***Editorial***

DOI: 10.5582/bst.2024.01342

# Applications of and issues with machine learning in medicine: Bridging the gap with explainable AI

**Kenji Karako**[1,*], **Wei Tang**[1,2]

[1] Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan;
[2] National Center for Global Health and Medicine, Tokyo, Japan.

**SUMMARY** In recent years, machine learning, and particularly deep learning, has shown remarkable potential in various fields, including medicine. Advanced techniques like convolutional neural networks and transformers have enabled high-performance predictions for complex problems, making machine learning a valuable tool in medical decision-making. From predicting postoperative complications to assessing disease risk, machine learning has been actively used to analyze patient data and assist healthcare professionals. However, the "black box" problem, wherein the internal workings of machine learning models are opaque and difficult to interpret, poses a significant challenge in medical applications. The lack of transparency may hinder trust and acceptance by clinicians and patients, making the development of explainable AI (XAI) techniques essential. XAI aims to provide both global and local explanations for machine learning models, offering insights into how predictions are made and which factors influence these outcomes. In this article, we explore various applications of machine learning in medicine, describe commonly used algorithms, and discuss explainable AI as a promising solution to enhance the interpretability of these models. By integrating explainability into machine learning, we aim to ensure its ethical and practical application in healthcare, ultimately improving patient outcomes and supporting personalized treatment strategies.

***Keywords*** machine learning, deep learning, explainable AI, medical applications

## 1. Introduction

In recent years, machine learning technologies, and particularly deep learning (*1*), have advanced rapidly. The emergence of techniques such as convolutional neural networks (*2,3*) and transformers (*4*), which are designed for image recognition and natural language processing, has enabled high-performance predictions even for complex problems. While terms like "deep learning" and "artificial intelligence (AI)" have gained popularity recently, these technologies are part of a broader category of machine learning. Machine learning is a technique where algorithms, using data, discover and learn patterns and features from that data and make predictions or classifications based on the learned results. A key feature of machine learning is that, instead of humans manually defining rules for predictions (*e.g.*, if a measurement is above 1, classify as A, otherwise classify as B), the algorithm itself identifies patterns from the collected data and its corresponding outcomes. By finding regularities within large datasets, machine learning enables accurate predictions. As machine learning technology has progressed, its applications have expanded to various fields, including medicine, where research utilizing machine learning is actively being conducted.

In the medical field, machine learning holds great potential. It has been used to predict postoperative outcomes based on patient measurements (*5,6*) and disease risk (*7*). The realization of predictive models using machine learning is expected to significantly contribute to the decision-making of medical professionals and to the treatment of patients. Despite its high potential for medical applications, machine learning faces a significant challenge in the form of the "black box" problem. The black box problem refers to the issue where the prediction results and processes generated by machine learning are not easily understandable by humans. As machine learning algorithms become more complex, their behavior becomes more difficult to interpret at a macro level, even though some aspects may be understood at a micro level. This complexity leads to situations where why a certain prediction was made or the thought process that underpinned it is unclear. This lack of transparency can be a major barrier to the acceptance of machine learning in the medical field, as physicians and patients may be reluctant to trust predictions with a rationale that is not clear.

A technology known as Explainable AI (XAI) (*8,9*) has been gaining attention as a way to address the black box problem. XAI involves analyzing machine learning models to clarify how predictions are made, identify trends in predictions, and provide reasoning for those predictions. By presenting the importance of various features in a manner that is understandable to humans, XAI helps to reveal the factors influencing the algorithm's outcomes. As a result, it should make machine learning more acceptable in the medical field.

The current study starts by presenting specific examples of medical applications of machine learning. Next, the mechanisms behind commonly used machine learning algorithms are described. Last, this study provides an in-depth explanation of explainable machine learning techniques as a solution to the black box problem. Through this discussion, we aim to share insights into the potential applications of machine learning in the healthcare field.

## 2. Machine learning applications in the medical field

With the rapid advancement of AI and deep learning in recent years, research utilizing machine learning for disease prediction, diagnosis, and prognosis prediction has been widely conducted in the medical field. These models are expected to analyze complex patient data and serve as tools to predict complications, recovery outcomes, and aid in decision-making with regard to treatment strategies. Here, we will describe specific examples of medical applications of machine learning that are anticipated to contribute to medical decision-making.

