Performance Analysis of Machine Learning and Statistical Models in Predicting Outcomes of Carbon Monoxide Poisoning: A Multicenter Retrospective Study

Po-Chun Chuang[1,2], Ye-In Chang[1], Chia-En Li[3]

[1] Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan.

[2] Department of Emergency Medicine, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan.

[3] Department of Information Management, Cheng Shiu University, Kaohsiung, Taiwan

Keyword: Carbon Monoxide Poisoning, Machine Learning, Predictive Models, Toxicology, Outcome Prediction

Abstract

Background: Carbon monoxide (CO) poisoning is a major public health issue worldwide, with symptoms ranging from mild discomfort to severe conditions like coma or death. Although there are various treatment options available, such as hyperbaric oxygen therapy (HBOT), predicting patient outcomes remains challenging due to the variability in clinical presentations and the lack of standardized guidelines in many countries. Traditional statistical models, like the COGAS and FIRED scores, have been used for outcome prediction, but with the rise of machine learning, there is potential for developing more accurate and reliable predictive tools.

Materials and Methods: This project is a multicenter retrospective study using data from the Chang Gung Research Database, focusing on adult patients (age ≥ 17) who were treated for CO poisoning between 2019 and 2022. We will compare the predictive performance of traditional statistical models (FIRED, COGAS, SOFA, PSS) with machine learning models such as Random Forest, Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP). The data cleaning process will include handling missing values using K-Nearest Neighbors (KNN) imputation, detecting and addressing outliers, and encoding categorical data. Feature selection will be guided by clinical expertise and algorithmic methods. The primary outcome will be the model's ability to predict poor outcomes, defined as death or a Glasgow Coma Scale (GCS) score of less than 13 at discharge. Model performance will be evaluated using metrics such as the Area Under the Curve (AUC), sensitivity, specificity, precision, recall, and F1-score.

Result: Preliminary results indicate that the FIRED score may continue to be a strong predictor, with an AUC of 0.93 in external validation datasets. However, machine learning models, particularly XGBoost and Random Forest, also show potential, delivering comparable or slightly lower AUCs (around 0.89). We anticipate further refining these models and exploring whether integrating machine learning techniques with traditional clinical predictors can enhance outcome prediction in CO poisoning cases. This study aims to equip clinicians with more precise tools for predicting patient outcomes, ultimately leading to better-targeted interventions and improved patient care.

Conclusion: In this study, the FIRED score remains a strong predictor in external validation, with an AUC of 0.93. Machine learning models, although not fine-tuned or carefully adjusted, showed similar performance to the expert-modified statistical models. Both approaches are clinically sufficient for predicting outcomes in patients with carbon monoxide (CO) poisoning.

Introduction

Carbon monoxide poisoning (COP) is a common, potentially fatal emergency with cases reported worldwide. Its clinical presentation varies, ranging from mild symptoms like headaches and dizziness to severe outcomes such as coma and death [1]. In the United States, approximately 50,000 people seek treatment for COP annually, with a mortality rate of 1-3% [2]. In Germany, COP is one of the most common types of poisoning, with a mortality rate of about 0.73 per 100,000 people [1, 3]. In Taiwan, a study using national health insurance data reported 24,046 cases of COP between 1999 and 2012, with 6,793 patients (28.2%) receiving hyperbaric oxygen therapy (HBOT) [4]. However, many countries, including China and Germany, lack national guidelines for COP treatment, leading to uncertainty in treatment effectiveness and timing, even after reducing carboxyhemoglobin (COHb) levels [3, 5].

Survivors of COP often experience long-term neurological and emotional sequelae, which may be linked to CO's impact on mitochondrial respiration, energy utilization, inflammation, and free radical generation, especially in the brain and heart [2]. The primary treatment for COP is HBOT, which can significantly reduce permanent neurological and emotional effects [2, 6]. In normal air (21% oxygen), the half-life of CO is 320 minutes; in 100% oxygen, it decreases to 90 minutes, and in a hyperbaric oxygen environment at 3 atmospheres, it can further reduce to 23 minutes [7]. Therefore, HBOT is a widely recognized treatment for severe COP cases [5, 6, 8-10].

