

Machine learning approach using electrochemical immunosensor data for precise classification of *Opisthorchis viverrini* infection

Received: 12 June 2024

Accepted: 17 November 2025

Published online: 03 December 2025

Cite this article as: Aye N.N.S., Daduang S., Tippayawat P. *et al.* Machine learning approach using electrochemical immunosensor data for precise classification of *Opisthorchis viverrini* infection. *Sci Rep* (2025). <https://doi.org/10.1038/s41598-025-29347-x>

Nang Noon Shean Aye, Sakda Daduang, Patcharaporn Tippayawat, Anchalee Techasen, Pornsuda Maraming, Paiboon Sithithaworn, Rungrueang Phatthanakun & Jureerut Daduang

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Machine learning approach using electrochemical immunosensor data for precise classification of *Opisthorchis viverrini* infection

Nang Noon Shean Aye^{1,2}, Sakda Daduang³, Patcharaporn Tipayawat², Anchalee Techasen^{2,4}, Pornsuda Maraming², Paiboon Sithithaworn^{4,5}, Rungrueang Phatthanakun^{6*}, Jureerut Daduang^{2*}

¹ Biomedical Sciences Program, Graduate School, Khon Kaen University, Khon Kaen 40002, Thailand; nangnoonshean.aye@kkumail.com

² Centre for Research and Development of Medical Diagnostic Laboratories, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen 40002, Thailand; patchatip@kku.ac.th (PT), pornsma@kku.ac.th (PM)

³ Division of Pharmacognosy and Toxicology, Faculty of Pharmaceutical Sciences, Khon Kaen University, Khon Kaen 40002, Thailand; sakdad@kku.ac.th

⁴ Cholangiocarcinoma Research Institute, Khon Kaen University, Khon Kaen 40002, Thailand; anchte@kku.ac.th

⁵ Department of Parasitology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand; paib_sit@kku.ac.th

⁶ Synchrotron Light Research Institute, Nakhon Ratchasima 30000, Thailand; rungrueang@slri.or.th*

* Corresponding authors. Tel. +66868511422, email: rungrueang@slri.or.th, jurpoo@kku.ac.th

Abstract— Opisthorchiasis, a major foodborne parasitic zoonotic disease in Thailand and neighboring countries, is caused by the carcinogenic liver fluke *Opisthorchis viverrini* (OV). Accurate classification of OV infection is critical for timely intervention and public health management. In this study, we propose a reliable machine learning (ML) classification model based on peak current data from an electrochemical immunosensor and additional patient condition features, which can facilitate intuitive decision-making without the need for expert personnel. This is the first report to classify OV infection using a ML algorithm integrated with electrochemical biosensor data. A total of 531 urine samples from both OV-positive and OV-negative individuals in endemic areas were analyzed using the immunosensor. We evaluated the effectiveness of six different ML models through cross-validation. Among these models, the decision tree and AdaBoost classifiers demonstrated outstanding performance, each achieving the highest accuracy of 90.65% (95% CI 0.89 – 0.91). The decision tree model yielded an F1 score of 91%, sensitivity of 95%, and specificity of 83%, while the AdaBoost model achieved an F1 score of 90%, sensitivity of 94%, and a higher specificity of 86%. The neural network model also performed excellently, with an accuracy of 89.72%

(95% CI 0.84 – 0.93), and an F1 score of 89%. The statistical comparison of the model's performance highlighted the significant difference between the top-performing models and the rest. These results underscore the significance of incorporating sensor data and ML to accurately classify OV infections and enable early diagnosis and intervention. By using this ML model, the status of OV infection can be detected by interpreting sophisticated raw electrochemical data. This implies that patients or medical staff with no prior experience with electrochemical sensors can nevertheless comprehend the disease condition with confidence. The proposed ML models hold promise for enhancing disease surveillance and control strategies in endemic regions and could thus assist medical professionals in the decision-making process and in addressing the burden of opisthorchiasis infection.

Keywords: *Opisthorchis viverrini*, Electrochemical immunosensor, Machine learning, Decision tree, AdaBoost

1. Introduction

Opisthorchis viverrini (OV) infection is a major foodborne parasitic disease and a neglected, clinically significant tropical disease ¹. Approximately 40 million people are infected with OV in the GMS region alone ^{2,3}, and more than 600 million people are at risk ⁴. The life cycle of OV can take place in multiple hosts, including freshwater snails (*Bythinia* spp.), freshwater fish (*Cyprinid* spp.) and fish-consuming hosts such as humans, cats, or dogs. The primary mechanism of infection is the ingestion of raw or inadequately cooked fish cuisines in Thailand. Notably, opisthorchiasis is the most critical risk factor associated with cholangiocarcinoma (CCA) and is classified as a group 1 biological carcinogen ⁵. Early detection, diagnosis and prevention of OV infection constitute an effective approach for addressing the high prevalence and control of the disease in endemic areas. Highly sensitive detection methods and precise classification of infections are urgently needed for timely intervention and surveillance of OV infections.

The gold standard for detecting OV infection is the formalin-ethyl acetate concentration technique (FECT), which is commonly used to quantify OV in fecal samples ⁶. However, the FECT has limitations, including low sensitivity and specificity, extensive sampling requirements, and the need for a skilled microscopist. Additionally, the FECT may fail to detect mild infections, and biliary tract obstruction can impede OV egg flow into feces, leading to false-negative results ⁷. Recent advancements in immunological and molecular diagnostic methods, such as enzyme-linked immunosorbent assay (ELISA), have been explored for the detection of somatic and excretory-secretory (ES) antigens ^{8,9}. Although these methods offer higher sensitivity and specificity over the FECT, their disadvantages include high testing costs, complex procedures, and the need for sophisticated instruments. Serological OV antibody testing is a highly accurate diagnostic technique but has drawbacks such as cross-reactivity and persistent antibodies post-treatment. Attempts to develop ELISA techniques for urine OV-ES antigen detection have shown promise but are hindered by detection limits and complex procedures, highlighting the need for a simple, cost-effective point-of-care biosensor for OV detection.

Since the successful implementation of the amperometric glucose sensor for diabetes monitoring, many electrochemical biosensors have been developed at the industrial level ¹⁰. The development of point-of-care biosensors has revolutionized onsite detection and disease screening in the community. Electrochemical biosensors have strong advantages, including cost-effectiveness and a small size allowing portability for point-of-care testing. A typical bioelectrochemical sensor is an analytical device that integrates a biorecognition component termed a receptor with an electrochemical transducer ¹¹. Our group recently developed an electrochemical immunosensor for sensitive detection

of OV antigens in urine samples ¹². The developed immunosensor exhibited the lowest limit of detection (LOD) of 1.5 ng mL⁻¹ for OV antigen detection in urine compared to 78 ng mL⁻¹ for detection by ELISA and 28.5 ng mL⁻¹ for detection by a rapid OV Ag detection test kit (OV-RDT). Compared with ELISA and OV-RDT, the developed electrochemical sensor demonstrated high diagnostic sensitivity and specificity (96% and 90%, respectively) for OV detection. However, the major challenges associated with electrochemical biosensors include low signal-to-noise (S/N) ratios, electrode fouling, chemical interference, complex signal processing, and matrix effects, which undermine the precision and accuracy of biosensors ¹³.

The incorporation of AI into biosensors has emerged as a solution for overcoming their limitations. Machine learning (ML), a critical aspect of AI, has been employed as a powerful tool for data analysis and processing in materials science ¹⁴. ML can effectively process big sensing data for complex matrices or samples and extract detectable analytical signals from noisy and low-resolution sensing data. Furthermore, proper implementation of ML methods can uncover correlations between sample parameters and sensing signals through data visualization and enable mining of interrelations between signals and bio events. ML-reinforced electrochemical biosensors can improve sensor performance by transforming unintelligible raw data into understandable and valuable information that can be applied in disease diagnosis, treatment evaluation, pathogen detection ¹⁵, health monitoring ¹⁶, and food safety ¹⁷. The common primary machine learning algorithms include artificial neural networks (ANNs), convolutional neural networks (CNNs), decision trees (DTs), random forests (RFs), support vector machines (SVMs) and k-nearest neighbors (kNNs) ¹³. Certain challenges associated with biosensors can be overcome by ML-based data analysis. Moreover, intelligent biosensors that can automatically predict concentrations of analytes or species can be optimized by incorporating ML.

In this study, we proposed a reliable ML classification model based on peak current data from an electrochemical immunosensor and additional patient condition features, which can facilitate intuitive decision-making without the need for expert personnel. This is the first report of classifying OV infection using an ML algorithm integrated with electrochemical biosensor data. Technically, mathematical calculations must be carried out by technicians on peak current changes obtained from sensors to translate these changes into concentrations at which doctors can interpret the status of OV infection. ML integration with biosensors can enable classification of the levels of OV infection based on raw sensor data, thereby obviating manual calculations. Moreover, intelligent OV classification can improve the accuracy of sensors and thus facilitate sensor measurements and data interpretation in a timely manner.

2. Materials and Methods

2.1. Electrochemical setup

The electrochemical immunosensor was fabricated by immobilizing a graphene oxide-monoclonal antibody (GO-mAb) conjugate onto screen-printed carbon electrodes (SPCEs) for 90 minutes. Excess areas were blocked with 0.1% BSA solution for 30 minutes. Electrochemical analysis of OV antigens in urine samples was performed by introducing seven microliters of the samples on modified SPCEs, followed by incubation for 60 minutes at room temperature. The voltammetric signals of OV antigens were collected by square wave voltammetry (SWV) with an applied potential of -0.6 to 0.6 V, an amplitude of 2 mV, a step potential of 5 mV and a frequency of 8 Hz in the presence of 5 mM $[\text{Fe}(\text{CN})_6]^{3-/4-}$ as a redox indicator. The peak currents before (I_0) and after (I_1) OV antigen detection were recorded by SWV. Peak current changes ($\Delta I = I_0 - I_1$) were calculated to identify correlations with OV antigen levels in the samples.