### 2.1. Prediction of Postoperative Complications

Machine learning is highly effective in predicting the risk of postoperative complications (*5*). For example, models have been proposed to assess the risk of severe postoperative complications such as pneumonia, acute kidney injury, deep vein thrombosis, and pulmonary embolism. By using data from 111,888 surgeries (including patient characteristics and clinical information), five different ML algorithms (logistic regression (*10*), support vector machine (SVM) (*11*), random forest (*12*), gradient boosting (*13*), and deep neural networks) were used to compare the accuracy with which postoperative complications were predicted. Results demonstrated that the combination of preoperative and intraoperative data provided the highest prediction accuracy, highlighting the effectiveness of machine learning as a tool for postoperative risk management.

### 2.2. Early prediction of diabetes and cardiovascular diseases

Machine learning models are also effectively utilized to predict diabetes and cardiovascular diseases (*7*). These models integrate a variety of data, such as family history, age, weight, blood pressure, cholesterol levels, and lifestyle habits (*e.g.*, smoking and exercise), to predict disease risk. Studies have constructed models using algorithms suited for linear relationships, such as linear regression and SVM, as well as algorithms that account for nonlinear relationships, like random forest and gradient boosting, to provide highly accurate predictions.

### 2.3. Prediction of postoperative outcomes

Machine learning has been used to predict postoperative outcomes. A study sought to predict four short-term adverse events – extended hospitalization, discharge to a location other than home, readmission within 30 days, and major complications – following anterior cervical discectomy and fusion surgery (*6*). The study explored model construction using five machine learning algorithms: TabPFN (*14*), TabNET (*15*), XGBoost (*16*), LightGBM (*17*), and Random forest. Random forest demonstrated the best performance of the five, with an AUROC ranging from 0.776 to 0.846. Estimating the risk of postoperative adverse events enables early personalized interventions for each patient, helping to manage a potential deterioration in their condition.

As demonstrated, various predictive studies using clinical data have been conducted. Table 1 summarizes additional studies related to the application of machine learning in the medical field, including the algorithms used and their purposes. Traditional techniques like linear regression and logistic regression were limited to linear problems. However, with the advancement of machine learning and improvements in the learning and predictive performance of various algorithms, these models can now be applied to more complex problems.

## 3. Representative machine learning algorithms commonly used in recent years

In the field of machine learning, various algorithms have been developed and are widely used. Among these, foundational and representative methods that can be used for classification and prediction include logistic regression, decision trees (*18*), random forest, gradient boosting, SVM, and deep learning. Logistic regression and decision trees are simple in their configuration and easy to interpret, but their predictive accuracy is relatively low. In contrast, algorithms such as gradient boosting and deep learning exhibit superior predictive performance, though they are more difficult to interpret.

### 3.1. Logistic regression

Logistic regression (*10*) is a commonly used algorithm in the medical field and is one of the fundamental algorithms in machine learning. It is particularly well-suited for binary classification tasks and operates similarly to linear regression. Logistic regression performs a weighted linear combination of the input explanatory variables and passes the result through a sigmoid function to predict probabilities between 0 and 1. The weights are parameters calculated based on the training data, and effectively determining these parameters enables predictive tasks to be performed. This process is known as learning. Simply put, learning involves finding the parameters of a function that can accurately represent the relationship between the observed explanatory variables and the target variable. This learning step is achieved through optimization techniques.

In optimization, a loss function is defined to represent the objective that needs to be minimized, and the parameters are adjusted to minimize this function. In logistic regression, the goal is to maximize the log-likelihood, which is transformed into a form that minimizes the loss function. Through this optimization, logistic regression finds the most plausible parameters that fit the characteristics of the training data, allowing it to make predictions for binary classification tasks. Logistic regression assumes that the problem is linearly separable, so it may not perform well when there is a nonlinear relationship between the explanatory and target variables. The simplicity of logistic regression, along with the interpretability provided by the weighting of each explanatory variable, has resulted in its widespread use in the medical field.

