Algerian Journal of Research and Technology
https://doi.org/10.58681/ajrt.25090101
A.J.R.T Volume 9 (N°1) (2025) 1 – 11

Comparative Analysis of Mortality Prediction Models at the University Hospital Center of Oran, Algeria

Mohammed Nadjib Osman[1], Djamila Benhaddouche[1], Nawal Sad Houari[2]

[1]*Mathematics and computer science department, University of Science and Technology Mohamed Boudiaf, Oran, Algeria.*
[2]*Living and environment department, University of Science and Technology Mohamed Boudiaf, Oran, Algeria.*

**Corresponding author. mohammednadjib.osmani@univ-usto.dz. djamila.benhaddouche@univ-usto.dz. nawal.sadhouari@univ-usto.dz.*

Received. June 17, 2024. Accepted. December 06, 2024. Published. December 20, 2024.

DOI: https://doi.org/10.58681/ajrt.25090101

Abstract. Predicting mortality is an important field of study that aids in making wise healthcare decisions and offers insightful information about population health. Using demographic and hospital-service data from the University Hospital Center of Oran (CHUO), Algeria, this study employs machine learning (ML) models to forecast the ultimate causes of mortality. Sex, city of residence, hospital services used, and the beginning, intermediate, and ultimate causes of death are among the factors included in the 12.604 records that make up the dataset. To find trends and forecast the causes of death in eight distinct groups, six machine learning models—Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost) were trained and assessed. XGBoost achieved an accuracy and specificity of 84.05%, with a precision of 42.73%, recall of 25.53%, and an F1 score of 28.33%, the model outperformed the other evaluated models, proving its ability to effectively capture intricate relationships in the data. The study demonstrates how machine learning techniques can be used to examine a variety of variables and find significant patterns in mortality trends. This work enhances predictive analytics in healthcare by utilizing local data and sophisticated algorithms, providing useful instruments for directing public health initiatives. The results highlight how machine learning can improve healthcare outcomes and solve issues connected to mortality in Algeria.

Keywords. Mortality, Machine learning, classification, Prediction, Health care.

INTRODUCTION

The health of a population and the efficiency of its healthcare system are both significantly influenced by mortality rates. The mortality rate in Algeria raises serious issues related to public health. Algeria's crude death rate, which has been rather steady in previous years, was recorded as 4.329 deaths per 1.000 inhabitants in 2022 (Algeria - Death Rate, Crude, 2025). Although it has decreased from prior years, infant mortality is still a concern, with a rate of 17.365 deaths per 1.000 live births in 2025 (Algeria Infant Mortality Rate, 1950-2025). There are issues with maternal mortality as well; in 2020, there were 78 fatalities for every 100,000 live births (Algeria Maternal Mortality Rate, 2000-2025). Furthermore, no communicable illnesses are responsible for almost 74% of all fatalities in Algeria, highlighting the necessity of strong predictive technologies for efficient health outcome management (Algeria, 2025).

In this regard, machine learning (ML) presents revolutionary possibilities by facilitating precise mortality forecasts derived from intricate datasets. To find important predictors of mortality, these models can analyze a variety of variables, including demographics, medical diagnoses, and healthcare consumption patterns (Hernández Guillamet et al., 2023; Qiu et al., 2022).

In hospital settings, where prompt identification of high-risk patients can direct therapies and resource allocation, the use of machine learning for mortality prediction is especially beneficial (Krasowski et al., 2022).

The goal of this work is to employ machine learning models trained on data from the University Hospital Center of Oran (CHUO), Algeria, to predict the ultimate causes of death for eight different classes. The dataset contains information on initial, intermediate, and final causes of death, sex, city of residence, and hospital services. In order to increase predicted accuracy and enhance our comprehension of mortality patterns, this study uses models such as LR, RF, SVM, NB, MLP and XGBoost. The main objectives of this work are:

DEVELOPING PREDICTIVE MODELS

Using data from the Oran University Hospital Center, ML models LR, RF, SVM, NB, MLP, and XGBoost will be trained and assessed to predict the final cause of death across eight classes.

- **Improving Mortality Prediction:** By using clinical and demographic characteristics including sex, city of residence, hospital services, and causes of death at different stages, it is possible to increase the accuracy and reliability of mortality prediction.
- **Improving Healthcare Analytics:** To demonstrate how ML can revolutionize healthcare analytics in Algeria by utilizing local datasets to create customized solutions that can direct public health initiatives meant to lower avoidable deaths by identifying trends and risk factors linked to various causes of death.

