DOI: https://doi.org/10.36094/sc.v92.2026.Faithful_Feature_Time_Explanations_for_ICU.Chakraborty.398

ARTICLE

Sci. and Cult. 92 (5–6) : 398-400 (2026)

FAITHFUL FEATURE–TIME EXPLANATIONS FOR ICU MORTALITY PREDICTION USING GRU AND INTEGRATED GRADIENTS

SANDIP CHAKRABORTY*

Deep temporal models can help predict ICU mortality from vital-sign time series, but their limited interpretability reduces clinical trust. This study presents an explainable framework for in-hospital mortality prediction using the PhysioNet/CinC 2012 dataset. The model uses 24 hourly time steps and seven vital signs: HR, SysABP, DiasABP, MAP, RespRate, Temp, and SaO2. A GRU-based classifier was trained on set-a and evaluated on the independent set-b, achieving an AUROC of 0.606 and an AUPRC of 0.223. To interpret the predictions, Integrated Gradients was used to generate feature-time attributions. Explanation faithfulness was assessed through an occlusion-based deletion test, where highly attributed points were masked and compared with random masking. IG-guided masking caused larger drops in prediction confidence, suggesting that the explanations reflect the model's actual decision process. This framework supports auditing and error analysis of ICU risk prediction models.