### 3.2. Decision trees

Like logistic regression, decision trees (*18*) are intuitive and easy-to-interpret machine learning algorithms. A decision tree classifies input data by recursively splitting it according to specific rules. The structure formed by these splits resembles a tree, as shown in Figure 1, which is why it is called a decision tree. Figure 1 illustrates a tree structure that predicts whether the temperature on a given day will exceed 22°C based on inputs such as weather and season. In this tree structure, the path is determined from the top of the tree, based on the values of the input data. If, for example, the weather is sunny, the model follows the path on the right, while if the weather is cloudy or rainy it follows the path on the left. This process is repeated until the model predicts whether the temperature will exceed 22°C. Constructing a tree structure that accurately represents the data is essential, and algorithms such as ID3 (*18*), CART (*19*), and C4.5 (*20*) have been proposed for this purpose.

A key strength of decision trees is that the tree structure clearly shows the criteria for making predictions and which features are used, making the model easy to interpret. Unlike logistic regression, decision trees can be applied to non-linear problems. However, decision trees are prone to overfitting, meaning that they may perform too well on the training data, resulting in poor performance on unseen data.

### 3.3. Random forest

Random forest (*12*) is an algorithm that improves prediction accuracy by combining multiple decision trees. While individual decision trees are prone to overfitting and may exhibit low predictive performance, a random forest generates multiple decision trees and aggregates their predictions to enhance accuracy. The term "forest" refers to the collection of decision trees. This approach of combining weak predictors – multiple decision trees – to improve overall performance is called ensemble learning. A random forest operates in three main steps: bootstrap sampling, decision tree construction, and prediction aggregation.

In the first step, bootstrap sampling, the training data are divided into several sub-datasets. In the second step, decision trees are constructed for each sub-dataset using randomly selected subsets of input features. By randomly sampling the features, each decision tree learns from a different combination of variables, increasing the diversity of the trees and helping to prevent overfitting. In the third step, the predictions from the various decision trees are aggregated, either through majority voting or by averaging, to make the final prediction. A random forest offers higher accuracy and is less prone to overfitting compared to individual decision trees. However, a disadvantage of this approach is that the model is more difficult to interpret. While a single decision tree can be easily understood, interpreting how the different variables interact to produce the final prediction is challenging when multiple trees are combined.

### 3.4. Gradient boosting

Gradient boosting (*13*) is another type of ensemble learning that combines multiple weak predictors (usually decision trees) to build a strong model. While a random forest aggregates the predictions of multiple decision trees, gradient boosting takes a different approach by sequentially creating decision trees, where each new tree is trained to correct the errors made by the previous ones. The process starts by creating an initial decision tree, which typically results in significant errors between the predicted values and the actual data. To address this, the errors between the predicted results and the actual values are calculated. A new decision tree is then trained to predict these errors. Combining the outputs of the initial tree and the subsequent tree, which focuses on correcting mistakes, improves the overall performance of the model.

This process of error correction is repeated, allowing the model to refine itself and reduce prediction errors

