Temporal Pointwise Convolution Networks for Length of Stay Prediction in the ICU

Emma Rocheteau, University of Cambridge
Pietro Liò, University of Cambridge
Stephanie Hyland, Microsoft Research
With thanks to the Armstrong Fund

Introduction

- We propose a new deep learning model – Temporal Pointwise Convolution (TPC) – which combines **temporal convolution and pointwise (1x1) convolution**, to solve the **length of stay and mortality prediction** tasks on the **eICU and MIMIC-IV** critical care datasets.
- We have achieved **significant performance benefits of 18-68%** (metric dependent) over the commonly used Long-Short Term Memory (LSTM) network, and the multi-head self-attention network known as the Transformer.

Data

There is large variability in the behaviour between different time series variables.

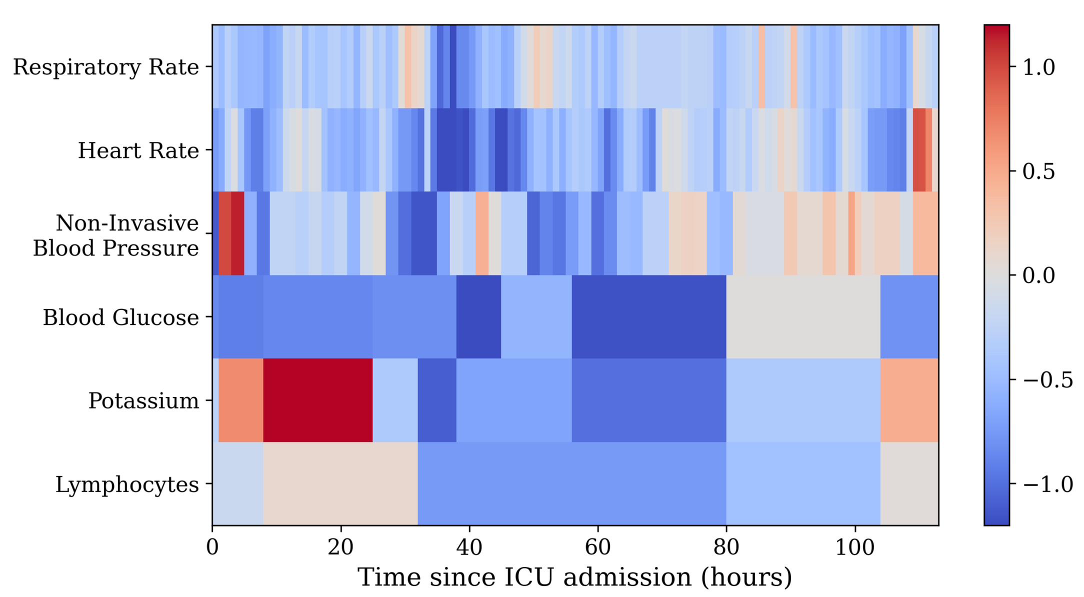

Figure 1: Example data from one patient.

Temporal Convolution

We use temporal convolution to extract **trends** in the data. We use skip connections so that the model can **choose the temporal receptive field**. We have **different parameters for each variable** to allow the model to tailor processing to each.

