PAPER • OPEN ACCESS

Machine learning to predict the antimicrobial activity of cold atmospheric plasma-activated liquids

View the article online for updates and enhancements.

MACHINE LEARNING
Science and Technology

Machine learning to predict the antimicrobial activity of cold atmospheric plasma-activated liquids

Mehmet Akif Özdemir[1,2,3] , Gizem Dilara Özdemir[1,2,3] , Merve Gül[2], Onan Güren[1] and Utku Kürşat Ercan[1,*]

[1] Department of Biomedical Engineering, Izmir Katip Celebi University, Cigli 35620, Izmir, Turkey
[2] Department of Biomedical Technologies, Graduate School of Natural and Applied Sciences, Izmir Katip Celebi University, Cigli 35620, Izmir, Turkey
[3] These authors contributed equally to this work.
* Author to whom any correspondence should be addressed.

E-mail: utkuk.ercan@ikcu.edu.tr

Abstract

Plasma is defined as the fourth state of matter, and non-thermal plasma can be produced at atmospheric pressure under a high electrical field. The strong and broad-spectrum antimicrobial effect of plasma-activated liquids (PALs) is now well known. The antimicrobial effects of PALs depend on many different variables, which complicates the comparison of different studies and determining the most dominant parameters for the antimicrobial effect. The proven applicability of machine learning (ML) in the medical field is encouraging for its application in the field of plasma medicine as well. Thus, ML applications on PALs could present a new perspective to better understand the influences of various parameters on their antimicrobial effects. In this paper, comparative supervised ML models are presented by using previously obtained data to predict the *in vitro* antimicrobial activity of PALs. A comprehensive literature search was performed, and 12 distinct features related to PAL-microorganism interactions were collected from 33 relevant articles to automatically predict the antimicrobial activity of PALs. After the required normalization, feature encoding, and resampling steps, two supervised ML methods, namely classification and regression, are applied to the data to obtain microbial inactivation (MI) predictions. For classification, MI is labeled in four categories, and for regression, MI is used as a continuous variable. Sixteen different classifiers and 14 regressors are implemented to predict the MI value. Two different robust cross-validation strategies are conducted for classification and regression models to evaluate the proposed method: repeated stratified *k*-fold cross-validation and *k*-fold cross-validation, respectively. We also investigate the effect of different features on models. The results demonstrated that the hyperparameter-optimized Random Forest Classifier (oRFC) and Random Forest Regressor (oRFR) provided superior performance compared to other models for classification and regression. Finally, the best test accuracy of 82.68% for oRFC and R^2 of 0.75 for the oRFR are obtained. Furthermore, the determined most important features of predictive models are in line with the outcomes of PALs reported in the literature. An ML framework can accurately predict the antimicrobial activity of PALs without the need for any experimental studies. To the best of our knowledge, this is the first study that investigates the antimicrobial efficacy of PALs with ML. Furthermore, ML techniques could contribute to a better understanding of plasma parameters that have a dominant role in the desired antimicrobial effect. Moreover, such findings may contribute to the definition of a plasma dose in the future.

1. Introduction

The term 'plasma' was first introduced by Irving Langmuir in 1928 and refers to the fourth state of matter, which can be generated under an electric field and is a partially ionized gas composed of photons, free electrons, ions, free radicals, reactive oxygen species (ROS), and reactive nitrogen species (RNS) [1]. Plasma can be produced at atmospheric pressure or under a vacuum and can be classified into two categories based on the thermal equilibrium between the electrons and heavy particles: thermal or hot plasma and non-thermal or cold plasma. Cold atmospheric plasma (CAP) could be generated at atmospheric pressure and room temperature under an externally applied electrical field. CAP applications can be summarized in three main categories; surface modifications, therapeutic applications, and biological decontamination [2]. Antimicrobial activity, blood clotting, tooth whitening, wound healing, and anticancer efficacy are the main biomedical applications of CAP that are reported in the literature [3]. Due to its strong antimicrobial activity against a wide spectrum of microorganisms, including antibiotic-resistant organisms, CAP is an emerging technology that is undergoing intensive research. One of the treatment methods of CAP, direct CAP treatment, has been used in sterilization and disinfection applications with the goal of microbial inactivation (MI) in a number of studies. Beside direct CAP treatment, liquids treated with plasma have been shown in the literature to have similar effects to direct CAP treatment by undergoing chemical modifications [4]. Reactive plasma-generated species such as ROS, RNS, free radicals, and electrons are transferred to the liquid by CAP treatment. Those reactive species may also lead to the formation of new species via interaction with the treated liquid [5]. The liquid that is activated by CAP treatment is called plasma-activated liquids (PALs).

The plasma-generated reactive species were linked to the strong and broad-spectrum antimicrobial activity of PALs [6]. Xiang *et al* conducted a comprehensive study on the antimicrobial efficacy of plasma-activated water (PAW) against a range of microbial strains [7]. In another study [8], different liquids, including deionized water (DIW), N-acetyl-cysteine, and phosphate-buffered saline (PBS), were activated via CAP treatment, and these PALs exhibited broad-spectrum antimicrobial activity. Furthermore, this research reveals that PALs have a long-lasting antimicrobial effect. Schmidt *et al* investigated the antimicrobial activities, pH value, and conductivity of several solutions (tap water, PBS, and physiological saline) that are activated with CAP [9]. Studies in the field of MI by PALs are expanding and becoming more prevalent as a consequence of the advantages they provide. However, due to the differences between various plasma generation systems, including electrical plasma parameters, electrode geometry and type, liquid type, treatment volume, etc comparison of the antimicrobial efficacies of PALs from different research laboratories can be difficult [10]. Furthermore, no well-established method for measuring, comparing, and predicting the antimicrobial effectiveness of PALs generated by different plasma systems has yet to be developed. Moreover, the inhibition activity of the produced PALs may differ depending on the liquids and microorganism strains. Thus, a method that would help to predict the antimicrobial strength and efficacy of a PAL is needed to achieve the desired antimicrobial activity by PALs [11].

Artificial intelligence (AI), a branch of computer science that is utilized in a wide range of disciplines and is receiving a lot of attention these days, is frequently employed in solving complicated problems, making decisions, and recognizing patterns [12]. Machine learning (ML), one of AI's sub-branches, is the ability of constructed algorithms to learn from incoming data. ML is a computational technique that automates the process of model construction by using algorithms to learn from input data and make predictions or decisions without explicit instructions. Furthermore, it should be noted that ML needs accurate knowledge about the in-depth comprehension of the underlying data and models. In fact, a thorough understanding of the data and the ability to interpret the results are crucial for the successful implementation and interpretation of ML models. ML makes predictions by extracting inferences from data using mathematical and statistical operations. ML algorithms are computational techniques used to analyze and identify patterns in data, and to construct predictive models based on that discriminative knowledge [13]. In scenarios where classical statistical methods are insufficient, ML provides an excellent option for analyzing a very large amount of data due to its adaptability. AI is widely used in many fields nowadays since these tools are simple and affordable [14]. Furthermore, investigations based on this foundation shed light on their approach, which is based on data inputs rather than physical test materials. They can also anticipate the impacts of materials whose effects have yet to be discovered using constructed models, contributing to prediction approaches without the need for complex optimization processes [12]. Antibiotic development [15] and antimicrobial resistance prediction for specific bacteria [16] have both benefited from ML approaches. In the medical and biomedical disciplines, ML models are commonly used. The proven applicability of ML algorithms in the medical field bodes well for their use in other fields, such as the prediction of antimicrobial activity with different agents. There are only a few studies that apply ML techniques to estimate antimicrobial activity. Shaban and Alkawareek sought to estimate the antibiofilm activity of antibiotics using three optimized ML models based on logistic regression (LR), decision tree (DT), and random forest (RF)

algorithms, using data manually obtained from the available literature for high-accuracy prediction of *in vitro* antibiofilm activity, where they achieved 67% $\pm$ 6.1% prediction accuracy for the LR model, 73% $\pm$ 5.8% for the DT model, and 74% $\pm$ 5% for the RF model [17]. Apart from that, there have been other studies on ML prediction involving nanoparticles and antimicrobial peptides. These studies collect data from the literature or databases and analyze them using ML algorithms to predict toxicity or antimicrobial activity [18, 19].

In the field of plasma medicine, determining the plasma treatment parameters, attaining the required antimicrobial effect, and optimizing plasma treatment to a specific standardization are critical challenges. To solve this problem, we propose an ML method to predict the antimicrobial activity of PALs against a wide spectrum of microorganisms. This tool uses CAP parameters, experimental setup parameters, and microorganism characteristics as inputs to predict antimicrobial efficacy. We compiled *in vitro* experimental data from published research and structured them into a comprehensive dataset. By reducing the number of trials and errors in laboratory studies, the current approach allows for the screening of PALs and the prediction of their ability to inactivate microorganisms, saving time and cost. To the best of our knowledge, no study has used ML algorithms to predict the antimicrobial efficacy of PALs. AI algorithms intended for use in the field of plasma medicine are expected to be pioneering studies in this field, paving the way for other researchers to use AI techniques in plasma medicine.

2. Methods

The visual framework of the proposed methodological structure is presented in figure 1. A detailed description of each stage is given in the following.

2.1. Data collection

A literature search was performed between December 2021 and March 2022 in the Scopus, ScienceDirect, PubMed, and Web of Science databases to find relevant articles on the antimicrobial activity of PALs which publication dates between 2008 and 2022. The keywords used in the literature search were determined as 'antimicrobial, antibacterial, inactivation, CAP, non-thermal plasma, cold plasma, PAW, PAL, plasma treated water, and plasma-treated liquid'. A detailed expression of the literature search is presented in table S1 in the supplementary data. In addition to using keywords to search databases, articles were acquired by conducting a manual literature search as well. A total of 307 relevant articles were obtained, and the titles and abstracts of each article were examined to determine the most appropriate publications.

The primary inclusion criteria for the studies are to determine the MI value as a log reduction by exposing PAL to microorganisms. The characteristics of the microorganisms and cold plasma specifications utilized in the experiments should also be indicated in the article. The article removal process was carried out, and a total of 33 papers were chosen. Afterward, the feature extraction stage was performed. Features such as CAP specifications, PAL characteristics, *in vitro* characteristics, and CAP treatment characteristics were manually extracted from the 33 articles [6, 8, 9, 20–49]. Any article which has missing value about the specified features was eliminated. As a result, a total of 762 observations were collected without any missing predictor values.

2.2. Predictors and outcome

In the field of plasma medicine and antimicrobial studies with PALs, plasma treatment type, gas type, discharge gap, PAL type, treatment volume, treatment time, microbial strain, initial microbial load, PAL/microorganism suspension volume ratio, contact time, post-storage time, and incubation temperature are the most widely used parameters, which were also used as predictors in the feature selection. Logarithmic values for MI were utilized for supervised ML models to relate these predictors to an outcome.

Before data preprocessing, 13 different features that were manually extracted from the relevant articles were modified. While the discharge gap, treatment volume, treatment time, initial microbial load, PAL/microorganism suspension volume ratio, exposure time, post-storage time, and incubation temperature features of the PALs were numeric data, the plasma type, gas type, liquid type, and microbial strain were nominal data among the features. The logarithm of the initial microbial load (n) was also taken because one of the inclusion criteria for an article was to define the MI as a logarithmic value. The input and output variables, their categories and types, as well as the total number of unique values for each variable, are shown in table 1. Also, the scatter plot of numerical predictors and outcome is presented in figure S1 in the supplementary data.

The MI value was also labeled and categorized to a certain standard for classification models. The MI value was determined based on the logarithm of the n of the bacterial solution because the n could vary from study to study. A computation approach was used to label the results as four different categorical data, as shown in table 2. Following the categorization of results, there are 211 observations in the 'Complete' class, 94 in the 'Strong' class, 169 in the 'Weak' class, and 288 in the 'None' class. The identified categorical classes
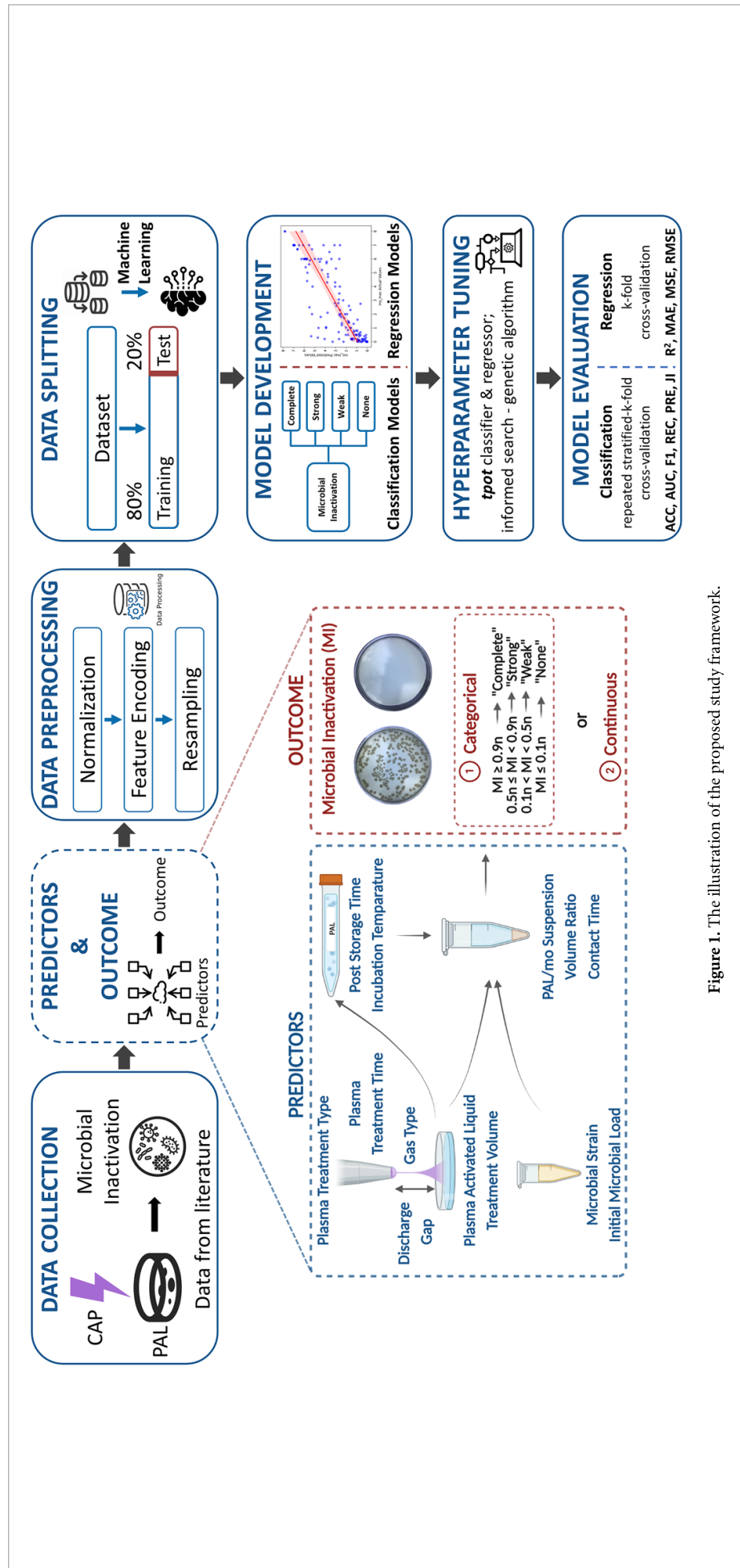
Figure 1. The illustration of the proposed study framework.

