



OPEN AI-powered model for predicting mortality risk in VA-ECMO patients: a multicenter cohort study

Shuai Wang^{1,7}, Sichen Tao^{2,7}, Ying Zhu¹, Qiao Gu¹, Peifeng Ni^{3,4}, Weidong Zhang⁶, Chenxi Wu⁶, Ruihan Zhao⁵, Wei Hu^{1✉} & Mengyuan Diao^{1,3,4,6✉}

Veno-arterial extracorporeal membrane oxygenation (VA-ECMO) is a critical life support technology for severely ill patients. Despite its benefits, patients face high costs and significant mortality risks. To improve clinical decision-making, this study aims to develop a non-invasive, efficient artificial intelligence (AI)-enabled model to predict the risk of mortality within 28 days post-weaning from VA-ECMO. A multicenter, retrospective cohort study was conducted across five hospitals in China, including all the patients who received VA-ECMO support between January 2020 and January 2024. Based on the innovatively selected 25 easily obtainable patient examination features as potentially relevant, this study involved developing ten predictive models using both classical and advanced machine learning techniques. The model's performance is evaluated using various statistical metrics and the optimal predictive model are identified. Feature correlations are analyzed using Pearson correlation coefficients, and SHapley Additive exPlanations (SHAP) are employed to interpret feature importance. Decision curve analysis is used to evaluate the clinical utility of the predictive models. The study included 225 patients, with 66 patients from one hospital forming the training cohort. Three validation cohorts were used: internal validation with 16 patients from the training hospital and external validation with 30 and 60 patients from the other 4 hospitals. The random forest model emerged as the best predictor of 28-day mortality, achieving an AUROC of 1.00 in the training cohort and 1.00, 0.97, and 0.93 in the three validation cohorts, respectively. Despite the limited training data, the developed model, eCMoML, demonstrated high accuracy, generalizability and reliability. The model will be available online for immediate use by clinicians. The eCMoML model, validated in a multicenter cohort study, offers a rapid, stable, and accurate tool for predicting 28-day mortality post-VA-ECMO weaning. It has the potential to significantly enhance clinical decision-making, helping doctors better assess patient prognosis, optimize treatment plans, and improve survival rates.

Keywords VA-ECMO, Medical artificial intelligence, Medical machine learning model, Mortality risk prediction

Veno-arterial extracorporeal membrane oxygenation (VA-ECMO) can provide temporary cardiac and respiratory life support for patients with cardiogenic shock that is refractory to conventional medical therapy, allowing myocardial recovery or serving as a bridge during the wait for heart transplantation or left ventricular assist device (LVAD) implantation^{1–3}. Despite the increasing use of VA-ECMO, its costs and complication rates remain high⁴. Many patients ultimately die⁵, or suffer long-term sequelae that severely impact their quality of life^{6,7}. Inappropriate use of VA-ECMO increases resource utilization and hospital costs and is associated with high mortality rates; therefore, it is necessary to identify risk factors for death early on and make predictions. However, so far there have been few studies reporting on mortality-related factors and long-term prognosis in patients with severe cardiogenic shock treated with VA-ECMO^{8–11}. Therefore, developing an efficient model for predicting mortality risk in this context is particularly important.

¹Department of Critical Care, Affiliated Hangzhou First People's Hospital, School of Medicine, Westlake University, Hangzhou 310006, China. ²Faculty of Engineering, University of Toyama, Toyama-shi 930-8555, Japan. ³Department of Critical Care, Zhejiang University of Medicine, Hangzhou 310006, China. ⁴Department of Critical Care, Hangzhou First People's Hospital, Hangzhou 310006, China. ⁵School of Mechanical Engineering, Tongji University, Shanghai-shi 200082, China. ⁶Department of Critical Care, The Fourth School of Clinical Medical, Zhejiang Chinese Medical University, Hangzhou First People's Hospital, Hangzhou 310006, China. ⁷Shuai Wang and Sichen Tao contributed equally to this work. ✉email: huwei@hospital.westlake.edu.cn; diaomengyuan@hospital.westlake.edu.cn

The deterioration of cardiac function affects all organs, including the liver. Systemic congestion and hypoperfusion caused by cardiogenic shock lead to liver dysfunction and increase markers of hepatic fibrosis¹². Studies have shown that reduced arterial perfusion and systemic congestion can cause liver dysfunction and fibrosis. Liver dysfunction in patients with cardiogenic shock is a condition known as cardiohepatic syndrome, which significantly impacts clinical outcomes and mortality risk^{13,14}. The Fibrosis-4 Index (FIB-4), which includes age, aspartate aminotransferase (AST), alanine aminotransferase (ALT), and platelet (PLT) count, is a non-invasive marker for hepatic fibrosis in chronic liver disease patients¹⁵. Recently, researchers have found that FIB-4 is a predictive marker for mortality and rehospitalization in cardiac disease patients^{16–18}. Professor Shibata N.'s research discovered that high FIB-4 at admission is an important predictor of all-cause mortality and rehospitalization in heart failure¹⁹. In recent years, the Fibrosis-5 Index (FIB-5), calculated using albumin (Alb), alkaline phosphatase (ALP), AST/ALT ratio, and PLT count has been considered a simple and easily obtainable non-invasive marker for hepatic fibrosis²⁰. Research by Japanese scholar Daichi Maeda et al. found that FIB-5 was significantly associated with poor prognosis in patients with cardiogenic shock, proving it to be a useful risk stratification marker²¹.

In recent years, several predictive models have been developed, but due to the difficulty in obtaining data, only a few models have undergone external validation and their predictive performance has only reached a moderate level^{1,22}. Therefore, there is still a need to develop new high-precision predictive models based on external validation. Additionally, using easily accessible patient characteristic information as model input is crucial for clinical applications. This will help cover the widest range of patients, minimize diagnostic time, promote patient self-health management, and achieve highly economical medical services. Machine learning (ML) algorithms predict clinical patient survival by analyzing high-dimensional predictors and their complex interactions and have become one of the most promising solutions for such data mining tasks. In this study, utilizing the multi-center VA-ECMO patient cohorts, based on 25 easily obtainable patient features, we establish an ML model, called eCMoML, that accurately predicts the risk of death within 28 days after VA-ECMO withdrawal. Since all feature variables required for eCMoML prediction can be obtained through simple blood tests and clinical inquiries, it can serve as a non-invasive and cost-effective risk calculator used in routine clinical work for real-time prediction of mortality risk. This will provide doctors with important clinical decision-making references to optimize prognosis assessment and treatment plans while improving patient survival rates. By setting up one internal validation cohort and two external validation cohorts from five hospitals' multi-center cohorts, we further validated the generalizability of eCMoML.

Method

Study population

This multicenter retrospective study included patients who received VA-ECMO treatment from January 2020 to January 2024 at five hospitals in China. Patients were divided into a training cohort and three validation cohorts. The training cohort consisted of patients from the First People's Hospital of Hangzhou; the internal validation cohort was part of the patient group from the First People's Hospital of Hangzhou; two independent external validation cohorts came from Huzhou Central Hospital, Hangzhou Geriatric Hospital, and Yuhang District Third People's Hospital (Validation Cohort 1) and the Second Affiliated Hospital of Zhejiang University (Validation Cohort 2).

Ethics statement

This research protocol complies with the ethical requirements of the 1975 Helsinki Declaration and has been approved by the ethics review committees of all participating hospitals. The patients/participants provided their written informed consent to participate in this study.

Inclusion and exclusion criteria

Main criteria

The study included all patients who received VA-ECMO support. In this study, we included all patients who underwent VA-ECMO treatment in five tertiary hospitals over the past five years (2020–2024) and screened individuals meeting the analysis criteria based on strict standards. The specific inclusion criteria were as follows. First, adult patients aged more than 18 years (pediatric ECMO was not included due to its distinct pathophysiological characteristics). Second, patients who successfully underwent VA-ECMO treatment and were successfully weaned from ECMO. Third, ECMO runtime more than 24 hours to exclude patients with short-term ECMO use that might not fully exert its effects, thereby avoiding potential bias in the analysis results. Fourth, availability of complete perioperative clinical data and laboratory test records, including biochemical parameters, hemodynamic indicators, inflammation-related markers, and organ function scores (such as the APACHE-II score), to ensure the quality of input variables in the model. Fifth, availability of complete 28-day follow-up data to determine the survival status of patients after weaning, ensuring the accuracy of the study objectives. Furthermore, to enhance the rigor of the study, we ensured that data collection adhered to the principles of Real-World Study. First, all VA-ECMO-related data were sourced from the Hospital Information System, the Intensive Care Unit Electronic Health Records, and records maintained by the multidisciplinary ECMO management team, thereby minimizing selection bias to the greatest extent possible. Second, the data collection process complied with the international TRIPOD+ AI guidelines (TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods) and strictly followed standardized data processing procedures to ensure consistency across multiple centers²³.

In below's subsections, we provide some additional necessary explanations regarding variable selection criteria for the reader's reference.

Overall principles for feature variable selection

The selection of variables in this study is based on the following core principles. First, incorporating medical consensus to ensure clinical relevance. All included variables have been considered potentially related to patient prognosis in previous ECMO-related studies or have been widely used in critical care for disease assessment and risk stratification. Second, ensuring data availability to enhance model applicability. Priority is given to easily accessible and standardized data from ICU monitoring and laboratory tests to ensure model reproducibility and clinical usability. Third, optimizing variable information utilization through machine learning methods. SHAP importance analysis and Pearson correlation analysis are employed to ensure that variables contribute meaningfully to prediction rather than serving as redundant information. Fourth, external validation assesses the robustness of the variables. Using multicenter data from five hospitals ensures the applicability of variables across different cohorts while avoiding center-specific biases that could affect model generalizability.

In terms of specific variable selection, we classified the 25 variables as follows and explained the reasons for their selection separately. Category 1: Medical History Variables (5 variables). Medical history variables are used to reflect the patient's baseline health status and long-term disease burden, which may affect ECMO prognosis. Gender is included because previous studies have shown that female ECMO patients have a higher risk of bleeding, which may impact survival rates. Age is one of the most commonly used indicators of ECMO prognosis, with older age generally associated with poorer survival rates. Diabetes is included because it affects vascular function and immune response, potentially increasing the risk of mortality in ECMO patients. Hypertension, a common chronic disease closely related to cardiovascular function, may influence hemodynamic stability during ECMO support. Smoking history is included because prior smoking may increase the risk of ECMO-related lung injury and is associated with cardiovascular disease risk. Category 2: Scoring Systems (3 variables). Scoring systems are used to comprehensively assess the severity of the patient's condition and enhance the interpretability of the model. The APACHE-II score is one of the most commonly used severity assessment systems in the ICU field. By integrating multiple physiological indicators, it can predict mortality risk in critically ill patients and is therefore included to provide comprehensive illness evaluation information. The FIB-4 score is used to assess liver fibrosis status and may play an important role in evaluating ECMO-related liver injury. The FIB-5 score has gained increasing attention in ICU research in recent years. It is associated with hepatic functional reserve and may have independent prognostic value in ECMO outcome prediction, thus it is also included. Category 3: Vital Signs and ECMO-Related Variables (3 variables). These variables are used to reflect the patient's physiological status and ECMO operation conditions. BMI is included because it may affect hemodynamic stability in ECMO patients, and some studies even support the "obesity paradox," which may influence ECMO prognosis. CPR duration is a key variable affecting ECMO prognosis, as prolonged cardiopulmonary resuscitation (CPR) duration indicates more severe ischemia–reperfusion injury, which is generally associated with lower survival rates. ECMO duration is also included because prolonged ECMO support may indicate increased disease severity and may be associated with a higher incidence of complications. Category 4: Hematological and Biochemical Indicators (14 variables). These variables are used to assess organ function, inflammatory status, and metabolic conditions. The AST/ALT ratio is included because liver injury is common in ECMO patients, and this ratio can be used to evaluate liver function. Alkaline phosphatase (ALP) reflects hepatobiliary function, and ECMO patients may experience cholestasis; thus, this indicator may have prognostic value. Albumin (ALB), as a comprehensive indicator of nutritional status and inflammatory response, has lower levels that may indicate poorer prognosis. Platelets (PLT) are used to assess coagulation function, and thrombocytopenia is common in ECMO patients, being associated with bleeding and thrombosis risks; therefore, it is included. BNP (B-type natriuretic peptide) is an important indicator of cardiac function, and its elevation suggests heart failure, which may have predictive value in ECMO patients. Gamma-glutamyl transferase (GGT) is an important marker of hepatobiliary function and is associated with inflammation and oxidative stress, hence its inclusion. Total bilirubin is used to assess liver function, and its elevation may indicate ECMO-related liver injury. Phosphate is a crucial component of cellular metabolism, and hypophosphatemia may affect myocardial contraction and neurological function, thus it is included. White blood cells (WBC) are key markers of inflammation and infection, and ECMO-related infections may impact patient survival. Neutrophils are an essential component of the immune system and are associated with inflammation and infection risk. Lymphocytes reflect immune status, and impaired immune function may influence ECMO prognosis. Eosinophils may be related to inflammatory responses and allergic diseases, and some studies suggest their role in ICU prognosis, thus they are included.

For overall principles for variable selection. In this study, variable selection is based on the following key principles.

Regarding data availability (Availability & Clinical Feasibility), we prioritize biochemical indicators that are easily obtainable in the ICU environment because they are objective and stable. Compared to dynamic physiological variables such as heart rate, blood pressure, and central venous pressure (CVP), which are more susceptible to interference, biochemical test results have stronger comparability across different hospitals. For example, although variables like blood pressure, heart rate, and CVP are important for ICU monitoring, they can be influenced by factors such as medications (vasopressors, sedatives), mechanical ventilation, and operational errors. In ECMO patients, these variables exhibit high variability and are affected by ECMO machine control, making them less reliable in reflecting the patient's true circulatory status. Therefore, we did not include these variables.