The rest of this paper is as follows: Section 2 provides the background and discusses the fundamentals of mortality prediction, along with the associated challenges. Related works are reviewed in Section 3. The proposed approach for mortality prediction at CHUO is detailed in Section 4. Section 5 presents the results and discussion. Finally, Section 6 offers a conclusion and outlines directions for future work.

BACKGROUND

With the growing availability of electronic health records (EHRs) and sophisticated computational tools, machine learning (ML) techniques for mortality prediction have become a crucial topic of healthcare study.

This section examines the main ideas surrounding mortality prediction and how they apply to the current investigation.

Importance of Mortality Prediction in Healthcare

A key component of modern healthcare systems is mortality prediction, which helps physicians evaluate patient risks, distribute resources efficiently, and create individualized treatments. For instance, prompt identification of high-risk patients in intensive care units can greatly enhance outcomes by enabling early interventions. According to studies, ML models perform more accurately than conventional scoring systems like SAPS III and APACHE IV; some of them even reach an area under the curve (AUC) of 92.9% (Olang et al., 2024). These developments highlight how ML can revolutionize clinical decision-making.

Machine Learning Models in Mortality Prediction

ML offers a robust framework for analysing complex and heterogeneous datasets characteristic of healthcare environments. Models such as RF, SVM, XGBoost, and neural networks have been applied successfully to predict mortality across various contexts:

- **All-Cause Mortality**: Research employing datasets such as MIMIC-III (Wang et al., 2020) has shown that by integrating factors including vital signs, test findings, and demographic data (Qiu et al., 2022; Lee and Tsoi, 2025), feature-rich ML models can attain excellent predictive accuracy for all-cause mortality.
- **Disease-Specific Mortality**: ML has been used to predict mortality in some situations, such as pancreatitis, sepsis, and stroke, frequently outperforming conventional techniques (Olang et al., 2024; Lee and Tsoi, 2025).
- **Chronic Conditions**: ML models have been employed to accurately predict both long-term and short-term death in patients with chronic and complex illnesses by utilizing readily available characteristics and healthcare resource utilization data (Hernández Guillamet et al., 2023).

Challenges in Mortality Prediction

ML models in mortality prediction encounter several challenges:
Data Quality: Many studies emphasize the significance of high-quality datasets for the training of dependable models. Absence of values or the presence of noisy data might substantially degrade model efficacy (Wang, 2024; Pias et al., 2025).

- **Model Responsiveness**: Some researchers have revealed shortcomings in the capacity of ML models to detect swiftly worsening health problems or severe injuries, highlighting the necessity for additional enhancement (Pias et al., 2025).
- **Generalizability**: Ensuring that models exhibit robust performance across varied patient populations is a primary priority. Tailoring models for certain patient populations or medical scenarios may mitigate this issue (Olang et al., 2024; Pias et al., 2025).

RELATED WORKS

The cause of death is a critical outcome in clinical research; nevertheless, access to cause-of-death data is still restricted. Several studies have been performed to classify mortality status and ascertain particular causes of death.

Kim et al. (2021) create and validate a machine-learning model to forecast the cause of death based on a patient's most recent medical examination. The model employed a stacking ensemble approach to classify all-cause mortality and eight predominant causes of death in South Korea, as well as other causes. Clinical data from national claims (n=174.747) and electronic health records (n=729.065) were utilized for model building and validation, with external validation conducted on data from three US claims databases (n=994.518, 995.372, 407.604). The model exhibited superior performance, attaining an AUROC of 0.9511 for