Table 1. Input and output variables, their categories, types, and the total numbers of unique values (A detailed expression of unique values of nominal variables is presented in Table S2 in supplementary data).

Category	Variables	Type	Unique value
CAP specifications	Plasma treatment type	Nominal	7
	Gas type	Nominal	8
	Discharge gap	Numeric	15
	Plasma treatment time	Numeric	21
PAL characteristics	Plasma activated liquid	Nominal	15
	Treatment volume	Numeric	14
In vitro characteristics	Microbial strain	Nominal	22
	Initial microbial load	Numeric	32
CAP treatment characteristics	PAL/mo suspension volume ratio	Numeric	12
	Contact time	Numeric	26
	Incubation temperature	Numeric	8
	Post storage time	Numeric	27
Output	Microbial inactivation	Numeric	126
		Categorical	4

Table 2. Categorization conditions for MI (microbial inactivation: outcome) and counts of class labels before applying SMOTE (n indicates the initial microbial load).

Condition	Output label	Counts
$\text{MI} \leqslant 0.1n$	None, N	288
$0.1n < \text{MI} < 0.5n$	Weak, W	169
$0.5n \leqslant \text{MI} < 0.9n$	Strong, S	94
$\text{MI} \geqslant 0.9n$	Complete, C	211

were used to express the MI value for the classification models. Regression models were also utilized to estimate the output as an exact value. Therefore, MI was used as a continuous variable in regression models.

2.3. Data preprocessing
The collected dataset for the creation of ML models was subjected to normalization, encoding, and resampling data preprocessing steps before the training in order to avoid some issues such as variability in unit sizes, the data imbalance effect, and data type inconsistency.

2.3.1. Normalization
The variability in units of measurement and magnitudes during model training may make optimization more challenging. In order to improve model performance, the normalization technique was used as a pre-processing step [50]. Data normalization is the process of converting the obtained data into distinct ranges. During the normalization step, Z-score, min-max scaler [51], max-absolute scaler, and robust scaler [52] were implemented. Different normalization strategies have been tested, taking into account that normalization may have an impact on the model's performance because the majority of the data is numerical. All normalization methods were implemented for both classification and regression models. The best results for classification and regression were determined empirically among different normalization methods. The best normalization technique was adopted for the continuation of the experimental studies. The initial numerical data distribution as well as the updated numerical data distributions that were produced through the utilization of the four distinct normalization approaches outlined above depicts in figure S2 in the supplementary data.

2.3.2. Feature encoding
The variables used in ML need to be converted into machine-readable data in order to train the model. The most popular method for achieving this transition among the several known techniques is one-hot encoding, which divides the column into multiple columns to transform the nominal data into numeric data. The nominal values are replaced by 1s and 0s, whether they have the specified feature or not [53]. Since the frequency of the variables was not taken into consideration, nominal values that were repeated numerous times were only counted once. Because some of the predictors in this study contain nominal data and have neither rank nor order, these features were transformed into numeric values with the one-hot encoding approach. In contrast, the categorical outcome, MI, was encoded gradually for use in classification models due to it contains a rank between labels, which have an increasing order from 'None' to 'Complete'. The

'Complete' class is superior to the other classes so it was encoded as '3', and the 'Strong', 'Weak', and 'None' classes were encoded as '2', '1', and '0', respectively.

2.3.3. Resampling

Classes are frequently uneven in studies using real data, such as the prevalence of diseases and experimental outcomes, and class imbalance is an important problem in studies using supervised ML [54]. Classifiers and regressors have less accurate predictions as a result of the large class difference. Correct classification and regression may be ensured by balancing the proportion of majority classes with minority classes [55]. There are a number of ML techniques, and numerous studies claim that balanced data can improve prediction performance. A type of oversampling technique called synthetic minority oversampling technique (SMOTE) was used for classification to overcome the imbalance issue in unbalanced and high-dimensional datasets. The SMOTE technique creates new minority class samples based on the original dataset that was generated at random from the closest neighbors joining the minority class data to expand the amount of data [56]. The label 'None' predominates the dataset with 288 observations, and there are four categories of MI that were assigned as outcomes. All other MI categories should therefore be equivalent to 288.

Synthetic minority over-sampling technique for regression (SMOTER) with Gaussian Noise (SMOGN) is another resampling method to eliminate the imbalance problem [57]. The SMOGN technique was used to resample the imbalanced dataset for use in the training of the regression models. The SMOTER and the introduction of Gaussian Noise are two oversampling techniques combined in the SMOGN approach. Based on the distance to the KNNs, SMOGN iterates through all unusual samples and selects between SMOTER's interpolation-based oversampling and Gaussian noise-based oversampling [58]. The data imbalance effect was eliminated by using the SMOTE and SMOGN techniques for classification and regression models, respectively, to resample the dataset into the classes and change the relative frequency of the other labels.

2.4. Data splitting

To evaluate the robustness of the trained ML model, the dataset should be split [59]. Hence, unknown test data which never used in the training step can be predicted fairly with the trained ML model. Splitting the dataset is the last step before developing the model and may be grouped as train-test split and validation splitting. Train-test splitting in ML is the process of dividing the data sample into two groups, a 'training set' and a 'test set', for the purposes of training and testing the model [53]. In this study, the data were randomly divided into two groups for classification and regression models: a training set including 80% of the data and a test set containing the remaining 20%. It's also crucial to validate the outputs of the ML models to assure their accuracy. In this study, two different robust validation techniques, namely repeated stratified k-fold cross-validation and k-fold cross-validation, were utilized for classification and regression models, respectively, by considering the training sets.

2.5. Model development

Studies that use a range of ML techniques to create a variety of models are documented in the literature. These methods can be split into two groups: supervised learning and unsupervised learning. Supervised learning is the process of developing algorithms that reliably identify data or forecast outcomes using labeled datasets [60]. Supervised learning has two subcategories: regression and classification. Both supervised ML techniques were used in this study to predict the categorical or continuous MI value. Models were developed using Python version 3.7.0 and *scikit-learn* version 1.0.2.

2.5.1. Classification models

In ML, the term 'classification' refers to a predictive modeling problem where a class label is predicted for a given set of inputs. The ML method links CAP parameters, PAL characteristics, and *in vitro* characteristics to the inhibition of bacteria and makes it possible to predict the MI efficacy of PALs. As candidates for our model, we tested a number of supervised classification algorithms to see which one could provide the most accurate prediction.

LR, Naïve Bayes classifier, k-nearest neighbor (KNN), DTs, RF, support vector machines (SVMs), and Boosting algorithms are some of the supervised learning methods. Understanding how a group of independent variables influence the result of the dependent variable is made easier with the help of LR [61]. A modest amount of training data is needed for the Naïve Bayes classifier in order to estimate the necessary parameters. When compared to other classifiers, they are reasonably fast [62]. A lazy learning technique called KNN saves every instance corresponding to the training set in n-dimensional space [63]. A DT has the benefit of being easy to comprehend and depict, and it also requires a very small amount of data preparation. The DTs drawback is that it can produce complex trees that might not categorize things well. Due to less over-fitting than DTs, the RF has the advantage of being more accurate [64]. The RF classifiers' only

drawback is that their implementation is fairly difficult [65]. SVM is memory-efficient and works incredibly well in high-dimensional domains since it only employs a portion of the training points in the decision function. The SVM's sole drawback is that the technique does not elicit probability estimates directly [66]. Boosting is an ensemble learning technique that turns weak learners into strong ones in order to improve the model's accuracy. It makes use of a variety of ML algorithms [67].

LR, linear discriminant analysis, Gaussian process classifiers, KNN, decision tree classifier, extra trees classifier, random forest classifier (RFC), Gaussian Naïve Bayes (GNB), Bernoulli Naïve Bayes (BNB), SVM, bagging classifier, extreme gradient boosting classifier (XGBC), AdaBoost classifier, histogram-based gradient boosting classifier (HGBC), and gradient boosting classifier algorithms were utilized as classifiers to predict categorical MI in the model development stage of the study.

2.5.2. Regression models
Regression is another subfield of supervised ML. It aims to create a model of the relationship between a certain number of features and a continuous target variable [68]. One of the most fundamental kinds of regression in ML is linear regression. A predictor variable and a dependent variable that are linearly related to each other compensate for the linear regression model. When there is a strong correlation between the independent variables, ridge regression (RR) is commonly used. Regularization and feature selection are both carried out through the least absolute shrinkage and selection operator (LASSO) regression. It limits the regression coefficient's maximum absolute value. With a small change, polynomial regression is identical to multiple linear regression. The value of the regression coefficients is determined via Bayesian regression using the Bayes theorem. Instead of locating the least squares, this regression method determines the posterior distribution of the features [69]. Furthermore, such algorithms DT, RF, boosting, and KNN can be adapted as regression models.

LASSO regression, RR, extreme gradient boosting regression (XGBR), LASSO-least-angle regression, k-neighbors regression, AdaBoost regression, extra trees regression (ETR), bagging regression, elastic net regression (ENR), linear support vector regression (LSVR), Bayesian ridge regression, multi-layer perceptron regression, random forest regression (RFR), and gradient boosting regression were used as regressors to determine which model could deliver the most accurate performance in the prediction of the exact value of MI.

2.6. Hyperparameter tuning
Hyperparameters are the variables that define the model architecture. The model's parameters, which are learned during training, cannot be manually set. Starting with random parameter values, a model modifies them throughout the training process. Hyperparameters, on the other hand, are variables that are chosen before the model is trained [70]. Hyperparameter values have the potential to increase or decrease model accuracy. Hyperparameter tuning is the process of searching for the ideal model architecture and its optimum parameters. It is a crucial phase in the model training process that allows the model to test various combinations of hyperparameters and make predictions using the optimal hyperparameter values [71]. For hyperparameter tuning, a variety of techniques are utilized, including grid search, random search, and informed search [72]. A genetic algorithm is a method for hyperparameter tuning that is based on an informed search and the real-world idea of genetics. The procedure begins by building a few models, selecting the best one, building other models that are similar to the best ones, and adding some randomness until the target accuracy is achieved. Grid and random search are both used in informed search and genetic algorithms. The *tpot* library [73] predicts the optimum hyperparameter values, and the evolutionary algorithm chooses the best model after learning from past iterations. However, it takes a lot of computational sources to compute [74]. In this study, the *tpot* classifier and *tpot* regressor was combined with a genetic algorithm as an informed search technique for classification and regression models, respectively. Thus, in addition to empirically determining the best accurate classifiers and regressors, not only the model selection has been strengthened, but also the model's parameters have been tuned with the genetic algorithm-based informed search used as additional validation.

2.7. Model evaluation
The robustness of the model was assessed using a variety of performance evaluation criteria. Accuracy (ACC), recall (REC), precision (PRE) *F1*-score (*F1*), area under the receiver operating characteristic (ROC) curve (AUC) [75], Jaccard index (JI), and elapsed time (ET) [76] were selected as the evaluation criteria for classification models. The confusion matrix and ROC curve were also provided for the best classification model. The formulas of the selected performance metrics are presented as follows;

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$REC = \frac{TP}{TP + FN} \tag{2}$$

$$PRE = \frac{TP}{TP + FP} \tag{3}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{4}$$

$$JI = \frac{TP}{TP + FP + FN} \tag{5}$$

where *TP*, *TN*, *FP*, and *FN* indicate predicted classes as true positive, true negative, false positive, and false negative, respectively. As this problem is a multiclass classification problem, *TP*s hold the diagonal values of corresponding classes, while off-diagonal values represent the errors in the confusion matrix. More precisely, *FN*s denote the total number in the corresponding rows excluding the *TP*s, *FP*s denote the total number in the corresponding columns excluding the *TP*s, and finally, *TN*s indicate the total numbers excluding the corresponding row and column. The multiclass classification problem can be handled by a set of many binary classification problems for each class. For instance, when considering the 'None' class, *TP* denotes the actual 'None' class is predicted to be 'None', while 'Weak' or others is to be considered as *FN*. Similarly, when considering the *REC*, or complement of Type II error, it holds the *TP*s ratio among those with corresponding classes, and e.g. for the 'Strong' class, Type II error denotes the MI incorrectly classified as 'Strong' but actually not. This might be applied to each class. Therefore, these metrics show the success of the trained models in the classification of MI considering the categorized groups (none, weak, strong, and complete), rather than the positive and negative discrimination of MI.

To evaluate the regression models' performance, robust statistical metrics such as the coefficient of determination or *R*-squared (R^2), mean absolute error (MAE), mean squared error (MSE), root-mean-squared error (RMSE) [12], and ET [77] were calculated. The formulas of the selected performance metrics are presented below;

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \tag{6}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}| \tag{7}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y})^2} \tag{8}$$

where *i* is the data point, *N* is the number of data points, and $\hat{y}$ and $\bar{y}$ indicate the predicted value of actual value (*y*) and mean value of *y*, respectively.

