waves 3 and 4.
- Inconsistency in some questions and their response options across 5 waves.
- Some join in Wave 1 or later and continue through all 5 waves or leave earlier.



ELSA dataset

- 14,966 participants across 5 waves.
- 61 features: age is continuous and the rest are categorical features.
- Demographic, disease, health and lifestyle factors.
- Missing values are filled with the mean for continuous features and modes for categorical features.
- 976 deaths after 5 waves.

	Wave1	Wave2	Wave3	Wave4	Wave5
Males	46%	55%	55%	55%	55%
Females	54%	45%	45%	45%	45%
Mean age	65.27	66.92	65.94	66.11	67.77

Percentage of people die at different ages

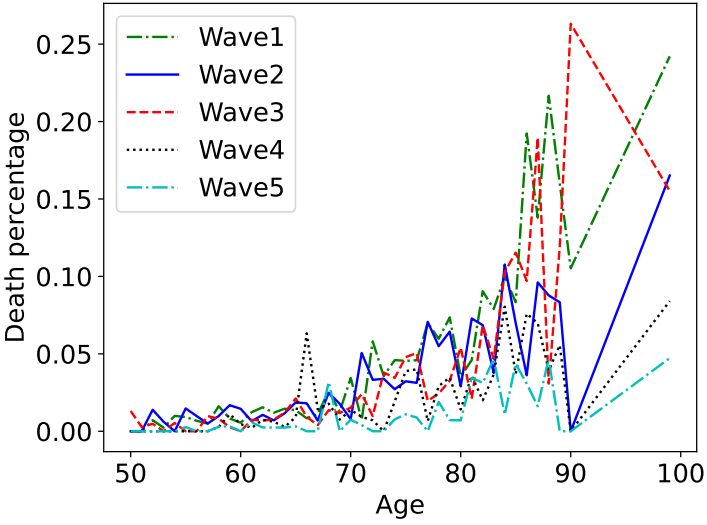
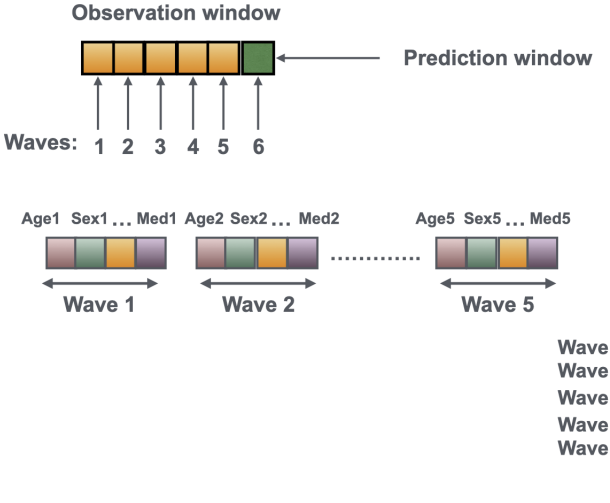


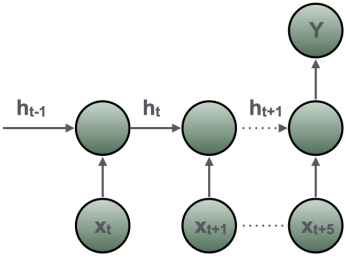
Figure: Ages are capped at 90 and recorded as 99.

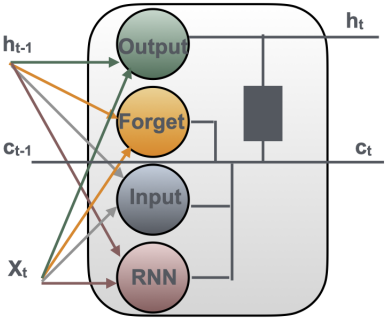
Data preparation

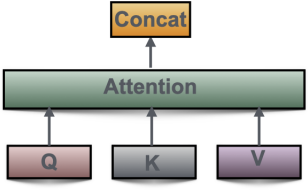


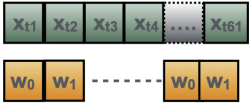
Zhao, Juan, QiPing Feng, Patrick Wu, et al. (2019)