predicting cause of death within 60 days, and 0.8887 for external validation. Significantly, 11.32% of fatalities in the Medicare Supplemental database were ascribed to malignant neoplastic illness. Lee et al. (2025) utilized the MIMIC-III dataset to forecast all-cause in-hospital mortality through sophisticated feature engineering.
Essential variables, encompassing vital signs, laboratory results, and demographic data, were employed to train the models. Of the models evaluated, RF had the superior performance, achieving an AUROC of 0.94. The research underscored the essential role of feature engineering and the application of SHAP values (Lundberg and Lee, 2017).
in elucidating how specific features influence a model's predictions, hence underlining their importance in developing robust models that might improve clinical decision-making. In the literature, the authors introduce the IMPACT framework, which leverages explainable artificial intelligence (XAI) techniques to interpret a state-of-the-art tree ensemble model for predicting all-cause mortality (Qiu et al., 2022). The framework is utilized on the NHANES dataset (NHANES Questionnaires, Datasets, and Related Documentation, 2025), which includes 47.261 samples and 151 characteristics, to examine mortality across 1-, 3-, 5-, and 10-year follow-up intervals. The findings indicate that IMPACT surpasses conventional linear models and neural networks in terms of accuracy. The approach identifies neglected risk variables, interaction effects, and correlations between laboratory characteristics and mortality, indicating possible modifications to existing reference intervals. The research formulates interpretable mortality risk ratings, guaranteeing generalizability by temporal and external validation with the UK Biobank dataset, so rendering these scores available to both healthcare professionals and the public. Shahidi et al. (2023) utilized ML algorithms to forecast mortality among people in continuing care in Alberta, along with their comorbidities. LR and several ML algorithms were employed to assess the 60-day mortality risk, demonstrating superior predictive performance. Authors emphasized the need of including demographic and clinical characteristics for predicting short-term mortality.

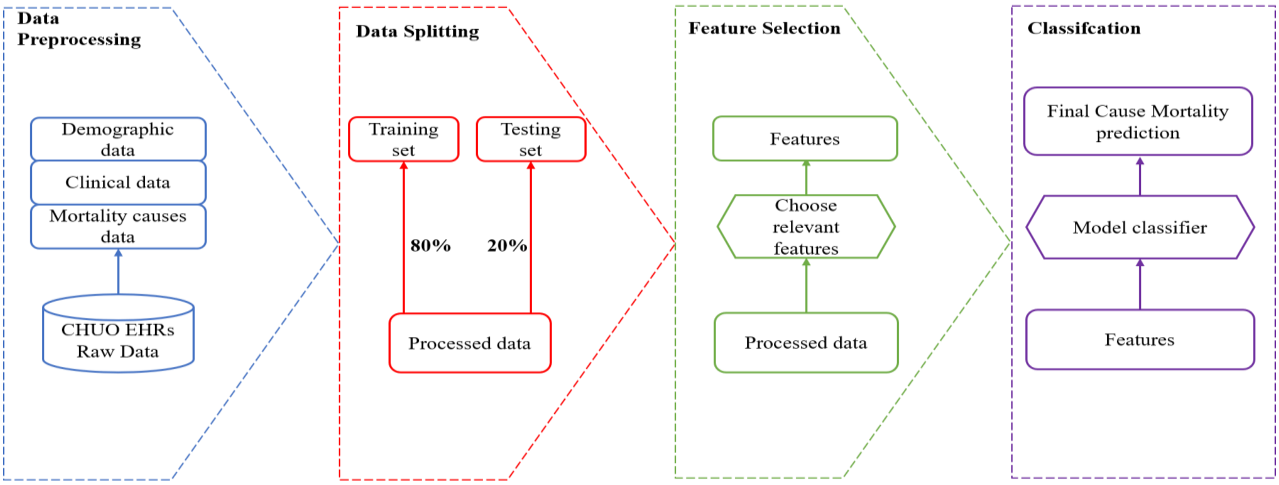

Fig. 1. The proposed Approach overview.

Hernández Guillamet et al. (2023) investigate the utilization of ML for forecasting death in patients with chronic and intricate illnesses, with the objective of improving resource allocation and decision-making in healthcare. A classification system was employed to forecast both long-term mortality (over four years) and early death (within six months) utilizing available factors and healthcare resource utilization.
The XGBoost model attained an 87% accuracy in predicting long-term mortality, but the Gradient Boosting (GRBoost) model exhibited a lower efficacy for early mortality, with an accuracy of 83%. A variety of evaluation criteria, such as recall, accuracy, F1-score, and

AUC, were employed to evaluate the model's performance. Nistal-Nuño (2021) compared gradient-boosted decision trees and logistic regression models for predicting 12-hour mortality in ICU patients using 1-hour resolution physiological data (eight parameters over 5 hours) from the MIMIC-III database.