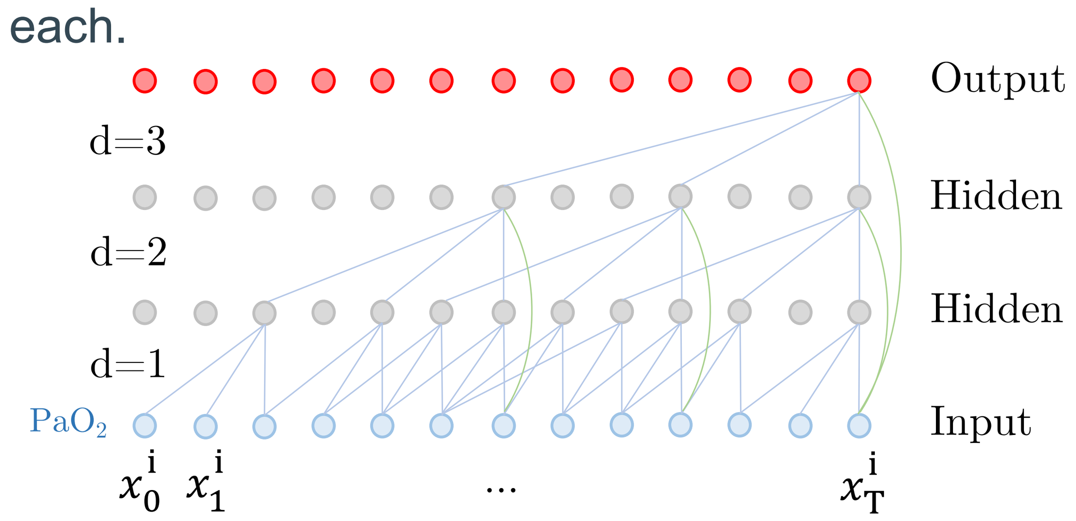

Figure 2: Temporal convolution with skip connections.

Pointwise Convolution

We use pointwise convolution to extract **relationships between features** at each timepoint.

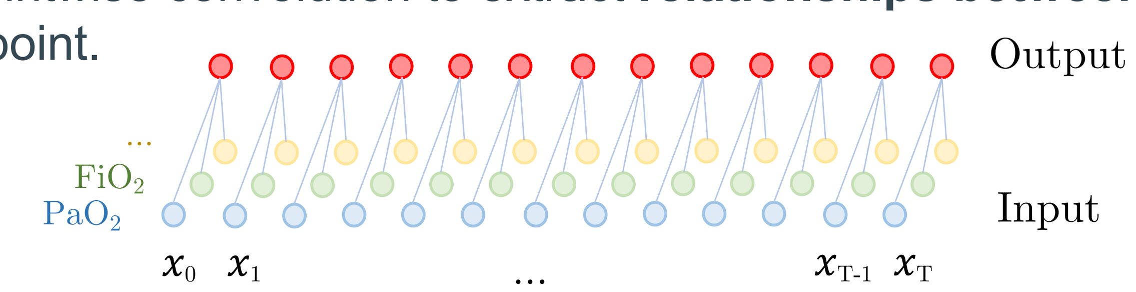

Figure 3: Pointwise convolution.

