Hospital AlmaMáter | GHIPS Gestión Hospitalaria | UNIVERSIDAD DE ANTIOQUIA Facultad de Medicina

Prediction of the risk of adverse clinical outcomes with machine learning techniques in patients with chronic no communicable diseases

Alejandro Hernández Arango [1,2]
María Isabel Arias [2,3]
Viviana Pérez [2]
Luis Daniel Chavarría [2,4]
Fabian A. Jaimes B. [5]

[1] University of Antioquia, School of Medicine, Department of Internal Medicine, Internist Master's student in Telehealth and Digital Medicine, Medellín, Colombia.

[2] Alma Mater Hospital of Antioquia, University of Antioquia. Medellín, Colombia.

[3] Living Lab Health Information Systems Professional. Medellín, Colombia.

[4] Data Scientist, National University. Medellín, Colombia.

[5] Professor, Department of Internal Medicine, Universidad de Antioquia, Medellín, Colombia.

Background

Growing demand for healthcare services due to chronic non-communicable diseases challenges health systems and generates high costs in Colombia and in the world. Risk stratification allows for the optimal allocation of finite resources and coordination of care levels based on risk profiles in chronic diseases. Previous work in this cohort validated a functional classification.

Clinical decision support (CDS) systems embedded in electronic health records can enhance medical decisions, streamline work, and improve patient and population outcomes.

Objective

Develop a real-time risk prediction methodology for adverse clinical outcomes using machine learning techniques and big data analysis in patients with chronic non-communicable diseases.

Create a prescriptive analytics dashboard as a clinical decision support system, allowing real-time interaction with predictions based on the clinical and epidemiological characteristics of the patients in the cohort.

Data Source

Retrospective cohort study on electronic health records (EHRs) from April 1, 2017, to December 31, 2020. Data collected and managed within the hospital, ensuring patient privacy and security.

Participants

Conducted in a high-complexity hospital complex in Medellín, Antioquia (Hospital Alma Máter de Antioquia) with ambulatory, in hospital and domiciliary data. Inclusion criteria: ≥18 years old and presenting at least one chronic disease according to ICD-10. Exclusion criteria: Patients with no clinical data due to absence of medical care or loss of follow-up.

Outcomes

In-hospital and out-of-hospital mortality. Hospitalization. Emergency consultations in the reference hospital.

Predictive Models

XGBoost and Elastic Net Regression were used to predict the three outcomes. Included 164 predictor variables, extracted from clinical data, laboratory results, and billing information.

Statistical Analysis

Data imputation: K-Nearest Neighbors (KNN) algorithm for variables with <20% missing data. Data preprocessing: Centering and scaling numerical variables, creating dummy variables for nominal variables. Models: Elastic Net Regression (combines L1 and L2 regularization) and XGBoost (gradient boosting with decision trees).

Model performance: AUCROC, sensitivity, specificity, PPV, NPV, calibration curves, 95% confidence intervals.

Results

Models compared using DeLong test for AUCROC differences. Risk groups created using Tableau software, allowing visualization and filtering of patients based on their predicted risk. (Data in images)