Variables related to clinical relevance must have been proven in existing ECMO studies or critical care literature to be potentially associated with mortality risk. For example, AST/ALT, lactate, and phosphate have been confirmed by multiple studies to be related to ECMO prognosis and were therefore included. Although composite scoring systems such as SOFA score and APACHE-II score are commonly used in clinical practice, some components (e.g., GCS score) may not be applicable during ECMO. Therefore, we selected key components (such as liver function, kidney function, and lactate levels) instead of the complete scores.

Regarding model interpretability. In medical artificial intelligence research, the clinical interpretability of models is crucial to ensure that doctors can understand how the model makes decisions. SHAP analysis (Shapley Additive Explanations) results show that certain laboratory indicators (such as FIB-4 and FIB-5) have high importance in the model. Therefore, retaining these variables allows the model to provide a more intuitive medical explanation when predicting ECMO prognosis. Additionally, these variables may represent important pathophysiological mechanisms (such as ECMO-related liver injury and chronic inflammatory burden), which have practical clinical significance for doctors interpreting the model's output.

Regarding the automatic feature selection capability of machine learning models (Feature Selection in ML), nonlinear models (such as XGBoost and Random Forest) can automatically assign feature weights. Therefore, compared to traditional statistical models (such as Cox regression), which have strict requirements for variable collinearity, ML models can tolerate a certain degree of variable redundancy and focus only on the most contributive variables through internal feature selection mechanisms. As a result, when selecting variables, we do not manually remove potentially relevant variables but instead allow the model to learn and autonomously filter out the most useful information.

Why indicators such as FIB-4 and FIB-5?

Regarding why indicators such as FIB-4 and FIB-5, which have some degree of overlap with other variables, are included as model inputs. FIB-4 and FIB-5 are liver fibrosis scoring systems calculated based on age and biochemical indicators (AST, ALT, PLT). Although they partially overlap with certain biochemical variables (such as AST and ALT), they still possess unique medical value. Their inclusion is based on the following considerations:

1. ECMO prognosis is closely related to liver function, and FIB-4 and FIB-5 may provide more comprehensive information than individual biochemical indicators. Liver injury in ECMO patients (ischemic hepatitis, ECMO-induced liver dysfunction) is an important factor in predicting ECMO prognosis. Studies have shown that high FIB-4 and FIB-5 scores may indicate a higher risk of ECMO-related liver dysfunction. Due to significant hemodynamic changes in the liver during ECMO, relying solely on AST and ALT may not fully reflect liver functional reserve and chronic damage. In contrast, FIB-4 and FIB-5 incorporate factors such as age and platelet count, offering a more holistic assessment.
2. The application of FIB-4 and FIB-5 in the ICU field is increasing, and their inclusion in ECMO research remains an emerging exploration. In recent years, the use of FIB-4 and FIB-5 in ICU scoring systems has gradually increased, with some studies suggesting that they may be effective tools for predicting the prognosis of patients with sepsis and acute liver injury. Currently, no studies have systematically evaluated the predictive value of FIB-4 and FIB-5 specifically for ECMO patients. Therefore, this study aims to explore whether these scores can play a role in ECMO patients. Through SHAP analysis, we found that FIB-4 and FIB-5 demonstrated high importance across multiple models, suggesting that they may serve as potential novel indicators for ECMO prognosis assessment and warrant further investigation.
3. FIB-4 and FIB-5 provide “nonlinear risk integration” information, enhancing model stability. In traditional statistical models (such as Cox), researchers typically use individual variables (e.g., AST, ALT), whereas in machine learning, the model can learn complex variable interactions. FIB-4 and FIB-5 are essentially “combinations of variables,” integrating age, platelets, AST, and ALT into a single score, which may outperform individual variables in certain cases. Even if some component variables (such as ALT and AST) have already been included, FIB-4 and FIB-5 may still provide additional risk stratification information. Therefore, we retained them during initial modeling.
4. Enhance the medical interpretability of the model. In the SHAP analysis, we found that FIB-4 and FIB-5 had high contributions in multiple models, suggesting that they may play an important role in ECMO prognosis prediction. Since these two scoring systems incorporate information from multiple variables (age, AST, ALT, PLT), they may better reflect the overall health status of ECMO patients compared to individual biochemical indicators. Incorporating these scores can help clinicians more intuitively understand the risk factors in ECMO prognosis prediction rather than relying solely on the model's “black-box” decisions.

Why indicators such as APACHE II?

Although the APACHE II score may overlap with some univariate variables, we still believe its inclusion has important clinical and methodological value, including: (1) the widespread application and acceptance of the APACHE II score in the ICU field; (2) the global assessment information it provides on patient condition; (3) the ability of machine learning models' feature selection mechanisms to automatically adjust redundant information; (4) enhanced clinical interpretability, making it easier for doctors to understand; and (5) potential prognostic value for ECMO independent of any single variable. A more detailed explanation is as follows:

1. As a classic tool for ICU prognosis assessment, the APACHE II score still holds significant clinical value. Since its introduction by Knaus et al. in 1985, APACHE II (Acute Physiology and Chronic Health Evaluation II) has been a standardized scoring system for ICU prognosis evaluation and is widely used in clinical practice and research. Although machine learning models in the AI era may develop more complex scoring systems, APACHE II remains one of the most widely accepted ICU scoring methods globally. Clinicians have a high level of understanding of it, and it has been validated as effective in multiple studies on critically ill patient prognosis assessment. Therefore, in this study, we included the APACHE II score to ensure the medical interpretability of our model and to make its predictive results more easily accepted and applied by clinicians.
2. The comprehensiveness of the APACHE II score allows it to provide global information different from that of a single physiological variable. The APACHE II score consists of multiple physiological and biochemical var-

- iables as well as chronic health conditions, including blood pressure, heart rate, PaO₂, serum creatinine, and GCS score, which may be related to ECMO prognosis. However, the APACHE II score is not merely a simple summation of these variables; rather, it is a comprehensive scoring system derived from statistical analysis and empirical weighting that can be used for an overall assessment of ICU patients' severity of illness. Compared with individual biochemical indicators (such as lactate or AST/ALT), the APACHE II score provides a systematic quantification of the patient's overall condition and therefore retains additional predictive value.
- Machine learning methods can automatically assess the importance of variables, avoiding redundancy that may affect model performance. In traditional statistical modeling (such as Cox regression), variable collinearity may lead to model instability, necessitating strict variable selection to prevent information redundancy. However, in modern machine learning methods (such as random forests and XGBoost), models can automatically adjust feature weights and select the most predictive variables through feature importance analysis (e.g., SHAP interpretation models). Therefore, even if there is overlap between the APACHE II score and certain individual variables, machine learning models can still balance their contributions, effectively utilizing the information provided by the APACHE II score without being negatively impacted by redundancy. In fact, SHAP analysis revealed that the APACHE II score had a high feature contribution across multiple models, indicating its independent predictive value in ECMO prognosis prediction.
 - Facilitating clinical application and interpretation by physicians enhances the model's acceptability. Although new scoring systems in the AI era may be more accurate, clinicians still heavily rely on the APACHE II score for daily ICU assessments. If our model can demonstrate the role of the APACHE II score in ECMO prognosis prediction, this score could be integrated into AI-driven decision support systems in the future, making the model more interpretable and increasing physicians' trust in its results. Additionally, during clinical translation of the model, incorporating traditional scoring systems (such as APACHE II) can reduce the learning curve, allowing ICU clinical teams to adopt and apply it more quickly.
 - The APACHE II score may have a specific interaction with ECMO prognosis. The mortality risk of ECMO patients is influenced by acute pathophysiological conditions (such as hypotension and acidosis), underlying diseases (such as chronic kidney disease and liver disease), and ECMO-related complications. As a comprehensive assessment tool, the APACHE II score may incorporate interactive information from these risk factors, thus potentially retaining independent value in ECMO prognosis prediction. Although some individual variables may already be included in the model, their predictive effect when entered separately may not be as strong as the integrated APACHE II score. Therefore, in this study, we retained the APACHE II score to explore its potential role in predicting ECMO prognosis.

Data collection

Upon admission, the following data were collected: age, gender, body mass index, cardiopulmonary resuscitation time, Simplified Acute Physiology Score (APACHE-II), hypertension, diabetes, smoking status as well as blood routine and biochemical indicators. FIB-4 and FIB-5 indices were also calculated. During hospitalization, the following variables were collected: start time of VA-ECMO and weaning time from VA-ECMO with follow-up until 28 days after weaning (mostly post-discharge follow-up). Some common clinical indicators such as blood pressure, heart rate, CVP, etc. were not collected in this study because of excessive confounding. This type of follow-up dataset has rarely been collected and included in research related to ECMO mortality prediction studies in the past.

Patient management

During the study period, all patients were managed for cardiogenic shock according to international guidelines. When VA-ECMO flow gradually recovered to a minimum of 1.0–1.5 L/min, echocardiography was performed to assess myocardial recovery. Weaning treatment was initiated when hemodynamic tolerance during the weaning trial was good, without the need for increased inotropic or vasopressor support, and when echocardiographic criteria (LVEF > 20–25%, time-velocity integral > 10 cm, lateral mitral annulus peak systolic velocity > 6 cm/s, satisfactory right ventricular systolic function without dilation²³) were met.

Overview of machine learning models

This study developed and evaluated 10 classical and advanced ML models to predict the mortality risk of VA-ECMO patients, including 5 advanced ensemble methods: Adaptive Boosting (ADB), Categorical Boosting (CatB), Light Gradient Boosting Machine (LGBM), Random Forest (RF), and Extreme Gradient Boosting (XGB); 1 widely used support vector method: Support Vector Machine (SVM); and 4 baseline methods: Decision Tree (DT), K-Nearest Neighbor (KNN), Logistic Regression (LR) and Gaussian Naive Bayes (GNB). Each model has its unique advantages and applicable scenarios. ADB improves overall prediction performance by combining multiple weak classifiers, increasing the weight of misclassified samples from the previous round in each iteration. Its advantage lies in being very effective for handling imbalanced datasets, significantly enhancing the performance of weak classifiers. CatB is an efficient gradient boosting decision tree algorithm developed by Yandex, particularly suitable for processing categorical features. Its advantages include good handling of categorical features and missing values, fast training speed, and prevention of overfitting. LGBM is a gradient boosting framework developed by Microsoft that optimizes speed and memory usage. Its advantage is high efficiency in handling large-scale datasets with fast training speed. RF makes predictions by constructing multiple decision trees and averaging their results, offering high robustness. Its advantage is effectively preventing overfitting while maintaining stable performance when dealing with large-scale datasets. XGB is an efficient gradient boosting framework that makes predictions through stepwise additive model optimization. Its advantages include excellent performance, suitability for large-scale data, and strong ability to handle missing data. XGB is very popular in renowned data science competitions like Kaggle due to its outstanding performance

and flexibility, and it excels in handling various complex problems. SVM classifies by finding the hyperplane with the maximum margin, suitable for high-dimensional data. Its advantage lies in its excellent performance on small samples and high-dimensional data. DT predicts by constructing a tree-like model and selecting optimal features for data partitioning. Its advantages are easy understanding and interpretation, making it suitable for handling nonlinear data. KNN classifies based on the nearest K neighbors without needing a training phase. Its advantage is simplicity and ease of use without assuming any distribution of the data. LR is a linear model that classifies by maximizing the likelihood function; its advantages are simplicity of the model, ease of implementation, and interpretation. GNB is based on Bayes' theorem assuming independence among features, suitable for continuous features; its advantages include fast computation speed and effectiveness on small datasets. In this study, we employed the most widely recognized optimization and search methods for various models, including Grid Search, Random Search, and Bayesian Optimization, to optimize the hyperparameters of different machine learning models and ensure optimal performance. To achieve this goal, we adopted three different hyperparameter optimization strategies. Grid Search is suitable for models with fewer hyperparameters by performing an exhaustive combination search to find the optimal parameter configuration. In this study, it was used for LR, DT, KNN, SVM, and ADB. Random Search is appropriate for models with a large hyperparameter space by randomly selecting parameter combinations to improve computational efficiency. In this study, it was applied to RF and XGB. Bayesian Optimization is ideal for high-dimensional complex models as it utilizes probabilistic modeling to efficiently identify the best hyperparameters. In this study, it was used for LGBM and CatB. In addition, GNB usually does not require the special use of optimization algorithms.

In the prognostic prediction study of ECMO patients, we face multiple challenges, including a relatively small patient population but high data quality, complex feature variable types, high requirements for clinical interpretability, and the need for models to balance stability and generalization ability. Therefore, our machine learning model selection strategy is not only data-driven but also highly integrated with medical knowledge. We adopt a "Knowledge Embedded Expert System" approach, enabling the model to more effectively learn ECMO prognostic risk factors while reducing dependence on large-scale data.

Criteria for selecting a machine learning model

To meet the specific needs of ECMO prognosis prediction, we selected 10 machine learning models, covering various types such as ensemble learning methods, support vector machines, and traditional statistical models, to ensure model robustness and broad applicability. First, we selected five ensemble learning methods, including Random Forest (RF), Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM), Categorical Boosting (CatB), and Adaptive Boosting (ADB). Among them, Random Forest (RF) constructs multiple decision trees and performs voting, effectively handling high-dimensional data while maintaining high interpretability. Extreme Gradient Boosting (XGB) adopts the gradient boosting algorithm, making it suitable for small-sample, high-dimensional data and capable of reducing the impact of redundant variables. Light Gradient Boosting Machine (LGBM) has higher computational efficiency compared to XGB, particularly for high-dimensional data. Categorical Boosting (CatB) is optimized for categorical variables, automatically handling missing values and preventing overfitting. Adaptive Boosting (ADB) improves model adaptability to small datasets by iteratively weighting weak classifiers. Next, to enhance model adaptability, we selected Support Vector Machine (SVM), which is well-suited for high-dimensional, small-sample learning and can effectively handle nonlinear classification problems, showing promising applications in ECMO prognosis prediction, where complex feature interactions exist. Additionally, we selected four benchmark methods, including Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB). Decision Tree (DT) is easy to interpret and suitable for handling nonlinear data, but a single decision tree is prone to overfitting. K-Nearest Neighbors (KNN) is based on the nearest neighbor rule and is suitable for cases where the data distribution follows clear patterns, but it has low computational efficiency. Logistic Regression (LR), as a classic binary classification model, is well-suited for linearly separable problems and serves as an important benchmark model. Gaussian Naive Bayes (GNB) assumes feature independence and is suitable for low-computation, fast prediction tasks. By selecting different types of models, we ensured a comprehensive exploration ranging from simple linear models to complex nonlinear ensemble learning methods, thereby optimizing the modeling strategy for ECMO prognosis prediction.