**Table 1. Overview of Studies Applying Machine Learning in the Medical Field**

| Study | Machine Learning Algorithms Used | Prediction Performance |
|---|---|---|
| Prediction of Acute Kidney Injury after Cardiac Surgery (*24*) | Logistic Regression, SVM, Random Forest (RF), XGBoost, RF + XGBoost | Area Under the Curve (AUC): 0.843 (RF + XGBoost) |
| Prediction of Postoperative Complications (Pneumonia, AKI, DVT, *etc.*) (*5*) | Gradient Boosting, Deep Neural Network (DNN), RF, SVM | AUC: 0.905 (Gradient Boosting) |
| Prediction of Acute Kidney Injury after Aortic Arch Surgery (*25*) | Logistic Regression, SVM, RF, Gradient Boosting | AUC: 0.8 (Gradient Boosting) |
| Risk Prediction of Diabetes and Cardiovascular Diseases (*7*) | Logistic Regression, SVM, RF, Gradient Boosting | AUC: 0.862 (XGBoost) |
| Prediction of Postoperative Outcomes (*6*) | TabPFN, TabNET, XGBoost, LightGBM, RF | AUC: 0.776 (RF) |
| Prediction of 30-day Postoperative Mortality Risk (*26*) | Convolutional Neural Network (CNN), DNN, RF, SVM | AUC: 0.867 (CNN) |
| Prediction of Postoperative Delirium (POD) in Elderly Patients (*27*) | Logistic Regression, RF, GBM, XGBoost, Ensemble | AUC: 0.783 (Logistic Regression) |
| Prediction of Mortality Risk after Hepatocellular Carcinoma Surgery (*28*) | Logistic Regression, RF, Gradient Boosting, Decision Tree | AUC: 0.803 (RF) |
| Prediction of Postoperative Survival in Gastric Cancer Patients (*29*) | Cox Regression, Random Survival Forest, DNN | AUC: 0.868 ((DNN) |
| Prediction of ICU Admission and 30-day Postoperative Mortality Risk (*30*) | RF, Gradient Boosting, SVM, Adaptive Boosting | AUPRC: 0.38 (Gradient Boosting) |

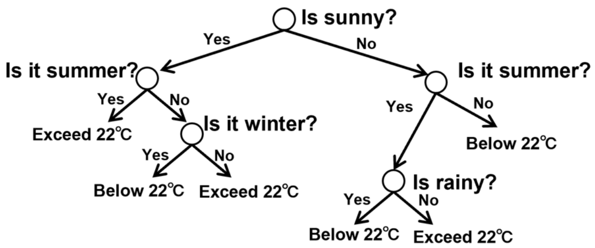


**Figure 1. Sample decision tree with splits and nodes.** This decision tree demonstrates how inputs such as season and weather conditions are used to predict whether the temperature will exceed 22°C. The tree branches represent the decision-making process, splitting based on the input features to arrive at a final prediction.

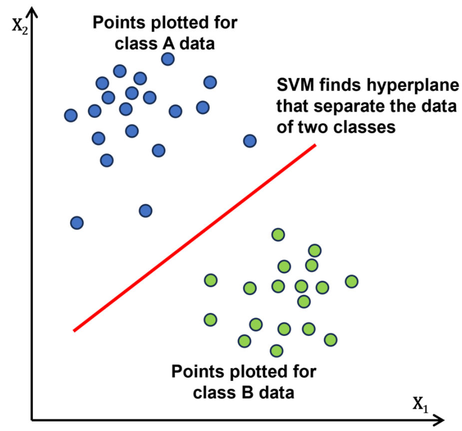


**Figure 2. Hyperplane separation of two classes using a support vector machine (SVM).** A SVM identifies the optimal hyperplane that maximizes the margin between the two classes.

with each iteration, ultimately resulting in a high level of predictive accuracy. Popular algorithms that implement gradient boosting include XGBoost (*16*) and LightGBM (*17*), which have been optimized for both performance and computational efficiency. This makes them suitable for large-scale datasets. Gradient boosting often produces more accurate models compared to decision trees and is less prone to overfitting. However, similar to a random forest, the combination of multiple decision trees makes interpreting how the model arrived at its predictions difficult, posing challenges in understanding the rationale behind the results.

3.5. SVM

SVM (*11*) is a powerful machine learning algorithm used for classification and regression problems. It works by finding an optimal boundary that separates the classes in the training data, which is then used to make predictions. For example, as illustrated in Figure 2, a dataset with two variables, X and Y, is plotted. The data belong to two classes (A and B), and each class is grouped within a certain region in a two-dimensional space. SVM finds the optimal boundary that best separates the two classes in this space. When new piece of data is plotted, if it falls on the side of the boundary corresponding to class A, it is predicted to be class A, and if it falls on the side corresponding to class B, it is predicted to be class B. While this example involves two variables and two dimensions, SVM can be extended to handle higher-dimensional data by increasing the number of explanatory variables.

Initially, SVM was designed for linear problems, but the algorithm has been improved to handle nonlinear problems as well. Linear problems are relatively easy to interpret, but interpretation becomes more challenging when dealing with nonlinear problems.