2.2. Data labeling

A total of 531 urine samples from both OV-positive and OV-negative individuals were collected from endemic areas in Northeastern Thailand and were analyzed using the immunosensor. Peak current changes (ΔI) were extracted from square wave voltammograms. Demographic information such as age, sex, and urinalysis results (urine glucose, ketones, protein, pH, specific gravity and red blood cell levels) were collected as clinical parameters. The characteristics of the collected data are shown in Table 1. Missing values, redundant information, and incompatible data strongly impacted model performance. Therefore, the data were preprocessed by data cleaning, data integration, and feature engineering. The categorical variables, such as Sex, were transformed into a numerical format using label encoding. Each unique category was assigned an integer value, and the original variable was preserved alongside the encoded version for reference. To ensure data completeness and consistency, all rows containing missing data were removed prior to model training. Inputs (features) were selected based on their correlations with OV infection. For supervised classification, OV-positive and OV-negative data were labeled as confirmed by the conventional ELISA.

Table 1 Dataset description.

No	Feature	Data type	Description
1	Age	Integer	Patient age when first diagnosed
2	Sex	Male = 1	Patient sex

		Female = 0	
3	Peak current changes	Float	Peak current changes measured in urine samples by the electrochemical immunosensor
4	Glucose	Integer	Routine urine examination analysis
5	Ketones	Integer	Routine urine examination analysis
6	Protein	Integer	Routine urine examination analysis
7	pH	Float	Routine urine examination analysis
8	Specific gravity	Float	Routine urine examination analysis
9	Blood	Integer	Routine urine examination analysis
10	Label	OV-positive = 1 OV-negative = 0	Classification of OV infection status confirmed by standard methods

2.3. Exploratory data analysis

The dataset contained 531 rows of data. The labels that we used were OV-positive and OV-negative, representing the infection status of the patients. The number of patients with each label is depicted in Figure 1 (a). A total of 204 patients were observed to have OV infection, and 327 individuals did not. The sex ratio is illustrated in Figure 1 (b). A total of 135 in 204 OV-infected patients were female, accounting for 66% of the patients, and 69 male patients accounted for 34%. The OV-negative individuals included 95 males and 232 females.

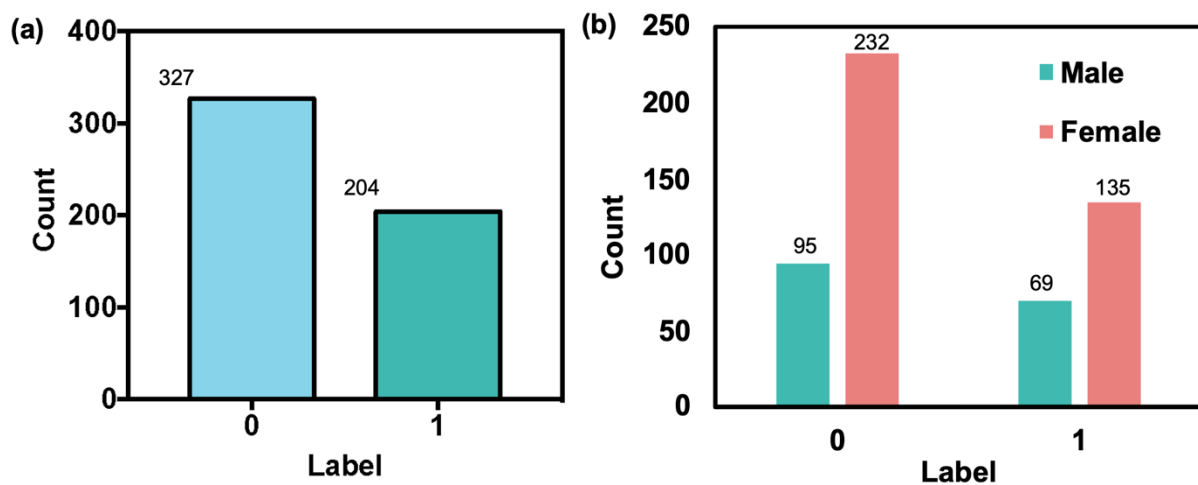


Figure 1 (a) Data distribution per class and (b) the sex proportion in each class. Labels: 0 = OV-negative, 1= OV-positive.

133 The distributions of age, peak current changes, pH and specific gravity for each label class are
 134 depicted in a box plot (see Figure 2). The mean age of the OV-infected patients (53.34 ± 14.03) was
 135 slightly greater than that of the noninfected individuals (45.19 ± 20.33). Peak current changes were
 136 significantly different between the two groups at $2.21 \pm 2.16 \mu\text{A}$ in the OV-positive patients and 1.10
 137 $\pm 2.15 \mu\text{A}$ in the OV-negative patients. The urine pH and specific gravity data did not differ
 138 significantly between the classes. Urine glucose, ketone and protein levels were within the normal
 139 ranges for both groups, with some outliers that were not associated with the disease (data not shown).

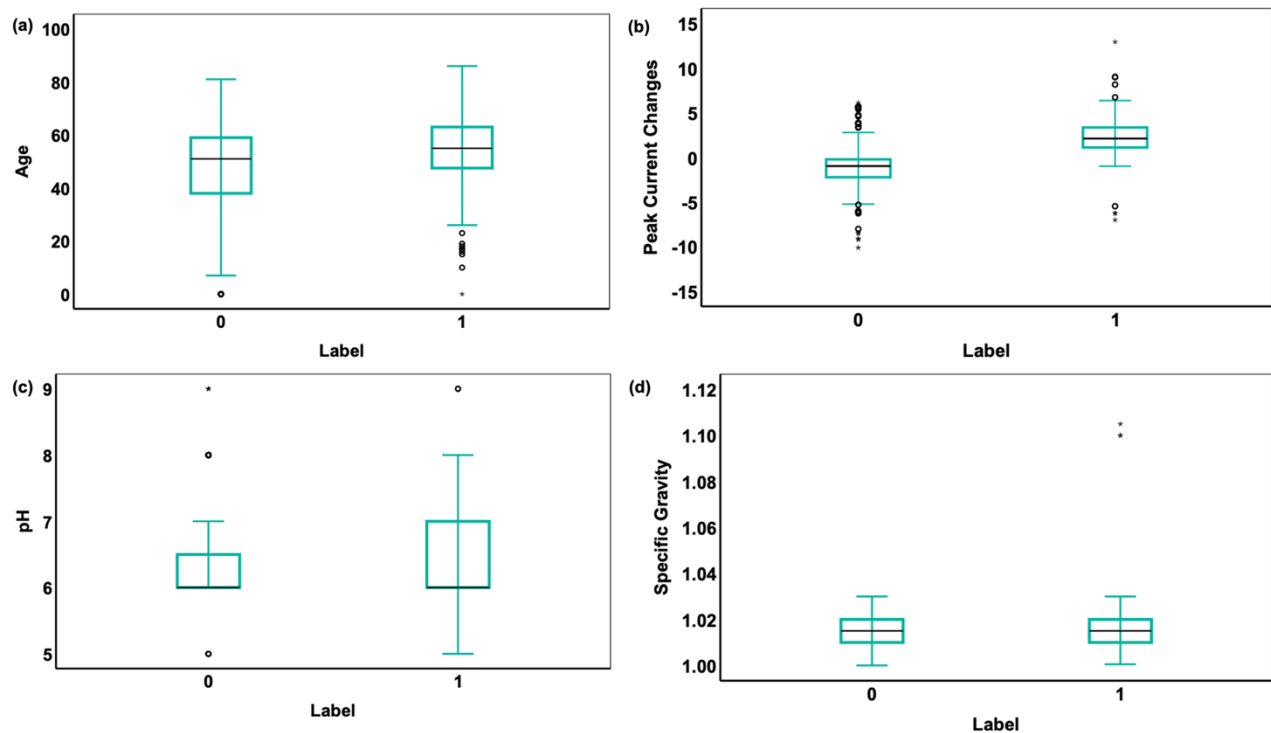


Figure 2 Data distributions of (a) age, (b) peak current changes, (c) pH and (d) specific gravity in each class. Labels: 0 = OV-negative, 1= OV-positive.

140 2.4. Sample size determination

141 For classification models, the minimum sample size required was estimated based on the rule
 142 of thumb that at least 10 events per variable (EPV) are needed for developing a reliable model. Using
 143 9 variables and binary classification in our full model, a minimum of 180 positive cases were required.
 144 Additionally, power analysis was performed assuming 80% power, a significance level of 0.05, and
 145 an expected AUC of 0.75, which indicated a minimum required sample size of 114 samples (57 per
 146 group). Our actual sample size of 531 (204 OV-positive and 327 OV-negative) was substantially
 147 larger than this requirement, providing adequate statistical power for our analyses.