Optimization of machine learning models

To ensure optimal model performance, we employed three different hyperparameter optimization strategies. For models with fewer hyperparameters (LR, DT, KNN, SVM, ADB), we used Grid Search to perform an exhaustive search for the best parameters. For models with a large hyperparameter space (RF, XGB), we applied Random Search, which selects parameter combinations randomly to improve computational efficiency. For high-dimensional complex models (LGBM, CatB), we utilized Bayesian Optimization to leverage probabilistic models for more efficient identification of optimal hyperparameters.

Application of knowledge-embedded expert system approach

Traditional machine learning methods typically rely on large-scale data for training. However, ECMO patients represent a highly specific population with limited sample sizes, making purely data-driven approaches prone to poor generalization. Therefore, we adopted a knowledge-embedded expert system approach that integrates clinical medical knowledge with machine learning modeling to enhance predictive performance and medical interpretability. First, our variable selection is based on clinical consensus rather than purely data-driven methods. For example, we included APACHE-II score, FIB-5, CPR duration, and ECMO duration instead of allowing the model to blindly select from hundreds of variables. This approach enables the model to directly leverage existing clinical knowledge, reducing reliance on extremely large training datasets and improving learning efficiency.

Second, composite scores provide high-level information that enhances model learning efficiency. For instance, the APACHE-II score is not merely a simple weighted sum of multiple physiological indicators but rather the result of long-term clinical experience. It offers more stable risk assessment capabilities than individual variables. By learning these scores, the model can more rapidly capture ECMO prognosis patterns, improving learning efficiency. Furthermore, this approach reduces computational burden and enhances generalization ability. Since variable selection is already informed by clinical knowledge, the model does not need to learn from numerous irrelevant variables, thereby minimizing wasted computational resources. This improves generalization performance, making the model more effectively applicable to different patient cohorts.

Primary objective

The main objective of this study is to develop an economical, rapid, high-precision, and highly generalizable model for predicting the risk of death within 28 days after VA-ECMO weaning based on easily accessible patient characteristics.

Statistical analysis

In the training queue, we selected 25 variables as input features for model training and learning. These variables include FIB-4, FIB-5, AST/ALT, ALP, ALB, PLT, AST, ALT, BMI, Gender, Age, CPR time, ECMO time, Diabetes, Hypertension, Smoking, APACHE-II, BNP, GGT, Total bilirubin, Phosphate, WBC, Neutrophil, Lymphocyte, and Eosinophil. In the training data, the proportion of different labels was ensured to be similar to reduce the impact of class imbalance on training effectiveness. To deeply understand each model's internal prediction mechanism and its dependence on features, we used SHapley Additive exPlanations (SHAP) for interpretive analysis of the developed models. SHAP is an interpretation method based on Shapley values that originated from cooperative game theory to allocate contributions of each participant to overall outcomes. When applied to machine learning models, SHAP provides a consistent and interpretable way to quantify feature importance by assigning contribution values of each feature to prediction results. We first developed and trained various models and then used model-specific SHAP interpreters for interpretability analysis. By calculating Shapley values for each feature, we obtained their contribution in predictions and could intuitively display positive or negative correlation impacts of each feature on prediction results. This is crucial for understanding the decision-making mechanisms of models and enhancing their transparency. The results can also help us identify potential data anomalies or model biases thereby further optimizing model performance. Additionally, Pearson correlation coefficients were used to obtain correlations between every pair of variables.

To comprehensively evaluate the diagnostic performance of machine learning models, we calculated and summarized the Area Under the Receiver Operating Characteristic Curve (AUROC), Sensitivity (SE), Specificity (SP), Positive Predictive Value (PPV), and Negative Predictive Value (NPV) across multiple cohorts. These performance metrics' distributions were estimated by resampling the data, allowing us to compute their 95% confidence intervals. To further assess the practical clinical value of the model, we employed decision curve analysis. Decision curve analysis evaluates a model's net benefit at different decision thresholds, helping determine its utility in actual decision-making processes. These analytical methods not only help reveal the diagnostic performance of models but also quantify their potential value in real-world clinical applications.

Results

Patient characteristics

A total of 225 patients received VA-ECMO support. Among them, 54 patients did not meet the inclusion criteria and were excluded. Finally, 172 patients were included in the analysis. According to the inclusion and exclusion criteria, there were 66 people (38.37%), 16 people (9.30%), 30 people (17.44%), and 60 people (34.88%) in the training cohort, validation cohort 1, validation cohort 2, and validation cohort 3 respectively who received VA-ECMO support for more than 24 hours (Fig. 1). In the training cohort, the number of death and survival samples is 31 and 35, respectively. In validation cohort 1, they are 8 and 8, respectively. In validation cohort 2, they are 20 and 10, respectively. In validation cohort 3, they are 36 and 24, respectively. Table 1 shows the baseline characteristics of each cohort. In the training cohort, the average age is 51.12 years, and the APACHE-II score is 26.09 ± 9.49 . Among them, 60.61% are male, and 15.15% have diabetes. In validation cohort 1, the average age is 51 years, with an APACHE-II score of 22.56 ± 8.29 ; 73.1% are male, and 6.25% have diabetes. In validation cohort 2, 73.33% are male with a median BMI of $24.73 \pm 2.63 \text{ kg/m}^2$, among which 23.33% have diabetes; in validation cohort 3, 75% are male, APACHE-II score is 29.47 ± 8.80 , BMI is $24.08 \pm 5.54 \text{ kg/m}^2$. In the 4 cohorts, the average FIB-4 values are 18.76, 44.66, 33.18, and 31.88 respectively. The average FIB-5 values are -16.80, -12.5, -16.92, and -17.2 respectively.

Development of machine learning models

We included a total of 25 patient characteristic variables. Figure 2 shows the Pearson correlation coefficients among the included features (clustered by Logical Regression learning based on correlation magnitude). It can be observed that AST is positively correlated with ALT, ALP with GGT, Diabetes with Hypertension, and Lymphocyte with PLT. Notably, we found that ALT, Neutrophil, WBC, ALB, and AST were positively correlated with ECMO time, indicating that these variables are also closely related to the duration of ECMO support. Additionally, ALT was negatively correlated with Age and Hypertension, suggesting that younger age and lower incidence of Hypertension are closely associated with lower ALT levels. Interestingly, we also found that FIB-4, FIB-5, and Phosphate were negatively correlated with ECMO time, providing a basis for our subsequent intervention studies. Based on these variables, we developed ten prediction models to predict the 28-day mortality risk after VA-ECMO decannulation. These models are based on ADB, CatB, DT, GNB, KNN, LR, LGBM, RF, SVM, and XGB, and their performance was tested in the four cohorts. The statistical results showed

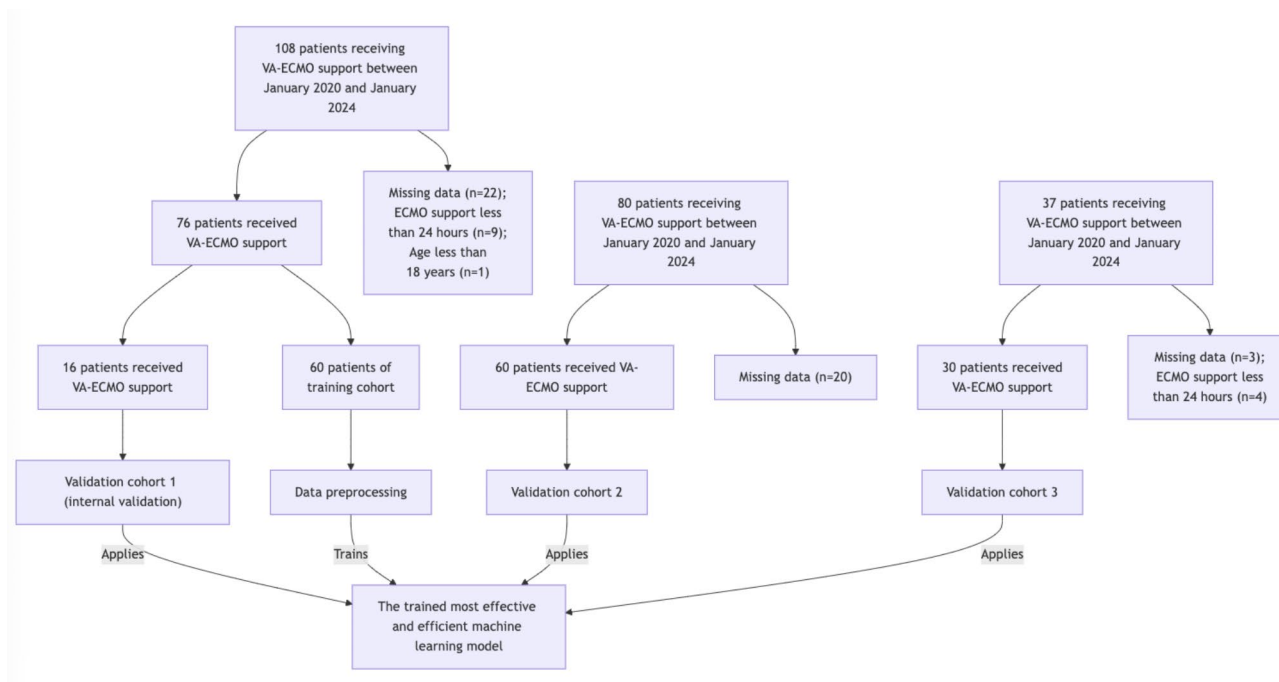


Fig. 1. Flow diagram of the study population.

that the AUROC range of these models in the training cohort was 0.66 to 1.0. In validation cohorts 1, 2, and 3, the AUROC ranged from 0.47 to 1.0, 0.52 to 0.97, and 0.54 to 0.93, respectively (see Fig. 3). From Table 2 and Fig. 3, it can be seen that the predictive performance of the KNN-based model was relatively low in both the training and validation cohorts. However, the prediction model based on RF achieved the highest AUROC (0.93–1) in all four cohorts, demonstrating the best performance.

Prediction performance of the RF-based model

In the training cohort, the accuracy, sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) of the RF model were all 1.0 (95% CI 1.0–1.0) (see Table 2). In validation cohort 1, the RF model also demonstrated excellent predictive performance, with accuracy, SE, and SP all being 1.0 (95% CI 1.0–1.0) (see Table 2). In validation cohort 2, the predictive accuracy of the RF model was 0.97, with SE and SP being 1.0 (95% CI 1.0–1.0) and 0.89 (95% CI 0.64–1.0), respectively (see Table 2). In validation cohort 3, the predictive accuracy of the RF model was 0.82, with SE being 0.83 (95% CI 0.70–0.95) and SP being 0.79 (95% CI 0.70–0.95) (see Table 2). The results of the RF prediction model were generally superior to those of the other nine models within the same cohort, preliminarily validating the outstanding predictive value of the RF-based model.

The relationship between variables and predicted results in the RF-based model

Since the RF-based prediction model performed the best in predicting the 28-day mortality risk of VA-ECMO patients after decannulation in both internal validation and multicenter external validation, its predictive paradigm deserves analysis and consideration as a clinical reference. In the decision curve analysis, the developed RF model demonstrated consistently high and stable net benefit across all probability thresholds in both the training cohort and the three validation cohorts (with a highly consistent pattern observed in the curves, see Fig. 4). This indicates that the model can provide reliable clinical decision support for patients across diverse populations and different medical centers, with significant practical value in decision-making. In the decision curve analysis, the developed RF model showed consistently high and stable net benefits across all probability thresholds in the training cohort and the three validation cohorts (see Fig. 4), indicating its high practical utility in decision-making. Decision curves for the other nine methods are shown in Appendix Fig. 13 through Appendix Fig. 22. Additionally, the SHAP method was used to interpret and analyze the relevance assigned to different input variables by the developed model. The results showed that BMI was significantly the most relevant factor in decision-making, followed by APACHE-II, FIB-5, BNP, Phosphate, and CRP time (see Fig. 5). According to the feature relevance heatmap of the RF model (see Fig. 6), we found that among the top three most relevant features, higher values of BMI and APACHE-II were positively correlated with higher mortality risk, while FIB-5 was negatively correlated with mortality probability. Other positively correlated mortality features included Phosphate, CRP time, FIB-4, ALP, Age, and Smoking, whereas BNP, PLT, WBC, Neutrophil, and ALT were negatively correlated with mortality. These findings not only help us understand the decision-making mechanism of the model but also provide valuable reference information for clinical practice, enabling better assessment of VA-ECMO patients' prognosis, optimization of treatment plans, and ultimately improving patient survival rates. SHAP feature relevance plots and heatmaps for other models are shown in Appendix Fig.