Models


Y
h_{t-1}
h_t
h_{t+1}
x_t
x_{t+1}
x_{t+5}


h_{t-1}
h_t
Output
Forget
Input
RNN
c_{t-1}
c_t
X_t

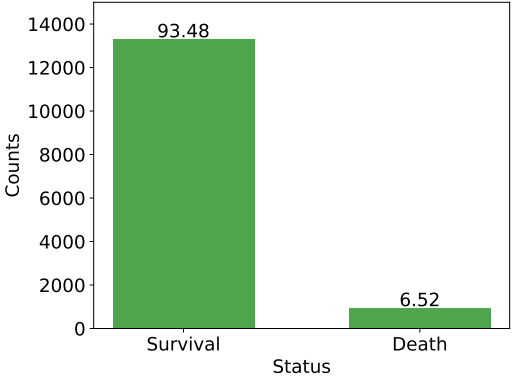

Concat
Attention
Q
K
V


x_{t1}
x_{t2}
x_{t3}
x_{t4}
....
x_{t61}
w_0
w_1
w_0
w_1

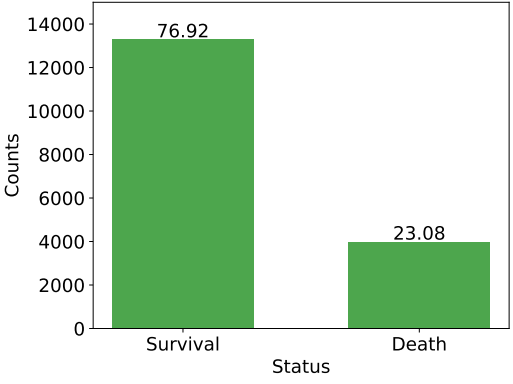
Imbalanced dataset

- Weighted loss function
- Generating synthetic data:
  - Under-representing the over-represented class
  - Over-representing the under-represented class

Death percentage

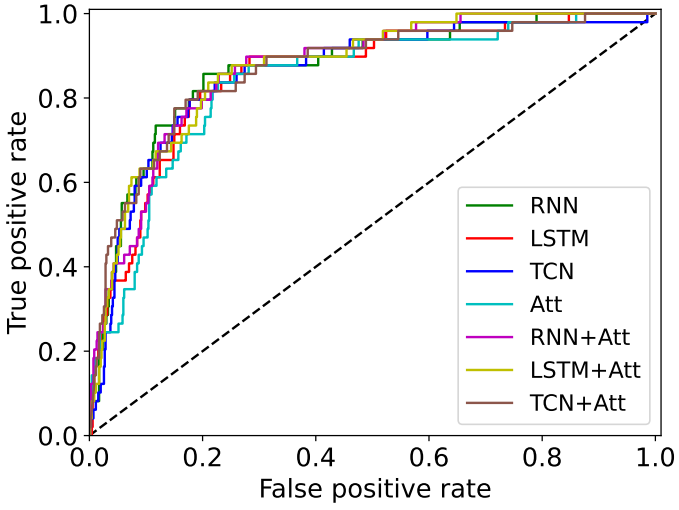


(a) Original data



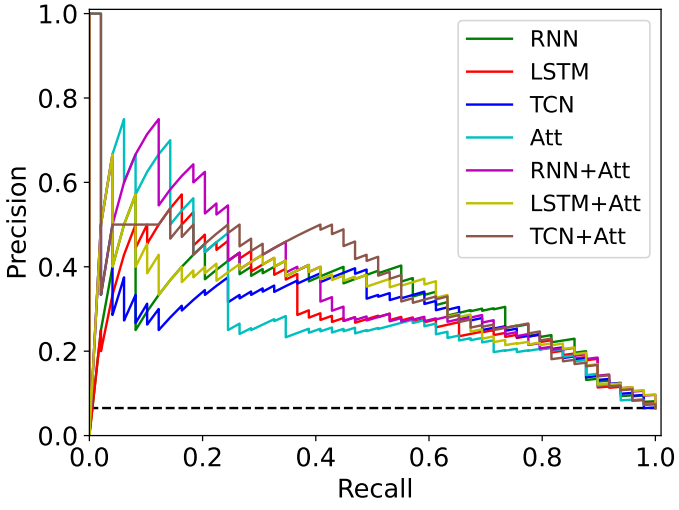
(b) Synthetic data

ROC (original data)



