IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 July 2025, accepted 12 September 2025, date of publication 22 September 2025, date of current version 7 October 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3612957

APPLIED RESEARCH

A Trustworthy Model With Uncertainty Management for Predicting Vascular Access Dysfunction in Hemodialysis Patients

CHIA-HSUN LIN[1,2], ZHENG LIN CHEN[3], CHUNG-KUAN WU[1,4,5], TSEN-CHE WU[3], AND CHING-WEN MA[3], (Member, IEEE)

[1]School of Medicine, Fu Jen Catholic University, Taipei, Taiwan
[2]Division of Cardiovascular Surgery, Department of Surgery, Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan
[3]College of Artificial Intelligence, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
[4]Division of Nephrology, Department of Internal Medicine, Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan
[5]Division of Digital Informatics Management, Department of Digital Medicine, Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan

Corresponding author: Ching-Wen Ma (machingwen@nycu.edu.tw)

This work was supported by the Shin Kong Wu Ho-Su Memorial Hospital and the National Yang Ming Chiao Tung University under Grant 112-SKH-NYCU-03 and Grant 2024SKHADR014.

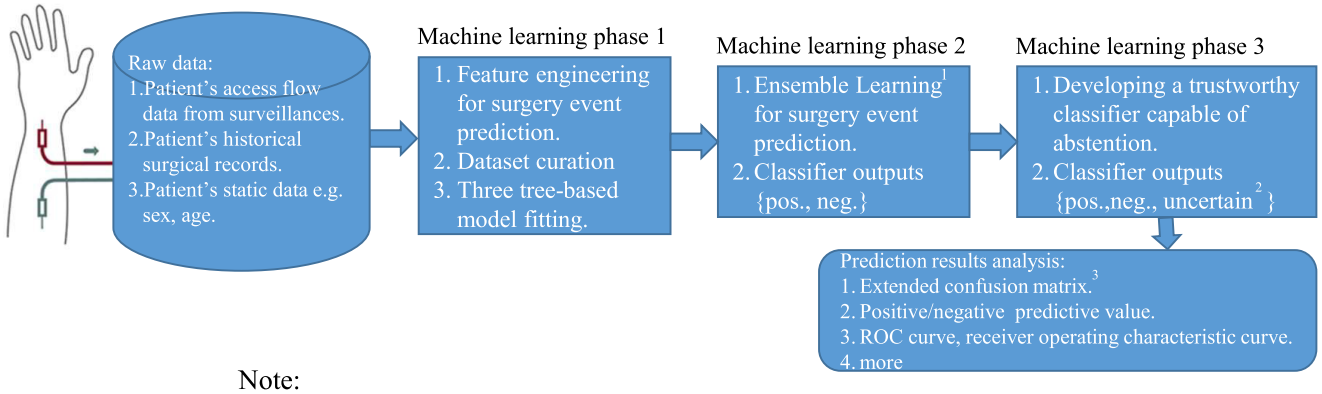
ABSTRACT Vascular access dysfunction is a prevalent and critical complication among hemodialysis patients, particularly in Taiwan, which has the highest proportion of dialysis patients globally. Early diagnosis and effective management are essential for improving patient outcomes. However, traditional diagnostic approaches, such as routine surveillance and fixed blood flow thresholds, often fail to reliably identify patients requiring timely surgical intervention. To address these challenges, this study proposes a trustworthy AI system utilizing an uncertainty-aware, tree-based machine learning framework to improve the assessment of vascular access dysfunction. The proposed framework incorporates advanced uncertainty management techniques to address both aleatoric uncertainty (data uncertainty) originating from inherent data variability, and epistemic uncertainty (model uncertainty) arising from model limitations. Aleatoric uncertainty is systematically quantified using a multipass perturbation strategy that simulates sample variability to capture the distribution of potential outcomes, while epistemic uncertainty is mitigated using ensemble methods. By leveraging the multipass perturbation strategy, the framework generates calibrated uncertainty estimations. To increase the reliability and trustworthiness of clinical decision-making, the framework assigns low-confidence (high-uncertainty) predictions to an 'Uncertain' category, thereby achieving a near-zero leakage rate, which is crucial in high-stakes clinical scenarios. Additionally, an extended confusion matrix and novel uncertainty metrics are introduced to comprehensively evaluate model performance. The system has been validated using real-world hospital datasets, demonstrating its deployable AI potential with superior predictive accuracy, sensitivity, and robustness compared to traditional methods such as the Kidney Disease Outcomes Quality Initiative (KDOQI) guidelines.

INDEX TERMS Deployable AI, uncertainty quantification, machine learning, toward zero leakage rate AI, tree-based models, trustworthy AI system, vascular access dysfunction.

I. INTRODUCTION

Hemodialysis is a life-saving treatment for patients with end-stage renal disease (ESRD), but it comes with significant challenges. Among these, vascular access dysfunction remains one of the most critical complications, directly affecting treatment efficacy and patient survival. This issue is especially pronounced in Taiwan, which has the highest proportion of dialysis patients globally. Early detection and timely intervention are therefore vital for effectively managing vascular access dysfunction and improving patient outcomes.

The associate editor coordinating the review of this manuscript and approving it for publication was Akansha Singh.



Note:

1. Ensemble learning involves soft voting among the three tree-based models.
2. 'Uncertain' means the classifier abstains when confidence is low in distinguishing between positive and negative cases.
3. Extended confusion matrix involves the number of cases of 'uncertain.'

FIGURE 1. Overview of the machine learning workflow for vascular access dysfunction prediction. The process involves three phases: (1) feature engineering from raw surveillance data, historical surgical records, and static patient attributes; (2) ensemble learning using tree-based classifiers (Decision Tree, Random Forest, XGBoost) with soft voting; and (3) development of a trustworthy classifier capable of abstaining (outputting "uncertain") under high uncertainty. Prediction analysis includes extended confusion matrix, positive and negative predictive values, and ROC curves.

Traditionally, vascular access dysfunction is diagnosed through routine surveillance methods, primarily relying on fixed access flow thresholds as guided by the KDOQI guidelines [1], [2], [3]. However, these methods often fail to consider patient-specific variations, potentially leading to delayed interventions, missed diagnoses, and overconfident misclassifications. Consequently, there is an urgent need for more sophisticated, personalized diagnostic tools.

Recent advancements in artificial intelligence (AI) offer promising solutions to address these limitations [4]. Although deep neural networks have gained attention in various applications, they are less effective when dealing with structured clinical data due to their limited interpretability. In contrast, tree-based models—such as Decision Trees, Random Forests, and XGBoost—excel in analyzing structured data [31], [32], [33]. These models are highly interpretable and can efficiently handle heterogeneous clinical features, making them particularly suitable for medical applications where transparency and trustworthiness are essential.

