# 12, 6, 3 and 2-Lead ECG Multi-label Classification using Ensemble Learning

Bjørn-Jostein Singstad, Pål Haugar Brekke

Oslo University Hospital
Oslo, Norway

**Background and aims:** In recent years it has been shown that Artificial Intelligence (AI) can be used to interpret electrocardiograms (ECG). However, it is still unclear whether, or to what extent, the performance of AI-based ECG classification models is affected by reducing the number of leads used by the models. This study aims to develop and compare AI-based ECG classification models based on 12, 6, 3 and 2 leads.

**Methods:** Four ensemble models were used to classify extracted ECG features from 12, 6, 3 and 2 lead ECGs. The models used a combination of Random Forests, Classifier Chains and Label Space Partitioning Classifiers to classify 27 diagnoses. The 12-lead model used 125 features, the 6-lead model used 65 features, the 3-lead model used 35 features and the 2-lead model used 25 features. The performance of each model was evaluated on a development set, consisting of 43,101 ECGs, using 10-folded stratified cross-validation (CV). The models were submitted to the organizers of PhysioNet/CinC Challenge who trained the models on the same development set and applied the models on a subset of a hidden test set.

**Results:** The results obtained by our team, CardiOUS, are summarised in Table 1.

Table 1. Scores obtained in the unofficial phase. The table shows the mean and standard deviation from train and validations (val) scores obtained by the 10-folded cross-validation (CV) performed on the development set. The test scores are from a subset of a hidden test set possessed by the organizers of PhysioNet/CinC Challenge.

| | 12-lead | 6-lead | 3-lead | 2-lead |
|---|---|---|---|---|
| CV (train) | $0.66 \pm 0.002$ | $0.65 \pm 0.001$ | $0.65 \pm 0.001$ | $0.65 \pm 0.002$ |
| CV (val) | $0.15 \pm 0.003$ | $0.15 \pm 0.003$ | $0.15 \pm 0.002$ | $0.15 \pm 0.005$ |
| Test score | 0.43 | 0.42 | 0.42 | 0.41 |

**Conclusion:** The results of this study show that the models did not show any significant increase in performance when adding more ECG leads. This result may be explained by the fact that the scored diagnoses are mainly rhythm disorders and thus several leads may be redundant.