IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 July 2024, accepted 28 July 2024, date of publication 1 August 2024, date of current version 9 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3436603

RESEARCH ARTICLE

Mastering Precision in Pivotal Variables Defining Wine Quality via Incremental Analysis of Baseline Accuracy

CEMIL EMRE YAVAS, (Member, IEEE), JONGYEOP KIM, (Member, IEEE), AND LEI CHEN, (Senior Member, IEEE)
Department of Information Technology, Georgia Southern University, Statesboro, GA 30460, USA
Corresponding author: Cemil Emre Yavas (cy02470@georgiasouthern.edu)

This work was supported by Dean's Office of the Allen E. Paulson College of Engineering and Computing at Georgia Southern University.

ABSTRACT This study investigates the application of machine learning (ML) algorithms to enhance the precision of wine quality assessment, focusing specifically on Portuguese red wine. Amidst the growing interest in leveraging artificial intelligence (AI) for sensory analysis, our research distinguishes itself by employing a rigorous methodological framework. Our approach, named the 'Incremental Analysis of Baseline Accuracy,' identifies the chemical variables most predictive of wine quality. This framework aims to streamline the predictive process by pinpointing key variables that significantly influence quality assessments. In this paper, we demonstrate the feasibility of a methodology that precisely determines the criticality of chemical inputs, both their exact values and their correct order, to identify which inputs significantly contribute to the quality assessment of a sensory perception, such as taste. The centerpiece of our paper is a vibrant 3D pie chart that illustrates the percentage criticality of different input variables for perceiving the quality of red wine. This chart symbolizes the essence of our paper: a 'pie' representing the empirical conclusion, not mere conjecture. Through this paper, we have shown that it is possible to quantify a qualitative, perceptual aspect like taste perception, which is often believed to be assessable only through subjective conjecture. Moreover, our findings, facilitated by the Incremental Analysis of the Baseline Accuracy method, demonstrate that this perception can be systematically quantified, challenging traditional assumptions about sensory analysis.

INDEX TERMS Algorithm comparison, artificial intelligence (AI), chemometric variables, data analysis, machine learning (ML), Portuguese red wine, predictive analytics, random forest model, sensory analysis, variable selection, wine quality assessment.

I. INTRODUCTION

The history of wine, with roots stretching back over millennia, is rich and varied. Evidence suggests its production dates as far back as 6000 BC in Georgia and 5000 BC in Iran. Wine has been revered since ancient times, seamlessly integrating into the fabric of human civilization and playing pivotal roles in religious ceremonies, social gatherings, and culinary traditions. Its journey from a simple fermented beverage to a complex symbol of culture, refinement, and scientific intrigue illustrates the dynamic relationship between humans and this esteemed nectar [1].

For much of its history, the quality of wine was assessed solely through sensory evaluation: sight, smell, and taste. These subjective methods, while the only means available, provided a window into the nuances of wine's character. The art of winemaking, saturated with tradition and passed through generations, was cloaked in mystique, with only a surface understanding of the scientific underpinnings of wine quality.

In the 19th century, significant advancements in the scientific analysis of wine were made with the introduction of the

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed F. Zobaa.

first acidity tests in the 1820s, crucial for assessing a wine's tartness and stability, and the standardization of methods to determine alcohol content. This enhancement improved our understanding of wine's fundamental characteristics. Progress continued into the 20th century with the evolution of analytical chemistry and microbiology, leading to advanced techniques that could quantify volatile acidity, citric acid, residual sugar, chlorides, sulfur dioxide levels, density, pH, sulfates, and alcohol content, providing a comprehensive overview of wine's chemical makeup. Despite these advancements, identifying variables most critical to wine quality remained challenging. Wine's intricacy, characterized by a vast spectrum of volatile compounds and a delicate flavor balance, made it difficult to single out key factors contributing to quality. This complexity fueled ongoing debates regarding the relative importance of physical, chemical, and sensory characteristics in determining wine's excellence, without reaching a conclusive answer.

The emergence of machine learning (ML) and artificial intelligence (AI) has ushered in a new era for wine quality analysis, enabling extensive datasets to be examined to uncover patterns and correlations beyond human detection. This research, focusing on red wine, represents a significant step forward by applying ML algorithms for a comprehensive analysis of the chemical properties intrinsic to wine, with the aim of identifying the most predictive factors of wine quality. This endeavor not only enhances our understanding of what constitutes exceptional wine but also advances methodologies for more efficient and economical quality testing. By focusing on crucial chemical properties, winemakers can refine quality control practices, uphold high standards, and minimize costs.

Major Contributions: The contributions to this paper can be summarized as follows: **First**, we employed a brute-force approach to training and testing ML models on all 11 possible combinations that determine the taste of wine, providing concrete figures on the top 10 combinations of variables which affect the accuracy of wine quality prediction. **Second**, we identified the ''baseline accuracy'' as the inherent predictive ability of our ML models to accurately identify the quality of wine samples without any specific input data. The baseline accuracy is the performance of the AI when it uses this foundational understanding to make educated guesses about wine quality, achieving an accuracy of 40% in identifying the quality of wine samples without direct information. This concept highlights the AI's capability to leverage its initial training to approximate the quality distribution of wine effectively, even when not provided with explicit input variables for each sample. **Third**, we calculated the criticality of each variable in assessing wine quality through a methodological framework called ''Incremental Analysis of Baseline Accuracy.'' We began by training our AI with a large dataset, which included various chemical input variables associated with wine samples. To evaluate the impact of each variable, we systematically added one variable at a time to the model and observed how the predictive accuracy changed from the baseline measure. We explored over 2,047 combinations of input variables, testing each combination multiple times to ensure reliability and consistency in our results. By measuring the change in accuracy with the inclusion of each variable, we identified those that significantly enhanced the model's predictive capacity. These impactful variables were then ranked and weighted according to their importance in improving accuracy. This detailed approach allowed us to quantify the criticality of each variable, thereby enhancing our understanding of their roles in the sensory perception of wine quality. We believe this attempt was the first in the field of wine machine learning.

II. BACKGROUND

A. RELATED WORK - RESEARCH MENTIONING CRITICAL VARIABLES

We have selected the domain of chemical tests on red wine as our primary focus because, through our review of research papers, we observed that machine learning algorithms can determine the quality of red wine with over 70% accuracy on a scale of 0 to 10. We reviewed 19 papers related to algorithms predicting wine quality using machine learning and selected nine that specified critical variables. Here, we summarize their work as follows.

(1) Yavas et al. [2] machine learning techniques such as Random Forest and Logistic Regression were used to analyze the taste profiles of Red Portuguese Wine. The research found that a few critical variables, notably volatile acidity and alcohol, are primary indicators for flavor prediction. This suggests potential for a more simplified approach in sensory analysis, with the Random Forest model yielding the highest accuracy in predictions. Machine Learning (Decision Tree, Random Forest, Logistic Regression, Support Vector Machine), Input Variables (11), Critical Variables (Volatile Acidity, Sulfates, Alcohol, Solid Acid), Wine Type (Red Portuguese Wine). This paper is accepted to SERA 2024.

(2) Cortez et al. [3] in the paper titled ''Using Data Mining for Wine Quality Assessment,'' the team applied machine learning algorithms such as Support Vector Machines, Multiple Regression, and Neural Networks to determine the quality of White Vinho Verde wine. The research identified alcohol, sulphates, pH, volatile acidity, and residual sugar as key variables. Findings indicated that the Support Vector Machine model was particularly effective in this context, offering valuable insights into the relationship between wine's physicochemical properties and sensory preferences. Machine Learning (Support Vector Machine, Multiple Regression, Neural Networks), Input Variables (11), Critical Variables (Alcohol, Sulphates, pH, Volatile Acidity, Residual Sugar), Wine Type (White Vinho Verde).

(3) Gupta's 2018 research [4], titled ''Selection of Important Features and Predicting Wine Quality Using Machine Learning Techniques,'' delved into the efficacy of feature selection in wine quality prediction using Linear Regression, Neural Networks, and Support Vector Machines. Gupta's

work highlighted the varied significance of physicochemical properties, such as volatile acidity and sulphates, in predicting quality in red wines, and fixed acidity and residual sugar in white wines. The study concluded that machine learning models yielded more accurate predictions when trained on a carefully selected subset of features, underscoring the critical role of feature selection in enhancing model performance. Machine Learning (Linear Regression, Neural Networks, Support Vector Machine), Input Variables (11), Critical Variables (Red wine: Volatile Acidity, Chlorides, Free Sulfur Dioxide, Total Sulfur Dioxide, pH, Sulphates, Alcohol; White: Fixed Acidity, Volatile Acidity, Residual Sugar, Free Sulfur Dioxide, Density, pH, Sulphates, Alcohol), Wine Type (Red and White).

(4) Anami et al. [5] concluded that the Support Vector Machine (SVM) outperforms other techniques with an error of 0.003 and a quality rate of 7.99 for predicting wine quality as good or bad. This approach is deemed useful for the wine industry for quality testing and assurance for customers. Machine Learning: Neural Networks, Logistic Regression, Support Vector Machine, Input Variables (4), Critical Variables (4: Volatile Acidity, Citric Acid, Residual Sugar, Free Sulfur Dioxide), Wine Type: Portuguese ''Vinho Verde'' Wine.

(5) Zhang et al. [6], showed that both CART and Random Forest models achieved high accuracy in predicting the quality of red wine. Quality was converted into a binary classification to enhance prediction accuracy. Feature importance analysis showed alcohol as a significant factor influencing wine quality. The combination of Logistic Regression with Random Forest did not significantly improve model accuracy over using Random Forest alone. Machine Learning (Decision Tree, Boosting, Classification and Regression Tree (CART), Random Forest), Input Variables (11), Critical Variables (Alcohol is the most influential), Wine Type (Red).

(6) Korade and Salunke [7], identified the Random Forest algorithm as the most effective in predicting wine quality, achieving the highest accuracy among the tested algorithms. The study also highlighted the importance of feature selection in improving prediction accuracy, with alcohol, volatile acidity, and sulphates being the most indicative of wine quality. Machine Learning (Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, AdaBoost Classifier, Gradient Boosting Classifier, K-Nearest Neighbors, Naive Bayes), Input Variables (11), Critical Variables (Alcohol, Volatile Acidity, Sulphates), Wine Type (White).

(7) Jain et al. [8], found that Random Forest and XGBoost algorithms were the most accurate for predicting wine quality, with feature selection identifying Alcohol, Sulfates, and Volatile Acidity as critical variables. The XGBoost model demonstrated 100% accuracy when trained and tested with selected features. The research also highlighted the importance of feature selection in improving model accuracy and performed hyperparameter tuning and clustering analysis to refine the prediction model further. Machine Learning: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Decision Trees (DT), AdaBoost, Gradient Boost, Input Variables (11), Critical Variables (Alcohol, Sulfates, Volatile Acidity), Wine Type (Red Portuguese ''Vinho Verde'' Wine).

(8) Olatunde David Akanbi et al. [9], found that the Random Forest model was the most accurate in predicting wine quality when validated using a 10-fold cross-validation technique. The study identified alcohol as the feature with the most significant impact on wine quality, suggesting that adjustments in alcohol level, along with fixed acidity, citric acid, and sulphates, can enhance wine quality. Conversely, volatile acidity and chlorides were found to contribute the least to wine quality. Machine Learning (Linear Regression, Neural Network, Naive Bayes Classification, Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), k-Nearest Neighbors (kNN), Support Vector Machines (SVM) with a linear kernel, Random Forest (RF)), Input Variables (11), Critical Variables (Alcohol (most contribution), Fixed Acidity, Citric Acid, Sulphates (significant contribution), Volatile Acidity, Chlorides (least contribution)), Wine Type (Red: Portuguese ''Vinho Verde'').

(9) Dahal et al. [10], noted that the Gradient Boosting Regressor outperformed other models with MSE, R, and MAPE of 0.3741, 0.6057, and 0.0873 respectively, demonstrating the effectiveness of statistical analysis in identifying key components controlling wine quality. The study emphasized alcohol as the main component influencing wine quality. Machine Learning: Ridge Regression, Support Vector Machine, Gradient Boosting Regressor, Artificial Neural Network, Input Variables (11), Critical Variables (Alcohol (highest correlation with wine quality), Citric Acid (lowest correlation)).

In Table 1, we summarize related work for better clarity.

TABLE 1. Summary of wine quality analysis works.

Work	Critical Variables	ML Techniques	Wine Type
[2]	Volatile Acidity, Sulfates, Alcohol, etc.	Decision Tree, RF, Logistic Regression, SVM	Red
[3]	Alcohol, Sulphates, PH, etc.	SVM, Regression, Neural Networks	White
[4]	Various acids and sulfates	Linear Regression, Neural Network, SVM	Red/White
[5]	Acidity, Sugar, Sulfur dioxide	Decision Tree, Boosting, CART, RF	Red
[6]	Alcohol influence	Tree, Boosting, CART, RF	Red
[7]	Alcohol, Acidity, Sulphates	Regression, Tree, RF, SVM, Boosting	White
[8]	Alcohol, Sulfates, Acidity	RF, XGBoost, DT, AdaBoost, Boost	Red
[9]	Alcohol, Acidity, Sulphates, etc.	Regression, Neural Network, Naive Bayes, etc.	Red
[10]	Alcohol, Citric Acid	Ridge, SVM, Boosting, ANN	Red

B. OTHER RELATED WORKS

(10) In their 2023 paper, Zaza et al. [11] found that the Support Vector Machine (SVM) outperformed all other models with a 96% accuracy rate after applying sampling methods to balance the dataset. Alcohol was identified as significantly impacting wine quality among the explored features. They used ML techniques including Random Forest, Support Vector Machine, Gradient Boosting, K-Nearest Neighbors, and Decision Tree.

(11) Angus [12] developed binary and multi-class neural network classifiers to predict wine quality based on physicochemical properties. The study achieved good prediction results and suggested that this makes it feasible to automate wine quality assessments without professional testers. The ML technique used was Neural Network (Binary and Multi-class classifiers).

(12) Bhardwaj et al. [13] found that the AdaBoost classifier achieved 100% accuracy in wine quality prediction using both features extracted by the XGB method and essential variables. The performance of the Random Forest significantly improved to 100% accuracy when using essential variables. The study highlighted the effectiveness of synthetic data generation and feature selection in predicting wine quality. They utilized ML techniques such as Adaptive Boosting (AdaBoost), Random Forest (RF), Extreme Gradient Boosting (XGB), Stochastic Gradient Decision Classifier (SGDC), Support Vector Machine (SVM), Gaussian Naive Bayes (GNB), Decision Tree Classifier (DTC), and K-Nearest Neighbors (KNN).

(13) In their study, Koranga et al. [14] found that both J48 and Random Forest algorithms exhibited high accuracy in the classification of white wine quality, achieving an accuracy of 99.895%. They noted that no single algorithm consistently outperformed others across all error measures in regression analysis, but J48 and MLP showed formidable performance. This study highlights the effectiveness of machine learning algorithms in predicting wine quality, indicating that it can be accurately achieved without relying on a single algorithm.

(14) Among the evaluated machine learning models, Mani et al. [15] concluded that Random Forest achieved the highest accuracy in predicting wine quality. The study emphasized the effectiveness of Random Forest for quality estimation, particularly in the food industry, due to its ability to handle complex datasets and provide robust predictions. They also highlighted the importance of model optimization for enhancing prediction accuracy. The ML techniques used in this study included Logistic Regression, Support Vector Machine (SVM), Adaboost, Decision Tree, and Random Forest.

(15) In their 2020 paper, Ye et al. [16] proposed a framework that combines MF-DCCA with XGBoost and LightGBM algorithms for predicting red wine quality ratings. The methodology demonstrated a higher accuracy (91.04%) compared to other machine learning algorithms on the same dataset. The study also highlighted the significant role of residual sugar in the complexity of red wine quality, while they identified volatile acidity and chlorides as weakly correlated attributes. They used ML techniques like XGBoost and LightGBM.

(16) Sinha and Kumar [17] found that among the tested algorithms, Random Forest exhibited the highest accuracy of 87.33%. Their study revealed that high quality is usually associated with low levels of volatile acidity, indicating its impact on wine quality. It also noted the significance of alcohol content in determining wine preferences among consumers. This work implemented ML techniques such as Logistic Regression, Stochastic Gradient Descent, Support Vector Classifier, and Random Forest.

(17) Aich et al. [18] found that feature selection-based feature sets provided better prediction accuracy than using all features. Accuracy ranged from 95.23% to 98.81% with different feature sets. Simulated Annealing-based feature selection outperformed Genetic Algorithm (GA)-based feature selection in achieving higher accuracy, with the SVM classifier showing the best performance among the tested classifiers. This work utilized ML techniques such as Nonlinear, Linear, Probabilistic Classifiers, Genetic Algorithm (GA)-based feature selection, and Simulated Annealing (SA)-based feature selection.

(18) Olatunde David Akanbi et al. [9] found that the Random Forest model was the most accurate in predicting wine quality when validated using a 10-fold cross-validation technique. They identified alcohol as the feature with the most significant impact on wine quality, suggesting that adjustments in alcohol level, along with fixed acidity, citric acid, and sulphates, can enhance wine quality. Conversely, they found that volatile acidity and chlorides contributed the least to wine quality. They utilized ML techniques such as Random Forest, Linear Regression, Neural Network, Naive Bayes Classification, Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), k-Nearest Neighbors (kNN), and Support Vector Machines (SVM) with a linear kernel.

(19) In his 2021 study, Zhou Tingwei [19] used active learning and the KNN algorithm for predicting red wine quality. The prediction accuracy reached nearly 90% after several iterations. The approach demonstrated the potential of active learning in reducing the need for large labeled datasets, with accuracy depending on various factors such as dataset size and iteration number. He used the ML technique K-Nearest Neighbor (KNN).

C. RELATED WORK ANALYSIS

Numerous studies have evaluated wine quality using Machine Learning, demonstrating varying degrees of success. However, among these studies, only nine have identified critical variables. To our knowledge, no study has investigated how these variables impact the judgment of wine quality. We address this gap by exploring every possible combination of variables to determine which set, when

used as inputs to an ML model, would yield the highest accuracy.