Nevertheless, even robust machine learning models can produce uncertain predictions, especially in complex clinical environments. To improve the reliability of predictions in high-stakes medical scenarios, it is essential to explicitly quantify and manage uncertainty. Uncertainty in predictions can be categorized into two main types: aleatoric uncertainty, which arises from inherent variability in the clinical data itself, and epistemic uncertainty, which stems from limitations within the predictive model.

Aleatoric uncertainty, reflecting intrinsic data variability such as measurement noise or natural fluctuations in clinical parameters, is irreducible and must be explicitly quantified. In this study, we address aleatoric uncertainty using a multipass perturbation strategy [25] that simulates data variability to better understand the distribution of potential outcomes.

Epistemic uncertainty, on the other hand, arises due to limitations in model structure or training data, such as insufficient data coverage or model bias. This type of uncertainty can be mitigated through ensemble methods that combine predictions from multiple models, thereby increasing overall robustness and reducing reliance on individual model biases.

By integrating these two uncertainty management strategies, the proposed framework generates more calibrated and reliable predictions. Importantly, the framework introduces an "Uncertain" prediction category, aligning our approach with the established practice of selective classification or reject-option classification [26], [27], [28], [29], [30]. This cautious approach improves clinical decision-making and aligns with the critical nature of medical diagnostics.

To provide clarity and context, we illustrate the entire machine learning workflow in Figure 1. The figure summarizes each stage, from feature engineering to uncertainty-aware prediction and evaluation, guiding readers through our methodological framework.

The main contributions of this study include:

- **Dual Uncertainty Management:** A comprehensive approach that quantifies aleatoric uncertainty through multipass perturbation and mitigates epistemic uncertainty using ensemble learning.
- **Trustworthy Clinical Decision-Making:** A trustworthy classifier designed to abstain from or reject low-confidence (high-uncertainty) predictions, ensuring safer and more reliable clinical decision-making.
- **Novel Extended Confusion Matrix:** An extended confusion matrix specifically developed to categorize uncertain predictions into detailed sub-categories (e.g.,

Uncertain Positive, Uncertain Negative), providing more granular insights than traditional rejection evaluation confusion matrices and simultaneously evaluating prediction accuracy and rejection performance.

- **Tailored Feature Engineering:** Careful selection and engineering of clinically relevant features specifically for vascular access dysfunction prediction, including (1) the trend of access flow (slope of measurement history), (2) the difference between current and previous measurements, and (3) detailed surgery history. The selection of these features was guided by careful observation of patients' surveillance data and surgical histories.
- **Real-world Validation:** Empirical validation demonstrating the framework's effectiveness using real-world hospital datasets, with substantial improvements in predictive accuracy and reliability compared to traditional diagnostic methods.

By addressing the critical challenges of uncertainty management and diagnostic precision, this research aims to support more reliable, interpretable, and actionable clinical decisions in managing vascular access dysfunction for hemodialysis patients.

II. RELATED WORKS

A. TRADITIONAL APPROACHES FOR MANAGING VASCULAR ACCESS DYSFUNCTION

1) KDOQI GUIDELINES AND PHYSICAL EXAMINATION

The KDOQI guidelines provide a standardized framework for managing vascular access in hemodialysis patients. These guidelines emphasize the importance of routine monitoring and surveillance, recommending fixed access flow thresholds; KDOQI guideline as key indicators for early detection of vascular access dysfunction. The guidelines have been widely adopted to standardize care and improve outcomes. However, they rely heavily on fixed thresholds and routine clinical evaluations, which may not account for patient-specific variability or dynamic changes in vascular access conditions [4].

Physical examination complements the KDOQI guidelines by enabling clinicians to detect early signs of stenosis or dysfunction through routine assessments, such as auscultation and palpation. Studies have shown that integrating vascular access surveillance with physical examinations improves the prediction of stenosis by providing a more comprehensive assessment of the patient's condition [5]. Despite these advantages, physical examination is subject to inter-clinician variability and often lacks the precision required for high-stakes clinical decision-making.

2) LIMITATIONS OF TRADITIONAL METHODS

While the combination of KDOQI guidelines and physical examination forms the cornerstone of current clinical practices, these methods have notable limitations. Fixed thresholds of access flow fail to capture the dynamic and individualized nature of vascular access dysfunction. Furthermore, physical examinations, though valuable, are prone to subjective interpretation, leading to potential inconsistencies in diagnosis. These limitations underscore the need for advanced approaches that integrate real-time data and personalized analytics.

3) TOWARD AI-ENHANCED SOLUTIONS

Recent advancements in AI have shown promise in addressing the shortcomings of traditional methods [4]. By leveraging data-driven models, AI systems can provide real-time insights, account for patient-specific variability, and increase diagnostic accuracy. The integration of uncertainty-aware AI frameworks, such as the one proposed in this study, offers a pathway to complement and improve existing practices while mitigating their inherent limitations.

B. MODERN MACHINE-LEARNING TECHNIQUES FOR VASCULAR ACCESS DYSFUNCTION

1) TREE-BASED MODELS AND ENSEMBLE LEARNING

Tree-based models such as Decision Trees [6], Random Forests [7], and Gradient Boosting [8], [9], [10], [11] have been widely applied in structured data analysis due to their interpretability and effectiveness in handling heterogeneous data. These models excel in clinical applications by offering transparent and reliable predictions, making them a valuable choice for tasks involving tabular medical data.

Ensemble learning techniques further enhance the performance of tree-based models by combining multiple learners to improve prediction accuracy and robustness [12]. Random Forests, which aggregate predictions from numerous decision trees, and Gradient Boosting methods, which iteratively optimize model performance, are particularly effective at mitigating overfitting and increasing generalizability. Ensemble methods also inherently address model uncertainty by leveraging diverse learners to produce more stable predictions.

2) LIMITATIONS IN HIGH-STAKES CLINICAL SCENARIOS

Despite these strengths, tree-based and ensemble methods alone are limited in their ability to quantify and manage predictive uncertainty, particularly in high-stakes clinical scenarios like vascular access dysfunction. These limitations necessitate frameworks that integrate uncertainty quantification to provide calibrated predictions and enable reliable decision-making [13], [14], [15], [16], [17].

C. UNCERTAINTY QUANTIFICATION AND TRUSTWORTHY AI

1) UNCERTAINTY QUANTIFICATION

Uncertainty quantification addresses ambiguous predictions by estimating confidence levels through calibrated uncertainty estimation [18], [19], [20], [21], [22], [23]. In this study, predictions with high uncertainty (i.e., low confidence) are categorized as ''uncertain,'' also referred to as indeter-

minate, abstention, or reject cases. By setting appropriate thresholds, this approach supports cautious decision-making and minimizes the risk of overconfident errors in critical applications.