Delayed neuropsychiatric sequelae (DNS) is another common and serious consequence of COP [11]. In a study of 387 patients in Korea, 26.1% developed DNS, with acute brain lesions (ABLM) observed on MRI closely associated with DNS occurrence [12, 13]. Specifically, the sensitivity of abnormal findings on diffusion-weighted imaging (DWI) was 75.2%, and specificity was 90.2%, indicating its potential as a predictor for DNS [12]. Additionally, lower Glasgow Coma Scale (GCS) scores and longer CO exposure times were identified as significant risk factors for DNS [13, 14].

Clinical indicators have been linked to poor neurological outcomes after COP [15, 16]. Several statistical models have been developed to predict DNS and adverse outcomes following COP. For example, a study in Korea involving 1,282 patients proposed the COGAS score (incorporating high creatine kinase, HBOT, lower GCS score, older age, and shock) to predict neurocognitive sequelae risk, achieving an AUC of 0.87 [17]. Another study with 192 patients showed that the PSS, initial SOFA score, and second SOFA score accurately predicted acute adverse outcomes in COP, with AUCs of 0.977, 0.945, and 0.978, respectively [18]. The FIRED score, which includes factors such as fire scene, intentional CO exposure, respiratory failure, GCS decline, and diabetes, effectively predicted poor outcomes with an AUC of 0.930 [19]. Additionally, a Chinese study successfully predicted DNS occurrence using a risk score model based on MRI abnormalities, initial GCS score, exposure duration, and CK levels, with an AUC of 0.99 [20].

With advancements in machine learning, more studies have started using these models to predict clinical outcomes. Common choices include Random Forest (RF) [21], Extreme Gradient Boosting (XGB) [22], and Multilayer Perceptron (MLP) [23, 24], known for their ability to handle non-linear features and complex data structures.

This study aims to use the Chang Gung Research Database [25] as an external validation dataset to compare the predictive abilities of existing statistical models with machine learning models. We seek to identify a more accurate and practical prediction method to enhance the prognosis of COP patients and provide more appropriate treatment recommendations.

Methods

In our previous paper [19] , we conducted a retrospective analysis of adult patients (age $\geq$ 17) who visited the emergency department at Chang Gung Memorial Hospital due to carbon monoxide poisoning (COP) between 2009 and 2018. In this study, a poor outcome was defined as death or a Glasgow Coma Scale (GCS) score of less than 13 at discharge. Through stepwise regression analysis, we developed the FIRED score, which includes five features: Fire scene (F), Intentional CO exposure (I), Respiratory failure (R), Every point of reduced GCS (E), and Diabetes mellitus (D). The FIRED score was shown to effectively predict outcomes in non-OHCA COP patients. A

FIRED score ≥ 10 was significantly associated with poor outcomes, achieving a sensitivity of 89.6%, specificity of 82.4%, and an AUC of 0.930.

This study aims to use the Chang Gung Research Database (CGRD) to analyze adult patients (age ≥ 17) who experienced COP between 2019 and 2022. This will serve as an external validation dataset to verify the accuracy of the FIRED score. Additionally, we will compare the performance of existing statistical models (such as COGAS, SOFA, PSS) with machine learning models (such as Random Forest, Extreme Gradient Boosting, and Multilayer Perceptron) in predicting COP outcomes, seeking a more accurate method to guide clinical treatment.

The detailed steps include:

1. **Ethical Review**:
   Approval has been obtained from the Chang Gung Medical Foundation Institutional Review Board (IRB), with the IRB number 202400111B0C501.

2. **Data Source**:
   Data will be sourced from the Chang Gung Research Database (CGRD). The study will include adult patients (age ≥ 17) who experienced COP between 2019 and 2022 across four CGRD hospitals (Linkou, Keelung, Chiayi, and Kaohsiung).

3. **Data Cleaning**:
   3-1. **Handling Missing Values**:
   Managing missing data is crucial. We use K-Nearest Neighbors (KNN) imputation [26] to fill in missing values. KNN identifies data points closest to the missing values (neighbors) and uses their average to replace the missing data. This method is suitable for datasets with similar distributions and intrinsic correlations, preserving overall trends and avoiding bias from random imputations.
   3**-2. Data Encoding**:
   Data encoding involves converting categorical and numerical data into a format suitable for model analysis. Categorical data will be transformed into numerical format using one-hot encoding or the pandas.get_dummies method in Python. Numerical data will be standardized, scaling values to a range (e.g., 0 to 1) to reduce errors and enhance model stability.

