Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Comprehensive Classification of Iris Flower Species: A Machine Learning Approach

Renas Rajab Asaad[1,2], Adnan Mohsin Abdulazeez[3]
renas.rekany@nawroz.edu.krd, adnan.mohsin@dpu.edu.krd
[1]Duhok Polytechnic University/Akre University for Applied Sciences - Technical College of Informatics-Akre- Department of Information Technology Duhok, KRG-Iraq
[2]Department of Computer Science, Nawroz University, Duhok, KRG-Iraq
[3]Technical College of Engineering-Duhok, Duhok Polytechnic University, Duhok, KRG-Iraq

Article Information

Submitted : 23 Jan 2024
Reviewed: 26 Jan 2024
Accepted : 10 Feb 2024

Keywords

Iris Flower Species, Decision Tree, Random Forest, SVM, KNN, XGBoost, Classification

Abstract

This study employs robust machine learning techniques to comprehensively assess the classification of Iris flower species. This study investigates the effectiveness of several machine learning algorithms in reliably classifying Iris flower species by utilizing a dataset that includes crucial morphological attributes such as sepal length, sepal width, petal length, and petal width. The algorithms under consideration are Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost. Every algorithm has its own distinct methodology for classification, where Decision Trees offer clear interpretability and Random Forest and XGBoost offer strong and complex ensemble techniques. The primary aim of this study is to assess and contrast different algorithms, considering not only their classification accuracy but also significant performance metrics including precision, recall, F1-score, ROC AUC, and specificity. This research provides valuable insights into the capabilities and constraints of each methodology when implemented on a meticulously organized and defined botanical dataset. It is expected that the results of this study will contribute significantly to the fields of artificial intelligence and botanical taxonomy, highlighting the capacity of these methods to accurately identify and categorize plant species.

A. Introduction

A subfield of artificial intelligence (AI), machine learning (ML) enables computers to autonomously learn and make decisions through the analysis and processing of data[1]. The development of algorithms capable of extracting insights from data and employing this knowledge to generate informed judgments or predictions constitutes machine learning. In the last few decades, machine learning has gained prominence in numerous industries, ranging from straightforward applications such as email categorization and speech recognition to more complex endeavors including medical analysis and financial forecasting[2].

Classification is an essential methodology in machine learning wherein algorithms are trained to precisely identify or categorize incoming data[3]. This approach is implemented in various real-world situations and is founded on the prediction of categorical results. In the domain of medical diagnosis, classification algorithms exhibit their utility through the differentiation of benign from malignant malignancies. In the financial sector, these algorithms are implemented to authenticate transactions and detect possible instances of fraudulent activity[4][5].

Classification algorithms are commonly taught using datasets that consist of varied input features and their related labels. The goal is for the algorithm to identify patterns or links in the data, allowing it to make precise predictions for fresh, unseen data points[6][7]. The learning method entails dividing the dataset into separate training and testing sets. The training set is utilized to instruct the model, whilst the testing set is kept aside for assessing its performance[8].

There is a diverse selection of categorization algorithms, each possessing distinct benefits and constraints. The choice of the best appropriate algorithm is contingent upon elements such as the attributes of the data and the precise requirements of the given task[9]. Several widely used categorization algorithms include Decision Trees, Random Forests, Support Vector Machines (SVMs), K-Nearest Neighbors (KNN), and Neural Networks (NN), among others[10].

The field of machine learning is constantly progressing, particularly in the area of classification, where there is continuous growth through the introduction of advanced approaches like ensemble methods and deep learning. These technological breakthroughs are facilitating the development of more precise predictive models and revealing profound insights within intricate datasets. Machine learning categorization techniques are advancing and have a significant impact on numerous sectors by providing intelligent, data-driven solutions and improving the capabilities of computers to learn[11][12].

This study introduces thorough investigation into the Iris flower species classification using advanced ML algorithms. This study delves into one of the most fundamental yet intriguing problems in the realm of botany and machine learning, employing a range of sophisticated algorithms to categorize Iris species based on various morphological features[13]. The Iris flower dataset, a cornerstone in the field of machine learning due to its clarity and well-defined features, serves as the foundation for this exploration.[14] The research utilizes a variety of algorithms, each bringing its unique strengths and methodologies to the task of species classification[15].

These machine learning algorithms offer diverse approaches to data classification and prediction. Decision Trees provide interpretable models based on simple decision rules, suitable for both numerical and categorical data[16]. Random Forest, an ensemble method, enhances prediction accuracy by combining multiple decision trees, making it

robust, accurate, and capable of handling missing data[17]. Support Vector Machine (SVM) excels in high-dimensional spaces by constructing hyperplanes for effective class separation, particularly valuable when class margin is crucial. K-Nearest Neighbors (KNN) relies on proximity-based classification, simplifying the process by grouping similar data points. Lastly, XGBoost, an eXtreme Gradient Boosting algorithm, employs decision trees within a gradient boosting framework, offering high efficiency, accuracy, and versatility for various data types[18][19].

The study aims not just to classify the Iris species accurately but also to compare the effectiveness, efficiency, and practicality of these algorithms in real-world scenarios[20]. By applying these diverse machine learning techniques, the research provides valuable insights into their comparative performance, including aspects like accuracy, precision, recall, and the ability to handle data complexities[21]. This comprehensive classification approach contributes significantly to the broader understanding of machine learning applications in biological classification, offering a robust framework for future studies in this intersecting field of botany and computational intelligence[22].

The study leverages uses advanced machine learning algorithms to sort Iris flower species into groups. To test how well the algorithms work, a dataset with important physical traits is used[23]. As well as accuracy, it rates algorithms like Decision Tree, Random Forest, SVM, KNN, and XGBoost based on performance metrics like recall, F1-score, ROC AUC, sensitivity, and precision. The goal of this study is to teach us about the pros and cons of each method for classifying plants, which will add to the fields of artificial intelligence and botanical taxonomy[24].

