**Classification of 12-lead ECGs using digital biomarkers and drawing from advances in representation learning**

David Assaraf1, Jeremy Levy1,2, Janmajay Singh3, Gal Yefet2, Armand Chocron1,2 and Joachim A. Behar1

1 Faculty of Biomedical Engineering, Technion-IIT, Haifa, Israel

2 Faculty of Electrical Engineering, Technion-IIT, Haifa, Israel

3 Independent scientist.

**Introduction:** The 12-lead electrocardiogram (ECG) is a standard tool used in medical practice for identifying cardiac pathologies. Because the necessary expertise to interpret this tracing is not readily available in all medical institutions or at all in some large areas of developing countries, there is a need to create a data-driven approach that can automatically decrypt the information contained in this multi-channel physiological time series and present it to a medical user in an interpretable manner in order to support clinical decision making.

**Methods:** As part of the unofficial phase of the challenge we developed an initial approach based on physiological features (“digital biomarkers”) used in cardiology. The training dataset consisted of 6,877 12-lead recordings. Two sets of features were engineered: (1) capturing the interval variation between consecutive heartbeats, commonly called heart rate variability (HRV) measures and (2) using morphological biomarkers (e.g. QT interval, QRS width). A total of 32 HRV features and 30 morphological biomarkers were implemented in python. A random forest (RF) model was trained using 6-fold cross validation to optimize the model hyperparameters. Stratification was performed over the classes types. Feature selection was performed using minimum redundancy maximum relevance (MRMR).

**Results and discussion:** A geometric\_mean = 0.47 +/- 0.02 score was obtained on the validation sets and a score of geometric\_mean = 0.52 was obtained on the hidden test set (listed “Technion\_AIMLAB” on the leaderboard). We will further improve the current feature engineering based approach, particularly by augmenting the set of morphological biomarkers. We will also develop a deep learning approach and combine both in a unified ensemble learning model. Part of our work will also focus on making the results of the classification process interpretable for the prospective clinical users.