2) ABSTENTION STRATEGIES AND TRUSTWORTHY AI

Techniques such as multipass perturbation, which simulates sample variability to capture a range of potential outcomes, and uncertainty-based rejection strategies [24], are critical for improving prediction reliability. Rejection strategies in particular enable the system to abstain from predictions when uncertainty exceeds a predefined threshold, ensuring robust performance even in complex or ambiguous scenarios.

3) INTERPRETABILITY METHODS

Interpretability methods such as local perturbation analysis [25] provide valuable insights into model behavior and feature contributions. In healthcare applications, interpretability is vital for fostering trust among clinicians and ensuring accountability in AI-assisted decision-making. For instance, visual explanations of rejected samples or uncertain predictions can highlight areas requiring further data collection or model refinement.

4) SUPERIORITY OF TREE-BASED MODELS IN STRUCTURED DATA APPLICATIONS

Research highlights the superior performance of tree-based models in structured or tabular data scenarios. Random Forests mitigate overfitting through variance reduction via bagging, while XGBoost leverages boosting techniques for enhanced accuracy and scalability [11]. Advanced models such as LightGBM and CatBoost further improve efficiency through selective sampling and categorical feature handling [9], [10]. Studies by Grinsztajn [31] and McElfresh and Khandagale [32] demonstrate that Gradient Boosted Decision Trees (GBDTs) outperform neural networks on medium-sized tabular datasets, where the latter often fail to generalize. Similarly, Ye et al. [33] report that while deep learning has made strides, tree-based models remain competitive, particularly in scenarios with limited data or computational resources. These strengths make tree-based models a robust, practical choice for structured data, where their adaptability and efficiency surpass neural networks.

5) SUMMARY

By integrating uncertainty quantification, abstention strategies, and interpretability into tree-based models, this study's framework ensures robust and trustworthy predictions tailored to clinical applications. These advancements address critical challenges in managing vascular access dysfunction, paving the way for safer and more reliable AI systems in healthcare.

III. METHODS

The study finally enrolled 412 ESKD patients receiving hemodialysis via functional arteriovenous access at a hemodialysis center of a medical center in 2018. Surveillance of vascular access was performed trimonthly to detect access flow volume from the time of enrollment until the patient transfer to other clinics, death or the end of study on December 31, 2021.

This study was conducted in accordance with the Declaration of Helsinki, and was approved by the Institutional Review Board of Shin Kong Wu Ho-Su Memorial Hospital (IRB No. 20221004R; approval date: January 20, 2023).

This section outlines the methods developed to increase the predictive accuracy, robustness, and trustworthiness of the proposed classification framework. The methodology is specifically designed to address both aleatoric and epistemic uncertainties, ensuring cautious and reliable predictions for clinical decision-making in vascular access dysfunction. The training of this framework consists of three main phases, as show in Figure 1:

1) **Feature Engineering and Tree-Based Model Design:** Based on observations of clinical surveillance data and surgical histories, we designed task-relevant features to improve prediction of surgical events. In particular, the trend of access flow (i.e., the slope of the measurement history) proved to be an informative indicator. Using these features, we developed three tree-based models: Decision Tree, Random Forest, and XGBoost.
2) **Ensemble Learning:** To increase predictive reliability and reduce epistemic uncertainty, we adopted an ensemble strategy that integrates outputs from the three tree-based models. This approach capitalizes on the strengths and diversity of individual learners to improve generalization and stability. Ensemble models are also inherently better calibrated in terms of expected calibration error (ECE), and in this framework, they provide a more robust estimation of uncertainty by mitigating biases from any single model.
3) **Uncertainty-Aware Classification with Rejection:** Aleatoric uncertainty is quantified by simulating data variability using Gaussian noise perturbation and computing prediction entropy. High-entropy outputs are flagged as uncertain. This allows the model to abstain (reject) from making predictions when confidence is insufficient. The final classification outcomes are defined as:
   - **'True'**: High-confidence prediction indicating vascular access dysfunction requiring intervention;
   - **'False'**: High-confidence prediction indicating absence of dysfunction;
   - **'Uncertain'**: Prediction rejected due to high uncertainty, ensuring clinical safety when the model lacks sufficient confidence.

These phases are elaborated in the following subsections.

A. FEATURE ENGINEERING, TREE-BASED MODEL DESIGN, AND ENSEMBLING

The proposed framework incorporates essential data preprocessing steps tailored for vascular access dysfunction

TABLE 1. Aligned raw data collected from patient surveillance. *Qa* means access flow.

Patient ID	Current Qa Value	Days Since Previous Exam	Surgery Performed
01	540	0	Yes
01	1430	91	No
01	1380	91	No

Note: A value denotes the measured vascular access flow (Q_a).

prediction and employs ensemble methods based on three tree-based models—Decision Tree, Random Forest, and XGBoost—combined using a weighted soft voting strategy to aggregate their predictions.

1) FEATURE ENGINEERING

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

a: DATA ALIGNMENT

In the original dataset, there is no Qa (access flow) 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. An aligned raw data example is showed in Table 1.

b: FEATURE CREATION

The KDOQI guidelines define three critical features related to vascular access dysfunction: (1) arteriovenous fistula (AVF) with a Qa value less than 400 mL/min or 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 improve 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. The sample data for machine learning algorithms derived from Table 1 are showed in Table 2.

2) TREE-BASED MODEL DESIGN AND ENSEMBLING

After the design of features and curation of sample data, we choose three tree-based models—Decision Tree, Random Forest, and XGBoost—for surgery prediction (i.e., y in Table 2). By leveraging model diversity, the ensemble mitigates the epistemic uncertainty that arises from limitations in individual models. To ensure reliable and generalized performance, the models were trained using K-fold cross-validation.

It is important to note that Decision Tree and Random Forest models are inherently less sensitive to class imbalance due to their splitting criteria, which can adapt to the distribution of classes in the training data. However, XGBoost requires additional balancing adjustments to achieve similar robustness.

The `scale_pos_weight` parameter in XGBoost was fine-tuned to ensure balanced learning. This parameter adjusts the weight of positive samples during training and is defined as:

$$\texttt{scale\_pos\_weight} = \frac{n_{\text{neg}}}{n_{\text{pos}}}, \tag{1}$$

where n_{neg} and n_{pos} represent the number of negative and positive samples, respectively, derived from the training data. This adjustment compensates for imbalanced datasets, ensuring that predictions for minority classes are appropriately weighted.

3) MODELING AND VALIDATION

We trained a soft-voting ensemble comprising a Decision Tree (max_depth = 5), a Random Forest (n_estimators = 50, max_depth = 4), and an XGBoost classifier (eta = 0.001, n_estimators = 50). To mitigate class imbalance, XGBoost's `scale_pos_weight`was set to the ratio of negative to positive samples in each training fold. We performed 3-fold cross-validation with shuffling (random_state = 42) and built separate models for AVF and AVG cohorts. Feature preprocessing included binarizing the PTA label, one-hot encoding of categorical variables (sex, access site, access type), removal of identifiers and surgical detail fields, column renaming aligned with clinical semantics (e.g., KDOQI-derived indicators), and imputing missing values for the interval since prior surgery.

