

Unique Protocol ID: 13011998

Brief Title: Volatilome and Single-Lead Electrocardiogram Optimize Ischemic Heart Disease
Diagnosis Using Machine Learning Models

Official Title: Biomarkers of the Exhaled Breath and Single-Lead Electrocardiography in the
Diagnosis of Myocardial Ischemia

NCT number: NCT06181799

Document date: 17.06.2024

Statistical Analysis Plan

The algorithms employ statistical methods including descriptive statistics, Student's t-test, chi-square test, ROC analysis, Shapiro-Wilk test, Mann-Whitney U test, Kruskal-Wallis test, Fisher's test, McNemar test and Lasso regression. Continuous variables express as mean \pm standard deviation for normal distribution or as medians for skewed distribution. Categorical variables are expressed as frequencies and percentages. The Kolmogorov-Smirnov test use to determine whether continuous variables were normally distributed. The relationships between quantitative variables test using Pearson or Spearman correlation, depending on the type of distribution. Analysis of differences in means between continuous variables of two categories perform using the Student's (Welch) t-test or the Mann-Whitney U-test, as appropriate; categorical variables were analyzed using the χ^2 test (or Fisher's exact test if this was not applicable). To assess the relationship between continuous variables of more than two categories, analysis of variance and the least significant difference test in post hoc tests will be used. The Kruskal-Wallis test was use in cases where the variables were skewed. A p-value of <0.05 was considere statistically significant. Multinomial regression logistic model analysis, or multivariate analysis of variance (MANOVA), or analysis of variance (ANOVA), use to determine whether there any statistically significant differences between two or more groups on a set of continuous dependent variables, as appropriate. For quantitative indicators, the distribution pattern (using the Shapiro-Wilk test), mean, standard deviation, median, interquartile range, 95% confidence interval, minimum and maximum values were determined. For categorical and qualitative features, the proportion and absolute number of values determine. Statistical processing performe using the R v4.2 programming language, Python v.3.10 [^R], Statistica 12 program. (StatSoft, Inc. (2014). STATISTICA (data analysis software system), version 12. www.statsoft.com.), and IBM SPSS Statistics version 28.0.1.1 (14), 2021, USA. And different models of machine learning model to assess the diagnostic accuracy of the volatile organic compound and the single lead electrocardiography in the diagnosis of ischemic heart disease.