

國立陽明交通大學
智慧系統與應用研究所
碩士論文

Institute of Intelligent Systems
National Yang Ming Chiao Tung University
Master Thesis

透過樹模型方法處理慢性血液透析患者血管通路功能障礙
的不確定性預測

Handling Indeterminate Predictions of Vascular Access Dysfunction in Chronic Hemodialysis
Patients through Tree-Based Approaches

研究生：陳政霖 (Zheng-Lin Chen)
指導教授：馬清文 (Ching-Wen Ma)

中華民國 一一四年一月
Jan 2025

透過樹模型方法處理慢性血液透析患者血管通路功能障礙的
不確定性預測

Handling Indeterminate Predictions of Vascular Access
Dysfunction in Chronic Hemodialysis Patients through
Tree-Based Approaches

研 究 生：陳政霖
指導教授：馬清文

Student: Zheng-Lin Chen
Advisor: Dr.Ching-Wen Ma

國立陽明交通大學
智慧系統與應用研究所
碩士論文

A Thesis
Submitted to Institute of Intelligent Systems
College of Artificial Intelligence
National Yang Ming Chiao Tung University
in partial Fulfilment of the Requirements
for the Degree of
Master
in
Artificial Intelligence

Jan 2025

Taiwan, Republic of China

中華民國 一一四年一月

Acknowledgement

首先，我誠摯感謝我的指導教授馬清文老師，無論我在研究過程中遇到任何困難，老師總是給予我清晰的方向和寶貴的建議，並以開放的態度歡迎學生與老師討論，讓我受益良多。

接著，我也感謝實驗室的同學們：士軒、佩珊、秋賢以及 David，以及 AICV 實驗室的昱昇。特別感謝士軒學長，他不僅是我進入 AI 領域的引路人，更是實驗室的明燈，指引著我們的未來。此外，也感謝成功大學電管所的晉瑋與景南，與我一同完成了多項跨領域合作的專案，讓我收穫了許多寶貴的經驗。

最後，我要特別感謝我的父母，感謝你們一路以來的支持與鼓勵，始終尊重並支持我追求自己的興趣與夢想。無論我面臨什麼挑戰，你們總是默默在背後陪伴著我，給予我無限的信心與力量。從小到大，你們不僅為我提供了最好的學習環境，更教會了我做人處事的道理。你們的栽培與愛護是我成長的最大動力，也是我堅持努力的源泉。沒有你們的付出，就不會有今天的我，衷心感謝你們的付出與陪伴。

透過樹模型方法處理慢性血液透析患者血管通路功能障礙的不確定性預測

學生：陳政霖

指導教授：馬清文 教授

國立陽明交通大學 智慧系統與應用研究所

摘 要

血管通路功能不全是透析患者常見且嚴重的併發症，尤其在透析患者比例全球最高的台灣，早期診斷和管理顯得尤為重要。傳統的診斷方法，如定期監測及固定血流量閾值的應用，難以找出潛在需要執行手術的病患。為解決此問題，本研究提出了一個具備不確定感知能力的樹狀機器學習框架，旨在改善血管通路功能不全的判斷方法。

該框架結合了 Decision Trees、Random Forests 及 XGBoost 等樹狀模型，並融入了先進的不確定性量化技術。通過多次擾動的方式模擬樣本的變化，實現了對不確定樣本的識別。本研究還引入了擴展混淆矩陣及創新的不確定性指標，從多角度對模型性能進行全面評估。從實驗與驗證結果表明，該框架相較於傳統方法（如 KDOQI guidelines）具有更高的預測準確性和敏感性。將模稜兩可的樣本分類為不確定樣本的能力提升了臨床決策的可靠性，避免了過於自信的誤判並減少了不必要的干預。本研究在醫院的真實資料集中進行模擬，驗證所提出之方法的性能。

關鍵詞: 血管通路功能不全、樹模型、機器學習、不確定性量化

Handling Indeterminate Predictions of Vascular Access Dysfunction in Chronic Hemodialysis Patients through Tree-Based Approaches

Student: Zheng-Lin Chen

Advisor: Dr. Ching-Wen Ma

Institute of Intelligent Systems
National Yang Ming Chiao Tung University

Abstract

Vascular access dysfunction is a common and serious complication in hemodialysis patients, particularly in Taiwan, which has the highest proportion of dialysis patients globally. Early diagnosis and management are critically important. Traditional diagnostic methods, such as routine surveillance and fixed blood flow thresholds, often fail to identify patients who may require surgical intervention. To address this issue, this study proposes an indeterminate-aware, tree-based machine learning framework aimed at improving the assessment of vascular access dysfunction.

The framework integrates tree-based models, including Decision Trees, Random Forests, and XGBoost, with advanced uncertainty quantification techniques. By employing a multi-pass perturbation approach to simulate sample variability, the framework effectively identifies indeterminate cases. Additionally, this study introduces an extended confusion matrix and innovative uncertainty metrics to comprehensively evaluate model performance from multiple perspectives. Experimental and validation results demonstrate that the proposed framework outperforms traditional methods, such as KDOQI guidelines, in terms of predictive accuracy and sensitivity. The ability to classify ambiguous samples as indeterminate enhances the reliability of clinical decision-making, avoiding overconfident misclassifications and reducing unnecessary interventions. The method's performance has been validated using real-world hospital datasets, confirming its efficacy and robustness.

Keywords: Vascular Access Dysfunction, Tree-Based Models, Machine Learning, Indeterminacy Quantification

Table of Contents

Acknowledgement	3
摘要	i
Abstract	ii
Table of Contents	iii
List of Figures	v
List of Tables	vi
1 Introduction	1
1.1 Background	1
1.2 Motivation	1
1.3 Goal	2
1.4 Contribution	3
2 Related Works	4
2.1 KDOQI Guidelines	4
2.2 Physical Examination	5
2.3 Tree-based Approaches	6
2.4 Ensemble Learning	7
2.5 Indeterminacy Handling and Interpretability	8
3 Method	10
3.1 Model Architecture	11
3.2 Multipass Indeterminacy Estimation Method	12
3.3 Indeterminate-Aware Data Classification	15
3.3.1 Condition True	16
3.3.2 Condition False	16
4 Experiments and Results	18
4.1 Dataset	18

4.1.1	Input Data	18
4.1.2	Output Data	19
4.1.3	Data Organization	20
4.2	Feature Engineering	22
4.2.1	Data Alignment	22
4.2.2	Feature Creation	22
4.3	Evaluation Metrics	25
4.3.1	Extended Confusion Matrix	25
4.3.2	Evaluation Metrics	25
4.4	Experiment Setup	27
4.5	Experiment Results	28
4.5.1	Arteriovenous Fistula Dataset	29
4.5.2	Arteriovenous Graft Dataset	34
5	Conclusion	39
	References	41

List of Figures

3.1	Indeterminacy Classification Model	11
4.1	Overview of patient data	21
4.2	Data alignment process	22
4.3	KDOQI guidelines feature method to determine whether to perform surgery . .	23
4.4	The slope of the Qa value difference between the current and previous measure- ments relative to the days elapsed since the previous check	24
4.5	Extended confusion matrix. a: True Positive (TP), b: False Negative (FN), c: False Positive (FP), d: True Negative (TN), e: Indeterminate Positive (IP), f: Indeterminate Negative (IN)	25
4.6	Threshold changes used to determine output of indeterminacy	29
4.7	AVF data distribution	30
4.8	AVF Dataset Extended Confusion Matrix	32
4.9	AVF Dataset ROC Curve	33
4.10	AVG data distribution	35
4.11	AVG Dataset Extended Confusion Matrix	37
4.12	AVG Dataset ROC Curve	38

List of Tables

4.1	The parameter settings in AVF Dataset	30
4.2	AVF Dataset Standard Metrics of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)	31
4.3	AVF Dataset Indeterminacy Metrics (I) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)	31
4.4	AVF Dataset Indeterminacy Metrics (II) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)	31
4.5	The parameter settings in AVG Dataset	34
4.6	AVG Dataset Standard Metrics of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)	35
4.7	AVG Dataset Indeterminacy Metrics (I) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)	35
4.8	AVG Dataset Indeterminacy Metrics (II) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)	36

Chapter 1

Introduction

1.1 Background

Hemodialysis is a life-sustaining treatment for patients with end-stage renal disease (ESRD), and vascular access dysfunction remains one of the most critical complications faced by these patients. In Taiwan, the global leader in the prevalence of dialysis patients, the early detection and management of vascular access dysfunction are paramount to ensuring treatment efficacy and patient survival. Traditional diagnostic methods, such as physical examinations and reliance on fixed flow rate thresholds provided by the KDOQI guidelines, have been the cornerstone for vascular access surveillance. However, these methods are limited in their ability to handle ambiguous cases or adapt to the nuanced variability in patient conditions, resulting in missed or delayed interventions.

Advances in machine learning, particularly tree-based models like Decision Trees, Random Forests, and XGBoost, have demonstrated promise in structured data classification. By incorporating indeterminacy quantification and interpretability techniques, these models offer an opportunity to refine the decision-making process in clinical settings, providing more precise and adaptive diagnostic frameworks.

1.2 Motivation

Despite the well-established guidelines for vascular access management, challenges persist in predicting vascular access dysfunction with high confidence. The variability in patient-specific factors, the limitations of flow rate thresholds, and the absence of mechanisms to manage indeterminacy in ambiguous cases highlight the need for advanced methodologies. The

integration of machine learning offers a pathway to address these challenges. However, most existing studies either focus solely on improving predictive accuracy or ignore the interpretability and indeterminacy of predictions. This gap motivates the need for a framework that not only enhances prediction reliability but also ensures that the decision-making process is transparent and actionable, particularly in high-stakes clinical environments.

1.3 Goal

The primary goal of this research is to develop an indeterminate-aware, tree-based machine learning framework for predicting vascular access dysfunction in dialysis patients. This framework seeks to address key limitations in traditional diagnostic methods by:

- Incorporating indeterminacy quantification to manage ambiguous cases.
- Applying machine learning models capable of providing clinically relevant predictions while maintaining a balance between accuracy and explainability.
- Integrating clinical guidelines, such as KDOQI, with data-driven approaches to enhance diagnostic precision and adaptability.

This work aims to improve the timeliness and accuracy of interventions, ultimately contributing to better patient outcomes and resource allocation in healthcare settings.