### 3.6. Deep learning

Deep learning (*1*) is a model that mimics the behavior of neurons in the brain, using artificial neurons as mathematical models. An artificial neuron receives inputs from explanatory variables, applies a weighted linear combination, and passes the result through an activation function, with the output serving as the neuron's response. If a sigmoid function is used as the activation function, this operation is nearly identical to logistic regression. In deep learning, as shown in Figure 3, multiple artificial neurons with the same input variables are constructed and treated as layers in a neural network. The output of one neural network layer is then used as the input for another, with multiple layers connected to form a full neural network. Each artificial neuron has weight parameters used in its computations, and adjusting these weights enables the neural network to achieve superior predictive performance.

The parameters are determined through training with data. Initially, random values are assigned as weights, and the output of the neural network is calculated based on the input data. The error between the predicted output and the actual values is then calculated, and the weights are adjusted to reduce the error. This process is repeated multiple times, gradually refining the parameters so that the network can accurately predict the correct output when given new input data. Conceptually, this can be viewed as a model composed of multiple connected logistic regression models.

Over the past decade, deep learning has been intensively researched, leading to various improvements and new network structures that have resulted in higher performance compared to other models. In particular, convolutional neural networks (*3*) for image recognition have gained prominence, while transformers have emerged for language processing and time-series analysis, with extensive research being conducted in these areas.

## 4. Explainable machine learning

Thus far, we have described representative machine learning algorithms. While each algorithm has its strengths and weaknesses, they all demonstrate a high level of performance. However, there is a significant challenge when applying these algorithms to the medical field: interpretability. Algorithms like logistic regression and decision trees are simple, making their predictions relatively easy to interpret. However, the more advanced algorithms developed in recent years, which exhibit excellent performance, are more complex and difficult to interpret, leading to a "black box" problem. Although the individual operations performed by the models can be understood at a micro level, interpreting the model as a whole is difficult. This challenge is known as the black box problem, and it is a significant issue in fields like medicine, where rationales for and explanations of diagnoses are especially important.

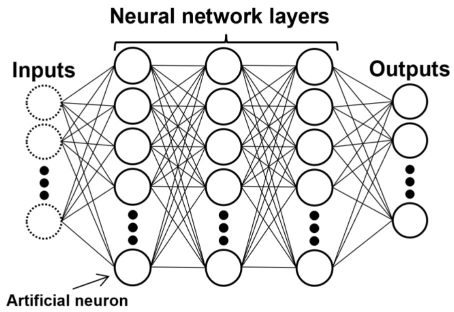


**Figure 3. Conceptual Structure of Neural Networks.** Each circle represents an artificial neuron that receives inputs from the preceding layer, applies a weighted linear combination, and passes the result through an activation function. Multiple layers of neurons are shown, where the output of one layer becomes the input to the next, forming a deep network.

To address the black box problem, efforts are underway to develop technologies that can explain the internal structure and decision-making processes of models in a way that humans can understand. These technologies are collectively referred to as XAI, and several approaches are emerging in this area (*8,9*). XAI primarily attempts to explain machine learning models from two perspectives: global and local explanations.

### 4.1. Global explanations

Global explanations aim to describe the overall characteristics of the model itself. A machine learning model learns from training data to obtain parameters and a structure that allows it to perform predictive tasks. By analyzing which explanatory variables the model emphasizes when making predictions, a technique called Feature Importance can be used to calculate and assess which variables are most important to the model. Another global interpretability approach involves constructing a simplified model that is easier to interpret and using that model to understand the behavior of the more complex model. For example, a simplified interpretable model, such as a decision tree or logistic regression, can be used to approximate the behavior of a deep learning model. The deep learning model, seen externally, functions as

