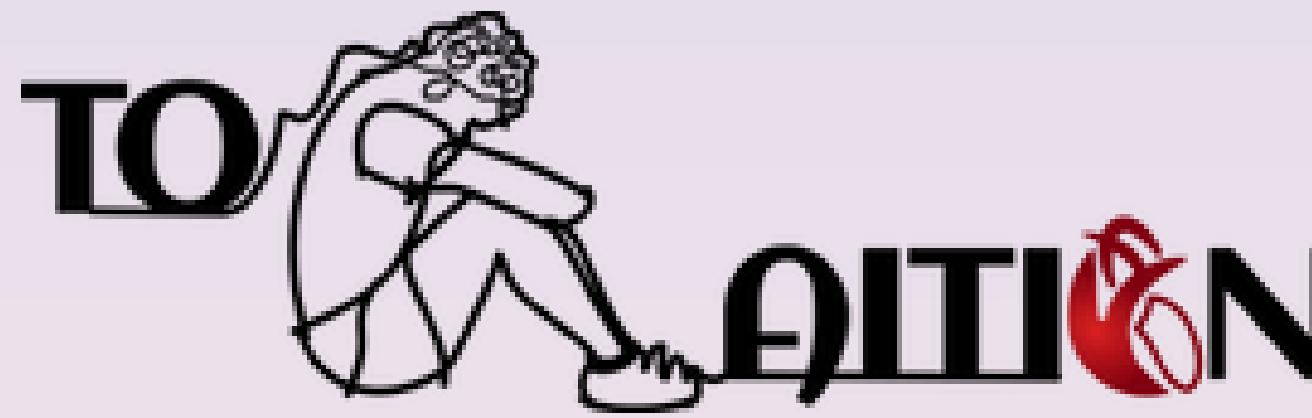


AI-based diagnosis of depression and cardiovascular disease comorbidity based on big data



V. Pezoulas¹, K. Dobretz², G. Ehret², D. Fotiadis¹, A. Sakellarios^{1,3,*}

¹Dept. of Materials Science and Engineering, University of Ioannina, Ioannina, Greece, ²Department of Internal Medicine Hôpitaux Universitaires de Genève, Genève, Switzerland, ³Dept. of Mechanical and Aeronautics Engineering, University of Patras, Patras, Greece

EPP019

INTRODUCTION

- Cardiovascular disease is a hazardous disorder for humans, accounting for about 17.9 million deaths every year worldwide, [1].
- One in three patients with cardiovascular disease (CVD) suffers from depression.
- CVD and depression develop in the same at-risk population.