In order to assess the robustness of the models, the mentioned evaluation metrics were calculated during the training, validating, and testing phases. Moreover, feature importance analysis was performed for the best classification and regression models. Feature importance analysis refers to a technique that calculates the weighted scores of all the predictors. The calculated scores describe which features are more relevant for trained models [78]. In this paper, an improved Gini Index-based feature importance analysis algorithm [79] was adopted to determine more relevant features. By using this improved version, feature scores were determined by calculating proportionally to the node purity of features instead of counting splits. Calculated scores basically denote a total increment in the purity of a node weighted by the number of samples and weighted by the probability of reaching the corresponding node. It means a higher score of purity indicates that the particular attribute has a greater impact on the prediction model for MI. Additionally, in order to determine whether there is a statistical significance in the performance of the optimized classification and regression models, the one-way ANOVA test was performed on each trained model's validation accuracies and R^2 scores [80].
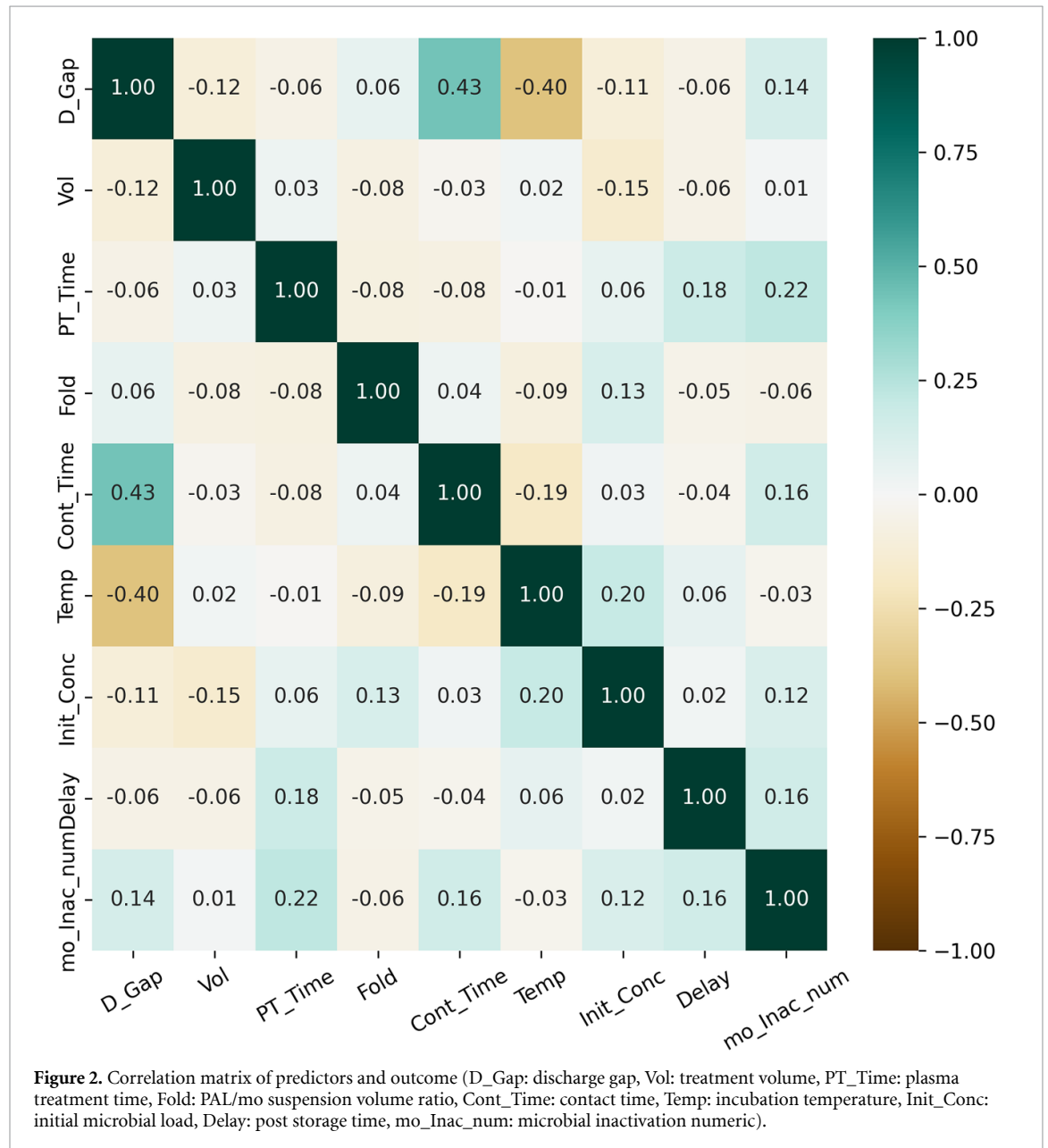
3. Results

3.1. Experimental setup

A total of collected data with 13×762 dimensions satisfied the required criterion without any missing values. Table 3 represents statistical data for numeric variables. The discharge gap has a mean of 13.71 ± 17.23 mm. Moreover, the range of this variable is between 0 and 81 mm. Plasma treatment time has a mean of 1165 ± 2771.48 s and a range of 0–14 400 s. Treatment volume ranges from 0.25 to 500 ml, with a mean of 45.52 ± 116.56 ml. Another variable, the initial microbial load, has a mean of 6.62 ± 0.91 log, with a range of 2–9 log. The PAL/mo suspension volume ratio is one of the most crucial variables, with a mean value of 109.25 ± 246.75 fold, ranging from 1 to 1000 fold. Contact time ranges from 0 to 1440 min and has a mean of 39.05 ± 131.02 min. Incubation temperature has a mean of $23.5 \pm 12.4\,^\circ$ C while ranging from -80 to $50\,^\circ$ C, having mostly positive values. The range of post-storage time is 0–6120 min, with a mean of 64.47 ± 345.56 min. Additionally, the detailed expression of nominal predictors with their unique values' labels and frequencies in the total observations are presented in table S2 in the supplementary data. The most prevalent type of plasma treatment, which is a nominal predictor among the included articles and considering 762 observations, was volume dielectric barrier discharge with a frequency of 40.6%. The second one was the plasma jet, with a frequency of 33.3%. Among the gas types, the air has a frequency of 78.5% and was the most used gas type among the articles. The second most used gas type was argon+oxygen with a frequency of 14.6%. When PALs were examined, DIW was the most commonly used among the 33 articles. DIW has a frequency of 51.7%, while the second-most-used liquid, saline, has a frequency of 18.4%. Two model organisms, *Escherichia coli (E.coli)* and *Staphylococcus aureus (S.aureus)*, which are widely practiced in experiments, were the two species with the highest frequency in our study when microbial strain is considered. The most frequently used microorganism in the included studies was the gram-negative model organism *E.coli*, which has a frequency value of 41.5%. The second most frequently used microorganism was the gram-positive model organism *S.aureus*, which has 31.1% frequency. All nominal variables were selected to build ML models even though they have diverse data distribution and some variable frequency rates are higher than others in order to observe the impact and contributions on the predictions.

In order to better understand the relationships between the numeric predictors as well as the outcome (MI), a correlation matrix was also created. A perceptible event is shown as color in two dimensions using the data visualization technique known as a correlation matrix. Based on hue or intensity, variation in color provides unmistakable visual cues as to how the event clusters or changes in space [81]. Figure 2 showed the results of the correlation matrix. The correlation matrix revealed correlations between predictors and their correlations with output. Examining the correlation matrix reveals that there was relatively no correlation between the predictors. Contact time, incubation temperature, and discharge gap might stand out as the most correlated features. The discharge gap and contact time have a positive correlation of 0.43 magnitude, and the incubation temperature and discharge gap have a negative correlation of 0.40 magnitude. As clearly stated in [82], a magnitude of correlation coefficients between 0.3 and 0.5 indicates a low correlation level, whereas a moderate level needs a higher magnitude, and so on. Moreover, a correlation coefficient of 0.5 magnitude was determined as a threshold between independent variables in a recent similar work [17]. Therefore, there is no significant correlation between the predictors. As a result, all predictors were preserved for model training because they might make crucial contributions to model development. Additionally, any of the predictors have no significant correlation with the outcome. Therefore, all predictors might contribute to the prediction of the outcome since the correlation coefficient threshold for elimination or preservation was not achieved. As a result of the correlation analysis, both the collected numerical and nominal predictors are essential and key variables that might have a significant biological impact on MI. It should also be noted that the feature importance analysis was conducted to examine the contributions of predictors to model development.

Preprocessing steps were used to prepare the data for ML models by distributing the data evenly, converting nominal data to numeric data, and avoiding data imbalance. After normalization and resampling, a total of 1152 observations were obtained. After resampling, a total of 922 observations (80%) were randomly selected for use in the training phase, while the remaining unseen 230 observations (20%) were used in the test phase. Also, 10% (92 observations) of the training data was used for validation in each fold. For the *n*-repeated stratified *k*-fold cross-validation strategy in the classification problem, n and k were selected as 3 and 10, respectively, which yields a total of 30 folds. For the *k*-fold cross-validation in regression models, *k* was chosen as 10. Following the model development, the generation size was set to 100, the population size to 150, and the offspring size to 20 for both the genetic algorithm-based *tpot* classifier and regressor. Both informed searches were conducted with a traditional ten-fold cross-validation strategy to determine the best pipeline. An Intel® Core™ i9-10 940X CPU and 64 GB RAM were used to develop all supervised ML models.
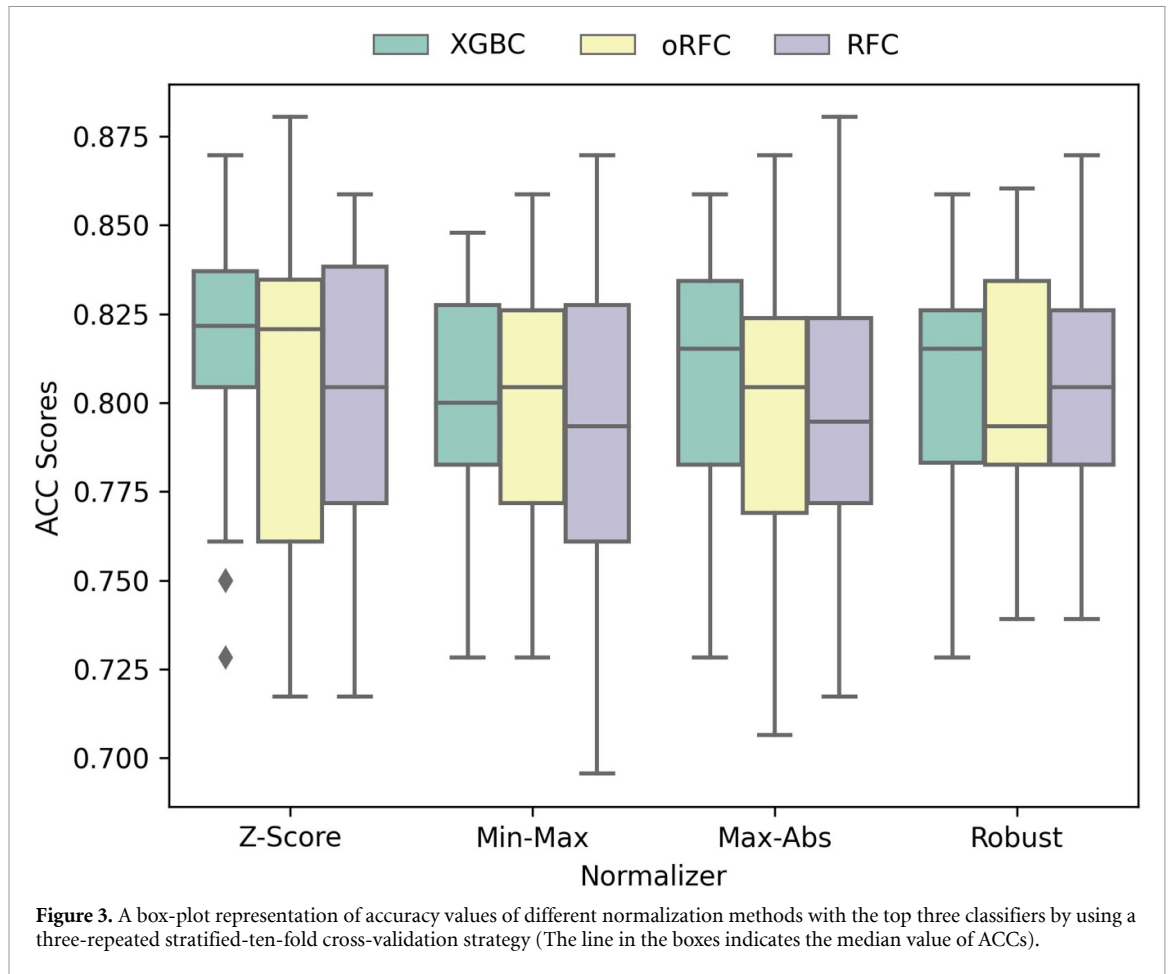
Table 3. Descriptive statistics of numeric variables.

Numeric variable	Unit	Count	Mean	SD	Min.	Max.
Discharge gap	mm	762	13.71	17.23	0	81
Plasma treatment time	sec	762	1165	2771.48	0	14400
Treatment volume	mL	762	45.52	116.56	0.25	500
Initial microbial load	log	762	6.62	0.91	2	9
PAL/mo suspension volume ratio	fold	762	109.29	246.75	1	1000
Contact time	min	762	39.05	131.02	0	1440
Incubation temperature	°C	762	23.5	12.4	−80	50
Post storage time	hour	762	64.47	345.56	0	6120
Microbial inactivation	log	762	2.83	2.77	0	9

Figure 2. Correlation matrix of predictors and outcome (D_Gap: discharge gap, Vol: treatment volume, PT_Time: plasma treatment time, Fold: PAL/mo suspension volume ratio, Cont_Time: contact time, Temp: incubation temperature, Init_Conc: initial microbial load, Delay: post storage time, mo_Inac_num: microbial inactivation numeric).

3.2. Classification results

To enhance the accuracy of the models, the numeric input variables underwent data normalization. Several methods, including Z-score, min-max scaler, max-absolute scaler, and robust scaler, were used as normalization techniques. The performance of three distinct classifiers and the various normalization techniques were compared. Figure 3 presents the accuracy scores obtained from the XGBC and RFC with unoptimized hyperparameters and optimized RFC (oRFC) models by using a three-repeated

Figure 3. A box-plot representation of accuracy values of different normalization methods with the top three classifiers by using a three-repeated stratified-ten-fold cross-validation strategy (The line in the boxes indicates the median value of ACCs).

stratified-ten-fold cross-validation strategy. oRFC is a hyperparameter-optimized version of conventional RFC by derived from the *tpot* classifiers. The average accuracy values after applying the *Z*-score normalization technique were 80.38%, 80.56%, and 81.29% for the XGBC, RFC, and oRFC models, respectively. The classifier with the highest accuracy value was oRFC, which accuracy value before normalization was 70.11% and a huge increase to 81.29% after normalization. Such a high accuracy change shows the effect of the magnitude difference of the numerical values on the performance of the model. Additionally, while oRFC yielded an average accuracy of 81.29% using *Z*-score normalization, 79.37%, 79.87%, and 80.35% accuracies were obtained by utilizing min-max scaler, max-absolute scaler, and robust scaler, respectively. This accuracy difference enabled *Z*-score selection among others and was implemented in the next stages of the study.

