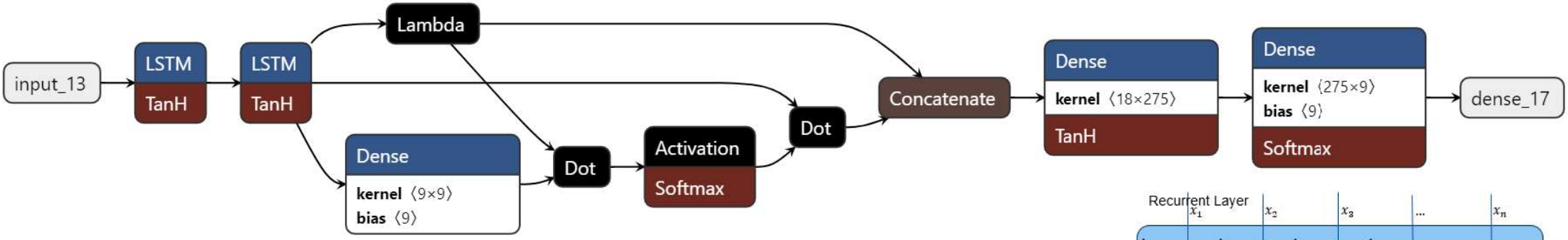
Attention and Explainability

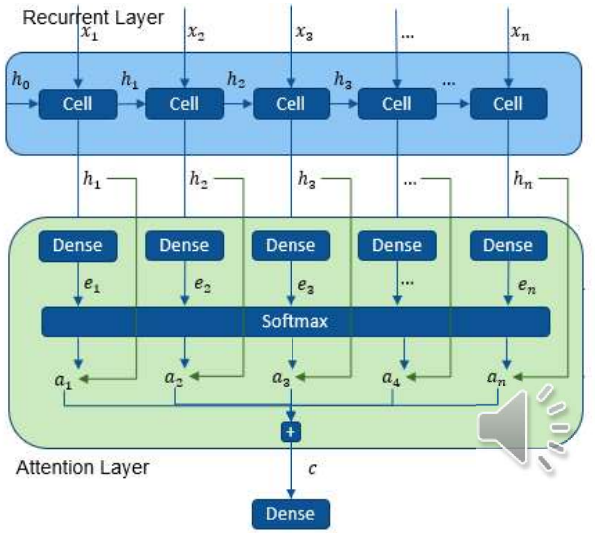
Dr. Fani Deligianni,
fani.deligianni@glasgow.ac.uk
Lecturer (Assistant Professor)
Lead of the Computing Technologies for Healthcare Theme
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

Attention in Clinical Time-Series

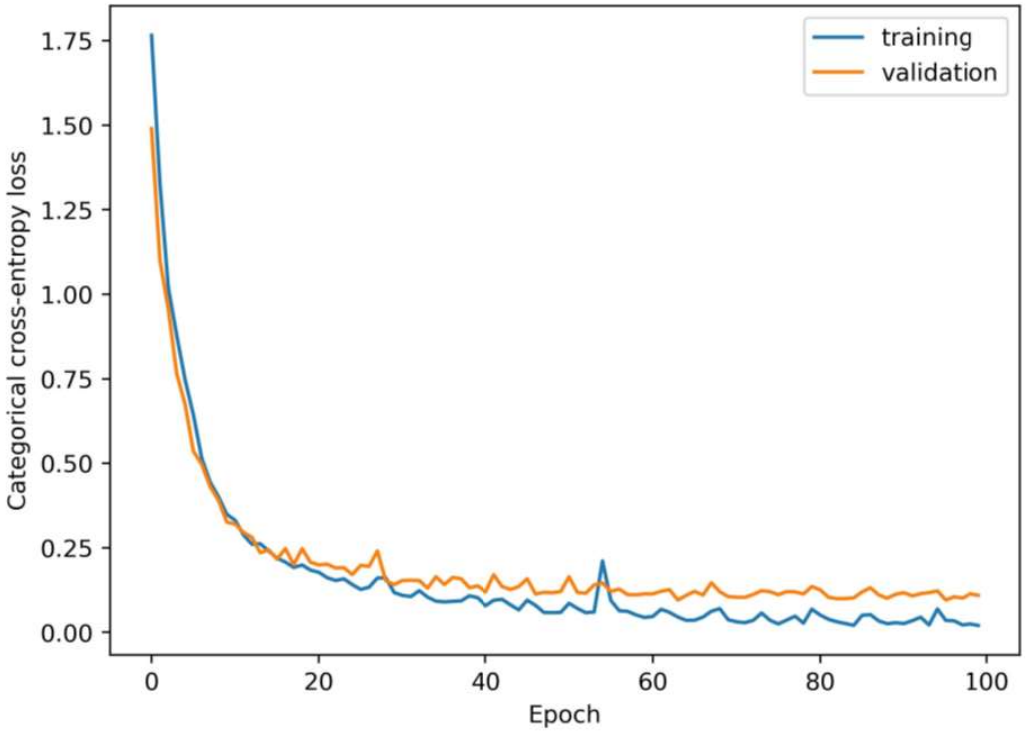
- Attention has been applied on electronic health records to enhance performance and stability of the deep learning process
- It has been used in conjunction with both CNN networks as well as RNN.
- Different flavors of attention, ie. self-attention and transformers
- MIMIC-III database is by far the most common way to test these methods on real-scale clinical datasets