This meticulously structured paper provides a clear and comprehensive analysis. The Introduction in Section 1 introduces the complex challenges faced by the iris flower. Section 2, the Literature Review, presents a comprehensive analysis of previous studies focused on the use of machine learning for iris flower classification. The Proposed Model in Section 3 outlines the methodological approach, detailing preprocessing techniques and classifier methodologies. Section 4, Dataset, provides an in-depth description of the data's composition and categorization. In Section 5, Evaluation Metrics, the paper discusses the criteria and approaches used to evaluate the model's performance. The Results and Discussion in Section 6 present a detailed analysis of the model’s effectiveness, offering comparisons with existing research. Section 7 discusses these comparative insights further. Finally, the Conclusion in Section 8 highlights the study’s significant contributions and implications for advancing plants through machine learning technologies.

B. Literature Review

The literature review highlights the significance of machine learning in Iris flower classification. Multiple studies utilize diverse machine learning algorithms, including XGBoost, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Trees, and Random Forests, to classify Iris flowers accurately. These studies emphasize the importance of accuracy, showcasing exceptional results approximate 100% accuracy. The rich biodiversity of Earth and the challenges posed by similar physical characteristics among species underscore the relevance of these classification efforts[25]. As the field of ML continues to evolve, Iris flower classification remains a valuable domain for exploring the capabilities of various algorithms and advancing the understanding of pattern recognition and classification[26].

(Zhou, 2022) [27] introduced Classification algorithms within the field of machine learning are crucial for categorizing data into distinct classes, making them invaluable for tasks like data mining. The process of selecting the most suitable classification model, one that offers both high accuracy and efficiency, is both important and challenging. To address this issue, this research paper conducts a comparative analysis of Multilayer Perceptron (MLP) in conjunction with various other ML methods, including KNN, SVM, Logistic Regression, Decision Trees, and Random Forests. All these models are trained to classify different types of Iris flowers to enhance our understanding of their capabilities. The performance of these models is evaluated using metrics like accuracy and confusion matrices. The experimental findings reveal that the confusion matrices generated by these diverse models exhibit similarities, but MLP stands out by consistently achieving higher accuracy. By employing an adequate number of nodes per hidden layer for the backpropagation process, MLP demonstrates exceptional analytical capabilities and delivers valuable projections. Given this superiority, it is anticipated that neural networks will continue to advance in their classification performance. Armed with an understanding of the strengths and limitations of various classification algorithms, we can look forward to the development of optimized models that better address real-world problems.

(Akter et al., 2022) [28] presented the Iris dataset is well-known, and it contains data with four attributes: Sepal.length, Sepal.width, Petal.length, and Petal.width. It is divided into three separate groups or subspecies, Sentosa, Versicolor, and Virginic, each with 50 samples. The measurements for these attributes are in centimeters. Ronald Fisher created this dataset in 1936, and it is easily accessible through the UCI dataset repository. This research will show how to solve classification problems using several methods such as K-means clustering, Random Forest decision, SVM, LR, KNN, and K-means. In addition, we investigate the translation of the four features into advanced features. We use the scikit-learn program for implementation. In this study, we use classification and regression techniques on the Iris dataset to find and analyze underlying patterns and correlations in the data.

(Taha Chicho et al., 2021) [29] this topic presented the classification is very common in modern machine learning, and it is widely used in a variety of domains such as face recognition, flower categorization, clustering, and other applications. The primary goal of this research project is to methodically arrange and classify a group of data objects. The study used the K-nearest neighbors, decision tree (j48), and random forest methods to accomplish this. Following that, it evaluates and compares their performance using the IRIS dataset. The results of this comparison analysis show that the K-nearest neighbors algorithm outperforms the other classifiers in terms of performance. Furthermore, when compared to the decision tree (j48), the random forest classifier outperforms it. Surprisingly, the research produced a 100% accuracy rate with no errors for the used classifier.

(Yehya Hussien, 2022) [30] introduced Creating a classification is a useful tool for predicting the group to which data instances belong, and the incorporation of ML approaches has aided in the classification process's simplification. The basic goal of this research is to use ANN and SVM to create models for categorizing Iris blossoms. Ronald Fisher, a British statistician and biologist, first created the Iris flower dataset, which is a multivariate dataset. It presents a significant barrier in taxonomic study, particularly in identifying between three iris flower species (Setosa, Versicolor, and Verginica) based on sepal and petal diameters. The main objective entails recognizing

patterns in the sizes of Iris flower sepals and petals and using these patterns to determine the flower's class. The results of the experiments demonstrated the usefulness of both ANN and SVM in effectively categorizing Iris flowers, with high accuracy rates of 98.66% and 97.79%, respectively.

(Swain, 2012) [31] showed the classification, a ML technique, is employed to predict the membership of data instances in specific groups or categories. In order to streamline the classification process, neural networks are being introduced. This research paper is centered around the classification of IRIS plants using Neural Networks as a tool. The core problem addressed in this study revolves around identifying the species of IRIS plants based on measurements of various plant attributes. The classification of the IRIS dataset involves the extraction of discernible patterns by analyzing the sizes of petals and sepals in IRIS plants. These patterns are then leveraged to make predictions and assign classes to different IRIS plant specimens. This approach not only aids in accurate classification but also sets the stage for more precise predictions of unknown data in the future. Artificial neural networks have proven to be highly effective in addressing a wide range of problems, including pattern classification, function approximations, optimization tasks, and associative memories. In this study, the focus is on training Multilayer feed-forward networks using the backpropagation learning algorithm.

