Prediction of Intradialytic Hypotension by Machine Learning: A Systematic Review

Nasrin Nikravangolsefid, M.D., M.P.H. 1,2, Jacob Ninan, M.D. 1, Hong Hieu Truong, M.D.1,2, Mariam Charkviani, M.D.1, Raghavan Murugan ,M.D.3, Gilles Clermont, M.D. 3, Kianoush B Kashani, M.D., M.S.1,2, , Juan Pablo Domecq Garces, M.D.1

1. Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Rochester, MN, USA
2. Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mayo Clinic, Rochester, MN, USA
3. Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Introduction

- Intradialytic hypotension (IDH) is a common complication associated with increased morbidity, mortality, and cardiovascular events.
- Several machine learning (ML) algorithms have been recently developed to predict IDH.
- We aimed to systematically review the ML models employed to predict IDH, their performance, methodological integrity, and clinical applicability.

Methods and Materials

- Pre-established protocol registered at the International Prospective Register of Systematic Reviews (PROSPERO ID: CRD42022362194).
- A comprehensive search: across 6 databases from inception to July 20, 2023.
- Two independent investigators (NN and JN) reviewed the articles, extracted data, and evaluated the risk of bias using the Prediction model Risk of Bias Assessment Tool (PROBAST).

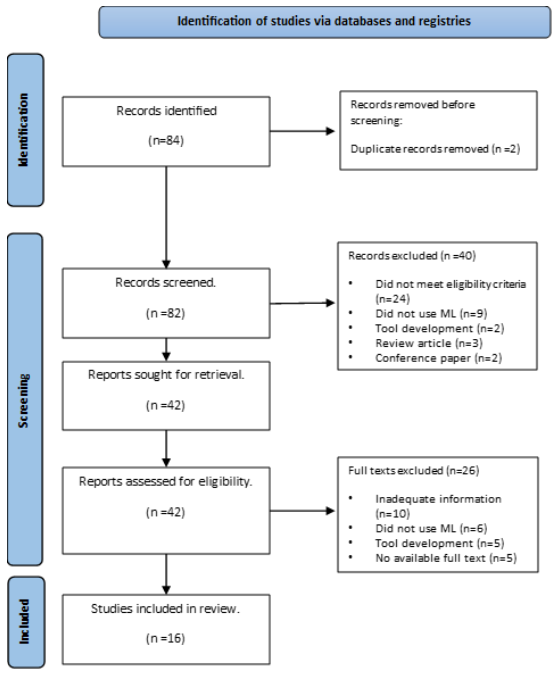

Figure 1. PRISMA flow chart

Table 1. Details of Machine Learning Model Assessment of Included Studies

Author/Year	ML models	Model Performance Evaluation	ML model: AUROC	Calibration	Validation	
					Internal	External
Gabutti et al. /2004	ANN	Sensitivity. Specificity, PPV	N/A	No	No	No
Gabutti et al./2005	ANN	AUROC	ANN: 0.68	No	Yes	No
Lin et al./ 2018	LASSO	AUROC, Sensitivity, Specificity	LASSO: 0.92	No	Yes	No
Huang et al./2020	Linear Regression RF, XGB, SVR, LASSO, Ensemble	R2, RMSE, MAE	N/A	No	No	No
Kang et al./2021	LR SVM DNN LGBM XGB	AUROC, Accuracy, F1 Score, Recall, Precision, F2 score, Specificity MCC	**XGB 0.83** DNN 0.82 SVM 0.81 LGBM 0.81 LR 0.81	Yes	Yes	No
Lee et al./2021	MLP LGBM LR RNN	AUROC, AUPRC, F1 score	**RNN: 0.94** LGBM: 0.93 MLP: 0.93 LR: 0.92	Yes	Yes	No
Tung et al./2021	GRU DNN LSTM	MAE, Accuracy, Specificity, Sensitivity	N/A	No	Yes	No
Elbasha et al./2022	ANN	AUROC, AUPRC, Precision, Recall, Accuracy, F1 score	ANN: 0.98	No	No	No
Kim et al./2022	LR XGB RF Deep learning (CNN)	AUROC AUPRC	**DL: 0.90** LR: 0.90 XGBM: 0.89 RF: 0.89	No	Yes	Yes
Mendoza- Pitti et al. /2022	LR RF MLP XGB	AUROC AUPRC F1 score Accuracy MCC	**XGB: 0.97** RF: 0.93 MLP: 0.93 LR: 0.85	No	Yes	No
Othman et al./2022	MLP, DT, KNN, SVCL, SVCR, RF, GBM	Accuracy, F1-score, Precision, Recall	N/A	Yes	Yes	No
Bae et al./2022	MLP	Accuracy, Sensitivity, Precision, MCC	N/A	No	No	No
Li et al./2022	bCOWOA-KELM	Accuracy, Specificity, Precision, F-measure	N/A	No	Yes	No
Dong et al./2023	LGBM SVM XGB MLP LDA TabNet	AUROC/ C-statistics	**LGBM: 0.82** LR: 0.81 TabNet: 0.80 XGB: 0.79 MLP: 0.75 SVM: 0.61 **XGB: 0.68** LGBM: 0.68 TabNet: 0.61 LR: 0.54 MLP: 0.52 SVM: 0.49	No	Yes	No
Lee et al./2023	RF XGB LR Deep learning	Accuracy, Recall, Specificity, Precision, F1 score, MCC, AUPRC, AUROC, NPV	**DL: 0.90** XGB: 0.87 RF: 0.86 LR: 0.85	No	Yes	No
Zhang et al./2023	XGB	AUROC, AUPRC, Precision	XGB: 0.88	Yes	Yes	No