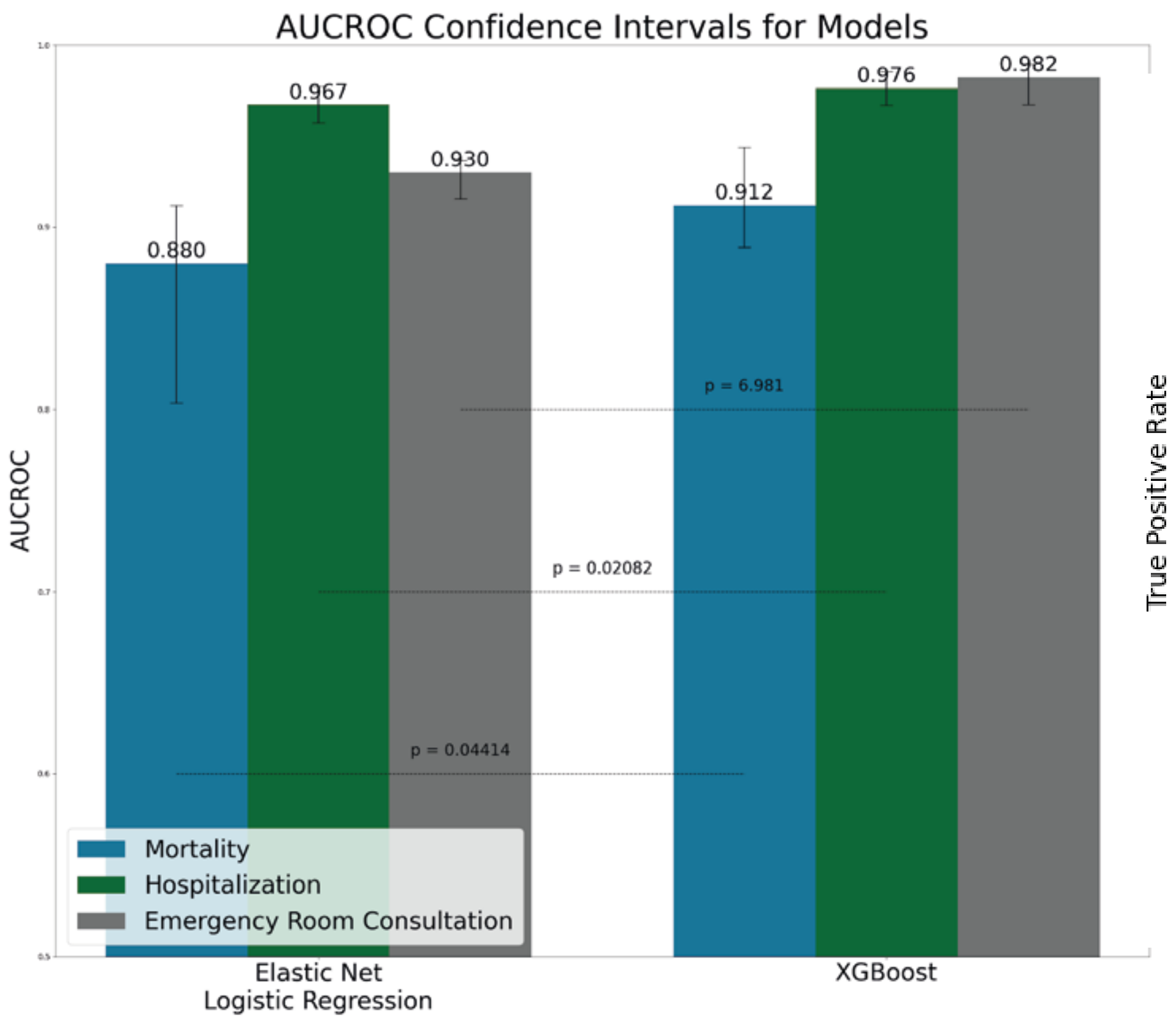
http://medicina.udea.edu.co
/Medicina_udea
@medicina_udea