(Aziz & Awrahman, 2023) [32] focused is supervised machine learning algorithms play a pivotal role in classification tasks. In this study, we focus on predicting the Iris plant's species using a variety of machine learning algorithms. The classification of Iris plants relies on various factors, including the length and width of certain properties. Horticultural experts have noted that some plants exhibit differences in physical attributes like size, shape, and color, making species recognition challenging. The Iris flower species consists of three unique subspecies: Versicolor, Setosa, and Virginica. This study utilizes machine learning methods to precisely categorize all types of Iris blooms, attaining a 100% accuracy rate for KNN, 95% for Random Forest (RF), 97% for Decision Tree (DT), and 98% for Logistic Regression (LR). The Iris dataset, easily accessible and deployed through Scikit tools, serves as the basis for developing predictive models for plants. In this work, machine learning methods such as LR, DT, KNN, and RF are utilized to create a strong prediction model.
(Srinivasarao et al., 2021) [33] demonstrated how, within the domain of Machine Learning, we identify Iris flower species by employing semi-automated techniques to extract knowledge from data. In the context of supervised learning, classification is utilized when the response variable is categorical in nature, comprising values that fall within a finite, unordered set. We employ the scikit-learn tools in order to streamline the classification process. This research paper is centered around the classification of Iris flowers using Machine Learning techniques with the assistance of scikit-learn tools. The central problem revolves around identifying Iris flower species based on measurements of various flower attributes. The process of classifying the Iris dataset involves the discovery of patterns by analyzing the sizes of petals and sepals in Iris flowers. These patterns function as the fundamental basis for generating predictions and allocating the suitable category to every Iris flower. In this manuscript, we employ data to train a machine learning model. When confronted with new, unobserved data, the predictive model utilizes the acquired knowledge from the training data to accurately classify the species of the Iris flower. This exemplifies the significant capability of learned patterns in the context of classification tasks.

(Shukla et al., n.d.) [34] The study revealed that the biodiversity of the Earth is extensive, with over 360,000 species coexisting in its ecosystems. Nevertheless, a significant number of these species exhibit very analogous physical attributes, including form, magnitude, and hue, so complicating their identification and categorization. The Iris flower species is an exemplary instance, consisting of three unique subspecies: Setosa, Versicolor, and Virginica. In order to tackle this classification difficulty, we employ the Iris dataset, which is a widely accessible and commonly utilized resource. The dataset comprises three distinct classes, with each class including 50 cases, rendering it a highly significant resource for machine learning applications. Through the utilization of machine learning methodologies, we can effectively and precisely detect and categorize the several subclasses of the Iris flower, surpassing mere estimations. The technique consists of three essential stages: segmentation, feature extraction, and classification. Our goal is to create a strong framework for automatically recognizing different flower classes by utilizing a range of machine learning algorithms, such as Neural Network, Logistic Regression, Support Vector Machine, and k-Nearest Neighbors. We prioritize achieving high accuracy in the classification process.

(Vatshayan, 2019) [35] worked on a ML, in the context of computer science, refers to a subfield that enables systems to gain information without relying on explicit programming. Supervised Machine Learning (SML) involves the use of algorithms to analyze externally provided data examples and create general hypotheses. These hypotheses are then used to produce predictions about future instances. In the context of SML, we face two types of learning: classification and regression. Both of these types belong to the category of supervised learning, where the response variable is categorical and can take values from finite discrete and continuous sets. In order to make the classification process more efficient, we utilize the scikit-learn tools. This study article primarily focuses on the categorization of Iris blossoms utilizing Machine Learning methodologies, employing scikit-learn tools as our software suite. This research aims to build a machine learning model using the Iris Dataset and evaluate the performance and accuracy of Iris classification using different Supervised Learning Algorithms.

(Rana et al., n.d.) [36] highlighted classification is a widely applied machine learning problem with diverse applications in fields such as security, agriculture, and industry. This research paper focuses on the classification of Iris flowers, utilizing Logistic Regression and Random Forest algorithms. To enhance the classification process, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are employed as feature extraction methods in both cases. In the second part of the study, a comparative analysis is conducted to evaluate the performance of both machine learning methods and both dimensionality reduction techniques. The comparative study findings highlight the superiority of LDA over PCA. Interestingly, when LDA is used, both Logistic Regression and Random Forest methods yield nearly identical results, showcasing the effectiveness of LDA in enhancing classification accuracy.

(Gupta et al., 2022) [37] investigated modern machine learning, which focuses on data categorization using advanced models and forecasting future outcomes based on these algorithms. It has been used in a variety of disciplines, including facial recognition, speech recognition, medical diagnostics (such as anticipating possible heart failure), sentiment analysis, and product recommendations. This study gives three categorization algorithms for reliably forecasting Iris flower species. The proposed model examines and preprocesses the dataset using Exploratory Data Analysis (EDA). To perform the prediction task, three classification models, namely "Logistic

Regression," "SVM," and "KNN," are used. The Iris dataset is used to test the three proposed models, which result in maximum accuracies of 96.43%, 98.21%, and 94.64%, respectively. This paper examines various supervised machine learning models that are suitable for detecting Iris flower species based on factors such as sepal width, sepal length, petal width, and petal length. It provides vital information about the performance of various models.

(Agarwal et al., 2022) [38] acknowledged the machine learning is undeniably one of the most dynamic and rapidly evolving fields in both research and industry. New techniques and methodologies continue to emerge at an astonishing rate, making it challenging for even experts to keep pace, let alone beginners. It's a well-established fact that incorporating diverse perspectives and inputs enhances the decision-making process, ultimately leading to more accurate results. In this context, the term "Ensemble Learning" within the realms of artificial intelligence (AI) and machine learning (ML) is recognized for its ability to enhance the capabilities, efficiency, robustness, and, most importantly, the accuracy of decision-making systems. Ensemble Learning is a powerful approach that leverages the collective wisdom of multiple models to improve the overall performance and reliability of machine learning systems.