4. **Model Validation**:
   Using the cleaned data, we will validate established statistical models such as COGAS,

FIRED, PSS, and SOFA scores [17-20]. Additionally, we will validate machine learning models like Random Forest, Extreme Gradient Boosting, and Multilayer Perceptron to

The Preliminary Results

This study collected 318 external validation cases, with a total of 40 features (318 rows × 40 columns). Preliminary validation shows that the FIRED score achieved an AUC of 0.93, which is comparable to other machine learning models. However, we still need to gather features required for COGAS, PSS, SOFA scores, and other models to complete these comparisons.

1. FIRED Score

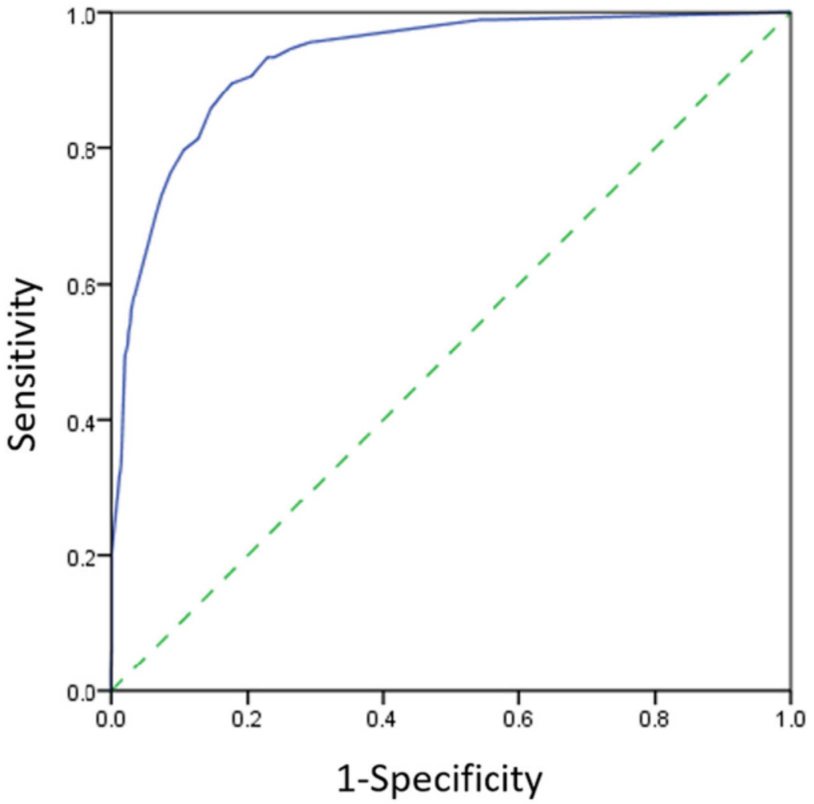

Figure Description: Performance of the FIRED score in the original dataset [19]. The area under the curve (AUC) is 0.930 (95% Confidence Interval, 0.911 to 0.949).

