MACHINE LEARNING-BASED PREDICTION OF MORTALITY AND RISK FACTORS IN PATIENTS WITH CHRONIC KIDNEY DISEASE DEVELOPED WITH DATA FROM 10000 PATIENTS OVER 11 YEARS

Jose Ibeas[1], Oscar Galles[2], Nuria Monill[1], Edwar Macias[2], Antoni Morell[2], Javier Serrano[2], Dolores Rexachs[2], Jose Vicario[2], Jordi Cokas[3] and Elisenda Martinez[4]

[1]Clinical, Interventional and Computational Nephrology group (CICN) of the Parc Taulí Research and Innovation Institute (I3PT), Barcelona, Spain, Nephrology. Parc Tauli University Hospital, Sabadell, Barcelona., Sabadell, Spain, [2]School of Engineering of Autonomous University of Barcelona, Spain, Barcelona, Spain, [3]The Register of Kidney Patients of Catalonia (RMRC)—Catalan Transplant Organization, OCATT., Barcelona, Spain and [4]Catalan Agency for Health Quality and Evaluation (AQuAS), Barcelona, Spain

BACKGROUND AND AIMS: Around the globe, over 850 million patients suffer from chronic kidney disease (CKD). These have associated with high mortality rates, in particular when undergoing renal replacement therapies (RRT) such as dialysis, reaching up to 10% a year, and therefore, are considered of a fragile status. CKD is also associated with cardiovascular complications that can cause mutual aggravation. Available clinical guidelines identify certain risk factors and predictive models, but those have not been tested and validated successfully for renal patients and consequently, there is for the identification of predictive factors and the prediction of mortality. This is caused by the limitations of current methodologies and statistics: current models simplify complex relationships by assuming a linear relation between risk factors and certain events, and so, there is a need for a new approach. Over the last years, a rise in Artificial Intelligence and Machine Learning has been seen, presenting an alternative for the first time. This project aimed to study the performance of different ML algorithms for the prediction of mortality and the identification of risk factors for CKD patients.

METHOD: Design: Retrospective analysis of a historical cohort from the Register of Renal Patients of Catalonia (RMRC) and the Catalan Agency for Health Quality and Evaluation. Group of 10 473 patients with CKD stages from first to RRT. Follow-up of 11 years, from January 2010 to December 2020. Inclusion criteria: >18 years. Training of an Extreme Gradient Boosting model, and comparison with other algorithms for the prediction of mortality at different times, using different follow-up periods for each patient.

Methodology:
Variables: i) Age, gender, body mass index, time for death (9), ii) Diagnoses (ICD-9/10) (26); iii) Laboratory variables (37) iv) All pharmacological treatments (46). For all executions, data was balanced using the SMOTETomek technique.
Analysis:

Table 1.

Metric Algorithm	Accuracy	Sensibility	Specificity	PPV	NPV	AUC
XGBClassifier	0.805	0.817	0.793	0.818	0.791	0.886
LGBMClassifier	0.815	**0.822**	0.810	0.825	**0.807**	0.898
CatBoostClassifier	**0.820**	0.809	**0.833**	**0.838**	0.803	**0.905**

REFERENCES

1. Pre-processing of the data collected and evaluation of the historical cohort for treatment of missings and data errors.
2. Exploration of variables. Training and testing of different variable groups to establish a balance between the number of patients and the information available
3. Use of multiple types of Machine Learning classifier algorithms: i) Extreme Gradient Boosting ii) Light Gradient Boosted Machines iii) CatBoost Classifier.
4. Samples randomization and data separation with a ratio of 80–20% for training and testing.
5. Comparison of the metrics obtained for each algorithm with a 5-iteration Cross-Validation system

RESULTS: The patient sample presented a mean of 68.2 ± 12.9 years and 65.8% were female and 34.2% male. Different follow-up and time windows were tested and the best results were obtained when using a 2-year period follow-up and a 4-year mortality prediction. The Area Under the Curve values obtained for each model were: XBGClassifier (0.89), LGBM Classifier (0.90), CatBoost Classifier (0.91). The 10 variables with major relevance according to the XBGClassifier (54.65% of the total weight of the 71 variables) and in this order, are cardiopathy, advanced chronic kidney disease, vasculopathy, age, neoplasia, transplant, digestive pathology, estimated glomerular filtration rate, high blood pressure. The results presented in *Figure* and *Table* correspond to the mean obtained for the 5-folds of the Cross-Validation.

CONCLUSION: Machine Learning techniques suppose an alternative to classical statistical methods, with a high predictive capacity for mortality. The possibility of generating algorithms with real-world data can allow the individualization of the mortality risk as well as the predictive factors.
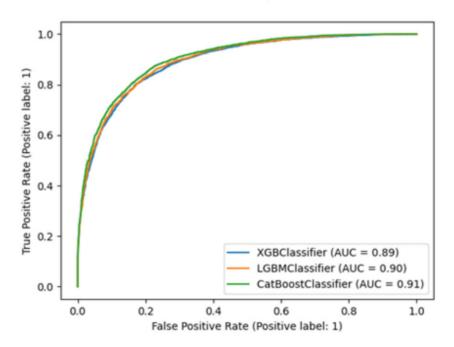
LONG-TERM RENAL OUTCOMES AFTER ST-ELEVATION MYOCARDIAL INFARCTION TREATED BY PERCUTANEAOUS CORONARY INTERVENTION

Marie Legras[1,2], Camille Amaz[3], Thomas Bochaton[2,4], Nathan Mewton[3,4], Christophe Soulage[2] and Fitsum Guebre-Egziabher[1,2]

[1]Hopital Edouard Herriot, Néphrologie, Lyon, France, [2]Laboratoire CarMeN, Institution National de la Santé et de la Recherche Médicale 1060 Université Lyon

ROC curve comparison