Predictive analysis of hospitalisations due to sickle cell disease using machine learning in Bahia

P2-J18

Cristina Canavarro Rodrigues Fuezi Britto[1,3], Jessidenes Teixeira de Freitas Mendes Leal[2,3], Ana Paula Pitanga Barbuda Prates[2,3], Brenda Lima Boaventura[2], Kezia Alves Mustafa[2], Stela Gleide Oliveira Santana Dilago[2,3], Ricardo Cardoso dos Santos[2,3], Paulo Jorge de Souza Vianna[2,3], Marli Simões de Araújo[3]

[1] University of São Paulo (USP), SP, Brazil, [2] Federal University of Bahia (UFBA), BA, Brazil, [3] Health Situation Room of Municipal Health Department of Salvador, BA, Brazil

The best-performing Machine Learning models were *Neural Networks* and *XGBoost*. *Race/ethnicity*, *Hydroxyurea* use, and *care type* were key predictors, indicating a strong link between race and hospitalization risk, and highlighting Hydroxyurea's effectiveness in reducing admissions.

BACKGROUND

Sickle Cell Disease (SCD) is a genetic and hereditary condition that affects the shape of red blood cells, leading to serious health problems such as pain crises, hospitalizations, and, in severe cases, death. Identifying variables that influence the risks of hospitalization, as well as predicting these outcomes, is crucial for providing anticipatory care and improving patients' quality of life. This study aims to develop and compare Machine Learning (ML) models to predict the risk of hospitalizations due to SCD and to assess the impact of Hydroxyurea use on reducing this outcome, using outpatient data from Bahia.

METHODS

Secondary data from patients with SCD residing in Bahia were analyzed, obtained from the Outpatient Information System (SIA) and the Hospital Information System (SIH), both part of the Brazilian Unified Health System (SUS). The data were integrated through probabilistic linkage using a composite key of anomalous data. To predict the risk of hospitalization, five ML models were compared: logistic regression, neural networks, random forest, decision tree, and Extreme Gradient Boosting (XGBoost). Hydroxyurea use was the primary predictor considered, while the dependent variable was the occurrence of hospitalization. Additionally, for the variables of interest, the odds ratio of hospitalizations based on individual characteristics was calculated.

RESULTS

Patients who used Hydroxyurea showed a lower risk of hospitalization. The ML models that showed the best performance according to the evaluated metrics were neural networks and XGBoost. Race/color, Hydroxyurea use, and the nature of care were the most important predictors for the models. This suggests that the race of the patient has a strong association with the risk of hospitalization due to SCD, as well as the effectiveness of Hydroxyurea in reducing hospitalization.

Performance Metrics for Predictive Models.

Algorithms	Accuracy	Accuracy Balanced	Log Loss	AUC*
XGBoost	0,716	0,697	0,530	0,790
Random Forest	0,715	0,698	0,550	0,790
Decision Tree	0,715	0,698	0,600	0,790
Logistic Regression	0,713	0,697	0,540	0,780
Neural Networks	0,720	0,712		

Data source: Original research.
*Area Under the Curve.

RESULTS CONTINUED

Importance of Predictors in Predictive Models

Algorithms	XGBoost	Random Forest	Decision Tree	Logistic Regression	Neural Networks
1º	Race/color	Character of Service	Character of Service	Character of Service	Character of Service
2º	Character of Service	Race/color	Hydroxyurea	Hydroxyurea	Race/color
3º	Hydroxyurea	Hydroxyurea	Race/color	Race/color	Age Group
4º	Establishment	Establishment	Establishment	Establishment	Hydroxyurea
5º	Age Group	Age Group	Age Group	Age Group	Sex*
6º	Sex*	Sex*	Sex*	Sex*	Establishment

Data source: Original research.
*Insignificant predictor.

CONCLUSIONS

The research revealed that race/color and the type of care (elective or emergency) are the main predictors of hospitalization for patients with Sickle Cell Disease in Bahia, highlighting the impact of institutional racism on healthcare access. Other relevant factors, such as the use of hydroxyurea, age group, and healthcare facility, were also identified, although with less influence. The application of artificial intelligence techniques, particularly machine learning models, was crucial in identifying these patterns with greater precision, emphasizing the need for public policies that address racial disparities and expand access to effective treatments like hydroxyurea, while considering the specific characteristics of patients with Sickle Cell Disease.

ADDITIONAL KEY INFORMATION

Author Contact Information: crisfuezi@hotmail.com.
Funding Source: Municipal Government of Salvador.
Conflicts of Interest: There are no conflicts of interest.
Acknowledgements: Municipal Health Department of Salvador (SMS) and the Municipal Department of Innovation and Technology (SEMIT/COGEL).

WCE WORLD CONGRESS OF EPIDEMIOLOGY 2024