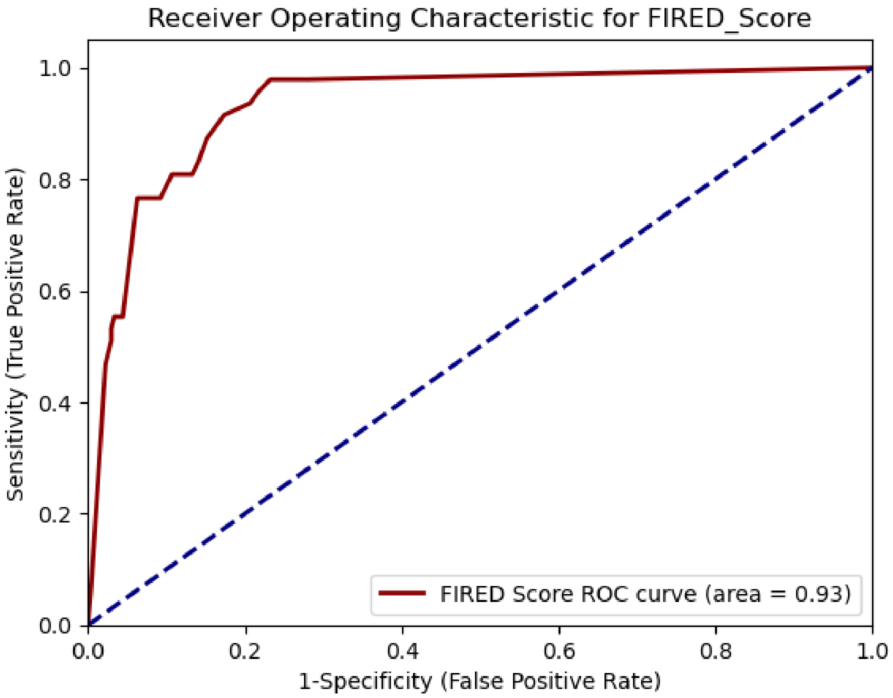

Figure Description: Performance of the FIRED score in the external validation data. AUC = 0.93.

2. XGBoost model

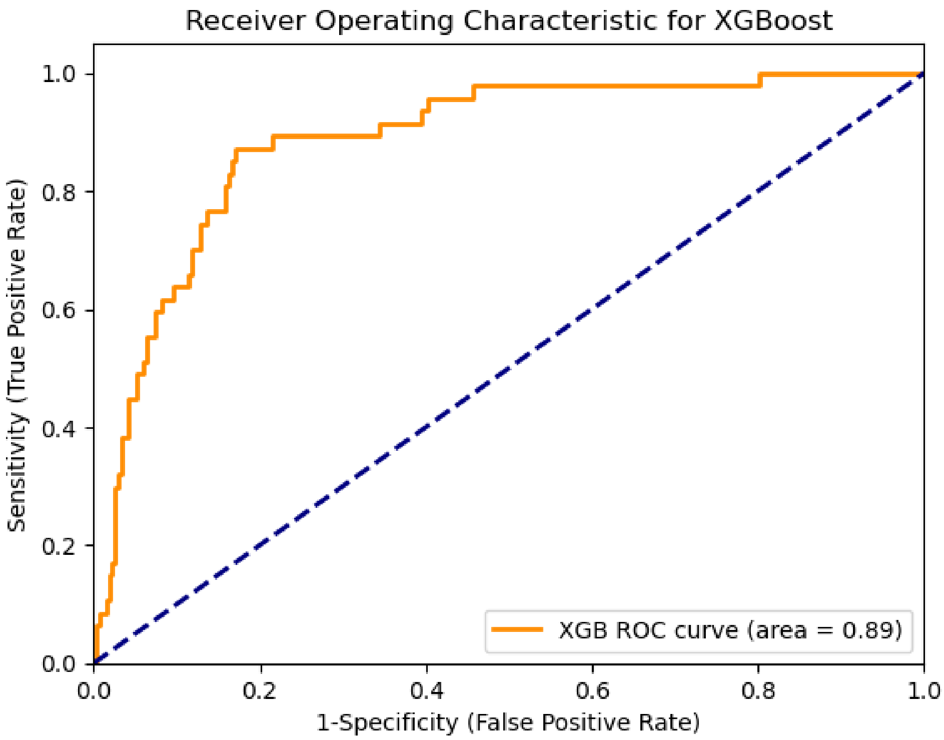

Figure Description: Performance of the XGBoost model on the external validation data. AUC = 0.89.

Table 1. The accuracy of external test data prediction in XGBoost model.				
	precision	recall	f1-score	support
0	0.93	0.92	0.93	271
1	0.57	0.60	0.58	47
accuracy = 0.87				318

Table Description: The accuracy of the external validation data in the XGBoost model is 87%.