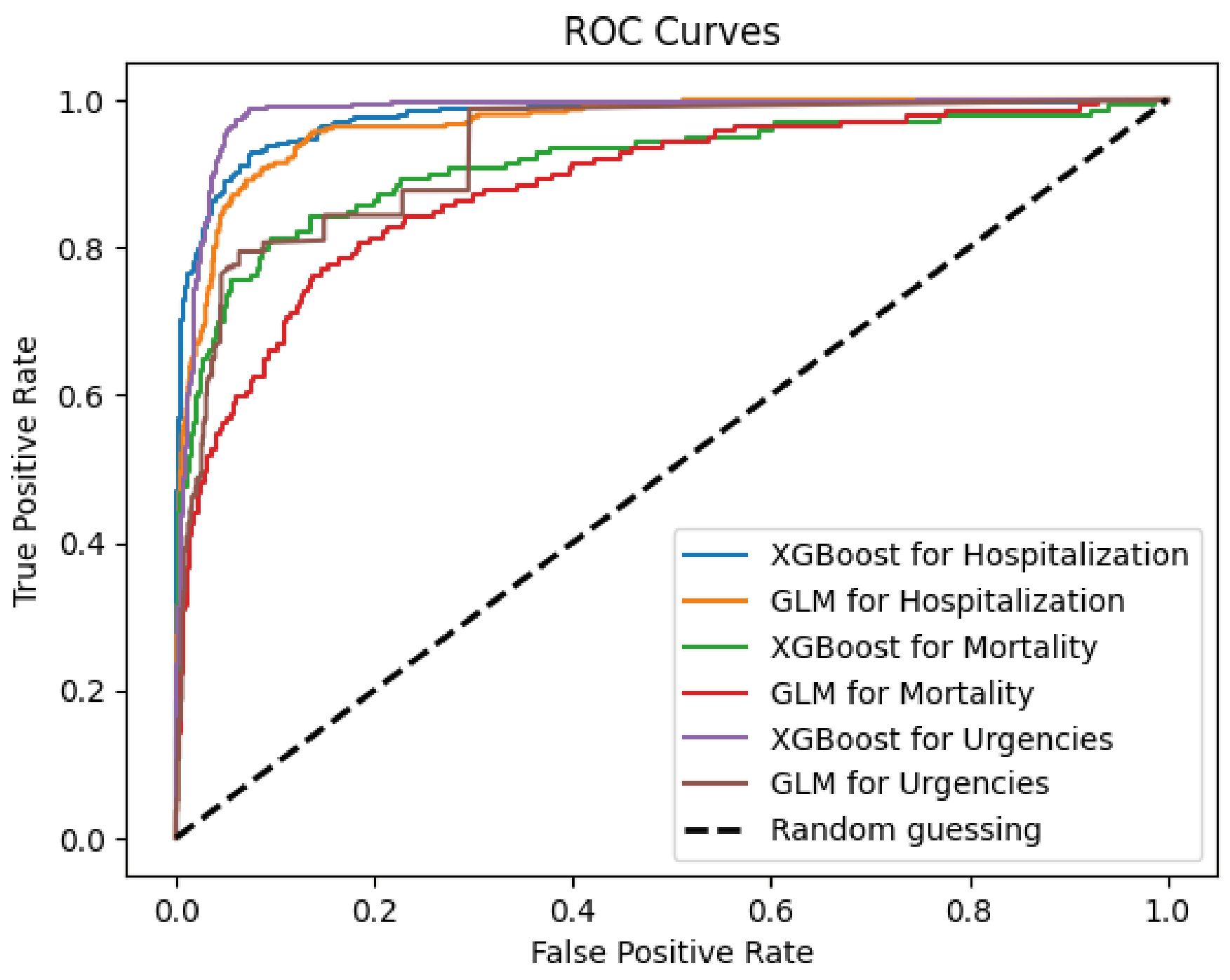
Conclusions

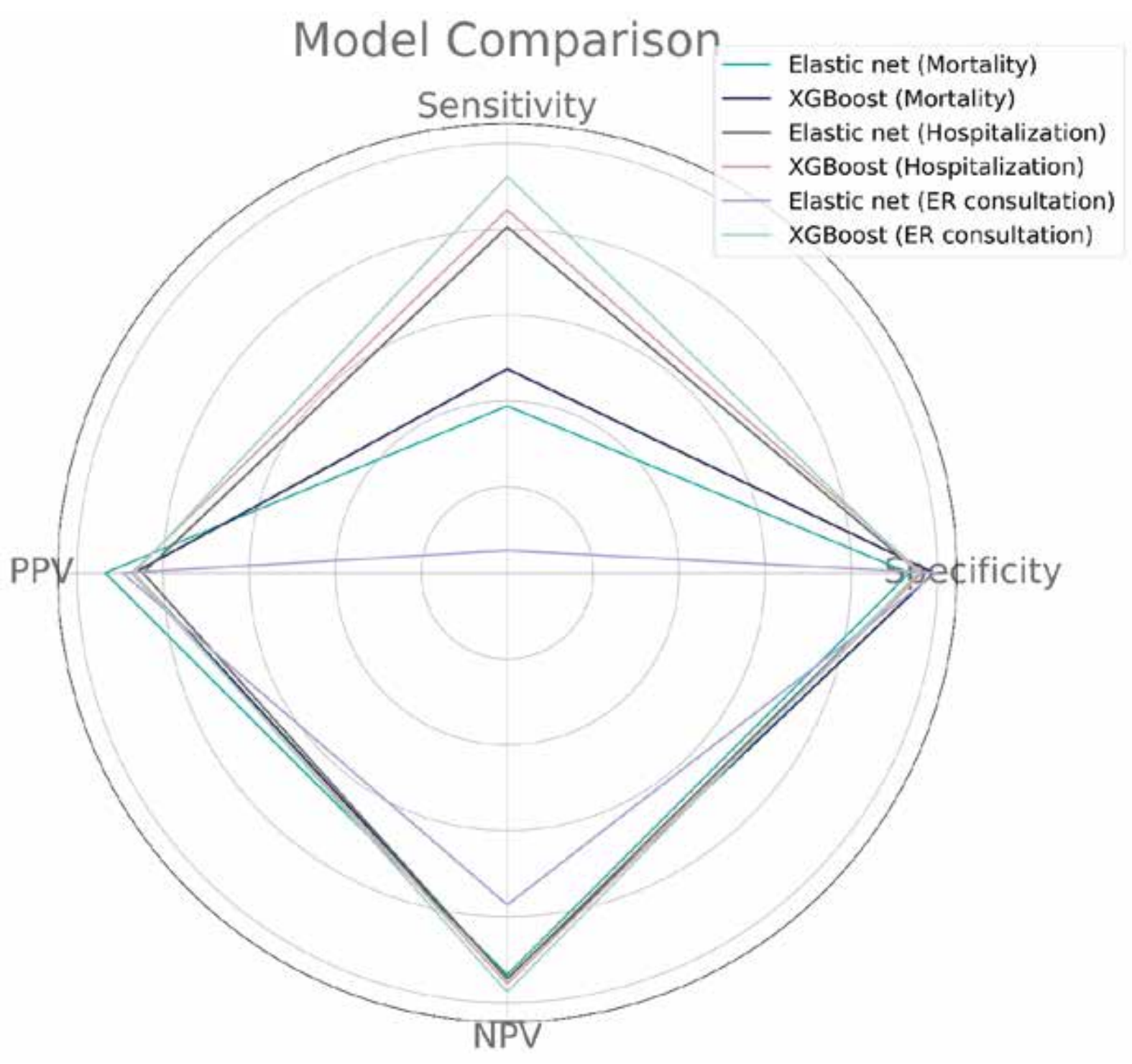
XGBoost model outperformed Elastic Net logistic regression in predicting mortality and hospitalization with statistical significance, while both models performed similarly for emergency room visits. Overall, the XGBoost model has the potential to be a tool for building clinical decision support systems that serve as useful prognostic models for decision-making in patients with chronic non-communicable diseases, based on an easy-to-use and visualize interface. Such tools should be evaluated and validated in future experimental intervention or prospective observational studies for safe implementation in clinical workflows