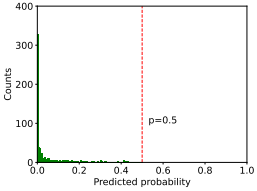
Models	AUC
RNN	0.869
LSTM	0.855
TCN	0.858
Att	0.839
RNN+Att	0.870
LSTM+Att	0.872
TCN+Att	0.867

Recall-Precision (original data)

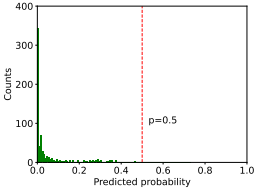


Models	AUC
RNN	0.314
LSTM	0.298
TCN	0.284
Att	0.311
RNN+Att	0.347
LSTM+Att	0.319
TCN+Att	0.365

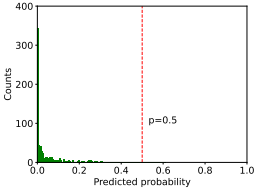
Predicted probability (original data)



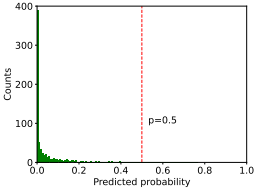
(a) RNN



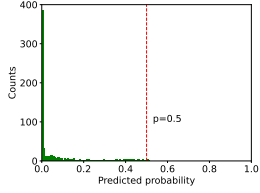
(b) LSTM



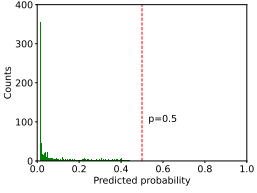
(c) TCN



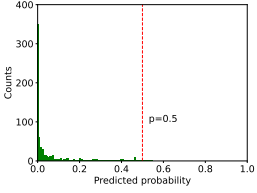
(d) Att



(e) RNN+Att

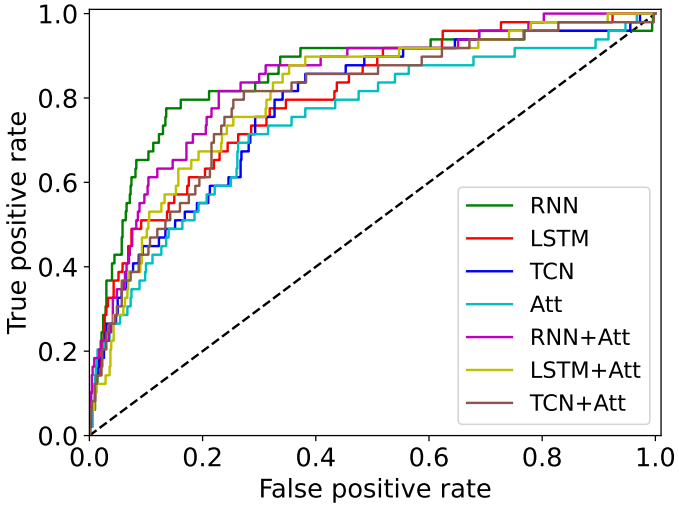


(f) LSTM+Att



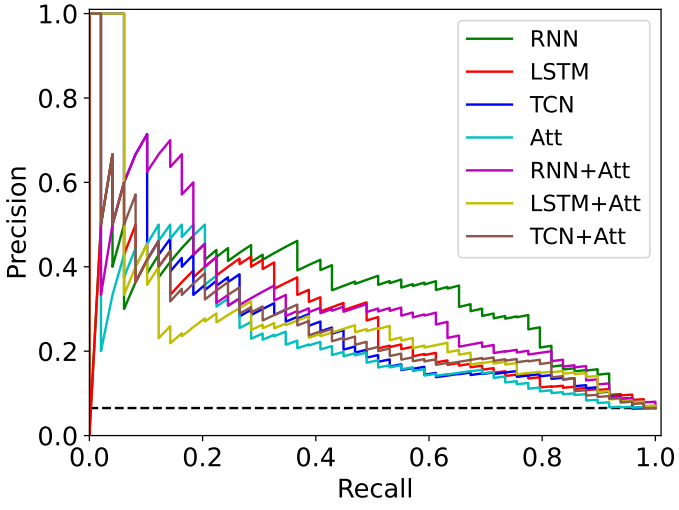
(g) TCN+Att

ROC (synthetic data)