1.4 Contribution

This research makes the following contributions:

1. **Framework Development:** Proposing a novel indeterminate-aware framework based on tree-based models (Decision Trees, Random Forests, and XGBoost) for predicting vascular access dysfunction in hemodialysis patients. The framework integrates indeterminacy quantification, interpretability, and clinical relevance.
2. **Indeterminacy Quantification and Evaluation:** Introducing a multipass perturbation approach to quantify prediction indeterminacy, enabling the categorization of cases into determinate and indeterminate groups. Additionally, the study evaluates model performance using an extended confusion matrix, which incorporates new metrics such as leakage, overkill, indeterminacy, and imperfection to provide a comprehensive assessment of diagnostic outcomes.
3. **Integration of Clinical Knowledge and Real-World Validation:** Combining clinical guidelines, such as KDOQI thresholds, with machine learning techniques to create a data-driven approach that enhances diagnostic precision and adaptability. The proposed framework is validated using real-world datasets from clinical settings, demonstrating its effectiveness over existing guidelines and methodologies in terms of accuracy, robustness, and clinical relevance.
4. **Clinical Decision Support:** Providing a scalable decision-support tool that aligns with the needs of dialysis care by improving the timeliness and accuracy of interventions. This tool ensures that ambiguous cases are carefully flagged, reducing risks associated with misdiagnoses and unnecessary procedures.

Chapter 2

Related Works

2.1 KDOQI Guidelines

The 2006 update to the KDOQI guidelines [1] offers a comprehensive framework for managing vascular access in hemodialysis patients, addressing aspects from patient evaluation to long-term access maintenance. This update, based on studies conducted between 1997 and 2005, aims to reduce vascular access-related morbidity and improve long-term functionality. The guidelines include evidence-based recommendations for creating and maintaining arteriovenous fistulas (AVFs) and grafts (AVGs), highlighting specific thresholds for vascular flow rates: AVFs with flow rates below 500 mL/min and AVGs below 600 mL/min should be further evaluated for potential stenosis or dysfunction.

However, relying solely on these flow rate thresholds may not be sufficient for optimal decision-making regarding interventions. Recent studies have explored the integration of artificial intelligence (AI) and machine learning techniques to enhance surgical decision-making processes. For instance, AI-enabled clinical decision [2] support systems have shown promise in improving surgical outcomes by analyzing complex datasets to predict patient-specific risks and recommend tailored interventions. These advancements suggest that incorporating AI into clinical practice could augment traditional guidelines, leading to more precise and individualized patient care.

In summary, while the KDOQI guidelines provide a solid foundation for vascular access management in hemodialysis patients, integrating AI and machine learning approaches holds the potential to further refine surgical decision-making, thereby enhancing patient outcomes.

2.2 Physical Examination

Regular vascular access blood flow (Qa) surveillance [3], combined with routine clinical monitoring and physical examination, has proven effective in predicting vascular access stenosis in chronic hemodialysis patients. By analyzing data from 397 patients with arteriovenous fistulas (AVF) and arteriovenous grafts (AVG), this approach identifies optimal Qa thresholds for detecting stenosis. Absolute thresholds of <500 mL/min for AVF and <600 mL/min for AVG demonstrate high predictive performance, achieving accuracy rates of 91.54% and 72.15%, respectively. Additionally, a stricter threshold of <400 mL/min for AVF is proposed as an alternative criterion for enhanced specificity. These absolute thresholds outperform relative thresholds, which incorporate both absolute Qa levels and a 25% decline from previous measurements, underscoring the reliability of absolute benchmarks in clinical decision-making.

Integrating these thresholds with physical examinations addresses the inherent limitations of clinical monitoring alone. While AVF patients benefit significantly from this combined approach, challenges remain for AVG patients due to structural differences in grafts, necessitating supplementary diagnostic tools such as duplex ultrasound. This framework not only improves early stenosis detection but also highlights the value of multidisciplinary collaboration in vascular access management. Furthermore, it establishes a foundation for adopting advanced techniques, such as machine learning models, to enhance prediction accuracy and inform timely interventions.

2.3 Tree-based Approaches

Tree-based machine learning models, such as Decision Trees, Random Forests, and XGBoost, represent foundational and advanced techniques widely used in structured data analysis. These methods have evolved to balance interpretability, computational efficiency, and predictive accuracy, addressing key challenges in traditional machine learning approaches.

Decision Trees [4] form the basis of tree-based methods, utilizing recursive splitting to categorize or predict outcomes. Despite their simplicity and high interpretability, Decision Trees are prone to overfitting, especially in complex, high-dimensional datasets. To address this limitation, ensemble techniques like Random Forests [5] and XGBoost [6] have emerged. Random Forests employ bagging (bootstrap aggregating), which involves training multiple decision trees on random subsets of data and aggregating their outputs. This method significantly reduces variance and enhances model stability, particularly in noisy or imbalanced datasets [7].

In contrast, XGBoost exemplifies the boosting paradigm, where weak learners are iteratively combined to minimize errors from previous iterations [8]. XGBoost introduces advanced features like sparsity-aware algorithms and regularized objectives, making it highly efficient and scalable for large datasets. These innovations have positioned XGBoost as a dominant method in machine learning competitions and real-world applications, from clinical diagnostics to financial modeling [9]. Recent comparative studies also highlight advancements in gradient boosting frameworks, such as LightGBM [10] and CatBoost [11]. While LightGBM excels in training speed through selective sampling techniques, CatBoost addresses prediction shifts caused by categorical feature encoding, further enhancing generalization performance. Together, these methods provide a comprehensive suite for structured data analysis, offering adaptability across diverse domains.

Tree-based models, such as XGBoost and Random Forests, often outperform neural networks on tabular data due to the unique characteristics of these datasets and the strengths of tree-based approaches. Tabular data typically consists of a mix of numerical and categorical features, irregular patterns, and skewed or heavy-tailed distributions, which tree-based mod-

els handle efficiently by learning piecewise constant functions. This allows them to capture irregular patterns, ignore irrelevant features, and adapt to data heterogeneity without extensive preprocessing or feature engineering. In contrast, neural networks struggle with these properties, requiring significant regularization and feature transformation to achieve comparable results.

Research supports these observations. Grinsztajn et al. (2022) [12] demonstrate that tree-based models excel on medium-sized datasets (10,000 samples), where neural networks often fail to generalize due to their reliance on smooth functions and their sensitivity to uninformative features. McElfresh and Khandagale (2023) [13] further highlight that Gradient Boosted Decision Trees (GBDTs) outperform neural networks on average, particularly in handling data irregularities and dataset-specific properties. Additionally, Ye et al. (2024) [14] find that while deep learning has improved, tree-based models remain competitive, especially in scenarios with limited data or computational resources. These advantages, combined with their computational efficiency and minimal hyperparameter tuning requirements, make tree-based models a robust and practical choice for structured datasets, where neural networks' strengths in image and text data do not translate effectively.

2.4 Ensemble Learning

Ensemble learning has become a cornerstone in machine learning for enhancing model robustness and improving predictive accuracy by combining multiple base classifiers [15]. Two primary ensemble strategies, bagging and boosting, form the foundation of many successful machine learning models. Bagging methods, such as Random Forest, focus on reducing variance by training multiple classifiers independently and aggregating their outputs. This approach ensures stability and robust performance, particularly in datasets with noise or complex distributions.

In this study, we leverage the strengths of these ensemble methods by employing a soft voting classifier that combines Decision Trees, Random Forests, and XGBoost [16, 17]. The soft voting approach integrates predictions probabilistically, assigning weights based on the confidence levels of individual models. This method captures both the diversity and the complementary

strengths of the base classifiers, ensuring a more balanced and accurate prediction framework.

The decision to use a soft voting ensemble stems from its proven ability to manage structured data effectively. Unlike neural networks, which may struggle with limited structured datasets and interpretability, tree-based ensembles offer clear advantages in handling imbalanced data, feature interactions, and hierarchical relationships inherent in clinical and structured datasets.

2.5 Indeterminacy Handling and Interpretability

In this study, tree-based models were employed to address challenges in structured data classification by integrating Indeterminacy quantification and enhancing model interpretability. While the methods used are specific to tree-based models, the approach is inspired by concepts and techniques explored in other machine learning and deep learning research [18].

Indeterminacy quantification is a critical aspect of this study, as it enables the model to effectively manage ambiguous cases and provide more reliable predictions. Insights from Indeterminacy modeling [19] provide a foundational framework, offering methodologies to quantify both aleatoric and epistemic indeterminacies and evaluate their impact on predictive accuracy. These concepts guided the implementation of metrics and approaches tailored for tree-based models to improve their reliability and adaptability in structured data classification.

Another key influence is the use of Indeterminacy-based rejection techniques [20], which demonstrate the utility of quantifying prediction Indeterminacy to enhance model selection and enable the integration of diverse classifiers. Inspired by these strategies, this study incorporates mechanisms to reallocate highly indeterminate predictions to an "indeterminate" category, ensuring more cautious decision-making. This approach is particularly valuable in clinical applications where incorrect classifications could lead to significant consequences.

The focus on interpretability in this study also draws from methods that emphasize local perturbations to approximate complex model behavior [21], providing interpretable explanations for individual predictions. These techniques influenced the design of Indeterminacy-aware explanations for tree-based models, facilitating a deeper understanding of decision boundaries and

feature contributions for each prediction.

Lastly, the study integrates principles inspired by methods that estimate Indeterminacy using randomness in model computations [22]. Although initially designed for neural networks, these concepts informed the multipass perturbation approach employed here, where noise is introduced into input features to analyze prediction variability. This method enables the approximation of Indeterminacy in ensemble tree-based models, extending the applicability of these ideas to structured data classification.

This integrated framework leverages existing advancements in Indeterminacy quantification and interpretability to adapt them for tree-based models, enabling robust and explainable predictions in complex real-world scenarios.

陽明交大
NYCU

Chapter 3

Method

This chapter describes the experimental methods and technical processes employed in this study and elaborates on how indeterminacy analysis can enhance the predictive capability and stability of the proposed models. The primary objective of this chapter is to propose an indeterminacy judgment method that predicts outcomes by introducing multiple instances of varying noise to the same sample, thereby quantifying indeterminacy and optimizing classification decisions accordingly.