Results

- Out of 84 screened articles: 16 studies included
- Inter-rater agreement: near perfect (Cohen's Kappa: 0.83, total agreement:91.6%)
- 14 studies: Retrospective, 1: Retrospective+ Prospective, 1: Prospective
- 10 studies: Asia (South Korea, Taiwan, China), 3: Europe (Spain, Switzerland), 2: Africa (Egypt) and 1 :USA
- 14,500 adult patients on HD and 2,349 patients on CRRT
- 12 studies: ESRD. 1:AKI, 3: did not mention population type
- 14 studies (87.5%): high risk of bias
- IDH prevalence reported between 1.2% and 51%
- IDH predictors: varied among studies (6 to 99 variables)
  - Vital signs (15 studies)
  - Dialysis setting measures (15 studies)
  - Demographics (12 studies)
  - Laboratory tests (12 studies)
- 38 ML models used to predict IDH
  - Neural Network models: in 13 studies
  - XGBoost: in 8 studies
  - Random Forest (RF) and Support vector machine (SVM) : in 4 studies
- AUROC ranged from 0.68 to 0.98
  - Artificial Neural Network (ANN) and XGBoost: the highest AUROCs of 0.98 and 0.97
- 12 studies: Internal validation
- 1 study: Internal and External validation, 4 studies: Calibration

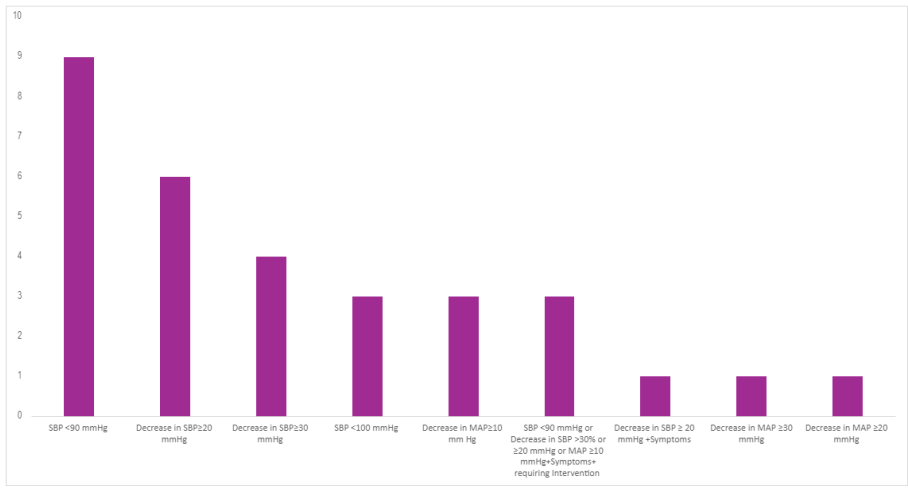

Figure 2. IDH definitions used in studies

Table 2. Results of Risk of Bias Assessment using the PROBAST tool

Author	Domain 1 (Participants)		Domain 2 (Predictors)		Domain 3 (Outcomes)		Domain 4 (Analysis)	Overall judgement	
	Risk of Bias	Applicability	Risk of Bias	Applicability	Risk of Bias	Applicability		Risk of Bias	Concern for Applicability
Gabutti et al.	High	High	Unclear	Unclear	Low	Low	High	**High**	**High**
Gabutti et al.	High	High	Low	Low	Low	Low	High	**High**	**High**
Lin et al.	Low	Unclear	Low	Low	High	High	High	**High**	**High**
Huang et al.	Low	Low	High	High	Unclear	Unclear	High	**High**	**High**
Kang et al.	High	Unclear	Low	Low	Low	Low	Unclear	**High**	Unclear
Lee et al.	Low	Unclear	Low	Low	Low	Low	Low	Low	Unclear
Tung et al.	High	Unclear	Low	Low	Unclear	Unclear	High	**High**	**High**
Elbasha et al.	High	High	Low	Low	Low	Unclear	Unclear	**High**	**High**
Kim et al.	Low	Unclear	Low	Low	Low	Low	Low	Low	Unclear
Mendoza-Pitti et al.	High	Unclear	High	Unclear	High	Unclear	Unclear	**High**	Unclear
Othman et al.	High	High	Unclear	Low	Low	Low	High	**High**	**High**
Bae et al.	High	High	Low	Low	Low	Low	High	**High**	**High**
Li et al.	High	High	Low	Low	Low	Low	High	**High**	**High**
Dong et al.	High	High	Low	Low	Low	Low	High	**High**	**High**
Lee et al.	High	Unclear	Low	Low	Low	Low	High	**High**	**High**
Zhang et al.	High	High	Low	Low	Low	Low	High	**High**	**High**

Conclusions

- ML models demonstrated strong performance in predicting IDH, offering the potential to alert clinicians and facilitate timely interventions for IDH prevention.
- The certainty of the current body of evidence is low due to
  - High risk of bias of the included studies
  - High heterogeneity among them
- External validation in multiple settings, accompanied by sample size calculation, model calibration, and the utilization of a consistent and uniform definition for IDH, is essential to enhance the robustness and applicability of these models.

THE 29TH INTERNATIONAL CONFERENCE ON
ADVANCES IN CRITICAL CARE NEPHROLOGY
AKI & CRRT 2024
MARCH 12-15, 2024 SAN DIEGO, CALIFORNIA