Variables	Training cohort (<i>n</i> = 66)	Validation cohort 1 (<i>n</i> = 16)	Validation cohort 2 (<i>n</i> = 30)	Validation 3 (<i>n</i> = 60)
Age, years	51.12 ± 17.55	51 ± 19.20	52.9 ± 15.25	50.25 ± 17.02
Gender, (<i>n</i> , %)	40, (60.61)	11, (68.75)	22, (73.33)	45, (75)
Smoking, (<i>n</i> , %)	22, (36.67)	5, (31.25)	4, (13.33)	26, 43.33
Diabete, (<i>n</i> , %)	10, (15.15)	1, (6.25)	7, (23.33)	8, (13.33)
Hypertension, (<i>n</i> , %)	25, (37.88)	9, (56.25)	15, (50)	17, (28.33)
BMI, kg/m ²	22.78 ± 2.81	23.17 ± 2.74	24.73 ± 2.63	24.08 ± 5.54
APACHE-II	26.09 ± 9.49	22.56 ± 8.29	31.5 ± 12.40	29.47 ± 8.80
ECMO time, days	7.03 ± 5.05	6.69 ± 4.11	4.93 ± 3.50	6.55 ± 3.15
CPR time, minutes	24.28 ± 43.18	17.06 ± 25.08	44.07 ± 47.32	38.05 ± 31.97
ALT, IU/L	524.80 ± 1032.98	402.25 ± 450.06	388.73 ± 525.58	425.53 ± 525.42
AST, IU/L	1000.12 ± 1451.91	982.75 ± 1386.42	667.8 ± 718.20	945.42 ± 892.95
AST/ALT	3.00 ± 1.95	2.26 ± 1.53	2.49 ± 1.56	2.90 ± 1.90
GGT, IU/L	60.21 ± 49.20	67.06 ± 71.18	63.9 ± 40.81	65.8 ± 38.39
ALP, IU/L	62.82 ± 31.16	79 ± 57.02	66.07 ± 34.53	61.17 ± 21.46
Albumin, g/L	27.34 ± 7.80	27.6 ± 7.70	21.45 ± 8.53	27.55 ± 7.96
Total bilirubina, umol/L	18.77 ± 14.98	20.24 ± 9.66	13.25 ± 9.43	27.25 ± 21.18
PLT, 10 ⁹ /L	176.41 ± 125.09	157.88 ± 106.80	132.47 ± 89.21	133.65 ± 80.29
WBC, 10 ⁹ /L	14.93 ± 7.79	13.06 ± 7.02	14.32 ± 9.42	15.34 ± 7.50
Neutrophil, 10 ⁹ /L	12.66 ± 7.06	11.16 ± 6.33	10.82 ± 7.79	13.25 ± 6.97
Lymphocyte, 10 ⁹ /L	3.62 ± 15.75	1.29 ± 1.10	1.76 ± 1.91	1.44 ± 1.45
Eosinophil, 10 ⁹ /L	0.06 ± 0.17	0.06 ± 0.14	0.27 ± 0.45	0.03 ± 0.06
Phosphate, mmol/L	1.32 ± 0.92	1.51 ± 0.80	7.26 ± 31.32	1.24 ± 0.64
BNP, pg/ml	8194.57 ± 10,163.61	7911.44 ± 9980.49	3759.17 ± 7662.86	7041.05 ± 9741.95
FIB-4	18.76 ± 19.49	44.66 ± 90.50	33.18 ± 46.08	31.88 ± 58.44
FIB-5	− 16.80 ± 12.61	− 12.50 ± 12.78	− 16.92 ± 10.87	− 17.20 ± 12.41
Death sample size	31	8	20	36
Survival sample size	35	8	10	24

Table 1. Baseline characteristics of the 4 cohorts. Values are expressed as (*n*, %) or mean ± standard deviation. HTN, Hyper-tension; BMI, Body Mass Index; APACHE-II, Acute Physiology and Chronic Health Evaluation; ECMO, Extracorporeal Membrane Oxygenation; CPR, Cardiopulmonary Resuscitation; ALT, Alanine Aminotransferase; AST, Aspartate Aminotransferase; GGT, γ -glutamyl Transpeptidase; ALP, Alkaline Phosphatase; ALB, Albumin; TB, Total Bilirubin; PLT, Platelet Count; WBC, White Blood Cell; BNP, Brain Natriuretic Peptide; FIB-4, The Fibrosis-4 Index; FIB-5, The Fibrosis-5 Index.

7 through Appendix Fig. 12. It is worth noting that some state-of-the-art machine learning models, such as ADB, CatB, and XGB, also achieved high stability and predictive accuracy. They all considered BMI as a highly relevant positive decision variable for mortality. This suggests that BMI should receive more attention in related clinical predictions and research.

In this study, we evaluated the performance of 10 machine learning models and found that ensemble learning based methods (such as Random Forest, XGBoost, and CatBoost) performed excellently on both the training set and external validation set, whereas some traditional machine learning methods (such as KNN and GNB) showed relatively weaker performance in predicting ECMO prognosis. However, we still decided to retain all models in the final results table based on the following considerations. First, providing benchmark comparison value: Some of the 10 different types of machine learning methods, despite their weaker predictive ability in this study, can serve as benchmark models for comparison with more complex algorithms. This helps highlight the practical advantages of advanced models (such as gradient boosting models) and informs future researchers about model performance levels to aid and inspire further studies. Second, revealing the impact of data characteristics: KNN relies on local distribution among samples, while GNB assumes conditional independence between features. Their lower performance suggests that ECMO prognosis prediction may exhibit highly nonlinear characteristics with potentially complex interactions between features. This observation aids future research in exploring feature engineering and modeling strategies for ECMO data more deeply. Third, maintaining research transparency and reproducibility: In medical AI research—as well as broader cutting-edge AI technology studies—fully presenting experimental results for all models enhances research transparency. It allows other researchers to reproduce or reference our experimental design and is considered a standard practice. Although some models performed poorly, their results still hold scientific significance and offer certain insights. Therefore, to provide a more comprehensive baseline analysis, enhance research transparency, offer references for future studies, and align with common practices in AI research papers, we have fully retained relevant key findings in this study.

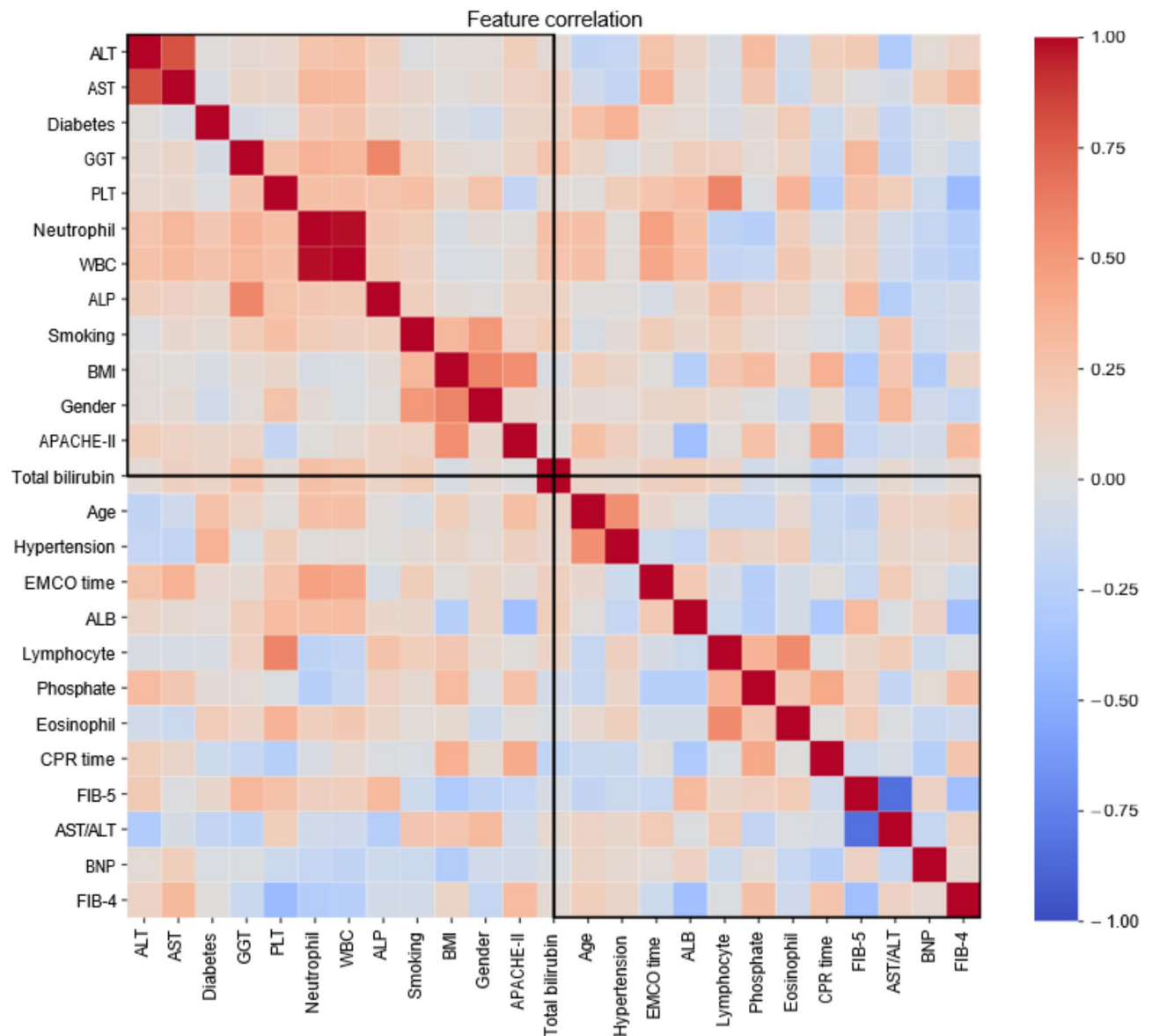


Fig. 2. Feature correlation heatmap between every two of the employed 25 easily obtainable patient feature variables.

Discussion Interpretation

In this study, we developed a medical model to predict the 28-day mortality risk of VA-ECMO patients after decannulation. The model based on RF showed the best performance and was established as an innovative and efficient AI-powered model that utilizes easily obtainable patient examination features to predict the mortality risk of VA-ECMO patients, termed eCMoML. The SHAP method applied in the decision relevance analysis aids in the multi-variable interpretability of eCMoML and the other nine developed prediction models, enhancing model transparency. Moreover, according to statistical test results, this model demonstrated robust predictive capabilities in both internal and external validations.

With the advancement of technology, the number of patients receiving ECMO support in China is gradually increasing, saving numerous lives. However, ECMO treatment is extremely costly, placing a burden on patients' families and society. Both patients' families and clinicians urgently need an easy-to-use model to predict the mortality of ECMO patients. In recent years, dozens of predictive models have been developed, but they suffer from issues such as small training sample sizes (fewer than 30 cases), being predominantly single-center cohorts, limited external validation, lower predictive performance, and difficulty in obtaining feature variables^{1,22}. Additionally, almost all models tend to underestimate mortality¹. To address these issues, we proposed, for the first time, using 25 easily obtainable laboratory or basic patient characteristics as feature variables for prediction in this study. We collected the most recent available samples and labels from 225 patients across five hospitals to construct an innovative multi-center and multi-cohort dataset. Furthermore, we developed a high-accuracy

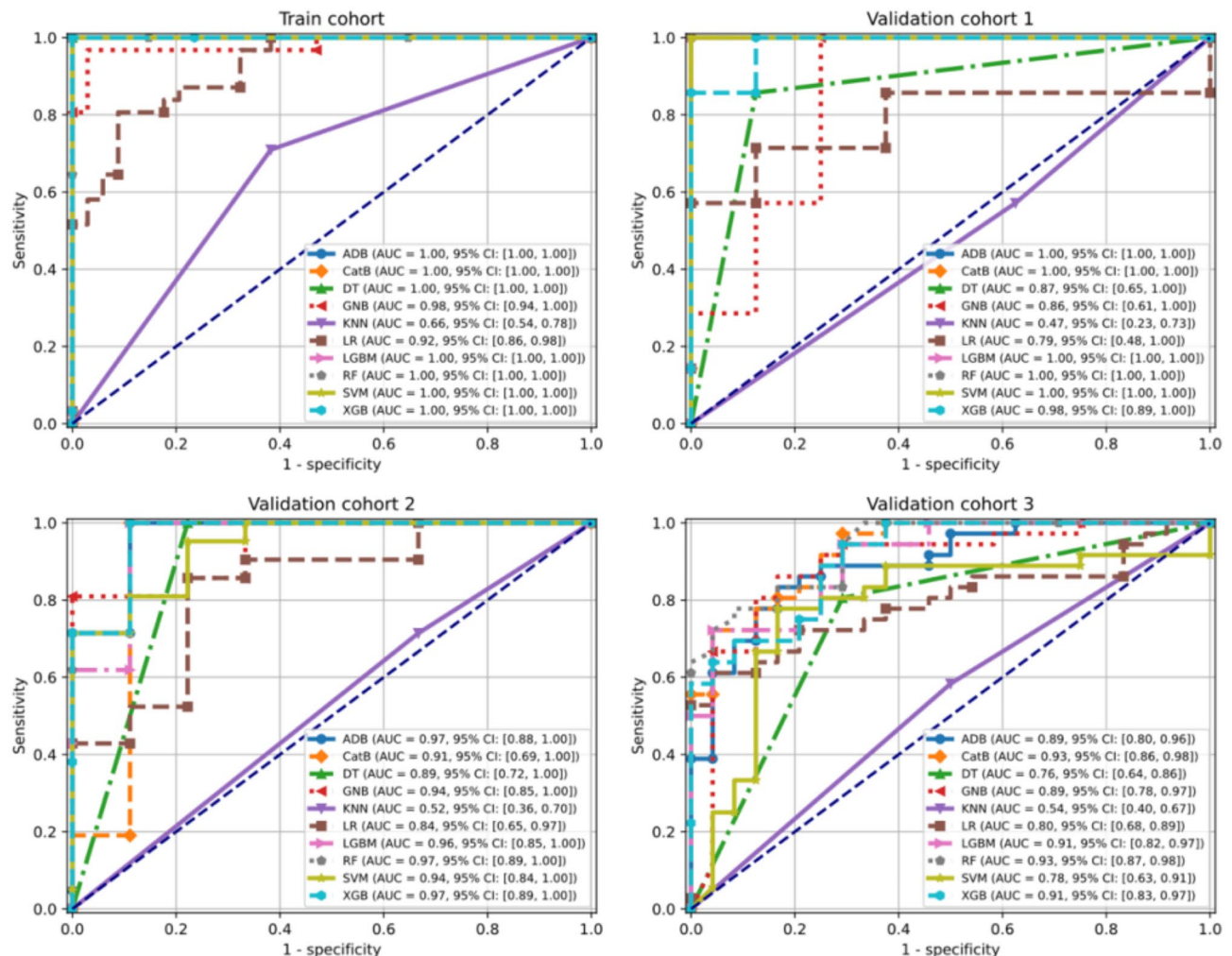


Fig. 3. The receiver operating characteristic curve plots of the developed machine learning models under the 4 cohorts.

predictive model (AUROC range up to 0.97–1) based on 10 classical and state-of-the-art models, and through statistical testing, we selected the best model (eCMoML) as a clinical decision support tool.