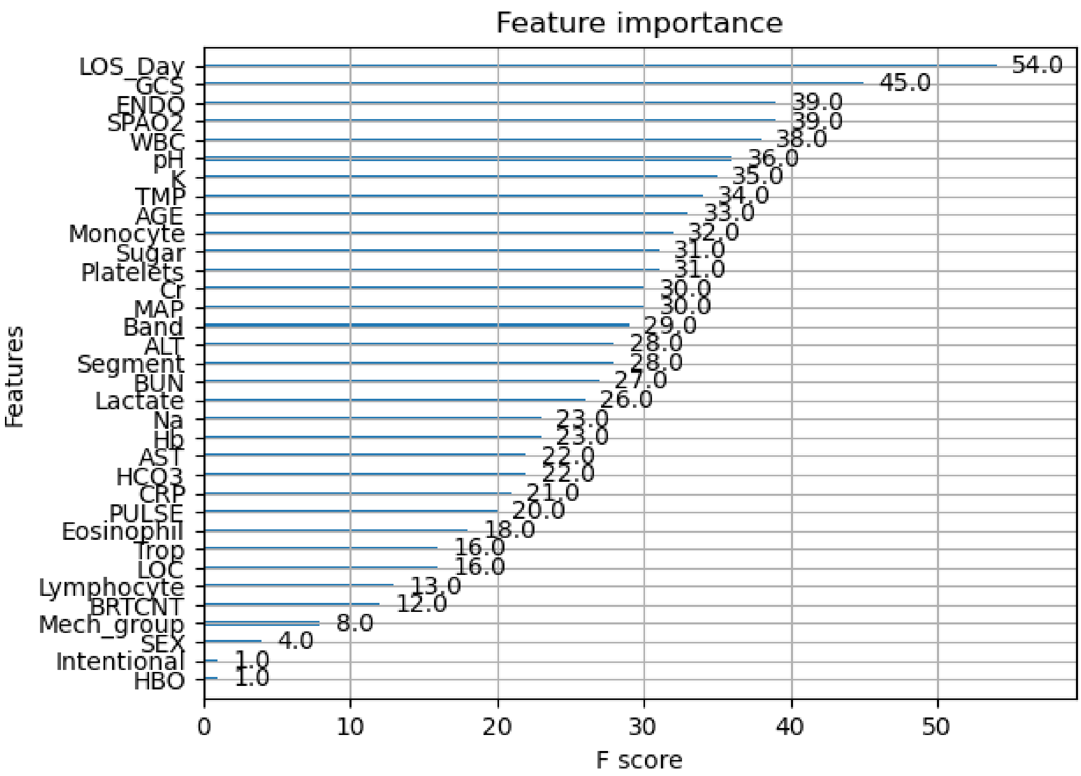

Figure Description: Feature selection in the XGBoost model for the external validation data.