Model Architecture

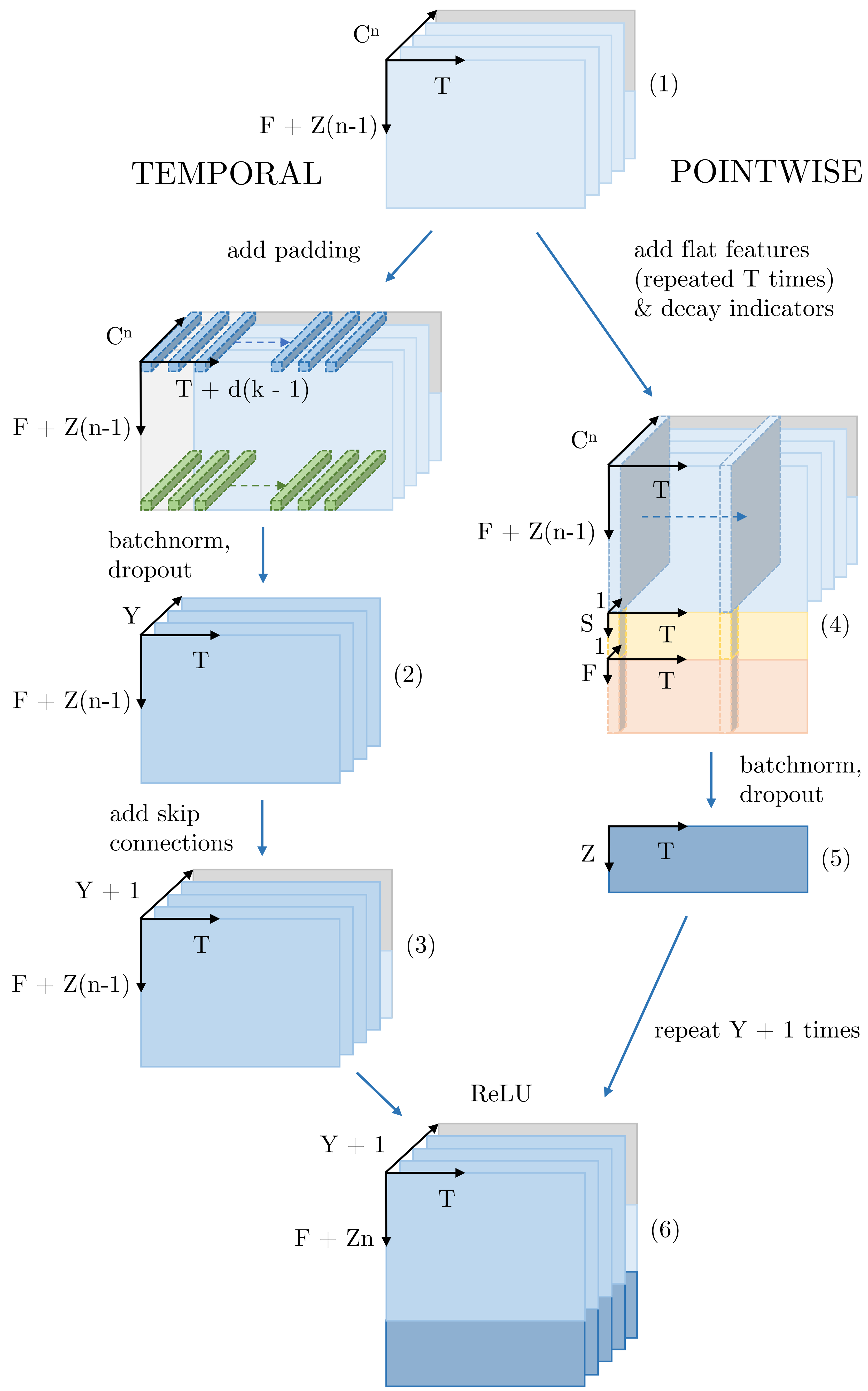

Figure 4: One layer of the TPC model. Temporal Convolution and Pointwise Convolution occur in parallel. In the temporal branch, independent parameters process each time series feature. See the paper for more details.

Results

The TPC model **significantly outperforms baselines** by very large margins.

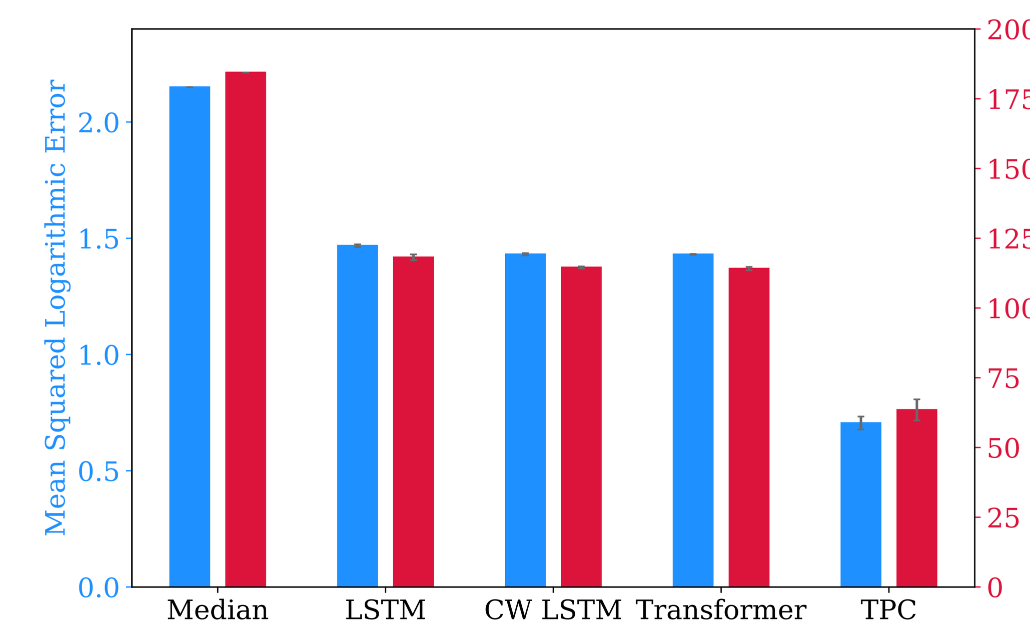

Figure 5: TPC performance compared to baselines.