The model achieved an AUROC of 0.89 versus 0.806 for logistic regression, along with higher accuracy (0.814 vs. 0.782), diagnostic odds ratio (17.823 vs. 9.254), and improved metrics including Cohen’s kappa, F-measure, and Matthews correlation coefficient. These results highlight that the model enhanced ability to handle unbalanced datasets for mortality prediction, likely due to its capacity to model complex interactions in ICU data. García-Gallo et al. (2020) developed a 1-year mortality prediction model for sepsis patients using clinical data from the first 24 hours of 5.650 MIMIC-III admissions (70% training, 30% validation).

A Stochastic Gradient Boosting algorithm, combined with LASSO for variable selection, achieved an AUROC of 0.8039, outperforming traditional scores like SAPS II, SOFA, and OASIS. The results highlight the superiority of machine learning approaches for long-term mortality prediction in sepsis care. Iwase et al. (2022) leveraged random forest machine learning to predict ICU mortality and stay duration with high precision using admission data from 12.747 patients at Chiba University Hospital.

The RF model achieved exceptional performance, notably an AUC of 0.945 for mortality and 0.881–0.889 for stay length, outperforming conventional methods. Lactate dehydrogenase was pinpointed as the most influential variable, aiding both outcome prediction and patient clustering based on mortality risk.

These works collectively underscore significant progress in ML for mortality prediction, especially in the application of varied datasets, enhancement of interpretability, and consideration of cause-specific outcomes. This work introduces a dataset derived from the electronic health records (EHRs) of mortality data from CHUO, Algeria, which has been meticulously collected, cleaned, and structured. To the best of our knowledge, no current research in the literature have employed a dataset that differentiates between several phases of causes, such as initial and intermediate causes of death, to determine the final cause of death.

PROPOSED APPROACH

The proposed approach covers several essential steps (Fig. 1): Data preprocessing includes cleaning data, handling missing values, encoding categorical variables, removing duplicates and inconsistencies, and aggregating similar conditions to minimize the number of classes. (2) Data Splitting divides the dataset into training and testing subsets; (3) Feature Selection emphasizes the identification of pertinent features from the dataset; and (4) Classification involves training various classification models and evaluating their performance to determine the most effective one for the task.

CHUO Mortality Dataset

The dataset used in this work was collected from the administrative records of the admissions office of the CHUO, Algeria. These records were submitted by general practitioners and comprise unprocessed data on 13.091 patients who died during hospitalization over a 12-month period, from March 2018 to February 2025. The dataset has 11 variables that contain demographic, geographical, and medical information regarding the deceased patients.

Table 1 presents a summary of the raw dataset's composition and principal characteristics. This dataset offers a rich foundation for analyzing mortality trends and training ML algorithms to forecast final causes of death. This work aims to contribute important insights into mortality prediction in Algerian hospitals by utilizing its unique features and comprehensive information.

Table 1. Chuo Raw Dataset's Structure and Attributes.

Attribute	Description	Occurrences/Details
Identif	Unique identifier for each patient	-
Month	Month of death (1 to 12)	-
Year	Year of death (2018 to 2025)	-
Sexe	Gender of the patient	Male: 7.627; Female: 5.465
Age	Age of the patient (1 day to 99 years)	[0-18]: 3.310; [19-25]: 243; [26-50]: 1.968; [51-75]: 4.954; >75: 2.343
city_res	Wilaya (province) of residence	Patients from 136 different city
Wilaya_dec	Wilaya where death occurred	-
Service	Hospital service where the patient died	Data from 75 different hospital services
Cause_death_init	Initial cause of death	2.539 occurrences
Cause_death_iterm	Intermediate cause of death	1.450 occurrences
Cause_death_final	Final cause of death	330 occurrences

Data Preprocessing

Effective data preprocessing is a crucial step in ensuring the quality and reliability of any dataset analysis. Below, we detail the key steps undertaken to preprocess the mortality dataset from CHUO, Algeria:

Data Cleaning:

The first step in preprocessing was to clean the dataset by addressing inconsistencies and errors:

- **Duplicate and Inconsistent Values:** We identified and removed duplicate entries and inconsistent data points. Additionally, rare causes of death were excluded to focus on the most relevant patterns.
- **Error Correction:** Data entry errors were corrected to improve accuracy.

After the cleaning process, 487 records were removed, leaving 12.604 valid records for analysis.