III. COMPARISON WITH EXISTING STUDIES

In this section, our intention was to conduct a comprehensive analysis of existing research on wine quality assessment, and compare various methods used in prior studies. By evaluating these studies, we pinpointed the strengths and weaknesses of previous research methodologies, and pinpointed the most effective variables employed in these studies. Comprehensive comparisons between existing works were provided, through which the strengths and weaknesses of prior research were scrutinized. Additionally, we discussed the effective variables and their respective definitions, clarifying the reasoning behind their selection. Moreover, we demonstrated comparative results with alternative approaches, to show how our proposed method aligns with existing techniques. This analysis not only places our study within a broader literature context, but it also highlights the innovative aspects of our approach, thereby addressing the limitations of existing methods and enhancing the accuracy and reliability of wine quality predictions.

A. COMPARISON WITH THE STUDY BY ZHAN ET AL.

Zhan et al. [20] investigated the forecasting of red wine quality utilizing machine learning algorithms, with specific emphasis on the effects of alcohol content, sulphates, total sulfur dioxide, and citric acid. Their study used the same dataset of Portuguese ''Vinho Verde'' red wine from 2009 as ours, but reclassified it into binary classes to denote low quality (ratings 1-5) and high-quality (ratings 6-10) wines. Four machine learning techniques, namely Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Naive Bayes, were employed. Among these, the Decision Tree technique exhibited superior performance, with an accuracy rate of 74.7%.

In contrast to their study, our research uses the full spectrum of wine quality ratings and employs a broader set of physicochemical attributes to enhance the precision of wine quality assessment, specifically pertaining to Portuguese red wine. Our study utilizes attributes such as fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality.

Zhan et al.'s primary strength lies in their comparative analysis of various machine learning techniques, and the identification of the most effective economic model for wine quality prediction. However, their binary classification approach may oversimplify the quality assessment process, compared to our detailed multi-class quality ratings.

B. COMPARISON WITH THE STUDY BY TINDAL ET AL.

Tindal et al. [21] emphasized mathematical modeling to enhance winemaking efficiency by examining the extraction and evolution of polyphenols, specifically anthocyanins and tannins, during the fermentation process. They applied dynamic and spatial mathematical models to comprehend the kinetics and spatial behavior of phenolic substances. Their study underscored the importance of elements like temperature, oxidation, and mixing regimes on the phenolic behavior over time.

Our study, in contrast, aims to identify the critical variables determining wine quality with the use of machine learning algorithms. Unlike Tindal et al., who focused on the extraction and reaction kinetics of phenolic substances, we concentrate on the application of various machine learning models to enhance the precision of wine quality assessment, related specially to Portuguese red wine. Our study uses physicochemical attributes as variables to predict wine quality, which include fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality.

Tindal et al.'s primary strength lies in their detailed exploration of phenolic extraction and subsequent reactions during winemaking, thereby providing valuable insights into process parameters affecting wine quality. However, their study does not incorporate sensory attributes or machine learning methods for quality prediction, which are significant aspects of our research.

C. COMPARISON WITH THE HOU ET AL. STUDY

Hou et al. [22] constructed a red wine rating credibility model using the Analytic Hierarchy Process (AHP) and performed cluster analysis to classify red grapes based on the evaluation results of red wine. They analyzed the correlation between the physical and chemical indicators of red grapes and red wine, concluding that aromatic substances play an essential role in red wine quality. Their study underscores the significance of sensory evaluation in determining wine quality and merges both sensory and physicochemical indicators for a comprehensive assessment.

Conversely, our study primarily concentrates on the physicochemical attributes of wine to predict quality using machine learning algorithms. While we acknowledge the significance of chemical indicators, we do not incorporate sensory attributes critical in Hou et al.'s study. Our methodology aims to deliver a more accessible and automated approach for quality assessment, utilizing machine learning to manage large datasets and identify key variables affecting wine quality.

The strength of the Hou et al. study resides in its comprehensive approach, combining sensory evaluation with statistical analysis to deliver a more holistic view of wine quality. However, it does not employ machine learning techniques for predictive modeling, which is a critical aspect of our research.

D. COMPARISON WITH THE COZZOLINO ET AL. STUDY

Cozzolino et al. [23] investigated the use of visible and infrared spectroscopy combined with chemometrics to measure phenolic compounds in grape and wine samples. Their study demonstrated that spectroscopic techniques could

simplify and reduce the analytical time for measuring a range of grape and wine analytes. Cozzolino et al. underscored the significance of phenolic compounds in determining the quality of red wines, as these compounds contribute to the color, taste, and mouthfeel of the wine.

Our study differs by focusing on the application of machine learning algorithms to predict wine quality based on physicochemical attributes rather than using spectroscopic techniques for phenolic measurement. While Cozzolino et al. provided a detailed methodology for phenolic measurement, our research aims to recognize and utilize pertinent variables that determine wine quality for predictive modeling.

The strength of the Cozzolino et al. study resides in its ability to deliver rapid and non-destructive analysis of phenolic compounds using advanced spectroscopic techniques. However, it does not incorporate machine learning for quality prediction, which is a critical aspect of our research.

E. COMPARISON WITH THE ARAPITSAS ET AL. STUDY

Arapitsas et al. [24] investigated the influence of sulfur dioxide (SO2) on wine flavanols and indoles concerning wine style and age. Their study focused on the chemical reactions between wine metabolites and SO2, recognizing and quantifying sulfonated derivatives of epicatechin, procyanidin B2, and various indoles in wines. They pointed out how these reactions affect wine quality, particularly in aged wines, where sulfonated flavanols and indoles become conspicuous markers.

In contrast, our study concentrates on recognizing the crucial variables that determine wine quality using machine learning algorithms. We do not specifically investigate the chemical interactions involving SO2 or the formation of sulfonated derivatives. Instead, we use a broader set of physicochemical attributes to predict wine quality, such as fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality.

The strength of the Arapitsas et al. study resides in its detailed observation of the impact of SO2 on wine chemistry and its implications on wine quality, particularly for aged wines. However, their study does not employ machine learning techniques for predictive modeling, which is a critical aspect of our research.

F. COMPARISON WITH THE STUDY BY LUO ET AL.

Luo et al. [25] investigated whether wine quality can be predicted by analyzing small volatile compounds. They collected and analyzed 157 commercial Shiraz wines from a competitive wine show over two years. They found significant correlations between specific volatiles and the panel's wine quality scores, although these correlations were not always consistent between years. Luo et al. used headspace solid-phase microextraction-gas chromatography-mass spectrometry (HS-SPME-GC-MS) to measure the volatile profiles of the wines. They suggested that while aroma determination is crucial, other factors like appearance and taste might be more influential in wine quality assessment.

Our study focuses on a broader set of physicochemical attributes for wine quality prediction using machine learning algorithms. While Luo et al. provided valuable insights into the relationship between volatile compounds and wine quality, our research aims to identify and utilize multiple physicochemical variables for a comprehensive predictive modeling approach.

The strength of Luo et al.'s study lies in its detailed analysis of volatile compounds and their impact on wine quality. However, they did not incorporate machine learning techniques for predictive modeling, a significant aspect of our research.

G. COMPARISON WITH THE STUDY BY STEIN ET AL.

Stein et al. [26] examined the quality of Cabernet Sauvignon wines determined by climatic attributes. They investigated how environmental conditions, such as rainfall, temperature, and solar radiation, influence the physicochemical properties of Cabernet Sauvignon wines. The research highlighted the importance of factors like total acidity, anthocyanins, pH, tannins, and total polyphenols for determining the sensory quality and aging potential of wines.

In contrast, our study focuses on identifying critical variables that determine wine quality using machine learning algorithms. While Stein et al. analyzed the impact of climatic conditions on wine quality, we employ a broader set of physicochemical attributes, such as fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality, to predict wine quality.

The strength of Stein et al.'s study lies in its detailed analysis of the relationship between climatic attributes and wine quality. However, they did not incorporate machine learning techniques for predictive modeling, a key aspect of our research.

H. COMPARISON WITH THE STUDY BY KURTANJEK

Kurtanjek [27] explored the application of causal artificial intelligence models to food quality data analysis. The study emphasized integrating theoretical field knowledge with process production, physicochemical analytics, and consumer organoleptic assessments. They used Bayesian networks and deep learning to infer causal relationships and intervention effects between process variables and consumer sensory assessments of food quality. The research highlighted the importance of addressing confounding effects in causal analysis and presented methodologies for predicting average causal effects (ACE) of process interventions on food quality.

In contrast, our study focuses on predicting wine quality using machine learning algorithms based on physicochemical attributes. While Kurtanjek's study applied causal AI models to infer relationships and intervention effects, we concentrate on identifying and utilizing critical physicochemical variables for predictive modeling. Our methodology leverages

machine learning to handle large datasets and improve precision in wine quality assessment.

The strength of Kurtanjek's study lies in its advanced causal modeling techniques, which provide insights into the causal effects of process variables on food quality. However, their study does not specifically focus on wine quality prediction using machine learning, which is the primary focus of our research.

I. COMPARISON WITH THE STUDY BY XU AND XU

Xu and Xu [28] conducted a study on the quality evaluation of Chinese red wine using a cloud model approach. Their study focused on sensory evaluation and the inherent uncertainties in sensory data. They utilized a cloud model to quantify the quality of wine based on sensory attributes, providing a way to handle the uncertainty and subjectivity inherent in sensory evaluations.

In contrast, our study utilizes machine learning algorithms to predict wine quality based on measurable physicochemical attributes rather than subjective sensory data. While Xu and Xu's cloud model addresses the uncertainty in sensory evaluations, our approach aims to provide an objective and consistent method for wine quality assessment.

The strength of Xu and Xu's study lies in its innovative use of a cloud model to address the complexities and uncertainties in sensory evaluation. However, it does not incorporate machine learning techniques for predictive modeling based on physicochemical data, which is a significant aspect of our research.

J. COMPARISON WITH THE STUDY BY ARCANJO ET AL.

Arcanjo et al. [29] conducted a study on the quality evaluation of red wines produced from Isabella and Ives grapes in Southern Brazil. They focused on analyzing physicochemical parameters, phenolic composition, and antioxidant activity. The study highlighted significant differences among wines in phenolic compound content, flavonoids, antioxidant activity, anthocyanin levels, and color parameters. The findings underscore the impact of the production region, grape variety, and enological practices on wine quality.

In contrast, our study employs machine learning algorithms to predict wine quality based on a comprehensive set of physicochemical attributes. While Arcanjo et al. focus on the detailed chemical analysis and antioxidant activity of specific grape varieties, our research aims to provide a broader predictive model for wine quality assessment using a diverse set of physicochemical variables.

The strength of Arcanjo et al.'s study lies in its thorough chemical analysis and the evaluation of antioxidant activity, which provide valuable insights into the health benefits and sensory characteristics of wines. However, their study does not utilize machine learning techniques for predictive modeling, which is a significant aspect of our research.

K. COMPARISON WITH THE STUDY BY PRESEROVA ET AL.

Preserova et al. [30] conducted a study on the phenolic profile and antioxidant activity in selected Moravian wines during the winemaking process using FT-IR spectroscopy. The study aimed to develop a rapid, robust method for monitoring total phenolic compounds (TPC) and total antioxidant activity (TAA) during different stages of wine production. They found significant variations in phenolic content and antioxidant activity among red, white, and rose wines, with red wines showing the highest levels of phenolics and antioxidant activity. The study also highlighted the changes in these compounds during different production stages, emphasizing the role of FT-IR spectroscopy combined with chemometrics as a reliable method for wine quality control.

In contrast, our study focuses on using machine learning algorithms to predict wine quality based on physicochemical attributes. While Preserova et al. utilize FT-IR spectroscopy to monitor phenolic compounds and antioxidant activity, our research leverages a broader set of physicochemical variables for predictive modeling.

The strength of Preserova et al.'s study lies in its methodological innovation, providing a rapid and non-destructive means of monitoring wine quality. However, their approach does not include machine learning techniques for predictive modeling, which is a key aspect of our research.

L. COMPARISON WITH THE STUDY BY SUPRIATNA ET AL.

Supriatna et al. [31] introduced an ensemble voting classifier for red wine quality classification using machine learning algorithms. Their study focused on enhancing classification accuracy by combining multiple models, including Random Forest and XGBoost, to leverage the strengths of each. The dataset used included various physicochemical attributes and quality ratings of red wines from Portugal's Vinho Verde region. They achieved an accuracy of 0.885 with their ensemble approach, significantly improving the classification performance compared to individual models.

In contrast, our study focuses on predicting wine quality based on a selected set of critical variables identified through feature importance analysis. While Supriatna et al. applied an ensemble method to improve accuracy, our approach aims to streamline the predictive process by focusing on key attributes that directly impact wine quality.

M. COMPARISON WITH THE STUDY BY SÁENZ-NAVAJAS ET AL.

Sáenz-Navajas et al. [32] investigated the effect of aroma perception on the taste and mouthfeel dimensions of red wines, correlating sensory and chemical measurements. The researchers used a rate-all-that-apply (RATA) sensory methodology with Spanish wine experts to evaluate 42 non-wooded red wines under two conditions: with and without aroma perception. The research focused on understanding how aroma influences mouthfeel and revealed that tannin

concentration and activity, along with pH, are significant predictors of mouthfeel dimensions related to dryness.

In contrast, our study employs machine learning algorithms to predict wine quality based on physicochemical attributes, not incorporating direct sensory evaluations or aroma analysis. While Sáenz-Navajas et al. provide valuable insights into sensory perceptions and their chemical correlations, our research focuses on leveraging machine learning for a more automated and scalable approach to wine quality assessment.

N. COMPARISON WITH THE STUDY BY OFOEDU ET AL.

Ofoedu et al. [33] conducted a comparative evaluation of the physicochemical, antioxidant, and sensory properties of domestic and foreign red wines to ascertain quality and authenticity. The study analyzed wines for parameters such as pH, total titratable acidity (TTA), total sugar, Brix, alcohol content, polyphenols, flavonoids, tannins, and antioxidant capacity using DPPH, FRAP, and TEAC assays. The sensory evaluation included attributes like color, taste, mouthfeel, flavor, and overall acceptability. The researchers found significant differences between domestic and foreign wines, with foreign wines generally exhibiting higher quality markers such as polyphenols, flavonoids, tannins, and antioxidant capacity.

In contrast, our study focuses on predicting wine quality using machine learning algorithms based on a comprehensive set of physicochemical attributes. While Ofoedu et al. perform a detailed analysis of the physicochemical and sensory properties, our research aims to leverage machine learning for predictive modeling, thereby offering a more automated and scalable approach to wine quality assessment.

O. COMPARISON WITH THE STUDY BY SEN ET AL.

Sen et al. [34] explored the combination of visible and mid-infrared (MIR) spectra for predicting the chemical parameters of wines. Their study used orthogonal partial least squares (OPLS) regression to analyze wine samples from twelve grape varieties across two harvest years. They aimed to improve the prediction of various chemical compounds in wines, including anthocyanins, total phenols, glycerol, the glycerol/ethanol ratio, malic acid, o-coumaric acid, and Â°Brix. The results demonstrated that combining visible and MIR spectra with multivariate methods improved the prediction of anthocyanins and total phenols compared to using MIR spectra alone.

In contrast, our study focuses on predicting wine quality using machine learning algorithms based on a comprehensive set of physicochemical attributes. While Sen et al. employed spectroscopic techniques to predict specific chemical parameters, our research leverages machine learning for broader predictive modeling of wine quality, making it more applicable for practical quality assessments in the wine industry.

1) STRENGTHS AND WEAKNESSES

Both Zhan et al., Tindal et al., Hou et al., Cozzolino et al., Arcanjo et al., Arapitsas et al., Luo et al., Stein et al., Kurtanjek, Xu and Xu, Preserova et al., Supriatna et al., Sáenz-Navajas et al., Ofoedu et al., and Sen et al. offer robust methodologies for understanding wine quality from different perspectives—machine learning models, mathematical modeling of phenolic extraction, sensory evaluation combined with statistical analysis, spectroscopic analysis of phenolic compounds, chemical interactions involving SO2, volatile compound analysis, climatic attributes, causal AI models, and cloud models for sensory evaluation. Our study combines elements of these approaches by utilizing machine learning to predict quality based on a variety of physicochemical attributes.

The primary strength of our study is its predictive capability, offering a practical tool for winemakers to assess wine quality based on easily measurable variables. However, it does not delve into the detailed chemical interactions and extraction processes during fermentation as explored by Tindal et al., incorporate sensory attributes as emphasized by Hou et al. and Sáenz-Navajas et al., employ spectroscopic techniques for phenolic measurement as used by Cozzolino et al. and Sen et al., investigate SO2 interactions as detailed by Arapitsas et al., focus on volatile compounds as investigated by Luo et al., analyze climatic attributes as investigated by Stein et al., apply advanced causal AI models as explored by Kurtanjek, address uncertainties in sensory evaluation as done by Xu and Xu, utilize FT-IR spectroscopy for rapid monitoring as demonstrated by Preserova et al., or provide a comparative analysis of domestic versus foreign wines as done by Ofoedu et al.

2) EFFECTIVE VARIABLES AND THEIR DEFINITIONS

Zhan et al. highlighted variables such as alcohol content, sulphates, total sulfur dioxide, and citric acid, which were chosen for their significant impact on wine quality. These variables were identified based on their strong correlation with wine quality, as shown through heatmap analysis.

Tindal et al. highlighted variables such as temperature, mixing regimes, and oxidation levels, which are critical in the phenolic extraction process. These variables are defined within the context of their impact on phenolic behavior, influencing wine color, flavor, and stability. The rationale behind their selection is based on their direct effect on the kinetic and spatial distribution of phenolics during fermentation.

Hou et al. focused on both physicochemical indicators (e.g., acidity, sugar content) and sensory attributes (e.g., aroma, taste) to evaluate wine quality. They emphasized the role of aromatic substances in determining quality and used cluster analysis to classify wine based on these attributes. The rationale behind their selection is to provide a comprehensive assessment that captures both chemical and sensory dimensions.

Cozzolino et al. highlighted the use of spectroscopic techniques (e.g., visible, near-infrared, mid-infrared) to measure phenolic compounds in grapes and wines. These variables were selected for their ability to provide rapid and non-destructive analysis of key phenolic compounds that influence wine quality.

Arcanjo et al. focused on phenolic composition, flavonoids, anthocyanins, and antioxidant activity. They selected these variables for their significant impact on wine quality, particularly in terms of sensory attributes like color, flavor, bitterness, and astringency, as well as health benefits due to antioxidant properties.