In most countries, including China, adult obesity rates are increasing year by year. Despite significant advances in our understanding of obesity, BMI remains the most commonly used indicator for diagnosing and classifying obesity²⁵. An increase in BMI is associated with higher blood pressure and diabetes, leading to an increased risk of coronary artery disease and stroke, as well as higher mortality rates in cardiovascular disease patients²⁶. In ICU patients, the concept of the obesity paradox has been proposed in recent years but is highly controversial. Our study found that a high BMI might result in a high mortality risk. In recent years, several studies have evaluated the relationship between obesity and sudden cardiac arrest. Some studies indicate that an elevated BMI may increase the risk of sudden cardiac arrest by 2–5 times^{27–29}. However, the impact of BMI on survival rates remains controversial. Some studies show that patients with higher BMI have lower survival rates^{30,31}, while others suggest no association between BMI and survival^{32,33}. In our dataset, most patients had cardiovascular disease. Since a high BMI represents a higher incidence of metabolic syndrome, this provides new evidence supporting the view that avoiding obesity can reduce the incidence of cardiovascular disease. Although the impact of BMI on the survival rates of VA-ECMO-supported patients has not been previously studied, our research fills this gap and finds that a high BMI likely contributes to increased mortality. Given that BMI affects not only the cardiovascular system but also multi-organ functions, leading to higher mortality risk, such related studies are crucial in clinical practice³⁴.

To provide more potentially useful reference information, we discuss three key issues and possible controversies regarding BMI in this study. First, there are two reasons why BMI was included in the model:

1. Basic physiological indicators and clinical experience. BMI has long been used to assess patients' metabolic status, body fat percentage, muscle mass, and inflammatory response and is widely applied in ICU prognosis research^{35,36}. In ECMO patients, BMI may affect prognosis by influencing hemodynamic stability, oxygenation efficiency, and the risk of complications such as deep vein thrombosis and pulmonary infection^{37,38}.

Model	AUROC	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Train cohort						
ADB	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
CatB	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
DT	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
GNB	0.98 (0.94–1.00)	0.89 (0.82–0.95)	0.81 (0.66–0.95)	0.97 (0.92–1.00)	0.96 (0.88–1.00)	0.85 (0.73–0.96)
KNN	0.66 (0.55–0.78)	0.66 (0.55–0.76)	0.71 (0.58–0.87)	0.62 (0.46–0.77)	0.63 (0.45–0.79)	0.70 (0.53–0.86)
LR	0.92 (0.85–0.97)	0.83 (0.74–0.91)	0.87 (0.77–0.99)	0.79 (0.65–0.90)	0.79 (0.66–0.91)	0.87 (0.75–0.97)
LGBM	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
RF	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
SVM	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
XGB	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
Validation cohort 1						
ADB	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
CatB	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
DT	0.87 (0.67–1.00)	0.87 (0.70–1.00)	0.86 (0.57–1.00)	0.88 (0.64–1.00)	0.86 (0.59–1.00)	0.88 (0.68–1.00)
GNB	0.86 (0.61–1.00)	0.87 (0.73–1.00)	1.00 (1.00–1.00)	0.75 (0.31–1.00)	0.78 (0.50–1.00)	1.00 (1.00–1.00)
KNN	0.47 (0.21–0.73)	0.47 (0.20–0.70)	0.57 (0.17–0.94)	0.38 (0.00–0.71)	0.44 (0.13–0.75)	0.50 (0.13–0.96)
LR	0.79 (0.48–1.00)	0.73 (0.53–0.93)	0.86 (0.63–1.00)	0.62 (0.33–1.00)	0.67 (0.29–1.00)	0.83 (0.50–1.00)
LGBM	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
RF	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
SVM	1.00 (1.00–1.00)	0.87 (0.73–1.00)	1.00 (1.00–1.00)	0.75 (0.46–1.00)	0.78 (0.39–1.00)	1.00 (1.00–1.00)
XGB	0.98 (0.89–1.00)	0.93 (0.80–1.00)	0.86 (0.55–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.89 (0.65–1.00)
Validation cohort 2						
ADB	0.97 (0.88–1.00)	0.93 (0.85–1.00)	0.95 (0.85–1.00)	0.89 (0.65–1.00)	0.95 (0.82–1.00)	0.89 (0.67–1.00)
CatB	0.91 (0.68–1.00)	0.97 (0.90–1.00)	1.00 (1.00–1.00)	0.89 (0.67–1.00)	0.95 (0.88–1.00)	1.00 (1.00–1.00)
DT	0.89 (0.71–1.00)	0.93 (0.87–1.00)	1.00 (1.00–1.00)	0.78 (0.46–1.00)	0.91 (0.81–1.00)	1.00 (1.00–1.00)
GNB	0.94 (0.83–1.00)	0.80 (0.67–0.90)	0.86 (0.68–1.00)	0.67 (0.39–1.00)	0.86 (0.68–1.00)	0.67 (0.35–1.00)
KNN	0.52 (0.35–0.71)	0.60 (0.40–0.77)	0.71 (0.50–0.89)	0.33 (0.04–0.68)	0.71 (0.58–0.92)	0.33 (0.00–0.65)
LR	0.84 (0.64–0.97)	0.80 (0.67–0.93)	0.86 (0.69–1.00)	0.67 (0.29–0.89)	0.86 (0.69–0.96)	0.67 (0.39–0.89)
LGBM	0.96 (0.85–1.00)	0.97 (0.90–1.00)	1.00 (1.00–1.00)	0.89 (0.68–1.00)	0.95 (0.87–1.00)	1.00 (1.00–1.00)
RF	0.97 (0.89–1.00)	0.97 (0.90–1.00)	1.00 (1.00–1.00)	0.89 (0.64–1.00)	0.95 (0.84–1.00)	1.00 (1.00–1.00)
SVM	0.94 (0.84–1.00)	0.90 (0.78–0.97)	0.95 (0.84–1.00)	0.78 (0.53–1.00)	0.91 (0.78–1.00)	0.88 (0.55–1.00)
XGB	0.97 (0.88–1.00)	0.97 (0.90–1.00)	1.00 (1.00–1.00)	0.89 (0.61–1.00)	0.95 (0.89–1.00)	1.00 (1.00–1.00)
Validation cohort 3						
ADB	0.89 (0.80–0.96)	0.83 (0.75–0.92)	0.86 (0.74–0.97)	0.79 (0.63–0.94)	0.86 (0.76–0.97)	0.79 (0.64–0.95)
CatB	0.93 (0.86–0.98)	0.80 (0.67–0.89)	0.83 (0.72–0.97)	0.75 (0.60–0.92)	0.83 (0.70–0.95)	0.75 (0.52–0.94)
DT	0.76 (0.64–0.87)	0.77 (0.67–0.86)	0.81 (0.68–0.90)	0.71 (0.51–0.86)	0.81 (0.64–0.93)	0.71 (0.52–0.88)
GNB	0.89 (0.79–0.97)	0.82 (0.72–0.91)	0.81 (0.66–0.90)	0.83 (0.69–0.96)	0.88 (0.76–0.97)	0.74 (0.62–0.87)
KNN	0.54 (0.42–0.66)	0.55 (0.41–0.68)	0.58 (0.44–0.75)	0.50 (0.28–0.67)	0.64 (0.48–0.80)	0.44 (0.21–0.58)
LR	0.80 (0.68–0.89)	0.72 (0.62–0.85)	0.72 (0.53–0.86)	0.71 (0.49–0.86)	0.79 (0.67–0.90)	0.63 (0.45–0.80)
LGBM	0.91 (0.81–0.96)	0.80 (0.72–0.92)	0.83 (0.73–0.94)	0.75 (0.54–0.89)	0.83 (0.69–0.94)	0.75 (0.53–0.90)
RF	0.93 (0.86–0.98)	0.82 (0.72–0.92)	0.83 (0.70–0.95)	0.79 (0.60–0.91)	0.86 (0.75–0.95)	0.76 (0.58–0.92)
SVM	0.78 (0.64–0.91)	0.77 (0.64–0.88)	0.83 (0.70–0.94)	0.67 (0.45–0.83)	0.79 (0.61–0.92)	0.73 (0.59–0.87)
XGB	0.91 (0.82–0.97)	0.80 (0.72–0.88)	0.83 (0.70–0.93)	0.75 (0.54–0.90)	0.83 (0.70–0.96)	0.75 (0.55–0.89)

Table 2. Performance evaluation of machine learning models for predicting 28-day mortality after VA-ECMO weaning. ADB, adaptive boosting; CatB, categorical boosting; DT, decision tree; GNB, gaussian naive bayes; KNN, K-nearest neighbor; LR, logistic regression; LGBM, light gradient boosting machine; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting. Recall = Sensitivity and Precision = Positive predictive value. Significant values are in bold.

2. Knowledge embedding and high generalization. We employ a knowledge embedding thoughts, directly incorporating relevant variables explicitly identified in existing clinical research findings into the model. This enables the model to learn key features more quickly and accurately in small-sample environments. This strategy not only enhances the interpretability of the model but also validates the possibility of achieving high performance and high generalization with limited data.

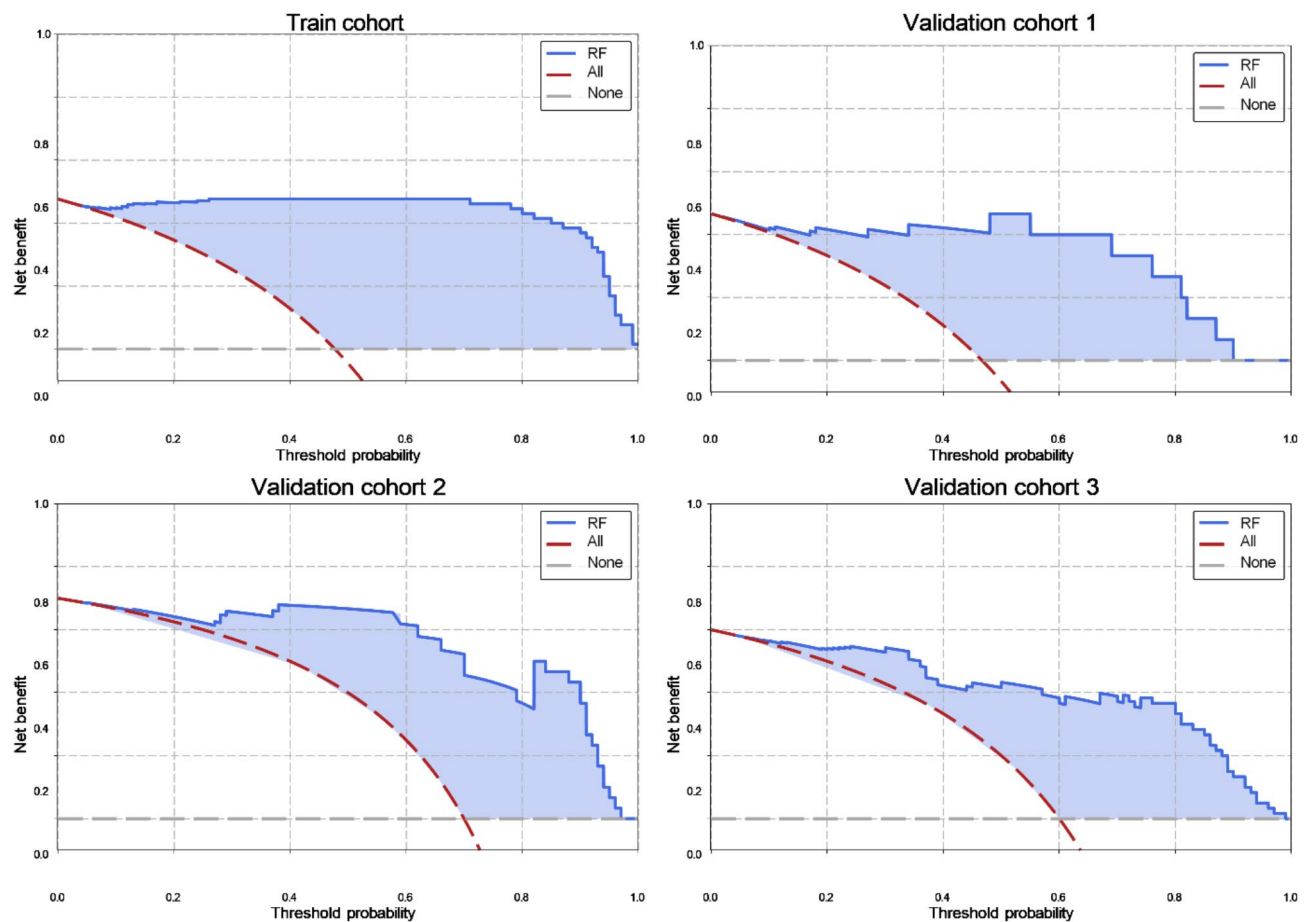


Fig. 4. The decision curve analysis of the random forest based prediction model.