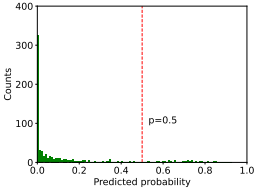
Models	AUC
RNN	0.851
LSTM	0.800
TCN	0.782
Att	0.744
RNN+Att	0.839
LSTM+Att	0.808
TCN+Att	0.794

Recall-Precision (synthetic data)

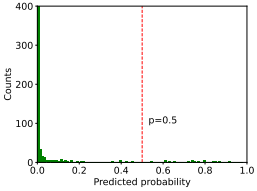


Models	AUC
RNN	0.327
LSTM	0.265
TCN	0.280
Att	0.232
RNN+Att	0.321
LSTM+Att	0.266
TCN+Att	0.259

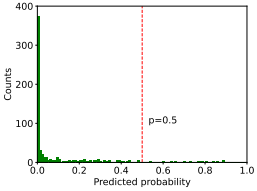
Predicted probability (synthetic data)



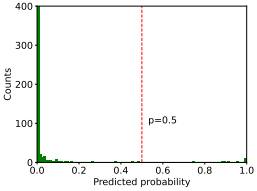
(a) RNN



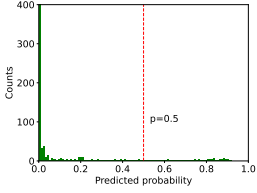
(b) LSTM



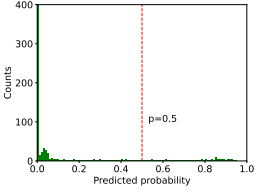
(c) TCN



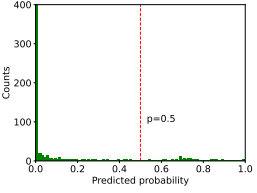
(d) Att



(e) RNN+Att

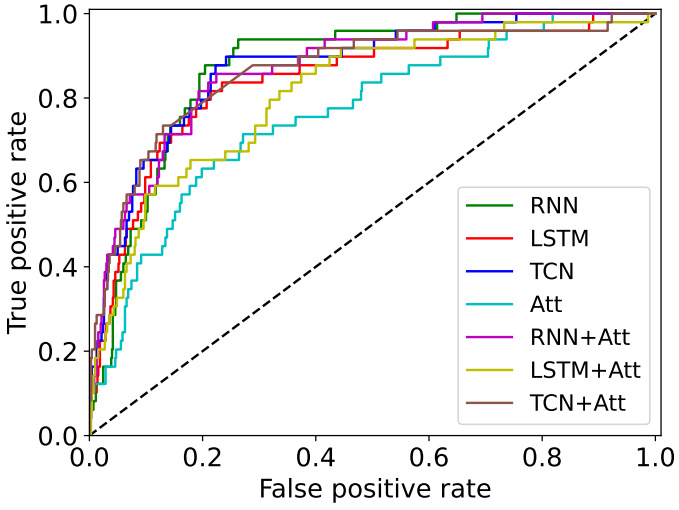


(f) LSTM+Att



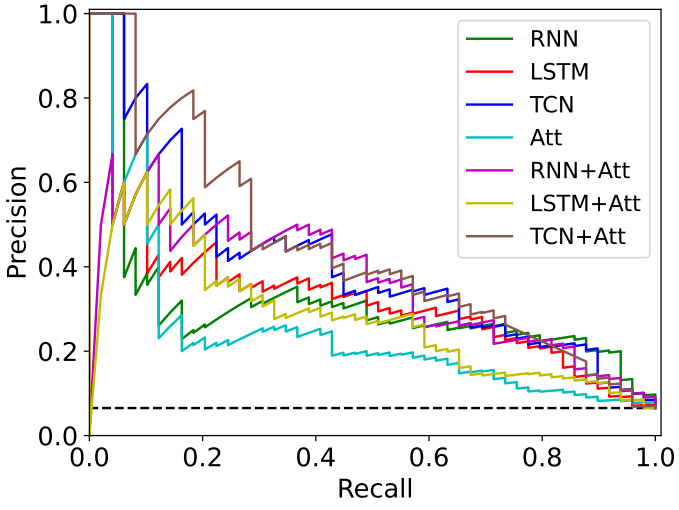
(g) TCN+Att

ROC (weighted loss)