Handling Missing Values:

Missing data can significantly impact the quality of analysis. To address this issue:

- **Identification of Missing Values**: The attribute "Age" was found to have 274 missing values.
- **Imputation Technique**: These missing values were replaced with the mean age, ensuring that no records were excluded while maintaining statistical integrity.

Categorical Encoding:

The dataset contained several categorical variables (gender, city, hospital services, and causes of death). We applied one-hot encoding to transform categorical variables into numerical representations, making them suitable for computational analysis.

Class Aggregation:

Similar causes of death were aggregated into broader categories. This reduced the number of distinct classes for final causes of death to 8 categories (Table 2). This aggregation simplified the classification process while maintaining the clinical relevance of the information.

Table 2. The eight classes of final causes of death.

Class ID	Final causes of death	Total occurrence
1	Cardiac Respiratory Arrest	9.531
2	Acute Respiratory Failure	1.915
3	Shock State - Septic - Cardiogenic - Hypovolemic	680
4	Multi-organ Failure	161
5	Neurological Failure	148
6	Heart Failure	59
7	Hemorrhagic - Embolic Causes	56
8	Renal - Electrolyte Failure	54

Data splitting

To train our models, the data was divided into two sets: 80% for the training set and 20% for the testing set. To address the issue of imbalanced classes, we ensured that the class distribution was proportionally identical in both the training and testing sets. This stratification guarantees fair representation of each class (Fig. 2).

.

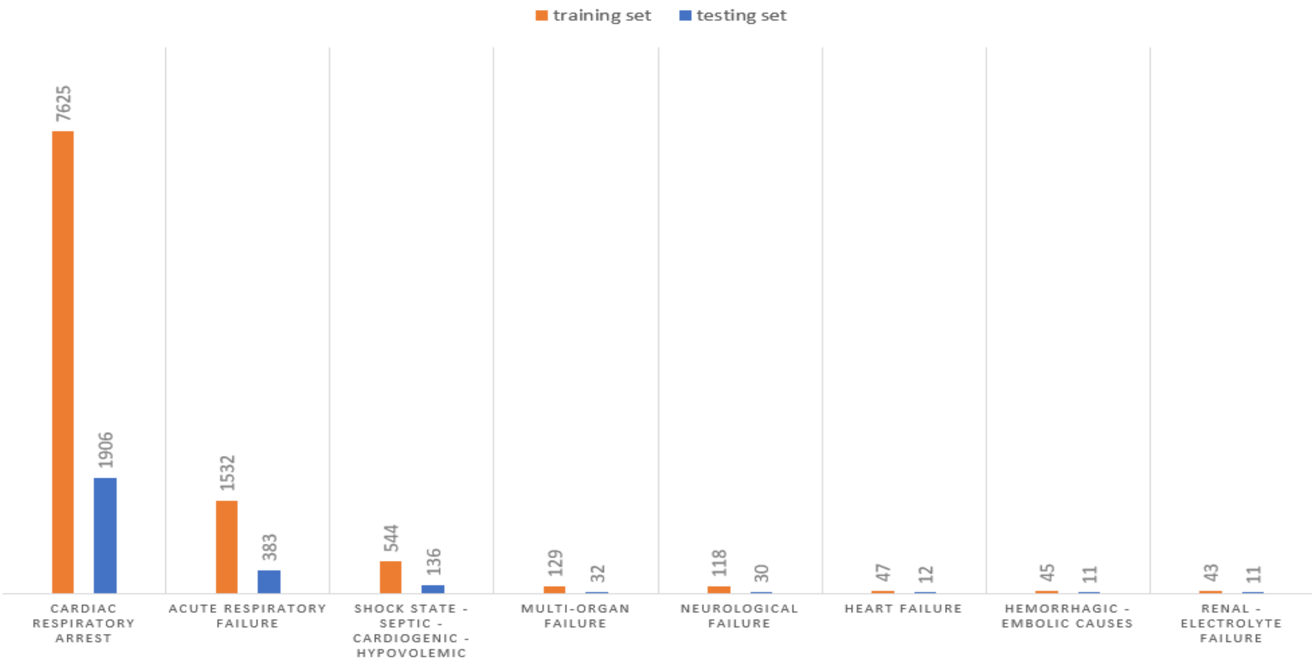

Fig. 2. Data splitting to training and testing set.