Second, the controversy regarding the “obesity paradox” and the relationship between BMI and prognosis is discussed as follows:

1. Differences in different disease contexts. While in some chronic diseases (such as heart failure), a high BMI may be associated with better prognosis, with some studies supporting the “obesity paradox”³⁹, in acutely critically ill patients (such as those dependent on ECMO), a high BMI may increase hemodynamic burden, reduce lung compliance, and exacerbate inflammatory responses, thereby increasing the risk of mortality⁴⁰.
2. ECMO specificity. Due to the extreme complexity of their condition, ECMO patients’ prognosis is influenced by multiple factors. Patients with a high BMI may face risks such as insufficient oxygenation, myocardial injury, and poor organ perfusion during ECMO treatment^{41,42}. Additionally, a high BMI may increase the risk of complications during ECMO procedures, such as deep vein thrombosis, pulmonary infections, and bleeding⁴³, as well as affect anticoagulation and anti-inflammatory treatment strategies⁴⁴.

Third, some of the latest literature also supports the view that BMI is detrimental to VA-ECMO prognosis. Multiple studies from the past four years have reached a clear conclusion that obesity increases mortality in patients receiving venoarterial extracorporeal membrane oxygenation treatment^{37,38,40,44}, with two papers published in 2025 and one in 2024. This finding directly supports our highly generalized model’s judgment that BMI is positively correlated with mortality in VA-ECMO patients, further corroborating our research perspective.

Overall, the role of BMI in critical care settings is complex, and different studies may yield different conclusions. Some studies support the obesity paradox (i.e., a high BMI may have a protective effect in ICU patients), while others suggest that a high BMI may increase ECMO-related complications. In critical illness, many studies have found that BMI has a protective effect, and obesity is becoming increasingly common in intensive care units. Recent research has shown that this proportion is around 20%^{35,36}. In 2022, Zhou et al.⁴⁵ also discussed the obesity paradox in critically ill patients, suggesting that obese patients may benefit from it. Similarly, Samir Jaber et al.⁴⁶ proposed that among hospitalized patients, intensive care unit patients, and chronic disease patients, the relationship between body mass index and blood pressure follows a “J” shape, as does the relationship between body mass index and mortality. Overweight and moderate obesity appear to have a protective effect compared to a normal body mass index. This suggests that being overweight and moderately obese has a protective effect compared to a normal body mass index or more severe obesity. However, this is highly controversial because the

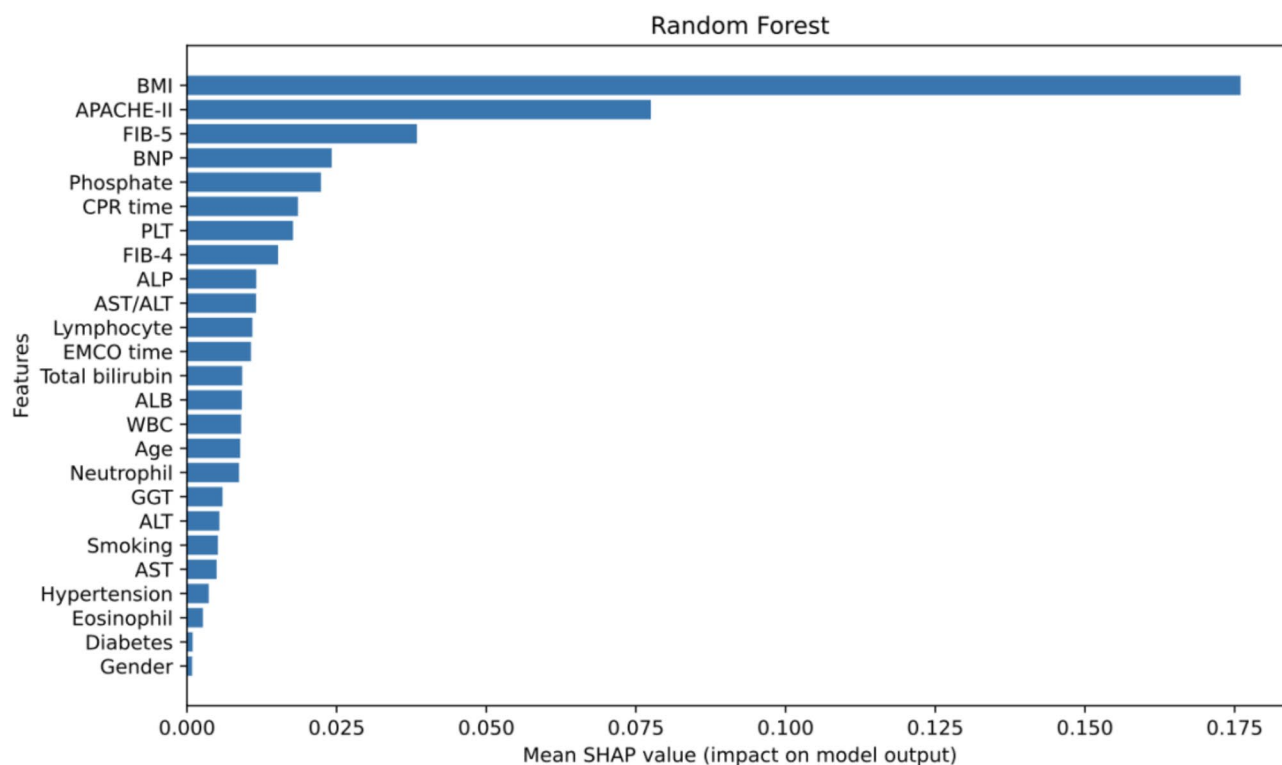


Fig. 5. SHAP feature correlation analysis of the Random Forest based prediction model.

risk of cardiovascular disease in obese individuals increases depending on the degree, distribution, and duration of obesity. Cardiovascular abnormalities related to fat mass include increased blood volume and cardiac output, leading to secondary ventricular hypertrophy and diastolic dysfunction, ultimately resulting in ventricular dilation (obesity cardiomyopathy). Atrial fibrillation is a common complication of obesity, and pulmonary hypertension should also be suspected (secondary to elevated left atrial pressure, hypoxia caused by obstructive sleep apnea, hypoventilation syndrome, or chronic thromboembolism). In addition to issues related to fat mass, pathological fat (visceral fat and ectopic fat) can affect the cardiovascular system either directly through immune and endocrine influences or indirectly through metabolic syndrome associated with hypertension (afterload), dyslipidemia, and ischemic heart disease⁴⁷. The acoustic performance of transthoracic echocardiography is often poor in patients with high BMI, hindering accurate image acquisition. Assessing hemodynamic instability may require transesophageal echocardiography. Uncalibrated noninvasive cardiac output measurement based on pulse contour analysis is popular in the intensive care unit but appears to be inaccurate in obese patients⁴⁸. This is not surprising, as converting pressure waveforms into cardiac output relies on algorithms that include vascular wall dynamic properties, which may undergo significant changes in obese patients. If it is deemed necessary to closely monitor cardiac output, a right heart catheter or transesophageal Doppler can be used. There is currently limited data on the interpretation of hemodynamic parameters in obese patients: a small study suggests that the hemodynamic parameters of obese patients should not differ from those of non-obese patients, provided these parameters are related to body surface area. These limitations affect the treatment of critically obese patients in the ICU. In extremely critical cases such as ECMO, our analysis of multiple machine learning models—validated for high generalization performance—consistently reveals that high BMI has a promoting and positively correlated inferential effect on mortality.

Serum phosphate levels in VA-ECMO supported patients have received less attention, primarily being studied in the context of chronic kidney disease. Numerous studies have shown a U-shaped relationship between serum phosphate levels and patient mortality, where both excessively high and low phosphate levels increase the risk of death and adverse cardiovascular events^{49,50,34,35}. A prospective, multicenter study in Japan found that patients had the highest mortality risk when serum phosphate levels were greater than 1.93 mmol/L^{51,36}. The COSMOS study found that the lowest mortality risk was observed at a serum phosphate level of 1.42 mmol/L^{52,37}. Research in China also indicated that serum phosphate levels below 0.8 mmol/L increased the risk of all-cause and cardiovascular mortality^{53,38}. These findings suggest that normal serum phosphate levels are associated with the best prognosis. Additionally, a meta-analysis involving 327,644 patients found a correlation between higher serum phosphate levels and increased mortality^{54,39}. Our study results also indicate that higher serum phosphate levels are associated with higher mortality. Therefore, in the management of VA-ECMO patients, attention should be paid to this indicator, and further evaluation is needed to determine if intervening in serum phosphate levels can improve patient outcomes.

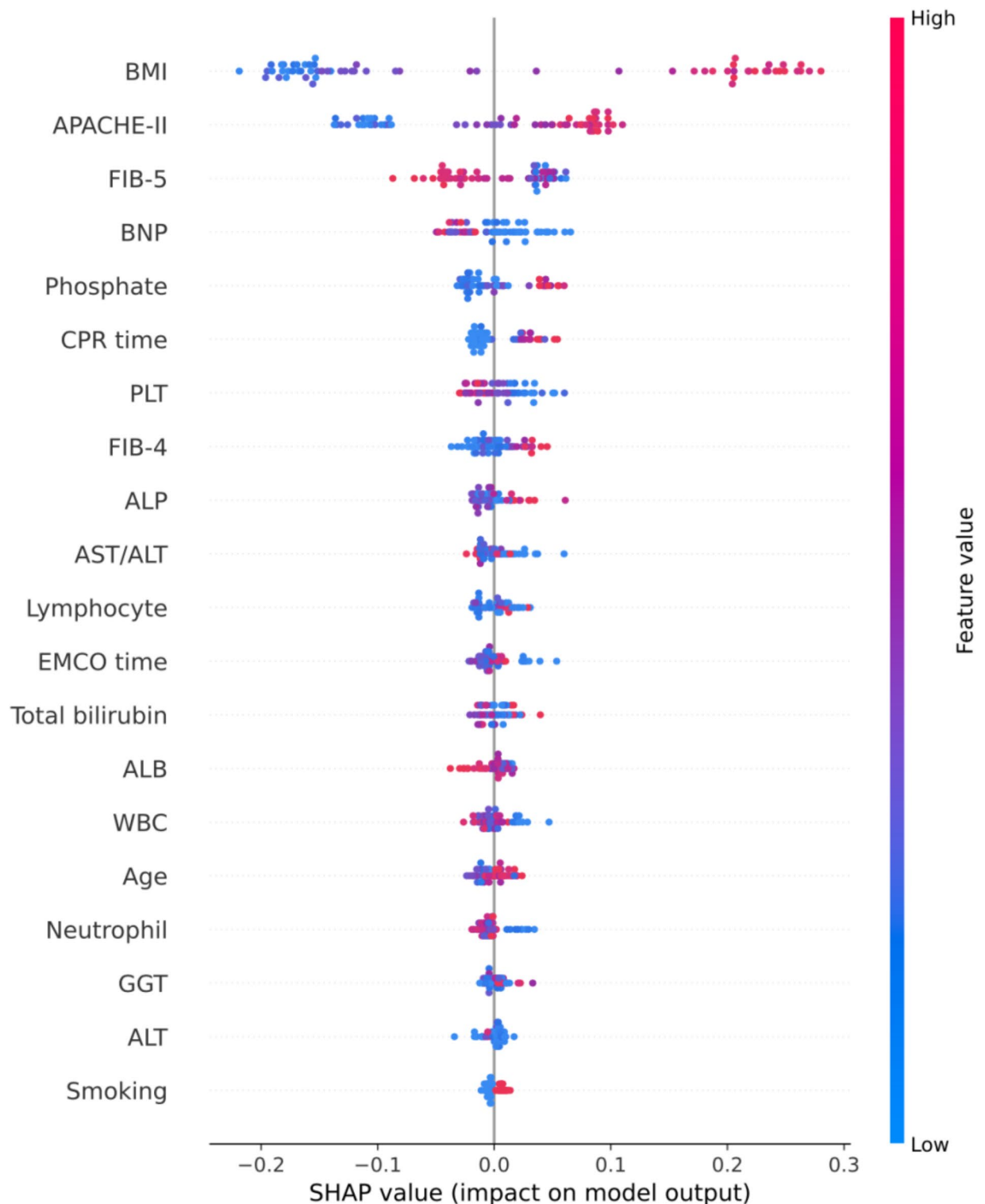


Fig. 6. SHAP feature correlation heatmap of Random Forest (RF).

Typically, in the early stages of VA-ECMO support, the heart and brain are the prioritized organs, rather than the liver. However, recent studies have found that liver function indicators can be used to predict mortality in VA-ECMO patients. Currently, the veno-arterial ECMO (SAVE) score, a prospectively validated VA-ECMO mortality prediction model, derives part of its predictive ability from liver function indicators, such as serum bilirubin and the critical values of AST and ALT¹¹. American clinicians have found that abnormal liver function indicators in VA-ECMO supported patients may be powerful predictors of mortality^{55,40}. Sern Lim's study^{56,41} on a small VA-ECMO cohort revealed a close link between the MELD-XI score and increased mortality. Cho et

al.^{57,42} reported similar findings in a study involving 49 VA-ECMO patients who underwent heart transplantation using the MELD score. Roth et al.^{58,43} found that elevated ALP and total bilirubin were the strongest predictors of 30-day mortality in a cohort of 240 patients who received VA-ECMO cannulation for cardiogenic shock post-myocardial infarction. Maxhera et al.^{59,44} also reported similar results in a cohort of VA-ECMO patients experiencing shock after LVAD surgery. These studies suggest that liver injury significantly impacts prognosis, likely due to coagulation abnormalities, poor wound healing, intractable peripheral vasodilation caused by abnormal nitric oxide metabolism, and progressive liver dysfunction leading to decreased immunity^{60,45}. Therefore, our model included liver function indicators such as ALT, AST, GGT, ALP, and total bilirubin, and used these indicators to compose FIB-4 and FIB-5 as variables for the prediction model. Our study found that higher FIB-4 and ALP levels were associated with higher mortality risk, while higher ALT levels were associated with lower mortality risk. The trained model achieved impressive predictive performance.