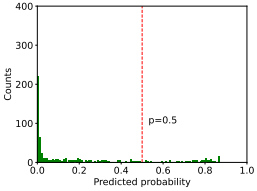
Models	AUC
RNN	0.875
LSTM	0.842
TCN	0.874
Att	0.764
RNN+Att	0.871
LSTM+Att	0.811
TCN+Att	0.865

Recall-Precision (weighted loss)

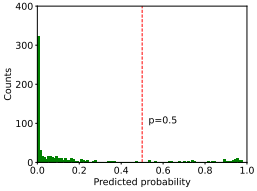


Models	AUC
RNN	0.313
LSTM	0.328
TCN	0.403
Att	0.242
RNN+Att	0.355
LSTM+Att	0.273
TCN+Att	0.439

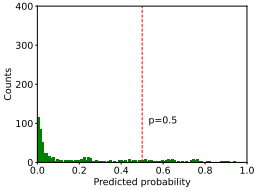
Predicted probability (weighted loss)



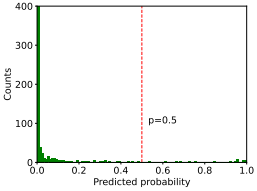
(a) RNN



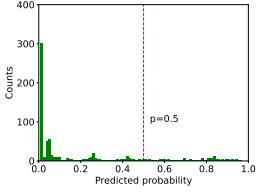
(b) LSTM



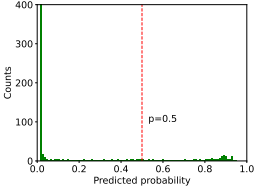
(c) TCN



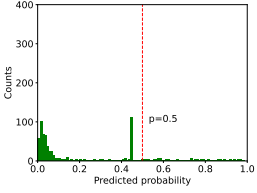
(d) Att



(e) RNN+Att



(f) LSTM+Att



(g) TCN+Att

Average accuracy for 10 repeatitions

Models	Original data	Synthetic data	Weighted loss
RNN	93.16 ± 0.838	88.52 ± 2.177	72.54 ± 5.291
LSTM	93.41 ± 0.427	87.16 ± 0.889	81.04 ± 1.941
TCN	93.20 ± 0.479	87.60 ± 0.636	77.49 ± 2.474
Att	90.98 ± 0.954	87.22 ± 0.702	80.05 ± 4.607
RNN+Att	93.22 ± 0.430	87.01 ± 1.061	77.06 ± 3.162
LSTM+Att	92.43 ± 0.886	87.65 ± 1.077	79.01 ± 2.963
TCN+Att	93.16 ± 0.540	86.82 ± 1.039	77.80 ± 2.713

Feature analysis

Statistical model	Machine learning models		Deep model
LR (coefficients)	RF (impurity)	GBT (permutation)	Att (scores)
Heart failure	Age	Age	Using a map (IADL)
Gardening (IADL)	Public tran	Vigorous sport	Alzheimer's
Dementia	Hearing	Heart failure	Marital status
Alzheimer's	Eyesight	Mod sport	Bathing (ADL)
Heart attack	Drinking	Cancer	Heart attack
Cancer	Marital status	Hearing noise	Diabetes
Sitting (Mobility)	Walking w h	Shopping (IADL)	Mild sport
Sleeping disorder	Employment	Drinking	Sex
Walking w h	Mod sport	Gardening (IADL)	Get up (Mobility)
Sex	Mild sport	Walking w help	Shopping (IADL)

Discussion

- Machine learning and deep learning models are more efficient than statistical models for making predictions as they capture non-linear relationships between features.
- The main purpose of statistical models is to interpret dependencies between dependent and independent variables rather than for classification or prediction.
- Qazvini (2023) used statistical models such as GLM and GLMM to study ELSA.
- Demographics, diseases, mobility, ADL, IADL and lifestyle factors are among the top 10 factors that affect mortality.