148 2.4.1. Learning curve analysis

A learning curve approach was employed to estimate the adequate training sample size required to achieve a model performance target of $AUC > 0.85$. This approach allows us to evaluate how model performance scales with increasing training data and to determine the point at which performance metrics converge. Each model was trained and validated using 10-fold stratified cross-validation to ensure that both classes were proportionally represented in each fold. The minimum training sample size to achieve $AUC > 0.85$ was estimated by computing the model performance across ten training set sizes, incrementally increasing from 10% to 100% of the total dataset (531 samples) in equal steps. The learning curve for each model was plotted as mean AUC versus training set size, with a horizontal reference line at $AUC = 0.85$. The AUC increased with increasing sample size in all models, reaching a plateau at 40 % of the sample size (~ 0.86 AUC). Therefore, a training set size of approximately 265–320 samples (i.e., ~ 40 –50% of total data) was sufficient for decision tree, AdaBoost, and neural network models to achieve or exceed the target AUC of 0.85.

2.5. Training and validation of classification machine learning models

A variety of ML classifiers were applied in our study to ensure the robustness of our models in interpreting correlations between risk factors and infection outcomes. Six ML classifiers commonly used on datasets composed of heterogeneous features were trained, including a decision tree, k-nearest neighbors, an AdaBoost, a random forest, naïve Bayes, and an artificial neural network. The implementation of these six classification algorithms on the dataset was evaluated by several metrics, including accuracy, precision, recall, F1 score, and cross-validation score. Five-fold cross-validation was implemented to protect against overfitting and ensure the generalizability of the model performance. To account for the different number of samples in the OV-positive and OV-negative groups, the cross-validation was stratified. This ensures that the proportion of cases from each class was maintained in every fold, leading to a more reliable estimate of model performance. In the training phase, all models were optimized by fine-tuning their hyperparameters using a 5-fold cross-validation scheme and a grid search algorithm, with the weighted F1-score as the scoring metric. This approach was applied to all models except Naïve Bayes. For the decision tree, the maximum depth was varied from 1 to 5 to control model complexity. The k-Nearest Neighbors algorithm was evaluated using different numbers of neighbors: 3, 5, and 7, which influence the smoothness of the decision boundary. For AdaBoost, multiple parameters were tuned, including the number of estimators (50, 100, 200), learning rates (0.01, 0.1, 0.5, 1), and the maximum depth of the base decision tree estimators (1, 2, 3), to balance between bias and variance. Random Forest models were tuned by adjusting the number of trees (50, 100, 200) and their maximum depth (5, 10, 20). Lastly, for the Neural Network, different hidden layer architectures (50, 25), (100, 50), and (200, 100) were

182 tested to determine the most effective network structure. These tuning strategies help ensure that each
 183 model is properly configured for optimal predictive accuracy. The code used to develop the models
 184 was written in Python and the open-source library scikit-learn. The overall ML framework is
 185 illustrated in Figure 3.

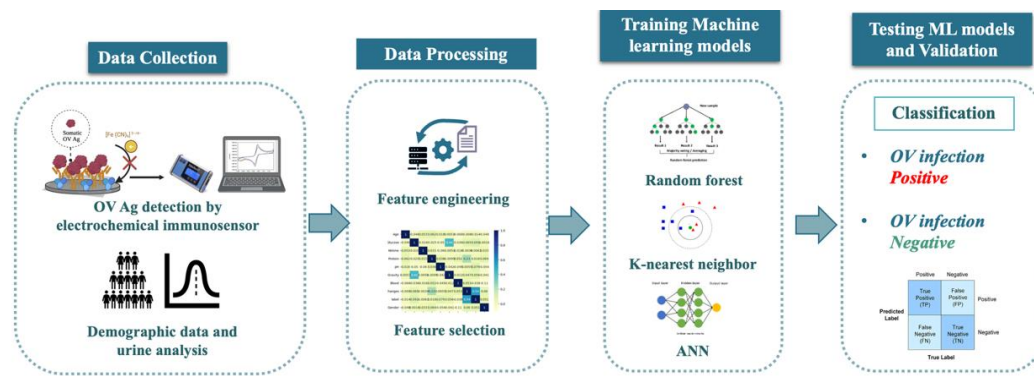


Figure 3 Schematic of the machine learning approach using electrochemical sensor data and additional clinical features for precise classification of *Opisthorchis viverrini* infection.

186 2.6. Machine learning models

187 Decision Tree

188 Decision tree consists of nodes that split the instance space into two or more subspaces according
 189 to a discrete function over the input attributes¹⁸. It works by recursively partitioning the dataset into
 190 subsets based on the values of input features, ultimately leading to a tree-like structure of decision
 191 nodes and leaf nodes. The decision tree model is prone to overfitting. Maximum depth governs the
 192 depth of the tree and can prevent overfitting. By using the grid search, a maximum depth of 1 was set
 193 to train the model.

194 k-Nearest neighbors

195 K-nearest neighbors algorithm is a data-classification method of estimating the likelihood that a
 196 data point will become a member of one group based on what group the data point nearest to it belongs
 197 to. The optimal k-value of 7, determined through fine-tuning, was used to train the model.

198 AdaBoost

199 AdaBoost works by combining the predictions of weak learners, which are typically simple
 200 models that perform slightly better than random guessing. Commonly, decision trees with a single
 201 split (stumps) are used as weak learners in AdaBoost. The AdaBoost algorithm is an iterative
 202 procedure that attempts to approximate the Bayes classifier by combining multiple weak classifiers.
 203 ¹⁹. The AdaBoost classifier was implemented using the scikit-learn library to perform supervised
 204 classification. The base estimator was a decision tree classifier with a maximum depth of 3, allowing

each weak learner to capture more complex patterns than a simple decision stump. The ensemble consisted of 200 weak learners, trained sequentially, with each learner focusing on the errors made by its predecessors. A learning rate of 0.01 was applied to reduce the contribution of each learner and mitigate the risk of overfitting, thereby promoting a more gradual and stable learning process.

Random Forest

Random Forest is an ensemble classifier based on decision tree models. It developed many trees and bootstrap technique and applied it to every tree from a set of training data. Every tree is present as a forest, and then it votes individually for the class outcome. Random Forest will vote for the majority class of the outcome from every tree²⁰. The optimized hyperparameters of 200 estimators and maximum depth 5 have been applied for training the model.

Naïve Bayes

Naïve Bayes is based on Bayes' theorem and assumes independence between features. It calculates the probability of each class given the features and selects the class with the highest probability.

Artificial neural networks

ANNs are computational models inspired by the human brain. They consist of interconnected layers of nodes (neurons). Neurons receive inputs, apply weights, pass the result through an activation function, and send the output to the next layer. In this study, a feedforward multi-layer perceptron (MLP) architecture was implemented using Scikit-learn's MLP Classifier. The proposed ANN model follows a multi-layer perceptron (MLP) architecture. This architecture involved an input layer with 3 neurons, two optimized hidden layers of 100, 50 neurons each, and an output layer. The hidden layers used the ReLU (Rectified Linear Unit) activation function, while the output layer employed a sigmoid activation function suitable for binary classification. The model was trained using the Adam optimizer, with a maximum of 500 iterations.

2.7. Model evaluation metrics

The performance of ML classifiers was evaluated using a range of metrics to gain a comprehensive understanding of their robustness and effectiveness. The common evaluation metrics for binary classification are accuracy, F1-score, sensitivity, and specificity, which express the percentage of correctly classified instances in the set of all the instances, the truly positive instances, the truly negative instances, or the instances classified as positive, respectively. Sensitivity is commonly referred to as recall²¹.

Recall (or sensitivity) measures the proportion of actual positives that were correctly identified by the model. Accuracy represents the overall percentage of correct predictions among all predictions made, giving an overall snapshot of model performance. F1-score is the harmonic mean of precision and recall, providing a balanced measure especially useful when the class distribution is uneven. Additionally, a Receiver Operating Characteristic (ROC) curve was generated, and the Area Under the Curve (AUC) was calculated. The ROC curve illustrates the trade-off between the true positive rate and the false positive rate across different classification thresholds. To further assess model robustness, cross-validation was employed, particularly useful when data is limited.

In practice, ML algorithms are evaluated using a confusion matrix that provides classes, such as the probability of true-negative (TN), true-positive (TP), false-negative (FN), and false-positive (FP). The confusion matrix was further utilized to evaluate the classification accuracy, precision, and recall. The significant performance indicators were obtained using the following equations [1 - 4]:

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad [1]$$

$$Recall = \frac{TP}{TP+FN} \quad [2]$$

$$Specificity = \frac{TN}{TN+FP} \quad [3]$$

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} \quad [4]$$

Confidence intervals for model performance metrics were estimated using the bootstrap percentile method with 1,000 resamples. The resulting 95% intervals, derived from the empirical distribution of the resampled statistics, were incorporated into the interpretation of model performance.

2.8. Code availability

The custom code used to perform the analysis in this study is provided as part of the replication package. It is available at supplementary B for review. It will be uploaded to Zenodo repository once the paper has been conditionally accepted.