In recent years, some studies have started to include certain blood routine indicators to assess the prognosis of VA-ECMO patients^{61,62,46,47}. One of the factors related to the mortality of VA-ECMO patients is systemic inflammation. To quantify systemic inflammation, basic and easily obtainable indicators such as WBC, PLT, and Neutrophil may be more meaningful.

In addition, in the SHAP analysis of this study, BMI and FIB-5 demonstrated high importance in predicting ECMO prognosis, suggesting that body mass index and liver function status may be key factors influencing ECMO outcomes. Although current medical research has not fully established a causal relationship between these variables and ECMO prognosis, we believe these findings can be used to optimize risk management for ECMO patients. Patients with higher BMI may require more stringent monitoring to prevent hemodynamic instability, thrombosis, and pulmonary complications associated with ECMO. Those with higher FIB-5 scores may need closer liver function management, such as optimizing fluid management strategies, avoiding hepatotoxic drugs, and monitoring liver injury during ECMO. Furthermore, future studies should further explore the specific role of BMI and FIB-5 in ECMO indications and integrate other clinical scoring systems (such as the SOFA score) to optimize patient management strategies.

Limitations

Regarding whether the model exhibits overfitting. The core characteristic of overfitting is that the model performs exceptionally well on training data but shows a significant decline in performance on independent test data not seen during training. In this study, we used three independent external validation cohorts with data from five different hospitals. Taking RF as an example, the developed model achieved an AUC of 1 in the training set; however, its AUC remained high (0.97 and 0.93) across all two external validation sets without a noticeable drop in performance. This result indicates that the model does not exhibit overfitting but rather maintains robust predictive capability on independent real-world data. During learning and training, if a model were overfitted, we would observe that its test AUC in the training set is significantly higher than that in external validation sets. However, in this study, the actual situation is that the test AUC values for external validation sets remain at high levels. In summary, we believe that our multi-center external validation results sufficiently demonstrate the generalizability of our model and do not support concluding overfitting merely because of an AUC = 1.0 in the training set. In fact, obtaining samples for ECMO-related research poses significant challenges, especially for highly specific clinical applications like VA-ECMO. Compared to other common diseases, the number of ECMO patients is inherently very limited, and their clinical management is extremely complex. Many hospitals are unable to provide sufficiently large-scale data. Therefore, our multicenter dataset represents the largest scale achievable under current real-world conditions. At the same time, we have rigorously screened the data to ensure its quality and consistency in order to enhance the model's clinical applicability. Moreover, in medical machine learning research, a model that maintains high generalization ability and accuracy even with a small training dataset is actually an advantage. This aligns with what machine learning and artificial intelligence strive for—the optimal solution that achieves maximum effectiveness with minimal learning cost. If a model requires tens of thousands of data points to produce reliable results, it may indicate that it compensates for its lack of learning capability through sheer data volume. Conversely, if a model demonstrates strong predictive performance on a small dataset, this suggests that its feature selection process, learning approach, and decision-making mechanism possess greater medical validity and generalizability.

Regarding the sample size of the internal validation cohort. One of the purposes of the internal validation set is to provide a basic reference for model tuning during training, while the final model evaluation primarily relies on multi-center external validation. In fact, when assessing a model's generalization ability, external validation results should be considered the primary reference standard, as they are far more meaningful than internal validation. In our external validation cohort, all available real patient data from multiple centers were included. This ensures that the class distribution in the training data aligns with the actual survival rate of VA-ECMO patients in multi-center hospitals (as all available data were collected and organized), thereby reducing discrepancies between sample label proportions and real-world proportions. Under these circumstances, an internal validation set can still verify whether and to what extent the developed model can effectively predict mortality for previously unseen patients within the same hospital (since ICU patient conditions and characteristics within a single hospital tend to be more similar). Therefore, having a smaller internal validation set does not affect research conclusions. On the contrary, due to both ECMO mortality prediction samples being scarce and despite internal validation providing some guidance on generalizability, external validation has stronger value in verifying generalization ability. To some extent, it also indirectly reflects insights into internal validation performance. As such, we allocated most of our available samples from internal validation into training to ensure higher-quality model performance while still obtaining an internal validation result. In summary, true demonstration and determination of a model's generalization capability come from external validation results. Given this context—despite having a relatively small number of samples in our internal validation set—the study

employed one independent internal cohort along with two external cohorts comprising four centers' samples. The AUC results across different datasets exhibited good consistency; therefore, we believe that under current experimental conditions, this model has been validated as possessing strong generalization ability and predictive performance without significant overfitting risk. In further summary, our proposed model, based on 25 input features that are easily obtainable in clinical diagnosis through reasonable selection, has a major advantage: it can achieve both high generalization and high predictive accuracy even with small-sample training.

Although our findings are promising, it is important to recognize their limitations. First, the eCMoML model was developed primarily using data from Chinese patients, thus further validation in different ethnic groups is necessary to ensure its generalizability. Secondly our sample size was small. Although we conducted a multicenter study and applied advanced algorithms for analysis, prospective international multicenter studies are needed to further validate the performance of the eCMoML model.

Usability of the model in the context of current care

According to the 2024 updated TRIPOD + AI statement, in this section, we further explain other model-related usage considerations.

How poor quality or unavailable input data should be assessed and handled when implementing the prediction model?

In this study, we have collected all available real patient data from five hospital centers as comprehensively as possible. The total number of low-quality and unusable data does not exceed 15 cases, each missing one of the 25 input features required by the research model. Unlike visual and image-based deep learning diagnostics, to maximize the authenticity of the data learned by the model and thereby obtain a high-quality predictive model with strong generalization ability, we directly removed all low-quality and unusable data. Although each sample in the relevant dataset is extremely valuable, experimental results indicate that the selection of highly relevant input variables and a high-quality model training process still resulted in a highly accurate predictive model with strong generalization ability. In other words, removing these data had no significant negative impact on the model and may have even played a positive role during training.

User interaction requirements and expertise level for input handling and model use

In a clinical setting, the effective application of AI prediction models depends not only on their technical performance but also on their usability and the level of expertise required from users. To ensure that eCMoML can be seamlessly integrated into ECMO clinical management, we evaluated its user interaction requirements as well as the level of expertise needed for input processing and model usage to ensure feasibility and reliability in real-world applications.

1. On the convenience of input processing and model usability. One significant advantage of eCMoML is that it relies solely on 25 clinical and laboratory variables, all of which are routinely collected indicators in ICU treatment. This eliminates the need for additional costly tests or complex imaging data, making the model highly implementable. Compared to models that depend on imaging or advanced biomarkers, eCMoML's input data is more accessible and applicable to ICUs with varying levels of medical resources, reducing the difficulty of data collection. In terms of data input, the model can be directly integrated with electronic health record (EHR) systems to enable automated data extraction, avoiding errors and workload associated with manual input. Furthermore, we recommend ensuring complete variable input as much as possible, rather than relying on data imputation. The rationale behind this strategy is as follows. First, the 25 selected feature variables are key indicators standardized in ICU clinical practice, inherently possessing high availability. Second, machine learning models are highly sensitive to input data quality, and data imputation may introduce potentially misleading information, compromising the stability of the model's predictive outcomes in real-world clinical settings. Third, given the high complexity of ECMO patients' conditions, inferring patient status based on missing data poses substantial risks. Therefore, at the study design stage, we ensured the high availability of the required feature variables to minimize the necessity of data imputation.
2. On applicable users and required expertise. The design goal of eCMoML is to serve as a clinical decision support tool rather than to replace clinicians' judgment. Therefore, users are not required to have specialized knowledge in machine learning or artificial intelligence. The model presents its predictive results in the form of a 28-day mortality risk score, allowing clinicians to integrate this information with the patient's condition for individualized decision-making. Specifically, For ICU attending physicians and ECMO teams: The model can assist in identifying high-risk patients and adjusting treatment strategies, such as enhancing monitoring, optimizing ECMO management, evaluating the timing of weaning, or considering alternative treatment plans. For residents and rotating physicians: The risk scores and SHAP-based explanations provided by the model can help junior doctors understand key factors in ECMO prognostic assessment, improving their diagnostic and treatment skills. For multidisciplinary teams (e.g., cardiology, hepatobiliary, nephrology departments): Since ECMO patients often experience multi-organ dysfunction, the model offers data-driven support to facilitate multidisciplinary collaborative decision-making, such as implementing enhanced liver protection measures for patients with impaired liver function.
3. On interpretation of prediction results and decision support. To enhance model transparency and user confidence, eCMoML employs SHAP interpretability analysis to provide detailed variable contributions for each patient's prediction results. This helps doctors understand which factors play a dominant role in specific cases. For example, if the APACHE-II score is the primary factor in an individual prediction, physicians can focus on assessing the overall severity of the patient's condition. If the FIB-5 score is high, it may indicate potential liver dysfunction, necessitating enhanced fluid management and liver protection measures. If BMI

is the key influencing factor, doctors can evaluate the patient's metabolic status to assess ECMO-related hemodynamic risks. Furthermore, future developments could incorporate risk trend analysis, using time-series representations of mortality risk scores to track disease progression trends, thereby optimizing long-term treatment strategies.

4. On future optimization directions. To further enhance the usability of eCMoML, we propose the following optimization directions. First, integration with ICU monitoring systems: embedding the model into ICU dashboards to enable real-time automatic updates of risk scores, thereby reducing the operational burden on physicians. Second, development of user-friendly mobile or web-based tools: creating a clinically accessible interface that allows doctors to quickly input key variables and obtain realtime prediction results. Third, targeted clinical training: offering short-term training courses or online instructional modules to help clinicians familiarize themselves with the model's usage and best practices, thereby improving its clinical acceptance.

In summary, eCMoML is designed to be simple, easy to use, and seamlessly integrated into clinical workflows without requiring doctors to have AI knowledge. It assists ICU physicians, residents, and multidisciplinary teams in ECMO prognostic risk assessment. The model requires 25 variables, all of which are standardized ICU collection indicators, and it is recommended to avoid data imputation to ensure the stability and accuracy of predictions. Future optimization efforts will focus on automated data input, real-time ICU monitoring integration, and intelligent decision support to further enhance the model's clinical value and scalability.

Next steps for future research

Since the training and validation data in this study are primarily based on Chinese patient data, it is necessary to verify its generalizability worldwide across different populations. Due to a series of objective issues such as ethics and privacy, despite our significant efforts and the collection of valuable samples from five different tertiary hospitals in China, collecting relevant samples from various countries remains extremely challenging. To determine whether the proposed model has racial or regional limitations, we leverage one major advantage validated in this study—its ability to achieve high generalizability and predictive accuracy even with a small sample size for training. In future research, an optimal approach would be to not only validate the trained model but also collect and use real-world samples from different regions globally for retraining so that the model can naturally adapt to potential variations in patient characteristics across regions. Additionally, retraining using mixed datasets from around the world may lead to a more universally applicable predictive model. It is important to note that our model has already demonstrated excellent generalizability among patients from different regions within the same country. Therefore, we cannot rule out the possibility that it may still exhibit high generalizability when applied to larger sample sizes. Furthermore, given its outstanding generalization capability with small-sample training, this model could potentially achieve even better predictive accuracy and broader applicability when supported by more extensive and comprehensive patient data worldwide. From another perspective, the model in this study predicts based on static variables at the time of decannulation. Therefore, future optimizations can be made in the following directions. First, time series modeling: Future studies can employ deep learning methods such as LSTM (Long Short-Term Memory) or Transformer to integrate dynamic variables during ECMO operation and enhance the model's real-time predictive capability. However, it is important to note that deep learning models typically require large amounts of data for support. Second, integrating EHR (Electronic Health Record) data for dynamic monitoring: Currently, ICU equipment can monitor hemodynamic parameters of ECMO patients in real time (e.g., blood lactate levels, blood gas analysis, cardiac output). In the future, these time-series data could be input into AI models to improve predictive accuracy. Third, individualized dynamic risk prediction: The current model relies on static assessments at decannulation; however, an AI-based early warning system incorporating dynamic data could be developed in the future to provide earlier risk assessments during ECMO operation.

Conclusion

This study developed an AI-driven model (eCMoML) based on 25 easily accessible patient characteristics and evaluated its ability to predict the 28-day post-extubation mortality risk in VA-ECMO patients through a multicenter, multi-cohort retrospective study. The model demonstrated stable predictive performance in internal validation and two independent external validation cohorts. Its potential clinical applicability was further verified using widely adopted statistical techniques and metrics, including the area under the receiver operating characteristic curve (AUROC), recall, positive predictive value (PPV), decision curve analysis (DCA), and SHAP interpretation. eCMoML can assist clinicians in assessing mortality risk after VA-ECMO weaning, thereby supporting individualized management and optimizing treatment decisions. However, we acknowledge certain methodological limitations in this study. First, as a retrospective study, the data sources may be constrained by region-specific ECMO treatment practices, which could affect the model's applicability to broader populations. Second, although multicenter external validation was conducted to ensure and comprehensively enhance the model's generalizability, further prospective studies remain necessary to validate the clinical utility of eCMoML in real-world settings. In the future, we plan to validate the model on a larger-scale international multicenter dataset and explore its performance across different ethnicities, ECMO indications, and ICU resource settings. Additionally, the SHAP importance analysis in this study has clearly identified key features such as BMI, APACHE II, and FIB-5 as critical factors in ECMO prognostic assessment, providing new insights for future ECMO risk stratification research. On the other hand, SHAP interaction effects may have potential in revealing complex feature interactions. However, given the computational complexity of SHAP interaction effects, challenges in clinical interpretability, and the fact that SHAP importance analysis already effectively meets model interpretability needs, this study did not incorporate interaction effect visualization. In the future, as larger datasets become available, we plan to explore the applicability of SHAP interaction effects

in extended studies and assess their practical value in enhancing clinical understanding and improving model transparency. With the continuous advancement of medical AI, future research can further integrate dynamic physiological parameters, timeseries data, deeplearning techniques, and advanced interpretability tools such as SHAP interaction effects to optimize individualized risk prediction strategies for ECMO patients and ensure the clinical feasibility of the model.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request. Data can be obtained by contacting jluiwangs@163.com. The proposed prediction model eCMoML will be available online at <https://github.com/SichenTao>, simply input the required features to immediately obtain the prediction results.