Model ablations studies reveal that:

- The **temporal convolutions are more important** than the pointwise convolutions. However, the best performing model (TPC) uses both.
- **Weight sharing** between temporal convolutions significantly hurts performance.
- The ***mean squared logarithmic error* (MSLE) is more appropriate** than mean squared error (MSE) for positively skewed tasks such as LOS.
- **Skip connections** significantly improve performance.

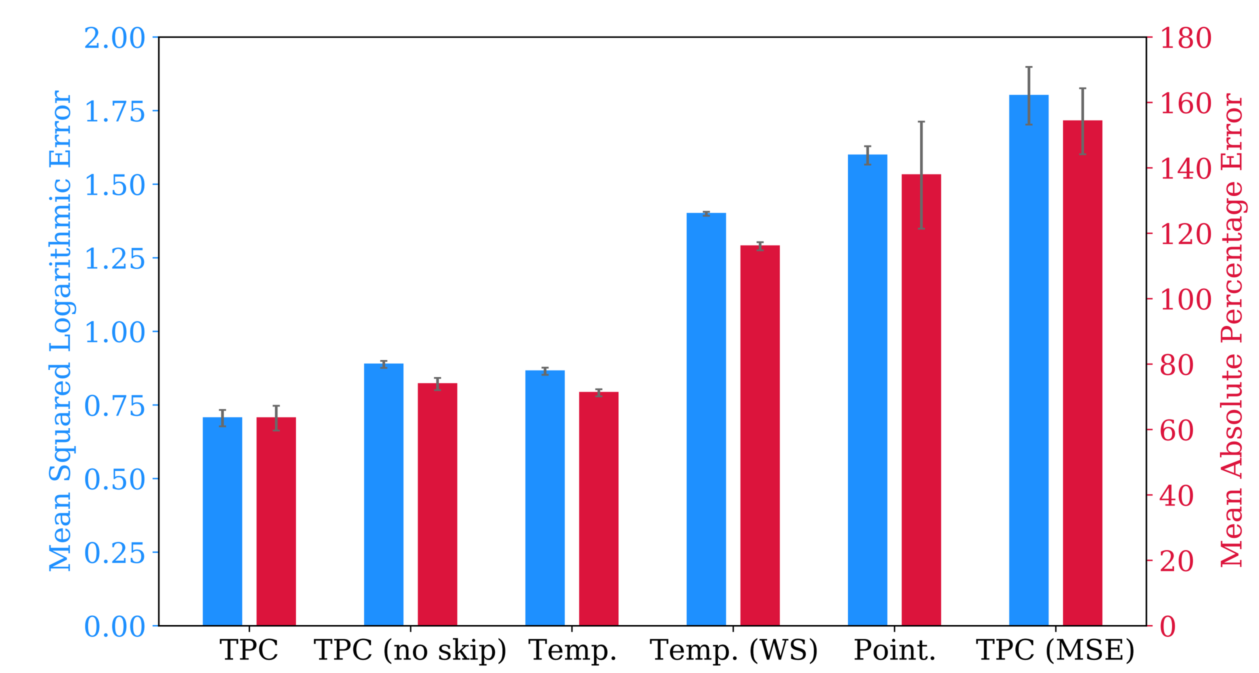

Figure 6: Ablation studies on the TPC model.

Take Home

The TPC model is well-equipped to analyse EHR time series containing missingness, different frequencies and sparse sampling. We believe that the following four aspects contribute the most to its success:

- The combination of **two complementary architectures** that are able to extract different features, both of which are important.
- The ability to **step over large time gaps**.
- The capacity to **specialise processing to each feature** (including the freedom to select the receptive field size for each).
- The **rigid spacing** of the temporal filters, making it easy to derive periodic trends.

SCAN ME