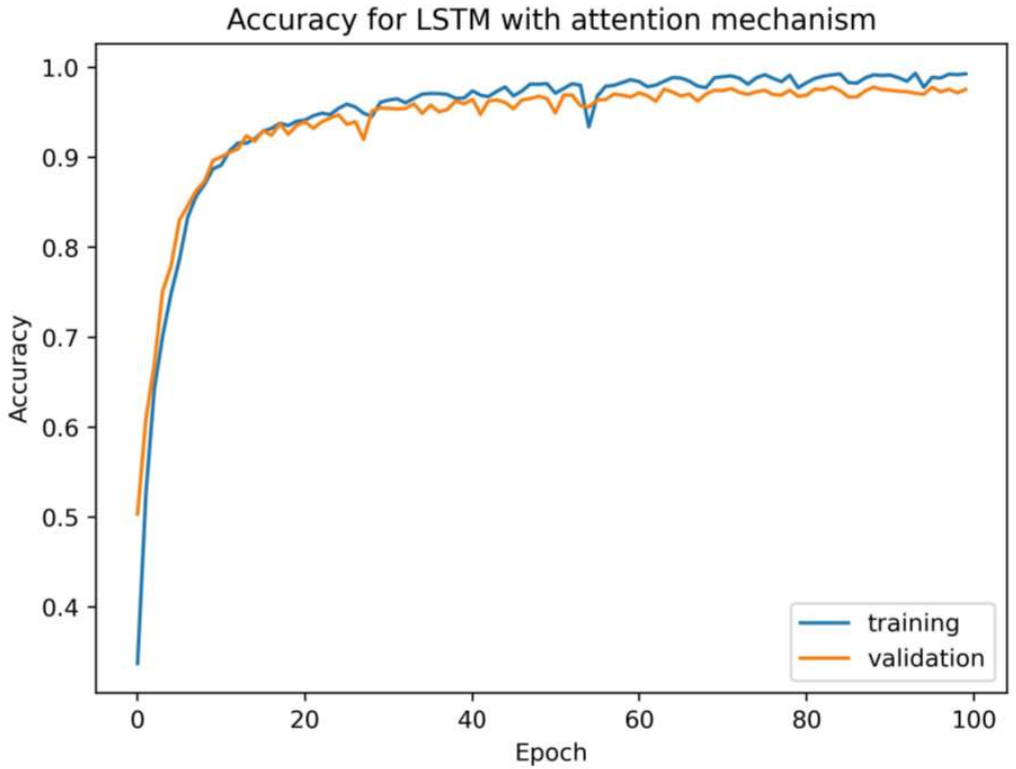
Attention Mechanism in Practice



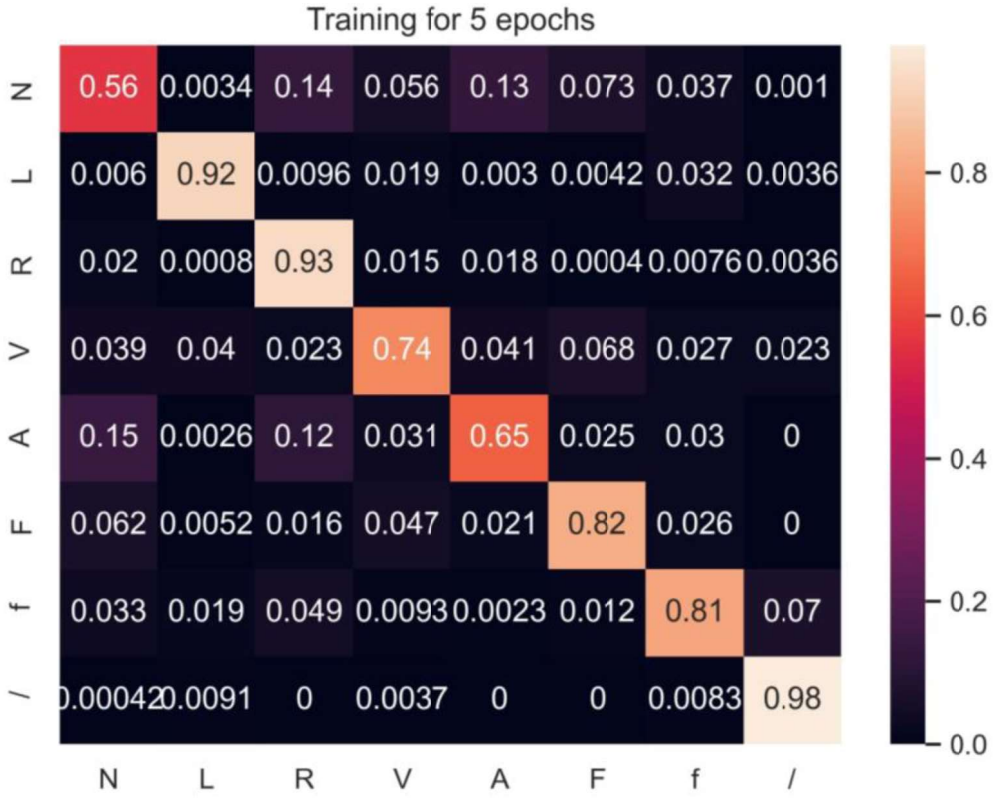


Training and Validation

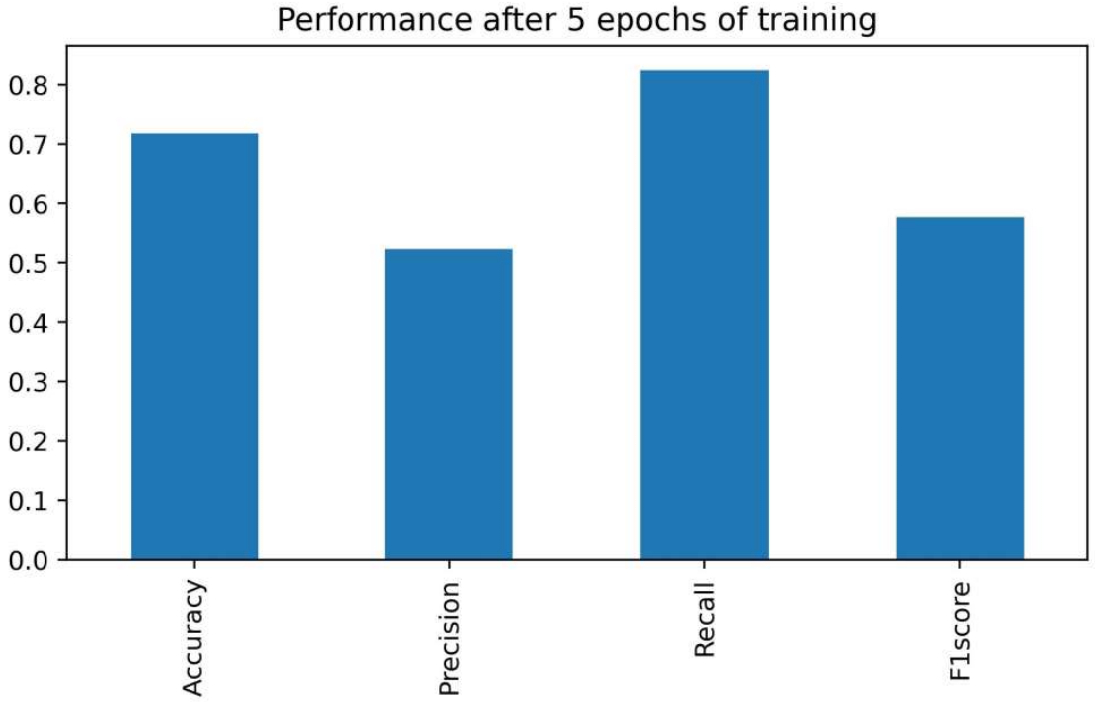


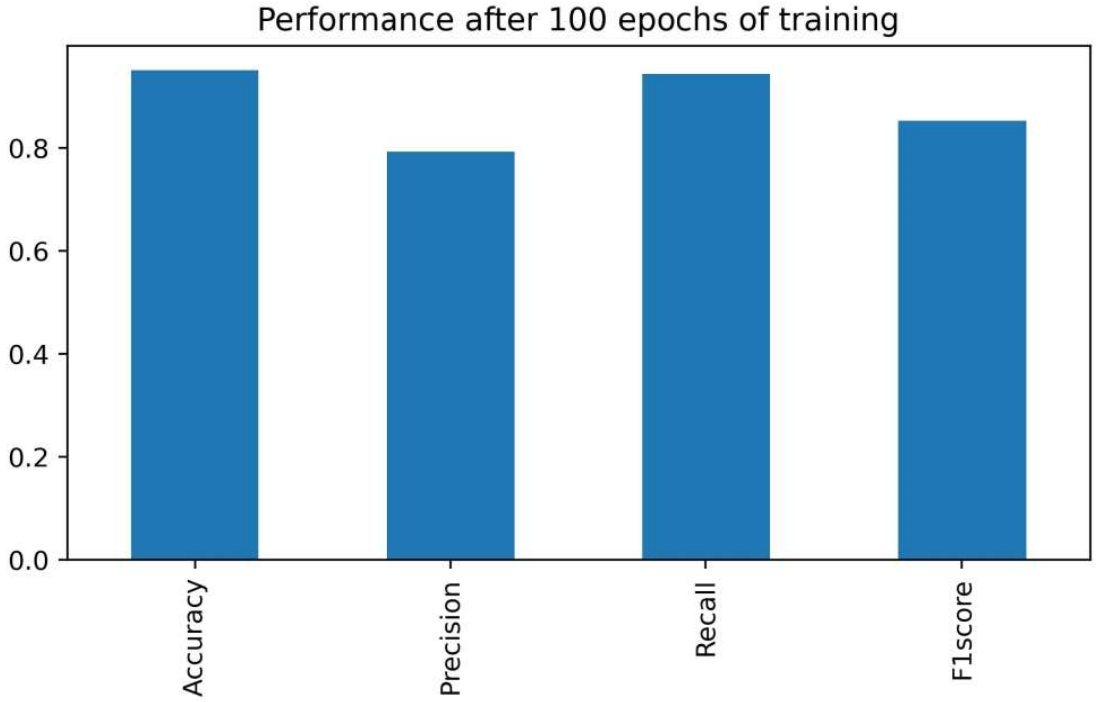


