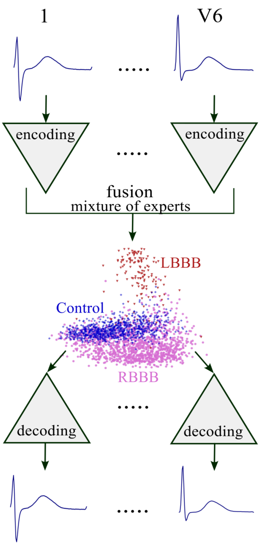
Multimodal learning and fusion from 12-Lead ECG

Anita Salvador*, Badih Ghattas, Nicolas Duchateau

Aix-Marseille University / CREATIS Lyon / IUF - France

Aims: Learning meaningful low-dimensional representations from medical data is essential for disease characterization and interpretability. For ECG signals, standard dimensionality reduction often ignores the structure of 12-lead ECGs, where each lead provides both redundant and complementary information. We specifically aim at quantifying the value of explicitly considering this structure with multimodal variational autoencoders (MVAEs), compared to simpler early or late fusion.



Methods: Each ECG lead was considered a separate modality. Signals were normalized temporally based on the QRS onset and averaged across cycles. We used a MVAE tailored for ECG information fusion (mixture-of-experts, 1D-convolutions) to learn a shared latent space capturing common and complementary information across all leads, while encoding/decoding each lead separately. We evaluated the quality of the learnt representations through their overall organization in the latent space and the separability between controls and known diseases. Comparisons included *early* fusion (single VAE on all leads considered as 12 channels) and *late* fusion (12 VAE on each channel independently followed by mixture-of-experts on the latent distributions).

Results: We considered 2587 patients from the public CPSC Database, within subgroups where the ECG shape is altered differently across each lead (controls [N=917], LBBB [N=158], RBBB [N=1512]). Our multimodal approach learns structured latent representations that reflect both shared and lead-specific information. The organization in the latent space better separates the subgroups (Hotelling T^2-test for controls vs. LBBB: 204.7 [MVAE / intermediate fusion], 181.5 [single VAE / early fusion], 106.6 [12 VAE / late fusion], and controls vs. RBBB: 322.8 [MVAE / intermediate], 166.7 [single VAE / early], 35.0 [12 VAE / late]).

Conclusion: Modeling 12-lead ECGs as a multimodal signal within a MVAE framework enables the learning of richer representations that consider redundancy and complementarity across leads, of high relevance for the characterization of subtle alterations of ECG shape due to disease.