3. Results

3.1. Electrochemical biosensing of urine samples

Field-collected urine samples with or without OV antigens were detected by the developed electrochemical immunosensor as detailed above. The resulting peak current with varying electrical potential was recorded. Examples of SWV curves for OV-positive and OV-negative individuals are shown in Figure 4. In the OV-positive samples, as the OV antigen present in the urine bound to the immobilized antibody on the immunosensor surface, the relative peak current was significantly reduced (Figure 4a). This phenomenon was caused by the antigen-antibody complex acting as a hindrance on the electrode surface, leading to less penetration of redox species and a subsequent reduction in the relative signal. Conversely, increasing peak current or no current changes were observed in the OV-negative samples due to the release of the immobilized antibody (Figure 4b). The square wave voltammograms generated the raw data signal. The raw data graphs were automatically calculated to extract peak current changes (ΔI) by Python programming. First, the desired peak current areas were defined by limiting the working voltage range from -0.1 to 0.6 V. The index of peak points and the lowest points were located, and the baseline was drawn between the two lowest points. The distance between the peak point and the baseline is defined as the peak height of the sensor. The distance difference between peak 1(baseline) and peak 2 (Sample) indicated the peak current change ($\Delta I/\mu A$). The resulting ΔI is directly proportional to OV antigen concentration as described in our previous study ¹². A standard curve was constructed to translate the ΔI into an understandable OV antigen concentration. The new dataset can further be tested to extract the information about ΔI value and/or OV antigen concentration from a single raw-spectra. The requirement of complicated signal processing and post-processing data analysis were overcome by the automated calculation programming.

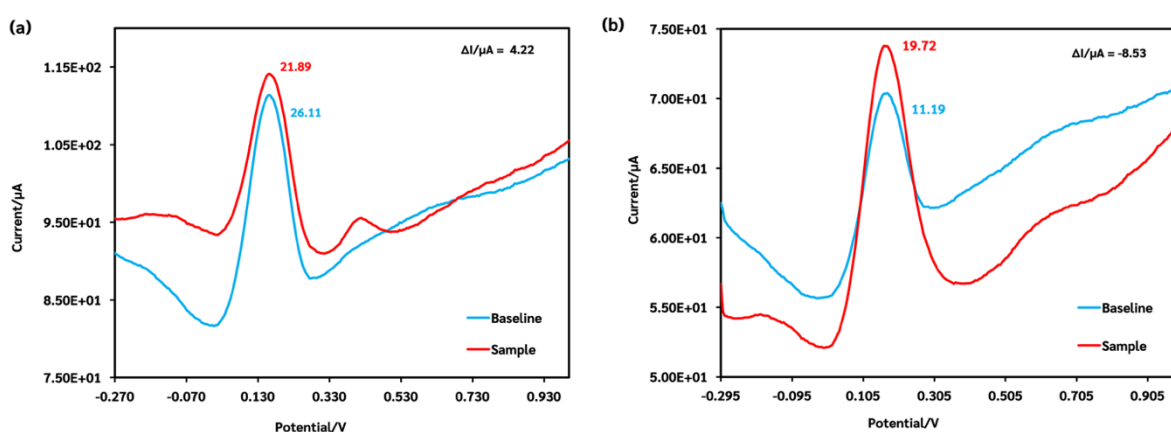


Figure 4 The square wave voltammograms of (a) OV-positive and (b) OV-negative urine samples showing the peak current changes (ΔI).

3.2. Feature selection

The associations between the collected input parameters and the labels were visualized by a correlation heatmap. The heatmap visualization can be seen in Figure 5. The peak current changes obtained from the electrochemical sensor measurements were most strongly correlated with OV infection (60%), followed by patient age (21%) and urine pH (14%) (Figure 5). The correlation between age and OV positivity was consistent with results from previous studies indicating that older age (> 60 years) is a risk factor for OV infection ^{22,23}.



Figure 5 Heatmap showing the correlations of all features with OV infection status (label).

3.3. Feature importance analysis

The permutation-based feature importance analysis quantifies how much model performance decreases when each feature's values are randomly shuffled, disrupting the true relationship between the feature and the target variable ²⁴. The permutation feature importance analysis (Figure 6) provides critical insights into the predictive power of each feature in our machine learning model for OV infection detection. The results demonstrate that peak current changes obtained from electrochemical sensor measurements were the most important predictor of OV infection, with a relative importance of approximately 0.68 (68%). This strongly supports our sensor-based approach, indicating that the electrochemical signals directly associated with OV infection biomarkers provide the most significant diagnostic information. This finding extends beyond simple correlation analysis by measuring the actual impact on model prediction accuracy when this feature is removed or altered.

Age emerged as the second most important feature with a relative importance of approximately 0.15 (15%), confirming our initial correlation analysis findings. Specific gravity ranked third in importance (approximately 6%), despite showing only a moderate correlation in the previous analysis. This elevated importance suggests that specific gravity may capture unique information about OV infection status that is not linearly correlated but becomes valuable when combined with other features in a complex model. Urine pH showed approximately 3% importance, which is substantially lower than its correlation coefficient (14%). This decrease in relative importance may indicate that while pH correlates with OV infection, much of its predictive power may be captured by other features, particularly peak current changes. The remaining features such as gender, protein, glucose, blood, and ketone, demonstrated minimal importance, validating our feature selection process that excluded these parameters from the final model. The permutation importance analysis strongly supports our three-feature model (peak current changes, age, and pH) while suggesting that specific gravity might also be considered in future model iterations. The clear dominance of peak current changes validates our electrochemical sensing approach as the primary diagnostic modality, with demographic and urinalysis parameters serving as valuable supplementary information.

In our ML model, the decision pathway can be explicated as a sequential process by which the algorithm evaluates key features to reach a prediction. When an input sample is processed, the model first heavily weighs the peak current changes, which represent electrochemical sensor responses to potential OV biomarkers in urine. Peak current changes above 0.11 μA strongly directed the model toward a positive OV infection classification, serving as the first and most critical decision point with the highest information gain. If the peak current changes are borderline, the model then considers age, with higher age (particularly >60 years) contributing additional risk weight toward OV positivity. Urine pH and other urinalysis parameters (e.g., protein, glucose) are given less influence, and generally do not affect the classification outcome unless the primary features are inconclusive.

In this study, feature selection was based on a combination of correlation analysis, permutation-based feature importance analysis, and domain expertise. Features with a Pearson correlation coefficient above 0.10 with the target label were considered for model input. Peak current change, age, and urine pH emerged as the most relevant variables. Domain knowledge also informed selection: age is a recognized risk factor for OV infection, and urinalysis parameters can influence electrochemical sensor output. While mutual information was not directly applied, permutation-based importance was used to capture feature relevance, accounting for both linear and non-linear contributions to the model.