Arapitsas et al. highlighted the impact of SO2 on flavanols and indoles, emphasizing the formation of sulfonated derivatives and their influence on wine quality, particularly in aged wines. These variables were selected to understand the chemical changes that occur during wine aging and the role of SO2 in modulating wine quality.

Luo et al. highlighted volatile compounds such as linalool, hexyl acetate, and 2-phenylethyl acetate, which were chosen for their significant correlations with wine quality scores. These variables were identified based on their impact on the aroma profile of the wine, which is a critical factor in quality assessment.

Stein et al. highlighted climatic attributes such as rainfall, temperature, and solar radiation, which were chosen for their significant impact on the physicochemical properties of wines. These variables were identified based on their influence on total acidity, anthocyanins, pH, tannins, and total polyphenols, which are crucial for wine quality.

Kurtanjek highlighted process variables such as temperature, pH, and fat content in dairy products, as well as alcohol and volatile acidity in wines, using Bayesian networks and deep learning to infer causal relationships and intervention effects. These variables were selected based on their significant impact on food quality and consumer sensory assessments.

Xu and Xu focused on sensory attributes such as taste, aroma, and appearance, using a cloud model to quantify these subjective measures. These variables were selected to address the uncertainty and subjectivity inherent in sensory evaluations.

Preserova et al. focused on total phenolic compounds (TPC) and total antioxidant activity (TAA) using FT-IR spectroscopy. These variables were selected for their significant impact on the quality and health benefits of wine. The rationale behind their selection is to provide a rapid, non-destructive method for monitoring changes in these compounds during the winemaking process.

Supriatna et al. focused on a comprehensive set of physicochemical attributes, including fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, and alcohol content. These variables were selected for their relevance to wine quality and ability to be measured consistently.

Sáenz-Navajas et al. emphasized the relationship between sensory attributes (e.g., mouthfeel dimensions like dry, silky, sticky, grainy, prickly, and oily) and chemical measurements (e.g., tannin concentration, tannin activity, pH, ethanol content, and spectroscopic measures). The rationale behind selecting these variables is to understand how different sensory dimensions are influenced by specific chemical components.

Ofoedu et al. focused on physicochemical properties such as pH, total titratable acidity (TTA), total sugar, Brix, alcohol content, polyphenols, flavonoids, tannins, and antioxidant capacity using DPPH, FRAP, and TEAC assays. These variables were selected to provide a comprehensive comparison of domestic and foreign red wines in terms of quality and authenticity.

Sen et al. highlighted the combination of visible and MIR spectra to predict chemical parameters like anthocyanins, total phenols, glycerol, the glycerol/ethanol ratio, malic acid, o-coumaric acid, and °Brix. These variables were selected for their ability to improve predictive accuracy using multivariate methods.

In our study, effective variables included chemical properties of the wine (e.g., acidity, sugar content) chosen for their significant impact on the overall quality of the wine. The rationale behind their selection is to capture critical chemical dimensions that influence wine quality.

3) PERFORMANCE COMPARISON

Our proposed method using machine learning algorithms shows promise in providing more precise and comprehensive wine quality assessments compared to the approaches reviewed by Zhan et al., Tindal et al., Hou et al., Cozzolino et al., Arcanjo et al., Arapitsas et al., Luo et al., Stein et al., Kurtanjek, Xu and Xu, Preserova et al., Supriatna et al., Sáenz-Navajas et al., Ofoedu et al., and Sen et al. While Zhan et al. provide a comparative analysis of machine learning techniques, Tindal et al.'s models are excellent for understanding the fundamental processes during fermentation, Hou et al.'s study provides a holistic view by integrating sensory evaluation, Cozzolino et al.'s research offers rapid phenolic measurement using spectroscopy, Arcanjo et al. focus on antioxidant activity and phenolic composition, Arapitsas et al. highlight the impact of SO2 on wine chemistry, Luo et al. focus on the influence of volatile compounds, Stein et al. analyze the impact of climatic attributes, Kurtanjek applies advanced causal AI models, Xu and Xu address uncertainties in sensory evaluation, Preserova et al. utilize FT-IR spectroscopy for rapid monitoring of phenolic compounds and antioxidant activity, Supriatna et al. demonstrate the effectiveness of ensemble classifiers, Sáenz-Navajas et al. explore cross-modal interactions between aroma and mouthfeel, Ofoedu et al. provide a comprehensive evaluation of domestic and foreign wines, and Sen et al. combine visible and MIR spectra for improved chemical parameter prediction. Our approach excels in practical application for quality control

and prediction in winemaking. The ability to analyze large datasets and adjust for numerous variables simultaneously gives our method a potential edge in real-world winemaking scenarios.

IV. PROPOSED MODEL

A. DATASET

This paper utilizes the renowned datasets from the University of California Irvine (UCI) Machine Learning Repository [35], with specific emphasis on the Wine Quality Data Set [36]. This dataset encompasses entries from two subsets of Vinho Verde wines from northern Portugal, which are categorized into red and white types. Our analysis will primarily focus on the subset consisting of 1,599 red wine samples, whose physicochemical features are presented in Table 2. Each sample is analyzed based on 11 distinct physicochemical traits and assigned a quality rating ranging from 0 (indicating very poor quality) to 10 (indicating excellent quality), each rating represented as an integer.

TABLE 2. Units of physicochemical features in wine quality dataset.

Feature	Unit
Fixed Acidity	g(tartaric acid)/dm^3
Volatile Acidity	g(acetic acid)/dm^3
Citric Acid	g/dm^3
Residual Sugar	g/dm^3
Chlorides	g(sodium chloride)/dm^3
Free Sulfur Dioxide	mg/dm^3
Total Sulfur Dioxide	mg/dm^3
Density	g/cm^3
pH	unitless
Sulphates	g(potassium sulphate)/dm^3
Alcohol	% volume
Quality	score (0-10)

B. STRATEGIES TO MITIGATE OVERFITTING

Overfitting is a common challenge in machine learning where a model performs well on the training data but performs poorly on unseen data. To address this, we propose two strategies: ensuring the interpretability of features and achieving better fitting and generalization capabilities.

1) ENSURING THE INTERPRETABILITY OF FEATURES

Interpretable features can significantly enhance a model's ability to generalize by providing insights into the relationship between the input variables and the target variable. Interpretability goes beyond just revealing feature importance; it involves understanding relationships between multiple features and how they interact to influence predictions [37]. This deeper exploration can unearth synergies or antagonisms between features, hence providing a more comprehensive understanding of the model's behavior [37]. Characterizing feature importance assists in explaining the primary factors affecting model outcomes, thus facilitating the feature extraction processes [38].

Various methods and algorithms such as Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanation have been employed to enhance the interpretability of machine learning models in different domains, including healthcare and finance [39], [40]. These techniques help in attributing output values to specific features, enabling researchers to analyze and interpret model decisions effectively [39], [40]. In our study, we focus on the following features from the wine quality dataset:

- **Fixed Acidity**: The level of non-volatile acids present in wine.
- **Volatile Acidity**: The amount of acetic acid in wine, which can lead to an unpleasant vinegar taste.
- **Citric Acid**: This contributes to the freshness and flavor of the wine.
- **Residual Sugar**: The amount of sugar remaining after fermentation stops.
- **Chlorides**: The amount of salt present in the wine.
- **Free Sulfur Dioxide**: The free form of SO2 acts as an anti-microbial and antioxidant.
- **Total Sulfur Dioxide**: The total amount of SO2 in both free and bound forms.
- **Density**: The density of the wine is related to its alcohol and sugar content.
- **pH**: A measure related to the acidity of the wine.
- **Sulphates**: These contribute to the wine's antimicrobial and antioxidant properties.
- **Alcohol**: The alcohol content of the wine.

By understanding these features, we can craft more interpretable models that highlight the significant predictors of wine quality.

2) ACHIEVING BETTER FITTING AND GENERALIZATION CAPABILITIES

Multiple techniques can help improve model's fitting and generalization capabilities:

- **Cross-Validation**: Implementing k-fold cross-validation aids in evaluating the model's performance on different subsets of the data, thereby providing a more robust estimate of its accuracy. This process is repeated multiple times to ensure the robustness and reliability of the model [41]. Cross-validation is crucial in preventing overfitting, a common issue in machine learning where the model performs well on the training data but fails to generalize to new, unseen data [42].
  Various forms of cross-validation exist, such as k-Fold Cross-Validation, where the data is divided into k subsets, and the model is trained and tested k times, each time using a different subset as the test set [43]. Another approach is leave-one-out cross-validation, where each data point is used as a test set once while the model is trained on the remaining data points [44]. These techniques aid in optimizing model hyperparameters, selecting the best features, and assessing the model's performance on unseen data [45].
  Cross-validation is not specific to any type of machine learning algorithm but is a widely accepted practice

in the field. It is extensively used in various domains such as in healthcare for disease prediction [46], image classification [44], environmental studies [47], and material science [45].

- **Ensemble Methods**: Methods like Random Forest and XGBoost aggregate multiple models to improve predictive performance and robustness. Ensemble methods in machine learning refer to supervised learning models that combine different individual models to improve predictive performance. These methods integrate diverse data sources and computational functions to enhance the accuracy and robustness of predictions [48]. Ensemble methods have been applied in various domains, such as healthcare for cancer prognosis and diagnosis [48], agriculture for crop yield prediction [49], and activity recognition using smartphone sensors [50].
  Among the basic ensemble methods, we have majority voting, where multiple models combine their predictions to decide the final output [51]. Other ways of creating ensembles include Basic Ensemble Method (BEM), Generalised Ensemble Method (GEM), and stacked generalised ensembles [49]. The goal of these methods is to leverage the strengths of different models to improve the overall performance.
  Ensemble learning addresses key challenges in supervised learning that include the statistical problem (limited training data), the computational problem (complex model search space), and the representational problem (model compatibility) [52]. By combining multiple models, an ensemble method can mitigate these issues and achieve higher accuracy compared to individual algorithms [53].
  Furthermore, ensemble methods can be tailored to specific applications, such as using stacking ensemble learning for predicting renal cell carcinoma with a high degree of accuracy [54]. Different ensemble strategies, such as iterative ensemble learning and attention network ensemble learning, can be employed to integrate information from diverse feature types and improve predictive modelling [55].
- **Hyperparameter Tuning**: Optimizing the model's hyperparameters ensures that it is neither too simple nor too complex, providing a balance between bias and variance. Hyperparameter tuning in machine learning involves selecting the optimal values for the hyperparameters within these models to enhance their performance [56]. Hyperparameters influence the learning process of machine learning models, and tuning them involves identifying the best values to make this learning process more effective [57]. This optimization aims to enhance the predictability of machine learning algorithms while minimizing the consumption of computational resources [58]. The automation of hyperparameter tuning reduces the manual effort often required to explore various configuration settings, leading to improved accuracy and reproducibility of machine learning models [59].
  Hyperparameter optimization is essential for deploying effective machine learning algorithms, as it significantly impacts their predictive power and generalization ability [60]. The process involves determining the best values for each hyperparameter in a machine learning algorithm to achieve the most effective results [61]. Proper tuning of hyperparameters can result in substantial performance gains, depending on the algorithm used [62]. Hyperparameters control the behavior of training algorithms and directly affect the performance of machine learning models [63].

By applying these strategies, our aim is to develop a model that fits the training data well, but also generalizes effectively to unseen data, thus providing reliable predictions.

C. INPUT VARIABLES

To comprehensively assess the impact of different factors on wine quality, our plan includes exploring all possible combinations of the 11 variables (excluding the quality variable itself) as inputs for our models. These 11 variables are acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH value, sulfates, and alcohol content. This approach entails calculating the total number of subsets of these 11 variables, enabling us to methodically evaluate how each combination of variables affects the quality of Portuguese wine.

When finding x-variable subsets from a set of n elements, we use the general formula for combinations:

$$C(n, x) = \frac{n!}{x!(n-x)!} \tag{1}$$

Here:

- $n!$ denotes the factorial of n, which is the product of all positive integers up to n,
- $x!$ denotes the factorial of x,
- $n - x$ represents the difference between n and x.

This formula calculates the number of different ways to choose x elements from a set of n elements, without considering the order of their selection.

For instance, with $n = 11$ and $x = 2$, there is a total of 55 2-variable subsets of the variables:

- (acidity, volatile acidity), (acidity, citric acid), …, (acidity, alcohol content),
- …, (volatile acidity, citric acid), …, (volatile acidity, alcohol content),
- …, (citric acid, residual sugar), …, (free sulfur dioxide, total sulfur dioxide),
- …, (sulfates, alcohol content).

Similarly, we will calculate subsets for all counts of variables, from $x = 1$ to $x = 11$. These subsets represent all potential pairings of variables that could be explored to assess their impact on the quality of Portuguese wine.

The subsequent table illustrates the number of ways to select various numbers of elements:

TABLE 3. The number of ways to choose elements.

Number of Elements Chosen	Number of Ways
To choose 1 element	11 ways
To choose 2 elements	55 ways
To choose 3 elements	165 ways
To choose 4 elements	330 ways
To choose 5 elements	462 ways
To choose 6 elements	462 ways
To choose 7 elements	330 ways
To choose 8 elements	165 ways
To choose 9 elements	55 ways
To choose 10 elements	11 ways
To choose all 11 elements	1 way
Total	2047 ways

The total number of ways to choose from 1 to 11 elements from a set of 11 elements equals 2,047. To show that the total number of subsets (including the empty set and the set itself) of a set with n elements equals to 2^n, one might consider a simple argument based on the concept of choice for each element in the set: each element can either be included in a subset or not. Since there are two choices for each of the n elements, the total number of subsets amounts to 2^n. After excluding the empty set, the solution becomes $2^n - 1 = 2047$.

D. EXPLORING THE DIVERSITY OF WINE QUALITY AND THE ACCURACY OF RANDOM PREDICTIONS

We investigated the number of distinct quality values within the original wine data and analyzed the outcomes of random predictions. The wine quality data contains six distinct values: 3, 4, 5, 6, 7, and 8.

If we were to divide the dataset into samples of 1280 and 320 using an 80%-20% ratio and then make random guesses for each quality in the 320 subset, we could calculate the expected number of correct guesses. With 6 distinct quality values, the probability of a correct guess for any given wine is 1/6. This probability suggests that we could expect, on average, 53 correct guesses out of 320.

$$\text{Exp. corr. guesses} = \text{Sample size} \times \text{Prob. of correct guess} \tag{2}$$

$$\text{Expected correct guesses for } 320 = 320 \times \frac{1}{6} \tag{3}$$

$$\text{Expected correct guesses for } 320 = \frac{320}{6} \approx 53.33 \tag{4}$$

Hence, we would expect to guess correctly approximately 53 times out of 320 random guesses. If any machine learning algorithm's accuracy is estimated to be close to 1/6, these predictions would be considered statistically insignificant.

To evaluate the statistical significance of the mean squared error (MSE) derived from machine learning algorithm predictions on a dataset of 320 wine samples, we need to calculate the expected MSE when guessing randomly among the six distinct quality values in the wine dataset.

Mean Squared Error (MSE) computation for random guesses requires evaluating the expected squared difference between the randomly guessed quality values and the actual quality values. Given our entirely random guesses, the expected squared error for a single prediction would be the average of the squared differences between a guessed value and all potential actual values—assuming a uniform distribution of quality values.

Let's calculate the expected MSE under these simplified assumptions:

$$E = n \times p \tag{5}$$

where:

- E is the expected number of correct guesses.
- n is the number of samples or guesses made (320 in this case).
- p is the probability of a correct guess, $\frac{1}{6}$.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

where:

- n is the number of samples.
- y_i is the actual value for the i-th sample.
- $\hat{y}_i$ is the predicted value for the i-th sample.

$$\text{Expected Squared Difference} = \frac{1}{|Q|^2}\sum_{q_1 \in Q}\sum_{q_2 \in Q}(q_1 - q_2)^2 \tag{7}$$

where:

- Q is the set of all possible quality values: 3, 4, 5, 6, 7, 8.
- $|Q|$ is the number of elements in Q (6 in this case).

Following these calculations:

$$\text{Expected Sq. Difference} = \frac{1}{36}\sum_{q_1 \in Q}\sum_{q_2 in Q}(q_1 - q_2)^2 \approx 5.83 \tag{8}$$

This value represents the Mean Squared Error (MSE) expected when making random guesses for the quality values, assuming those guesses are uniformly distributed across the set Q. The expected MSE for our random guesses, given a uniform distribution of actual quality values, would be approximately 5.83. This value represents the average squared difference between the guessed quality values and the actual quality values across all possible guesses and actual values. This information is vital as we delve deeper into the accuracy and efficacy of machine learning predictions in subsequent chapters.

V. METHODOLOGY

Figure 1 illustrates our methodology, which involves iterating through 2,047 combinations of variable subsets using Logistic Regression, Support Vector Machines (SVM), Decision Tree Classifiers, and Random Forest Models to determine

their accuracies. For each subset, we divide our wine dataset into a training set and a testing set, allocating eighty percent of the data randomly to the training set and the remaining twenty percent to the testing set. The training dataset is used to train the models, while the testing dataset is used to assess their performance. We employ various machine learning models, including Logistic Regression, Support Vector Machine (SVM), Decision Tree Classifier, and Random Forest Models, to analyze the dataset. The performance of each model is evaluated by calculating its accuracy and Mean Squared Error (MSE).

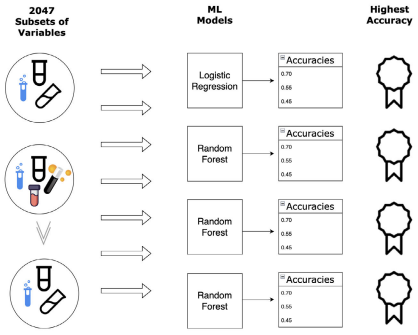


FIGURE 1. Our methodology.