4) SUMMARY

This integrated approach provides a robust and interpretable foundation for the inclusion of uncertainty quantification within the framework.

B. UNCERTAINTY-AWARE CLASSIFICATION WITH REJECTION

1) MULTIPASS UNCERTAINTY ESTIMATION AND REJECTION

To quantify prediction uncertainty, a multipass estimation method was developed. This approach captures aleatoric uncertainty by simulating variability in the input data through Gaussian noise perturbation and analyzing the model's response.

For each test instance, we performed $n = 10$ stochastic passes by injecting Gaussian noise into the A-value feature ($\sigma \in \{10, 20, \ldots, 100\}$), aggregating mean probabilities and variances from the ensemble. An entropy-based threshold $\theta_u = \mu \cdot \log(2)$ ($\mu \in \{0.90, 0.95, 1.00\}$) flagged

TABLE 2. Sample data curation with derived features from Table 1.

Sample ID	Current Qa Value x_0	Days Since Previous Exam x_1	Previous Qa Value x_1	Previous Surgery x_2	Qa Value Slope x_3	Qa Value Difference x_4	Other Features x_n	Label y
0	540	0	Padding	Padding	Padding	Padding	...	Yes
1	1430	91	540	Yes	$\frac{1430-540}{91} = 78$	890	...	No
2	1380	91	1430	No	$\frac{1380-1430}{91} = -0.55$	-50	...	No

Note: A value denotes the measured vascular access flow (Q_a).

high-uncertainty cases as *Uncertain*. For low-entropy cases, class means and variances defined non-overlapping confidence ranges to assign *True* or *False*; otherwise, the sample was rejected and labeled as *Uncertain*.

Steps in Algorithm 1:

1) **Perturbation**: Gaussian noise ($\epsilon \sim N(\mu, \sigma^2)$) is added to each input sample to simulate variability in the data. The values of μ and σ^2 are chosen empirically or determined using automated techniques, such as Bayesian optimization, to ensure optimal performance.
2) **Prediction**: Perturbed samples are processed through the three tree-based models. For Random Forest, the soft voting mechanism is used. The three predictions of 'True' and 'False' probabilities are collected in P_y. They are further averaged by W and appended to $\hat{P}$ to produce probability predictions for 'True' and 'False' classes for each sample.
3) **Aggregation**: For each input, the mean (p_u) and variance (p_{σ^2}) of the predicted probability is calculated across all passes, providing a comprehensive measure of uncertainty.

By leveraging this multipass approach, the framework ensures that uncertainty estimates are well calibrated to reflect variability in the data. As a result, the method inherently aligns with the principles of uncertainty calibration, ensuring that predictions flagged as uncertain correspond to cases with genuinely higher variability or ambiguity.

Algorithm 1 Multipass Uncertainty Estimation

Input: $\hat{x}$ (Input sample), μ (Noise mean), σ^2 (Noise variance), $W = [w_1, w_2, w_3]$ (Weights of estimators), n (Number of passes), *Estimators* = [Decision Tree, Random Forest, XGBoost]

Output: p_μ (Means of probabilities of both "True" and "False"), p_{σ^2} (Variances of probabilities of both "True" and "False")

Initialize: $\hat{P} \leftarrow []$
for $i = 1$ to n **do**
  Sample $\epsilon_i \sim N(\mu, \sigma^2)$
  $\hat{x_i} \leftarrow \hat{x} + \epsilon_i$
  $P_{y_i} \leftarrow []$
  for each *clf* in *Estimators* **do**
    if *clf* == "Random Forest" **then**
      $v_{clf} \leftarrow$ VotingClassifier(*clf*, voting = 'soft')
    else
      $v_{clf} \leftarrow$ Classifier(*clf*)
    end if
    $p_{y_i,clf} \leftarrow v_{clf}$.PredictProb($\hat{x_i}$) ▷ $p = [p_T, p_F]$
    Append $p_{y_i,clf}$ to P_{y_i}
  end for
  $\hat{p_i} \leftarrow$ average(P_{y_i}, W)
  Append $\hat{p_i}$ to $\hat{P}$
end for
$p_\mu \leftarrow$ mean($\hat{P}$) ▷ $p_\mu = [p_{T,\mu}, p_{F,\mu}]$
$p_{\sigma^2} \leftarrow$ variance($\hat{P}$) ▷ $p_{\sigma^2} = [p_{T,\sigma^2}, p_{F,\sigma^2}]$
return p_μ, p_{σ^2}

2) UNCERTAINTY-AWARE CLASSIFICATION WITH REJECTION

Based on the results of multipass uncertainty estimation, samples are further categorized into three classes: 'True,' 'False,' or 'Uncertain,' by Algorthim 2, Uncertain-Aware Data Classification. This classification is performed using entropy and confidence thresholds, where entropy is used to assess overall uncertainty and the confidence range is determined using mean probabilities and variance.

Steps in Algorithm 2:

1) **Entropy Calculation**: Compute the entropy $H(y_p, \mu)$ of the probability distribution to measure overall uncertainty. Samples with entropy exceeding a predefined threshold (θ_u) are classified as 'Uncertain.'

$$H(y_p, \mu) = -(p_{T,\mu} \log p_{T,\mu} + p_{F,\mu} \log p_{F,\mu}) \quad (2)$$

2) **Confidence Evaluation**: For samples with low entropy, their mean probabilities (p_u) and variances (p_{σ^2}) are used to classify them as 'True' or 'False.' The variance term defines a range of confidence for each class, ensuring that predictions with high variability are safely rejected.
3) **Classification Rules:**
   - A sample is classified as 'True' if:

$$p_{T,\mu} > p_{F,\mu} \text{ and } p_{T,\mu} - \sqrt{p_{T,\sigma^2}} > p_{F,\mu} + \sqrt{p_{F,\sigma^2}} \quad (3)$$

     This ensures that the predicted probability for the 'True' class is not only greater than the 'False' class but also falls within a narrow, confident range.
   - A sample is classified as 'False' if:

$$p_{F,\mu} > p_{T,\mu} \text{ and } p_{F,\mu} - \sqrt{p_{F,\sigma^2}} > p_{T,\mu} + \sqrt{p_{T,\sigma^2}} \quad (4)$$

Algorithm 2 Uncertain-Aware Data Classification

Input: $p_\mu = [p_{T,\mu}, p_{F,\mu}]$ (Mean of probabilities)
$p_{\sigma^2} = [p_{T,\sigma^2}, p_{F,\sigma^2}]$ (Variance of probabilities)
θ_μ (Uncertainty threshold)
Output: "True", "False" or "Uncertain"

Compute entropy: $H(y_p, \mu) \leftarrow -(p_{T,\mu} \log p_{T,\mu} + p_{F,\mu} \log p_{F,\mu})$
if $H(y_p, \mu) > \theta_\mu$ **then**
  return "Uncertain"
else
  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**
    return "True"
  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**
    return "False"
  else
    return "Uncertain"
  end if
end if

This ensures similar confidence in predictions for the 'False' class.