Various classifiers were utilized to choose the most accurate one among them. The different classifier results were presented in figure 4 as classifiers versus accuracy scores. RFC, XGBC, and LGBMC were the first three classifiers, which led to average validation accuracy scores of 80.47%, 80.23%, and 78.63%, respectively. The three classifiers with the lowest validation accuracy scores were GNB, SVM, and BNB, with average accuracy values of 38.17%, 46.20%, and 46.90%, respectively. For this study, RFC is the most accurate model that may provide the most precise predictions. It should be noted that the accuracy values may vary due to the stochastic nature of ML models and the fact that the dataset is randomly shuffled each time. While the accuracy value of the RFC model in the normalization comparison was 80.56%, the accuracy value in comparing classifiers was 80.47% with an acceptable variance. Table 4 summarizes the performance metrics of various classifiers for validation and training results. ACC and AUC were provided for training results, and ACC, F1, REC, PRE, JI, AUC, and ET were provided for validation results. Please note that F1, REC, and PRE values were provided as macro scores. There was a 42.3% accuracy difference between the most accurate model (RFC) and the least accurate model (GNB). To evaluate whether a model has statistical significance, a one-way ANOVA test was performed and the $p - value$ threshold was determined as .001. Statistical analyses demonstrated that while RFC and XGBC accuracy scores were statistically significant ($p < .001$) among other classifiers, no significance ($p > .999$) was found for other classifiers. Despite LGBMC and HGBC having an accuracy value of 78.63% and 78.50%, which is close to the RFC accuracy score, there was no statistical difference for both classifiers. For the RFC, the average training ACC and AUC were $99.00\% \pm 0.26\%$ and 1.00 ± 0.00, respectively. The validation ACC was $80.47\% \pm 4.23\%$, *F1*-score was $80.47\% \pm 4.07\%$, REC was

Figure 4. A box-plot representation of obtained validation accuracy values of different classifiers by using a three-repeated stratified-ten-fold cross-validation strategy (The line in the boxes indicates the median value of ACCs).
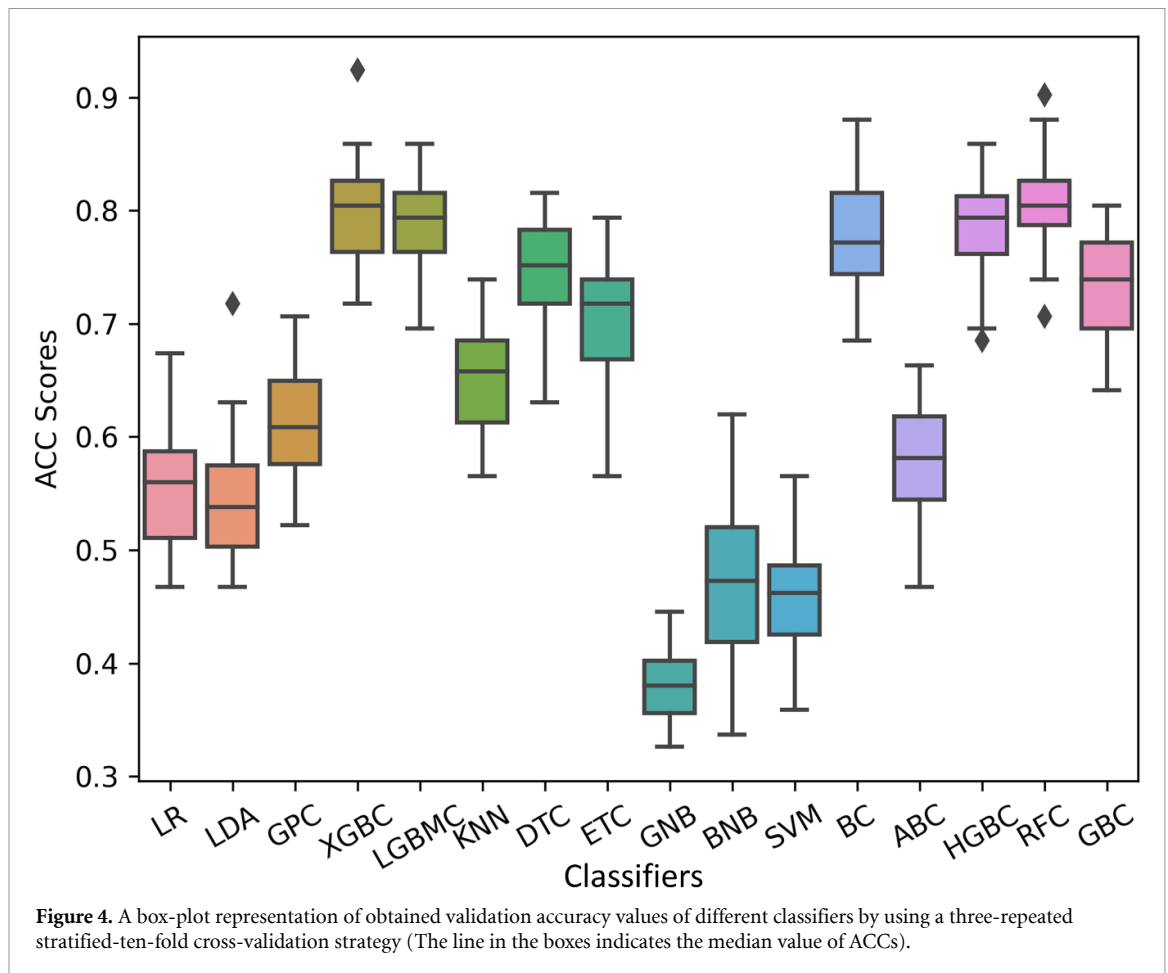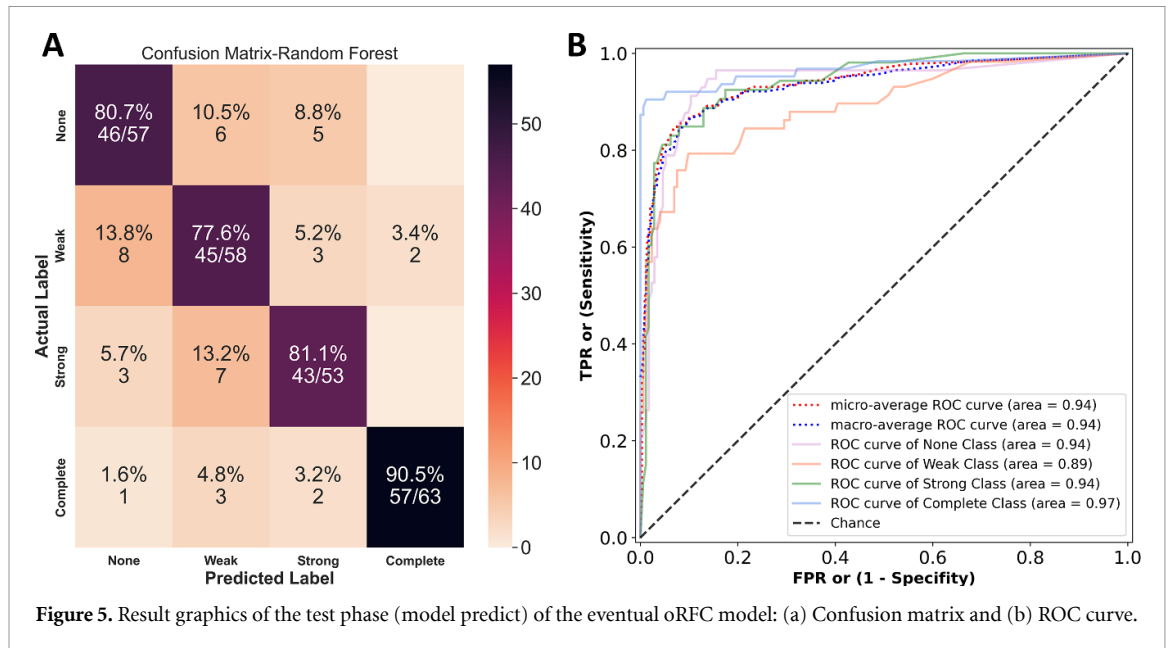
Table 4. Average performance metrics of three-repeated stratified-ten-fold cross-validation strategy for different classifiers (ACC, F1, REC, and PRE values are given as percentages (%) and the ET unit is s.).

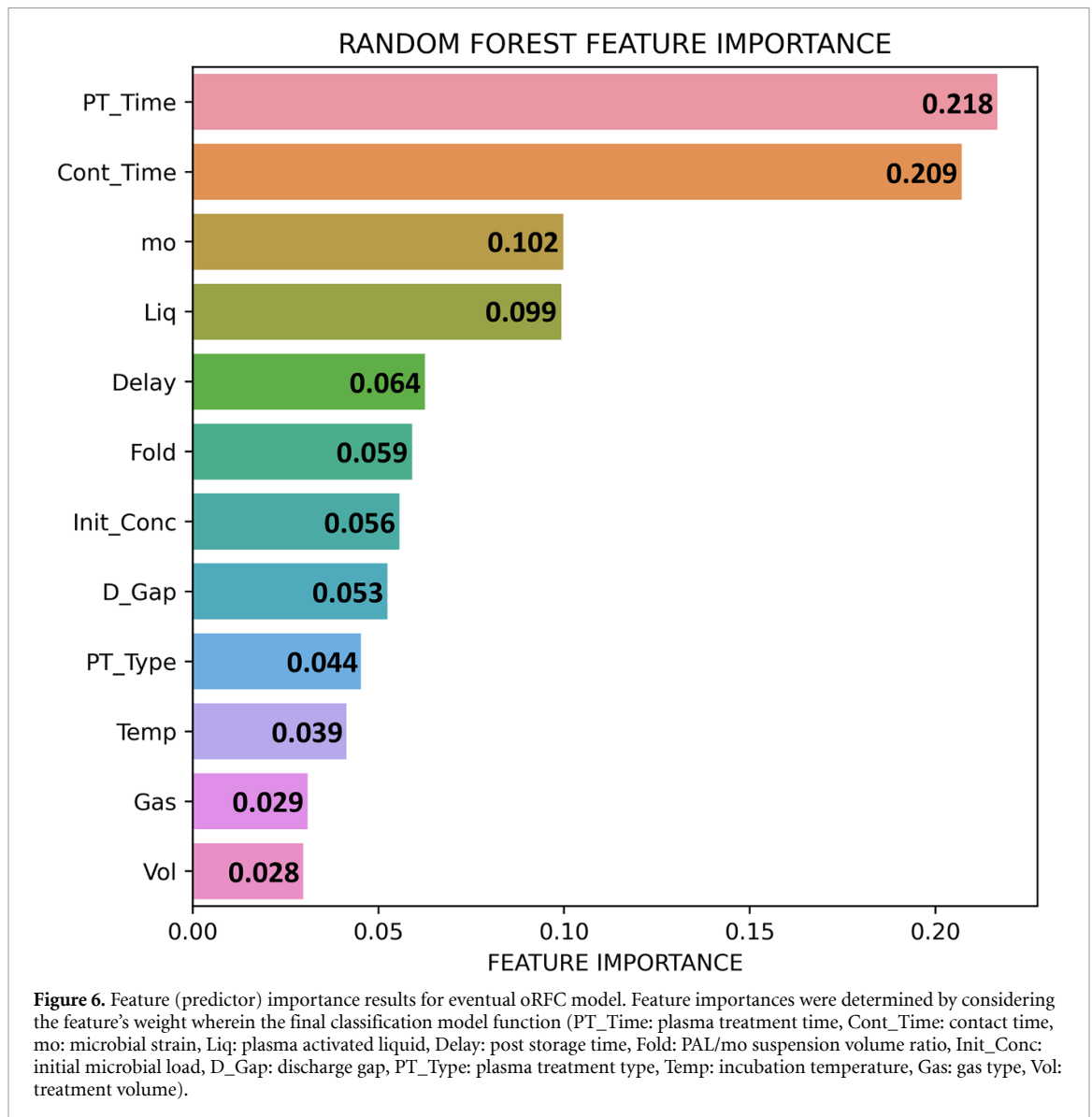| Classifier | Train | | Validation | | | | | | | |
	ACC	AUC	ACC	F1	REC	PRE	JI	AUC	ET	*p*-value
LR	55.40	0.82	55.53	55.40	55.50	57.17	0.39	0.77	1.28	$>.999$
LDA	60.20	0.83	54.70	54.43	54.60	56.43	0.38	0.78	1.62	$>.999$
GPC	70.07	0.90	61.20	61.03	61.20	62.13	0.45	0.83	92.17	$>.999$
XGBC	98.20	1.00	80.23	80.23	80.23	80.70	0.68	0.94	9.76	$<.001$
LGBMC	97.00	1.00	78.63	78.60	78.57	79.33	0.65	0.94	9.76	.250
KNN	77.40	0.95	65.20	64.93	65.37	66.23	0.49	0.85	3.16	$>.999$
DTC	99.00	1.00	74.73	74.63	74.77	75.43	0.60	0.84	1.20	$>.999$
ETC	99.00	1.00	70.37	70.30	70.30	71.00	0.55	0.81	1.07	$>.999$
GNB	40.52	0.76	38.17	31.80	37.97	45.57	0.20	0.72	1.14	$>.999$
BNB	51.30	0.76	46.90	46.73	46.97	47.53	0.31	0.72	1.17	$>.999$
SVM	48.03	0.77	46.20	45.53	46.23	50.47	0.30	0.74	14.77	$>.999$
BC	97.50	1.00	77.63	77.57	77.63	78.23	0.64	0.92	2.50	.781
ABC	62.20	0.81	58.17	58.13	58.23	59.37	0.42	0.77	6.35	$>.999$
HGBC	97.00	1.00	78.50	78.47	78.50	79.20	0.65	0.93	132.98	.331
RFC	99.00	1.00	**80.47**	80.47	80.47	81.03	0.68	0.94	10.50	$<.001$
GBC	88.00	0.98	73.47	73.43	73.47	74.37	0.59	0.91	22.27	$>.999$

80.47% ± 4.18%, PRE was 81.03% ± 4.00%, JI was 0.68 ± 0.06, and the AUC was 0.94 ± 0.02. The mentioned performance metrics of the RFC model in each fold are presented in table S3 in the supplementary data. In the second-most accurate model, the XGBC results, the ACC values are in line with the RFC. But, XGBC yielded better ET than RFC with 9.76 s. RFC was determined as the best classification model and was used for the hyperparameter tuning and testing phase of the study. The statistical tests also supported the superiority of the RFC algorithm, among others.

Figure 5. Result graphics of the test phase (model predict) of the eventual oRFC model: (a) Confusion matrix and (b) ROC curve.