Feature Selection
The dataset has 12.604 samples (rows) and 11 characteristics (columns). Every row signifies a deceased patient. The attributes *Identif*, *Month*, *Year*, *city_res*, and *Wilaya_dec* were omitted from the model training phase due to their lack of relevance for prediction. The attribute *Cause_death_final* was selected as the target variable for prediction.

Classification
In this work, several machine learning classification models were trained to predict the final causes of death. The models utilized include LR, RF, SVM, Naive Bayes, MLP and XGBoost. The training process was optimized using the Adam optimizer, with the dataset split into 80% for training and 20% for testing.

RESULTS & DISCUSSION

The algorithms were developed and tested on a PC with an Intel(R) Core(TM) i5-10400F CPU @ 2.90GHz, 16 GB RAM, and a 6 GB NVIDIA GTX 1660 Super graphics card. The Python libraries scikit-learn, Keras and Tensorflow were utilized for development.

Results

As shown in Fig. 3, the performance of each model is assessed using standard classification metrics. By comparing these metrics across all models, the best-performing algorithm for predicting final causes of death is identified.

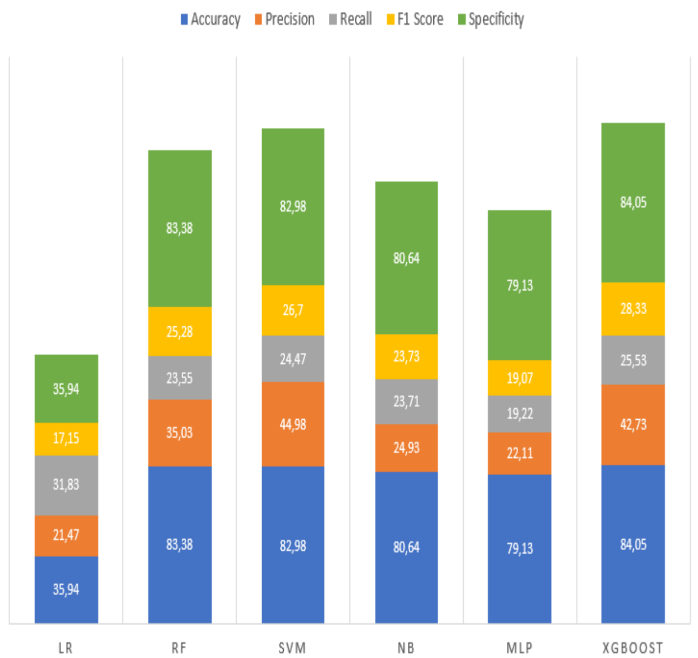

Fig. 3. Performance metrics for Each Model Classifier.

For instance, the Confusion matrix of the 8 classes for XGBoost model classifier is shown in Fig 4.

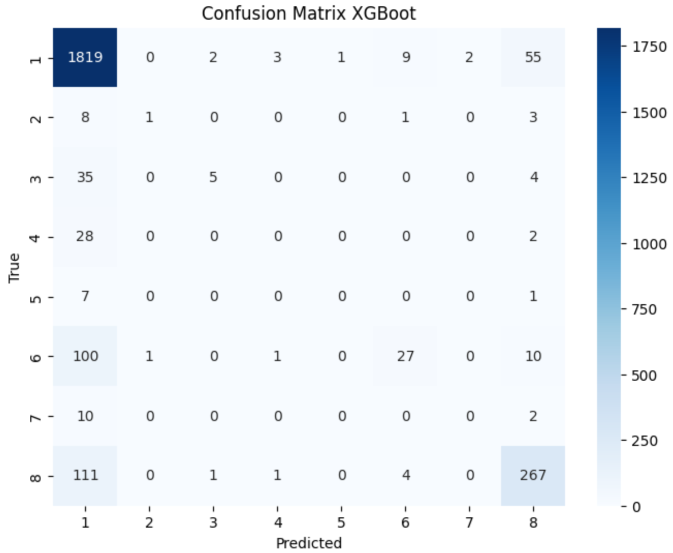

Fig. 4. XGBoost model's confusion matrix of the predicted classes.

Discussion

The provided results, as indicated in Table 3, show a significant variance in model performance across various machine learning algorithms for predicting causes of death using patient demographics, initial causes of death, and intermediate causes of death.