(Pachipala et al., 2022) [39] delved an AI approach called "Order" that is used to forecast the attendance of a crowd in informational events. Neural networks are implemented as a means to improve the precision of predictions in this particular environment. The main objective of this paper, however, revolves around the categorization of Iris plants using neural networks. The classification of Iris plant species relies on the precise measurement of the breadth of both the sepal and petal. Furthermore, the investigation delves into the utilization of sepal length and petal length measures as a means to enhance the precision of classification in subsequent endeavors. Artificial neural networks, specifically neural networks, are widely utilized in diverse applications including pattern identification, job scheduling, forecasting future job advancements, and job remembering. The main aim of this study is to categorize the many species of Iris flowers by utilizing a dataset specifically curated by biologists for the goal of studying flower kinds using data mining techniques and neural network classifiers. The primary obstacle in this study involves devising an innovative methodology for categorizing Iris flowers and discerning their respective varieties, with the intention of assisting biologists in their research through the utilization of sophisticated ML methods. This problem is classified as a classification problem, and the proposed solution entails utilizing a supervised machine learning classification technique, namely the Random Forest Classification technique, applied to the Iris dataset. The experimental findings demonstrate a fundamental error rate of 0.01067, a training duration of 0.691 milliseconds, and the utilization of four neurons in the neural network, confirming the efficacy of the suggested methodology.

(Pawar et al., 2022) [40] underscored pattern detection using is a crucial component of machine learning, and accurately identifying these patterns is essential for tackling performance-related difficulties. Machine Learning (ML), a component of Artificial Intelligence (AI), plays a crucial role in activities such as categorization, grouping, and forecasting. This research study focuses on the creation of an innovative classification technique specifically designed to categorize Iris plants and accurately characterize their flower patterns. An improved ensemble model has been developed by the researchers in order to detect and categorize these patterns. In the early stages, a variety of models are implemented, including Decision Tree, OneR, Adaboost, Random

Forest, and Bayesnet. To enhance performance, an ensemble model that has been proposed is implemented throughout the classification procedure.

(Pinto et al., 2018) [41] Classification was highlighted as a critical and indispensable element of machine learning, functioning as the primary method for processing data. Machine learning comprises innumerable classification algorithms, such as Backpropagation, Decision Trees, Naive Bayes, Neural Networks, Artificial Neural Networks, Multi-layer Perceptrons, Multi-class Classification, Support Vector Machines, and K-Nearest Neighbors. The principal objective of this study is to furnish comprehensive explanations of three unique classification methodologies. The Scikit-learn library and the Iris dataset are utilized to implement these methods. The principal objective of this work is to evaluate the Iris dataset using classification and regression techniques, with an emphasis on analyzing and examining patterns derived from the size of the flower's sepal and petal. The results of the study show that the Support Vector Machine (SVM) classifier outperforms the K-Nearest Neighbors (KNN) and Logistic Regression models in terms of accuracy, highlighting its utility in classification applications[42].

In summary, these studies collectively illustrate the evolving landscape of the iris flower, where various machine learning and artificial intelligence technologies are being increasingly integrated to enhance the accuracy and reliability of iris flower classification, promising significant improvements in plant care and better monitoring outcomes.

C. Proposed Model

The proposed model for iris flower classification is developed to assess the performance of Decision Tree, Support Vector Machine (SVM), Random Forest, KNN, and XGBoost. The Dataset is fed into the aforementioned classifiers in four forms as mentioned below.

1- Raw Dataset without Preprocessing and Feature Selection.
2- Dataset after preprocessing with MinMaxScaler.
3- Dataset after feature selection using SelectKBest.
4- Dataset after preprocessing using MinMaxScaler and feature selection using SelectKBest.

Afterwards the data is split and run through the classifiers.