- A sample is classified as 'Uncertain' if:
  - The entropy $H(y_p, \mu)$ exceeds the threshold θ_u; or
  - Neither of the above conditions for 'True' or 'False' are satisfied.

3) PURPOSE OF THE VARIANCE TERM

The variance term ($\sqrt{p_{\sigma^2}}$) is incorporated to define a range of confident outcomes for each prediction. Low variance indicates consistent predictions across perturbed samples, enabling confident classification. High variance, on the other hand, reflects uncertainty in the model's predictions, leading to cautious rejection of ambiguous cases.

This combination of entropy thresholds and confidence evaluation ensures that the framework avoids overconfident misclassifications while maintaining robust and reliable predictions tailored to high-stakes clinical scenarios.

4) ALGORITHM SUMMARIZATION

Algorithm 1, Multipass Uncertainty Estimation, quantifies prediction variability by generating probability distributions over perturbed input samples, capturing aleatoric uncertainty. Algorithm 2, Uncertainty-Aware Classification with Rejection, leverages entropy and confidence thresholds to categorize predictions into "True," "False," or "Uncertain," ensuring robust and cautious decision-making in ambiguous cases.

Together, these algorithms address the inherent limitations of traditional fixed-threshold diagnostic approaches by offering adaptive, uncertainty-informed predictions. This framework is particularly effective for managing borderline cases, increasing reliability and utility in high-stakes clinical applications.

IV. EXPERIMENTS AND RESULTS

A. DATASET DESCRIPTION

This study uses a hospital dataset comprising 5860 records from 412 patients undergoing clinical monitoring and surveillance. The data is divided into AVF and AVG subsets, representing native and synthetic vascular access types, respectively. Each record includes constant clinical features such as age, sex, and comorbidities, and dynamic procedural features such as access site and Qa values. The output data indicate whether a patient underwent Percutaneous Transluminal Angioplasty (PTA) intervention (surgery).

- **AVF Dataset**
  The AVF dataset consisted of 4991 cases, of which 10.9% required Percutaneous Transluminal Angioplasty (PTA) interventions.
- **AVG Dataset**
  The AVG dataset contained 869 cases and had a higher PTA intervention rate of 25.1%.

As a result, the dataset is imbalanced with respect to requiring interventions ('True') or not requiring interventions ('False'). While this imbalance affects the performance of XGBoost, it can be compensated through the ratio of positive to negative samples. In fact, we adjusted the parameter `scale_pos_weight` as equation (1) of the XDBoost algorithm to deal with the dataset imbalance issue. For Decision Tree and Random Forest, the imbalance issue is not critical because they are optimized according to leaf purity rather than loss minimization.

B. FEATURE ENGINEERING

The data from these 412 patients that were used to create features included constant data, serial data, and derived data.

The constant data includes patient-specific clinical characteristics, which remain consistent over time. The selected features for this study are demographic data (including age and sex) and comorbidity data (including diabetes mellitus [DM], hypertension [HTN], dyslipidemia [Dyslipid], coronary artery disease [CAD], acute myocardial infarction [AMI], cerebral vascular accident [CVA], peripheral artery occlusive isease [PAOD], and heart failure [HF]). These constant features capture the clinical background of the patient, providing valuable context for predicting procedural outcomes.

The serial data focus on vascular-access-related features that vary over time, including access site location, type of vascular access (AVF or AVG), and Qa value. As discussed in section III-A, the derived data were used to capture trends and patterns in vascular access performance, including previous Qa values, recent surgical indicators, and slope of Qa, among others. These serial and derived features provided clinically relevant metrics that increased the predictive power and interpretability of the model.

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

FIGURE 2. Extended confusion matrix. a, True Positive (TP); b, False Negative (FN); c, False Positive (FP); d, True Negative (TN); e, Uncertain Positive (UP); f, Uncertain Negative (UN).

TABLE 3. AVF dataset evaluation metrics for baseline (KDOQI guidelines), our method (Without uncertain), and our method (With uncertain).

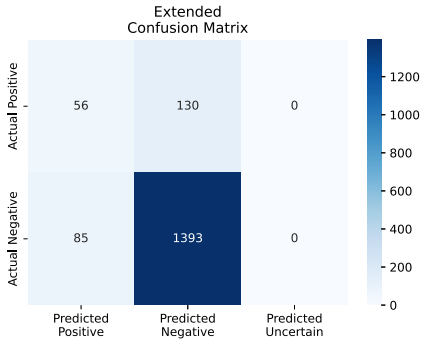
Metric	Baseline (KDOQI Guidelines)	Our Method (Without Uncertain)	Our Method (With Uncertain)
Accuracy	0.872 ± 0.001	0.895 ± 0.004	**0.913 ± 0.001**
PPV (Positive Predictive Value)	0.394 ± 0.011	0.589 ± 0.028	**0.841 ± 0.043**
NPV (Negative Predictive Value)	0.918 ± 0.003	0.904 ± 0.002	0.914 ± 0.001
Error Rate	0.129 ± 0.01	0.105 ± 0.004	**0.081 ± 0.002**
Leakage Rate	0.078 ± 0.012	0.093 ± 0.001	**0.08 ± 0.002**
Overkill Rate	0.051 ± 0.031	0.011 ± 0.003	**0.002 ± 0.002**
Uncertain Rate	-	-	0.073 ± 0.027

TABLE 4. AVG dataset evaluation metrics for baseline (KDOQI guidelines), our method (Without uncertain), and our method (Uncertain).