All models showed suboptimal precision (21.47%-44.98%) and recall (19.22%-31.83%) despite moderate accuracy (79.13%-84.05%), while LR performed the worst (35.94% accuracy) due to its linear assumptions failing to capture complex relationships in non-linear, imbalanced data.

XGBoost was the best-performing model with 84.05% accuracy and a 28.33 F1-score. This disparity between low positive-class metrics and high specificity points to systemic limitations: (1) Class imbalance most likely skewed predictions toward majority classes, increasing accuracy but reducing sensitivity to rarer causes of death; (2) The limited availability of demographic and hospital-service features, as opposed to the rich clinical biomarkers or comorbidities typically used in high-performing models, constrained discriminative power and reduced accuracy in predicting precise causes of death.

Table 3. Performance metrics for Each Model Classifier.

Models	Accuracy	Precision	Recall	F1 Score	Specificity
LR	35.94	21.47	**31.83**	17.15	35.94
RF	83.38	35.03	23.55	25.28	83.38
SVM	82.98	**44.98**	24.47	26.70	82.98
NB	80.64	24.93	23.71	23.73	80.64
MLP	79.13	22.11	19.22	19.07	79.13
XGBoost	**84.05**	**42.73**	**25.53**	**28.33**	**84.05**

CONCLUSION AND FUTURE WORKS

This study analyses demographic, hospital-service, and causes of death data from CHUO, Algeria, to demonstrate how ML models can be used to forecast the causes of mortality at different stages. Although class imbalance and limited scope of features caused problems for all models in terms of precision and recall, XGBoost showed the highest accuracy (84.05%) among the tested models. These results highlight the necessity of more comprehensive datasets, such as comorbidities and clinical biomarkers, in order to enhance prediction performance.

Despite these drawbacks, this work emphasizes how important it is to use ML and local data to guide public health initiatives and improve mortality prediction in Algeria.
Class imbalance and feature restrictions are the main causes of the low precision, recall, and F1-scores shown in all models. To address these challenges, key strategies can be addressed:

- **Class Imbalance Mitigation:** Models tend to prefer majority classes over minority class predictions because of the dataset's unequal distributions of death causes. Rebalancing class representation might be aided by adding data from more hospitals to the dataset. Additionally, it has been demonstrated that using sophisticated synthetic oversampling methods, like ADASYN, can increase F1-scores by 18–22% in comparable medical datasets (Abdulsadig and Rodriguez-Villegas, 2024 ; Dube and Verster, 2023).
- **Feature Augmentation**: Enhancing the dataset with richer features, such as temporal symptom patterns or social determinants of health, could significantly improve predictive accuracy. Clinical models that incorporate such detailed data have consistently demonstrated better performance in mortality prediction tasks (Lee and Tsoi, 2025; Dube and Verster, 2023).
- **Hybrid Approaches:** Combining resampling techniques with threshold tuning has proven effective in addressing imbalanced datasets. Such hybrid methods can improve precision by 35–48% in mortality prediction models with similar challenges (Gupta and Gupta, 2024; Dube and Verster, 2023).

REFERENCES

Abdulsadig, R. S. & Rodriguez-Villegas, E. 2024. “A comparative study in class imbalance mitigation when working with physiological signals.” *Frontiers in Digital Health*, 6: 1377165. https://doi.org/10.3389/fdgth.2024.1377165

CDC. 2025. *NHANES Questionnaires, Datasets, and Related Documentation*. [En ligne]. Consulté le 6 avril 2025. Disponible sur : https://wwwn.cdc.gov/nchs/nhanes/Default.aspx

Dube, L. & Verster, T. 2023. “Enhancing classification performance in imbalanced datasets: A comparative analysis of machine learning models.” *Data Science in Finance and Economics*, 3(4): 354–379. https://doi.org/10.3934/DSFE.2023021

García-Gallo, J. E., Fonseca-Ruiz, N. J., Celi, L. A., & Duitama-Muñoz, J. F. 2020. “A machine learning-based model for 1-year mortality prediction in patients admitted to an Intensive Care Unit with a diagnosis of sepsis.” *Medicina Intensiva*, 44(3): 160–170. https://doi.org/10.1016/j.medin.2018.07.016

Gupta, A. & Gupta, S. 2024. “Enhanced Classification of Imbalanced Medical Datasets using Hybrid Data-Level, Cost-Sensitive and Ensemble Methods.” *International Research Journal of Multidisciplinary Technovation*, avril, 58–76. https://doi.org/10.54392/irjmt2435