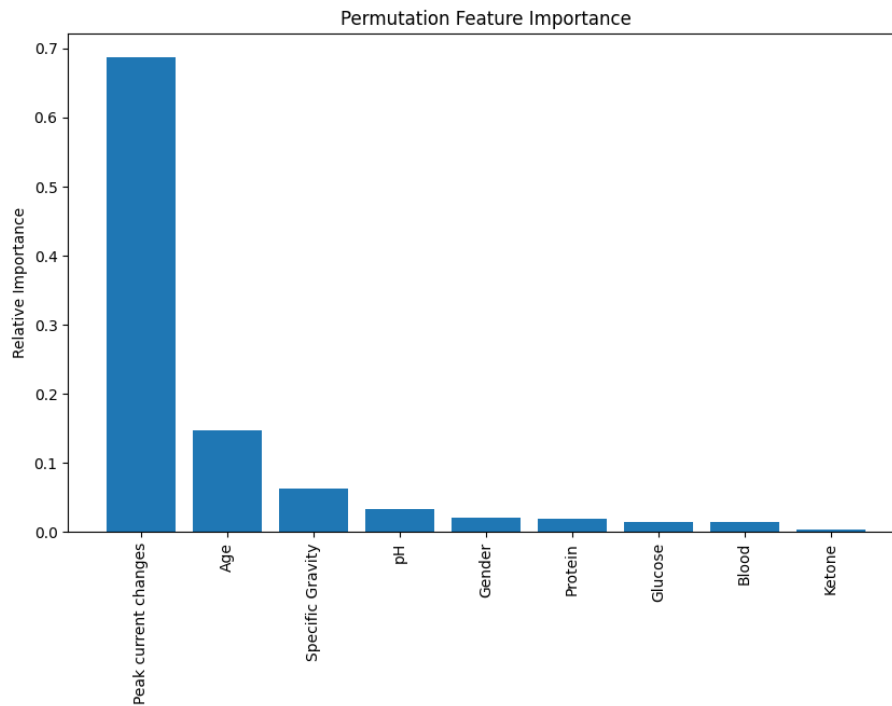


Figure 6 Permutation feature importance for the clinical and sensor features

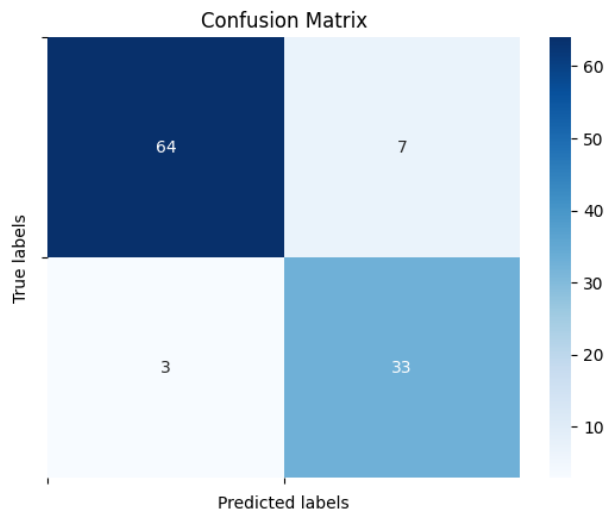
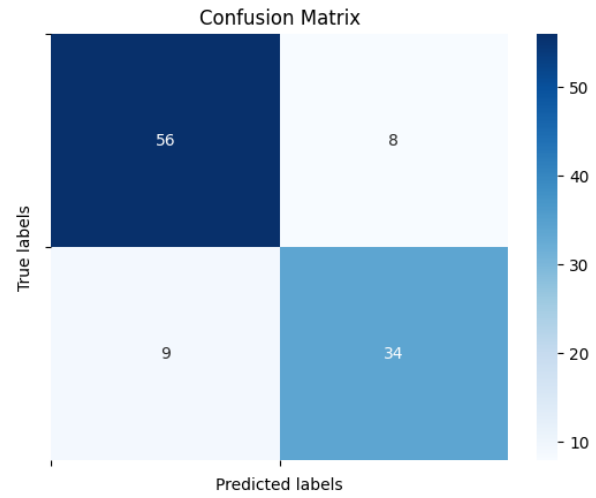
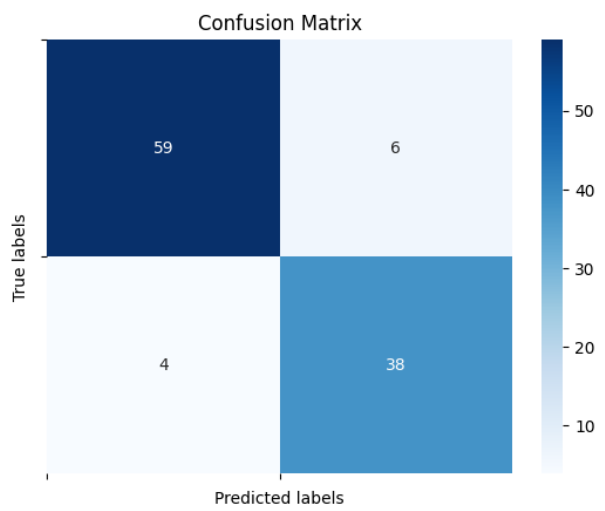
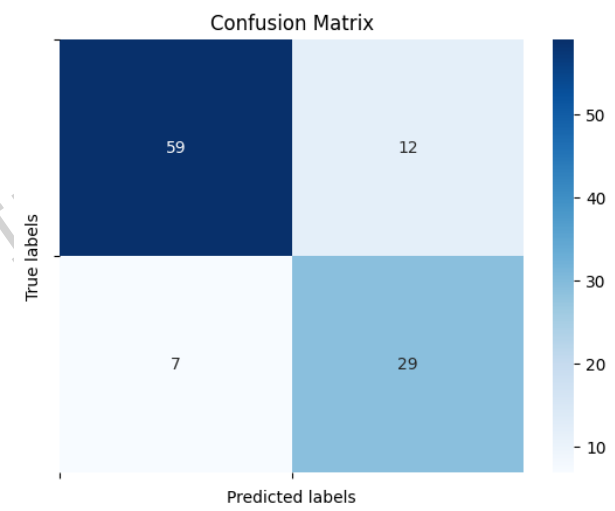
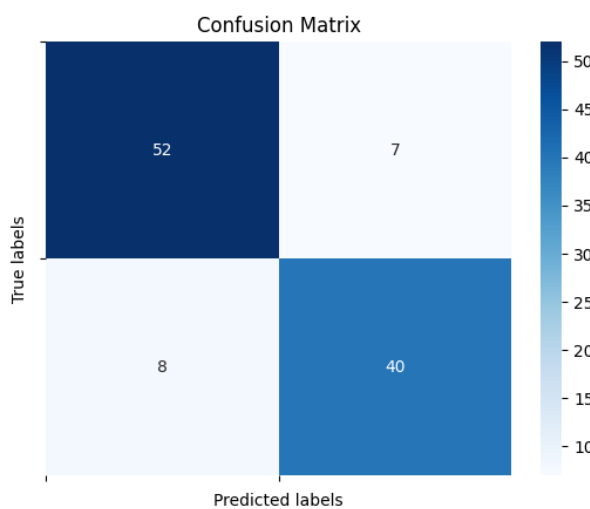
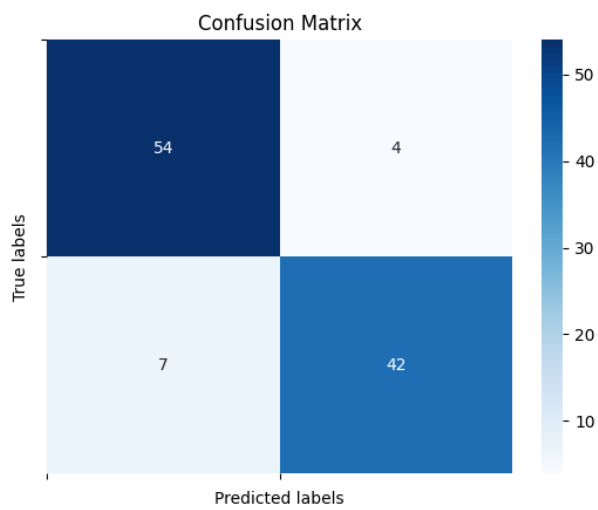
3.4. Classification performance and validation of ML models

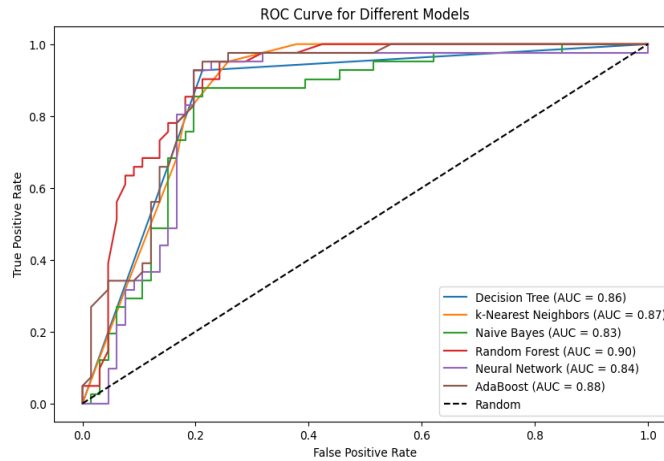
The performance of ML classifiers using different input variables was described in Table 2. Most classifiers, such as decision trees, AdaBoost, naïve Bayes, and neural networks, achieved accuracy levels above 80% and excellent AUC values exceeding 0.84 when using only peak current changes as an input. Eliminating the peak current changes reduced accuracy by 31%, while removing the clinical feature decreased the accuracy by 2%, confirming the hierarchical importance of these features in the decision process. While the decision tree model can effectively distinguish between OV-positive and OV-negative samples using sensor features alone, its accuracy decreases when clinical features are employed without sensor data. There is no significant difference in performance between using 9 features and 3 features, but notably the 3 selected features provided the highest accuracy. The available clinical data are observed to be insufficient to enable the precise classification. As a result, the developed ML model demonstrated the significance of incorporating sensor data and machine learning to accurately classify OV infections and enable early diagnosis and intervention. This comprehensive analysis of model interpretability and decision pathways provides transparent insight into how our ML model utilizes the measured features to detect OV infection. The clear dominance of peak current changes in the decision process, supported by demographic and urinary parameters at specific decision thresholds, validates our sensor-based approach while providing clinicians with an understandable framework for interpreting model outputs in real-world diagnostic settings.

Table 2 Multimodal classification performance using different input features

ML classifier	Clinical Features		Peak current changes		All 9 features		Selected 3 features	
	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC
Decision Tree	57.01 (0.55 – 0.59)	0.56 (0.52 – 0.61)	81.30 (0.77 – 0.85)	0.84 (0.83 – 0.86)	89.72 (0.84 – 0.93)	0.85 (0.82 – 0.88)	90.65 (0.89 – 0.91)	0.86 (0.83 – 0.88)
k-Nearest Neighbors	62.61 (0.59 – 0.66)	0.65 (0.63 – 0.69)	79.43 (0.77 – 0.82)	0.85 (0.83 – 0.86)	81.31 (0.77 – 0.84)	0.89 (0.88 – 0.89)	84.11 (0.83 – 0.88)	0.87 (0.85 – 0.90)
AdaBoost	57.94 (0.57 – 0.60)	0.64 (0.61 – 0.67)	81.30 (0.79 – 0.83)	0.87 (0.86 – 0.88)	87.85 (0.85 – 0.89)	0.88 (0.87 – 0.90)	90.65 (0.89 – 0.91)	0.88 (0.84 – 0.92)
Naïve Bayes	55.14 (0.52 – 0.55)	0.58 (0.56 – 0.60)	83.48 (0.79 – 0.87)	0.89 (0.85 – 0.93)	78.50 (0.75 – 0.81)	0.87 (0.83 – 0.91)	82.24 (0.79 – 0.86)	0.83 (0.81 – 0.86)
Random Forest	44.58 (0.42 – 0.46)	0.59 (0.55 – 0.63)	76.63 (0.74 – 0.79)	0.83 (0.81 – 0.85)	85.05 (0.80 – 0.90)	0.90 (0.87 – 0.92)	85.98 (0.80 – 0.89)	0.90 (0.87 – 0.93)
Neural network	58.87 (0.56 – 0.60)	0.63 (0.62 – 0.64)	80.38 (0.76 – 0.84)	0.84 (0.81 – 0.87)	87.85 (0.83 – 0.90)	0.90 (0.87 – 0.92)	89.72 (0.84 – 0.93)	0.84 (0.82 – 0.85)

The six ML classifiers, including decision tree, k-nearest neighbors, AdaBoost, naïve Bayes, random forest and neural network models, were compared. The evaluation metrics for assessing the performance of each ML model was described in Table 3. Almost all the models exhibited high accuracy for OV classification. The decision tree models exhibited excellent performance, with the highest accuracy of 90.65% (95% CI 0.89 – 0.91), an F1 score of 91% (95% CI 0.90 – 0.92), and a cross-validation score of 88.21% (95% CI 0.87 – 0.89) for predicting OV infection. The AdaBoost and neural network models also exhibited high accuracies of 90.65% (95% CI 0.89 – 0.91) and 89.72% (95% CI 0.84 – 0.93), respectively. The performance of all ML models and the confusion matrix of each ML model were illustrated (Figure 7 a-f). The AdaBoost model was observed to have a greater AUC (0.88) and specificity (0.86) than the decision tree model (0.86, 0.83) but had a lower F1 score and cross-validation score. We also created receiver operating characteristic (ROC) curves and calculated the area under the ROC curves (AUC) to assess the performance of the classification methods using the set of features and hyperparameters that produced the highest accuracy scores (see Figure 7g).