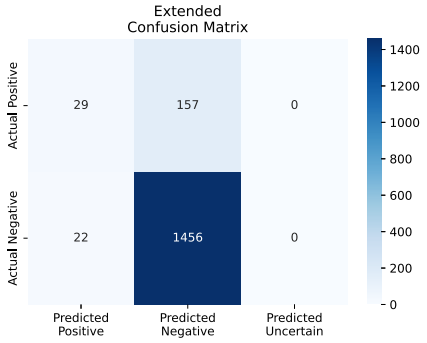
Metric	Baseline (KDOQI Guidelines)	Our Method (Without Uncertain)	Our Method (With Uncertain)
Accuracy	0.791 ± 0.018	0.765 ± 0.023	**0.818 ± 0.019**
PPV (Positive Predictive Value)	0.685 ± 0.044	0.695 ± 0.045	**0.74 ± 0.137**
NPV (Negative Predictive Value)	0.809 ± 0.026	0.788 ± 0.028	**0.828 ± 0.031**
Error Rate	0.208 ± 0.032	0.235 ± 0.023	**0.147 ± 0.017**
Leakage Rate	0.186 ± 0.011	0.19 ± 0.027	**0.128 ± 0.025**
Overkill Rate	0.048 ± 0.016	0.032 ± 0.017	**0.02 ± 0.011**
Uncertain Rate	-	-	0.192 ± 0.007

C. EVALUATION METRICS

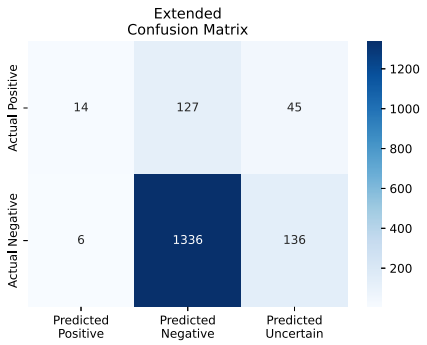
In this study, we define an extended confusion matrix to show the classification results. The extended confusion matrix, shown in Figure 2, categorizes predictions into six groups: True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN), Uncertain Positive (UP), and Uncertain Negative (UN). This expansion enables a more detailed assessment of prediction uncertainty and misclassification risks, which are particularly critical in clinical decision-making. A wide range of metrics can be derived from the extended confusion matrix for comprehensively evaluating model performance, managing uncertainty, and minimizing misclassification risk.



(a) Baseline (KDOQI Guidelines)



(b) Our Method (Without Uncertain)



(c) Our Method (With Uncertain)

FIGURE 3. AVF dataset extended confusion matrix.

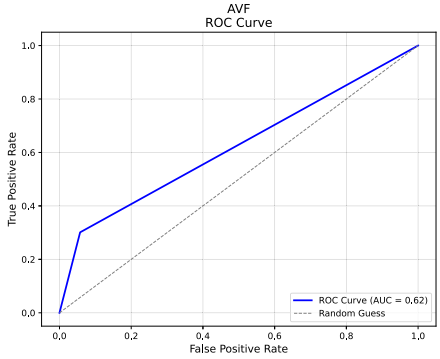
These metrics include Positive Predictive Value (PPV), Negative Predictive Value (NPV), Error Rate, Leakage Rate, Overkill Rate, and Uncertainty Rate. By leveraging these metrics, the framework ensures robust and reliable assessments tailored to high-stakes clinical scenarios.

- **Standard Metrics**:

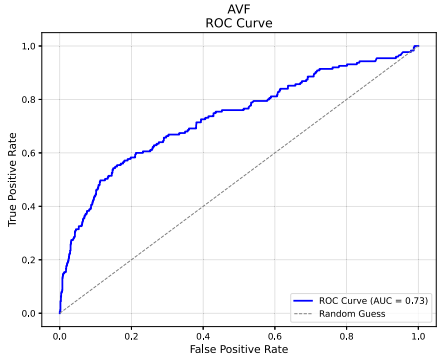
$$Accuracy = \frac{a+d}{a+b+c+d} \tag{5}$$

$$PPV = \frac{a}{a+c} \tag{6}$$

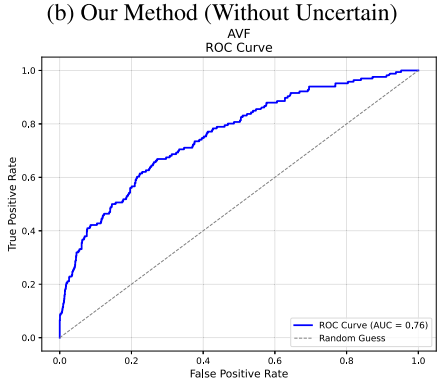
$$NPV = \frac{d}{b+d} \tag{7}$$



(a) Baseline (KDOQI Guidelines)

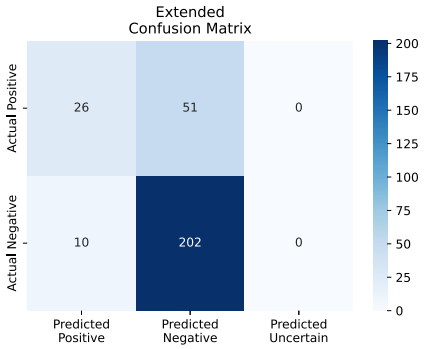


(b) Our Method (Without Uncertain)

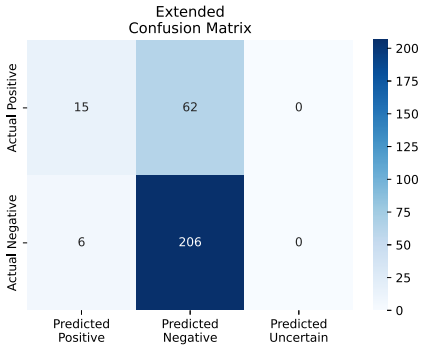


(c) Our Method (With Uncertain)

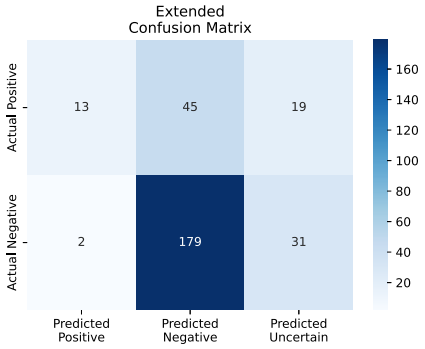
FIGURE 4. AVF dataset ROC curve comparison.



(a) Baseline (KDOQI Guidelines)



(b) Our Method (Without Uncertain)



(c) Our Method (With Uncertain)

FIGURE 5. AVG dataset extended confusion matrix.

- **Uncertainty Metrics**:

$$All = a + b + c + d + e + f \tag{8}$$

$$ErrorRate = \frac{b + c}{All} \tag{9}$$

$$LeakageRate = \frac{b}{All} \tag{10}$$

$$OverkillRate = \frac{c}{All} \tag{11}$$

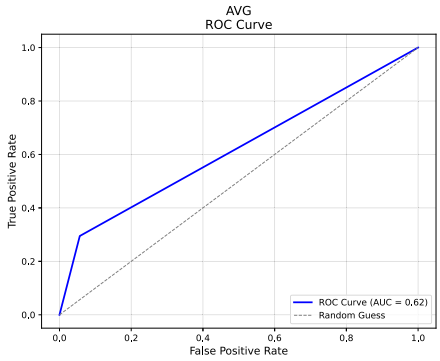
$$UncertainRate = \frac{e + f}{All} \tag{12}$$

Furthermore, the ROC curves are also plotted for evaluating the performance of the proposed algorithms and KDOQI guidelines.