Received: 27 November 2024; Accepted: 17 March 2025

Published online: 26 March 2025

References

- Pladet, L. C. et al. Prognostic models for mortality risk in patients requiring ECMO. *Intensive Care Med.* **49**, 131–141 (2023).
- Thiele, H., Ohman, E. M., Desch, S., Eitel, I. & De Waha, S. Management of cardiogenic shock. *Eur. Heart J.* **36**, 1223–1230 (2015).
- Abrams, D., Combes, A. & Brodie, D. Extracorporeal membrane oxygenation in cardiopulmonary disease in adults. *J. Am. Coll. Cardiol.* **63**, 2769–2778 (2014).
- Zangrillo, A. et al. A meta-analysis of complications and mortality of extracorporeal membrane oxygenation. *Crit. Care Resusc.* **15**, 172–178 (2013).
- Karagiannidis, C. et al. Extracorporeal membrane oxygenation: evolving epidemiology and mortality. *Intensive Care Med.* **42**, 889–896 (2016).
- Cheng, R. et al. Complications of extracorporeal membrane oxygenation for treatment of cardiogenic shock and cardiac arrest: A meta-analysis of 1,866 adult patients. *Ann. Thorac. Surg.* **97**, 610–616 (2014).
- Hodgson, C. L. et al. Long-term quality of life in patients with acute respiratory distress syndrome requiring extracorporeal membrane oxygenation for refractory hypoxaemia. *Crit. Care* **16**, 1–10 (2012).
- Doll, N. et al. Five-year results of 219 consecutive patients treated with extracorporeal membrane oxygenation for refractory postoperative cardiogenic shock. *Ann. Thorac. Surg.* **77**, 151–157 (2004).
- Kim, H. et al. Efficacy of veno-arterial extracorporeal membrane oxygenation in acute myocardial infarction with cardiogenic shock. *Resuscitation* **83**, 971–975 (2012).
- Sheu, J.-J. et al. Early extracorporeal membrane oxygenator-assisted primary percutaneous coronary intervention improved 30-day clinical outcomes in patients with ST-segment elevation myocardial infarction complicated with profound cardiogenic shock. *Crit. Care Med.* **38**, 1810–1817 (2010).
- Schmidt, M. et al. Predicting survival after ECMO for refractory cardiogenic shock: The survival after veno-arterial-ECMO (save)-score. *Eur. Heart J.* **36**, 2246–2256 (2015).
- Laribi, S. & Mebazaa, A. Cardiohepatic syndrome: Liver injury in decompensated heart failure. *Curr. Heart Fail. Rep.* **11**, 236–240 (2014).
- Alvarez, A. M. & Mukherjee, D. Liver abnormalities in cardiac diseases and heart failure. *Int. J. Angiol.* **8**, 135–142 (2011).
- Poelzl, G. et al. Liver dysfunction in chronic heart failure: Prevalence, characteristics and prognostic significance. *Eur. J. Clin. Invest.* **42**, 153–163 (2012).
- Sterling, R. K. et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* **43**, 1317–1325 (2006).
- Sato, Y. et al. Liver stiffness assessed by fibrosis-4 index predicts mortality in patients with heart failure. *Open Heart* **4**, e000598 (2017).
- Maeda, D. et al. Fibrosis-4 index reflects right-sided filling pressure in patients with heart failure. *Heart Vessels* **35**, 376–383 (2020).
- Nakashima, M. et al. Fibrosis-4 index reflects right ventricular function and prognosis in heart failure with preserved ejection fraction. *ESC Heart Fail.* **8**, 2240–2247 (2021).
- Shibata, N. et al. Impact of predictive value of fibrosis-4 index in patients hospitalized for acute heart failure. *Int. J. Cardiol.* **324**, 90–95 (2021).
- Attallah, A. M., Shiha, G. E., Omran, M. M. & Zalata, K. R. A discriminant score based on four routine laboratory blood tests for accurate diagnosis of severe fibrosis and/or liver cirrhosis in Egyptian patients with chronic hepatitis C. *Hepatol. Res.* **34**, 163–169 (2006).
- Maeda, D. et al. Prognostic value of the liver fibrosis marker fibrosis-5 index in patients with acute heart failure. *ESC Heart Fail.* **9**, 1380–1387 (2022).
- Van Diepen, S. et al. Contemporary management of cardiogenic shock: A scientific statement from the American Heart Association. *Circulation* **136**, e232–e268 (2017).
- Collins, G. S. et al. Tripod+ ai statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **8**, 385 (2024).
- Aissaoui, N. et al. Predictors of successful extracorporeal membrane oxygenation (ECMO) weaning after assistance for refractory cardiogenic shock. *Intensive Care Med.* **37**, 1738–1745 (2011).
- Ortega, F. B., Lavie, C. J. & Blair, S. N. Obesity and cardiovascular disease. *Circul. Res.* **118**, 1752–1770 (2016).
- Lu, Y. et al. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: A pooled analysis of 97 prospective cohorts with 1.8 million participants. *Lancet (London, England)* **383**, 970–983 (2013).
- Plourde, B., Sarrazin, J.-F., Nault, I. & Poirier, P. Sudden cardiac death and obesity. *Expert Rev. Cardiovasc. Ther.* **12**, 1099–1110 (2014).
- Finocchiaro, G. et al. Obesity and sudden cardiac death in the young: Clinical and pathological insights from a large national registry. *Eur. J. Prevent. Cardiol.* **25**, 395–401 (2018).
- Ohlsson, M. A., Kennedy, L. M. A., Juhlin, T. & Melander, O. Midlife risk factor exposure and incidence of cardiac arrest depending on cardiac or non-cardiac origin. *Int. J. Cardiol.* **240**, 398–402 (2017).
- Sung, C.-W. et al. Obese cardiogenic arrest survivors with significant coronary artery disease had worse in-hospital mortality and neurological outcomes. *Sci. Rep.* **10**, 18638 (2020).
- Geri, G. et al. Influence of body mass index on the prognosis of patients successfully resuscitated from out-of-hospital cardiac arrest treated by therapeutic hypothermia. *Resuscitation* **109**, 49–55 (2016).
- Chavda, M. P., Bihari, S., Woodman, R. J., Secombe, P. & Pilcher, D. The impact of obesity on outcomes of patients admitted to intensive care after cardiac arrest. *J. Crit. Care* **69**, 154025 (2022).

33. Chavda, M. P., Pakavakis, A. & Ernest, D. Does obesity influence the outcome of the patients following a cardiac arrest?. *Indian J. Crit. Care Med. Peer-Rev.* **24**, 1077 (2020).
34. Anderson, M. R. & Shashaty, M. G. Impact of obesity in critical illness. *Chest* **160**, 2135–2145 (2021).
35. Sakr, Y. et al. Being overweight is associated with greater survival in ICU patients: Results from the intensive care over nations audit. *Crit. Care Med.* **43**, 2623–2632 (2015).
36. De Jong, A. et al. Medical versus surgical icu obese patient outcome: A propensity-matched analysis to resolve clinical trial controversies. *Crit. Care Med.* **46**, e294–e301 (2018).
37. Ng, W. W. S. et al. Impact of obesity on outcomes in patients receiving extracorporeal membrane oxygenation: A systematic review and meta-analysis. *Int. J. Artif. Organs* **8**, 0391 (2025).
38. Pai, C.-H. et al. Does obesity matter in patients receiving venoarterial extracorporeal membrane oxygenation? The u-shaped relationship between body mass index and mortality after extracorporeal cardiopulmonary resuscitation. *Surgery* **178**, 108928 (2025).
39. Simati, S., Kokkinos, A., Dalamaga, M. & Argyrakopoulou, G. Obesity paradox: Fact or fiction?. *Curr. Obes. Rep.* **12**, 75–85 (2023).
40. Dana, R. et al. Obesity and mortality in critically ill covid-19 patients with respiratory failure. *Int. J. Obes.* **45**, 2028–2037 (2021).
41. Rudym, D. et al. Mortality in patients with obesity and acute respiratory distress syndrome receiving extracorporeal membrane oxygenation: The Multicenter Ecmobesity Study. *Am. J. Respir. Crit. Care Med.* **208**, 685–694 (2023).
42. Huang, X. & Lin, X. Impact of obesity on outcomes of extracorporeal membrane oxygenation support: A systematic review and meta-analysis. *BMC Pulm. Med.* **24**, 157 (2024).
43. Peetermans, M. et al. Impact of bmi on outcomes in respiratory ecmo: An elso registry study. *Intensive Care Med.* **49**, 37–49 (2023).
44. McCloskey, C. G., Hatton, K. W., Furfaro, D. & Engoren, M. Obesity is associated with increased mortality in patients undergoing venoarterial extracorporeal membrane oxygenation. *Crit. Care Med.* **8**, 10–1097 (2024).
45. Zhou, D., Wang, C., Lin, Q. & Li, T. The obesity paradox for survivors of critically ill patients. *Crit. Care* **26**, 198 (2022).
46. Schetz, M. et al. Obesity in the critically ill: A narrative review. *Intensive Care Med.* **45**, 757–769 (2019).
47. Piche, M.-E., Poirier, P., Lemieux, I. & Despres, J.-P. Overview of epidemiology and contribution of obesity and body fat distribution to cardiovascular disease: An update. *Progress Cardiovasc. Dis.* **61**, 103–113 (2018).
48. Tejedor, A. et al. Accuracy of vigileo/flotrac monitoring system in morbidly obese patients. *J. Crit. Care* **30**, 562–566 (2015).
49. Tentori, F. et al. Mortality risk for dialysis patients with different levels of serum calcium, phosphorus, and pth: The dialysis outcomes and practice patterns study (dopps). *Am. J. Kidney Dis.* **52**, 519–530 (2008).
50. Kalantar-Zadeh, K. et al. Survival predictability of time-varying indicators of bone disease in maintenance hemodialysis patients. *Kidney Int.* **70**, 771–780 (2006).
51. Fukagawa, M. et al. Abnormal mineral metabolism and mortality in hemodialysis patients with secondary hyperparathyroidism: Evidence from marginal structural models used to adjust for time-dependent confounding. *Am. J. Kidney Dis.* **63**, 979–987 (2014).
52. Fernandez-Martin, J. L. et al. Improvement of mineral and bone metabolism markers is associated with better survival in haemodialysis patients: The cosmos study. *Nephrol. Dial. Transpl.* **30**, 1542–1551 (2015).
53. Zhu, J.-G. et al. Association between extreme values of markers of chronic kidney disease: Mineral and bone disorder and 5-year mortality among prevalent hemodialysis patients. *Blood Purif.* **45**, 1–7 (2018).
54. Palmer, S. C. et al. Serum levels of phosphorus, parathyroid hormone, and calcium and risks of death and cardiovascular disease in individuals with chronic kidney disease: A systematic review and meta-analysis. *JAMA* **305**, 1119–1127 (2011).
55. Masha, L. et al. Yellow means caution: Correlations between liver injury and mortality with the use of va-ecmo. *ASAIO J.* **65**, 812–818 (2019).
56. Sern Lim, H. Baseline meld-xi score and outcome from veno-arterial extracorporeal membrane oxygenation support for acute decompensated heart failure. *Eur. Heart J. Acute Cardiovasc. Care* **5**, 82–88 (2016).
57. Cho, Y. H. et al. Extracorporeal life support as a bridge to heart transplantation: Importance of organ failure in recipient selection. *ASAIO J.* **61**, 139–143 (2015).
58. Roth, C. et al. Liver function predicts survival in patients undergoing extracorporeal membrane oxygenation following cardiovascular surgery. *Crit. Care* **20**, 1–7 (2016).
59. Maxhera, B. et al. Survival predictors in ventricular assist device patients with prior extracorporeal life support: Selecting appropriate candidates. *Artif. Organs* **38**, 727–732 (2014).
60. Rastan, A. et al. Autopsy findings in patients on postcardiotomy extracorporeal membrane oxygenation (ECMO). *Int. J. Artif. Organs* **29**, 1121–1131 (2006).
61. Roth, S. et al. Neutrophil-lymphocyte-ratio, platelet-lymphocyte-ratio and procalcitonin for early assessment of prognosis in patients undergoing va-ecmo. *Sci. Rep.* **12**, 542 (2022).
62. Diakos, N. A. et al. Systemic inflammatory burden correlates with severity and predicts outcomes in patients with cardiogenic shock supported by a percutaneous mechanical assist device. *J. Cardiovasc. Transl. Res.* **14**, 476–483 (2021).

Author contributions

Study concept and design: Shuai Wang, Sichen Tao and Ying Zhu; data collection and technical support: Sichen Tao, Qiao Gu, Peifeng Ni, Weidong Zhang, Chenxi Wu, Ruihan Zhao; reviewers: Shuai Wang, Sichen Tao, Mengyuan Diao and Wei Hu, Ruihan Zhao. Acquisition and validation of basic data: Shuai Wang and Peifeng Ni. All authors have read and approved the final manuscript.

Funding

This research was partially supported by Science and Technology Program of Traditional Chinese Medicine in Zhejiang Province (Grant: 2024ZL718), Zhejiang Provincial Medical and Health Technology Project (Grant: WKJZJ-2315), Science and Technology Development Project of Hangzhou (Grant: 202204A10), Medical and Health Technology Project of Hangzhou (Grant: Z20240021), Construction Fund of Medical Key Disciplines of Hangzhou (Grant: OO20200485), the Japan Science and Technology Agency (JST) Support for Pioneering Research Initiated by the Next Generation (SPRING) (Grant Number: JPMJSP2145), and the Tongji University Support for Outstanding Ph.D Student Short-Term Overseas Research Funding (Grant Number: 2023020043).

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement

This study was approved by the local institutional review boards. The patients/participants provided their written informed consent to participate in this study.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-94734-3>.

Correspondence and requests for materials should be addressed to W.H. or M.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025