A. PSEUDO CODE

1. ***Import necessary libraries***
2. ***Set parameters***
CV_Folds: Number of cross-validation folds
number_of_tests: Number of tests to perform
param_grids: Dictionary containing parameter grids for Random Forest, Logistic Regression, SVM, and Decision Tree models
thresholds: Minimum acceptable score for each model type
3. ***Define function*** *evaluate_model_with_grid_search*:
Inputs: df_subset (subset of features), model (the ML model), target (the target variable), param_grid (grid of parameters for the model)
Process:
Split data into training and test sets
Normalize feature values
Initialize and fit GridSearchCV with model and parameters
Make predictions on test set
Calculate and return micro f1 score, mean squared error, and best parameters
4. ***Load the dataset and preprocess:***
Read dataset
Separate features and target
Define models dictionary with instances of Random Forest Classifier, Logistic Regression, SVC, and DecisionTreeClassifier
5. ***Main Loop:***
Iterate over each model in models dictionary
for each model **do**
for each test in number_of_tests **do**
for each combination of feature subsets **do**
Use *evaluate_model_with_grid_search* to get model performance metrics
if score exceeds predefined threshold **then**
Update best score, parameters, and combination
Save test number, model name, best feature combination, score, MSE, and best parameters to a CSV file
end if
end for
end for
end for
end for
6. ***Output completion message***

In our paper, we enhanced the default machine learning accuracies by approximately 1-2% through the tuning of hyperparameters. By pushing the boundaries of ML algorithms and rigorously analyzing our results at these limits, we identified the most robust solutions. Here's an overview of the key strategies we employed and the benefits they offered:

- **Hyperparameter Tuning with GridSearchCV:** We employed GridSearchCV to systematically explore a wide range of hyperparameter combinations for several models, including Random Forest Classifier, Logistic Regression, SVC, and DecisionTreeClassifier. This methodical approach allowed us to pinpoint the optimal settings for each model, contributing to the accuracy improvements.
- **Feature Selection:** By examining all possible combinations of features from the dataset, we were able to identify the most predictive subsets for each model. This improved model accuracy and provided insights into which features were most significant for predicting outcomes, enhancing our understanding of the data.
- **Normalization:** We applied MinMaxScaler to normalize the feature values, ensuring that our models were not biased by the scale of the data.
- **Evaluation Metrics:** Focusing on both the f1 score and mean squared error allowed us to have a comprehensive view of model performance beyond simple accuracy measures.
- **Iterative Testing:** By running a specified number of tests for each model and feature combination, we were able to iteratively refine our models and hyperparameters. This approach also enabled us to estimate the stability and robustness of each model under different configurations.
- **Result Analysis and Reporting:** For each model and test iteration, we recorded the best feature combination, score, mean squared error, and the corresponding hyperparameters. This detailed record-keeping facilitated deeper analysis of what worked best and why, allowing us to extract actionable insights and best practices for future research.

- **Performance Thresholds:** We set minimum performance thresholds to focus our efforts on the most promising model configurations.

Through these strategies, we deepened our understanding of model behavior at its performance limits.

B. ESTABLISHING THE OPTIMAL ML MODEL FOR WINE QUALITY ANALYSIS

In this part of our research, we compared the Random Forest, Logistic Regression, SVM, and Decision Tree models to select the one with the highest accuracy for subsequent critical variable analysis. We conducted five tests using the Random Forest, Decision Tree, Support Vector Machine, and Logistic Regression models, each with 2047 subsets. In our research, we included the highest accuracies from the first 10 results of each test. The results are given in Appendix A for Random Forest, in Appendix B for Decision Tree, in Appendix C for Logistic Regression, and in Appendix D for Support Vector Machine.

Our study conclusively demonstrates the superiority of the Random Forest model as the most effective predictor of Portuguese red wine quality, assessed on a scale from 0 to 10. Contrary to our initial assumption that alternative machine learning algorithms might excel in smaller subsets, the Random Forest model uniformly outperformed other models, including Logistic Regression, which is traditionally favored for its simplicity and effectiveness with fewer variables. Across all tested subsets, the Random Forest model consistently emerged as the top performer, leading us to the definitive conclusion that further exploration of alternative machine learning algorithms is unnecessary when the Random Forest provides superior results.

C. PRECISION-RECALL AND ROC ANALYSIS FOR RANDOM FOREST

To further demonstrate the applicability of our results, we included Precision-Recall (PR) and Receiver Operating Characteristic (ROC) figures specific to the wine quality dataset. For this part of the study, the target variable, 'quality', was binarized to classify wines as ''good'' (quality >= 7) or ''not good'' (quality < 7). The PR and ROC curves provide visual evidence of the model's performance. These analyses were performed using the Random Forest Classifier.

1) MATHEMATICAL BACKGROUND

a: PRECISION-RECALL CURVE

- **Precision** (Positive Predictive Value) is defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

where TP is the number of true positive predictions and FP is the number of false positive predictions.

- **Recall** (Sensitivity or True Positive Rate) is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

where TP is the number of true positive predictions and FN is the number of false negative predictions.

- The Precision-Recall curve is a plot of Precision (y-axis) versus Recall (x-axis) for different threshold values. A curve closer to the top-right corner indicates better performance, with high precision and high recall.

b: ROC CURVE

- **True Positive Rate** (Recall or Sensitivity) is defined as:

$$\text{TPR} = \frac{TP}{TP + FN} \tag{11}$$

- **False Positive Rate** is defined as:

$$\text{FPR} = \frac{FP}{FP + TN} \tag{12}$$

where TN is the number of true negative predictions.

- The ROC curve in Figure 3 is a plot of the True Positive Rate (y-axis) versus the False Positive Rate (x-axis) for different threshold values. A curve closer to the top-left corner indicates better performance. The Area Under the Curve (AUC) provides a single measure of overall model performance, with values closer to 1 indicating better performance.

2) PRECISION-RECALL CURVE

The Precision-Recall curve in Figure 2 plots Precision against Recall for different threshold values. Precision is the ratio of true positive predictions to the sum of true positive and false positive predictions, indicating the accuracy of positive predictions. Recall (Sensitivity) is the ratio of true positive predictions to the sum of true positive and false negative predictions, indicating the coverage of actual positive instances. Precision, also known as positive predictive value, measures the proportion of correctly predicted positive instances among all instances predicted as positive, while recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances out of all actual positive instances [64]. This curve is particularly useful when dealing with imbalanced datasets, where one class significantly outnumbers the other, as it provides a more informative picture of model performance compared to ROC curves in such scenarios [65]. The Precision-Recall curve is especially beneficial for evaluating classification performance in situations where the data is heavily imbalanced in favor of the negative class, such as in information retrieval systems [66].

The area under the Precision-Recall curve (AUPRC) is a common metric used to assess model performance, especially in cases involving rare events, as it is not dependent on model specificity and is suitable for imbalanced datasets [67]. This metric, along with other evaluation metrics like accuracy, F1 score, and the area under the ROC curve, provides a comprehensive assessment of a model's predictive capabilities [68]. The Precision-Recall curve is often used in conjunction with other metrics like the receiver operating characteristic (ROC)

curve to evaluate prediction performance in various domains, including healthcare for tasks such as predicting stroke events or classifying skin lesions [69], [70].

A curve closer to the top-right corner indicates better performance, with high precision and high recall. For this wine dataset, high precision means that most of the wines predicted as ''good'' are actually ''good,'' and high recall means that most of the ''good'' wines are correctly identified.

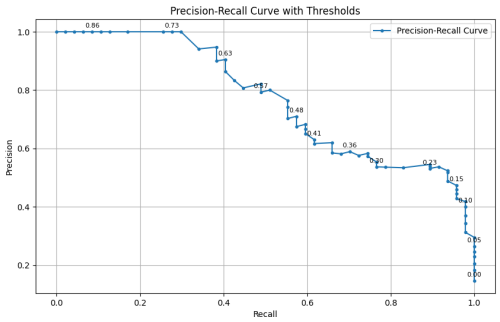


FIGURE 2. Precision-Recall Curve for Wine Quality Prediction.

Analysis: The Precision-Recall curve in Figure 2 shows that the model maintains high precision and recall at lower thresholds, indicating that the Random Forest Classifier effectively identifies good quality wines with minimal false positives. However, as the recall increases, precision decreases, suggesting that the model encounters more false positives at higher thresholds.

3) ROC CURVE

The Receiver Operating Characteristic (ROC) curve in Figure 3 plots the True Positive Rate (Recall) against the False Positive Rate for different threshold values. The False Positive Rate is the ratio of false positive predictions to the sum of false positive and true negative predictions. A curve closer to the top-left corner indicates better performance, with a high true positive rate and a low false positive rate.

The Receiver Operating Characteristic (ROC) curve is a fundamental tool in machine learning for evaluating the performance of classification models, particularly in diagnostic medicine studies [71]. The ROC curve plots the true positive rate against the false positive rate as the discrimination threshold varies, providing a visual representation of a model's ability to distinguish between classes [72]. It is widely used in biomedical research to assess the performance of diagnostic tests, where the area under the ROC curve (AUC) is a preferred measure as it does not depend on specific discrimination thresholds [73].

ROC curves are not limited to binary classification tasks but can also be extended to regression problems, survival analysis, and outlier detection [74], [75] [76]. They have been applied in various fields such as biology, medicine, and computer science to assess classifier performance and understand the trade-off between sensitivity and specificity [77], [78]. ROC curves have been utilized in the context of visual search models and in the evaluation of virtual screening and docking performance [79], [80].

ROC curves can be used to compare different algorithms or models by analyzing the area under the curve (AUC) and assessing discrimination capabilities [81], [82].

For this wine dataset, a high True Positive Rate means that most of the ''good'' wines are correctly identified, and a low False Positive Rate means that few ''not good'' wines are incorrectly classified as ''good.'' The Area Under the Curve (AUC) for the ROC curve provides a single measure of overall model performance, with values closer to 1 indicating better performance.

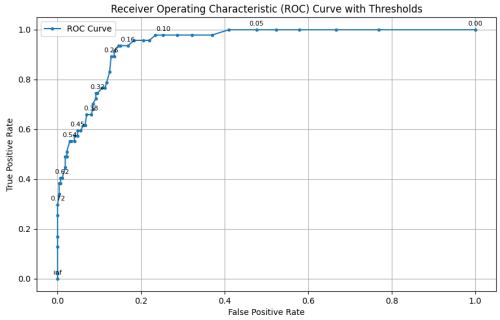


FIGURE 3. ROC curve for wine quality prediction.

Analysis: The ROC curve in Figure 3 indicates excellent model performance, with the Random Forest Classifier achieving a high true positive rate and a low false positive rate. The curve's proximity to the top-left corner and the high AUC value suggest that the model is highly effective in distinguishing between good and not good quality wines.

VI. REFINING THE MODEL: SINGLE VARIABLE ANALYSIS FOR WINE QUALITY

Upon examining Table 4, we observe that even single variables exhibit surprisingly high accuracy. One might anticipate their accuracies to be around 1/6, yet they significantly surpass this expectation. This raises a question: does their performance deem them critical for evaluating wine quality? It might seem that accuracies exceeding 1/6 indicate critical importance, but this is not necessarily the case. To grasp the underlying dynamics, an analysis of the wine quality data is essential.

The histogram in Fig 4 reveals an uneven distribution of test samples across different wine qualities. The learning model identifies a higher probability of wine samples being rated 5, 6, or 7, even when trained with variables that are ostensibly unrelated. In scenarios where the model is uncertain, it tends to predict a quality of 5, achieving accuracy superior to random guessing. A purely random guess would correctly predict more than 53 cases (approximately 53.33 qualities), whereas our model impressively forecasts around 160 correct guesses, all while maintaining a remarkably low Mean Squared Error (MSE) of about 0.6. This phenomenon was further elaborated in the section 'Exploring

TABLE 4. Average scores and MSE of single variables for random forest model across 5 tests.

Variable	Average Score	Average MSE
Alcohol	0.545625	0.664375
Sulphates	0.484375	0.760625
Total Sulfur Dioxide	0.482500	0.760625
Density	0.476875	0.913125
Free Sulfur Dioxide	0.473125	0.924375
Volatile Acidity	0.471875	0.845625
Citric Acid	0.460625	0.860625
Fixed Acidity	0.460000	0.945000
Residual Sugar	0.439375	0.915000
pH	0.433750	0.913750
Chlorides	0.423750	0.962500

the Diversity of Wine Quality and the Accuracy of Random Predictions' of our document, where we demonstrate that the MSE of random guesses, not based on a quality scale of 0 to 10 but on the actual observed quality levels of 3, 4, 5, 6, 7, and 8, can reach as high as 5.83.

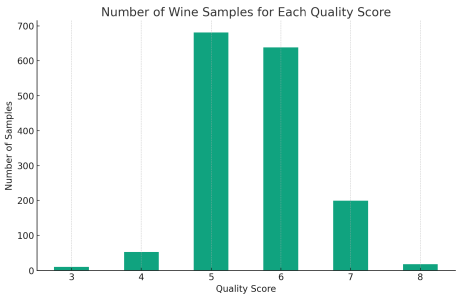


FIGURE 4. Histogram for wine quality.

The Random Forest Model is an ensemble learning method used for classification and regression tasks. Random Forest is a renowned machine learning algorithm introduced by Breiman in 2001 [83]. It aggregates multiple decision trees into a "forest" to improve prediction accuracy and combat overfitting [83]. The Random Forest model is commonly utilized for regression and nonparametric classification tasks based on the decision tree algorithm [84]. It constructs a series of binary trees using recursive partitioning, enabling it to effectively handle nonlinear data [85]. The model randomly selects training datasets to create numerous classification and regression trees, with predictions being generated by consolidating the outcomes from each tree [86].

Random Forest models are recognized for their adaptability and predictive capabilities [87]. They are particularly well-suited for medical data modeling and have demonstrated effectiveness in accurately modeling biomedical data for various applications [88]. The algorithm excels at capturing nonlinear relationships in data, making it suitable for intricate datasets with complex relationships [89]. Random Forest models find applications in various fields such as genetics, environmental science, and finance due to their robustness and efficiency [90], [91] [92]. It operates by constructing a multitude of decision trees during the training process and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The model divides the data into subsets to train each tree, ensuring that the trees are diverse and reduce overfitting.

The process is as follows:

1) **Bootstrap Aggregating (Bagging):** Random Forest starts by performing bootstrap aggregating, or bagging, where it randomly selects a subset of the training data with replacement to train each tree. This process ensures that each tree learns from a different portion of the data, making the model more robust.
2) **Building Decision Trees:** Each tree in a Random Forest grows by making decisions based on the values of different features in the data. A decision (or split) is made at each node of a tree, dividing the data into two subsets. This process continues recursively, forming a tree structure with nodes and leaves.
   - *Nodes:* Each node represents a condition on a single feature, designed to split the dataset into two so that similar response values end up in the same set.
   - *Leaves:* Leaves represent the final outcomes or decision results. In classification tasks, a leaf represents a class label; in regression tasks, it represents a continuous outcome.
3) **Random Feature Selection:** Unlike a single decision tree, Random Forest introduces more randomness when it splits nodes by selecting a random subset of the features. This strategy makes the trees more diverse and leads to more robust overall predictions.
4) **Aggregating Trees:** After training, predictions are made by aggregating the predictions of the ensemble of trees. For classification tasks, this typically means taking the majority vote among all trees. For regression tasks, it means taking the average of the predictions.

In summary, the Random Forest Model creates a 'forest' of diverse trees to make predictions, thereby reducing the risk of overfitting and improving prediction accuracy.

Therefore, predictions made by a Random Forest, even when trained with the least important variable, are not random. The model is informed beforehand that the quality of the collected Portuguese wines typically ranges between 5 and 7. Given this, how can we determine which variables are critical when even single variables, or no variables at all, lead to high accuracy?

The Random Forest Model understands the probabilities of wine quality being 3, 4, 5, 6, 7, and 8, as detailed in Table 5. However, its predictive capability exceeds mere probability matching. We will demonstrate that while mere guessing based on these probabilities yields an accuracy of around 0.35, the Random Forest's predictions fluctuate around an accuracy of 0.40. This improvement is due to the model's ability to discern patterns in the output variable that transcend simple probabilistic distributions.

TABLE 5. Probabilities of wine quality.

Quality	Probability (%)
3	0.63
4	3.31
5	42.59
6	39.90
7	12.45
8	1.13

These percentages indicate the distribution of wine quality ratings in the dataset.

To illustrate our argument that machine learning algorithms can predict the quality of wine with high accuracy even when the input variable is unrelated, simply by leveraging the probabilities of the outcomes, we have crafted a concise Python script. This script, depicted in Figure 5, generates numbers (output_a) based on the precise probabilities of our dataset, while simultaneously making guesses (guess_b) with identical probabilities. Upon executing this script 320 times, we observe accuracies that align closely with those of machine learning models, approximately 0.35. Consequently, this empirically derived accuracy of 0.35 serves as a foundational benchmark for our predictions, substantiating the capability of machine learning algorithms to forecast outcomes effectively through statistical probabilities inherent in the output, independent of direct relationships with input variables.

```python
import numpy as np

# Define the probabilities and numbers according to the dataset
probabilities = [0.0063, 0.0331, 0.4259, 0.399, 0.1245, 0.0113]
numbers = [3, 4, 5, 6, 7, 8]

# Ensure the probabilities sum to 1
prob_sum = sum(probabilities)
if prob_sum != 1.0:
    probabilities = [p/prob_sum for p in probabilities]

# Initialize the counter for coincidences (accurate guesses)
coincidences = 0

# Iterate 320 times
for _ in range(320):
    # Machine A output (or the "true" number based on dataset probabilities)
    output_a = np.random.choice(numbers, p=probabilities)
    # Machine B guess (guessing numbers with the same probabilities)
    guess_b = np.random.choice(numbers, p=probabilities)

    # Increment coincidences if guess matches the output
    if output_a == guess_b:
        coincidences += 1

# Calculate and print the accuracy
accuracy = coincidences / 320
print(f"Accuracy: {accuracy}")
```

FIGURE 5. Estimating machine learning accuracy through probability matching.

In Table 6, we observe the accuracy outputs of the Python script in Fig 5, which guesses the qualities by solely considering the given probabilities.

TABLE 6. Accuracy values.

Test#	Accuracy
1	0.3625
2	0.340625
3	0.3375
4	0.33125
5	0.340625
6	0.334375

- Without the application of any intelligent method, the baseline predictive accuracy is approximately 0.16.
- Incorporating human intelligence into the process increases accuracy to around 0.35.
- Utilizing machine learning algorithms without any variables yields a baseline accuracy of about 0.40. Setting the `random_state` variable to 42 for the sake of repeatability in our experiments, we achieve an accuracy of exactly 0.4062.

The results are summarized in Table 7.

TABLE 7. Comparative predictive accuracies.