3. Random Forest model

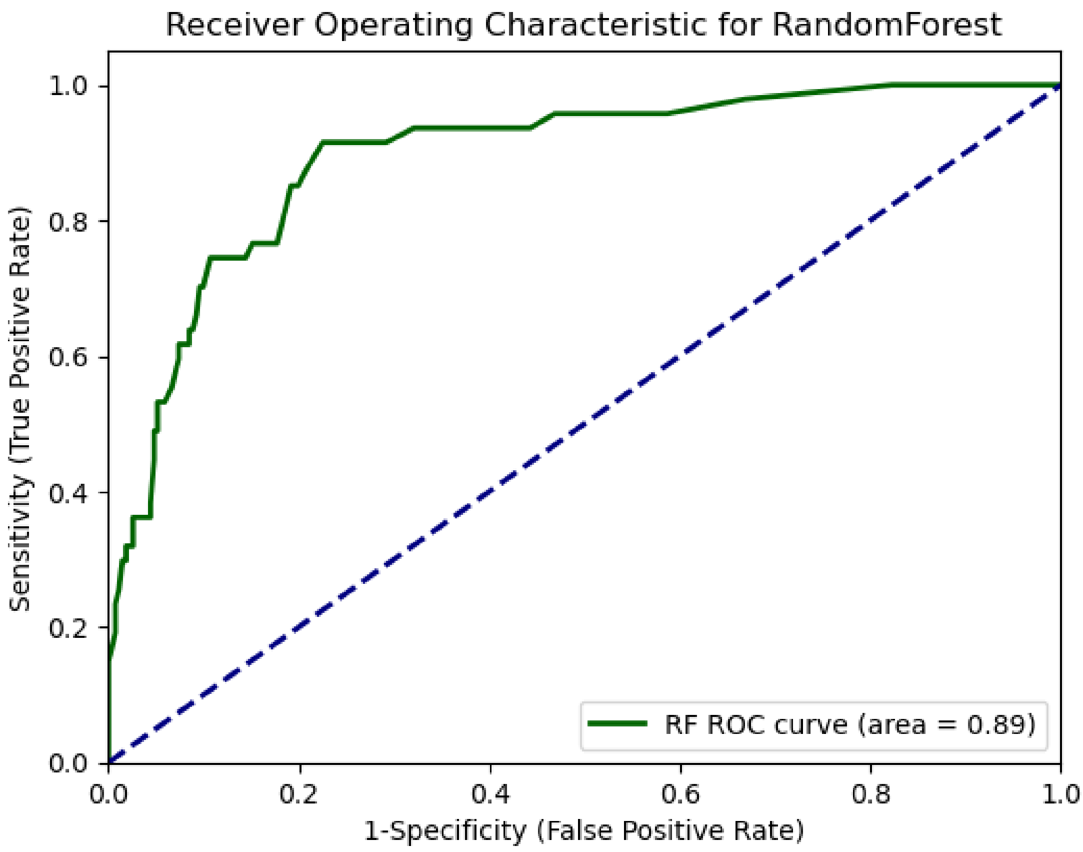

Figure Description: Performance of the Random Forest model on the external validation data. AUC = 0.89.

Table 2. The accuracy of external test data prediction in Random Forest model.				
	precision	recall	f1-score	support
0	0.93	0.93	0.93	271
1	0.59	0.62	0.60	47
accuracy = 0.88				318

Table Description: The accuracy of the external validation data in the Random Forest model is 88%.

4. MLP Classifier

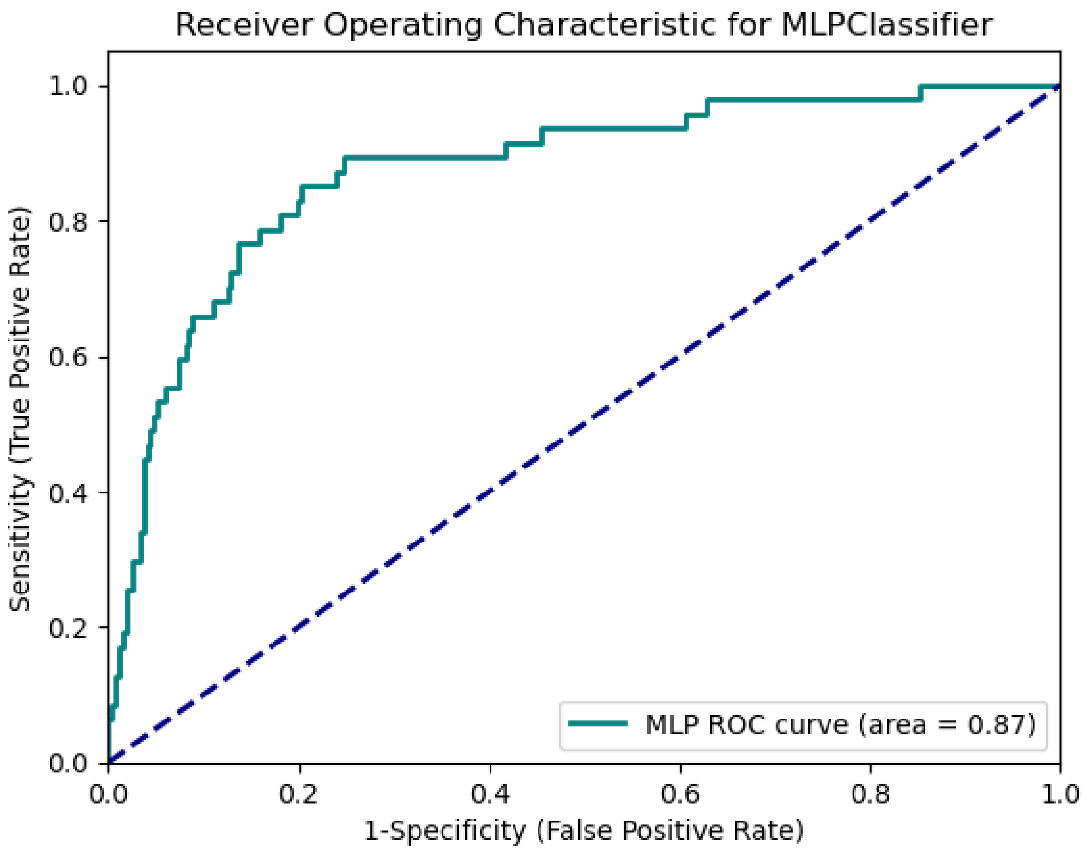

Figure Description: Performance of the MLP classifier on the external validation data. AUC = 0.87.