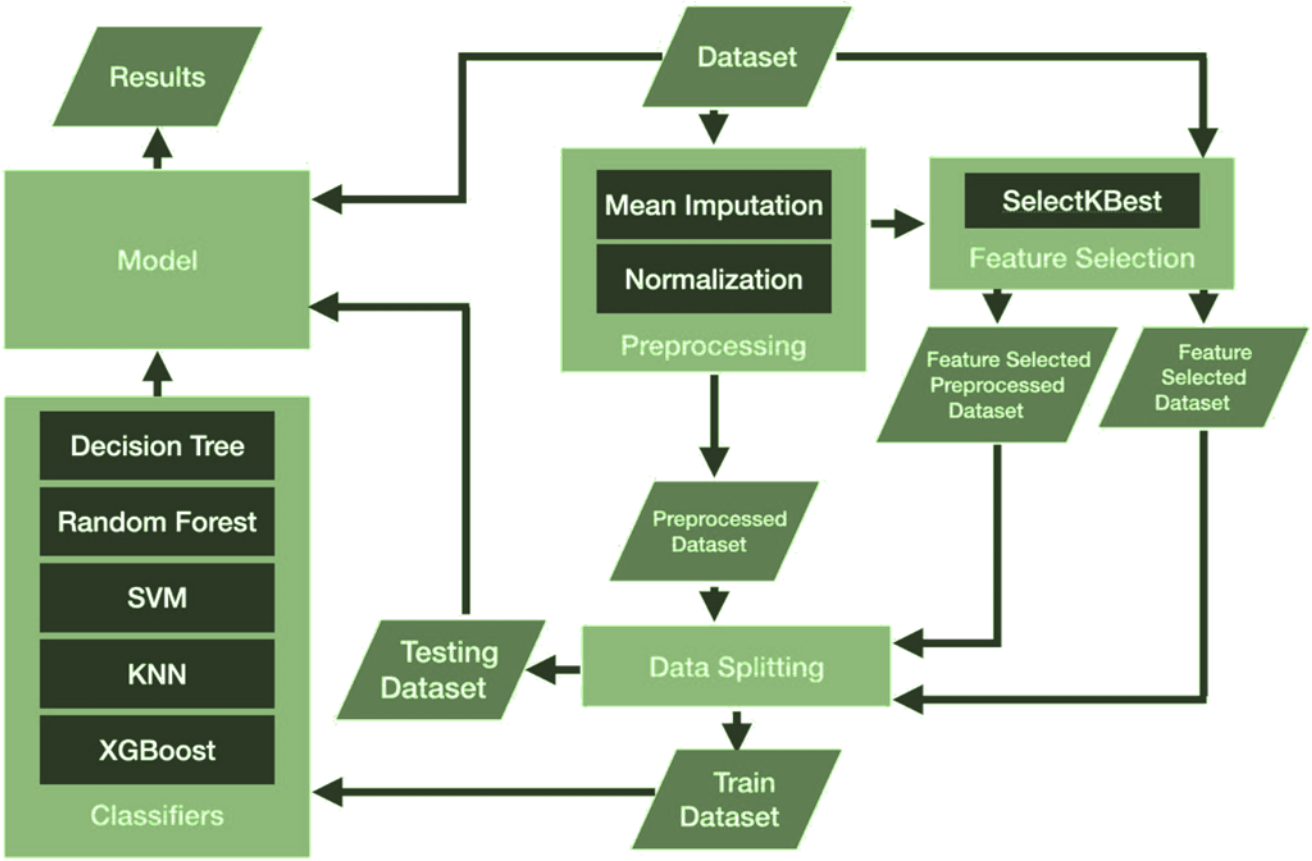

Figure 1. Proposed Model

1. Preprocessing: Normalization

Normalization and scaling are essential preprocessing steps in machine learning, especially for models dealing with complex data such as iris species classification. Normalization adjusts the scale of data attributes, ensuring that each feature contributes equally to the model's learning process. This is crucial because features with larger numerical ranges could disproportionately influence the model's performance. Techniques like MinMaxScaler, StandardScaler, and RobustScaler are employed to rescale data to a specific range or to adjust it based on statistical properties such as mean and standard deviation. Normalization thus ensures that each feature contributes approximately proportionately to the final prediction. In the case of the mentioned dataset, which likely contains a variety of features, normalization would ensure that no single feature dominates the model due to its scale, leading to more accurate and stable model performance[43].

2. Feature Selection

Feature selection plays a pivotal role in the preprocessing phase of machine learning models, particularly in the context of the dataset we are discussing, which might be involved in areas like biological classification or customer segmentation. This process involves pinpointing the most impactful features from a possibly extensive dataset, a step that significantly enhances the efficacy of predictive models. In our model's context, feature selection techniques are crucial to distill the essence of the data, focusing on the most informative attributes[44].

Among various techniques, Mutual Information (MI) emerges as an exceptionally efficient method for our dataset. This statistical, filter-based approach is adept at

quantifying the relevance of each feature by assessing its correlation with the target variable. By presuming that features with stronger correlations to the class labels are more valuable for classification tasks, MI helps in refining the feature set. This is particularly important for our dataset, which might contain a mix of highly relevant and less pertinent features. By applying MI, we can ensure that the machine learning models, be it Decision Trees or SVMs, are trained on the most informative features, leading to improved accuracy and generalizability in predictions. In supervised classification of multivariate time series, such as in iris species classification using iris flower data, mutual information is used to measure the relevance of each feature subset. This method of feature selection using mutual information while maintaining or increasing classification accuracy it can significantly reduce the number of features[45].

3. Classifiers

The proposed model relies on an array of classifiers. Namely, Decision tree, Random Forest, Support Vector Machine, KNN, XGBoost.

Sure, here are descriptions for each of the mentioned machine learning algorithms:

3.1. Decision Tree:

A Decision Tree is a type of machine learning algorithm that is used to perform classification and regression tasks under the guidance of a supervisor. The algorithm operates by iteratively dividing the dataset into smaller groups, using the most influential trait or characteristic as the basis for partitioning at each step. The procedure results in the creation of a hierarchical structure resembling a tree. Each internal node in the tree represents a judgment or test, while each leaf node represents a class label or numerical value. Decision Trees possess a high degree of interpretability and visualizability, rendering them valuable for comprehending the decision-making procedure within the model[46].

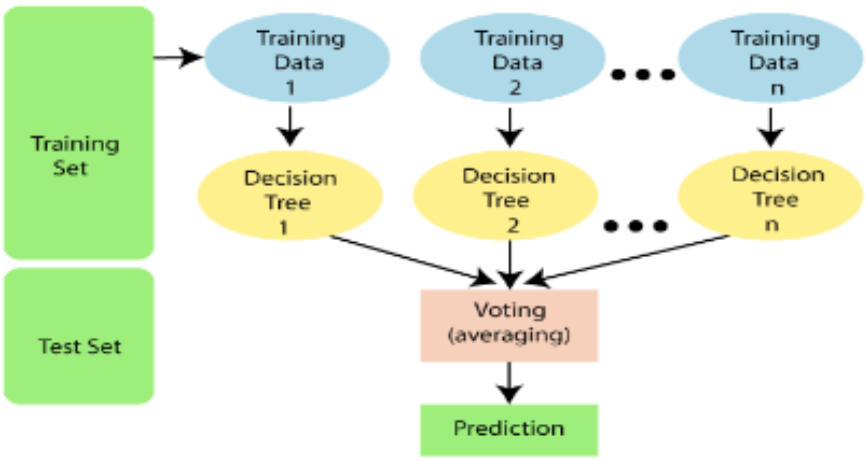

Figure 2. Decision Tree

3.2. Random Forest:

Random Forest is an ensemble learning method that combines multiple Decision Trees to improve prediction accuracy and reduce overfitting. It works by creating a collection of Decision Trees, each trained on a different subset of the data and using random feature subsets. During prediction, the results from individual trees are aggregated (e.g., through voting for classification or averaging for regression) to make a final prediction. Random Forest is known for its robustness and ability to handle complex datasets.[48]