The methodology presented in this chapter is structured into three main components. First, Section 3.1 outlines the data preprocessing techniques, the design of the machine learning models, and the training process used in this study. Second, Section 3.2 explains how multiple passes through perturbed samples are used to analyze sample indeterminacy and compute the probability means and variances of the predictions. Finally, Section 3.3 demonstrates how classification decisions are made based on the indeterminacy evaluation results, categorizing samples into "determinate" (True/False) or "indeterminate" (Indeterminate) classes.

In modern clinical applications, there is an increasing demand for diagnostic precision and robustness, particularly in the early detection of vascular access dysfunction. Traditional diagnostic approaches often rely on fixed blood flow thresholds, which can lead to misdiagnoses or missed diagnoses due to fluctuations in individual values. The primary goal of this study is to address this problem by introducing the concept of indeterminacy analysis, which adjusts classification decisions to ensure that the system provides more cautious diagnostic outcomes for ambiguous samples. The framework proposed in this chapter effectively addresses the limitations of traditional diagnostic methods, enabling the system to classify clear samples accurately while also offering indeterminacy evaluations for borderline cases, thereby enhancing its utility in clinical applications.

3.1 Model Architecture

In this chapter, we introduce the proposed Indeterminacy Classification Model designed to classify samples, as illustrated in Figure 3.1.

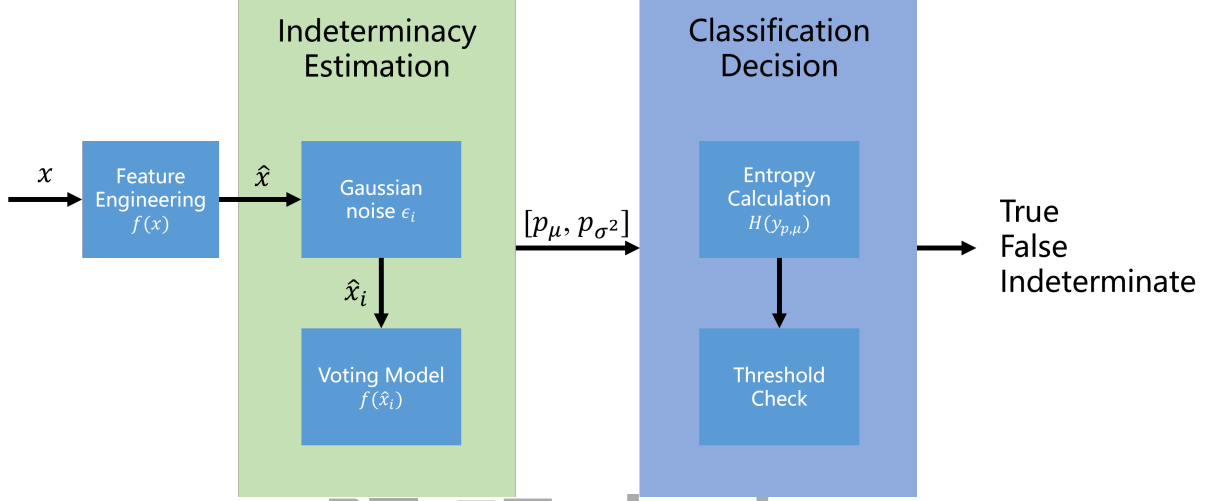


Figure 3.1: Indeterminacy Classification Model

The structured dataset underwent preprocessing steps, which included removing redundant information, feature creation, and aligning the data. These steps ensure that the dataset is clean and suitable for training machine learning models. For a detailed explanation of the preprocessing steps, please refer to Chapter 4.2.

The models utilized in this study are tree-based machine learning classifiers, specifically:

- **Decision Tree:** A fast and interpretable model; however, it is prone to overfitting when applied to high-dimensional data.
- **Random Forest:** An ensemble of decision trees that effectively reduces model variance and improves prediction stability.
- **XGBoost:** A gradient boosting framework renowned for its computational efficiency and strong classification performance, particularly in imbalanced datasets.

To leverage the strengths of these models, we implemented a weighted soft voting strategy, combining the prediction probabilities from each model. This approach allows the ensemble

to capture diverse perspectives from the individual classifiers. Furthermore, to address the imbalance between positive and negative samples, we adjusted the `scale_pos_weight` parameter in XGBoost, ensuring that the model assigns appropriate importance to minority class samples. All models were trained and evaluated using K-fold cross-validation, which ensures robust and reliable performance.

After constructing the soft voting model, it was integrated into the Multipass Indeterminacy Estimation Method and the Indeterminate-Aware Data Classification process. These methods enhance the classification framework by quantifying and addressing indeterminacy, further improving the model's applicability in complex scenarios.

3.2 Multipass Indeterminacy Estimation Method

To quantify the indeterminacy of each sample, we propose the Multipass Indeterminacy Estimation Method. The key idea is to generate multiple perturbed versions of each sample by adding Gaussian noise $\epsilon = N(\mu, \sigma^2)$. Each perturbed version of the sample is then passed through the weighted soft voting classifier, which predicts the probabilities for True and False. The predicted probabilities for True and False across all perturbed versions are then aggregated to calculate their respective means (p_u) and variances (p_{σ^2}). This process is repeated for each sample individually.

The main steps of this method are:

1. **Perturbed Sample Generation:** Adding Gaussian noise to the input features to simulate data indeterminacy, creating multiple perturbed versions of the sample.
2. **Multi-Model Prediction:** Passing each perturbed version through the soft voting classifier, which integrates predictions from Decision Tree, Random Forest, and XGBoost to generate probabilities for *True* and *False*.
3. **Probability Mean and Variance Calculation:** Aggregating the predicted probabilities for True and False across all perturbed versions by calculating their respective means (p_u) and variances (p_{σ^2}).

The use of multiple perturbed versions of a sample considers the fact that a patient's current blood flow rate is not fixed but continuously fluctuating. By generating multiple perturbed inputs representing variations in blood flow rates for the same patient, the model can simulate real-world conditions. This approach enables the model to provide more confident and robust predictions under varying circumstances.

陽明交大
NYCU

Algorithm 3.1 Multipass Indeterminacy Estimation

Input:

\hat{x} : Input sample, μ : Noise mean, σ^2 : Noise variance

$W = [w_1, w_2, w_3]$: Weight of estimator, n : Number of estimation times

Estimators: [*DecisionTree*, *RandomForest*, *XGBoost*]

Output:

p_μ : Mean of probability, p_{σ^2} : Variance of probability

Initialize:

$\hat{P} = [\quad]$

```
1: for estimate  $i$  from 1 to  $n$  do
2:    $\epsilon_i \sim N(\mu, \sigma^2)$ 
3:    $\hat{x}_i \leftarrow \hat{x} + \epsilon_i$ 
4:    $P_{y_i} = [ \quad ]$ 
5:   for  $clf$  in Estimators do
6:      $v_{clf} = \text{VotingClassifier}(clf, \text{voting} = 'soft')$ 
7:      $p_{y_i, clf} = v_{clf}.PredictProb(\hat{x}_i) \# p = [p_T, p_F]$ 
8:     Append  $p_{y_i, clf}$  to  $P_{y_i}$ 
9:   end
10:   $\hat{p}_i = \text{average}(P_{y_i}, W)$ 
11:  Append  $P_{y_i}$  to  $\hat{P}$ 
12: end
13:  $p_\mu = \text{mean}(\hat{P}) \# p_\mu = [p_{T, \mu}, p_{F, \mu}]$ 
14:  $p_{\sigma^2} = \text{variance}(\hat{P}) \# p_{\sigma^2} = [p_{T, \sigma^2}, p_{F, \sigma^2}]$ 
15: return  $p_\mu, p_{\sigma^2}$ 
```

Algorithm 3.1 demonstrates the process of performing Multipass Indeterminacy Estimation. This method involves iteratively sampling noise ϵ from a standard Gaussian distribution to simulate variability. The noise ϵ is then combined with the original blood flow rate of the sample, producing perturbed blood flow rates. These perturbed values are subsequently passed through

an ensemble learning model with a soft voting mechanism, where the predictions from individual models are weighted and averaged to generate probability distributions for the perturbed blood flow rates. Finally, the mean and variance of these probability distributions are calculated to determine the likelihood of whether the sample requires surgical intervention. This approach quantifies the indeterminacy inherent in the prediction, ensuring robust and confident decision-making.

After processing all perturbed versions of a sample, the method moves on to the next sample, repeating the process until all samples are evaluated. This approach effectively captures data indeterminacy and provides robust and reliable prediction outputs.

3.3 Indeterminate-Aware Data Classification

Based on the results of the multipass method, we propose an Indeterminate-Aware Data Classification strategy. This method classifies samples into three categories: *True*, *False*, and *Indeterminate*.

Algorithm 3.2 demonstrates classifying which data are determinate or which data are indeterminate. This method is implemented in two steps. First, the information entropy $H(y_{p,u})$ of the probability distribution is calculated to quantify its uniformity. If the entropy exceeds a predefined threshold θ_u , the sample is classified as *Indeterminate*, indicating insufficient confidence in the prediction.

$$H(y_p, \mu) = -(p_{T,\mu} \log p_{T,\mu} + p_{F,\mu} \log p_{F,\mu}) \quad (3.1)$$

Second, for samples with entropy below the threshold, their probability distributions' means and variances are further evaluated to classify them as *True* or *False*. Samples with minimal mean differences and large variances are reclassified as *Indeterminate* to reduce misclassification risks.

The classification decisions in Indeterminate-Aware Data Classification are guided by the probability means $p_{T,\mu}, p_{F,\mu}$ and variances $p_{T,\sigma^2}, p_{F,\sigma^2}$ of the predicted probabilities for "True"

and "False." These two equations are used to classify a sample as "True" or "False" while accounting for indeterminacy in the predictions.

3.3.1 Condition True

This equation determines whether the sample is classified as "True." It includes two conditions:

$$p_{T,\mu} > p_{F,\mu} \text{ and } p_{T,\mu} - \sqrt{p_{T,\sigma^2}} > p_{F,\mu} + \sqrt{p_{F,\sigma^2}} \quad (3.2)$$

- $p_{T,\mu} > p_{F,\mu}$: The mean probability of "True" is greater than that of "False."
- $p_{T,\mu} - \sqrt{p_{T,\sigma^2}} > p_{F,\mu} + \sqrt{p_{F,\sigma^2}}$: This ensures that even after accounting for the variability (as represented by the standard deviation), the confidence in "True" exceeds that of "False." The subtraction and addition of the square root of variances serve to introduce a conservative buffer for decision-making.