Discussion

The results of our study show that the FIRED score is still a strong tool for predicting outcomes in carbon monoxide poisoning (COP) cases, with an AUC of 0.93 in the external validation data. This performance is similar to that of the machine learning models we tested, including XGBoost (AUC = 0.89), Random Forest (AUC = 0.89), and MLP Classifier (AUC = 0.87). The fact that the FIRED score can perform so well, even without the complexity of machine learning, highlights its value in clinical settings.

However, the machine learning models did show potential, especially since they can automatically handle complex data and select important features. Despite this, their performance was only slightly lower than the FIRED score. One reason for this could be the lack of additional patient information, such as MRI results and medical history, which we plan to include in future research. We also acknowledge that our machine learning models were not fully fine-tuned, which may have affected their performance.

Conclusion

In this study, the FIRED score proved to be a reliable predictor for COP outcomes, with an AUC of 0.93, which is as effective as the machine learning models we tested. While machine learning has the potential to improve predictions, traditional scores like FIRED are still useful in clinical practice. Future work will focus on gathering more detailed data and refining our models to make predictions even more accurate.

Reference

[1] L. Eichhorn, M. Kieback, D. Michaelis, M. Kemmerer, B. Juttner, and K. Tetzlaff, "[Treatment of carbon monoxide poisoning in Germany : A retrospective single center analysis]," *Anaesthesist,* vol. 68, no. 4, pp. 208-217, Apr 2019, doi: 10.1007/s00101-019-0544-8. Behandlung von Kohlenmonoxidvergiftungen in Deutschland : Eine retrospektive Single-Center-Analyse.

[2] J. J. Rose *et al.*, "Carbon Monoxide Poisoning: Pathogenesis, Management, and Future Directions of Therapy," *Am J Respir Crit Care Med,* vol. 195, no. 5, pp. 596-606, Mar 1 2017, doi: 10.1164/rccm.201606-1275CI.

[3] M. Juttner *et al.*, "[Evaluation and assessment of the health care process in patients with carbon monoxide poisoning in Germany]," *Z Evid Fortbild Qual Gesundhwes,* vol. 140, pp. 1-13, Mar 2019, doi: 10.1016/j.zefq.2018.12.002. Evaluation und Beurteilung des Versorgungsprozesses von Patienten mit Kohlenmonoxidvergiftung in Deutschland.

[4] C. C. Huang *et al.*, "Impact of Hyperbaric Oxygen Therapy on Subsequent Neurological Sequelae Following Carbon Monoxide Poisoning," *J Clin Med,* vol. 7, no. 10, Oct 13 2018, doi: 10.3390/jcm7100349.

[5] L. Eichhorn, M. Thudium, and B. Juttner, "The Diagnosis and Treatment of Carbon Monoxide Poisoning," *Dtsch Arztebl Int,* vol. 115, no. 51-52, pp. 863-870, Dec 24 2018, doi: 10.3238/arztebl.2018.0863.

[6] K. Sethuraman and S. R. Thom, "Hyperbaric oxygen should be used for carbon monoxide poisoning," *Br J Clin Pharmacol,* vol. 89, no. 3, pp. 939-941, Mar 2023, doi: 10.1111/bcp.15605.

[7] M. E. Hanley and P. H. Patel, "Carbon Monoxide Toxicity," in *StatPearls*. Treasure Island (FL), 2024.

[8] S. C. Liao, Y. C. Mao, K. J. Yang, K. C. Wang, L. Y. Wu, and C. C. Yang, "Targeting optimal time for hyperbaric oxygen therapy following carbon monoxide poisoning for prevention of

delayed neuropsychiatric sequelae: A retrospective study," *Journal of the neurological sciences,* vol. 396, pp. 187-192, Jan 15 2019, doi: 10.1016/j.jns.2018.11.025.

[9] X. M. Xu *et al.*, "Management of delayed encephalopathy after CO poisoning: An evidence-based narrative review," *Medicine (Baltimore),* vol. 98, no. 49, p. e18199, Dec 2019, doi: 10.1097/MD.0000000000018199.

