



## Maksym A. Jopek

Medical University of Gdansk  
Poland

### Optimizing RNA normalization techniques in platelet RNA-based liquid biopsies for improved ovarian cancer detection using machine learning

#### Abstract:

**Introduction:** The reliability of machine-learning (ML) integrated platelet RNA-based diagnostic tests is heavily influenced by the sample processing methods used during model training. In this study, we assessed the impact of three RNA normalization protocols and their impact on classification accuracy.

**Materials and Methods:** Data were collected from two distinct cohorts. The first cohort included publicly available samples (390 healthy donors - HD, 144 ovarian cancer - OC, 1484 samples from patients with 17 other cancer types) from In't Veld et al. 2022 [1] and served as a training and validation dataset. The second cohort consisted of samples from HD (n=7), benign controls - BC (n=25), and OC patients (n=32) served as a test set. Gene counts from the test cohort were normalized using three different approaches: 1) normalizing the test samples independently, 2) normalizing the test samples alongside the training data, and 3) normalizing each test sample separately with the training data.

**Results:** The logistic regression model scored 68.63% sensitivity, at a 100% specificity threshold on the validation set. Using the first normalization method for the test samples, the model scored 66.3% balanced accuracy while classifying the HD, BC, and OC samples. The second approach resulted in 79.9% balanced accuracy. The third approach resulted in the highest balanced accuracy of 81.3%.

**Conclusion:** This study highlights how normalization protocol affects the diagnostic process of RNA sample classification using ML models. Minimizing batch effects by normalizing each test sample with training data separately proved to be the most accurate approach.

#### Biography

**Maksym Jopek** is currently a 3rd year PhD student at the Intercollegiate Faculty of Biotechnology at the University of Gdansk and Medical University of Gdansk, specializing in the Division of Translational Oncology. Their research is focused on the applications of artificial intelligence in liquid biopsy for cancer detection. In addition to their PhD studies, Maksym Jopek consults on medical research projects at the Centre of Biostatistics and Bioinformatics Analysis at the Medical University of Gdansk and works as a bioinformatician at the Clinic of Arterial Hypertension and Diabetology at the University Clinical Centre in Gdansk.