**(a) Decision Tree****(b) k-Nearest Neighbors****(c) AdaBoost****(d) Naïve Bayes****(e) Random Forest****(f) Neural Networks**



(g) ROC-AUC analysis

Figure 7 Confusion matrices showing the binary classification of OV infection by (a) decision tree, (b) k-nearest neighbors, (c) AdaBoost, (d) naïve bayes, (e) random forest, (f) neural network and (g) ROC curve for different ML classifiers

Table 3 Evaluation metrics and 95% confidence interval (CI) of the ML classifiers using selected features (Peak current changes, age, urine pH)

Model	Accuracy (%) (95 % CI)	F1 Score (95 % CI)	Recall (95 % CI)	Specificity (95 % CI)	AUC (95 % CI)	Cross-validation Scores (%) (95 % CI)
Decision Tree	90.65 (0.89 – 0.91)	0.91 (0.90 – 0.92)	0.96 (0.90 – 0.98)	0.83 (0.80 – 0.86)	0.86 (0.83 -0.88)	88.21 (0.87 – 0.89)
k-Nearest Neighbors	84.11 (0.83 – 0.88)	0.84 (0.74 – 0.93)	0.86 (0.84 – 0.88)	0.81 (0.78 – 0.86)	0.87 (0.85 – 0.90)	83.95 (0.79 – 0.86)
AdaBoost	90.65 (0.89 – 0.91)	0.90 (0.88 – 0.92)	0.94 (0.92 – 0.96)	0.86 (0.84 -0.89)	0.88 (0.84 – 0.92)	85.60 (0.85 – 0.89)
Naïve Bayes	82.24 (0.79 – 0.86)	0.82 (0.81 -0.84)	0.89 (0.88 – 0.93)	0.71 (0.69 – 0.74)	0.83 (0.81 – 0.86)	82.31 (0.79 – 0.86)
Random Forest	85.98 (0.80 – 0.89)	0.85 (0.80 – 0.88)	0.87 (0.84 – 0.90)	0.85 (0.81 – 0.89)	0.90 (0.87 – 0.93)	86.08 (0.82 – 0.90)
Neural Network	89.72 (0.84 – 0.93)	0.89 (0.86 – 0.91)	0.89 (0.87 – 0.90)	0.91 (0.88 – 0.94)	0.84 (0.82 – 0.85)	87.50 (0.86 – 0.89)

3.5. Synthetic Minority Over-sampling Technique

Synthetic Minority Over-sampling Technique (SMOTE) algorithm is an extended algorithm for imbalanced data proposed by Chawla²⁵. SMOTE generates new synthetic instances by interpolating

between a data point and one of its nearest neighbors from the same class, rather than duplicating existing samples. In our study, SMOTE was applied using the imblearn library to address class imbalance in the training data. This technique helps to expand the decision boundary of the minority classes in feature space. In our case, SMOTE adjusted the training set so that the number of positive and negative samples was equal. We verified the resampled dataset to ensure this balance was achieved. The oversampling process was configured with a random state of 42 to ensure reproducibility and an auto sampling strategy, which automatically increases the number of samples in each minority class to match the majority class (class 0). The k-neighbors parameter was set to 5, meaning that for each minority class sample, five nearest neighbors were considered during the generation of synthetic samples. We evaluated the performance of all machine learning models with and without SMOTE (see Table S1). The decision tree model, along with the other classifiers, showed no significant improvement in performance metrics following resampling. This indicates that the mild imbalance in our dataset did not adversely affect model performance.

3.6. Statistical test

For the statistical comparison of model performance, the Friedman non-parametric test was employed first to reject the null hypothesis, and then proceeded with a post-hoc analysis based on the Wilcoxon-Holm method as suggested by Ismail et al., 2019²⁶. The Friedman test utilizes a multi-model approach, simultaneously evaluating all models to provide a broader perspective on their relative performance. Across all metrics, a statistically significant difference ($p < 0.001$) was achieved. The Wilcoxon-Holm post hoc test with Critical Differences (CD) was conducted to have an overall statistical comparison, ensuring more robust results, addressing potential violations of normality assumptions, small sample sizes, and the influence of outliers. In the critical difference (CD) diagram for accuracy (Figure 8), three distinct performance groups of ML models can be observed. The top-performing groups consist of AdaBoost, decision tree, and neural network (positioned on the right side of the scale), and they are not significantly different from each other. However, the bottom group includes random forest, k-nearest neighbors, and Naïve Bayes, indicating the worst performance. The overall performance of AdaBoost, decision tree, and artificial neural network are significantly better than random forest, k-nearest neighbors, and Naïve Bayes ($p < 0.01$).

Paired t-tests were further conducted to compare the best-performing classifiers. The results showed no statistically significant differences in accuracy: Decision Tree vs. AdaBoost ($p = 0.888$), Decision Tree vs. ANN ($p = 0.081$), and AdaBoost vs. ANN ($p = 0.108$). These findings support the conclusion that, despite achieving the highest performance metrics, the differences among these

models are not statistically significant. This aligns with the CD diagram, which also identified all three as the best-performing models.

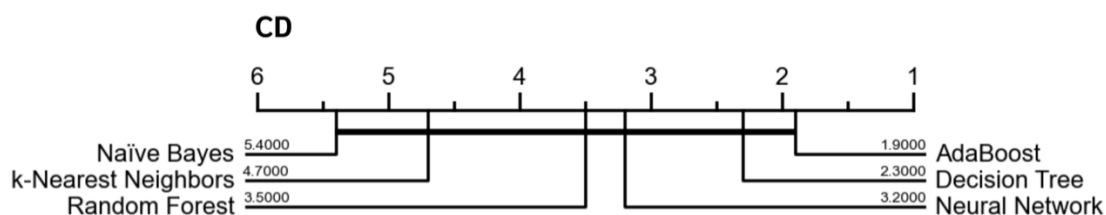


Figure 8 Critical Difference diagram based on the Wilcoxon-Holm test to detect pairwise significance between the performance achieved by the considered classifier (accuracy).

4. Discussion

Many studies have reported promising results in disease prediction using data-driven ML models. Intelligent electrochemical biosensors for pathogen detection, biomarker detection, and other medical applications have been constructed. Electrochemical biosensors can be classified into voltammetric, impedimetric and amperometric biosensors. Voltammetric sensors can detect biomolecules based on direct interaction with the electrical interface through which analyte information is obtained by varying a potential and then measuring the resulting current. Many forms of voltammetry exist based on potential variation, such as polarography (DC voltage)²⁷, linear sweep, differential staircase, normal pulse, reverse pulse, and differential pulse, among others. Dean SN et al. (2019) developed machine learning and deep learning models to classify chemicals from cyclic square wave voltammetry data. Electrochemical impedance spectroscopy (EIS) measurements are more complicated and require circuit fitting for data processing. Recently, Wang et al. (2023) developed a machine learning-assisted electrochemical impedance sensor for the classification of three types of bacteria. They used six EIS parameters as inputs and trained the model with SVM, DT, random forest, naïve Bayes, and AdaBoost classifiers²⁸. Ali et al. (2018) also applied an ANN, linear discrimination analysis (LDA) and the maximum likelihood method to classify *Escherichia coli* and *Salmonella typhimurium* strains using impedance signals¹⁵. In our study, we used SWV to measure OV antigen in clinical samples and extracted peak current changes as features for OV infection classification. Urine OV antigen detection was previously reported to be more sensitive than fecal examination for the diagnosis of opisthorchiasis (i.e., 17–44% of individuals with no observed fecal eggs had positive antigen tests)^{29,30}. The SWV peak current changes were found to be correlated with OV antigen concentrations and were ultrasensitive for OV infection detection in urine samples.

Despite the development of promising ML models integrated with electrochemical biosensors, their practical use on clinical samples remains challenging. Most studies have been conducted in spiked samples rather than in real clinical settings. Moreover, the data used to train models are often retrospectively collected and preselected. The scarcity of prospective clinical trials and validation on available datasets rather than on actual patients in the clinical setting has led to limited confidence in the validity of AI for real-world applications³¹. We tested urine samples collected from endemic areas of opisthorchiasis, ensuring an actual clinical setting for the development of the ML model.

Various machine learning approaches have been applied to identify OV infection using clinical features and sensor data. Logistic regression (LR) was reported to outperform other ML classifiers in terms of predictive tasks in several previous studies^{32–34}. Herein, we demonstrated that the decision tree, AdaBoost and neural network methods achieved greater accuracy than the LR method. Moreover, the ML approach for OV detection enhanced the accuracy of the electrochemical immunosensor from 84% to 91%, signifying the potential utility of the ML model.