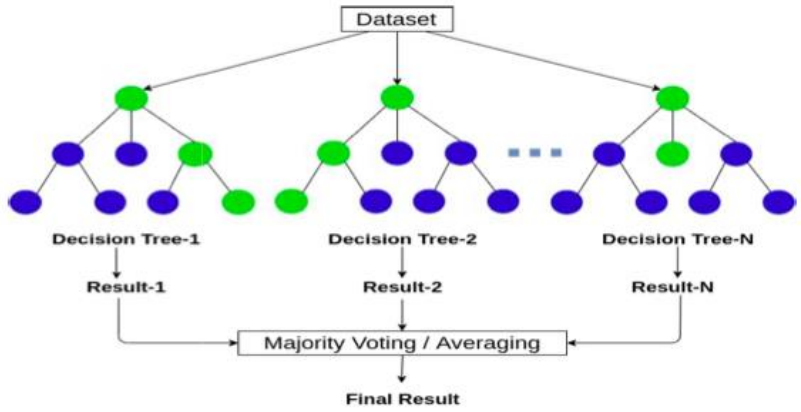

Figure 3. Random Forest [29]

3.3. Support Vector Machine (SVM):

SVM, is a robust supervised machine learning technique that is utilized for both classification and regression applications. The algorithm identifies the most effective hyperplane (decision boundary) that maximizes the separation between several classes in a space with many dimensions. SVM is capable of processing both linear and nonlinear data by employing various kernel functions. It is efficient for categorizing data points into separate classes and is particularly valuable when working with datasets that exhibit evident class differentiation. [49]

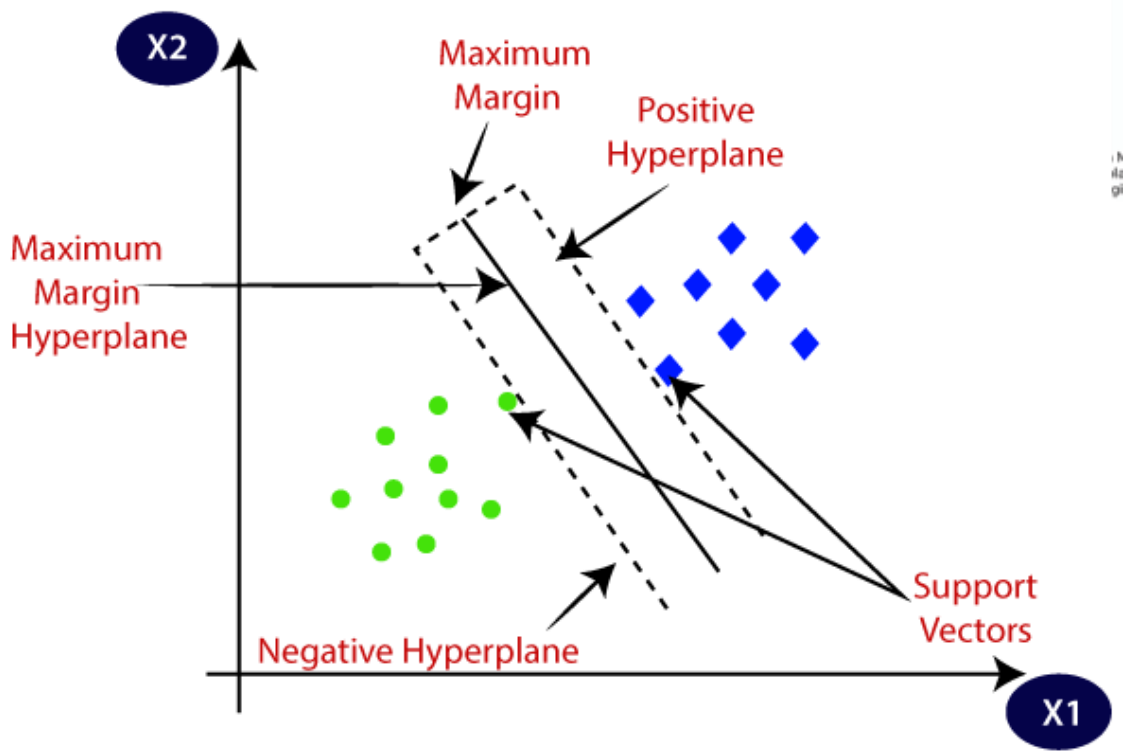

Figure 2. SVM [50]

3.4. K-Nearest Neighbors (KNN):

KNN is a straightforward and intuitive machine learning method usually employed for classification jobs. It categorizes data points by comparing their similarity to the k-nearest neighbors in the training dataset. The variable "k" denotes the quantity of

nearest neighbors that are taken into consideration[51]. KNN utilizes distance metrics, such as Euclidean distance, to ascertain similarity and allocate a class label by means of majority voting among the closest neighbors. The algorithm is non-parametric and its performance can be influenced by the selection of the "k" value. [52]