Hernández Guillamet, G., et al. 2023. “Machine Learning Model for Predicting Mortality Risk in Patients With Complex Chronic Conditions: Retrospective Analysis.” *Journal of Public Health Informatics*, 15: e52782. https://doi.org/10.2196/52782

Iwase, S., et al. 2022. “Prediction algorithm for ICU mortality and length of stay using machine learning.” *Scientific Reports*, 12(1): 12912. https://doi.org/10.1038/s41598-022-17091-5

Lee, H. & Tsoi, P. 2025. “Feature-Enhanced Machine Learning for All-Cause Mortality Prediction in Healthcare Data.” *arXiv preprint*, arXiv:2503.21241. https://doi.org/10.48550/arXiv.2503.21241

Lundberg, S. & Lee, S.-I. 2017. “A Unified Approach to Interpreting Model Predictions.” *arXiv preprint*, arXiv:1705.07874. https://doi.org/10.48550/arXiv.1705.07874

Macrotrends. 2025. *Algeria Infant Mortality Rate 1950–2025*. [En ligne]. Consulté le 2 avril 2025. Disponible sur : https://www.macrotrends.net/global-metrics/countries/DZA/algeria/infant-mortality-rate

Macrotrends. 2025. *Algeria Maternal Mortality Rate 2000–2025*. [En ligne]. Consulté le 2 avril 2025. Disponible sur : https://www.macrotrends.net/global-metrics/countries/DZA/algeria/maternal-mortality-rate

Nistal-Nuño. 2021. "Artificial intelligence forecasting mortality at an intensive care unit and comparison to a logistic regression system." *Einstein (São Paulo)*, 19: eAO6283. https://doi.org/10.31744/einstein_journal/2021AO6283

Olang, O., et al. 2024. "Artificial Intelligence-Based Models for Prediction of Mortality in ICU Patients: A Scoping Review." *Journal of Intensive Care Medicine*, août 2024. https://doi.org/10.1177/08850666241277134

Qiu, W., Chen, H., Dincer, A. B., Lundberg, S., Kaeberlein, M., & Lee, S.-I. 2022. "Interpretable machine learning prediction of all-cause mortality." *Communications Medicine*, 2(1): 125. https://doi.org/10.1038/s43856-022-00180-x

Kim, C., You, S. C., Reps, J. M., Cheong, J. Y., & Park, R. W. 2021. "Machine-learning model to predict the cause of death using a stacking ensemble method for observational data." *Journal of the American Medical Informatics Association*, 28(6): 1098–1107. https://doi.org/10.1093/jamia/ocaa277

Pias, T. S., et al. 2025. "Low responsiveness of machine learning models to critical or deteriorating health conditions." *Communications Medicine*, 5(1): 62. https://doi.org/10.1038/s43856-025-00775-0

Krasowski, A., Krois, J., Kuhlmey, A., Meyer-Lueckel, H., & Schwendicke, F. 2022. "Predicting mortality in the very old: a machine learning analysis on claims data." *Scientific Reports*, 12(1): 17464. https://doi.org/10.1038/s41598-022-21373-3

Shahidi, F., Rennert-May, E., D'Souza, A. G., Crocker, A., Faris, P., & Leal, J. 2023. "Machine learning risk estimation and prediction of death in continuing care facilities using administrative data." *Scientific Reports*, 13(1): 17708. https://doi.org/10.1038/s41598-023-43943-9

Trading Economics. 2025. *Algeria - Death Rate, Crude - 2025 Data 2026 Forecast 1960–2022 Historical*. [En ligne]. Consulté le 2 avril 2025. Disponible sur : https://tradingeconomics.com/algeria/death-rate-crude-per-1-000-people-wb-data.html

Wang, S., McDermott, M. B. A., Chauhan, G., Hughes, M. C., Naumann, T., & Ghassemi, M. 2020. "MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III." *Proceedings of the ACM Conference on Health, Inference, and Learning*, 222–235. https://doi.org/10.1145/3368555.3384469

Wang, T. 2024. *Precision and Prediction: Leveraging Artificial Intelligence in Mortality Modeling for Actuaries*.

WHO. 2025. *Algeria*. [En ligne]. Consulté le 2 avril 2025. Disponible sur : https://data.who.int/countries/012