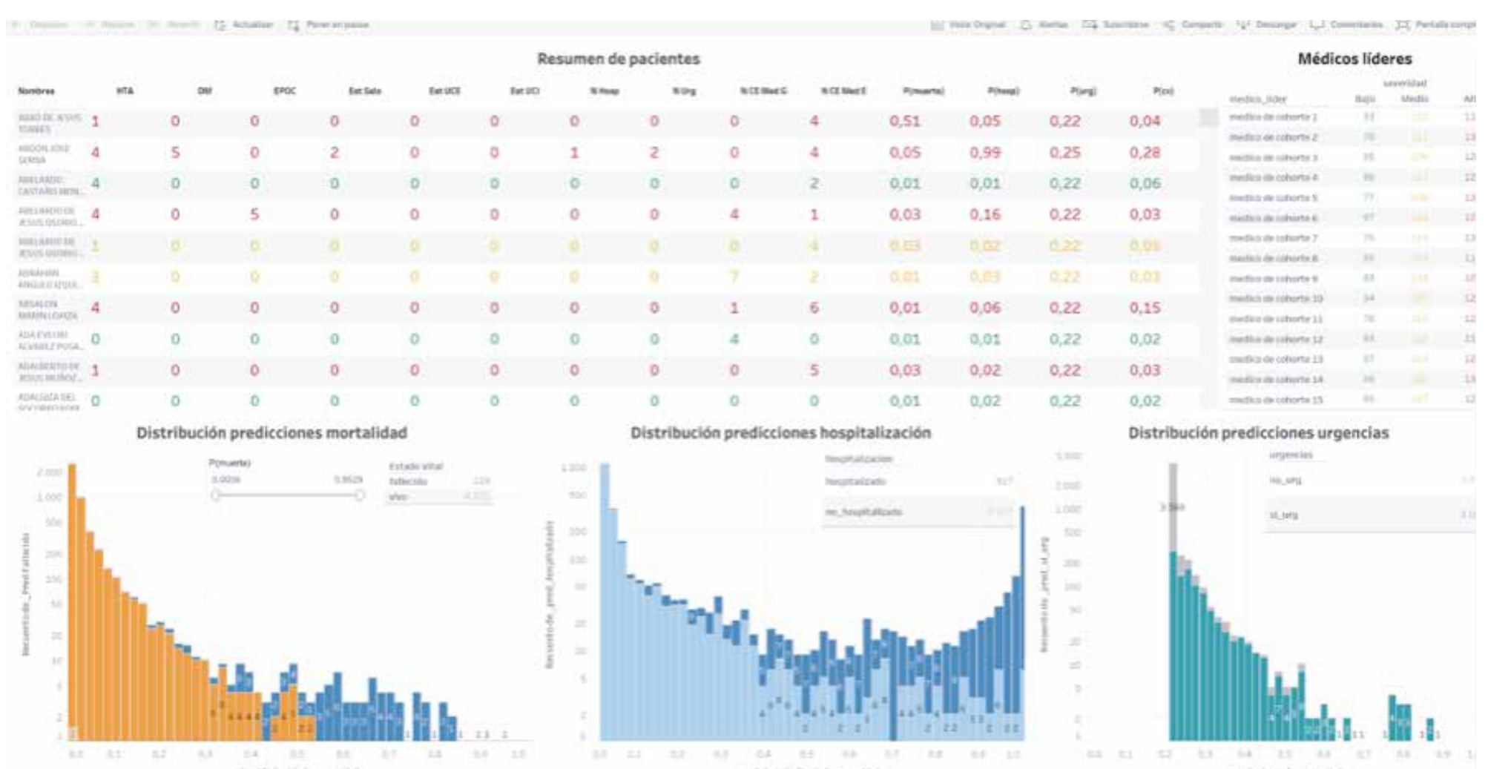
Note

Models compared using DeLong test for AUCROC differences. Risk groups created using Tableau software, allowing visualization and filtering of patients based on their predicted risk. (Data in images)









Resumen de pacientes
Médicos líderes
Distribución predicciones mortalidad
Distribución predicciones hospitalización
Distribución predicciones urgencias

References

Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligencie: DECIDE-AI. Nat Med 2022 May;28(5):924–33. Available from: http://dx.doi.org/10.1038/s41591-022-01772-9

Forrest IS, Petrazzini BO, Duffy Á, Park JK, Marquez-Luna C, Jordan DM, et al. Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. Lancet [Internet]. 2023 Jan 21;401(10372):215–25. Available from: http://dx.doi.org/10.1016/S0140-6736(22)02079-7

García-Arango V, Osorio-Ciro J, Aguirre-Acevedo D, Vanegas-Vargas C, Clavijo-Usuga C, Gallo-Villegas J. Validación predictiva de un método de clasificación funcional en adultos mayores. Rev Panam Salud Publica 2021 Apr 30;45:e15.

MacKay EJ, Stubna MD, Chivers C, Draugelis ME, Hanson WJ, Desai ND, et al. Application of machine learning approaches to administrative claims data to predict clinical outcomes in medical and surgical patient populations. PLoS One [Internet]. 2021 Jun 3;16(6):e0252585. Available from: http://dx.doi.org/10.1371/journal.pone.0252585