[10] B. Juttner *et al.*, "S2k guideline diagnosis and treatment of carbon monoxide poisoning," *Ger Med Sci,* vol. 19, p. Doc13, 2021, doi: 10.3205/000300.

[11] I. S. Choi, "Delayed neurologic sequelae in carbon monoxide intoxication," *Archives of neurology,* vol. 40, no. 7, pp. 433-5, Jul 1983, doi: 10.1001/archneur.1983.04050070063016.

[12] S. B. Jeon *et al.*, "Acute Brain Lesions on Magnetic Resonance Imaging and Delayed Neurological Sequelae in Carbon Monoxide Poisoning," *JAMA Neurol,* vol. 75, no. 4, pp. 436-443, Apr 1 2018, doi: 10.1001/jamaneurol.2017.4618.

[13] S. Nah *et al.*, "Cerebral White Matter Lesions on Diffusion-Weighted Images and Delayed Neurological Sequelae after Carbon Monoxide Poisoning: A Prospective Observational Study," *Diagnostics (Basel),* vol. 10, no. 9, Sep 16 2020, doi: 10.3390/diagnostics10090698.

[14] S. Choi, S. Nah, and S. Han, "Correlation between Time to Hyperbaric Oxygen Therapy and Delayed Neurological Sequelae in Acute Carbon Monoxide Poisoning Patients," *Diagnostics (Basel),* vol. 14, no. 2, Jan 15 2024, doi: 10.3390/diagnostics14020186.

[15] H. Lee, H. Kang, B. S. Ko, J. Oh, T. H. Lim, and Y. Cho, "Initial creatine kinase level as predictor for delayed neuropsychiatric sequelae associated with acute carbon monoxide poisoning," *Am J Emerg Med,* vol. 43, pp. 195-199, May 2021, doi: 10.1016/j.ajem.2020.02.054.

[16] Y. Zhang, Q. Lu, J. Jia, D. Xiang, and Y. Xi, "Multicenter retrospective analysis of the risk factors for delayed neurological sequelae after acute carbon monoxide poisoning," *Am J Emerg Med,* vol. 46, pp. 165-169, Aug 2021, doi: 10.1016/j.ajem.2020.06.090.

[17] S. H. Kim, Y. Lee, S. Kang, J. H. Paik, H. Kim, and Y. S. Cha, "Derivation and Validation of a Score for Predicting Poor Neurocognitive Outcomes in Acute Carbon Monoxide Poisoning," *JAMA network open,* vol. 5, no. 5, p. e2210552, May 2 2022, doi: 10.1001/jamanetworkopen.2022.10552.

[18] I. J. Wang *et al.*, "Poison severity score and sequential organ failure assessment score: Carbon monoxide poisoning prognosis," *PLoS One,* vol. 14, no. 3, p. e0212025, 2019, doi: 10.1371/journal.pone.0212025.

[19] Y. J. Chi, H. Y. Pan, F. J. Cheng, Y. I. Chang, and P. C. Chuang, "Experience of carbon monoxide poisoning and the outcome predicting score: A multicenter retrospective study," *Am J Emerg Med,* vol. 58, pp. 73-78, Aug 2022, doi: 10.1016/j.ajem.2022.05.012.

[20] S. Yang, H. Liu, Q. Peng, J. Li, and Q. Liu, "Predicting scale of delayed neuropsychiatric sequelae in patients with acute carbon monoxide poisoning: A retrospective study," *Am J Emerg Med,* vol. 52, pp. 114-118, Feb 2022, doi: 10.1016/j.ajem.2021.10.056.

[21] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.

[22] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the*

22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785-794.

[23] G. E. Hinton, "Connectionist learning procedures," in *Machine learning*: Elsevier, 1990, pp. 555-610.

[24] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] M. S. Tsai *et al.*, "Chang Gung Research Database: A multi-institutional database consisting of original medical records," *Biomed J*, vol. 40, no. 5, pp. 263-269, Oct 2017, doi: 10.1016/j.bj.2017.08.002.

[26] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, "k-Nearest Neighbor Classification," in *Data Mining in Agriculture*. New York, NY: Springer New York, 2009, pp. 83-106.