The permutation-based importance analysis supports the biological and clinical relevance of the selected features. The dominant influence of peak current changes confirms the biosensor's efficacy in detecting OV-related biomarkers in urine. The decision-making process in the model follows a hierarchical weighting strategy, where dominant features drive primary decisions, while less important features refine or validate the classification. This layered decision pathway mirrors clinical reasoning, where strong diagnostic signals (e.g., a highly elevated biomarker) are supported by demographic context (e.g., patient age) and auxiliary findings. These findings have important implications for the real-world implementation of our diagnostic method in resource-limited settings. The overwhelming importance of the sensor-based measurement suggests that even in cases where complete patient data might be unavailable, the electrochemical readings alone could provide substantial diagnostic value. However, the meaningful contribution of age and, to a lesser extent, specific gravity and pH, indicates that incorporating these readily available parameters can enhance diagnostic accuracy. Overall, the feature importance results provide transparency into the model's decision-making process and reinforce confidence in its applicability for field-deployable OV infection screening. Future work will include broader clinical features to further enhance model robustness.

In northeastern Thailand, where the samples were collected, raw fish consumption is common, resulting in the highest prevalence rates of OV infection and OV infection-associated CCA worldwide³⁵. Hence, a critical solution to address the burden of opisthorchiasis in Thailand requires not only efficient diagnosis and treatment methods but also multisectoral and multidisciplinary involvement³⁶. Early diagnosis of OV infection plays a major role in first-line defense strategies. However, due

to limited numbers of health care workers and technicians, mass screening in endemic areas may be challenging, potentially resulting in delayed diagnoses. Therefore, the proposed ML model is intended to assist health care practitioners (HCPs) in decision-making processes, especially in uncertain situations. In our study, samples were collected from diverse demographics, including both urban and rural areas within endemic regions, to enhance model generalizability. Training with data from multiple endemic areas improved the model's applicability beyond localized settings. However, the lack of data from non-endemic areas and other populations is limited.

The formalin-ethyl acetate concentration technique (FECT) is considered the gold standard for detecting *Opisthorchis viverrini* infection, particularly in cases of moderate to heavy infection. However, FECT has notable limitations such as a lack of sensitivity in detecting light infections, inability to detect in case of bile-duct obstruction, low specificity in areas where morphologically similar flukes (MIFs) are endemic (such as the Mekong Basin), and the requirement of extensive sample collection³⁷. Additionally, FECT detects infection only at the egg stage, which typically appears four weeks post-infection, thus limiting its utility for early detection³⁸. Molecular techniques provide high specificity for OV detection, but the sensitivity can be variable depending on the number of eggs in the faeces. Moreover, this method is costly, requires a specialized technician, and projects the risk of false negatives due to PCR inhibitors in feces³⁹. In contrast, antigen detection methods, such as ELISA, enable earlier diagnosis and intervention, which is essential for reducing the risk of severe complications, including cholangiocarcinoma. A comparison with PCR or FECT would provide additional context for predictive capabilities. However, these tests were not concurrently performed in this study due to practical constraints, including cost and sample handling limitations. Using ELISA as a reference method for OV classification aligned the most with our study's objective of enabling timely detection and intervention. However, the use of ELISA as the gold standard for labeling may introduce sampling and label bias due to its limited sensitivity and specificity. Misclassifications, particularly near detection thresholds or weak antigen expression, can lead to systematic labeling errors that affect model training and evaluation. If the training set is biased toward samples that ELISA classifies more easily, this may reduce generalizability. Therefore, the reliability of ELISA-based labels should be critically assessed, and complementary validation methods considered where feasible.

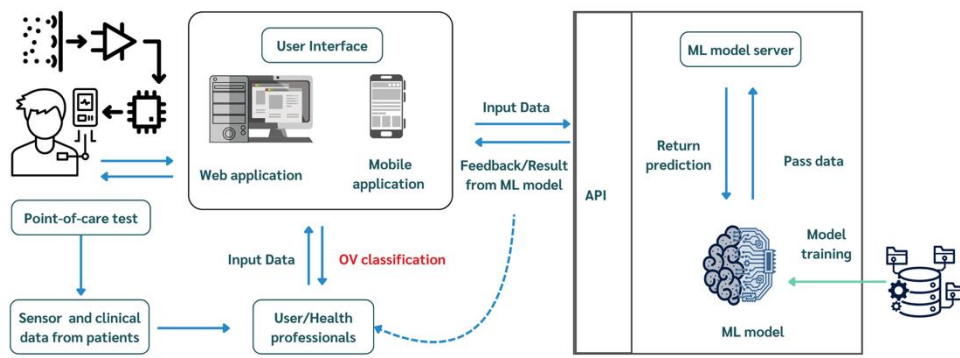


Figure 9 The proposed architecture of machine learning model deployment

4.1. Real-world deployment strategy

During recent years, research on the application of machine learning models in the healthcare sector has grown tremendously. It aims to develop and improve disease prediction, diagnosis, prognosis, treatment, and personalized medicine⁴⁰. Establishing a time-saving and accurate diagnostic method is crucial in the surveillance, prevention, and control of parasitic diseases. Several implementation strategies are proposed to facilitate the real-world deployment of the proposed ML model for OV infection classification. A modular machine learning (ML) deployment architecture for disease classification tasks will be implemented to facilitate real-time clinical decision support. The system architecture, illustrated in Figure 9, consists of five primary components: data acquisition, user interface, API integration, ML inference server, and model training pipeline. The user can input the clinical and sensor data manually or data will be collected directly from point-of-care devices into the application. This interface ensures usability and accessibility in clinical environments. ML server will apply ML algorithm locally or through cloud computing and return disease classification outputs via a secure API. The web application will provide the infection status to the user in a laboratory setting, assisting in decision-making. The public health professionals and non-expert users in rural areas will be empowered by the accessible mobile application. However, several challenges may be interfaced, including the requirement for standardization and improving the generalizability of ML models across different populations. Before widespread implementation, concerns around data protection, internet connectivity, and compliance with regulatory standards must be addressed. Regardless of these drawbacks, the proposed system holds significant potential tools to assist healthcare providers in decision making, particularly in resource-limited or high-prevalence settings.

5. Limitations

The major limitations of the study are the small number of features associated with the disease and the lack of external validation. More features associated with OV infection, such as laboratory results, epidemiological data and socioeconomic status, will be further applied to train the models. Furthermore, the size of our dataset limited the model's predictive capabilities, which can be mitigated by increasing the number of patients in each class for further validation procedures. At present, external validation is not feasible due to the limited availability of independent datasets and the fact that electrochemical biosensing for *Opisthorchis viverrini* infection is currently confined to our testing center. As such, the peak current data were limited to other populations. While external validation on an independent dataset would further strengthen the results, it was not performed in this study as it is a single-center study. Instead, five-fold cross-validation was applied to provide a more reliable and unbiased estimate of model performance than a single hold-out split, with classifier parameters carefully optimized for each model. Furthermore, our study was based on field testing rather than retrospective clinical data, which made it challenging to include external samples from outside the endemic areas. Future multicenter collaborations will be conducted to enable broader data collection and enhance the generalizability and reliability of our model. The sensor's variability and instability over time can potentially affect model performance. To ensure consistent electrochemical signals, regular calibration and the use of internal standards are essential. Although our current dataset did not exhibit significant evidence of performance drift, addressing this issue is critical for long-term or real-world applications. Future work may include the implementation of drift detection and correction algorithms, as well as periodic recalibration protocols, to sustain the reliability and robustness of the model over time.

6. Conclusion

In conclusion, this is the first study to classify OV infection using an ML approach based on electrochemical sensor data and patient information. The interpretability of model is demonstrated by the clear dominance of peak current changes in the decision process, supported by demographic and urinary parameters at specific thresholds. The application of this ML model allows for the interpretation of complex raw electrochemical signals to determine OV infection status. This means that medical personnel or patients without prior knowledge of electrochemical sensors can easily understand the disease condition. The decision tree classifier outperformed all other models, achieving an accuracy of 90.65%, a F1 score of 91%, recall of 96%, specificity of 83%, and ROC-AUC of 0.86. Similarly, the AdaBoost model demonstrated excellent performance, with a higher specificity of 86% and a superior ROC-AUC of 0.88 compared to the decision tree. The statistical analyses evaluated the significant difference in the model's performance. The ML model can be

trained to continuously learn and adapt to new data, improving its predictive capabilities over time. Moreover, the proposed ML model is not limited to OV infection and can be applied for the classification of other parasitic infections. A mobile or web application for real-world deployment was proposed in this study. In the future, this study can be used as a prototype to develop a health care strategy for OV-infected patients. The proposed ML model holds promise for enhancing disease surveillance and control strategies and assisting HCPs in decision-making processes in endemic regions.

Author Contributions: Conceptualization, N.N.S.A, R.P, and J.D; Formal analysis, N.N.S.A; Investigation, N.N.S.A; Methodology, N.N.S.A, R.P ; Supervision, P.M, R.P, and J.D; Writing – original draft, N.N.S.A ; Writing – review and editing, N.N.S.A, S.D, P.T, A.T, P.M, P.S, R.P, and J.D. Resources – P.S.

Funding: This research was supported by NSRF under the Basic Research Fund of Khon Kaen University through the Cholangiocarcinoma Research Institute: CARI-BRF64-47 and the Fundamental Fund of Khon Kaen University and the National Science, Research, and Innovation Fund (FF2566).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Centre for Ethics in Human Research, Khon Kaen University (HE664025).