Method	Predictive Accuracy
No intelligence applied	0.16
Probability matching	0.35
ML algorithms (baseline)	0.40

Indeed, once we grasp what we're working with, identifying the critical variables becomes quite straightforward. The critical variables would be those that elevate the base accuracy levels from 0.40 to the higher accuracy levels of 0.75. Although accuracy levels of 0.40 are also notable, they are solely based on the predictive capability of the ML models by observing the pattern of the output, not the influence of individual input variables. The improvement in accuracy from 0.40 to 0.75 is attributable to the combination of input variables.

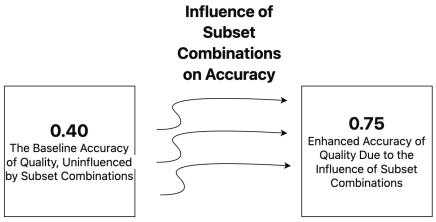


FIGURE 6. This is where we will conduct our critical variable analysis.

As illustrated in Figure 6, our next step involves identifying the critical variables capable of elevating the base accuracy level from 0.4 to 0.75. These variables represent the key factors our brains use to assess wine quality.

VII. SQUARES TO GRAPES: UNRAVELING THE MYSTIQUE OF ML WITH SAVANNAH'S GEOMETRY AND PORTUGUESE WINE

To conclusively demonstrate that 0.40 serves as a baseline accuracy for machine learning (ML) algorithms, we propose a unique experiment. Our aim is to further explore the impact of seemingly irrelevant variables on model predictions. Specifically, we hypothesize that even when trained on data that seems entirely unrelated to the target variable, ML algorithms can still achieve a significant level of accuracy. To test this hypothesis, we examine an unconventional predictor: the number of squares in Savannah, Georgia, and its relationship to the quality of wine in Portugal.

Our experiment involves modifying a dataset on the quality of red wine to include only one predictor variable: the number of squares in Savannah, which is 22. By stripping the dataset to this single, seemingly irrelevant factor, we aim to challenge the algorithms' predictive capabilities.

The methodology follows these steps:

1) Read data from a CSV file named 'winedata.csv', which now includes only one input variable: the number of squares in Savannah, Georgia.
2) Designate the 'quality' column as the target variable.
3) Split the dataset, allocating 80% for training and 20% for testing.
4) Normalize the feature data to ensure that the input variable is scaled properly for the machine learning models.
5) Apply four machine learning algorithms: Logistic Regression, Support Vector Machine (SVM), Decision Tree Classifier, and Random Forest Classifier.
6) Report the accuracy and Mean Squared Error (MSE) values for each model to evaluate their performance.

As shown in Table 8, contrary to expectations, the results show:

TABLE 8. Performance of machine learning algorithms.

Algorithm	Accuracy	MSE
Logistic Regression	0.4062	1.1219
SVM	0.4062	1.1219
Decision Tree	0.4062	1.1219
Random Forest	0.4062	1.1219

These outcomes suggest that ML algorithms can predict outcomes with a certain baseline accuracy, purely based on the distribution inherent in the target variable, regardless of the relevance of the predictor. This baseline accuracy of 0.40 underscores the potential of ML algorithms to identify patterns even in seemingly unrelated data. Our experiment paves the way for identifying 'critical variables', those that can elevate accuracy from this baseline of 0.40 to higher levels, such as 0.75, thereby revealing the factors that are genuinely predictive of the target variable's outcomes.

We have presented an unconventional approach to examining the predictive capabilities of ML algorithms. This exploration of the impact of seemingly irrelevant variables on model predictions challenges traditional perceptions of data relevance in ML.

Our experiment, which utilizes the number of squares in Savannah, Georgia, as a predictor for the quality of wine in Portugal, uniquely illustrates this hypothesis. By demonstrating that ML algorithms can predict wine quality with a baseline accuracy of 0.40, based solely on the distribution inherent in the target variable, we have highlighted an important aspect of ML, the capability of algorithms to identify patterns and make predictions even in the absence of directly relevant data.

This idea of a 'baseline accuracy' suggests that there is a foundational level of predictive capability in ML algorithms, independent of the specific characteristics of the predictor variables. This opens up a broader discussion about the nature of prediction in ML and the identification of 'critical variables.' These are the variables that, when included in the model, can significantly elevate its accuracy beyond the baseline. Understanding these variables can lead to more effective and efficient predictive models.

High accuracy does not always imply that the variable is directly related; it can also suggest that there is a pattern in the quality of Portuguese wines. This pattern could be attributed to the wine production techniques involving Portuguese grapes, which yield wine in a consistent manner, or because the selection process for quality assessment inherently follows a specific pattern.

VIII. IDENTIFYING CRITICAL VARIABLES: ADDING INCREMENTAL CONTRIBUTIONS ON TOP OF BASELINE ACCURACY

A. ESTABLISHING THE BASELINE

First, we establish the baseline for gauging the significance (or criticality) of each involved variable. If no variables are under consideration, we set a preliminary accuracy level—measured by the micro F1 score—at 0.4062 (our baseline above) to serve as a reference point.

Baseline accuracy in machine learning refers to the performance level of a model before any learning or training is applied to it. Studies have shown that the baseline accuracy of machine learning models can significantly impact their performance [93], [94].

B. EXPLORING ALL POSSIBLE COMBINATIONS

We meticulously examine every possible combination of variables, ranging from individual variables to the collective analysis of all 11 variables, to evaluate their predictive power concerning wine quality. This examination begins with individual variables (single variable subsets), such as acidity or sugar level, before progressing to their various combinations, for example, analyzing both acidity and sugar level simultaneously.

C. DETERMINING VARIABLE IMPORTANCE (CRITICALITY)

For each variable combination, we deploy the Random Forest algorithm (with `random state=42`) to predict wine quality and subsequently calculate the micro F1 score. This process enables us to ascertain the extent of each variable's contribution to the predictive accuracy. In scenarios involving a single variable, the assessment is straightforward: we compare its score against the baseline (0.4062). For combinations comprising multiple variables, we sequentially remove one variable at a time, recompute the score, and observe the resulting deviation. The magnitude of this deviation indicates the significance of the omitted variable.

D. ILLUSTRATIVE EXAMPLE

Consider a scenario where we analyze a combination of acidity and alcohol, yielding a score of 0.7. Upon excluding acidity and relying solely on alcohol, the score reduces to 0.65, revealing that acidity's contribution resulted in a 0.05 increase. Conversely, when alcohol is removed, leaving only acidity, the score drops to 0.6, indicating that alcohol's presence contributed to a 0.1 increase in the score.

E. ITERATIVE PROCESS AND UPDATES

This methodology is repeatedly applied across all variable combinations, with the significance of each variable being updated whenever its inclusion in a combination leads to an improved score.

F. CONCLUSION

Upon the completion of this comprehensive analysis, we report the cumulative significance (or criticality) of each variable. This scaled outcome elucidates those variables that are paramount in predicting wine quality, as determined by their contribution to the accuracy of predictions.

For the critical variable analysis, we consistently set `random_state` to 42 in the Random Forest method across all methodologies, including when calculating the baseline, to ensure reproducibility, as shown in Table 9.

Let's reiterate our methodology. We began by evaluating the Random Forest Model using a variable unrelated to the study, set to any constant value, to determine the initial accuracy devoid of the effects of input variables, thereby establishing a baseline accuracy. This baseline represents the accuracy achievable in the absence of any input variables. Subsequently, we explored all possible combinations of variable subsets, assigning a criticality value to each variable based upon its contribution to the accuracy of the subset. For instance, consider a subset containing pH value and chlorine, which achieves an accuracy of 0.6. If alcohol is added to this subset and the accuracy increases to 0.7, this indicates that the contribution of alcohol is 0.1, prompting us to increase alcohol's criticality value by 0.1. This incremental analysis was repeated for every subset throughout the iteration process. Following this analysis, we calculated criticality values for all variables, which are listed in descending order in Table 9.

By understanding the baseline accuracy, we can accurately interpret these numbers as the contributions of each variable to the overall accuracy. Specifically, when the random state is set to 42, the cumulative contribution of alcohol to the overall accuracy across all subset combinations is exactly 48.19.

In Table 9, we note the cumulative criticality of variables.

Algorithm 1 outlines the pseudocode for calculating the cumulative criticality of each variable. It initiates by establishing a baseline accuracy, then assesses the criticality of each variable. This evaluation hinges on observing incremental improvements in accuracy when each variable is included into the subsets as it iterates through all possible combinations.

TABLE 9. Cumulative criticality of variables.

Variable	Cum. Criticality
alcohol	48.1917
sulphates	23.8251
density	22.5084
volatile acidity	17.2375
total sulfur dioxide	16.8334
free sulfur dioxide	11.7500
residual sugar	9.8063
pH	9.6021
chlorides	9.3792
fixed acidity	5.5980
citric acid	4.6084

Algorithm 1 Calculate Cumulative Criticalities in Wine Quality Prediction

Initialize data variables and load the dataset:

- Load the dataset from 'winequality-red.csv'
- Define X as features (all columns except 'quality')
- Define y as the target ('quality')

Initialize the criticalities dictionary with variables as keys and zeroes as values

Set the base F1 score for zero variable subset to 0.4062

Set `random_state` to 42

Define a function to calculate the micro F1 score:

- Split the data into training and testing sets
- Initialize the Random Forest Classifier
- Train the classifier on the training data
- Predict the target variable for the test data
- Calculate and return the micro F1 score

Iterate through all possible subsets of variables from 1 to 11:

- For each subset:
  - Convert the subset to a list and select the corresponding columns from X
  - Calculate the current F1 score
  - If the subset has only one variable:
    - Update the variable's criticality
  - If the subset has more than one variable:
    - For each variable in the subset:
      - Create a new subset without the current variable
      - Calculate the F1 score without the current variable
      - Update the variable's criticality

Output total criticalities for each variable

IX. CRITICALITY ANALYSIS IN ACTION: QUANTIFYING THE IMPORTANCE OF PLAYERS IN SOCCER FOR ENHANCED TEAM DYNAMICS

Numerous research papers have produced varying orders of criticality for variables, many of which use the same wine

dataset as our study. In our research, we do not only rank the variables by their criticality, but we also assign a value to each variable. This approach quantifies the extent to which one variable is more important than another, as outlined in Table 10.

To simplify the understanding of how to solve the incremental criticality problem using the baseline accuracy methodology, we present an analogy involving a soccer team. Suppose you are the coach of soccer Team A, and you intend to understand the criticality of your star players. While you recognize they are all significant, the hierarchy of their importance remains unclear.

A. QUANTIFYING STARDOM: A CRITICALITY ANALYSIS APPROACH TO RANK STAR PLAYERS BY IMPORTANCE

Imagine that we have 11 star players, and we aim to assign a criticality value to each, recognizing their unique contributions. Despite knowing they are all stars, empirically ranking their importance proves challenging. Our goal is to determine which player is more critical than another and to quantify the difference in their importance by percentage. This process involves conducting a criticality analysis to assign a criticality value to each star.

Suppose we are asked to predict the winner between two teams in an upcoming weekend match. Initially, lacking specific information, we estimate Team A's chances of winning against Team B in a soccer match at 40% based on prior knowledge—this 40% serves as our baseline accuracy.

Upon learning that a star player, John, will play for Team A, we revise our prediction to a 42% chance of Team A's victory, attributing this 2% increase to John's critical solo contribution.

$$\text{John:} \quad +2\% \quad \text{(for solo group contribution)}$$

When we find out that another star player, Michael, will also play for Team A, but without any other stars, we adjust our forecast to a 43% chance of victory, starting from 40%. This 3% increase is attributed to Michael's criticality.

$$\text{Michael:} \quad +3\% \quad \text{(for solo group contribution)}$$

It is then revealed that Michael will play alongside John. This information leads us to adjust our prediction to a 45% winning chance for Team A. Since John's presence alone resulted in a 42% chance, this incremental 3% for the first two-player combination is added to Michael's score.

$$\text{Michael:} \quad +3\% \quad \text{(for the first two-player group)}$$

Similarly, the increase from 43% to 45% is allocated to John for his contribution to the first two-player combination.

$$\text{John:} \quad +2\% \quad \text{(for the first two-player group)}$$

By exploring all combinations of the 11 star players, we aim to reward significant contributions more accurately, culminating in a criticality value for each player. As we progress, the number of ways a player can combine with others, with each player receiving one criticality point for their individual contribution, plus additional points for contributions in pairs, triples, etc., up to the full team, is detailed as follows:

- **First run:** Each player receives 1 criticality value for their individual contribution.
- **Second run:** Each player receives 10 additional criticality values for contributing alongside each of the other ten players.
- **Third run:** Each player receives 45 more criticality values for their contributions to three-player combinations.
- **Fourth run:** An additional 120 criticality values are assigned for contributions to four-player combinations.
- **Fifth run:** Each player receives 210 extra criticality values for their contributions to five-player combinations.
- **Sixth run:** An additional 252 criticality values are given to each player for contributing to six-player combinations.
- **Seventh run:** Each player receives another 210 criticality values for contributing to seven-player combinations (mirroring the values in the fifth run).
- **Eighth run:** An additional 120 criticality values are allocated for contributing to eight-player combinations (mirroring the values in the fourth run).
- **Ninth run:** Each player earns a further 45 criticality values for contributing to nine-player combinations (mirroring the values in the third run).
- **Tenth run:** An additional 10 criticality values are awarded for participation in ten-player combinations (mirroring the values in the second run).
- **Eleventh run:** Each player is given one criticality value for contributing to the full team, reflecting a contribution that includes all players.

To calculate the cumulative criticality value for each player, we consider their individual and collective contributions across various scenarios. These values are then summed to determine their total criticality score, with each player's score calculated 1,025 times.

Suppose your star players are volatile acidity, citric acid, residual sugar, chlorides, sulfur dioxide levels, density, pH, sulfates, and alcohol content. Their cumulative importance is reflected in Table 9 above.

To thoroughly understand the methodology, it is essential to grasp what occurs after each run, including how the criticality value of variables changes. To this end, we refer to Table 31 in Appendix E. This table is of utmost importance as it provides insight into the underlying dynamics of criticality analysis.

Let's delve deeper into our analysis, assuming our goal is to understand the relative importance of each variable by percentage. To accurately compare the resulting total values for each variable, we must scale our results by the number of criticality values received in each run (1, 10, 45, 120, etc.), as shown previously. By doing so, we construct the most critical table of this research paper, found as Table 32

in Appendix E. This approach ensures a fair comparison, highlighting the significance of each variable in our study.

TABLE 10. Critical variable comparison post-scaling.

Variable	Criticality
Alcohol	0.5671
Sulfates	0.3116
Density	0.3068
Volatile Acidity	0.2738
Total Sulfur Dioxide	0.2479
Free Sulfur Dioxide	0.2291
pH	0.1620
Residual Sugar	0.1503
Citric Acid	0.1328
Chlorides	0.1319
Fixed Acidity	0.1026

As seen in Table 10, alcohol (0.5671) leads the list in its impact on wine quality, followed by sulfates (0.3116), density (0.3068), volatile acidity (0.2738), total sulfur dioxide (0.2479), free sulfur dioxide (0.2291), pH value (0.1620), residual sugar (0.1503), citric acid (0.1328), chlorides (0.1319), and rounding off the list, fixed acidity (0.1026), illustrates the nuanced and complex chemistry that shapes the quality of wine.

Given the criticality scores in Table 10 of various factors in wine quality prediction, we can draw several compelling comparisons. For example, alcohol, with a score of 0.5671, is significantly more important than most other factors, being approximately 5.52 times more critical than fixed acidity, which has the lowest score of 0.1026. This difference underlines the important role alcohol content plays in determining wine quality compared to the foundational aspect of fixed acidity.

Another intriguing comparison is between sulfates and density. With scores of 0.3116 and 0.3068 respectively, these two factors hold almost equal significance. This observation suggests that both the concentration of sulfates—affecting microbial stability and antioxidant properties—and the density—indicative of the wine's body and alcohol content—play nearly equal roles in determining the wine's overall quality.

Volatile acidity, scoring 0.2738, is less critical than alcohol, sulfates, and density, but it is more pivotal than factors such as total sulfur dioxide (0.2479) and free sulfur dioxide (0.2291). This ranking implies that while the acidity level, which can contribute to the wine's aroma, is important, it is not as decisive as alcohol content or the wine's body.

Looking further down the list, we see that chlorides and citric acid are nearly equally critical, with scores of 0.1319 and 0.1328, respectively. This comparison indicates that the chloride content—which can influence saltiness—and citric acid—which contributes to the wine's freshness and flavor—hold almost equal significance in affecting wine quality. Still, they are less critical than the previously mentioned factors.

These comparisons help us appreciate the nuanced interplay of various chemical properties in determining the quality of wine, with alcohol emerging as a particularly important factor.

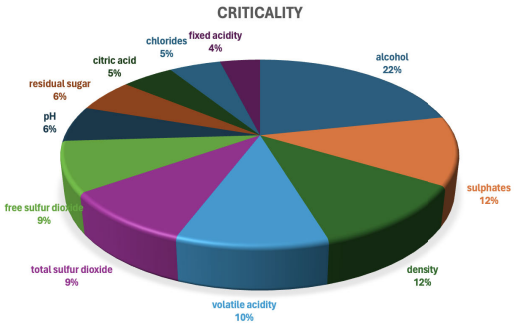


FIGURE 7. Critical variable comparison post-scaling.

X. CONCLUSION

In our research, we explored the use of machine learning (ML) algorithms to improve the accuracy of assessing wine quality, with a particular focus on Portuguese red wine. As interest in applying artificial intelligence (AI) for sensory analysis increases, our study distinguishes itself through the adoption of a comprehensive methodological approach. This method aimed to pinpoint the chemical variables that most effectively predict wine quality. By examining a dataset of diverse physicochemical properties, our objective was to refine the prediction process by highlighting the variables that substantially impact quality evaluations.

Central to our investigative endeavor was the ambition to dissect the influence of various factors on wine quality, prompting an exhaustive analysis across all possible combinations of the 11 variables. This approach facilitated a nuanced understanding of how variable interplays affect the quality of Portuguese wine.

The Random Forest model emerged as the keystone in our predictive analysis, outshining other machine learning algorithms by a significant margin. This finding underscored the model's ability to navigate the complexity of variable interactions, thereby solidifying its position as the premier tool for wine quality prediction.