Performance Evaluation - Confusion Matrices

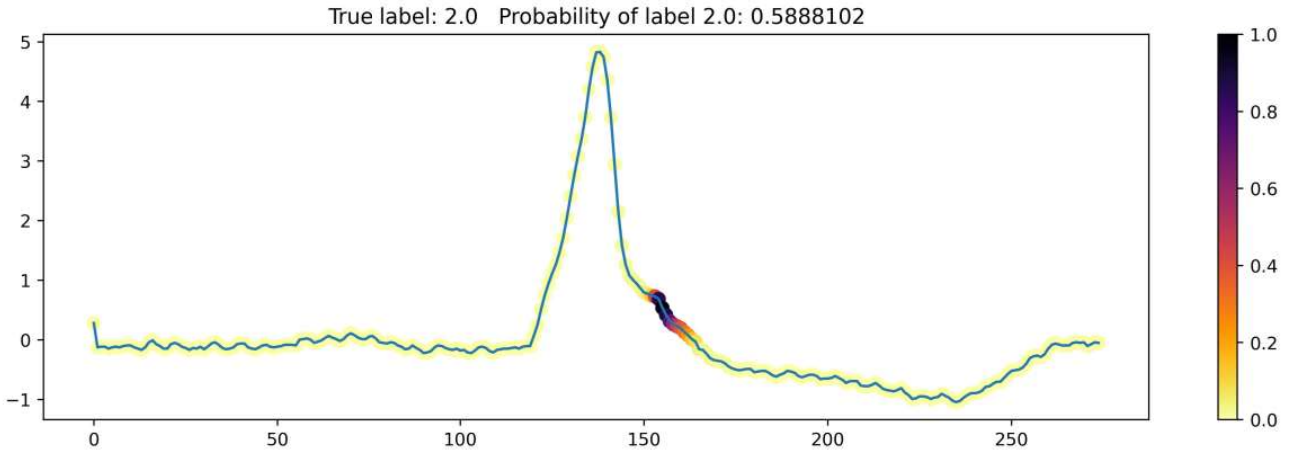


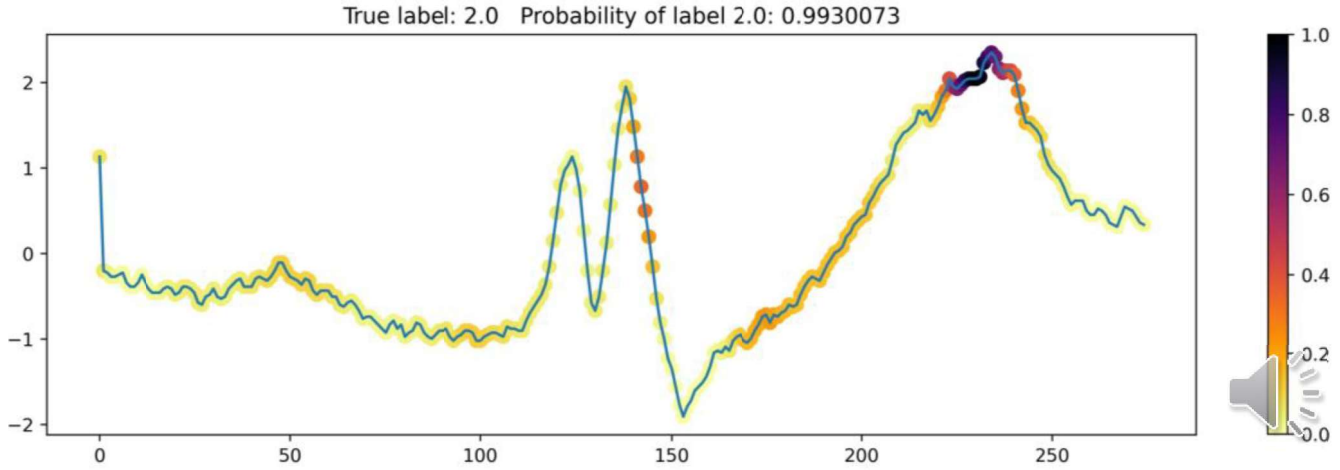
Performance Evaluation



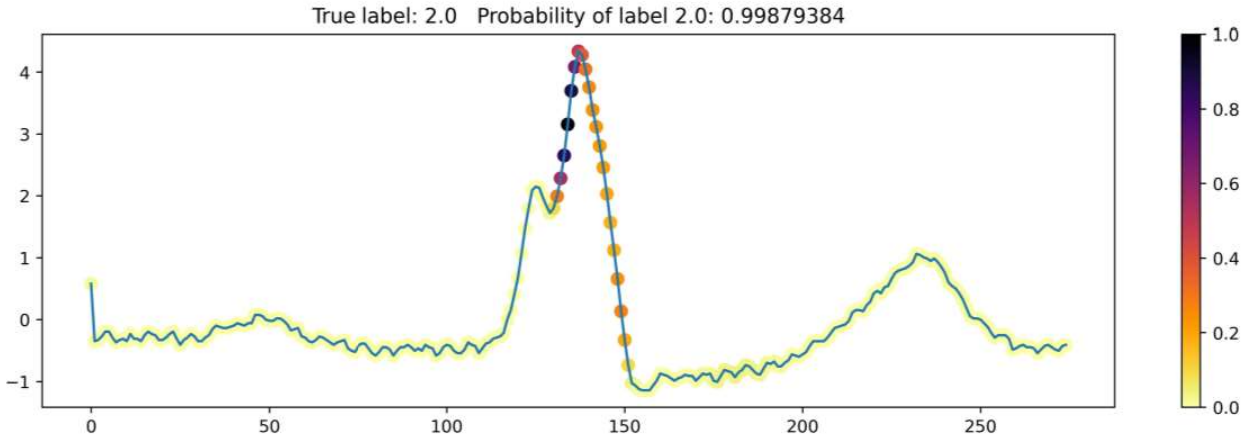


