

A. Czajkowska¹, K. Pietrowska¹, M. Czajkowski², J.B. Godzien¹, K. Jurczuk², P. Krasnicki³, E.T. Grochowski³, M. Ciborowski¹, Z. Mariak³, M. Kretowski², A. Kretowski^{1,4}, D.A. Dmuchowska³

1) Clinical Research Centre, Medical University of Białystok, Poland; 2) Faculty of Computer Science, Białystok University of Technology, Poland; 3) Department of Ophthalmology, Medical University of Białystok, Poland; 4) Department of Endocrinology, Diabetology and Internal Medicine, Medical University of Białystok, Poland. anna.czajkowska1@umb.edu.pl

INTRODUCTION

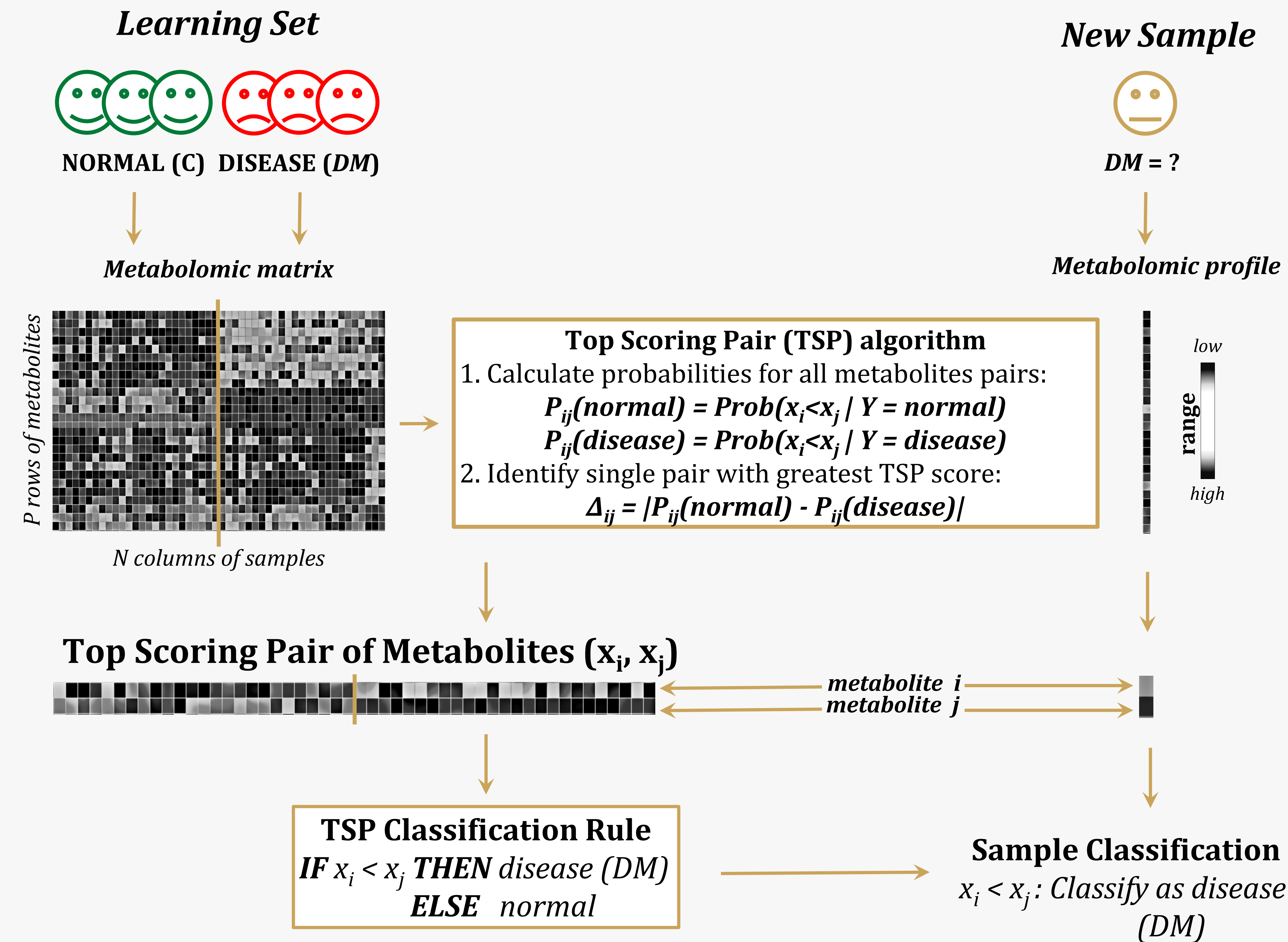
Cataract, an ocular lens opacity associated with a decrease in visual acuity, is a major cause of vision impairment and blindness worldwide. Epidemiological studies demonstrated that patients with diabetes mellitus are up to five times more likely to develop cataract than non-diabetic individuals. Growing evidence suggests that a longer duration and poor metabolic control of diabetes are the main risk factors for cataract formation¹. We tackle this problem with a novel machine learning approach as a complement to statistical analysis.

METHODS

Aqueous humour samples collected during cataract surgeries from 16 patients with diabetes and 19 non-diabetic controls were analyzed with LC-QTOF-MS (in both ion modes) using hydrophilic interactions (HILIC) and reversed-phase (RP) chromatography².

A novel machine learning classifier, known as Relative Dependency Analysis³ (RDA) was applied to identify discriminative patterns in the data. The algorithm induces a decision tree with splitting nodes that consider relative fraction changes between metabolites in each patient. RDA appeared insensitive to the commonly used data normalization and standardization procedures, and rules generated by this classifier based on relations within a small group of compounds can serve as biological switches.

RELATIVE DEPENDENCY ANALYSIS



For example, for HILIC(-) dataset, a simple rule perfectly classifies C & DM patients:

$$208.05827@5.9131513 > 10 * 459.1847@7.505553$$

(Methyltetrahydrofolic acid)

FALSE

TRUE

NORMAL (C)

DISEASE (DM)

RESULTS & CONCLUSIONS

The RDA analysis included all data sets: HILIC(+), HILIC(-), RP(+), and RP(-). Performed tests showed that RDA was capable of identifying patterns composed of the rules with 2 to 6 metabolic features with the accuracy of 91.4%, 100%, 97.1%, and 94.2%, respectively, each of the data sets. The rules used by RDA, included new metabolic features, previously identified as non-significant on univariate or multivariate analyses used commonly for metabolomics data. Upon their identification, these new signals might improve our knowledge of the relationship between diabetes and increased risk of cataract development. The bioinformatical approach used in this study could minimize the influence of interindividual metabolic differences on the statistical outcome. Application of RDA to metabolomics data could add considerably to the results of a typical statistical analysis, identifying additional relationships between metabolites that play a role in the pathogenesis of a given medical condition.

REFERENCES

- [1] Kiziltoprak H. et al. Cataract in diabetes mellitus. World J Diabetes. 2019,15;10(3):140-153.
- [2] Pietrowska K. et al. An exploratory LC-MS-based metabolomics study reveals differences in aqueous humor composition between diabetic and non-diabetic patients with cataract. Electrophoresis, 2018,39:1233-1240.
- [3] Czajkowski M. et al. Tree Based Advanced Relative Expression Analysis. Lecture Notes in Computer Science, 2020,12139:496-510.