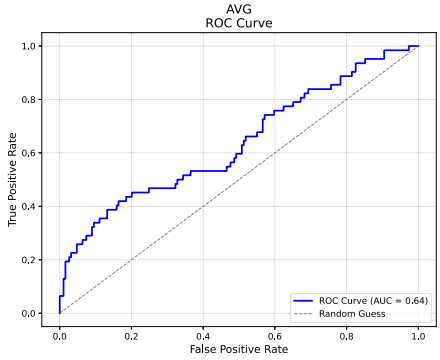
D. EXPERIMENT RESULTS

The evaluation metrics for both the AVF and AVG datasets are summarized in Tables 3 and 4 and visualized in Figures 3, 4, 5, and 6. A detailed discussion of the findings is provided below:

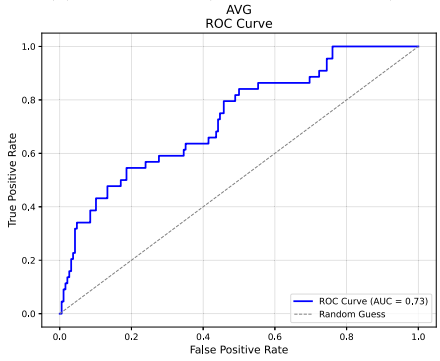
- **AVF Dataset**
  As shown in Table 3, along with Figures 3 and 4, the AVF dataset includes 4,991 cases, 10.9% of which required Percutaneous Transluminal Angioplasty (PTA) interventions. The proposed uncertainty-aware tree-based ensemble model achieved the highest accuracy of 91.3%, surpassing the KDOQI guidelines (87.2%) and the tree-based ensemble model without uncertain classifications (89.5%). This improvement was accompanied



(a) Baseline (KDOQI Guidelines)



(b) Our Method (Without Uncertain)



(c) Our Method (With Uncertain)

FIGURE 6. AVG dataset ROC curve comparison.

by a reduction in the Error Rate to 8.1% and the Leakage Rate to 8.0%. Moreover, the Area Under the Curve (AUC) increased significantly from 0.62 for the KDOQI guidelines to 0.76 for the uncertainty-aware tree-based ensemble model. These results highlight the model's ability to effectively manage ambiguous cases, thereby reducing misclassification risk and increasing prediction reliability.

- **AVG Dataset**
  As presented in Table 4 and Figures 5 and 6, the AVG dataset comprises 869 cases, with a higher PTA intervention rate of 25.1%. This dataset posed unique challenges due to its smaller size and greater variability. The uncertainty-aware tree-based ensemble model demonstrated superior performance, achieving an accuracy of 81.8%, compared to 79.1% for the baseline model. The Error Rate was reduced to 14.7%, while the Leakage Rate was minimized to 12.8%. Additionally, the AUC improved from 0.62 (KDOQI) to 0.73, underscoring the model's robustness in handling datasets with higher variability and imbalance.

These findings underscore the effectiveness of the proposed uncertainty-aware framework in improving diagnostic performance for vascular access dysfunction. By incorporating uncertain classifications, the model not only improves predictive accuracy but also reduces the occurrence of overconfidence errors, ensuring more reliable and cautious decision-making in clinical settings.

Compared 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 uncertainty-aware model consistently outperformed both the KDOQI guidelines and the model without uncertain classifications. The results emphasize the model's adaptability and robustness, making it a valuable tool for supporting clinical decision-making in vascular access management.

According to the 2006 and 2019 KDOQI guidelines, clinicians are recommended to make decisions regarding interventional procedures for dysfunction arteriovenous access based on the clinical symptoms and signs of vascular dysfunction and critical access blood flow (Qa). However, in real-world clinical practice for those hemodialysis patients with borderline or critical access blood flow (Qa) without clinical symptoms and signs or those hemodialysis patients with mild clinical symptoms and signs, this framework can integrate these uncertainties into the existing guideline recommendation to achieve precise treatment decisions in the future.

V. CONCLUSION AND FUTURE WORKS

Taiwan, which has the highest global proportion of dialysis patients, faces significant challenges in managing vascular access dysfunction—a critical complication of hemodialysis. Early and accurate diagnosis is essential for mitigating risks and improving patient outcomes. This study introduces an uncertainty-aware, tree-based machine-learning framework designed to increase diagnostic precision while addressing uncertainties inherent in clinical data.

The proposed methodology integrates Multipass Uncertainty Estimation with Uncertainty-Aware Data Classification, enabling robust predictions that account for ambiguous cases. Additionally, an extended confusion matrix and novel metrics—such as leakage, overkill, and uncertainty—provide a comprehensive evaluation framework, emphasizing reliable and cautious decision-making.

Experimental results underscore the model's superior performance, demonstrating notable improvements in metrics such as accuracy, AUC, and PPV across both the AVF and AVG datasets. The inclusion of uncertain classifications effectively reduces error rates and increases diagnostic reliability, outperforming traditional methods [5].

By categorizing ambiguous predictions into an 'Uncertain' class, the model supports cautious decision-making and mitigates misclassification risk. Example. A patient with AVF shows Qa close to the threshold and a conflicting recent slope; the model returns Uncertain, prompting repeat Qa within the next session or vascular sonography before intervention scheduling. This approach aligns with the clinical demand for reliable, interpretable, and actionable diagnostic tools that balance precision with caution.

This study used a single-center dataset (412 patients, 5,860 records), which may limit external generalizability. Future work will include multi-center external validation with patient-level stratification and site-level effects to assess transportability across institutions.

In conclusion, this study highlights the potential of combining machine learning with indeterminacy analysis to improve clinical decision-making in complex datasets. The proposed framework offers a scalable foundation for advancing medical AI, enabling more accurate diagnoses, minimizing unnecessary interventions, and supporting timely treatments for dialysis patients.

ACKNOWLEDGMENT

The computation resource was supported in part by aiForge of National Yang Ming Chiao Tung University.

REFERENCES

[1] R. Navuluri and S. Regalado, "The KDOQI 2006 vascular access update and fistula first program synopsis," *Seminars Interventional Radiol.*, vol. 26, no. 2, pp. 122–124, Jun. 2009, doi: 10.1055/s-0029-1222455.

[2] "Vascular access 2006 work group clinical practice guidelines for vascular access," *Amer. J. Kidney Diseases*, vol. 48, no. 1, pp. S176–S247, 2006, doi: 10.1053/j.ajkd.2006.04.029.

[3] C. E. Lok, T. S. Huber, T. Lee, S. Shenoy, A. S. Yevzlin, K. Abreo, M. Allon, A. Asif, B. C. Astor, M. H. Glickman, J. Graham, L. M. Moist, D. K. Rajan, C. Roberts, T. J. Vachharajani, and R. P. Valentini, "KDOQI clinical practice guideline for vascular access: 2019 update," *Amer. J. Kidney Diseases*, vol. 75, no. 4, pp. S1–S164, Apr. 2020, doi: 10.1053/j.ajkd.2019.12.001.