**Table 2. Summary of Explainable AI Techniques***

| Explanation Type | Method/Technique | Description | Use Case |
|---|---|---|---|
| Global Explanation | Feature Importance | Evaluates which variables the model emphasizes for prediction and identifies important ones. | Analyzing how certain features impact predictions across the model. |
| | Surrogate Models (decision trees, logistic regression) | Simplifies complex models (*e.g.*, deep learning) by approximating them with interpretable models like decision trees or logistic regression. | Using a simple model to explain the behavior of complex models. |
| Local Explanation | Shapley Additive Explanations Local Interpretable Model-agnostic Explanations | Provides a local explanation by showing which features contributed and how much to a specific prediction. | • Understanding key factors for predicting based on input data.<br>• Helping a physician understand why a specific prediction was made for a patient. |
| | Influence Functions | Calculates how individual training samples influenced a specific prediction. | Identifying which past cases in training data most influenced a given diagnosis. |

*The methods are categorized into global explanations, which provide insights into the overall behavior of a model, and local explanations, which offer case-specific rationales for individual predictions.

a predictor that outputs some result when given input data. By collecting the outputs from various inputs and using these data to train a decision tree, the tree will approximate the behavior of the deep learning model. The decision tree can then be visualized, helping to explain why certain outputs are predicted based on specific input variables. This type of global explanation can help identify important explanatory variables and can be used to improve models. Moreover, if the explanations provided by the model align with existing research, this can enhance the model's validity and credibility.

### 4.2. Local explanations

Local explanations, in contrast, provide insights into specific predictions made by the model when given particular input data. For instance, if a machine learning model predicts the presence or absence of a disease based on electronic health record data, a physician might have difficulty understanding why the model predicted that the patient has the disease or why it predicted that the patient does not. Local explanations provide explanations for these individual cases. There are several methods of providing local explanations, but two commonly used approaches are described here. One method identifies the main factors that contributed to the prediction. If, for example, a model predicts that a patient has diabetes, XAI might indicate that blood sugar levels and hemoglobin in the electronic health record were particularly high, indicating which factors the model considered important. XAI techniques like Shapley Additive Explanations (SHAP) (*21*) and Local Interpretable Model-agnostic Explanations (LIME) (*22*) are used to achieve this. Another approach involves finding similar past cases to provide an explanation. If, for example, a model predicts that a patient has diabetes, XAI might search through the training data to find similar cases and present a rationale such as "the selected patient was also diagnosed with diabetes under similar conditions". Influence Functions (*23*) are often used to provide these explanations. Influence Functions calculate how much each training sample contributed to a given prediction. Applying this method to the mode enables determination of which training data samples were most influential in shaping the model. By reviewing the most influential samples that relate to diabetes, one can understand which past data the model relied on when making its prediction. Influence Functions can also be used to improve models by identifying abnormal data that disproportionately influence the model's predictions. Such data might represent outliers.

Therefore, XAI techniques are being proposed to provide both global and local explanations, and they are being used to improve the interpretability of machine learning models. Table 2 summarizes the XAI methods discussed thus far. While XAI is still an evolving field, it is steadily providing a foundation for offering rational explanations, addressing the black box problem, and facilitating the practical use of machine learning in the medical field.

## 5. Conclusion

Thanks to the advent of deep learning in particular, machine learning has demonstrated great potential in various fields, including medicine. Its ability to analyze large, complex datasets and make accurate predictions offers significant advantages in predicting diseases, diagnosing conditions, and assisting in treatment planning. However, the use of machine learning in the medical field still faces important challenges particularly with regard to the interpretability of these models.

Traditional models like logistic regression and decision trees are relatively simple, making their predictions easier to explain. In contrast, more advanced models such as gradient boosting, random forest, and deep learning—despite their superior predictive accuracy—tend to behave like "black boxes." This lack of transparency is a major obstacle in the medical field, where clinicians and patients need to understand the rationale behind predictions for them to be accepted and trusted. The development of XAI techniques is crucial to addressing this issue. XAI aims to bridge the gap between the high performance of modern machine learning models and the need for understandable, interpretable predictions. The development of XAI tools such as SHAP, LIME, and Influence Functions allows for the use of machine learning in medicine with greater confidence. These tools not only offer transparency but also reinforce the reliability and validity of the models, helping to align predictions with established medical knowledge.