We assessed the performance of machine learning models against a random guess generator that reflected the dataset's quality distribution probabilities. Our findings revealed that the predictions from the Random Forest model, even when using the least significant variable, were not arbitrary. ML models, despite processing an unrelated input variable, could detect patterns linked to the probabilities of the target variable.

This discovery prompted a critical question: How can we identify the variables that are truly essential for assessing wine quality, especially when minimal or unrelated variables yield significant accuracy? Our subsequent analysis aimed to identify these crucial variables, classifying them as

TABLE 11. Top 10 scores for random forest test number 1.

Combination	Score	MSE
fixed acidity, free sulfur dioxide, total sulfur dioxide	0.746875	0.365625
chlorides, free sulfur dioxide, total sulfur dioxide	0.743750	0.446875
volatile acidity, citric acid, free sulfur dioxide	0.743750	0.350000
volatile acidity, free sulfur dioxide, total sulfur dioxide	0.740625	0.387500
volatile acidity, citric acid, residual sugar, total sulfur dioxide	0.737500	0.356250
volatile acidity, residual sugar, total sulfur dioxide	0.737500	0.362500
fixed acidity, volatile acidity, citric acid, residual sugar	0.737500	0.446875
volatile acidity, residual sugar, free sulfur dioxide	0.737500	0.384375
citric acid, free sulfur dioxide, total sulfur dioxide	0.734375	0.365625
fixed acidity, volatile acidity, citric acid, chlorides	0.734375	0.356250

TABLE 12. Top 10 scores for random forest test number 2.

Combination	Score	MSE
fixed acidity, volatile acidity, citric acid, residual sugar, chlorides	0.750000	0.306250
fixed acidity, citric acid, free sulfur dioxide, total sulfur dioxide	0.750000	0.315625
volatile acidity, citric acid, residual sugar, chlorides	0.746875	0.318750
fixed acidity, volatile acidity, citric acid, chlorides, free sulfur dioxide	0.746875	0.396875
volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide	0.746875	0.318750
volatile acidity, citric acid, chlorides, total sulfur dioxide	0.743750	0.359375
volatile acidity, citric acid, residual sugar, chlorides	0.743750	0.321875
volatile acidity, chlorides, total sulfur dioxide, free sulfur dioxide	0.740625	0.359375
fixed acidity, citric acid, residual sugar, chlorides, free sulfur dioxide	0.740625	0.325000
volatile acidity, chlorides, free sulfur dioxide	0.740625	0.325000

TABLE 13. Top 10 scores for random forest test number 3.

Combination	Score	MSE
fixed acidity, citric acid, residual sugar, free sulfur dioxide	0.756250	0.306250
fixed acidity, volatile acidity, chlorides, total sulfur dioxide	0.753125	0.375000
fixed acidity, citric acid, chlorides, free sulfur dioxide	0.746875	0.318750
citric acid, residual sugar, chlorides, free sulfur dioxide	0.746875	0.346875
fixed acidity, volatile acidity, residual sugar, chlorides	0.743750	0.350000
fixed acidity, volatile acidity, residual sugar, free sulfur dioxide	0.743750	0.368750
volatile acidity, citric acid, residual sugar, chlorides	0.740625	0.315625
fixed acidity, citric acid, total sulfur dioxide	0.737500	0.371875
fixed acidity, volatile acidity, citric acid, residual sugar	0.734375	0.412500
citric acid, chlorides, total sulfur dioxide, density	0.734375	0.350000

TABLE 14. Top 10 scores for random forest test number 4.

Combination	Score	MSE
volatile acidity, citric acid, chlorides, total sulfur dioxide	0.756250	0.365625
volatile acidity, citric acid, residual sugar, total sulfur dioxide	0.753125	0.384375
fixed acidity, residual sugar, sulphates, alcohol	0.750000	0.325000
fixed acidity, citric acid, free sulfur dioxide, total sulfur dioxide	0.746875	0.328125
volatile acidity, citric acid, residual sugar, chlorides	0.743750	0.356250
citric acid, residual sugar, chlorides, total sulfur dioxide	0.743750	0.400000
fixed acidity, volatile acidity, residual sugar, chlorides	0.743750	0.359375
fixed acidity, citric acid, residual sugar, free sulfur dioxide	0.743750	0.356250
fixed acidity, volatile acidity, residual sugar, chlorides, free sulfur dioxide	0.743750	0.337500
fixed acidity, volatile acidity, citric acid, residual sugar	0.740625	0.393750

key contributors to enhancing prediction accuracy. Critical variables are those that contribute more significantly than others to improving the base accuracy from about 0.4 to over 0.75.

TABLE 15. Top 10 scores for random forest test number 5.

Combination	Score	MSE
fixed acidity, volatile acidity, chlorides, free sulfur dioxide	0.753125	0.331250
fixed acidity, volatile acidity, citric acid, residual sugar, chlorides	0.750000	0.306250
chlorides, total sulfur dioxide, sulphates, alcohol	0.750000	0.278125
fixed acidity, volatile acidity, citric acid, residual sugar, free sulfur dioxide	0.743750	0.321875
fixed acidity, volatile acidity, chlorides, free sulfur dioxide, sulphates	0.743750	0.331250
fixed acidity, citric acid, residual sugar, total sulfur dioxide	0.743750	0.340625
volatile acidity, citric acid, residual sugar, chlorides	0.743750	0.331250
fixed acidity, total sulfur dioxide, sulphates, alcohol	0.743750	0.303125
fixed acidity, volatile acidity, citric acid, residual sugar	0.740625	0.400000
volatile acidity, residual sugar, total sulfur dioxide	0.740625	0.403125

TABLE 16. Top 10 scores for decision tree test number 1.

Combination	Score	MSE
volatile acidity, residual sugar, free sulfur dioxide	0.703125	0.481250
fixed acidity, volatile acidity, residual sugar, chlorides	0.681250	0.625000
citric acid, chlorides, free sulfur dioxide, density	0.678125	0.496875
citric acid, chlorides, total sulfur dioxide, pH	0.678125	0.575000
volatile acidity, citric acid, residual sugar, total sulfur dioxide	0.678125	0.496875
fixed acidity, citric acid, residual sugar, chlorides, free sulfur dioxide	0.675000	0.575000
chlorides, free sulfur dioxide, total sulfur dioxide, density	0.675000	0.503125
fixed acidity, volatile acidity, residual sugar, chlorides, free sulfur dioxide	0.671875	0.471875
volatile acidity, citric acid, residual sugar, total sulfur dioxide, density	0.671875	0.621875
citric acid, chlorides, free sulfur dioxide, density, sulphates	0.671875	0.550000

TABLE 17. Top 10 scores for decision tree test number 2.

Combination	Score	MSE
volatile acidity, residual sugar, chlorides, free sulfur dioxide	0.703125	0.450000
volatile acidity, citric acid, total sulfur dioxide, density	0.687500	0.478125
volatile acidity, citric acid, chlorides, density	0.684375	0.490625
residual sugar, chlorides, sulphates, alcohol	0.678125	0.556250
volatile acidity, citric acid, residual sugar, total sulfur dioxide	0.671875	0.506250
fixed acidity, residual sugar, density, pH, alcohol	0.671875	0.643750
fixed acidity, residual sugar, total sulfur dioxide, density	0.668750	0.503125
fixed acidity, citric acid, residual sugar, chlorides	0.668750	0.596875
volatile acidity, residual sugar, chlorides, total sulfur dioxide	0.668750	0.515625
citric acid, chlorides, free sulfur dioxide, total sulfur dioxide, density	0.665625	0.503125

TABLE 18. Top 10 scores for decision tree test number 3.

Combination	Score	MSE
citric acid, residual sugar, chlorides, total sulfur dioxide	0.703125	0.428125
fixed acidity, volatile acidity, chlorides, free sulfur dioxide	0.681250	0.493750
volatile acidity, citric acid, total sulfur dioxide, density	0.681250	0.503125
fixed acidity, citric acid, residual sugar, free sulfur dioxide	0.678125	0.537500
fixed acidity, volatile acidity, free sulfur dioxide, density	0.678125	0.490625
fixed acidity, volatile acidity, residual sugar, chlorides	0.678125	0.559375
volatile acidity, citric acid, residual sugar, chlorides	0.675000	0.546875
volatile acidity, citric acid, residual sugar, pH	0.675000	0.468750
volatile acidity, citric acid, residual sugar, chlorides, density	0.675000	0.496875
citric acid, residual sugar, chlorides, density, pH	0.671875	0.509375

In our paper, we went beyond simply listing the essential variables; we also assigned criticality values to each, enabling the quantification of their relative importance. Alcohol was identified as having the most substantial impact on wine

TABLE 19. Top 10 scores for decision tree test number 4.

Combination	Score	MSE
fixed acidity, residual sugar, chlorides, total sulfur dioxide	0.684375	0.528125
fixed acidity, citric acid, residual sugar, chlorides	0.681250	0.537500
volatile acidity, residual sugar, free sulfur dioxide	0.678125	0.450000
fixed acidity, volatile acidity, citric acid, residual sugar	0.675000	0.650000
fixed acidity, volatile acidity, residual sugar, chlorides	0.675000	0.618750
volatile acidity, citric acid, chlorides, total sulfur dioxide	0.671875	0.603125
citric acid, free sulfur dioxide, alcohol	0.671875	0.568750
volatile acidity, residual sugar, chlorides, total sulfur dioxide	0.668750	0.606250
volatile acidity, citric acid, residual sugar, sulphates	0.668750	0.490625
fixed acidity, volatile acidity, residual sugar, chlorides, free sulfur dioxide	0.665625	0.537500

TABLE 20. Top 10 scores for decision tree test number 5.

Combination	Score	MSE
fixed acidity, volatile acidity, citric acid, residual sugar	0.700000	0.453125
fixed acidity, chlorides, free sulfur dioxide, total sulfur dioxide	0.687500	0.509375
volatile acidity, citric acid, chlorides, free sulfur dioxide	0.681250	0.531250
fixed acidity, total sulfur dioxide, density, sulphates	0.678125	0.509375
volatile acidity, residual sugar, free sulfur dioxide	0.675000	0.543750
volatile acidity, citric acid, residual sugar, pH	0.671875	0.459375
fixed acidity, chlorides, free sulfur dioxide, pH	0.668750	0.506250
volatile acidity, residual sugar, chlorides, free sulfur dioxide	0.668750	0.606250
fixed acidity, volatile acidity, free sulfur dioxide, density	0.668750	0.625000
citric acid, residual sugar, free sulfur dioxide, density	0.668750	0.468750

TABLE 21. Top 10 scores for logistic regression test number 1.

Combination	Score	MSE
chlorides, sulphates, alcohol	0.671875	0.465625
volatile acidity, citric acid, residual sugar, free sulfur dioxide	0.665625	0.434375
fixed acidity, volatile acidity, residual sugar, chlorides	0.650000	0.440625
volatile acidity, chlorides, free sulfur dioxide, alcohol	0.646875	0.465625
residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide	0.646875	0.496875
volatile acidity, free sulfur dioxide, total sulfur dioxide, alcohol	0.643750	0.459375
chlorides, pH, sulphates, alcohol	0.643750	0.496875
fixed acidity, residual sugar, chlorides, free sulfur dioxide	0.643750	0.500000
volatile acidity, citric acid, residual sugar, free sulfur dioxide, alcohol	0.643750	0.484375
fixed acidity, chlorides, free sulfur dioxide, density	0.640625	0.471875

TABLE 22. Top 10 scores for logistic regression test number 2.

Combination	Score	MSE
fixed acidity, volatile acidity, chlorides, free sulfur dioxide	0.659375	0.387500
fixed acidity, volatile acidity, residual sugar, chlorides	0.653125	0.437500
fixed acidity, chlorides, total sulfur dioxide, density	0.646875	0.465625
fixed acidity, volatile acidity, chlorides, free sulfur dioxide, alcohol	0.643750	0.421875
fixed acidity, volatile acidity, residual sugar, chlorides, free sulfur dioxide	0.643750	0.450000
fixed acidity, volatile acidity, citric acid, residual sugar, chlorides	0.643750	0.496875
volatile acidity, free sulfur dioxide, total sulfur dioxide, alcohol	0.637500	0.390625
volatile acidity, chlorides, free sulfur dioxide, alcohol	0.634375	0.487500
fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide	0.631250	0.471875
volatile acidity, citric acid, residual sugar, chlorides	0.631250	0.453125

quality, with a value of 0.5671. It was followed by Sulphates (0.3116), Density (0.3068), Volatile Acidity (0.2738), Total Sulfur Dioxide (0.2479), Free Sulfur Dioxide (0.2291), pH Value (0.1620), Residual Sugar (0.1503),

TABLE 23. Top 10 scores for logistic regression test number 3.

Combination	Score	MSE
volatile acidity, citric acid, residual sugar, chlorides	0.668750	0.415625
volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide	0.662500	0.412500
volatile acidity, citric acid, residual sugar, free sulfur dioxide, total sulfur dioxide	0.653125	0.421875
fixed acidity, volatile acidity, total sulfur dioxide, density	0.646875	0.456250
volatile acidity, chlorides, free sulfur dioxide, total sulfur dioxide	0.646875	0.446875
volatile acidity, residual sugar, chlorides, total sulfur dioxide, density	0.646875	0.390625
volatile acidity, citric acid, residual sugar, chlorides, density	0.640625	0.434375
volatile acidity, citric acid, residual sugar, free sulfur dioxide, density	0.640625	0.468750
fixed acidity, volatile acidity, citric acid, residual sugar	0.637500	0.446875
volatile acidity, citric acid, free sulfur dioxide, density	0.634375	0.450000

TABLE 24. Top 10 scores for logistic regression test number 4.

Combination	Score	MSE
fixed acidity, residual sugar, free sulfur dioxide, density	0.650000	0.471875
residual sugar, chlorides, total sulfur dioxide, density	0.646875	0.534375
volatile acidity, citric acid, residual sugar, total sulfur dioxide, density	0.640625	0.425000
fixed acidity, volatile acidity, citric acid, residual sugar, density	0.640625	0.490625
fixed acidity, volatile acidity, citric acid, chlorides, density	0.634375	0.487500
fixed acidity, volatile acidity, chlorides, free sulfur dioxide, density	0.631250	0.453125
fixed acidity, volatile acidity, citric acid, free sulfur dioxide, density	0.631250	0.481250
sulphates, alcohol	0.631250	0.425000
fixed acidity, residual sugar, chlorides, free sulfur dioxide, density	0.631250	0.471875
volatile acidity, citric acid, residual sugar, alcohol	0.628125	0.409375

TABLE 25. Top 10 scores for logistic regression test number 5.

Combination	Score	MSE
volatile acidity, chlorides, total sulfur dioxide, density	0.662500	0.421875
volatile acidity, citric acid, residual sugar, chlorides	0.653125	0.450000
volatile acidity, residual sugar, total sulfur dioxide, density	0.643750	0.456250
fixed acidity, volatile acidity, citric acid, chlorides, density	0.643750	0.446875
fixed acidity, residual sugar, density, pH, alcohol	0.643750	0.418750
fixed acidity, volatile acidity, citric acid, residual sugar, density	0.640625	0.471875
fixed acidity, citric acid, chlorides, density, pH	0.640625	0.425000
volatile acidity, citric acid, residual sugar, chlorides, density	0.637500	0.518750
fixed acidity, citric acid, free sulfur dioxide, density	0.637500	0.484375
citric acid, residual sugar, chlorides, free sulfur dioxide, density	0.637500	0.453125

TABLE 26. Top 10 scores for SVM test number 1.

Combination	Score
residual sugar, chlorides, density, sulphates, alcohol	0.646875
fixed acidity, volatile acidity, citric acid, chlorides, free sulfur dioxide	0.643750
fixed acidity, volatile acidity, citric acid, free sulfur dioxide, density	0.640625
fixed acidity, volatile acidity, residual sugar, chlorides, free sulfur dioxide	0.637500
fixed acidity, volatile acidity, citric acid, chlorides, density	0.637500
fixed acidity, pH, sulphates, alcohol	0.634375
volatile acidity, citric acid, free sulfur dioxide, alcohol	0.634375
fixed acidity, volatile acidity, citric acid, alcohol	0.631250
volatile acidity, citric acid, residual sugar, total sulfur dioxide, density	0.631250
fixed acidity, volatile acidity, residual sugar, density, alcohol	0.628125

TABLE 27. Top 10 scores for SVM test number 2.

Combination	Score
volatile acidity, chlorides, free sulfur dioxide, density	0.646875
volatile acidity, citric acid, free sulfur dioxide, density	0.646875
fixed acidity, volatile acidity, citric acid, chlorides, density	0.646875
fixed acidity, volatile acidity, chlorides, total sulfur dioxide, density	0.643750
fixed acidity, citric acid, residual sugar, total sulfur dioxide, density	0.643750
fixed acidity, volatile acidity, residual sugar, chlorides, density	0.643750
fixed acidity, volatile acidity, citric acid, chlorides, free sulfur dioxide	0.643750
residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density	0.640625
chlorides, total sulfur dioxide, density, pH, sulphates, alcohol	0.640625
fixed acidity, volatile acidity, free sulfur dioxide, density	0.640625

TABLE 28. Top 10 scores for SVM test number 3.

Combination	Score
fixed acidity, volatile acidity, citric acid, residual sugar	0.665625
residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide	0.653125
fixed acidity, volatile acidity, citric acid, chlorides	0.646875
volatile acidity, citric acid, total sulfur dioxide, density	0.643750
fixed acidity, volatile acidity, citric acid, residual sugar, density	0.640625
volatile acidity, residual sugar, chlorides, total sulfur dioxide	0.640625
volatile acidity, citric acid, chlorides, free sulfur dioxide	0.637500
citric acid, free sulfur dioxide, total sulfur dioxide, density	0.637500
fixed acidity, volatile acidity, citric acid, residual sugar, free sulfur dioxide	0.637500
volatile acidity, citric acid, residual sugar, chlorides	0.637500

TABLE 29. Top 10 scores for SVM test number 4.