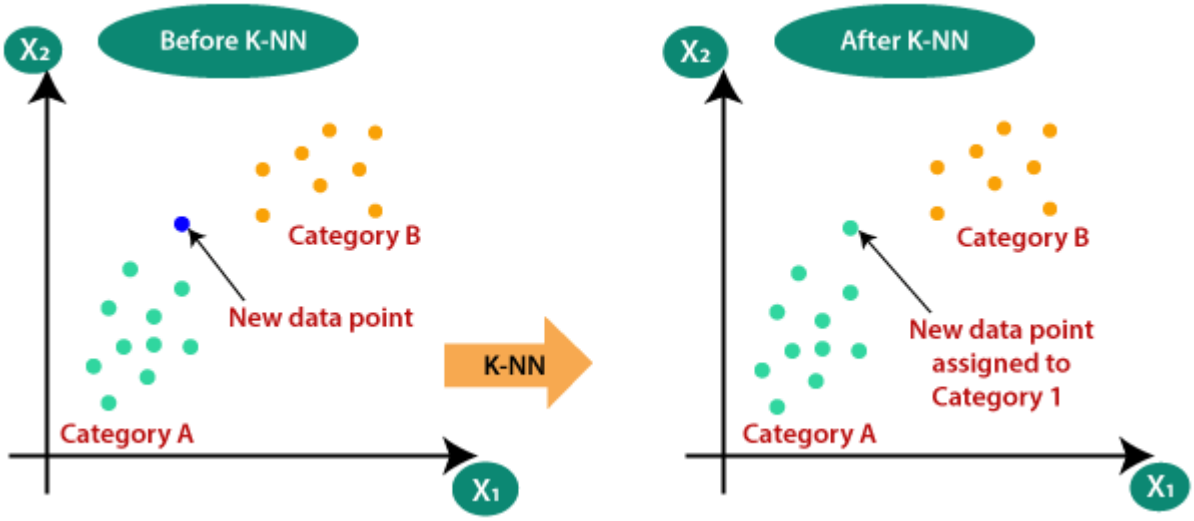

Figure 5. KNN [53]

3.5. XGBoost:

XGBoost, short for Extreme Gradient Boosting, is an ensemble learning method that belongs to the gradient boosting family. It's widely used in both classification and regression tasks and is known for its high performance. XGBoost builds an ensemble of decision trees sequentially, where each subsequent tree corrects the errors of the previous ones. It employs a gradient descent optimization technique to minimize a specified loss function, making it highly efficient and capable of handling large datasets. XGBoost often outperforms other algorithms in machine learning competitions and real-world applications.[54]

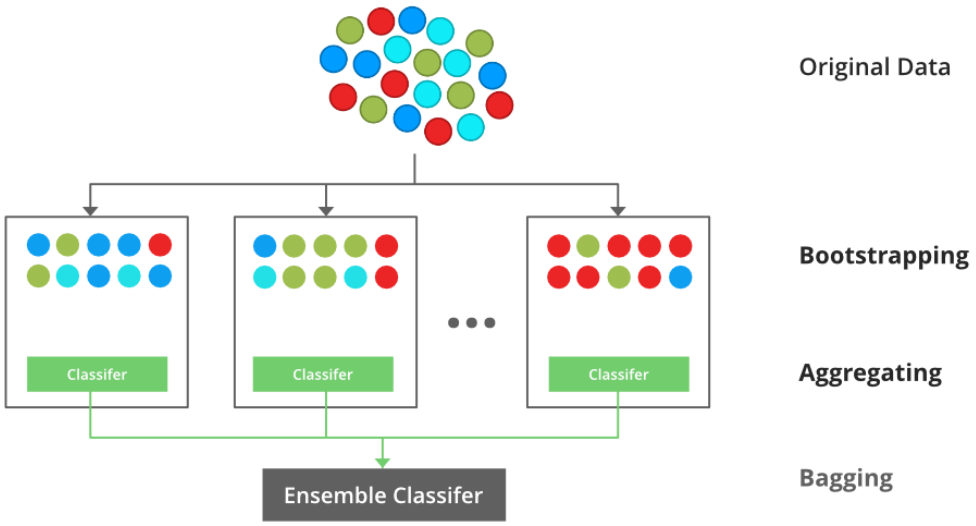

Figure 6. XGBoost [55]

D. Dataset

The dataset mentioned above, presumably an extended version of the classic Iris dataset, represents a more comprehensive approach to species classification with its inclusion of additional features like elevation, soil type, and various calculated ratios and areas. This enriched dataset not only includes the traditional measurements such as sepal length, sepal width, petal length, and petal width but also integrates ecological aspects like soil type and elevation, offering a multidimensional perspective on iris species classification. Such enhancements potentially allow for more nuanced and accurate modeling, enabling the application of advanced machine learning techniques like Random Forest, SVM, and XGBoost. These classifiers, when applied to the dataset, would be expected to discern patterns not just from the basic morphological traits of the iris flowers but also from their ecological contexts, thereby potentially increasing the accuracy and robustness of species classification. This extended dataset, therefore, represents a significant step forward in leveraging machine learning for botanical studies, offering a richer set of variables for understanding and predicting the categorization of iris species.

E. Evaluation Metrics

The model performance is measured for each of the dataset forms according to Accuracy, Precision, Recall, F1-Score, ROC AUC (Receiver Operating Characteristic Area Under the Curve), and Confusion Matrix.

$$\text{Accuracy} = \frac{(True\ Positives + True\ Negatives)}{Total\ Samples} \quad (1)$$

$$\text{Precision} = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

$$\text{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

$$\text{F1 Score} = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (4)$$

The ROC curve and the Area Under the Curve (AUC) are critical in assessing the performance of binary classification models. The ROC curve graphically displays the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) at various thresholds, illustrating the balance between sensitivity (recall) and specificity. The AUC, which ranges from 0 to 1, measures the model's capacity to distinguish between classes[56]. A perfect model would have an AUC of 1, while an AUC of 0.5 indicates no predictive ability, akin to random guessing. This metric is especially useful for evaluating models in scenarios with imbalanced classes, as it provides a comprehensive assessment independent of class distribution.[57]

A confusion matrix is an important statistical tool for visually depicting and evaluating the effectiveness of a categorization system. The tabular style displays the number of correct and incorrect forecasts in comparison to the actual results. True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are the four elements of the matrix. Many performance indicators, including as accuracy, precision, recall, and the F1 score, rely on these components. The confusion matrix configuration is very useful for identifying specific categories of errors in the model, such as persistent misclassifications, which can provide substantial insights for

implementing future improvements. It provides vital insights into the model's performance across classes, which is critical for understanding how it operates in datasets with uneven class distribution.[58]

F. Results and discussion

This section highlights the acquired results of the model along with a detail description of the dataset and the evaluation metrics used.