As machine learning continues to evolve, integrating these explainability techniques will be essential to ensuring its practical and ethical use in healthcare. The future of medicine may increasingly rely on machine learning, and with it, explainable models may become an indispensable tool to enhance both diagnostic accuracy and decision-making processes. Through these advances, machine learning can greatly help to improve patient care, facilitate personalized treatment strategies, and aid healthcare professionals in making informed decisions.

*Funding*: None.

*Conflict of Interest*: The author has no conflicts of interest to disclose.

## References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2016; 521:436-444.
2. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998; 86:2278-2324.
3. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017; 60:84-90.
4. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L. Attention is all you need. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
5. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, Kannampallil T, Abraham J. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. JAMA Netw Open. 2021; 4:e212240.
6. Karabacak M, Bhimani AD, Schupper AJ, Carr MT, Steinberger J, Margetis K. Machine learning models on a web application to predict short-term postoperative outcomes following anterior cervical discectomy and fusion. BMC Musculoskelet Disord. 2024; 25:401.
7. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak. 2019; 19:211.
8. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. J Artificial Intel Res. 2021; 70:245-317.
9. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Comput Surv. 2018; 51:Article 93.
10. Cox D. The regression analysis of binary sequences (with discussion). J Roy Stat Soc. 1958; 20:215-242.
11. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995; 20:273-279.
12. Breiman L. Random forests. Mach Learn. 2001; 45:5-32.
13. Hastie T, Tibshirani R, Friedman J. Boosting and additive trees. In: The Elements of Statistical Learning. Springer; New York, NY. 2009.
14. Hollmann N, Müller SG, Eggensperger K, Hutter F. TabPFN: A transformer that solves small tabular classification problems in a second. In: Proc Internatl Conf Learning Representations; 2022.
15. Arık S, Pfister T. TabNet: Attentive interpretable tabular learning. In: Proc 35th AAAI Conf Artificial Intel. 2021; 35: AAAI-21 Technical Tracks 8.
16. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; pages 785-794.
17. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T. LightGBM: A highly efficient gradient boosting decision tree. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
18. Quinlan JR. Induction of decision trees. Mach Learn. 1986; 1:81-106.
19. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Chapman and Hall; 2017; pages 1-368.
20. Salzberg SL. Book Review: *C4.5: Programs for Machine Learning* by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. Mach Learn. 1994; 16:235-240.
21. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
22. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; pages 1135-1144.
23. Koh PW, Liang P. Understanding black-box predictions *via* influence functions. In: Proceedings of the 34th International Conference on Machine Learning, 2017; PMLR 70:1885-1894.
24. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, Chen KL, Yang CY, Lee OK. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. Crit Care. 2020; 24:478.
25. Lei G, Wang G, Zhang C, Chen Y, Yang X. Using machine learning to predict acute kidney injury after aortic arch surgery. J Cardiothorac Vasc Anesth. 2020; 34:3321-3328.
26. Fritz BA, Cui Z, Zhang M, He Y, Chen Y, Kronzer A, Ben Abdallah A, King CR, Avidan MS. Deep-learning model for predicting 30-day postoperative mortality. Br J Anaesth. 2019; 123:688-695.

27. Song YX, Yang XD, Luo YG, Ouyang CL, Yu Y, Ma YL, Li H, Lou JS, Liu YH, Chen YQ, Cao JB, Mi WD. Comparison of logistic regression and machine learning methods for predicting postoperative delirium in elderly patients: A retrospective study. CNS Neurosci Ther. 2023; 29:158-167.
28. Wang Y, Ji C, Wang Y, Ji M, Yang JJ, Zhou CM. Predicting postoperative liver cancer death outcomes with machine learning. Curr Med Res Opin. 2021; 37:629-634.
29. Zeng J, Song D, Li K, Cao F, Zheng Y. Deep learning model for predicting postoperative survival of patients with gastric cancer. Front Oncol. 2024; 14:1329983.
30. Chiew CJ, Liu N, Wong TH, Sim YE, Abdullah HR. Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission. Ann Surg. 2020; 272:1133-1139.


Received November 20, 2024; Accepted December 3, 2024.

*Address correspondence to:
Kenji Karako, Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan.
E-mail: tri.leafs@gmail.com

Released online in J-STAGE as advance publication December 8, 2024.