After hyperparameter tuning, oRFC was obtained by using the predefined parameters to improve model accuracy. By optimizing the RFC model using the parameters obtained by the *tpot* classifier, a new model was developed. The optimized parameters were determined, such as the maximum depth of 10, maximum features of 0.3, minimum samples leaf of 7, minimum samples split of 10, and estimators count of 100 for the oRFC model. Prior to hyperparameter tuning, the accuracy value of RFC was obtained as 79.29% by using the default parameters provided in *scikit-learn* such as minimum samples leaf of 1, minimum samples split of 2, estimators count of 10, no maximum depth limit, and the number of maximum features limited by the square root of the input feature count. After hyperparameter tuning, the accuracy value for the oRFC model was enhanced to 81.29%. Model accuracy was clearly impacted by normalization methods and hyperparameter tuning. The oRFC model was tested using the 20% test data that was allocated during the data splitting stage after the model's performance had been established. In the test phase, the oRFC model achieved 82.68% ACC, 82.55% *F1* macro, 82.47% REC, and 82.71% PRE. Furthermore, JI, AUC, Cohen's kappa, and Matthew's correlation coefficients were calculated as 0.71, 0.88, 0.77, and 0.77, respectively. A visual representation of the tree structure of the oRFC model is presented in figure S3 in the supplementary data.

In this study, a confusion matrix was employed to assess the effectiveness of the classification model. By comparing the actual values with those predicted by the trained model, the confusion matrix evaluates how well the classification model works well [83]. Figure 5(a) shows the confusion matrix created using the preprocessed dataset with the oRFC model. The distribution of the four indicators in the dataset-TP, FP, TN, and FN-was shown in the confusion matrix. The confusion matrix was generated in the testing phase with the use of the previously indicated four output labels: None, Weak, Strong, and Complete. The confusion matrix's 'TP' and 'TN' values were those instances where the expected value and the actual value matched. The accuracies of the four output labels for the oRFC model were 80.7% for 'None', 77.6% for 'Weak', 81.1% for 'Strong', and 90.5% for 'Complete'. The estimated values have high match percentages with the true values, which were represented by the darker boxes, and low match percentages with the mismatched values, which were represented by the lighter boxes, proving that the chosen model is a reliable model. The overall test accuracy was yielded as 82.68%. Figure 5(b) shows the ROC curve that compares the rate of FP values to TP values. The gold standard of the classification has the largest AUC, and the sensitivity of the model rises as this value approaches 1. The area under the 'Complete' class' curve with a value of 0.97 has the most sensitivity, whereas the area under the 'Weak' class with a value of 0.89 has the lowest. Finally, the average test AUC score was computed as 0.94.

The final analysis for the classification model is presented in figure 6 which represents the findings of the feature or predictor importance analysis. The weights of the predictors' coefficients, which are part of the ML model function, were used to construct a measure of feature importance. Then the predictors were ordered, starting with the highest coefficient. The highest weight represents the most contributed feature to the building of the best oRFC model. Hence, the greatest ones have a tremendous impact on the accurate prediction of MI. In comparison to other variables, plasma treatment time, contact time, microbial strain,

Figure 6. Feature (predictor) importance results for eventual oRFC model. Feature importances were determined by considering the feature's weight wherein the final classification model function (PT_Time: plasma treatment time, Cont_Time: contact time, mo: microbial strain, Liq: plasma activated liquid, Delay: post storage time, Fold: PAL/mo suspension volume ratio, Init_Conc: initial microbial load, D_Gap: discharge gap, PT_Type: plasma treatment type, Temp: incubation temperature, Gas: gas type, Vol: treatment volume).

liquid type, and post-storage time were identified as more significant features; it may be inferred that these attributes have the greatest impact on the output. Despite the fact that the effects of treatment volume, gas type, incubation temperature, plasma treatment type, and discharge gap were not as significant as the others, the distributions of coefficients were not only weighted on certain ones. This indicates all predictors may have a significant impact on not only the building of the classification model but also on the accurate prediction of MI. These results were also in line with the correlation map results of the numerical data.

3.3. Regression results

The dataset was resampled by using SMOGN as a data preprocessing step before the training of the regression models because the data distribution was not balanced. The SMOGN algorithm, which mixes oversampling with Gaussian noise [58], is based on the SMOTER method. The density of the outcome of the dataset after oversampling with the SMOGN approach is shown in figure 7 together with the initial data density. The majority of the original data distribution ranged between 0 and 2 for MI. Following the application of SMOGN, the data was distributed more uniformly and nearly in a Gaussian distribution. The data imbalance problem was reduced using the SMOGN technique, and then all data was resampled to produce a homogenous Gaussian data distribution. The SMOGN implementation has affected R^2 scores. While the average R^2 value was 0.68 before resampling, with the 0.04 increments, this value increased to 0.72 after the SMOGN application. Although the effect of the SMOGN on the validation R^2 scores seems not to be too significant, the resampling method might have a clear impact on the prediction of new incoming data. Therefore, the importance of the resampling method is undeniable.

Figure 7. The data distribution graph of before and after the SMOGN method.
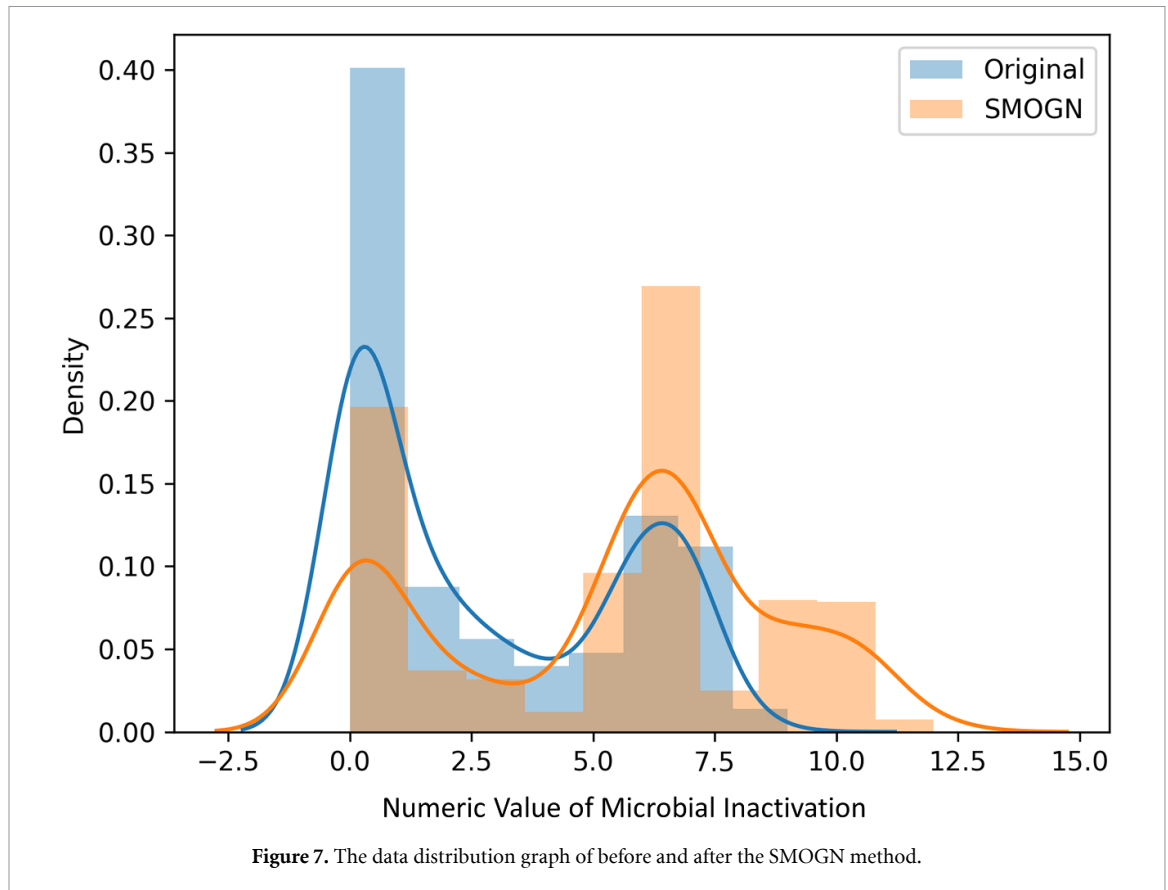
To enhance regression performance, similar to the classification problem, normalization methods like *Z*-score, min-max, max-absolute, and robust scaler were performed. The highest-performing top three regressors were chosen to compare the effect of the normalization techniques to the R^2 scores. Figure 8 presents the R^2 scores obtained from the XGBR and RFR with unoptimized hyperparameters and the hyperparameter-optimized version of the RFR (oRFR) models by using a traditional ten-fold cross-validation strategy. The oRFR which is a similar version of the oRFC model, was derived by using the *tpot* regressors. The average R^2 values after applying the robust scaler normalization technique were 0.68, 0.68, and 0.72 for the XGBR, RFR, and oRFR models, respectively. The regressor with the highest average R^2 score was oRFR, which R^2 value before normalization was 0.64 and increased to 0.72 after normalization. While the average R^2 value was achieved for the best regressor, oRFR, as 0.72, other normalization methods yielded the average R^2 scores as 0.71 for all three normalization methods (*Z*-score, min-max, and max-abs). The robust scaler normalization method was selected to be implemented in the next stages of the study since it was the most effective method on the R^2 scores.

Following the data preprocessing steps, regression models were trained for the optimized dataset. The various regressor results were presented in figure 9 as regressor versus R^2 scores. The three best R^2 scores belong to RFR, BR, and ETR, with 0.72, 0.69, and 0.68 scores, respectively. The three regressors with the lowest validation R^2 scores were ENR, LASSO, and LSVR, with average R^2 values of 0.00, 0.05, and 0.06, respectively. As with classification results, the RF-based regressor was the most accurate one. Due to the stochastic nature of the model, R^2 values may vary between normalization and regression results similar to the classification scenarios. Table 5 summarizes the performance metrics of a conventional ten-fold cross-validation strategy for various regressors. While R^2, MAE, and RMSE were provided for training results, R^2, MAE, MSE, RMSE, and ET were provided for validation results. According to the results, the RFR model has the lowest error and the greatest R^2 score when compared to the other regressors, with average R^2, MAE, MSE, and RMSE values of 0.72 ± 0.09, 0.33 ± 0.04, 0.28 ± 0.09, and 0.53 ± 0.09, respectively. There was a 0.72 R^2 value difference between the most accurate model (RFR) and the least accurate model (ENR). One-way ANOVA was performed, and the results demonstrated that while the RFR and BR models' R^2 scores were statistically significant ($p < .001$), no significance ($p > .999$) was found for other regressors. Despite the ETR and XGBR models both having the R^2 value of 0.68, which is close to the RFR R^2 score, there was no statistical difference for both regressors. For the RFR, the average training R^2, training MAE, and training RMSE were 0.95 ± 0.01, 0.13 ± 0.01, and 0.23 ± 0.01, respectively. The mentioned performance metrics of

Figure 8. A box-plot representation of R^2 scores of different normalization methods with the top three regressors by using a ten-fold cross-validation strategy (The line in the boxes indicates the median value of R^2s).

Figure 9. A box-plot representation of obtained validation R^2 scores of different regressors by using a ten-fold cross-validation strategy (The line in the boxes indicates the median value of R^2s).
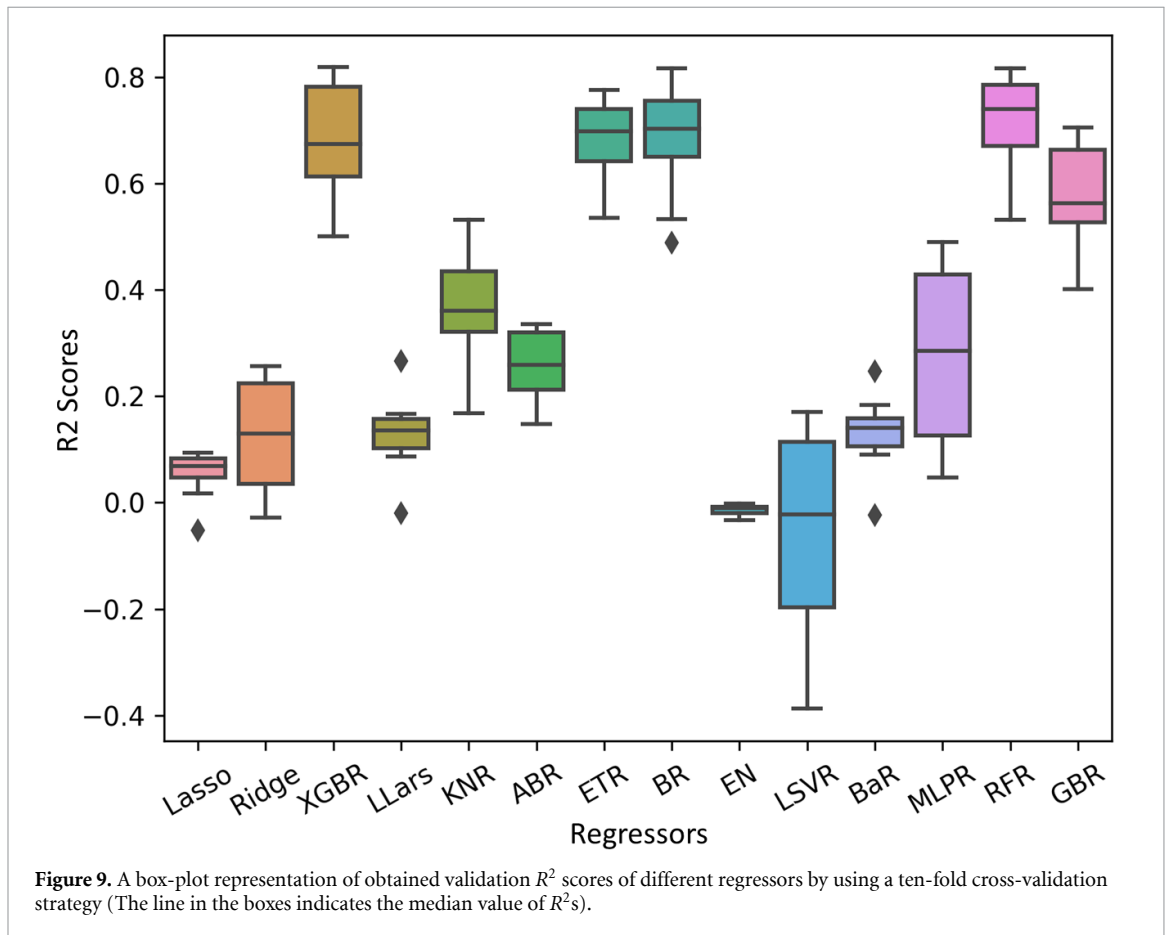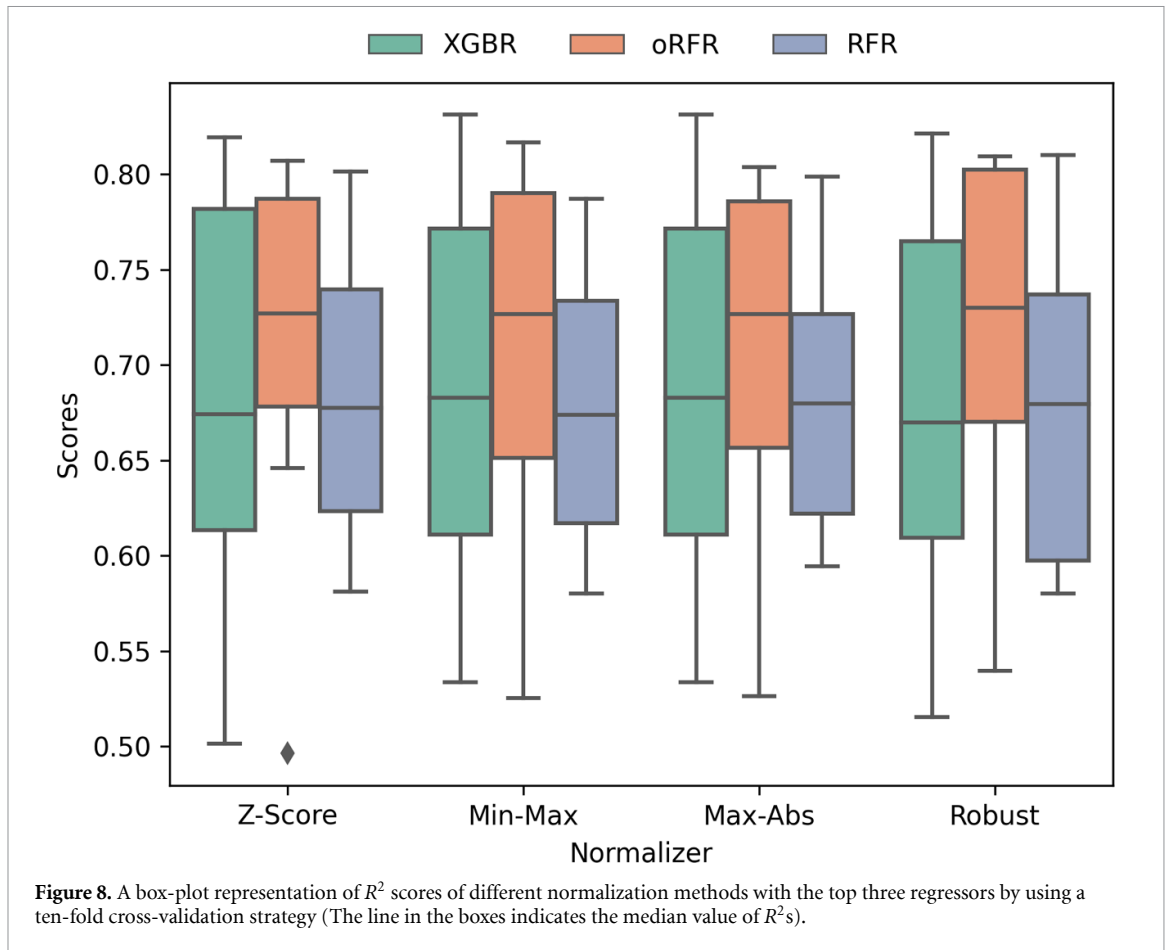
Table 5. Average performance metrics of ten-fold cross-validation strategy for different regressors (ET unit is s.).