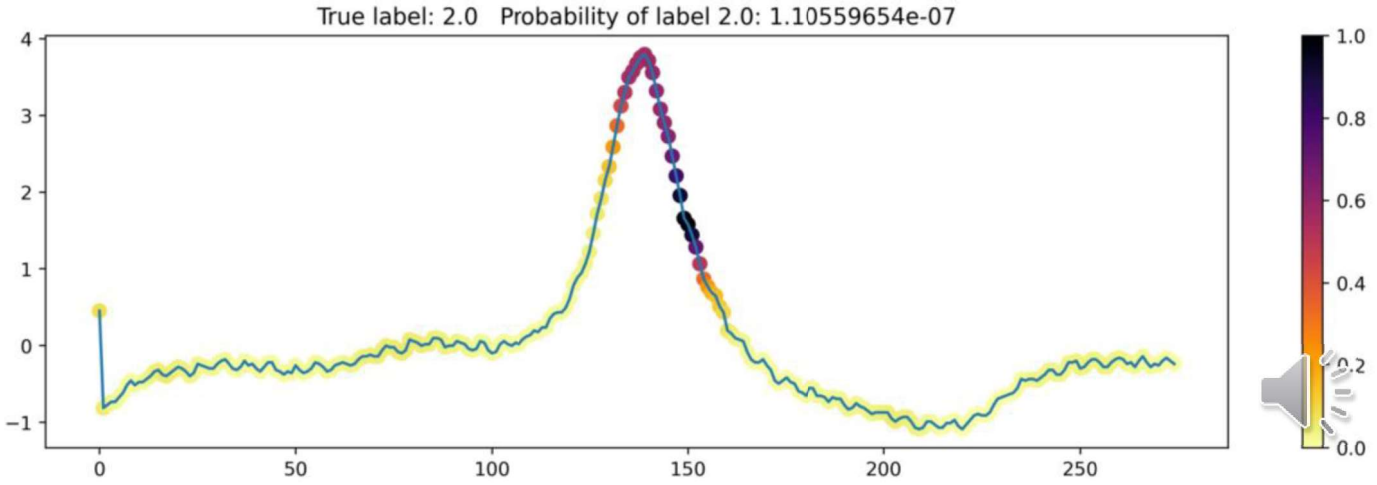
Attention Weights after 5 Epochs



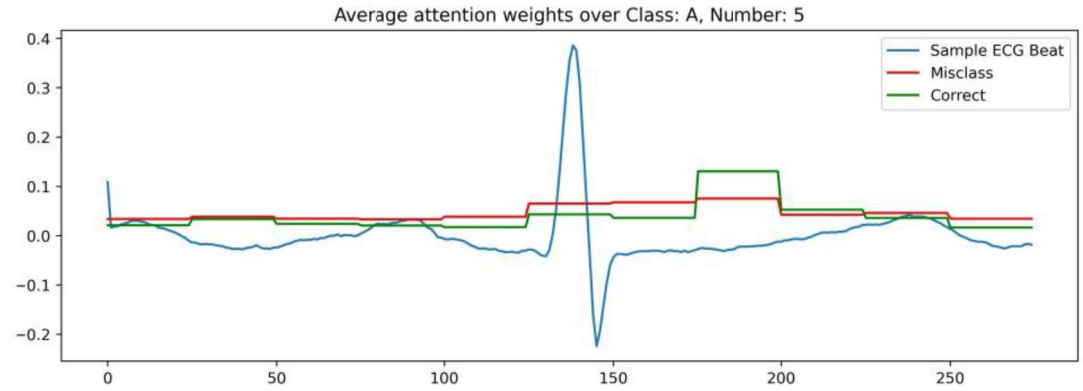


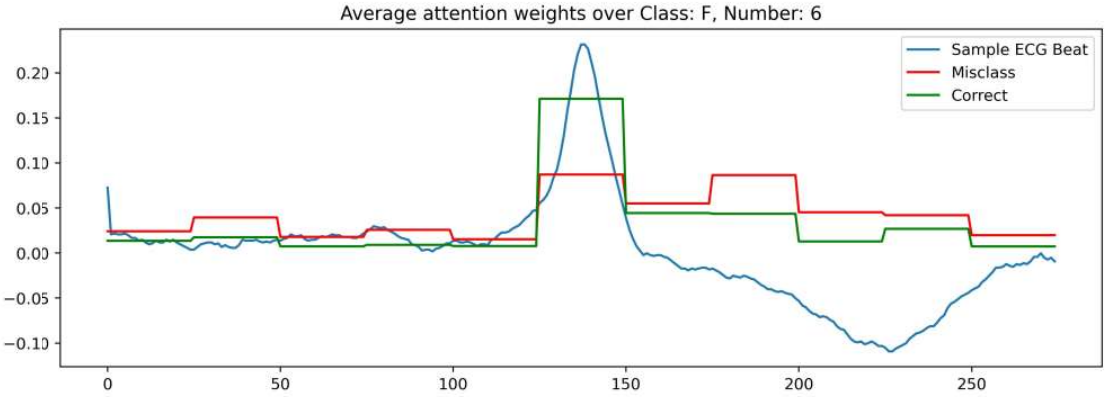
Attention Weights after 100 Epochs