Combination	Score
fixed acidity, volatile acidity, free sulfur dioxide, density	0.665625
fixed acidity, volatile acidity, citric acid, chlorides, density	0.659375
fixed acidity, volatile acidity, chlorides, total sulfur dioxide, density	0.650000
fixed acidity, volatile acidity, citric acid, residual sugar, chlorides	0.650000
fixed acidity, volatile acidity, citric acid, residual sugar, density	0.640625
fixed acidity, volatile acidity, citric acid, residual sugar, free sulfur dioxide, density	0.640625
volatile acidity, citric acid, chlorides, total sulfur dioxide, density	0.637500
volatile acidity, citric acid, residual sugar, free sulfur dioxide, density	0.637500
volatile acidity, citric acid, residual sugar, chlorides, density	0.634375
fixed acidity, volatile acidity, residual sugar, chlorides, density	0.634375

TABLE 30. Top 10 scores for SVM test number 5.

Combination	Score
fixed acidity, volatile acidity, total sulfur dioxide, density	0.653125
volatile acidity, residual sugar, free sulfur dioxide, density	0.643750
volatile acidity, density, pH, alcohol	0.640625
fixed acidity, citric acid, residual sugar, chlorides, density	0.637500
fixed acidity, volatile acidity, citric acid, chlorides, density	0.637500
residual sugar, total sulfur dioxide, density, sulphates, alcohol	0.634375
fixed acidity, volatile acidity, citric acid, residual sugar, density	0.634375
fixed acidity, volatile acidity, citric acid, chlorides, free sulfur dioxide, density	0.634375
fixed acidity, volatile acidity, citric acid, free sulfur dioxide, density	0.634375
fixed acidity, residual sugar, chlorides, free sulfur dioxide, density	0.631250

TABLE 31. Iterative updates to criticality scores.

Variable	Total	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10	Run 11
alcohol	48.1917	0.1021	0.9396	4.8833	7.4938	11.2083	11.2250	7.4979	3.7125	0.9458	0.1875	-0.0042
sulphates	23.8251	0.0667	0.7604	3.6083	3.9417	5.8563	5.1104	3.0229	1.4000	0.1417	-0.0688	-0.0146
density	22.5084	0.0626	0.8479	3.5521	3.6146	5.4458	4.8042	2.8229	1.1958	0.1750	0.0104	-0.0229
volatile acidity	17.2376	0.0542	0.6896	3.0979	2.7188	3.6917	2.9271	2.2563	1.2979	0.4688	0.0417	-0.0063
total sulfur dioxide	16.8334	0.0667	0.6771	3.1417	2.8396	4.1125	3.1813	2.0604	0.7917	0.0292	-0.0417	-0.0250
free sulfur dioxide	11.7501	0.0855	0.5125	2.5708	1.5563	2.1375	2.1167	1.7688	0.9479	0.0875	-0.0208	-0.0125
residual sugar	9.8063	0.0396	0.2938	2.1938	1.2479	2.5396	1.6792	1.3354	0.6021	-0.0563	-0.0688	0.0000
pH	9.6021	0.0376	0.3479	2.1938	1.3604	2.0625	1.9333	0.9646	0.7313	-0.0313	0.0000	0.0021
chlorides	9.3792	0.0001	0.3667	2.3958	1.1417	1.8771	1.5688	1.1896	0.7854	0.0521	-0.0021	0.0042
fixed acidity	5.5980	0.0188	0.4854	2.6292	1.2083	1.2521	0.4458	0.0292	-0.1875	-0.1688	-0.0875	-0.0271
citric acid	4.6084	0.0480	0.4917	2.1875	0.6375	0.9500	0.4083	-0.1063	0.2313	-0.1688	-0.0542	-0.0167

TABLE 32. Scaled iterative updates to criticality scores.

Variable	Total	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10	Run 11
alcohol	0.5671	0.1021	0.0940	0.1085	0.0624	0.0534	0.0445	0.0357	0.0309	0.0210	0.0188	-0.0042
sulphates	0.3116	0.0667	0.0760	0.0802	0.0328	0.0279	0.0203	0.0144	0.0117	0.0031	-0.0069	-0.0146
density	0.3068	0.0626	0.0848	0.0789	0.0301	0.0259	0.0191	0.0134	0.0100	0.0039	0.0010	-0.0229
volatile acidity	0.2738	0.0542	0.0690	0.0688	0.0227	0.0176	0.0116	0.0107	0.0108	0.0104	0.0042	-0.0062
total sulfur dioxide	0.2479	0.0667	0.0677	0.0698	0.0237	0.0196	0.0126	0.0098	0.0066	0.0006	-0.0042	-0.0250
free sulfur dioxide	0.2291	0.0855	0.0513	0.0571	0.0130	0.0102	0.0084	0.0084	0.0079	0.0019	-0.0021	-0.0125
pH	0.1620	0.0375	0.0348	0.0488	0.0113	0.0098	0.0077	0.0046	0.0061	-0.0007	0.0000	0.0021
residual sugar	0.1503	0.0396	0.0294	0.0488	0.0104	0.0121	0.0067	0.0064	0.0050	-0.0012	-0.0069	0.0000
citric acid	0.1328	0.0480	0.0492	0.0486	0.0053	0.0045	0.0016	-0.0005	0.0019	-0.0037	-0.0054	-0.0167
chlorides	0.1319	0.0000	0.0367	0.0532	0.0095	0.0089	0.0062	0.0057	0.0065	0.0012	-0.0002	0.0042
fixed acidity	0.1026	0.0188	0.0485	0.0584	0.0101	0.0060	0.0018	0.0001	-0.0016	-0.0037	-0.0087	-0.0271

Citric Acid (0.1328), Chlorides (0.1319), with Fixed Acidity (0.1026) being the least impactful among the factors listed.

APPENDIX A
RESULTS FROM THE RANDOM FOREST MODEL

see Tables 11–15.

APPENDIX B
RESULTS FROM THE DECISION TREE MODEL

See Tables 16–20.

APPENDIX C
RESULTS FROM THE LOGISTIC REGRESSION MODEL

See Tables 21–25.

APPENDIX D
RESULTS FROM THE SUPPORT VECTOR MACHINE MODEL

See Tables 26–30.

APPENDIX E

See Tables 31 and 32.

REFERENCES

[1] P. McGovern, M. Jalabadze, S. Batiuk, M. P. Callahan, K. E. Smith, G. R. Hall, E. Kvavadze, D. Maghradze, N. Rusishvili, L. Bouby, O. Failla, G. Cola, L. Mariani, E. Boaretto, R. Bacilieri, P. This, N. Wales, and D. Lordkipanidze, "Early neolithic wine of Georgia in the South Caucasus," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 48, pp. E10309–E10318, Nov. 2017.

[2] C. E. Yavas, J. Kim, and L. Chen, "Exploring flavors through AI: The future of culinary taste prediction," in *Proc. 22nd IEEE/ACIS Int. Conf. Softw. Eng., Manage. Appl. (SERA)*, Honolulu, HI, USA, May/Jun. 2024.

[3] P. Cortez, J. Teixeira, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Using data mining for wine quality assessment," in *Discovery Science* (Lecture Notes in Computer Science), vol. 5808, J. Gama, V. S. Costa, A. M. Jorge, and P. B. Brazdil, Eds., Berlin, Germany: Springer, 2009, pp. 66–79. [Online]. Available: http://link.springer.com/10.1007/978-3-642-04747-3_8

[4] Y. Gupta, "Selection of important features and predicting wine quality using machine learning techniques," *Proc. Comput. Sci.*, vol. 125, pp. 305–312, Jan. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877050917328053

[5] B. S. Anami, K. Mainalli, S. Kallur, and V. Patil, "A machine learning based approach for wine quality prediction," in *Proc. 2nd Asian Conf. Innov. Technol. (ASIANCON)*, Pune, India, Aug. 2022, pp. 1–6.

[6] H. Zhang, Z. Wang, J. He, and J. Tong, "Construction of wine quality prediction model based on machine learning algorithm," in *Proc. 5th Int. Conf. Artif. Intell. Virtual Reality (AIVR)*, Kumamoto, Japan. New York, NY, USA: ACM, Jul. 2021, p. 6, doi: 10.1145/3480433.3480443.

[7] N. Korade and M. Salunke, "Identification of appropriate machine learning algorithm to predict wine quality," *Int. J. Sci. Res. Eng. Manag.*, vol. 5, no. 5, pp. 1–7, 2021.

[8] K. Jain, K. Kaushik, S. K. Gupta, S. Mahajan, and S. Kadry, "Machine learning-based predictive modelling for the enhancement of wine quality," *Sci. Rep.*, vol. 13, no. 1, Oct. 2023, Art. no. 17042. [Online]. Available: https://www.nature.com/articles/s41598-023-44111-9

[9] O. D. Akanbi, T. M. Faloni, and S. Olaniyi, "Prediction of wine quality: Comparing machine learning models in R programming," *Int. J. Latest Technol. Eng., Manage. Appl. Sci.*, vol. 11, no. 9, pp. 1–6, 2022.

[10] K. R. Dahal, J. N. Dahal, H. Banjade, and S. Gaire, "Prediction of wine quality using machine learning algorithms," *Open J. Statist.*, vol. 11, no. 2, pp. 278–289, 2021.

[11] S. Zaza, M. Atemkeng, and S. Hamlomo, "Wine feature importance and quality prediction: A comparative study of machine learning algorithms with unbalanced data," Oct. 2023, *arXiv:2310.01584*.

[12] D. Angus. (2023). *Modeling Wine Quality From Physicochemical Properties*. Stanford University. [Online]. Available: https://mailto: dangus@stanford.edu

[13] P. Bhardwaj, P. Tiwari, K. Olejar, W. Parr, and D. Kulasiri, "A machine learning application in wine quality prediction," *Mach. Learn. Appl.*, vol. 8, Jun. 2022, Art. no. 100261.

[14] M. Koranga, R. Pandey, M. Joshi, and M. Kumar, "Analysis of white wine using machine learning algorithms," *J. Wine Sci.*, vol. 34, no. 2, pp. 120–135, 2024.

[15] S. Mani, R. A. Krishnankutty, S. Swaminathan, and P. Theerthagiri, "An investigation of wine quality testing using machine learning techniques," *IAES Int. J. Artif. Intell.*, vol. 12, no. 2, p. 747, Jun. 2023.

[16] C. Ye, K. Li, and G.-Z. Jia, "A new red wine prediction framework using machine learning," *J. Phys., Conf. Ser.*, vol. 1684, no. 1, Nov. 2020, Art. no. 012067.

[17] A. Sinha and A. Kumar, "Wine quality and taste classification using machine learning model," *Int. J. Innov. Res. Appl. Sci. Eng.*, vol. 4, no. 4, pp. 715–721, Oct. 2020.

[18] S. Aich, A. A. Al-Absi, K. L. Hui, and M. Sain, "Prediction of quality for different type of wine based on different feature sets using supervised machine learning techniques," *ICACT Trans. Adv. Commun. Technol.*, vol. 7, no. 3, pp. 1122–1127, 2018.

[19] Z. Tingwei, "Red wine quality prediction through active learning," *J. Phys., Conf. Ser.*, vol. 1966, no. 1, Jul. 2021, Art. no. 012021.

[20] B. Zhan, "Forecasting red wine quality: A comparative examination of machine learning approaches," *Appl. Comput. Eng.*, vol. 32, no. 1, pp. 58–65, Jan. 2024. [Online]. Available: https://ace.ewapublishing.org/article/004c9e38802c463e986f015b91a79436

[21] R. A. Tindal, D. W. Jeffery, and R. A. Muhlack, "Mathematical modelling to enhance winemaking efficiency: A review of red wine colour and polyphenol extraction and evolution," *Austral. J. Grape Wine Res.*, vol. 27, no. 2, pp. 219–233, Apr. 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/ajgw.12488

[22] G.-L. Hou, B. Ge, L.-L. Sun, and K.-X. Xing, "A study on wine sensory evaluation by the statistical analysis method," *Czech J. Food Sci.*, vol. 38, no. 1, pp. 1–10, Feb. 2020. [Online]. Available: http://cjfs.agriculturejournals.cz/doi/10.17221/438/2017-CJFS.html

[23] D. Cozzolino, "The role of visible and infrared spectroscopy combined with chemometrics to measure phenolic compounds in grape and wine samples," *Molecules*, vol. 20, no. 1, pp. 726–737, Jan. 2015. [Online]. Available: https://www.mdpi.com/1420-3049/20/1/726

[24] P. Arapitsas, G. Guella, and F. Mattivi, "The impact of SO_2 on wine flavanols and indoles in relation to wine style and age," *Sci. Rep.*, vol. 8, no. 1, p. 858, Jan. 2018. [Online]. Available: https://www.nature.com/articles/s41598-018-19185-5

[25] J. Luo, P. Zhang, Y. T. Loo, J. Ma, S. Wu, P. J. Marriott, and K. Howell, "Can wine quality be predicted by small volatile compounds? A study based on performance of wine show entries and their volatile profiles," *Flavour Fragrance J.*, vol. 37, no. 6, pp. 354–363, Nov. 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/ffj.3720

[26] T. Stein, I. R. Carvalho, V. J. Szareski, R. G. S. Zocche, F. Zocche, K. G. Aloy, L. D. V. D. Santos, H. C. G. B. Martins, F. Lautenchleger, and V. Q. D. Souza, "Quality of cabernet sauvignon wines determined by the variability of climatic attributes," *J. Agricult. Sci.*, vol. 10, no. 8, p. 380, Jul. 2018. [Online]. Available: http://www.ccsenet.org/journal/index.php/jas/article/view/73781

[27] Ž. Kurtanjek, "Causal artificial intelligence models of food quality data," *Food Technol. Biotechnol.*, vol. 62, no. 1, pp. 102–109, Mar. 2024. [Online]. Available: https://www.ftb.com.hr/archives/1859-causal-artificial-intelligence-models-of-food-quality-data

[28] Q. Xu and K. Xu, "Quality evaluation of Chinese red wine based on cloud model," *J. Food Biochem.*, vol. 43, no. 10, Oct. 2019, Art. no. e12787. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/jfbc.12787

[29] N. M. D. O. Arcanjo, I. A. Neri-Numa, T. K. A. Bezerra, F. L. H. D. Silva, G. M. Pastore, and M. S. Madruga, "Quality evaluation of red wines produced from the Isabella and Ives cultivar (*Vitis labrusca*): Physicochemical parameters, phenolic composition and antioxidant activity," *Food Sci. Technol.*, vol. 37, no. 2, pp. 184–192, May 2017. [Online]. Available: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-20612017000200184&lng=en&tlng=en

[30] J. Preserova, V. Ranc, D. Milde, V. Kubistova, and J. Stavek, "Study of phenolic profile and antioxidant activity in selected moravian wines during winemaking process by FT-IR spectroscopy," *J. Food Sci. Technol.*, vol. 52, no. 10, pp. 6405–6414, Oct. 2015. [Online]. Available: http://link.springer.com/10.1007/s13197-014-1644-8

[31] D. J. I. Supriatna, H. Saputra, and K. Hasan, "Enhancing the red wine quality classification using ensemble voting classifiers," *Infolitika J. Data Sci.*, vol. 1, no. 2, pp. 42–47, Oct. 2023. [Online]. Available: https://heca-analitika.com/ijds/article/view/95

[32] M.-P. Sáenz-Navajas, S. Ferrero-del-Teso, D. W. Jeffery, V. Ferreira, and P. Fernández-Zurbano, "Effect of aroma perception on taste and mouthfeel dimensions of red wines: Correlation of sensory and chemical measurements," *Food Res. Int.*, vol. 131, May 2020, Art. no. 108945. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0963996919308312

[33] C. E. Ofoedu, E. O. Ofoedu, J. S. Chacha, C. I. Owuamanam, I. S. Efekalam, and C. G. Awuchi, "Comparative evaluation of physicochemical, antioxidant, and sensory properties of red wine as markers of its quality and authenticity," *Int. J. Food Sci.*, vol. 2022, pp. 1–17, Oct. 2022. [Online]. Available: https://www.hindawi.com/journals/ijfs/2022/8368992/

[34] I. Sen, B. Ozturk, F. Tokatli, and B. Ozen, "Combination of visible and mid-infrared spectra for the prediction of chemical parameters of wines," *Talanta*, vol. 161, pp. 130–137, Dec. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0039914016306312

[35] M. Kelly, R. Longjohn, and K. Nottingham. (2024). *The UCI Machine Learning Repository*. Accessed: Jun. 1, 2024. [Online]. Available: https://archive.ics.uci.edu

[36] A. C. P. Cortez. (2009). *Wine Quality*. [Online]. Available: https://archive.ics.uci.edu/dataset/186

[37] F. Yu, C. Wei, P. Deng, T. Peng, and X. Hu, "Deep exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles," *Sci. Adv.*, vol. 7, no. 22, May 2021, Art. no. eabf4130. [Online]. Available: https://www.science.org/doi/10.1126/sciadv.abf4130

[38] Y. Liu, W.-J. Li, X. Zhang, M. Lewenstein, G. Su, and S.-J. Ran, "Entanglement-based feature extraction by tensor network machine learning," *Frontiers Appl. Math. Statist.*, vol. 7, Aug. 2021, Art. no. 716044. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fams.2021.716044/full

[39] M. S. Park, H. Son, C. Hyun, and H. J. Hwang, "Explainability of machine learning models for bankruptcy prediction," *IEEE Access*, vol. 9, pp. 124887–124899, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9529166/

[40] A. Ejmalian, A. Aghaei, S. Nabavi, M. A. Darabad, A. Tajbakhsh, A. A. Abin, M. E. Moghaddam, A. Dabbagh, A. Jahangirifard, E. Memary, and S. Sayyadi, "Prediction of acute kidney injury after cardiac surgery using interpretable machine learning," *Anesthesiol. Pain Med.*, vol. 12, no. 4, Sep. 2022, Art. no. e127140. [Online]. Available: https://brieflands.com/articles/aapm-127140.html

[41] Y. Huimin, "Research on parallel support vector machine based on spark big data platform," *Sci. Program.*, vol. 2021, pp. 1–9, Dec. 2021. [Online]. Available: https://www.hindawi.com/journals/sp/2021/7998417/

[42] N. Foreman and M. Evett, "Preventing overfitting in GP with Canary functions," in *Proc. 7th Annu. Conf. Genet. Evol. Comput.*, Washington, DC, USA, Jun. 2005, pp. 1779–1780. [Online]. Available: https://dl.acm.org/doi/10.1145/1068009.1068307

[43] M. Zakariyah and U. Zaky, "Analysis of machine learning algorithm for sleep apnea detection based on heart rate variability," *Jurnal Informatika*, vol. 10, no. 2, p. 173, Nov. 2022. [Online]. Available: https://jurnalnasional.ump.ac.id/index.php/JUITA/article/view/14575