Only when both conditions are met does the sample get classified as "True." This approach reduces the risk of misclassification due to overlapping confidence intervals.

3.3.2 Condition False

This equation mirrors Equation 3.2 but applies to the classification of "False." Similarly, it requires:

- $p_{F,\mu} > p_{T,\mu}$: The mean probability of "False" is greater than that of "True."
- $p_{F,\mu} - \sqrt{p_{F,\sigma^2}} > p_{T,\mu} + \sqrt{p_{T,\sigma^2}}$: After accounting for variability, the confidence in "False" must still dominate over "True."

$$p_{F,\mu} > p_{T,\mu} \text{ and } p_{F,\mu} - \sqrt{p_{F,\sigma^2}} > p_{T,\mu} + \sqrt{p_{T,\sigma^2}} \quad (3.3)$$

When both conditions are satisfied, the sample is classified as "False." Like the previous equation, this ensures a robust classification by considering prediction indeterminacy.

Algorithm 3.2 Indeterminate-Aware data classification

Input:

p_μ : Mean of probability, which is $[p_{T,\mu}, p_{F,\mu}]$

p_{σ^2} : Variance of probability, which is $[p_{T,\sigma^2}, p_{F,\sigma^2}]$

θ_μ : Threshold used to determine output of indeterminacy

Output:

"True", "False" or "Indeterminate"

```

1:  $H(y_p, \mu) \leftarrow -(p_{T,\mu} \log p_{T,\mu} + p_{F,\mu} \log p_{F,\mu})$ 
2: if  $H(y_p, \mu) > \theta_\mu$  then
3:   return "Indeterminate"
4: else
5:   if  $p_{T,\mu} > p_{F,\mu}$  and  $p_{T,\mu} - \sqrt{p_{T,\sigma^2}} > p_{F,\mu} + \sqrt{p_{F,\sigma^2}}$  then
6:     return "True"
7:   else if  $p_{F,\mu} > p_{T,\mu}$  and  $p_{F,\mu} - \sqrt{p_{F,\sigma^2}} > p_{T,\mu} + \sqrt{p_{T,\sigma^2}}$  then
8:     return "False"
9:   else
10:    return "Indeterminate"

```

This classification strategy effectively handles boundary samples, minimizing erroneous predictions while providing higher confidence in clinical decision-making.

Chapter 4

Experiments and Results

4.1 Dataset

This study utilizes a dataset provided by a hospital, containing 5,860 records from periodic patient examinations. The dataset includes data from 412 patients and is categorized into three types of data to support the analysis: constant data, serial data, and output data. Furthermore, the data is divided into two subsets based on the type of vascular access: Arteriovenous Fistula (AVF) and Arteriovenous Graft (AVG).

4.1.1 Input Data

1. **Constant Input (Clinical Features):** The constant input includes patient-specific clinical characteristics, which remain consistent over time. The selected features for this study are as follows:

Demographics :

- Age
- Gender

Comorbidities :

- Diabetes Mellitus (DM)
- Hypertension (HTN)
- Dyslipidemia (Dyslipid)
- Coronary Artery Disease (CAD)

- **Acute Myocardial Infarction (AMI)**
- **Cerebral Vascular Accident (CVA)**
- **Peripheral Artery Occlusive Disease (PAOD)**
- **Heart Failure (HF)**

These features capture the clinical background of the patient, providing valuable context for predicting procedural outcomes.

2. **Serial Input (Procedural Features):** The serial input focuses on vascular access-related features that vary over time. The selected features include:

- **Location of the Access Site**
- **Type of Vascular Access (AVF or AVG)**
- **QA value:** A clinical measurement indicative of vascular access functionality.

By combining these features with constant inputs, the dataset offers a comprehensive representation of both static and dynamic aspects of patient health. Additionally, the dataset is divided into two subsets based on the "Type of Vascular Access":

- **Arteriovenous Fistula (AVF) Dataset**
- **Arteriovenous Graft (AVG) Dataset**

4.1.2 Output Data

The output data represents the procedural outcome, specifically whether the patient underwent a percutaneous transluminal angioplasty (PTA). The label is determined based on the following two features:

- **PTA_S (Simple PTA):** Indicates that the PTA procedure addressed only a single site of stenosis.

- **PTA_C (Complex PTA):** Indicates that the PTA procedure addressed two or more sites of stenosis.

The final label, PTA, is derived by considering the union of these two features, effectively capturing all cases where a PTA procedure was performed, regardless of its complexity.

4.1.3 Data Organization

To ensure consistency and clarity:

- The dataset contains 5,860 records collected from 412 patients during regular clinical monitoring.
- Constant input features (clinical data) and serial input features (vascular access data) are combined for each patient.
- Based on the "Type of Vascular Access," all data is stratified into two groups:
 - **Arteriovenous Fistula (AVF):** For patients with native fistulas.
 - **Arteriovenous Graft (AVG):** For patients with synthetic grafts.

This structured dataset provides a solid foundation for building machine learning models that incorporate both clinical and procedural data to predict procedural outcomes effectively.

To better understand the distribution of patients in the dataset, we analyzed the proportions of patients with Arteriovenous Fistula (AVF) and Arteriovenous Graft (AVG), as well as the distribution of patients who underwent Percutaneous Transluminal Angioplasty (PTA) procedures.

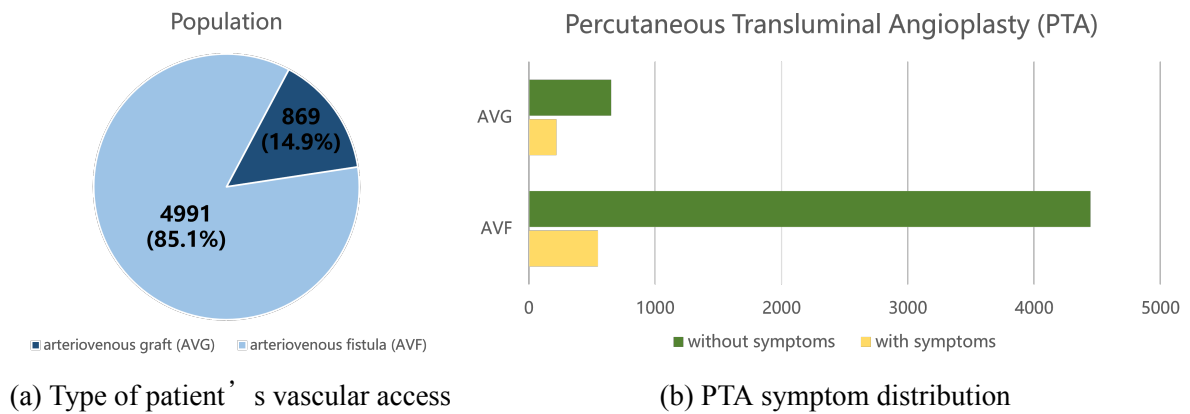


Figure 4.1: Overview of patient data

Figure 4.1(a) illustrates the distribution of the dataset based on the type of vascular access. Among the 5,860 records, 85.1% (4,991 records) belong to patients with AVF, while 14.9% (869 records) are from patients with AVG. This highlights a significant predominance of AVF cases in the dataset, with AVF records outnumbering AVG records by approximately 6 times. Figure 4.1(b) shows the distribution of patients who underwent PTA procedures, further divided into those with and without symptoms. For AVF patients, the majority did not undergo surgery, with 4,445 (89.1%) cases categorized as "no PTA," while 546 (10.9%) cases underwent PTA. Among those who underwent PTA, the majority were asymptomatic. For AVG patients, 651 (74.9%) cases did not undergo PTA, while 218 (25.1%) cases underwent PTA. Similar to AVF, most AVG patients who underwent PTA procedures were asymptomatic. However, the proportion of AVG patients undergoing PTA (25.1%) is significantly higher compared to AVF (10.9%). The data highlights the disparity in procedural intervention rates between the two vascular access types. AVF patients appear less likely to undergo PTA compared to AVG patients.

These visualizations emphasize the diversity of patient profiles in the dataset and provide a foundation for the analysis and modeling performed in subsequent sections.

4.2 Feature Engineering

To ensure the dataset is suitable for analysis and accurately represents the clinical scenarios, two feature engineering methods were applied: Data Alignment and Feature Creation. These steps enhance the model’s ability to capture meaningful patterns and improve prediction performance.

4.2.1 Data Alignment

In the original dataset, there is no Qa value recorded for the day of surgery. Instead, the dataset provides the difference in days between the current record and the previous one. To address this, the record immediately preceding the surgery was selected to represent the surgical day. This approach aligns the data temporally, ensuring that the features used in the model correspond to the patient’s most recent clinical status before the procedure. Figure 4.2(b) illustrates the data alignment process.

