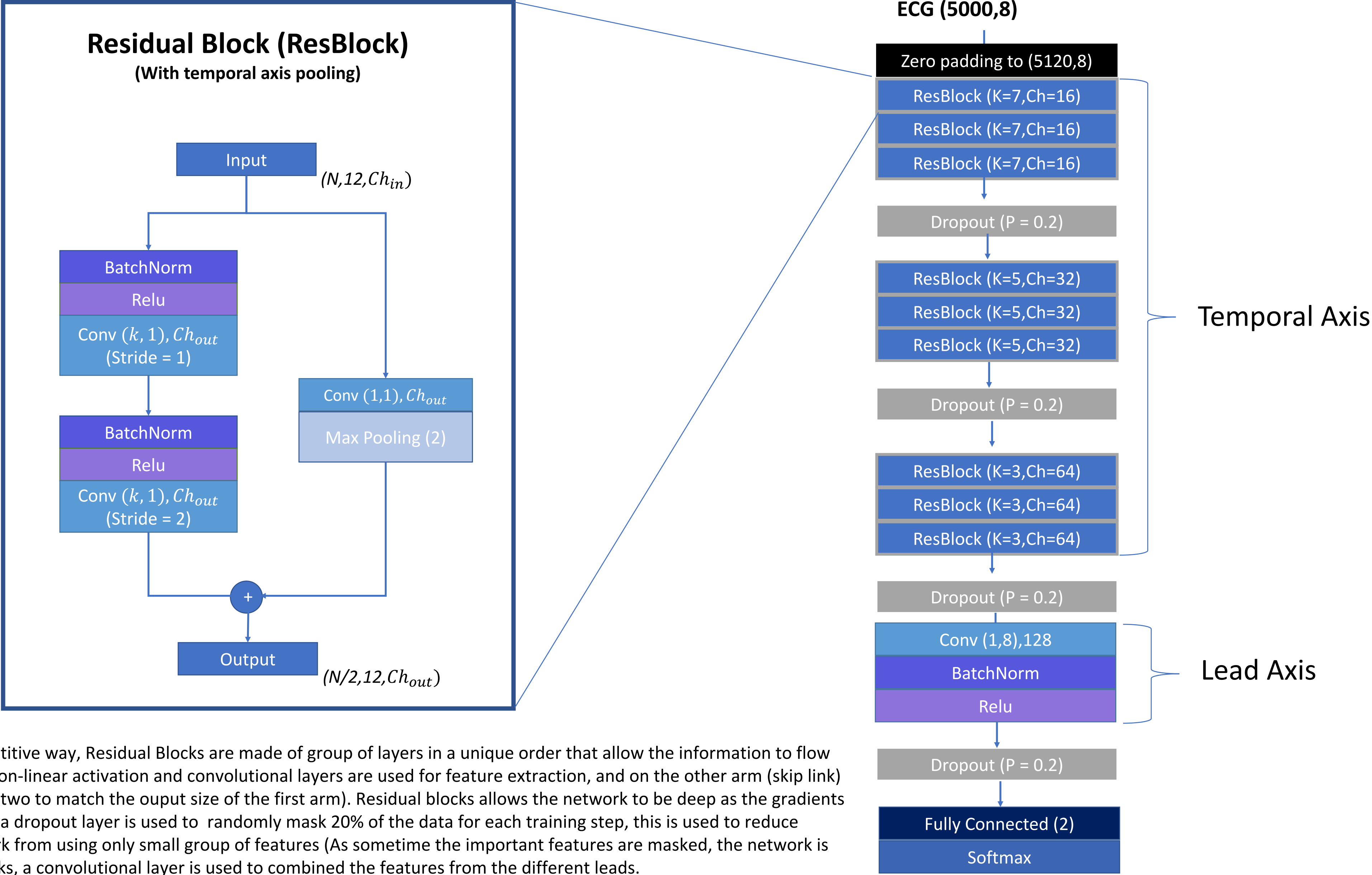# THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplemental Figure

Residual Block (ResBlock)
(With temporal axis pooling)

ECG (5000,8)

Zero padding to (5120,8)

ResBlock (K=7,Ch=16)
ResBlock (K=7,Ch=16)
ResBlock (K=7,Ch=16)

Dropout (P = 0.2)

ResBlock (K=5,Ch=32)
ResBlock (K=5,Ch=32)
ResBlock (K=5,Ch=32)

Dropout (P = 0.2)

ResBlock (K=3,Ch=64)
ResBlock (K=3,Ch=64)
ResBlock (K=3,Ch=64)

Dropout (P = 0.2)

Temporal Axis

Conv (1,8),128
BatchNorm
Relu

Lead Axis

Dropout (P = 0.2)

Fully Connected (2)
Softmax

Input $(N,12,Ch_{in})$

BatchNorm
Relu
Conv $(k,1), Ch_{out}$ (Stride = 1)

Conv $(1,1), Ch_{out}$
Max Pooling (2)

BatchNorm
Relu
Conv $(k,1), Ch_{out}$ (Stride = 2)

+

Output $(N/2,12,Ch_{out})$

The network is using different layers structured in a repetitive way, Residual Blocks are made of group of layers in a unique order that allow the information to flow in parallel, in one arm, 6 layers of batch normalisation,non-linear activation and convolutional layers are used for feature extraction, and on the other arm (skip link) the data is flowing directly (downsampling by a factor of two to match the ouput size of the first arm). Residual blocks allows the network to be deep as the gradients can flow through the skip link., after each tree Resblock, a dropout layer is used to randomly mask 20% of the data for each training step, this is used to reduce overfitting and acts as a regulizer, preventing the network from using only small group of features (As sometime the important features are masked, the network is forced to learn other features), After the 9 residual blocks, a convolutional layer is used to combined the features from the different leads.
Batch normalisation blocks are used across the model to reduce covariates shift and normalizes the data.
The "Relu" activation function and the "max pooling" are being used to allow the model to represent non-linear functions and helps learn more complex features. It also creates a certain buffer between the different layers, as if all of the layers were linear, the whole model would collapse to a shallow linear model. The use of max pooling help reduce the temporal resolution as we learn more features.