[44] E. I. Zacharaki, V. G. Kanas, and C. Davatzikos, "Investigating machine learning techniques for MRI-based classification of brain neoplasms," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 6, no. 6, pp. 821–828, Nov. 2011. [Online]. Available: http://link.springer.com/10.1007/s11548-011-0559-3

[45] J. Jeon, N. Seo, J.-G. Jung, S. B. Son, and S.-J. Lee, "Machine learning prediction for cementite precipitation in austenite of low-alloy steels," *Mater. Trans.*, vol. 63, no. 10, pp. 1369–1374, Oct. 2022. [Online]. Available: https://www.jstage.jst.go.jp/article/matertrans/63/10/63_MT-MB2022009/_article

[46] S. Dankwa and W. Zheng, "Special issue on using machine learning algorithms in the prediction of kyphosis disease: A comparative study," *Appl. Sci.*, vol. 9, no. 16, p. 3322, Aug. 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/16/3322

[47] A. C. Just, Y. Liu, M. Sorek-Hamer, J. Rush, M. Dorman, R. Chatfield, Y. Wang, A. Lyapustin, and I. Kloog. (Dec. 2019). *Gradient Boosting Machine Learning to Improve Satellite-Derived Column Water Vapor Measurement Error*. [Online]. Available: https://amt.copernicus.org/preprints/amt-2019-308/amt-2019-308.pdf

[48] K. Kavousi, L. Mirsadeghi, R. H. Hosseini, A. M. Banaei-Moghaddam, and S. R. Beh-Afarin. (Jun. 2019). *A Post-Method Condition Analysis of Using Ensemble Machine Learning for Cancer Prognosis and Diagnosis: A Systematic Review*. [Online]. Available: https://www.researchsquare.com/article/rs-1536/v1

[49] M. Shahhosseini, G. Hu, S. Khaki, and S. V. Archontoulis, "Corn yield prediction with ensemble CNN-DNN," *Frontiers Plant Sci.*, vol. 12, Aug. 2021, Art. no. 709008. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpls.2021.709008/full

[50] K. N. K. A. Rahim, I. Elamvazuthi, L. I. Izhar, and G. Capi, "Classification of human daily activities using ensemble methods based on smartphone inertial sensors," *Sensors*, vol. 18, no. 12, p. 4132, Nov. 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/12/4132

[51] Y. Liu, S. Ye, X. Xiao, C. Sun, G. Wang, G. Wang, and B. Zhang, "Machine learning for tuning, selection, and ensemble of multiple risk scores for predicting type 2 diabetes," *Risk Manage. Healthcare Policy*, vol. 12, pp. 189–198, Nov. 2019. [Online]. Available: https://www.dovepress.com/machine-learning-for-tuning-selection-and-ensemble-of-multiple-risk-sc-peer-reviewed-article-RMHP

[52] Z. Lu, X. Wu, and J. C. Bongard, "Active learning through adaptive heterogeneous ensembling," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 2, pp. 368–381, Feb. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/6731529/

[53] B. Shao, N. Lavesson, V. Boeva, and R. K. Shahzad, "A mixture-of-experts approach for gene regulatory network inference," *Int. J. Data Mining Bioinf.*, vol. 14, no. 3, p. 258, 2016. [Online]. Available: http://www.inderscience.com/link.php?id=74876

[54] A. Doganer, C. Çolak, F. Küçükdurmaz, and C. Ölmez, "Prediction of renal cell carcinoma based on ensemble learning methods," *Middle Black Sea J. Health Sci.*, vol. 7, no. 1, pp. 104–114, Apr. 2021. [Online]. Available: http://dergipark.org.tr/en/doi/10.19127/mbsjohs.889492

[55] Y. Xu, X. Zhao, S. Liu, and W. Zhang, "Predicting long non-coding RNAs through feature ensemble learning," *BMC Genomics*, vol. 21, no. S13, p. 865, Dec. 2020. [Online]. Available: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07237-y

[56] A. Chatzimparmpas, R. M. Martins, K. Kucher, and A. Kerren, "VisEvol: Visual analytics to support hyperparameter search through evolutionary optimization," *Comput. Graph. Forum*, vol. 40, no. 3, pp. 201–214, Jun. 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/cgf.14300

[57] A. Biswas, R. Vasudevan, M. Ziatdinov, and S. V. Kalinin, "Optimizing training trajectories in variational autoencoders via latent Bayesian optimization approach," *Mach. Learn., Sci. Technol.*, vol. 4, no. 1, Mar. 2023, Art. no. 015011. [Online]. Available: https://iopscience.iop.org/article/10.1088/2632-2153/acb316

[58] K. Senagi and N. Jouandeau, "A non-deterministic strategy for searching optimal number of trees hyperparameter in random forest," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2018, pp. 73–80. [Online]. Available: https://fedcsis.org/proceedings/2018/drp/202.html

[59] A. Yasmin, W. Haider Butt, and A. Daud, "Ensemble effort estimation with metaheuristic hyperparameters and weight optimization for achieving accuracy," *PLoS ONE*, vol. 19, no. 4, Apr. 2024, Art. no. e0300296. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0300296

[60] H. S. Salem, M. A. Mead, and G. S. El-Taweel, "Particle swarm optimization-based hyperparameters tuning of machine learning models for big COVID-19 data analysis," *J. Comput. Commun.*, vol. 12, no. 3, pp. 160–183, 2024. [Online]. Available: https://www.scirp.org/journal/doi.aspx?doi=10.4236/jcc.2024.123010

[61] E. K. Yilmaz and H. Bakir, "Hyperparameter tunning and feature selection methods for malware detection," *Politeknik Dergisi*, vol. 27, no. 1, pp. 343–353, Feb. 2024. [Online]. Available: http://dergipark.org.tr/en/doi/10.2339/politeknik.1243881

[62] P. Gijsbers, F. Pfisterer, J. N. van Rijn, B. Bischl, and J. Vanschoren, "Meta-learning for symbolic hyperparameter defaults," in *Proc. Genetic Evol. Comput. Conf. Companion*, Lille, France, Jul. 2021, pp. 151–152. [Online]. Available: https://dl.acm.org/doi/10.1145/3449726.3459532

[63] R. O. Ogundokun, S. Misra, M. Douglas, R. Damaševičius, and R. Maskeliūnas, "Medical Internet-of-Things based breast cancer diagnosis using hyperparameter-optimized neural networks," *Future Internet*, vol. 14, no. 5, p. 153, May 2022. [Online]. Available: https://www.mdpi.com/1999-5903/14/5/153

[64] K. M. Beyene, D. Chen, and Y. G. Kifle, "A novel nonparametric time-dependent precision–recall curve estimator for right-censored survival data," *Biometrical J.*, vol. 66, no. 3, Apr. 2024, Art. no. 2300135. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/bimj.202300135

[65] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, Pittsburgh, PA, USA, 2006, pp. 233–240. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1143844.1143874

[66] K. Zuva, "Evaluation of information retrieval systems," *Int. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 3, pp. 35–43, Jun. 2012. [Online]. Available: http://www.aircsse.org/journal/jcsit/0612csit04.pdf

[67] H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich, "The area under the precision-recall curve as a performance metric for rare binary events," *Methods Ecol. Evol.*, vol. 10, no. 4, pp. 565–577, Apr. 2019. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13140

[68] M. Xie, Y. Deng, Z. Wang, Y. He, X. Wu, M. Zhang, Y. He, Y. Liang, and T. Li, "Development and assessment of novel machine learning models to predict the probability of postoperative nausea and vomiting for patient-controlled analgesia," *Sci. Rep.*, vol. 13, no. 1, p. 6439, Apr. 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-33807-7

[69] J. Xie, Z. Wang, Z. Yu, B. Guo, and X. Zhou, "Ischemic stroke prediction by exploring sleep related features," *Appl. Sci.*, vol. 11, no. 5, p. 2083, Feb. 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/5/2083

[70] B. K. Jang and Y. R. Park, "Development and validation of adaptable skin cancer classification system using dynamically expandable representation," *Healthcare Informat. Res.*, vol. 30, no. 2, pp. 140–146, Apr. 2024. [Online]. Available: http://e-hir.org/journal/view.php?doi=10.4258/hir.2024.30.2.140

[71] L. Tang, P. Du, and C. Wu, "Compare diagnostic tests using transformation-invariant smoothed ROC curves," *J. Stat. Planning Inference*, vol. 140, no. 11, pp. 3540–3551, Nov. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378375810002739

[72] S. Pérez-Fernández, P. Martínez-Camblor, P. Filzmoser, and N. Corral, "nsROC: An R package for non-standard ROC curve analysis," *R J.*, vol. 10, no. 2, p. 55, 2019. [Online]. Available: https://journal.r-project.org/archive/2018/RJ-2018-043/index.html

[73] Y. T. Van Der Schouw, H. Straatman, and A. L. M. Verbeek, "ROC curves and the areas under them for dichotomized tests: Empirical findings for logistically and normally distributed diagnostic test results," *Med. Decis. Making*, vol. 14, no. 4, pp. 374–381, Oct. 1994. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0272989X9401400408

[74] J. Hernández-Orallo, "ROC curves for regression," *Pattern Recognit.*, vol. 46, no. 12, pp. 3395–3411, Dec. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0031320313002665

[75] X. Song, X. Zhou, and S. Ma, "Nonparametric receiver operating characteristic-based evaluation for survival outcomes," *Statist. Med.*, vol. 31, no. 23, pp. 2660–2675, Oct. 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/sim.5386

[76] P. F. Evangelista, M. J. Embrechts, and B. K. Szymanski, "Data fusion for outlier detection through pseudo-ROC curves and rank distributions," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Vancouver, BC, Canada, Jul. 2006, pp. 2166–2173. [Online]. Available: https://ieeexplore.ieee.org/document/1716379

[77] C. Feng, T. Xiang, Z. Yi, X. Meng, X. Chu, G. Huang, X. Zhao, F. Chen, B. Xiong, and J. Feng, "A deep-learning model with the attention mechanism could rigorously predict survivals in neuroblastoma," *Frontiers Oncol.*, vol. 11, Jul. 2021, Art. no. 653863. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fonc.2021.653863/full

[78] M. Vuk and T. Curk, "ROC curve, lift chart and calibration plot," *Adv. Methodol. Statist.*, vol. 3, no. 1, pp. 89–108, Jan. 2006. [Online]. Available: https://mz.mf.uni-lj.si/article/view/44

[79] R. D. Clark and D. J. Webster-Clark, "Managing bias in ROC curves," *J. Comput.-Aided Mol. Des.*, vol. 22, nos. 3–4, pp. 141–146, Mar. 2008. [Online]. Available: http://link.springer.com/10.1007/s10822-008-9181-z

[80] D. P. Chakraborty, "ROC curves predicted by a model of visual search," *Phys. Med. Biol.*, vol. 51, no. 14, pp. 3463–3482, Jul. 2006. [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/51/14/013

[81] C.-L. Lee, W.-J. Liu, and S.-F. Tsai, "Development and validation of an insulin resistance model for a population with chronic kidney disease using a machine learning approach," *Nutrients*, vol. 14, no. 14, p. 2832, Jul. 2022. [Online]. Available: https://www.mdpi.com/2072-6643/14/14/2832

[82] Y. Huang, Z. Chen, X. Wang, C. Zheng, L. Shao, Y. Tian, X. Cao, Y. Tian, R. Gao, L. Zhang, and Z. Wang, "Comparison of the three most commonly used metabolic syndrome definitions in the Chinese population: A prospective study," *Metabolites*, vol. 13, no. 1, p. 12, Dec. 2022. [Online]. Available: https://www.mdpi.com/2218-1989/13/1/12

[83] S. Chen, F. Tao, C. Pan, X. Hu, H. Ma, C. Li, Y. Zhao, and Y. Wang, "Modeling quality changes in Pacific white shrimp (*Litopenaeus vannamei*) during storage: Comparison of the Arrhenius model and random forest model," *J. Food Process. Preservation*, vol. 45, no. 1, Jan. 2021, Art. no. e14999. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/jfpp.14999

[84] Y. Xu, W. Li, Y. Yang, S. Dong, F. Meng, K. Zhang, Y. Wang, L. Ruan, and L. Zhang, "Deep learning-based prediction of coronary artery calcium scoring in hemodialysis patients using radial artery calcification," *Seminars Dialysis*, vol. 37, no. 3, pp. 234–241, May 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/sdi.13191

[85] R. Liu, Y. Liu, F. Zhang, J. Wei, and L. Wu, "A cuproptosis random forest cox score model-based evaluation of prognosis, mutation characterization, immune infiltration, and drug sensitivity in hepatocellular carcinoma," *Frontiers Immunol.*, vol. 14, Mar. 2023, Art. no. 1146411. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1146411/full

[86] R. Sun and C. Li, "A risk prediction model of hard landing based on random forest algorithm," in *Proc. E3S Web Conf.*, vol. 257, 2021, p. 02080. [Online]. Available: https://www.e3s-conferences.org/10.1051/e3sconf/202125702080

[87] A. Si, J. Zhang, Y. Zhang, E. Kazuva, Z. Dong, Y. Bao, and G. Rong, "Debris flow susceptibility assessment using the integrated random forest based steady-state infinite slope method: A case study in Changbai mountain, China," *Water*, vol. 12, no. 7, p. 2057, Jul. 2020. [Online]. Available: https://www.mdpi.com/2073-4441/12/7/2057

[88] C. Gao, "Balancing interpretability and performance: Optimizing random forest algorithm based on point-to-point federated learning," *J. Electr. Syst.*, vol. 20, no. 7s, pp. 2389–2400, May 2024. [Online]. Available: https://journal.esrgroups.org/jes/article/view/3990

[89] C. Liu, "Implied volatility forecasting for American options based on random forest regressor, linear regression model," *Adv. Econ., Manage. Political Sci.*, vol. 85, no. 1, pp. 154–160, May 2024. [Online]. Available: https://aemps.ewapublishing.org/article/b4d1c32756df4fb2a6156c3e6c535246

[90] B. A. Goldstein, E. C. Polley, and F. B. S. Briggs, "Random forests for genetic association studies," *Stat. Appl. Genet. Mol. Biol.*, vol. 10, no. 1, p. 32, Jan. 2011. [Online]. Available: https://www.degruyter.com/document/doi/10.2202/1544-6115.1691/html

[91] A. J. Blomberg, L. Li, J. D. Schwartz, B. A. Coull, and P. Koutrakis, "Exposure to particle beta radiation in greater Massachusetts and factors influencing its spatial and temporal variability," *Environ. Sci. Technol.*, vol. 54, no. 11, pp. 6575–6583, Jun. 2020. [Online]. Available: https://pubs.acs.org/doi/10.1021/acs.est.0c00454

[92] X. Liu, L. Wang, Y. Hu, and S. Hua, "Prediction of thermal insulation performance of vacuum glass based on extreme random forest model," *Proc. SPIE*, vol. 12287, pp. 635–641, Oct. 2022. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12287/2640729/Prediction-of-thermal-insulation-performance-of-vacuum-glass-based-on/10.1117/12.2640729.full

[93] A. K. Rahimi, M. Ghadimi, A. H. van der Vegt, O. J. Canfell, J. D. Pole, C. Sullivan, and S. Shrapnel, "Machine learning clinical prediction models for acute kidney injury: The impact of baseline creatinine on prediction efficacy," *BMC Med. Informat. Decis. Making*, vol. 23, no. 1, p. 207, Oct. 2023. [Online]. Available: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-023-02306-0

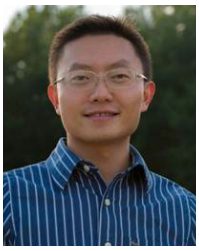
[94] S. Ogawa, "Predicting dropout from cognitive behavioral therapy for panic disorder using machine learning algorithms," *J. Clin. Med. Res.*, vol. 16, no. 5, pp. 251–255, May 2024. [Online]. Available: http://www.jocmr.org/index.php/JOCMR/article/view/5167

CEMIL EMRE YAVAS (Member, IEEE) received the B.Sc. degree in industrial engineering from Bosphorus University and the M.B.A. degree from Istanbul Bilgi University.

He is a Distinguished Industrial Engineer with over 30 years of global experience in the IT industry, having been actively involved in the field since the beginning of the Internet era. He founded Turkline, a leading digital services company, where he managed a diverse team of over 100 professionals and led numerous award-winning projects. He excels in full-stack software development, with specific expertise in mobile device development and home automation systems. He is also the manager for prestigious IT investment trusts, overseeing projects that primarily focus on enterprise resource planning (ERP) systems. His leadership in this area has significantly contributed to the strategic enhancement of business processes and operational efficiency for numerous organizations.

Mr. Yavas has received over 50 international awards for web design and development, including prestigious accolades, such as the W3 Awards, the Interactive Media Awards (IMA), and the Golden Spider Awards.

JONGYEOP KIM (Member, IEEE) received the B.Sc. degree in computer science from Korea National Open University, Seoul, South Korea, and the M.Sc. and Ph.D. degrees in computer science from Oklahoma State University. From 2016 to 2020, he was an Assistant Professor of computer science with Southern Arkansas University. Since August 2020, he has been an Assistant Professor of information technology with Georgia Southern University. He has extensive experience in the public sector, managing information systems for South Korean Military, where he held crucial positions, including the Assistant Director of the Information Management Division, Daejeon Chungnam and Gyeongin Regional Military Manpower Administrations, and an Officer at the Military Manpower Administration Headquarters. His research interests include data analytics, cybersecurity, and the application of IT solutions to complex systems. His combined academic and professional experiences provide him with robust theoretical knowledge and practical insights into information management and security.



LEI CHEN (Senior Member, IEEE) received the Ph.D. degree in computer science and software engineering from Auburn University. He is currently a tenured Professor and the Graduate Program Director with Georgia Southern University. He has authored over 100 peer-reviewed scholarly works, including notable publications in IEEE INTERNET OF THINGS JOURNAL and *Mobile Networks and Applications* (Springer). His book *Wireless Network Security: Theories and Practices* (Springer), has been widely appreciated in the field. His research interests include security, privacy, and digital forensics of networks and systems. His contributions extend to serving as an editor for high-impact journals and chairing numerous international conferences, with his work supported by U.S. National Security Agency (NSA) and the National Science Foundation (NSF).