PURPOSE

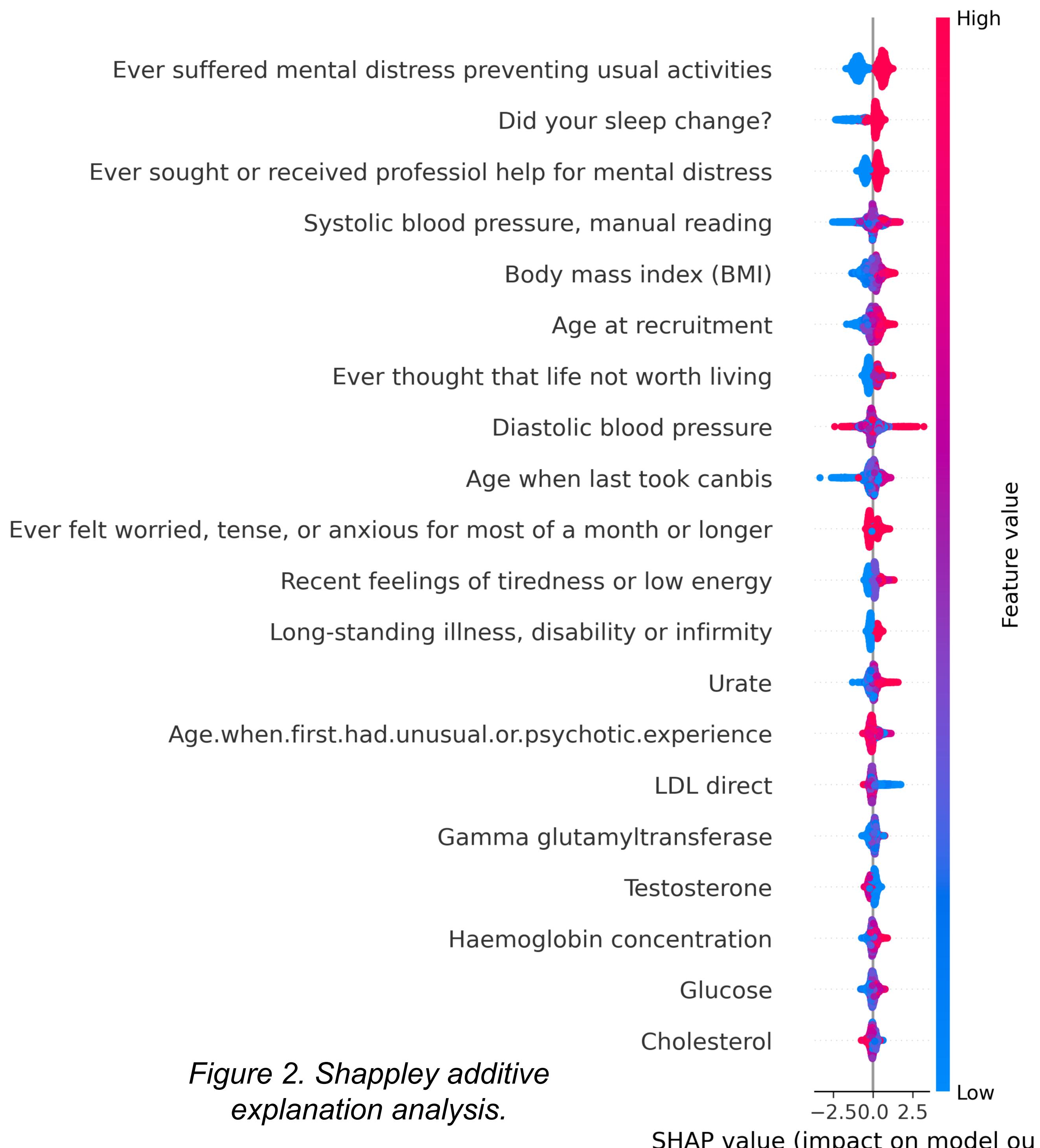
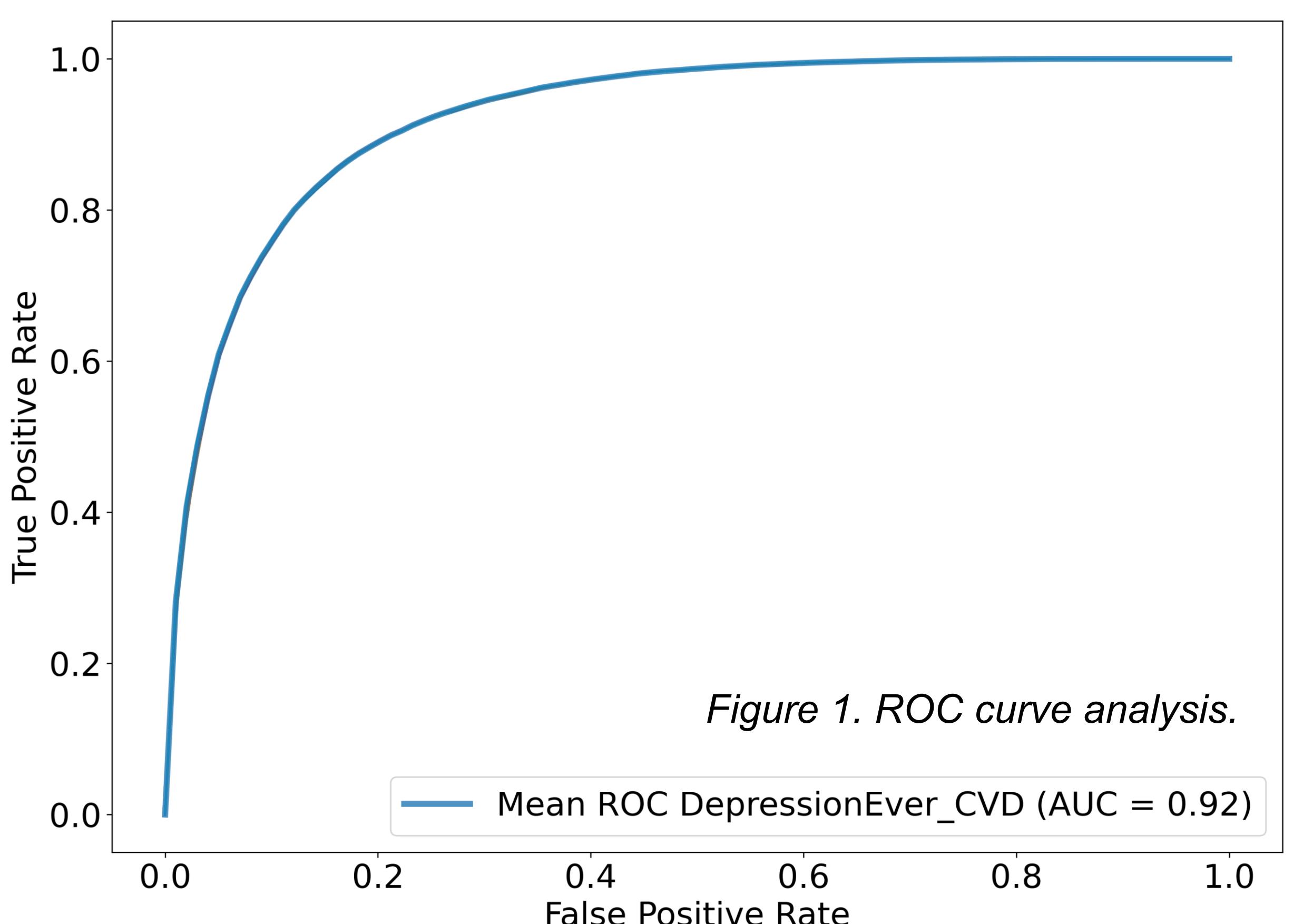
- The aim of this work is to diagnose depression and CVD with high accuracy using only simple biomarkers, which enable the applicability of the model easily in the clinical practice.
- To compare modern (e.g. XGBoost) and standard algorithmic models (e.g. LR).
- To provide AI-based explainable results to clinicians.