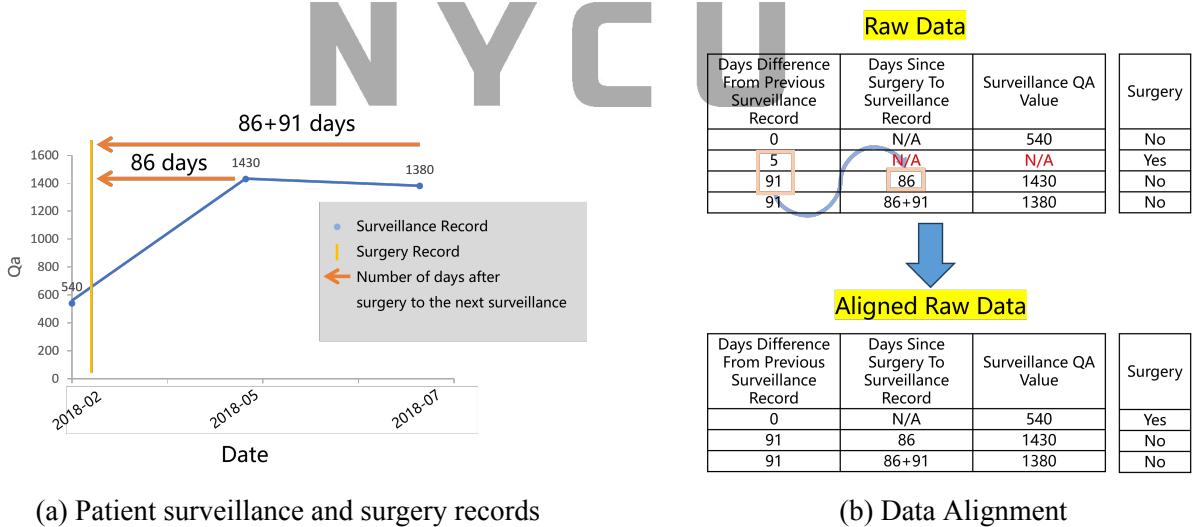


Figure 4.2: Data alignment process

4.2.2 Feature Creation

As shown in Figure 4.3, the KDOQI (Kidney Disease Outcomes Quality Initiative) guidelines define three critical features related to vascular access dysfunction: (1) arteriovenous fis-

tula (AVF) with a Qa value less than 400mL/min to 500 mL/min, (2) arteriovenous graft (AVG) with a Qa value less than 600 mL/min, and (3) cases where the Qa value is less than 1000 mL/min and has dropped by 25%. These features serve as clinically significant indicators for evaluating the performance and functionality of vascular access. Inspired by these guidelines, additional features were created to enhance the model's ability to predict surgical outcomes. These newly created features capture variations in blood flow that may indicate vascular access dysfunction, aiming to provide the model with clinically meaningful variables that align with real-world diagnostic practices.

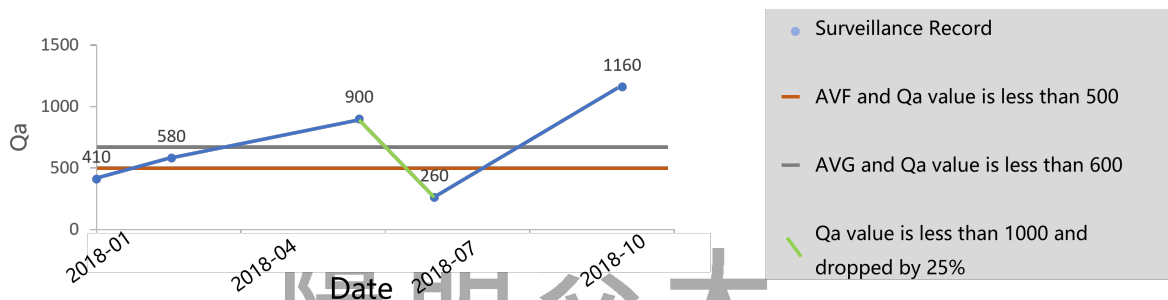


Figure 4.3: KDOQI guidelines feature method to determine whether to perform surgery

To enhance the model's ability to capture historical trends and assess the current state of vascular access functionality, several new features were created. These features provide each sample with information from previous measurements, enabling the model to analyze temporal patterns and changes. The created features are as follows:

- **Qa Value from the Previous Check:** This feature includes the Qa value recorded in the patient's most recent prior check, providing context on the patient's previous vascular access condition.
- **Previous Surgery Indicator:** A binary feature indicating whether a surgical procedure was performed during the previous check. This feature helps identify the impact of recent interventions on the current condition.
- **Slope of Qa Value Difference:** This feature calculates the slope of the Qa value difference between the current and previous measurements, normalized by the number of days

elapsed since the last check. It represents the rate of change in Qa value over time, offering a quantitative measure of vascular access deterioration or improvement. The calculation process for this feature is illustrated in Figure 4.4.

- **Difference in Qa Values:** The absolute difference between the current and previous Qa values. This feature captures the magnitude of change in vascular access functionality between two consecutive checks.

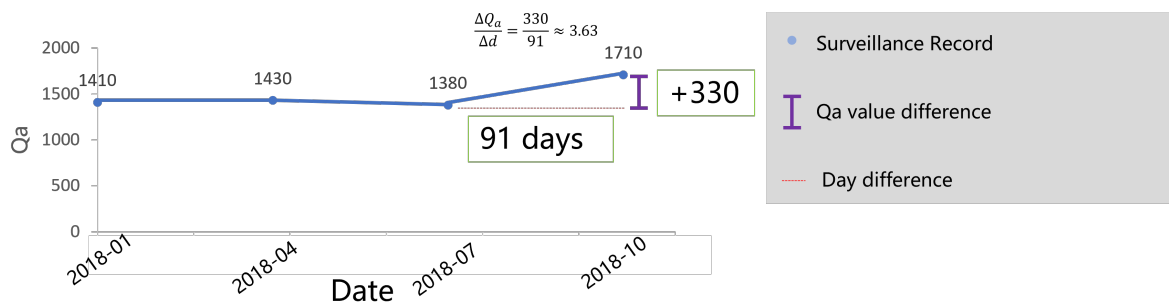


Figure 4.4: The slope of the Qa value difference between the current and previous measurements relative to the days elapsed since the previous check

By incorporating these newly created features, each sample is enriched with its historical context, enabling the model to assess trends and variations effectively. The slope and Qa value difference features, in particular, allow the model to detect significant changes in vascular access functionality, which are critical for predicting surgical needs.

4.3 Evaluation Metrics

To comprehensively evaluate the performance of the proposed model, we designed an Extended Confusion Matrix that accounts for an additional category of predictions: Indeterminate. This extended framework enables a more nuanced evaluation of model performance, particularly in cases where predictions are inconclusive.

4.3.1 Extended Confusion Matrix

Figure 4.5 shows the Extended Confusion Matrix, which expands upon the traditional confusion matrix by including two new categories:

- **IP (Indeterminate Positive):** Actual positive cases predicted as indeterminate.
- **IN (Indeterminate Negative):** Actual negative cases predicted as indeterminate.

	Predicted Positive	Predicted Negative	Predicted Indeterminate
Actually Positive	a	b	e
Actually Negative	c	d	f

Figure 4.5: Extended confusion matrix. a: True Positive (TP), b: False Negative (FN), c: False Positive (FP), d: True Negative (TN), e: Indeterminate Positive (IP), f: Indeterminate Negative (IN)

4.3.2 Evaluation Metrics

Based on Figure 4.5 the Extended Confusion Matrix, we derived several evaluation metrics to measure model performance, including both standard and indeterminacy-aware metrics:

- **Standard Metrics:**

$$Accuracy = \frac{a + d}{a + b + c + d} \quad (4.1)$$

$$PPV = \frac{a}{a + c} \quad (4.2)$$

$$NPV = \frac{d}{b + d} \quad (4.3)$$

• **Indeterminacy Metrics:**

$$All = a + b + c + d + e + f \quad (4.4)$$

$$Error = \frac{b + c}{All} \quad (4.5)$$

$$Leakage = \frac{b}{All} \quad (4.6)$$

$$Overkill = \frac{c}{All} \quad (4.7)$$

$$Indeterminate = \frac{e + f}{All} \quad (4.8)$$

$$Imperfection = \frac{b + c + e + f}{All} \quad (4.9)$$

$$IndeterminateRecall = \frac{a + e}{a + b + e} \quad (4.10)$$

$$Harmonic_Score = \frac{\frac{w'_1}{\exp(\frac{b}{All})} + \frac{w'_2}{\exp(\frac{c}{All})} + \frac{w'_3}{\exp(\frac{e+f}{All})}}{3} \quad (4.11)$$

Positive Predictive Value (PPV) evaluates the proportion of predicted positive cases that are actually positive, emphasizing the model's ability to minimize unnecessary interventions. Negative Predictive Value (NPV) measures the proportion of predicted negative cases that are truly negative, reducing the risk of missed diagnoses. Error Rate quantifies the proportion of all incorrect predictions, providing an overall misclassification rate. Beyond these standard metrics, Leakage Rate captures the proportion of actual positive cases classified as negative, highlighting the risk of failing to detect cases needing intervention, while Overkill Rate measures the proportion of actual negative cases classified as positive, addressing the tendency for over-treatment. Indeterminacy Rate reflects the proportion of predictions classified as indeterminate, enabling the model to flag ambiguous cases for further evaluation. Combining these, the Imperfection Rate aggregates the error and indeterminacy rates to assess the total proportion of samples either misclassified or marked as indeterminate. Additionally, Indeterminate Recall extends the

traditional recall metric by including indeterminate positive cases, ensuring the model prioritizes identifying all potential positive cases. Finally, the Harmonic Score provides a balanced measure of leakage, overkill, and indeterminacy rates using weighted components, offering an integrated assessment of the model's ability to minimize errors, reduce unnecessary interventions, and manage indeterminacy effectively. Together, these metrics form a comprehensive framework for evaluating both the predictive performance and reliability of the model.

4.4 Experiment Setup

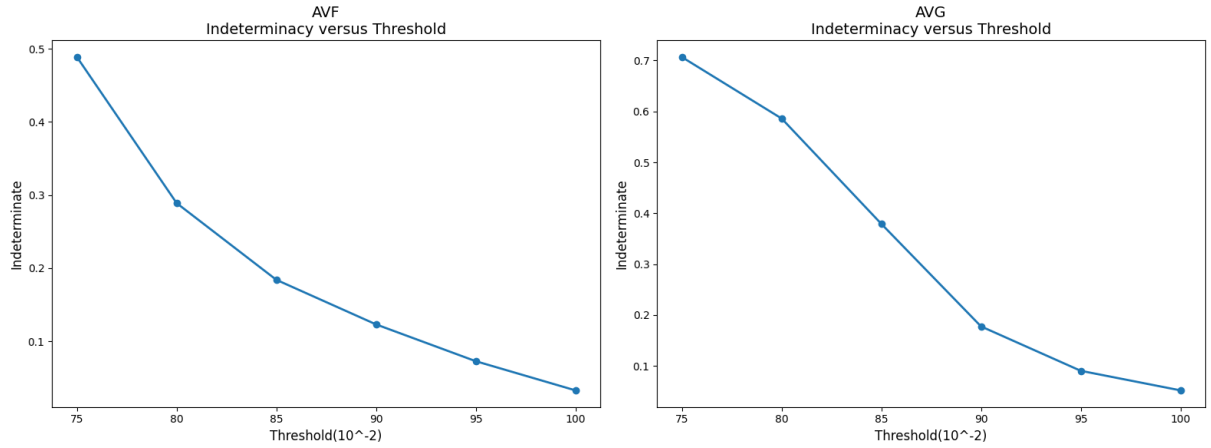
In this study, we employed an ensemble learning approach by combining Decision Tree, Random Forest, and XGBoost models into a Soft Voting model. The Decision Tree serves as a base model, providing interpretability for feature impact; Random Forest enhances model robustness through the aggregation of predictions from multiple trees; and XGBoost further improves prediction accuracy with its efficient gradient boosting algorithm. The Soft Voting model integrates the predictions from these individual models using a weighted voting mechanism, achieving overall performance improvement and effectively handling the complexity and indeterminacy in the dataset.