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets used and/or analysed during the current study belong to Cholangiocarcinoma Research Institute, Khon Kaen University, Thailand. The datasets used to test the model are available in the supplementary.

Acknowledgments: The author would like to express gratitude for a Khon Kaen University ASEAN-GMS scholarship and sample resources from Cholangiocarcinoma Research Institute.

601 **Conflicts of Interest:** The authors declare no conflicts of interest.

602

603

ARTICLE IN PRESS

604 References:

- 605 1. Ahmad, A. *et al.* Intracellular synthesis of gold nanoparticles by a novel alkalotolerant
 606 actinomycete, *Rhodococcus* species. *Nanotechnology* **14**, 824–828 (2003).
- 607 2. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. *Int. J.*
 608 *Cancer* **118**, 3030–3044 (2006).
- 609 3. Lun, Z.-R. *et al.* Clonorchiasis: a key foodborne zoonosis in China. *Lancet Infect. Dis.* **5**, 31–41
 610 (2005).
- 611 4. Sripa, B. *et al.* Opisthorchiasis and Opisthorchis-associated cholangiocarcinoma in Thailand and
 612 Laos. *Acta Trop.* **120 Suppl 1**, S158-168 (2011).
- 613 5. Sithithaworn, P., Yongvanit, P., Duenngai, K., Kiatsopit, N. & Pairojkul, C. Roles of liver fluke
 614 infection as risk factor for cholangiocarcinoma. *J. Hepato-Biliary-Pancreat. Sci.* **21**, 301–308
 615 (2014).
- 616 6. Qian, M.-B. *et al.* Accuracy of the Kato-Katz method and formalin-ether concentration technique
 617 for the diagnosis of *Clonorchis sinensis*, and implication for assessing drug efficacy. *Parasit.*
 618 *Vectors* **6**, 314 (2013).
- 619 7. Johansen, M. V., Sithithaworn, P., Bergquist, R. & Utzinger, J. Towards improved diagnosis of
 620 zoonotic trematode infections in Southeast Asia. *Adv. Parasitol.* **73**, 171–195 (2010).
- 621 8. Wongratanacheewin, S., Sermswan, R. W. & Sirisinha, S. Immunology and molecular biology of
 622 *Opisthorchis viverrini* infection. *Acta Trop.* **88**, 195–207 (2003).
- 623 9. Haswell-Elkins, M. R. *et al.* Immune responsiveness and parasite-specific antibody levels in
 624 human hepatobiliary disease associated with *Opisthorchis viverrini* infection. *Clin. Exp.*
 625 *Immunol.* **84**, 213–218 (1991).
- 626 10. Bettazzi, F., Marrazza, G., Minunni, M., Palchetti, I. & Scarano, S. Chapter One - Biosensors
 627 and Related Bioanalytical Tools. in *Comprehensive Analytical Chemistry* (eds Palchetti, I.,
 628 Hansen, P.-D. & Barceló, D.) vol. 77 1–33 (Elsevier, 2017).

- 629 11. Menon, S., Mathew, M. R., Sam, S., Keerthi, K. & Kumar, K. G. Recent advances and
630 challenges in electrochemical biosensors for emerging and re-emerging infectious diseases. *J.*
631 *Electroanal. Chem.* **878**, 114596 (2020).
- 632 12. Aye, N. N. S. *et al.* Functionalized graphene oxide–antibody conjugate-based electrochemical
633 immunosensors to detect *Opisthorchis viverrini* antigen in urine. *Mater. Adv.* (2024)
634 doi:10.1039/D3MA01075A.
- 635 13. Cui, F., Yue, Y., Zhang, Y., Zhang, Z. & Zhou, H. S. Advancing Biosensors with Machine
636 Learning. *ACS Sens.* (2020) doi:10.1021/acssensors.0c01424.
- 637 14. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning
638 in materials informatics: recent applications and prospects. *Npj Comput. Mater.* **3**, 1–13 (2017).
- 639 15. Ali, S. *et al.* Disposable all-printed electronic biosensor for instantaneous detection and
640 classification of pathogens. *Sci. Rep.* **8**, 5920 (2018).
- 641 16. Zeng, Z. *et al.* Nonintrusive Monitoring of Mental Fatigue Status Using Epidermal Electronic
642 Systems and Machine-Learning Algorithms. *ACS Sens.* **5**, 1305–1313 (2020).
- 643 17. Uwaya, G. E., Sagrado, S. & Bisetty, K. Smart electrochemical sensing of xylitol using a
644 combined machine learning and simulation approach. *Talanta Open* **6**, 100144 (2022).
- 645 18. Baghdadi, N. A. *et al.* Advanced machine learning techniques for cardiovascular disease early
646 detection and diagnosis. *J. Big Data* **10**, 144 (2023).
- 647 19. Hastie, T., Rosset, S., Zhu, J. & Zou, H. Multi-class AdaBoost. *Stat. Interface* **2**, 349–360
648 (2009).
- 649 20. Kumari, S., Kumar, D. & Mittal, M. An ensemble approach for classification and prediction of
650 diabetes mellitus using soft voting classifier. *Int. J. Cogn. Comput. Eng.* **2**, 40–46 (2021).
- 651 21. Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning.
652 *Sci. Rep.* **14**, 6086 (2024).

- 653 22. Srithai, C., Chuangchaiya, S., Jaichuang, S. & Idris, Z. M. Prevalence of *Opisthorchis viverrini*
654 and Its Associated Risk Factors in the Phon Sawan District of Nakhon Phanom Province,
655 Thailand. *Iran. J. Parasitol.* **16**, 474–482 (2021).
- 656 23. Thaewnongiew, K. *et al.* Prevalence and risk factors for *Opisthorchis viverrini* infections in
657 upper Northeast Thailand. *Asian Pac. J. Cancer Prev. APJCP* **15**, 6609–6612 (2014).
- 658 24. Nirmalraj, S. *et al.* Permutation feature importance-based fusion techniques for diabetes
659 prediction. *Soft Comput.* (2023) doi:10.1007/s00500-023-08041-y.
- 660 25. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority
661 Over-sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
- 662 26. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. & Muller, P.-A. Deep learning for
663 time series classification: a review. *Data Min. Knowl. Discov.* **33**, 917–963 (2019).
- 664 27. Heyrovský, J. The development of polarographic analysis. *Analyst* **81**, 189–192 (1956).
- 665 28. Wang, C. *et al.* Machine learning-assisted cell-imprinted electrochemical impedance sensor for
666 qualitative and quantitative analysis of three bacteria. *Sens. Actuators B Chem.* **384**, 133672
667 (2023).
- 668 29. Worasith, C. *et al.* Comparing the performance of urine and copro-antigen detection in
669 evaluating *Opisthorchis viverrini* infection in communities with different transmission levels in
670 Northeast Thailand. *PLoS Negl. Trop. Dis.* **13**, e0007186 (2019).
- 671 30. Worasith, C. *et al.* Effects of day-to-day variation of *Opisthorchis viverrini* antigen in urine on
672 the accuracy of diagnosing opisthorchiasis in Northeast Thailand. *PLOS ONE* **17**, e0271553
673 (2022).
- 674 31. Thomas, L. B., Mastorides, S. M., Viswanadhan, N. A., Jakey, C. E. & Borkowski, A. A.
675 Artificial Intelligence: Review of Current and Future Applications in Medicine. *Fed. Pract.*
676 *Health Care Prof. VA DoD PHS* **38**, 527–538 (2021).

- 677 32. Christodoulou, E. *et al.* A systematic review shows no performance benefit of machine learning
678 over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **110**, 12–22 (2019).
- 679 33. van der Ploeg, T., Austin, P. C. & Steyerberg, E. W. Modern modelling techniques are data
680 hungry: a simulation study for predicting dichotomous endpoints. *BMC Med. Res. Methodol.*
681 **14**, 137 (2014).
- 682 34. Jiang, Y. *et al.* Cardiovascular Disease Prediction by Machine Learning Algorithms Based
683 on Cytokines in Kazakhs of China. *Clin. Epidemiol.* **13**, 417–428 (2021).
- 684 35. Alsaleh, M. *et al.* Characterisation of the Urinary Metabolic Profile of Liver Fluke-Associated
685 Cholangiocarcinoma. *J. Clin. Exp. Hepatol.* **9**, 657–675 (2019).
- 686 36. Sripan, B. *et al.* Toward integrated opisthorchiasis control in northeast Thailand: The Lawa
687 project. *Acta Trop.* **141**, 361–367 (2015).
- 688 37. Worasith, C. *et al.* Advances in the Diagnosis of Human Opisthorchiasis: Development of
689 *Opisthorchis viverrini* Antigen Detection in Urine. *PLoS Negl. Trop. Dis.* **9**, e0004157 (2015).
- 690 38. Maitongngam, K., Tipayamongkhogul, M. & Kosaisavee, V. Diagnostic accuracy of
691 *Opisthorchis viverrini* antigen methods for human opisthorchiasis: Systematic review and Meta-
692 analysis. *Thai J. Public Health* **53**, 465–481 (2023).
- 693 39. Khuntikeo, N. *et al.* Current Perspectives on Opisthorchiasis Control and Cholangiocarcinoma
694 Detection in Southeast Asia. *Front. Med.* **5**, (2018).
- 695 40. Panch, T., Szolovits, P. & Atun, R. Artificial intelligence, machine learning and health systems.
696 *J. Glob. Health* **8**, 020303 (2018).