METHODS

- Data from 502,379 participants in the UK Biobank were utilized in this work.
- A subset of the participants has a mental assessment using questionnaires about the presence of depression. CVD assessment was also available for the majority of the patients.
- 126,033 participants had clinical assessment of both depression and CVD. From these, 8,925 patients had both comorbid conditions.
- A data curation pipeline was applied to automatically remove outliers and duplicated fields in the input dataset.
- An AI-empowered pipeline was developed to classify patients at higher risk for depression and CVD.
- Hybrid boosting ensembles, including the XGBoost algorithm with a customized hybrid loss function was trained on the curated data, to reduce training and testing loss and to avoid overfitting effects.
- Down-sampling method was applied producing a new balanced 1:1 ratio dataset.
- The AdaBoost (adaptive boosting), Random Forests and Extreme Gradient boosting trees (XGBoost) were applied.
- A nested cross-validation process was applied for hyperparameter optimization (3-fold cross validation) and model validation (10-fold cross validation).
- Advanced feature selection methods, including the Shapley additive explanation (SHAP) exploratory analysis were utilized to identify predictors with positive or negative impact to have both the comorbid conditions.

RESULTS

- The XGBoost classifier had the best performance among all tested classifiers.
- The results were 0.85, 0.88, 0.81 and 0.92 for the accuracy, sensitivity, specificity and AUC, respectively.
- Several social questionnaires are highly ranked for the diagnosis.
- BMI and age are important factors for the diagnosis.
- Blood biomarkers are also included in the model for diagnosis of CVD and depression.



CONCLUSIONS

- Monitoring patients with a history of cardiovascular disease is very important for the prediction of depression.
- Depression and CVD can be predicted with 85% accuracy.

REFERENCES

- [1] World Health Organization. (2017). Cardiovascular Diseases (CVDs). [Online]. Available online: <https://www.who.int/health-topics/cardiovasculardiseases/>
- [2] Ferrucci L, Fabbri E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat Rev Cardiol.* 2018 Sep;15(9):505-522.
- [3] Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. *Eur Heart J.* 2006 Dec;27(23):2763-74.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 848146, as part of TO_AITION project.

This paper has been financed by the funding programme "MEDICUS", of the University of Patras

CONTACT INFORMATION
Email: ansakel13@gmail.com