All three models in the Soft Voting ensemble were used with default parameters, except for the `scale_pos_weight` parameter in XGBoost, which was set based on the ratio of patients requiring surgery to those not requiring surgery in the dataset. This adjustment was made to address the data imbalance and ensure the model accurately captured the underrepresented class. The experiments were conducted using a K-fold cross-validation approach with three iterations, and the final results were averaged for consistency. All computations and model training were performed on a high-performance NVIDIA Tesla V100 GPU to ensure efficiency.

4.5 Experiment Results

This section summarizes the experimental results for the Arteriovenous Fistula (AVF) and Arteriovenous Graft (AVG) datasets under various parameter settings. By comparing the performance of the KDOQI guidelines with the proposed model, including scenarios where the model incorporates and excludes indeterminate classifications, we evaluate the advantages of our approach in improving diagnostic precision and managing indeterminacy. Key insights are derived from the results of the extended confusion matrix, ROC curves, and the evaluation metrics introduced earlier, highlighting the proposed model's ability to provide robust and clinically meaningful predictions.

Figures 4.6 illustrate the relationship between the indeterminacy rate and the threshold parameter (θ_μ) for the AVF and AVG datasets, respectively. As the threshold increases, the proportion of samples classified as indeterminate decreases, reflecting the model's growing confidence in its predictions. For the AVF dataset, the indeterminacy rate drops from 50% at a threshold of 0.75 to below 10% at a threshold of 1.00, while the AVG dataset follows a similar trend, starting at 70% and decreasing to under 10%. The higher initial indeterminacy rate for the AVG dataset highlights the increased indeterminacy associated with its smaller sample size and higher proportion of surgical cases. These trends underscore the importance of threshold tuning to balance model confidence and sensitivity to indeterminate cases. A lower threshold captures more ambiguous samples, aiding cautious decision-making, while a higher threshold emphasizes definitive classifications, improving specificity. Based on these findings, we selected a threshold of 0.95, ensuring that the indeterminacy rate remains below 10% for both datasets. This choice achieves a balance between minimizing indeterminate cases and maintaining high prediction confidence, aligning with the clinical need for precise yet cautious decision-making.



(a) Indeterminacy Rate versus Threshold for the AVF Dataset (b) Indeterminacy Rate versus Threshold for the AVG Dataset

Figure 4.6: Threshold changes used to determine output of indeterminacy

4.5.1 Arteriovenous Fistula Dataset

The analysis of the Arteriovenous Fistula (AVF) dataset highlights the significant advantages of the proposed indeterminate-aware model in improving clinical decision-making and addressing indeterminacies in vascular access predictions. This section discusses the experimental results, parameter optimization, and model evaluation, emphasizing the practical implications of the findings.

The parameter settings for the model, determined using Bayesian optimization to maximize accuracy, are presented in Table 4.1. This optimization process identified the best combination of parameters to enhance the model's predictive performance.

Table 4.1: The parameter settings in AVF Dataset

Parameter Settings	
Number of estimation times: n	10
Threshold used to determine output of indeterminacy: θ_μ	0.95
Noise mean: μ	0
Noise variance: σ^2	100
Weight of estimator: $W = [w_1, w_2, w_3]$	[1,1,1]
Weight of harmonic score: $W' = [w'_1, w'_2, w'_3]$	[1,1,1]

Figure 4.7 illustrates the distribution of the AVF dataset. Out of 4,991 cases, 546 (10.9%) required PTA interventions, while 4,445 (89.1%) did not undergo surgery. This imbalance highlights the dataset's focus on monitoring cases to avoid unnecessary surgical procedures.

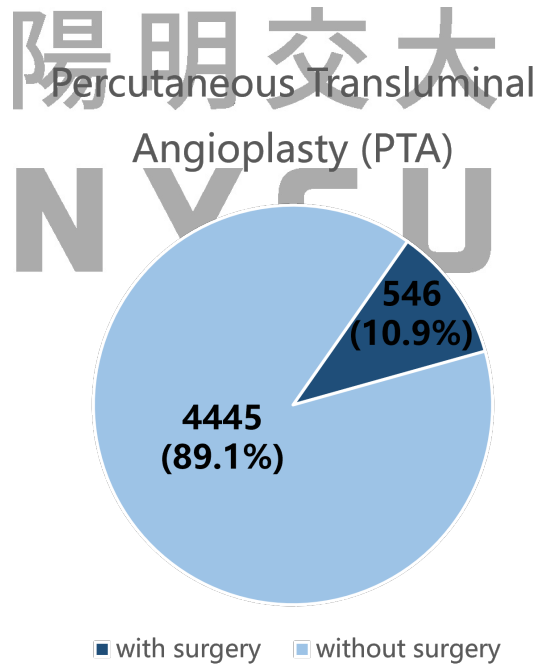


Figure 4.7: AVF data distribution

The comparison of evaluation metrics is detailed in Table 4.2, 4.3, and 4.4, showing results for the KDOQI guidelines, our proposed model without indeterminate classification, and our model with indeterminate classification:

Table 4.2: AVF Dataset Standard Metrics of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)

	Accuracy	PPV	NPV
Baseline	0.872 ± 0.001	0.394 ± 0.011	0.918 ± 0.003
Our Method(Without indeterminate)	0.895 ± 0.004	0.589 ± 0.028	0.904 ± 0.002
Our Method(indeterminate)	0.913 ± 0.001	0.841 ± 0.043	0.914 ± 0.001

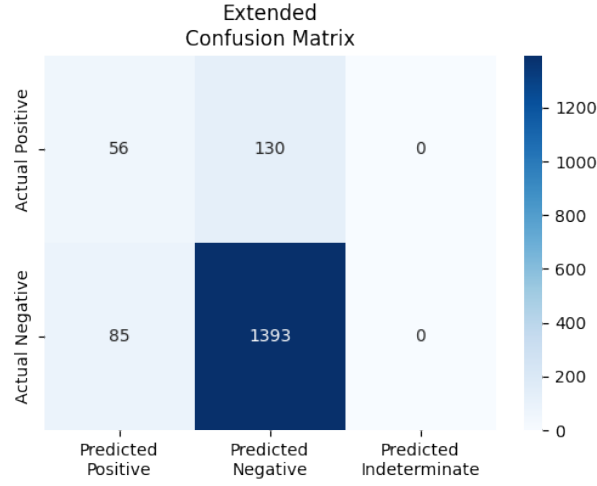
Table 4.3: AVF Dataset Indeterminacy Metrics (I) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)

	Error	Leakage	Overkill	Indeterminate
Baseline	0.129 ± 0.01	0.078 ± 0.012	0.051 ± 0.031	-
Our Method(Without indeterminate)	0.105 ± 0.004	0.093 ± 0.001	0.011 ± 0.003	-
Our Method(indeterminate)	0.081 ± 0.002	0.08 ± 0.002	0.002 ± 0.002	0.073 ± 0.027

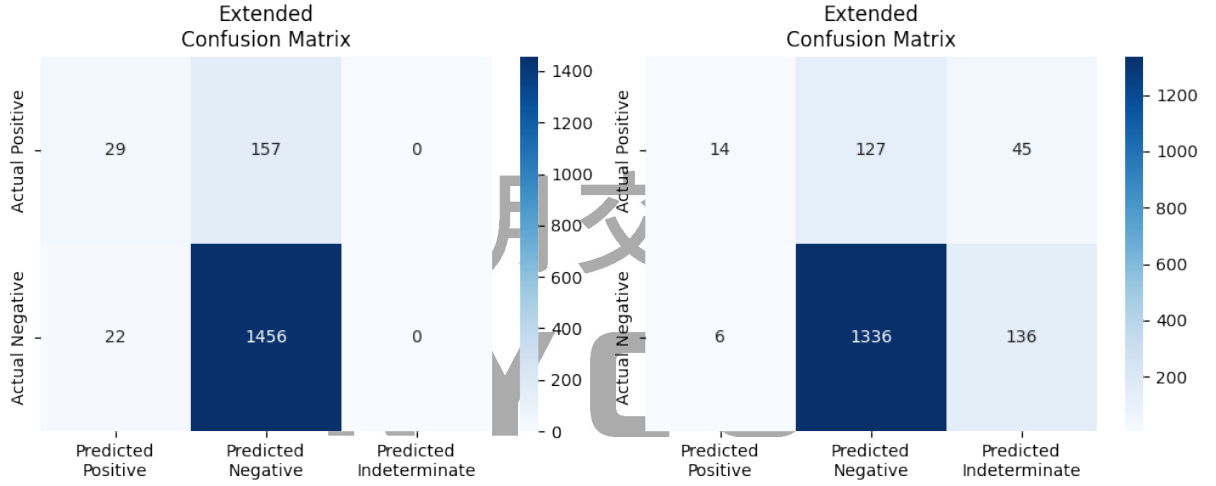
Table 4.4: AVF Dataset Indeterminacy Metrics (II) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)

	Imperfection	Indeterminate recall	Harmonic score
Baseline	0.129 ± 0.01	0.261 ± 0.039	0.958 ± 0.002
Our Method(Without indeterminate)	0.105 ± 0.004	0.146 ± 0.014	0.967 ± 0.001
Our Method(indeterminate)	0.154 ± 0.021	0.292 ± 0.03	0.951 ± 0.008

The extended confusion matrix, as shown in Figure 4.8, provides deeper insights into how the model handles ambiguous cases. Unlike the KDOQI guidelines, which classify all cases as either positive or negative, the indeterminate-aware model assigns indeterminate cases to a separate "Indeterminate" category. This approach prevents overly confident misclassifications, particularly for borderline cases where the probability distributions lack clear separation. By reallocating indeterminate predictions, the model reduces the likelihood of errors, ensuring a more cautious and clinically meaningful decision-making process.