Regressor	Train			Validation					
	R^2	MAE	RMSE	R^2	MAE	MSE	RMSE	ET	*p*-value
LASSO	0.08	0.88	0.96	0.05	0.89	0.95	0.97	0.18	$> .999$
RR	0.28	0.72	0.86	0.13	0.79	0.87	0.93	0.17	$> .999$
XGBR	0.96	0.10	0.19	0.68	0.36	0.32	0.56	1.35	.009
LLars	0.19	0.80	0.91	0.13	0.82	0.87	0.93	0.23	$> .999$
KNR	0.59	0.47	0.64	0.37	0.62	0.63	0.79	0.23	$> .999$
ABR	0.31	0.74	0.84	0.26	0.76	0.74	0.86	0.50	$> .999$
ETR	0.98	0.02	0.14	0.68	0.31	0.32	0.56	3.03	.003
BR	0.93	0.14	0.26	0.69	0.34	0.31	0.55	0.49	$< .001$
ENR	0.00	0.92	1.01	0.00	0.92	1.02	1.01	0.18	$> .999$
LSVR	0.16	0.64	0.92	0.06	0.76	1.07	1.03	0.22	$> .999$
BaR	0.20	0.79	0.90	0.13	0.82	0.87	0.93	0.48	$> .999$
MLPR	0.56	0.50	0.66	0.28	0.65	0.72	0.84	17.45	$> .999$
RFR	0.95	0.13	0.23	**0.72**	0.33	0.28	0.53	3.05	$< .001$
GBR	0.73	0.42	0.52	0.58	0.52	0.42	0.64	0.99	.952

the RFR model in each fold are presented in table S4 in the supplementary data. In the second most accurate model, BR has the R^2 value of 0.69, which is 0.03 lower than RFR. MAE, MSE, and RMSE values were close to RFR. However, BR yielded better ET than RFR with 0.49 s. RFR was determined as the best regression model and used for the hyperparameter tuning and testing phase of the study. Similar to the classification results, the statistical tests also supported the superiority of the RF-based algorithm, among others.

By employing the predefined parameters to increase the model's R^2 score after hyperparameter tuning, the oRFR model was produced. The optimum parameters for the oRFR model were as follows: a maximum depth of 8, a maximum feature of 0.5, a minimum sample leaf of 8, a minimum sample split of 12, and an estimator count of 100. Before hyperparameter tuning, the RFR model's R^2 score was obtained as 0.68 by using the default parameters provided in *scikit-learn* such as a maximum feature of 1, a minimum sample leaf of 1, a minimum sample split of 2, an estimator count of 10 and no maximum depth limit. The R^2 value for the oRFR model was increased to 0.72 after hyperparameter tuning. The oRFR model was tested using the 20% test data that was allocated during the data splitting stage after the model's performance had been established. In the test phase, the oRFR model was achieved nearly 0.75 R^2, 0.32 MAE, 0.25 MSE, and 0.50 RMSE value. Furthermore, the maximum error and variance scores were calculated as 1.72 and 0.75, respectively.

Additionally, figure 10 presents the plot of predicted MI values versus actual MI values. The model's ability to predict the measured outcome (MI) variable can be examined using the graph. The residual of the fit is not directly plotted on either axis in the graph. Instead, the graph shows the predicted y value on the y-axis and the actual y value (recorded in the data table) on the x-axis. The standard diagonal line that centers the graph is the fit line for the best regressor. In this instance, the vertical distance between the plotted point and the red line of identity serves as a representation of the residual (the horizontal distance can also be used as these distances will always be the same for each point). The graph demonstrates that the model performs better at predicting actual values at lower and medium Y values (between 0 and 5), which are closer to the RFR fit line, while predictions are further off at higher Y values (points farther from the fit line). Since each data point is close to the predicted regression line, it can be concluded that the chosen RFR model fits the data quite well, with a test R^2 value of nearly 0.75. The graphic demonstrates the close agreement between model predictions and actual data, demonstrating the reliability of the regression model.
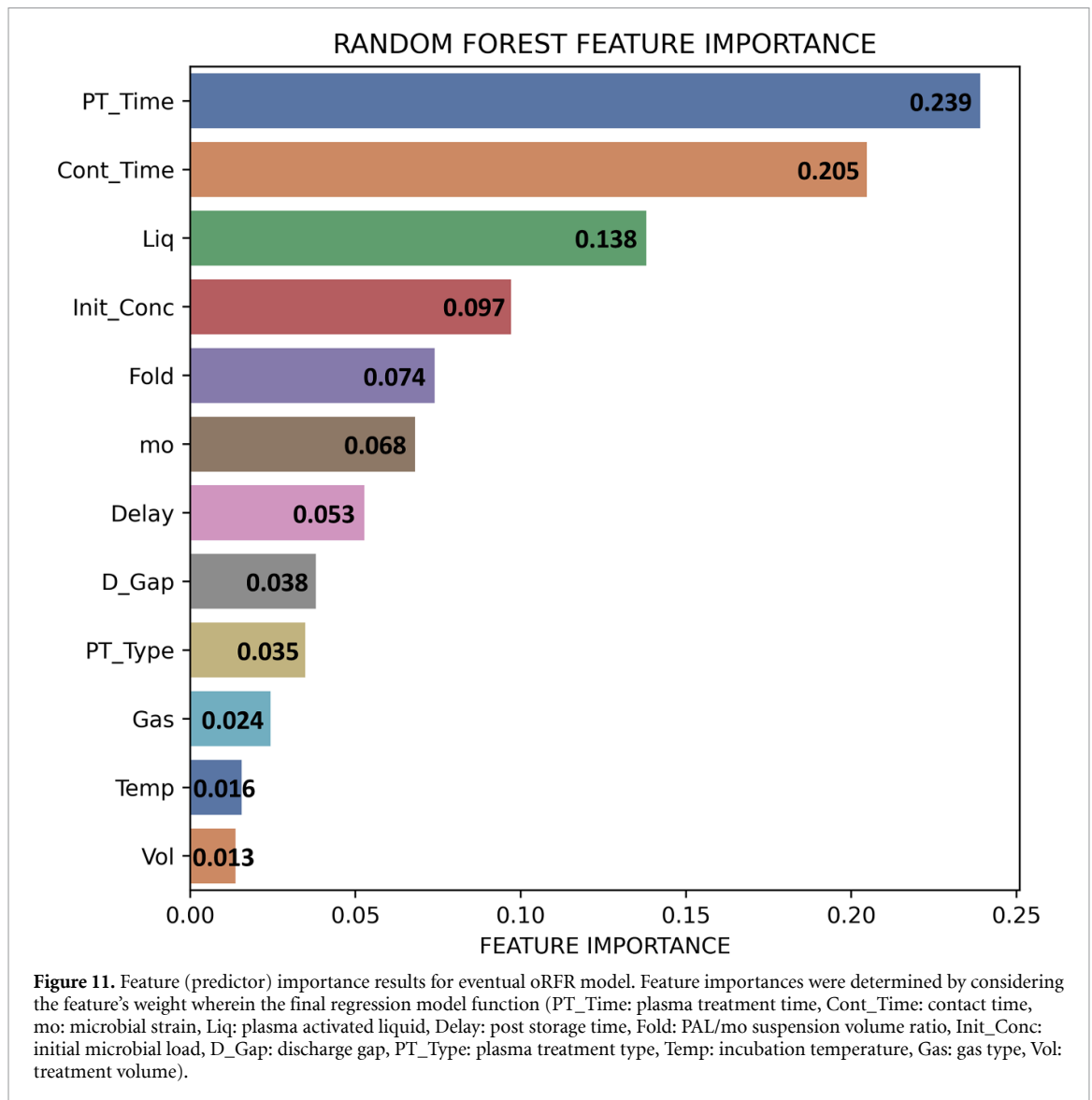
The objective of this study's feature importance analysis was to quantify the impact of the predictors on the output of the regression model. The scores from the feature importance analysis shed light on the dataset and model and may enhance the model's predictive performance [12]. Figure 11 presents the outcomes of the feature importance analysis. Similar to the classification problem, feature importance analysis was performed using the model function's coefficients. In comparison to other predictors, plasma treatment time, contact time, liquid type, and initial microbial load were determined to be more significant features; it may be inferred that these features had the most impact on the outcome. In contrast, the regression model is less impacted by treatment volume, incubation temperature, gas type, and plasma treatment type. Compared with the feature importance graph obtained as a result of the classification model, the four features (plasma treatment type, gas type, incubation temperature, and treatment volume) that had the least impact on the output were the same as the regression model's feature importance, even though they were in different orders. Furthermore, the two most important features of both models were the contact time and plasma

Figure 10. Regression plot of the test phase (model predict) of eventual oRFR model ($p < .001$).
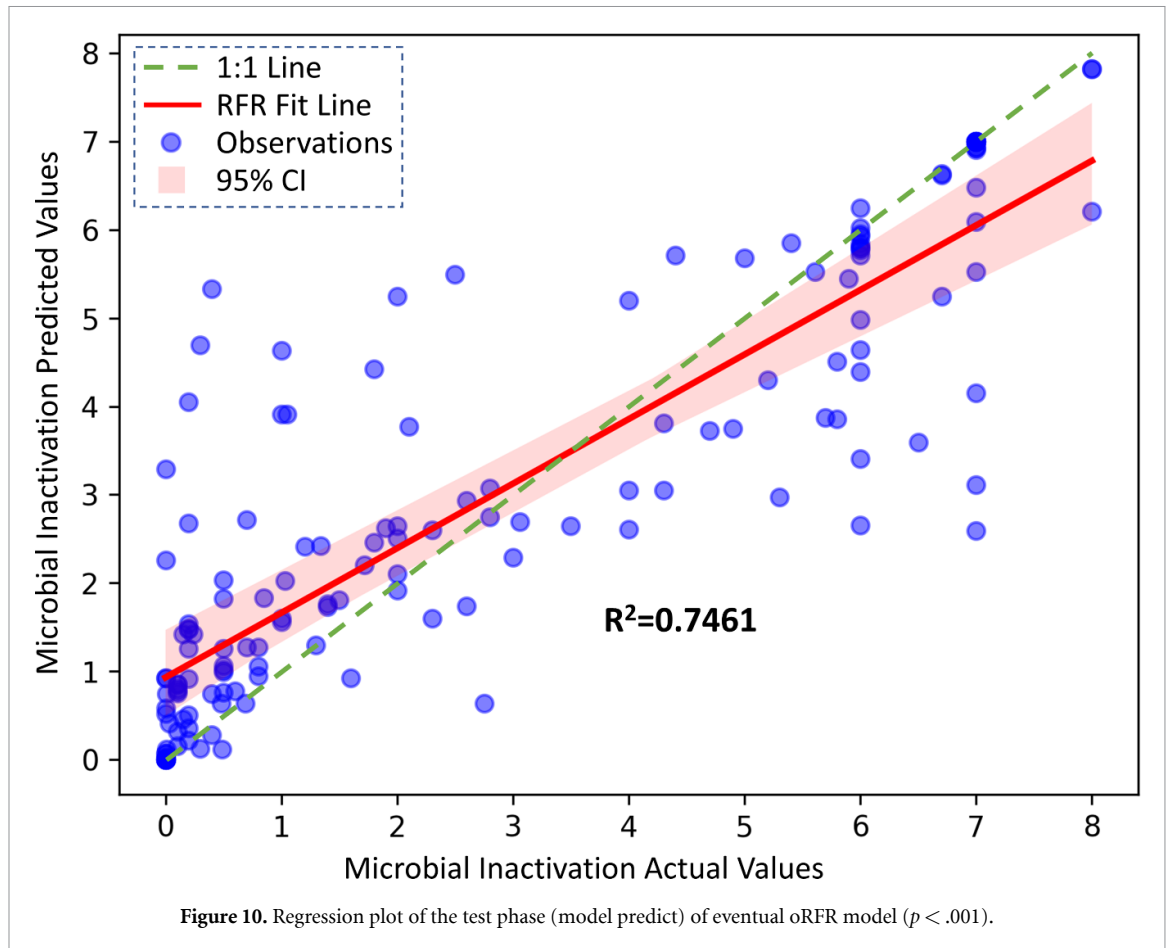
treatment time. When compared to other features, PAL ranked fourth in the classification model but third in the regression model. Microbial strain and initial microbial load are the only two features that have changed for the top four features in the feature importance analysis for both models. The initial concentration was the fourth important feature in the regression model, while the microbial strain was the third feature for classification. Although little differences exist, The regression features' importance was in line with the classification features importance analysis. The feature importance analysis results, as well as correlation graphs, demonstrated that all predictors might have a significant impact on the building of supervised ML models.