[4] 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 Educ.-J. Assoc. Surgical Educ.*, vol. 3, no. 1, p. 56, May 2024, doi: 10.1007/s44186-024-00257-2.

[5] C.-K. Wu and C.-H. Lin, "Integrating vascular access surveillance with clinical monitoring for stenosis prediction," *J. Nephrol.*, vol. 37, no. 2, pp. 461–470, Nov. 2023, doi: 10.1007/s40620-023-01799-2.

[6] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, 1st ed., Boca Raton, FL, USA: CRC Press, 1984, doi: 10.1201/9781315139470.

[7] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/a:1010933404324.

[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," 2016, *arXiv:1603.02754*.

[9] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. NIPS*, 2017.

[10] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient boosting with categorical features support," 2018, *arXiv:1810.11363*.

[11] C. Bentéjac, A. Csörgo, and G. Martínez-Muñoz, "A comparative analysis of XGBoost," 2019, *arXiv:1911.01914*.

[12] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[13] E. Begoli, T. Bhattacharya, and D. Kusnezov, "The need for uncertainty quantification in machine-assisted medical decision making," *Nature Mach. Intell.*, vol. 1, no. 1, pp. 20–23, Jan. 2019.

[14] C. Chen, J. Liang, F. Ma, L. Glass, J. Sun, and C. Xiao, "UNITE: Uncertainty-based health risk prediction leveraging multi-sourced data," in *Proc. Web Conf.*, Apr. 2021, pp. 217–226.

[15] N. Tomašev et al., "Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records," *Nature Protocols*, vol. 16, no. 6, pp. 2765–2787, Jun. 2021.

[16] V. Edupuganti, M. Mardani, S. Vasanawala, and J. Pauly, "Uncertainty quantification in deep MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 239–250, Jan. 2021.

[17] B. Lambert, F. Forbes, S. Doyle, H. Dehaene, and M. Dojat, "Trustworthy clinical AI solutions: A unified review of uncertainty quantification in deep learning models for medical image analysis," *Artif. Intell. Med.*, vol. 150, Apr. 2024, Art. no. 102830.

[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," 2023, *arXiv:2305.04933*.

[19] M. W. Dusenberry, D. Tran, E. Choi, J. Kemp, J. Nixon, G. Jerfel, K. Heller, and A. M. Dai, "Analyzing the role of model uncertainty for electronic health records," in *Proc. ACM Conf. Health, Inference, Learn.*, Apr. 2020, pp. 204–213.

[20] V. Jahmunah, E. Y. K. Ng, R.-.-S. Tan, S. L. Oh, and U. R. Acharya, "Uncertainty quantification in DenseNet model using myocardial infarction ECG signals," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2023, Art. no. 107308.

[21] T. Mimori, K. Sasada, H. Matsui, and I. Sato, "Diagnostic uncertainty calibration: Towards reliable machine predictions in medical domain," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020, pp. 3664–3672.

[22] T. Abe, E. K. Buchanan, G. Pleiss, and J. P. Cunningham, "Pathologies of predictive diversity in deep ensembles," *Trans. Mach. Learn. Res.*, 2023.

[23] M. S. Ayhan, L. Kühlewein, G. Aliyeva, W. Inhoffen, F. Ziemssen, and P. Berens, "Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection," *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101724.

[24] 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, p. 396, Jan. 2022, doi: 10.3390/electronics11030396.

[25] M. Tulio Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," 2016, *arXiv:1602.04938*.

[26] C. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 1, pp. 41–46, Jan. 1970, doi: 10.1109/TIT.1970.1054406.

[27] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017.

[28] Y. Geifman and R. El-Yaniv, "SelectiveNet: A deep neural network with an integrated reject option," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019.

[29] K. Hendrickx, L. Perini, D. Van der Plas, W. Meert, and J. Davis, "Machine learning with a reject option: A survey," *Mach. Learn.*, vol. 113, no. 5, pp. 3073–3110, May 2024, doi: 10.1007/s10994-024-06534-x.

[30] M. Swaminathan, P. K. Yadav, O. Piloto, T. Sjöblom, and I. Cheong, "A new distance measure for non-identical data with application to image classification," *Pattern Recognit.*, vol. 63, pp. 384–396, Mar. 2017, doi: 10.1016/j.patcog.2016.10.018.

[31] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on tabular data?" 2022, *arXiv:2207.08815*.

[32] D. C. McElfresh, S. Khandagale, J. Valverde, G. Ramakrishnan, M. Goldblum, and C. White, "When do neural nets outperform boosted trees on tabular data?" in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2023, pp. 76336–76369.

[33] 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," 2024, *arXiv:2407.00956*.

CHIA-HSUN LIN received the M.D. degree from Taipei Medical University, Taipei, Taiwan, in 1991. He is currently a Professor with the School of Medicine, Fu Jen Catholic University and the Director of the Department of Cardiovascular Surgery, Shin Kong Wu Ho-Su Memorial Hospital. His clinical and research interests include cardiovascular surgery, dialysis vascular access, and aortic and peripheral vascular interventions. He is the Honorary President of Taiwan Vascular Surgery Society.

ZHENG LIN CHEN received the M.S. degree from the Institute of Intelligent Systems, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. His research interests include machine learning and indeterminate predictions.

CHUNG-KUAN WU received the M.D. degree from Kaohsiung Medical University, Kaohsiung, Taiwan, in 2002, and the Ph.D. degree in clinical medicine from the National Yang-Ming University, Taipei, Taiwan, in 2020. He is currently an Associate Professor with the School of Medicine, Fu Jen Catholic University; and an attending Physician with the Department of Nephrology, Shin Kong Wu Ho-Su Memorial Hospital. He is also the Director of the Hemodialysis Center and the Dialysis Vascular Access Management Center. His research interests include nephrology, dialysis, vascular access, and hemodynamics.

TSEN-CHE WU received the B.S. degree in medical imaging and radiological sciences from Chung Shan Medical University, Taichung, Taiwan. He is currently pursuing the M.S. degree with the Institute of Computational Intelligence, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. His research interests include uncertainty-aware machine learning and trustworthy AI systems.

CHING-WEN MA (Member, IEEE) received the M.S. degree from the Department of Control Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 1992, and the Ph.D. degree from the Department of Electrical and Control Engineering, National Chiao Tung University, in 1998. Previously, he was a Technical Manager with MediaTek Inc., Hsinchu. He is currently an Associate Professor with the College of Artificial Intelligence, National Yang Ming Chiao Tung University. He is actively engaged in both academic research and industry-academia collaborative projects. His work has been published in leading venues, such as IEEE, UAI, and WACV. His research interests include uncertainty quantification in AI models, hallucination phenomena in large language models, zero-shot image generation using diffusion models, and related topics.