(a) baseline

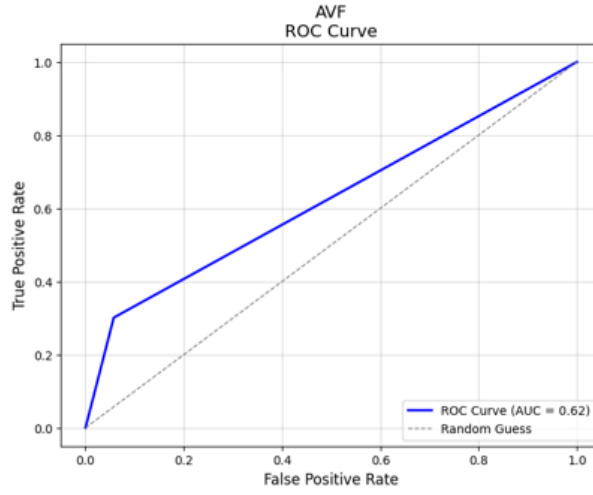


(b) Our Method (Without indeterminate)

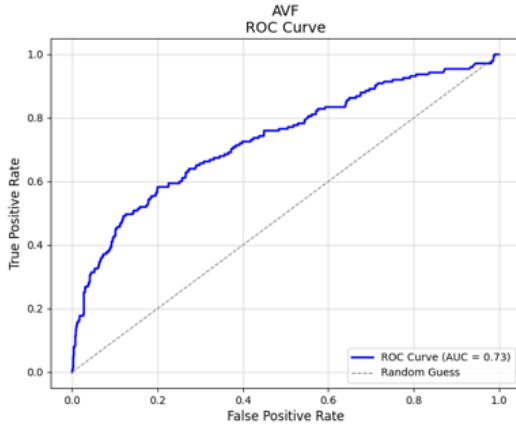
(c) Our Method (Indeterminate)

Figure 4.8: AVF Dataset Extended Confusion Matrix

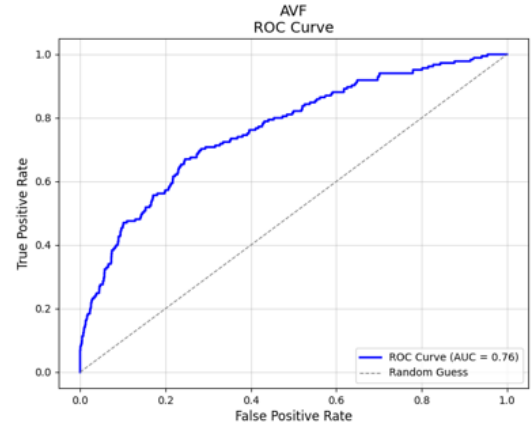
The ROC curve comparisons, as presented in Figure 4.9, further validate the proposed model's superior discriminative ability. The baseline KDOQI guidelines achieved an AUC of 0.62, reflecting its limited predictive capability. In contrast, the proposed model without indeterminate classification improved to an AUC of 0.73, and the inclusion of indeterminate classifications further enhanced the AUC to 0.76. This progression demonstrates that the model's ability to distinguish between positive and negative cases improves significantly when indeterminacy is explicitly accounted for.



(a) baseline



(b) Our Method (Without indeterminate)



(c) Our Method (Indeterminate)

Figure 4.9: AVF Dataset ROC Curve

The results from the AVF dataset illustrate the effectiveness of incorporating indeterminate classifications into the prediction framework. The proposed model not only surpasses the KDOQI guidelines in terms of accuracy and precision but also introduces a critical layer of indeterminacy management, which is essential for clinical applications. By balancing standard predictive metrics with indeterminacy-specific measures, the model ensures that ambiguous cases are flagged for further review, reducing the risk of misdiagnoses and unnecessary interventions. These findings underscore the clinical potential of the indeterminate-aware approach for managing vascular access in hemodialysis patients.

4.5.2 Arteriovenous Graft Dataset

The analysis of the Arteriovenous Graft (AVG) dataset provides complementary insights into the proposed model’s performance, particularly in a dataset with fewer samples and a higher proportion of surgical interventions compared to the AVF dataset. This section discusses the dataset characteristics, model evaluation, and the implications of the results.

As shown in Table 4.5, the parameters for the AVG dataset were optimized using Bayesian optimization, which maximized accuracy.

Table 4.5: The parameter settings in AVG Dataset

Parameter Settings	
Number of estimation times: n	10
Threshold used to determine output of indeterminacy: θ_μ	0.95
Noise mean: μ	0
Noise variance: σ^2	80
Weight of estimator: $W = [w_1, w_2, w_3]$	[1,1,1]
Weight of harmonic score: $W' = [w'_1, w'_2, w'_3]$	[1,1,1]

The AVG dataset comprises a total of 869 records, significantly fewer than the 4,991 records in the AVF dataset, as shown in Figure 4.10. Among these records, 218 (25.1%) involve PTA interventions, while 651 (74.9%) represent non-surgical cases. The higher proportion of surgical cases in AVG compared to AVF (10.9%) aligns with clinical observations that AVG is often associated with shorter durability and higher complication rates, necessitating more frequent interventions.

Percutaneous Transluminal Angioplasty (PTA)

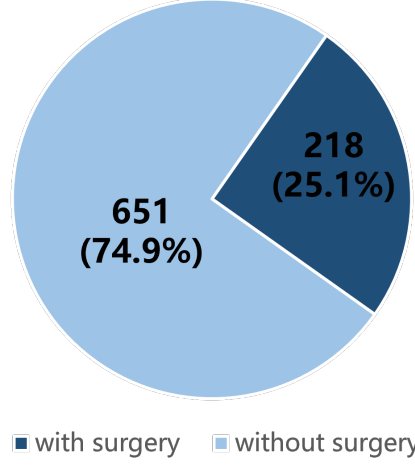


Figure 4.10: AVG data distribution

Tables 4.6, 4.7, and 4.8 show the results of the comparison of metrics between KDOQI guidelines and our method on the AVG dataset.

Table 4.6: AVG Dataset Standard Metrics of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)

	Accuracy	PPV	NPV
Baseline	0.791 ± 0.018	0.685 ± 0.044	0.809 ± 0.026
Our Method(Without indeterminate)	0.765 ± 0.023	0.695 ± 0.045	0.788 ± 0.028
Our Method(indeterminate)	0.818 ± 0.019	0.74 ± 0.137	0.828 ± 0.031

Table 4.7: AVG Dataset Indeterminacy Metrics (I) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)

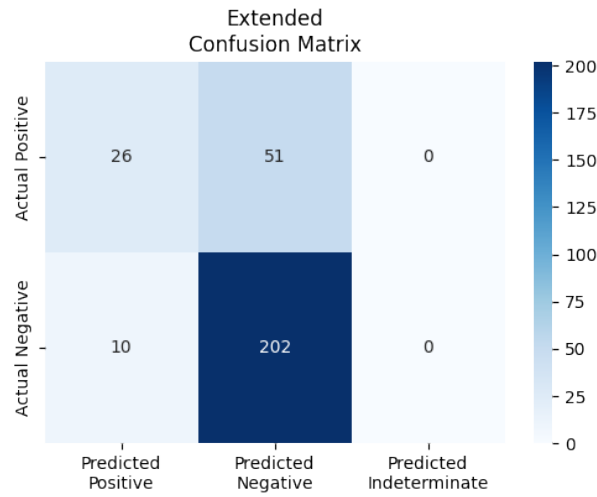
	Error	Leakage	Overkill	Indeterminate
Baseline	0.208 ± 0.032	0.186 ± 0.011	0.048 ± 0.016	-
Our Method(Without indeterminate)	0.235 ± 0.023	0.19 ± 0.027	0.032 ± 0.017	-
Our Method(indeterminate)	0.147 ± 0.017	0.128 ± 0.025	0.02 ± 0.011	0.192 ± 0.007

Table 4.8: AVG Dataset Indeterminacy Metrics (II) of Baseline, Our Method(Without indeterminate), Our Method(indeterminate)

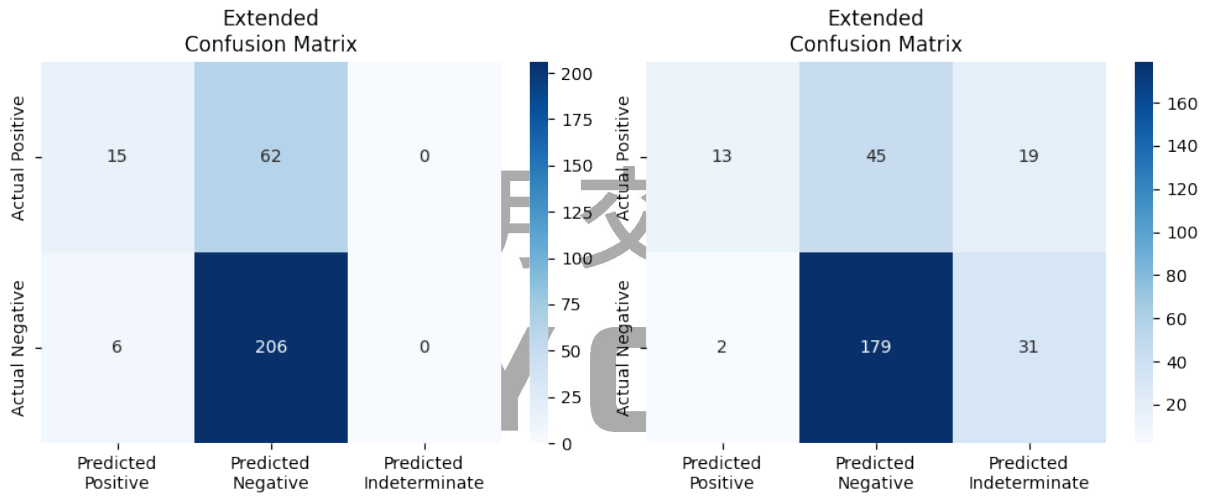
	Imperfection	Indeterminate recall	Harmonic score
Baseline	0.208 ± 0.032	0.262 ± 0.07	0.908 ± 0.018
Our Method(Without indeterminate)	0.235 ± 0.023	0.247 ± 0.059	0.928 ± 0.013
Our Method(indeterminate)	0.34 ± 0.011	0.496 ± 0.059	0.895 ± 0.013

The extended confusion matrix in Figure 4.11 shows how the proposed model reallocates a significant proportion of indeterminate predictions to the "Indeterminate" category. Compared to AVF, the higher surgical proportion in AVG is evident in the relative distribution of predicted positive and negative cases. The indeterminate-aware approach ensures that borderline cases are cautiously flagged, avoiding overconfidence in predictions.