4. Discussion

In order to predict the MI of PALs, this study presented the implementation of ML in the field of plasma medicine, from data collecting through model evaluation. According to the studies in the literature, the outcomes of quantitative and qualitative analyses to determine the antimicrobial activity of PALs depend on a number of factors. The microbial strain and type of PAL both affect PAL's antimicrobial activity [7]. Additionally, the power, frequency, exposure period, pulse form, electrode geometry, and other parameters that affect plasma device efficiency could change the antimicrobial effect of PALs [84]. Therefore, it is challenging to compare the antimicrobial activity of PALs produced by various devices, and the optimization of parameters required to achieve a remarkable antimicrobial effect is both time-consuming and costly. Determining the plasma treatment parameters, achieving the desired antimicrobial effect, and tailoring plasma treatment to a particular standardization are significant challenges in the field of plasma medicine. The definition of 'plasma dose' will be aided by standardization, which is both a crucial problem and a necessity in the field of plasma medicine. The issue of defining the 'plasma dose' using AI is growing in significance despite the fact that there is numerous research for plasma dose assessment in the literature [85, 86]. In this study, it is expected that the widespread use of AI in the medical and biomedical fields today will enable the targeted standardization of plasma medicine. To the best of our knowledge, no studies have previously employed ML models to predict the antimicrobial activity of PALs. Some research in the literature claims that PALs' antimicrobial activity is significantly influenced by a number of different factors. Plasma

Figure 11. Feature (predictor) importance results for eventual oRFR model. Feature importances were determined by considering the feature's weight wherein the final regression model function (PT_Time: plasma treatment time, Cont_Time: contact time, mo: microbial strain, Liq: plasma activated liquid, Delay: post storage time, Fold: PAL/mo suspension volume ratio, Init_Conc: initial microbial load, D_Gap: discharge gap, PT_Type: plasma treatment type, Temp: incubation temperature, Gas: gas type, Vol: treatment volume).

treatment time, contact time, liquid type, microbial strain, and PAL/mo suspension volume ratio are a few examples of these characteristics [6, 20, 24, 28]. This study's prediction model is based on the different research groups' features that are examining the antimicrobial efficacy of PALs.

In this study, ML was based on various preprocessing steps as well as different classification or regression models due to the select the best combination for building and training the models. Also, the obtained results for the training, validation, and testing phases were evaluated in a comprehensive and fair way. Therefore, many classification and regression scenarios, parameter tuning approaches, cross-validation strategies, and statistical analyses were carried out, and robust statistical metrics were provided. For this direction, different normalizing strategies were utilized to prepare the data for model training. The models trained using the original data before the normalization techniques have shown inferior prediction accuracy and R^2 scores. The main cause of this is that the created dataset has wide unit ranges, and before training the model, normalization techniques limit the ranges of the dataset for each feature to make model training easier. For classification and regression models, four alternative normalization techniques were tested. The precision capabilities of both models improved after normalization, however, while the *Z*-score normalization method revealed better accuracy for classification, the robust scaler method yielded better performance for regression models by using the same feature set.

In ML, the high accuracy of the created model is very important both for its usage and for evaluating its functionality in terms of the realization of the target and its suitability for real life. Especially, the high precision of ML studies aimed to be used in the field of health is a very necessary and desirable feature [87]. As mentioned before, there are only a few studies reported in the literature for the prediction of different biological outcomes by using ML techniques, and they have a great impact on their fields due to the

advantages of decreasing the experimental cost and time consumption. Nonetheless, their prediction performance is limited. For instance, Shaban and Alkawareek implemented supervised ML for predicting the antibiofilm activity of antibiotics using three models: regression, DT, and RF. The models utilized 18 inputs and 1 outcome, and the prediction accuracy was reported as 67% $\pm$ 6.1%, 73% $\pm$ 5.8%, and 74% $\pm$ 5%, respectively, based on 580 observations [17]. Mirzaei *et al* developed a predictive model for the antibacterial activity of nanoparticles with 11 inputs using a dataset of 1176 observations with missing values. Their regression model showed an R^2 value of 0.78 [12]. In another study [19], the RF model was employed to predict the neurotoxicity of nanoparticles with 12 different inputs and 1 output, based on a dataset of 603 observations without missing values. The ACC score of 72% was achieved based on the presented model. Overall, when other results of similar studies are taken into account and compared, the accuracy and R^2 values outperformed those presented in this study, and both of the created supervised ML models are more promising than other studies when considering the prediction performance.

The results of this study revealed that oRFC and oRFR were the most effective models for classification and regression. Furthermore, no overfitting has occurred for any supervised ML models. Due to their efficient operation on highly dimensional data and their excellent accuracy, RF models are a prominent data analysis tool in research for medical and biomedical applications [88]. RF models are typically applied to high-dimensional data sets without a linear relationship between the variables used as inputs and outputs [89]. Because the characterization of the data used in this study is high-dimensional and non-linear, it is appropriate for the RF model. In line with the literature, in this study, encouraging performances were obtained compared to other classification and regression algorithms by using RF-based models. Also, the RF-based models, which are structured on ensemble learning and trees, might have had superior performance to other models because of the data distribution.

Additionally, the trained oRFC model yielded an accuracy of 82.68%, whereas the final oRFR model achieved a R^2 value of 0.75 by predicting the unseen test data. For the comparison of two different supervised ML techniques, the best classification model (oRFC), in which the output variable (MI value) is categorized, outperformed the regression model (oRFR), in which the MI value is provided as a continuous variable. Regression models' prediction capability may be less than that of classification models because regression models attempt to estimate the exact MI values and MI has a wide distribution. For the field of plasma medicine, it may be more crucial to estimate the output value as the exact value when MI is considered as a biological output, but the model's prediction ability has been enhanced by the categorizing strategy utilized in the study. It is obvious that there is a trade-off between the ability to predict outcomes and the categorization of the MI variable. Additionally, both regression models using MI as a continuous variable and classification models using MI as a categorical variable have both been developed and discussed for usage. Classification or regression techniques may be preferred for the desired task by considering the advantages and disadvantages.

On the other hand, according to state-of-the-art studies, the plasma treatment time [20], contact time [24], liquid type [6], and initial microorganism concentration [28] are the variables that have the biggest impacts on the antimicrobial efficacy of PALs. They were picked as part of the study's data acquisition parameters as a consequence. On the basis of the results of the feature importance analysis, the plasma treatment time, contact time, and PAL type variables were shown to be the most significant features for obtaining a remarkable antimicrobial activity of PALs and highly impactful on the created models. These three variables, which were found to be the most important ones among other predictors, are also consistent with the literature. It is well known that antimicrobial activity directly correlates with plasma treatment time. Depending on the type and volume of PAL, it has been reported in the literature that the effectiveness of MI increases as the time of treatment of the liquid with CAP increases [5, 7]. MI is influenced by the length of time the generated PAL is in contact with the bacterial solution. During the period of contact, the plasma content that was transferred from the CAP to the liquid participates in a number of biochemical processes with microorganisms. Although the ideal duration for these reactions should be standardized, in some liquids it is vital to do so in order to prevent the plasma species' scavenging impact [9]. An extremely useful indicator of antimicrobial activity is the liquid form of PAL generated for a particular antimicrobial application. Changing the type of liquid also affects the liquid's chemical composition. It has been found that the molecules in this chemical structure interact with the reactive species in the plasma's active component to create new molecules [90]. The utilization of PALs as chemically-modified fluids for therapeutical applications is an area of active research in the plasma medicine field [91]. Related studies available in the literature have primarily focused on examining the chemical composition of PALs and identifying the dominant species present in the liquids, which are produced by different plasma systems and discharges [92]. In this study, we aimed to investigate the feature importance analysis of the parameters that are already accepted by research groups as 'optimal'. We assumed that the optimum conditions reported in the literature were sufficient for inducing specific chemical modifications, and our analysis aimed to identify the impact of those modifications on the resulting biological outcomes. Moreover, we acknowledge that the understanding

of the relationship between the chemicals formed because of the plasma discharge and the dominant species present in PALs, as well as their impact on biological outcomes, is an ongoing study that we aim to perform. In order to fully realize the potential of PALs as a therapeutic modality, further studies are needed to understand the underlying mechanisms of chemical modifications in PALs and the relationship between plasma parameters and the resulting chemical and biological outcomes [93].

The fourth important feature was the microbial strain for classification models, and it was the initial microbial load for regression models. Both are associated with *in vitro* characteristics. Also, both the microbial strain and initial microbial load features are crucial when comparing and evaluating the antimicrobial activity of PALs that are designed to target study [94]. Studies in the literature indicate that these characteristics, which are essential to the antimicrobial efficacy of PALs, correspond to the parameters identified by the trained model's feature importance analysis. This correlation demonstrated how well the training model fit and predicted the experimental data. Moreover, the classification model has been rebuilt with the four most important features to see whether the accuracy of the model has changed. The oRFC model yielded a 65.80% accuracy value by using only the top four important predictors, which were determined by the feature importance analysis. Furthermore, it was decided that testing the remaining eight features was necessary because the accuracy of the model built with the four most important features was quite low. Therefore, the oRFC model was rebuilt with the remaining eight features to observe the accuracy-changing effect and achieved an accuracy value of 58.87%. Both rebuilt models' accuracy values were far below the eventual test result of 82.68%. The findings from the rebuilt model demonstrate that all predictors have an impact on the model, even though the features with the largest and lowest contributions to the model were noted following feature importance analysis.

Despite the encouraging results of the present study, a number of limitations should be considered. First, the trained model was created using the data from the articles that were used. The data from the research in the literature were the only sources on which the prediction model was built. The models' accuracy and R^2 score were unaffected by this condition, but it prevented the model from being validated using actual data. Because the accuracy of the studies that were performed is a requirement for the prediction model that was presented in this study, this constraint needs to be removed. It should be also noted that although there are no significantly distributed/outlier values within the cross-validation stages, the best model for classification, RFC, has ± 4.23 SD between min. of 71.00% and max. of 90.00% accuracy in folds, and the best model for regression, RFR, has ± 0.09 SD between min. of 0.53 and max. of 0.82 R^2 in folds (see tables S3 and S4). After the model was developed, it should be tested by obtaining real data and comparing the prediction results to the experimental validation. The lack of observations is another limitation of this study. The amount of data used for this study should be increased, and the effect of the observation variance on the prediction ability of ML models should be decreased in this way. Higher data sizes in ML denote the model's robustness. In order to increase model validation, a dataset with a lot more observations is required. Briefly, a higher-dimensional dataset and *in vitro* tests are required to more accurately examine the antimicrobial effectiveness of PALs.

AI is effective in several biomedical fields, which encourages its application in plasma medicine. This study is the first to show how ML modeling can qualitatively predict the antimicrobial activity of PALs in the field of plasma medicine. This research is a pioneering study for the future development of qualitative and quantitative prediction models in plasma medicine applications via ML. The study's final results are promising and also capable of automatically predicting the MI for both models. The handling of input data which has diverse distributions by ML, one of the subfields of AI, also demonstrated remarkable performance in terms of computational cost and prediction capability. The results demonstrated that an adroit combination of ML techniques with CAP-related data might have a significant impact on plasma medicine, besides considering that AI studies are widely used and of crucial importance in other biomedical research fields. A comparison of real-time experimental results with prediction model results will demonstrate the applicability of ML in plasma medicine and demonstrate how it relates to daily life. In conclusion, an ML model that can automatically predict the MI value with high prediction ability has been created. It should also be noted that experimental studies are expensive, time-consuming, and difficult to show the experimental outcome. The obtained findings encourage the potential for merging ML techniques with plasma medicine applications to define the 'plasma dose' as well as adopting AI approaches to the prediction of other important parameters used in plasma medicine in the future. Also, the results are in line with the literature on the manner of biological aspects of CAP.