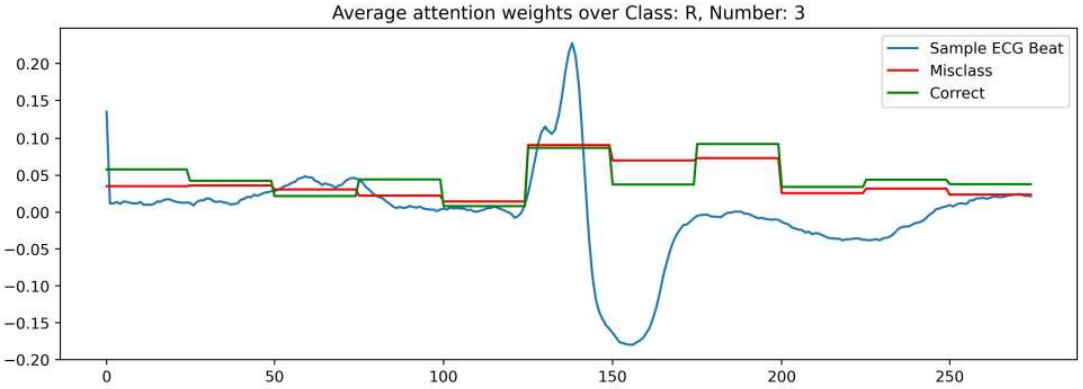


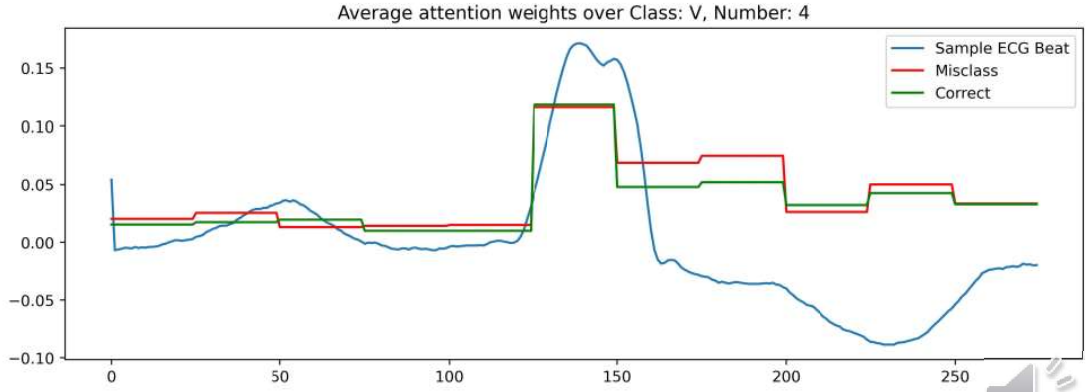


Average Attention Weights









Summary

- Attention has been mostly applied to enhance performance
- Attention weight visualisations offer some degrees of understanding of the deep network decision process
- Attention has been very little studied in terms of explainability

References

- Niu et al. ‘A review on the attention mechanism of deep learning’, Neurocomputing, 2021
- Harerimana et al. ‘A deep attention model to forecast the Length Of Stay and the in-hospital mortality right on admission from ICD codes and demographic data’, Journal of Biomedical Informatics, 2021