陽明交大
NYCU



(a) baseline

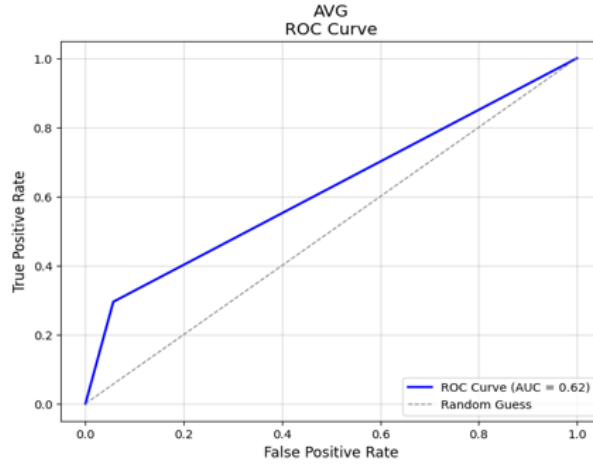


(b) Our Method (Without indeterminate)

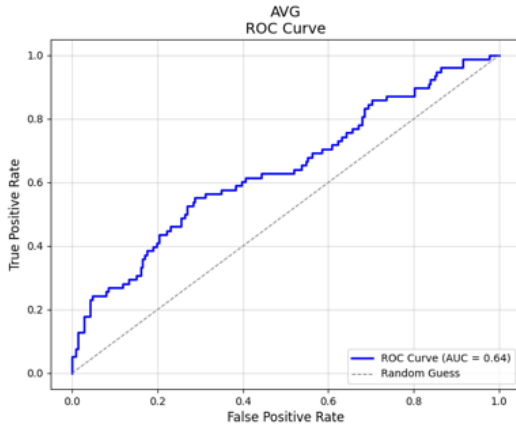
(c) Our Method (Indeterminate)

Figure 4.11: AVG Dataset Extended Confusion Matrix

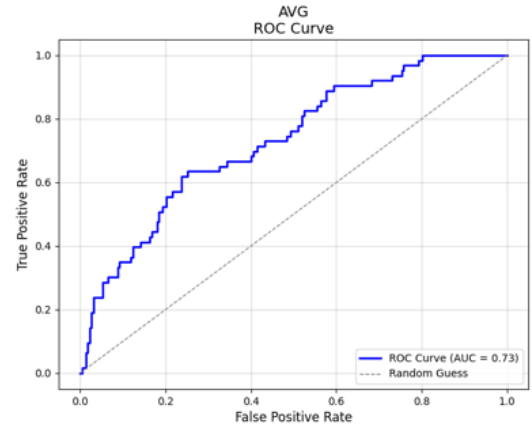
The ROC curve analysis, illustrated in Figure 4.12, shows the progression of AUC values across configurations. The baseline KDOQI guidelines achieved an AUC of 0.62, reflecting its limited predictive capability. The proposed model without indeterminate classification achieved an AUC of 0.64, while the inclusion of indeterminate classifications further improved the AUC to 0.73. These results demonstrate that explicitly accounting for indeterminacy enhances the model's ability to distinguish between positive and negative cases.



(a) baseline



(b) Our Method (Without indeterminate)



(c) Our Method (Indeterminate)

Figure 4.12: AVG Dataset ROC Curve

In comparison to the AVF dataset, the AVG dataset presents unique challenges due to its smaller size and higher surgical intervention rate. Despite these differences, the proposed indeterminate-aware model consistently outperformed both the KDOQI guidelines and the model without indeterminate classifications. The results emphasize the model's adaptability and robustness, making it a valuable tool for supporting clinical decision-making in vascular access management.

Chapter 5

Conclusion

Taiwan, with the world's highest proportion of dialysis patients, faces significant challenges in managing vascular access dysfunction, a common complication among these patients. Early detection and accurate diagnosis are critical for reducing the risks associated with this condition and improving patient outcomes. This study proposes an indeterminate-aware tree-based machine learning framework to enhance diagnostic precision and address the inherent indeterminacies in clinical data.

Our methodology integrates Multipass Indeterminacy Estimation and an Indeterminate-Aware Data Classification approach to provide robust predictions while accounting for ambiguous cases. By introducing an extended confusion matrix and novel metrics such as leakage, overkill, indeterminacy, and imperfection, the study captures a comprehensive evaluation of model performance. These metrics not only highlight the strengths of the proposed model in reducing misclassification but also ensure cautious decision-making, particularly for borderline cases.

Our model demonstrates superior predictive performance across multiple metrics. The inclusion of indeterminate classifications further enhances diagnostic reliability, reflected in improved AUC, PPV, and reduced error rates. For example, in both AVF and AVG datasets, our approach outperformed baseline methodologies [3], achieving significant improvements in accuracy, harmonic scores, and indeterminacy handling.

This research emphasizes the importance of accounting for indeterminacy in clinical decision-making. By reallocating ambiguous predictions to an "Indeterminate" category, the model avoids overly confident misclassifications, providing a balanced and cautious framework for diagnosis. The proposed approach not only improves prediction results but also aligns with the clinical need for reliable, interpretable, and actionable decision support systems.

In conclusion, the study demonstrates that integrating machine learning with indeterminacy

analysis offers a powerful tool for managing complex clinical datasets. The proposed methodology has the potential to enhance patient care by enabling more accurate and informed decisions, reducing unnecessary interventions, and ensuring timely treatment for dialysis patients. This framework can serve as a foundation for future advancements in medical AI, particularly in areas requiring precise and cautious diagnostic strategies.

陽明交大
NYCU

References

- [1] R. Navuluri and S. Regalado, “The kdoqi 2006 vascular access update and fistula first program synopsis,” *Semin Intervent Radiol*, vol. 26, no. 2, pp. 122–124, 2009.
- [2] J. A. Balch, B. Shickel, A. Bihorac, G. R. Upchurch, and T. J. Loftus, “Integration of AI in surgical decision support: improving clinical judgment,” *Global Surgical Education - Journal of the Association for Surgical Education*, vol. 3, no. 1, p. 56, 2024. [Online]. Available: <https://doi.org/10.1007/s44186-024-00257-2>
- [3] C. Wu and C. Lin, “Integrating vascular access surveillance with clinical monitoring for stenosis prediction,” *J Nephrol*, vol. 37, pp. 461–470, 2024. [Online]. Available: <https://doi.org/10.1007/s40620-023-01799-2>
- [4] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, 1st ed. Chapman and Hall/CRC, 1984. [Online]. Available: <https://doi.org/10.1201/9781315139470>
- [5] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>
- [6] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” *arXiv preprint arXiv:1603.02754*, 2016.
- [7] A. Choudhury, A. Mondal, and S. Sarkar, “Searches for the bsm scenarios at the lhc using decision tree based machine learning algorithms: A comparative study and review of random forest, adaboost, xgboost and lightgbm frameworks,” *arXiv preprint arXiv:2405.06040*, 2024.
- [8] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A comparative analysis of xgboost,” *arXiv preprint arXiv:1911.01914*, 2019.

- [9] A. Yan, “A Survey on Tree-Based Machine Learning Methods: The Path to XGBoost and Generalized Random Forests,” *University of California, Riverside*, May 2020, available at ayan013@ucr.edu.
- [10] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “Lightgbm: A highly efficient gradient boosting decision tree,” 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
- [11] A. Veronika Dorogush, V. Ershov, and A. Gulin, “Catboost: gradient boosting with categorical features support,” *arXiv preprint arXiv:1810.11363*, 2018.
- [12] L. Grinsztajn, E. Oyallon, and G. Varoquaux, “Why do tree-based models still outperform deep learning on tabular data?” *arXiv preprint arXiv:2207.08815*, 2022.
- [13] D. McElfresh, S. Khandagale, J. Valverde, V. Prasad C, G. Ramakrishnan, M. Goldblum, and C. White, “When do neural nets outperform boosted trees on tabular data?” pp. 76 336–76 369, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/f06d5ebd4ff40b40dd97e30cee632123-Paper-Datasets_and_Benchmarks.pdf
- [14] H.-J. Ye, S.-Y. Liu, H.-R. Cai, Q.-L. Zhou, and D.-C. Zhan, “A closer look at deep learning methods on tabular datasets,” *arXiv preprint arXiv:2407.00956*, 2024.
- [15] I. D. Mienye and Y. Sun, “A survey of ensemble learning: Concepts, algorithms, applications, and prospects,” *IEEE Access*, vol. 10, pp. 99 129–99 149, 2022.
- [16] J. Kim and S. Choi, “Automated machine learning for soft voting in an ensemble of tree-based classifiers,” 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:219631676>
- [17] Z. Yang, A. Sudjianto, X. Li, and A. Zhang, “Inherently interpretable tree ensemble learning,” *arXiv preprint arXiv:2410.19098*, 2024.

- [18] V. Nemani, L. Biggio, X. Huan, Z. Hu, O. Fink, A. Tran, Y. Wang, X. Zhang, and C. Hu, “Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial,” *arXiv preprint arXiv:2305.04933*, 2023.
- [19] A. F Psaros, X. Meng, Z. Zou, L. Guo, and G. Em Karniadakis, “Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons,” *arXiv preprint arXiv:2201.07766*, 2022.
- [20] M. Barandas, D. Folgado, R. Santos, R. Simão, and H. Gamboa, “Uncertainty-based rejection in machine learning: Implications for model development and interpretability,” *Electronics*, vol. 11, no. 3, 2022. [Online]. Available: <https://www.mdpi.com/2079-9292/11/3/396>
- [21] M. Tulio Ribeiro, S. Singh, and C. Guestrin, ““why should i trust you?”: Explaining the predictions of any classifier,” *arXiv preprint arXiv:1602.04938*, 2016.
- [22] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1050–1059.