5. Conclusions

In conclusion, in the present study, we have utilized ML techniques to predict the antimicrobial activity of the PALs, which depends on the various plasma treatment parameters and hard to make comparisons in between the different studies. We conducted comprehensive and various robust training and test scenarios, as

well as statistical analysis, for building and testing supervised ML models to evaluate fairly the results. Our results revealed that the generated models in the present study can predict the antimicrobial activity of PALs with a test accuracy of 82.68% for the categorical outcome and with a test R^2 score of 0.75 for the exact values of the outcome based on the available literature. Also, RF-based models demonstrated superiority among other algorithms for the prediction of MI in both regression and classification problems. Furthermore, the importance of features that was determined through the model was in line with the literature, where many studies have shown the plasma treatment time as one of the primary parameters for the antimicrobial effect of PALs. Besides, the study has certain limitations, such as the dependence on data from the literature for model building, the lack of observations, and the need for higher-dimensional datasets and *in vitro* tests to examine the antimicrobial effectiveness of PALs more accurately. Despite these limitations, this study is the first to adequately utilize ML techniques in the field of plasma medicine to the best of our knowledge. By considering the wide range of applications of CAP in medicine and biology, ML techniques may assist in better understanding the biological outcomes of CAP applications. Furthermore, ML applications in plasma medicine may contribute to defining the 'plasma dose' which is a contemporary concept in the field of plasma medicine and thought to be a crucial concept to convey various CAP applications into clinical practice.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

CRediT

Mehmet Akif Özdemir: Conceptualization, Methodology, Software, Validation, Investigation, Visualization, Writing-Original Draft, Writing-Review & Editing. **Gizem Dilara Özdemir:** Conceptualization, Methodology, Investigation, Visualization, Data Curation, Formal Analysis, Writing-Original Draft, Writing-Review & Editing. **Merve Gül:** Data Curation, Writing-Original Draft. **Onan Güren:** Conceptualization, Writing-Review & Editing. **Utku Kürşat Ercan:** Conceptualization, Methodology, Investigation, Funding Acquisition, Writing-Original Draft, Writing-Review & Editing, Supervision. All authors read and approved the final manuscript.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID iDs

Mehmet Akif Özdemir ● https://orcid.org/0000-0002-8758-113X
Gizem Dilara Özdemir ● https://orcid.org/0000-0002-3682-3733
Utku Kürşat Ercan ● https://orcid.org/0000-0002-9762-2265

References

[1] Moreau M, Orange N and Feuilloley M 2008 *Biotechnol. Adv.* **26** 610–7
[2] Metelmann H R, Von Woedtke T and Weltmann K D 2018 *Comprehensive Clinical Plasma Medicine: Cold Physical Plasma for Medical Application* (Berlin: Springer)(https://doi.org/10.1007/978-3-319-67627-2)
[3] Oztan M O, Ercan U K, Aksoy Gokmen A, Simsek F, Ozdemir G D and Koyluoglu G 2022 *Sci. Rep.* **12** 1–15
[4] Kim S and Kim C H 2021 *Biomedicines* **9** 1700
[5] Kaushik N K *et al* 2019 *Biol. Chem.* **400** 39–62
[6] Tsoukou E, Bourke P and Boehm D 2018 *Plasma Med.* **8** 299–320
[7] Xiang Q, Fan L, Li Y, Dong S, Li K and Bai Y 2022 *Crit. Rev. Food Sci. Nutrition* **62** 2250–68
[8] Ercan U K, Wang H, Ji H, Fridman G, Brooks A D and Joshi S G 2013 *Plasma Process. Polym.* **10** 544–55
[9] Schmidt M, Hahn V, Altrock B, Gerling T, Gerber I C, Weltmann K D and von Woedtke T 2019 *Appl. Sci.* **9** 2150
[10] Cheng J, Chen Q, Fridman G and Ji H-F 2019 *Sens. Actuators Rep.* **1** 100001
[11] Mesbah A and Graves D B 2019 *J. Phys. D: Appl. Phys.* **52** 30LT02
[12] Mirzaei M, Furxhi I, Murphy F and Mullins M 2021 *Nanomaterials* **11** 1774

[13] Rajkomar A, Dean J and Kohane I 2019 *New Engl. J. Med.* **380** 1347–58

[14] Furxhi I, Murphy F, Mullins M and Poland C A 2019 *Toxicol. Lett.* **312** 157–66

[15] Serafim M S M, Kronenberger T, Oliveira P R, Poso A, Honorio K M, Mota B E F and Maltarollo V G 2020 *Expert Opin. Drug Discovery* **15** 1165–80

[16] Liu Z *et al* 2020 *Front. Microbiol.* **11** 48

[17] Shaban T F and Alkawareek M Y 2022 *Comput. Biol. Med.* **140** 105065

[18] Furxhi I, Murphy F, Mullins M, Arvanitis A and Poland C A 2020 *Nanomaterials* **10** 116

[19] Furxhi I and Murphy F 2020 *Int. J. Mol. Sci.* **21** 5280

[20] Kojtari A, Ercan U, Smith J, Friedman G, Sensenig R, Tyagi S, Joshi S, Ji H and Brooks A 2013 *J. Nanomed. Biother. Discovery* **4** 120

[21] Hong Q, Dong X, Yu H, Sun H, Chen M, Wang Y and Yu Q 2021 *Dental* **3** 1–7

[22] Smet C, Govaert M, Kyrylenko A, Easdani M, Walsh J L and Van Impe J F 2019 *Front. Microbiol.* **10** 1539

[23] Shen J, Tian Y, Li Y, Ma R, Zhang Q, Zhang J and Fang J 2016 *Sci. Rep.* **6** 1–10

[24] Zhao Y M, Ojha S, Burgess C, Sun D W and Tiwari B 2020 *J. Appl. Microbiol.* **129** 1248–60

[25] Li Y, Pan J, Ye G, Zhang Q, Wang J, Zhang J and Fang J 2017 *Eur. J. Oral Sci.* **125** 463–70

[26] Liu J, Yang C, Cheng C, Zhang C, Zhao J and Fu C 2021 *Bioengineered* **12** 4605–19

[27] Oehmigen K, Winter J, Hähnel M, Wilke C, Brandenburg R, Weltmann K D and von Woedtke T 2011 *Plasma Process. Polym.* **8** 904–13

[28] Kamgang-Youbi G, Herry J M, Brisset J L, Bellon-Fontaine M N, Doubla A and Naïtali M 2008 *Appl. Microbiol. Biotechnol.* **81** 449–57

[29] Kamgang-Youbi G, Herry J-M, Meylheuc T, Brisset J-L, Bellon-Fontaine M-N, Doubla A and Naitali M 2009 *Lett. Appl. Microbiol.* **48** 13–18

[30] Dezest M, Bulteau A L, Quinton D, Chavatte L, Le Béchec M, Cambus J P, Arbault S, Nègre-Salvayre A, Clement F and Cousty S 2017 *PLoS One* **12** e0173618

[31] Simon S, Salgado B, Hasan M I, Sivertsvik M, Fernández E N and Walsh J L 2022 *Plasma Chem. Plasma Process.* **42** 377–93

[32] Zhou R, Zhou R, Prasad K, Fang Z, Speight R, Bazaka K and Ostrikov K K 2018 *Green Chem.* **20** 5276–84

[33] Chiappim W, Sampaio A da Gça, Miranda F, Fraga M, Petraconi G, da Silva Sobrinho A, Kostov K, Koga-Ito C and Pessoa R 2021 *Water* **13** 1480

[34] Hänsch M A, Mann M, Weltmann K D and Von Woedtke T 2015 *J. Phys. D: Appl. Phys.* **48** 454001

[35] Rathore V, Patel D, Butani S and Nema S K 2021 *Plasma Chem. Plasma Process.* **41** 871–902

[36] Joshi I, Salvi D, Schaffner D W and Karwe M V 2018 *J. food Prot.* **81** 1472–80

[37] Ma M, Zhang Y, Lv Y and Sun F 2020 *J. Phys. D: Appl. Phys.* **53** 185207

[38] Naïtali M, Kamgang-Youbi G, Herry J M, Bellon-Fontaine M N and Brisset J L 2010 *Appl. Environ. Microbiol.* **76** 7662–4

[39] Qi Z, Tian E, Song Y, Sosnin E A, Skakun V S, Li T, Xia Y, Zhao Y, Lin X and Liu D 2018 *Plasma Chem. Plasma Process.* **38** 1035–50

[40] Royintarat T, Seesuriyachan P, Boonyawan D, Choi E H and Wattanutchariya W 2019 *Curr. Appl. Phys.* **19** 1006–14

[41] Suganuma R and Yasuoka K 2018 *Jpn. J. Appl. Phys.* **57** 046202

[42] Tian Y, Ma R, Zhang Q, Feng H, Liang Y, Zhang J and Fang J 2015 *Plasma Process. Polym.* **12** 439–49

[43] Traylor M J, Pavlovich M J, Karim S, Hait P, Sakiyama Y, Clark D S and Graves D B 2011 *J. Phys. D: Appl. Phys.* **44** 472001

[44] Wu S, Zhang Q, Ma R, Yu S, Wang K, Zhang J and Fang J 2017 *Eur. Phys. J. Spec. Top.* **226** 2887–99

[45] Xiang Q, Wang W, Zhao D, Niu L, Li K and Bai Y 2019 *Food Control* **106** 106741

[46] Ye G, Zhang Q, Pan H, Wang G, Li Y, Pan J, Wang J, Zhang J and Fang J 2013 Efficiency and mechanism of pathogenic bacteria inactivation by non-thermal plasma activated water *Proc. 21st Int. Symp. on Plasma Chemistry* p 1

[47] Zhang Q, Liang Y, Feng H, Ma R, Tian Y, Zhang J and Fang J 2013 *Appl. Phys. Lett.* **102** 203701

[48] Zhang Q, Ma R, Tian Y, Su B, Wang K, Yu S, Zhang J and Fang J 2016 *Environ. Sci. Technol.* **50** 3184–92

[49] Joshi S G *et al* 2015 *Adv. Biosci. Biotechnol.* **6** 49

[50] Ksiazek W, Abdar M, Acharya U R and Pławiak P 2019 *Cogn. Syst. Res.* **54** 116–27

[51] Pandey A and Jain A 2017 *Int. J. Comput. Netw. Inf. Secur.* **9** 36

[52] Ahsan M M, Mahmud M P, Saha P K, Gupta K D and Siddique Z 2021 *Technologies* **9** 52

[53] Yang X, Gong Y, Waheed N, March K, Bian J, Hogan W R and Wu Y 2019 Identifying cancer patients at risk for heart failure using machine learning methods *AMIA Annual Symp. Proc.* vol 2019 (American Medical Informatics Association) pp 933–41

[54] Tanha J, Abdi Y, Samadi N, Razzaghi N and Asadpour M 2020 *J. Big Data* **7** 1–47

[55] Alghamdi M, Al-Mallah M, Keteyian S, Brawner C, Ehrman J, Sakr S and Liu B 2017 *PLoS One* **12** 1–15

[56] Chawla N V, Bowyer K W, Hall L O and Kegelmeyer W P 2002 *J. Artif. Intell. Res.* **16** 321–57

[57] Branco P, Torgo L and Ribeiro R P 2017 Smogn: a pre-processing approach for imbalanced regression *1st Int. Workshop on Learning With Imbalanced Domains: Theory and Applications* (PMLR) pp 36–50

[58] Steininger M, Kobs K, Davidson P, Krause A and Hotho A 2021 *Mach. Learn.* **110** 2187–211

[59] Ozdemir M A, Degirmenci M, Izci E and Akan A 2021 *Biomed. Eng. / Biomed. Tech.* **66** 43–57

[60] Kaur P and Sharma M 2019 *J. Med. Syst.* **43** 1–30

[61] Christodoulou E, Ma J, Collins G S, Steyerberg E W, Verbakel J Y and Van Calster B 2019 *J. Clin. Epidemiol.* **110** 12–22

[62] Xu S 2018 *J. Inf. Sci.* **44** 48–59

[63] Zhang M L and Zhou Z H 2007 *Pattern Recognit.* **40** 2038–48

[64] Charbuty B and Abdulazeez A 2021 *J. Appl. Sci. Technol. Trends* **2** 20–28

[65] Hengl T, Nussbaum M, Wright M N, Heuvelink G B and Gräler B 2018 *PeerJ* **6** e5518

[66] Sørensen L, Nielsen M and Initiative A D N *et al* 2018 *J. Neurosci. Methods* **302** 66–74

[67] Rahman S, Irfan M, Raza M, Moyeezullah Ghori K, Yaqoob S and Awais M 2020 *Int. J. Environ. Res. Public Health* **17** 1082

[68] Shipe M E, Deppen S A, Farjah F and Grogan E L 2019 *J. Thoracic Dis.* **11** S574–84

[69] Fahrmeir L, Kneib T, Lang S and Marx B D 2021 *Regression Models Regression* (Berlin: Springer) pp 23–84

[70] Schratz P, Muenchow J, Iturritxa E, Richter J and Brenning A 2019 *Ecol. Modelling* **406** 109–20

[71] Degirmenci M, Ozdemir M, Izci E and Akan A 2021 *IRBM* **43** 422–433

[72] Andonie R 2019 *J. Membr. Comput.* **1** 279–91

[73] Olson R S, Urbanowicz R J, Andrews P C, Lavender N A, Kidd L C and Moore J H 2016 Automating biomedical data science through tree-based pipeline optimization *Applications of Evolutionary Computation: 19th European Conf., EvoApplications 2016 (Porto, Portugal)* vol Part I, ed G Squillero and P Burelli (Springer International Publishing) pp 123–37

[74] Ippolito P P 2022 Hyperparameter tuning *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies and Applications* (Berlin: Springer) pp 231–51

[75] Ozdemir M A, Ozdemir G D and Guren O 2021 *BMC Med. Inf. Decis. Mak.* **21** 1–20

[76] McGuinness K and O'connor N E 2010 *Pattern Recognit.* **43** 434–44

[77] Tohka J and Van Gils M 2021 *Comput. Biol. Med.* **132** 104324

[78] Altmann A, Toloşi L, Sander O and Lengauer T 2010 *Bioinformatics* **26** 1340–7

[79] Shang W, Huang H, Zhu H, Lin Y, Qu Y and Wang Z 2007 *Expert Syst. Appl.* **33** 1–5

[80] Ozdemir M A, Kisa D H, Guren O and Akan A 2022 *Biomed. Signal Process. Control* **77** 103787

[81] Steiger J H 1980 *Psychol. Bull.* **87** 245

[82] Koo T K and Li M Y 2016 *J. Chiropr. Med.* **15** 155–63

[83] Haghighi S, Jasemi M, Hessabi S and Zolanvari A 2018 *J. Open Source Softw.* **3** 729

[84] Monetta T, Scala A, Malmo C and Bellucci F 2011 *Plasma Med.* **1** 205–14

[85] Bonzanini A D, Shao K, Stancampiano A, Graves D B and Mesbah A 2022 *IEEE Trans. Radiat. Plasma Med. Sci.* **6** 16–32

[86] Sakai O, Kawaguchi S and Murakami T 2022 *Jpn. J. Appl. Phys.* **61** 070101

[87] Fleuren L M *et al* 2020 *Intensive Care Med.* **46** 383–400

[88] Garg A and Mago V 2021 *Comput. Sci. Rev.* **40** 100370

[89] Li C, Liao C, Meng X, Chen H, Chen W, Wei B and Zhu P 2021 *Patient Prefer. Adherence* **15** 691–703

[90] Verlackt C, Van Boxem W and Bogaerts A 2018 *Phys. Chem. Chem. Phys.* **20** 6845–59

[91] Griseti E, Merbahi N and Golzio M 2020 *Cancers* **12** 721

[92] Bradu C, Kutasi K, Magureanu M, Puač N and Živković S 2020 *J. Phys. D: Appl. Phys.* **53** 223001

[93] Zhou R, Zhou R, Wang P, Xian Y, Mai-Prochnow A, Lu X, Cullen P J, Ostrikov K K and Bazaka K 2020 *J. Phys. D: Appl. Phys.* **53** 303001

[94] Lee M J, Kwon J S, Jiang H B, Choi E H, Park G and Kim K M 2019 *Sci. Rep.* **9** 1–13