1. Obtained Results

In this research paper, the application of various ML models, on a dataset of 1201 individual records from iris species exams, was thoroughly evaluated. These records, classified into three categories Iris Setosa, Iris Versicolor, and Iris Virginica, provided a diverse and comprehensive basis for assessing the effectiveness of each model in distinguishing between different states of iris species. Below is a detailed presentation of the results achieved.

Table 1. Results Without preprocessing and feature selection

Classifier	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Confusion Matrix	Specificity
Decision Tree	0.971	0.971	0.971	0.971	0.977	[[68, 0, 0], [0, 81, 3], [0, 4, 84]]	0.973
Random Forest	0.975	0.975	0.975	0.975	0.999	[[68, 0, 0], [0, 81, 3], [0, 3, 85]]	0.977
SVM	0.946	0.948	0.946	0.946	0.995	[[68, 0, 0], [0, 81, 3], [0, 10, 78]]	0.951
KNN	0.967	0.967	0.967	0.967	0.996	[[68, 0, 0], [0, 82, 2], [0, 6, 82]]	0.969
XGBoost	0.979	0.979	0.979	0.979	0.998	[[68, 0, 0], [0, 82, 2], [0, 3, 85]]	0.981

In this table, all classifiers perform commendably well despite the absence of preprocessing and feature selection. This high performance could be attributed to the inherent simplicity and distinctiveness of the Iris dataset's features. The Random Forest and XGBoost classifiers show slightly superior performance compared to others, indicating their robustness even without preprocessing.

Table 2. Results after preprocessing without feature selection

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Confusion Matrix	Specificity
Decision Tree	0.971	0.971	0.971	0.971	0.977	[[68, 0, 0], [0, 81, 3], [0, 4, 84]]	0.973
Random Forest	0.983	0.983	0.983	0.983	0.999	[[68, 0, 0], [0, 82, 2], [0, 2, 86]]	0.984
SVM	0.992	0.992	0.992	0.992	1.0	[[68, 0, 0], [0, 83, 1], [0, 1, 87]]	0.992
KNN	0.988	0.988	0.988	0.988	0.996	[[68, 0, 0], [0, 84, 0], [0, 3, 85]]	0.989
XGBoost	0.979	0.979	0.979	0.979	0.998	[[68, 0, 0], [0, 82, 2], [0, 3, 85]]	0.981

In this table, the performance of all classifiers, particularly SVM and KNN, has significantly improved post preprocessing. This enhancement suggests that standardizing the features and handling missing values or outliers (if any) in the dataset

greatly benefited these models. The improvement in SVM's ROC-AUC to a perfect score of 1.0 demonstrates the impact of preprocessing on models sensitive to feature scales and distributions.

Table 3. Results after feature selection without preprocessing

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Confusion Matrix	Specificity
Decision Tree	0.954	0.954	0.954	0.954	0.964	[[68, 0, 0], [0, 78, 6], [0, 5, 83]]	0.957
Random Forest	0.967	0.968	0.967	0.967	0.996	[[68, 0, 0], [0, 83, 1], [0, 7, 81]]	0.97
SVM	0.95	0.952	0.95	0.95	0.995	[[68, 0, 0], [0, 81, 3], [0, 9, 79]]	0.954
KNN	0.962	0.965	0.962	0.962	0.988	[[68, 0, 0], [0, 83, 1], [0, 8, 80]]	0.967
XGBoost	0.962	0.964	0.962	0.962	0.995	[[68, 0, 0], [0, 82, 2], [0, 7, 81]]	0.966

The accuracy and other metrics slightly decrease for some models compared to Table 1. This could indicate that the selected features, while reducing model complexity, might have omitted some informative attributes. Despite this, the classifiers maintain a high level of performance, suggesting that the most relevant features for classification were retained.

Table 4. Results after preprocessing and feature extraction

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Confusion Matrix	Specificity
Decision Tree	0.958	0.958	0.958	0.958	0.968	[[68, 0, 0], [0, 79, 5], [0, 5, 83]]	0.961
Random Forest	0.967	0.968	0.967	0.967	0.996	[[68, 0, 0], [0, 83, 1], [0, 7, 81]]	0.97
SVM	0.954	0.955	0.954	0.954	0.995	[[68, 0, 0], [0, 81, 3], [0, 8, 80]]	0.958
KNN	0.958	0.961	0.958	0.958	0.988	[[68, 0, 0], [0, 83, 1], [0, 9, 79]]	0.963
XGBoost	0.962	0.964	0.962	0.962	0.995	[[68, 0, 0], [0, 82, 2], [0, 7, 81]]	0.966

This table shows a balanced approach with both preprocessing and feature selection applied. The results are comparable to Table 2, which highlights that feature selection, when combined with preprocessing, does not necessarily enhance the model performance further for this specific dataset. Random Forest consistently shows high accuracy and ROC-AUC across all tables, indicating its robustness and suitability for this dataset.

In summary, while preprocessing generally improves model performance by standardizing the dataset, feature selection does not always lead to a significant enhancement, possibly due to the Iris dataset's inherent simplicity. The choice of classifier also plays a crucial role, with ensemble methods like Random Forest and XGBoost consistently performing well across different preprocessing and feature selection scenarios.

G. Comparison with other studies

The study performance of the ML models shows a distinct improvement in specific areas when directly compared with similar metrics available from previous studies.

Table 5. Comparison with Previous Studies

Study	*Model*	*Accuracy*	*Precision*	*Recall*	*F1-Score*	*ROC AUC*	*Specificity*
This Study	**DT**	**0.971**	**0.971**	**0.971**	**0.971**	**0.977**	**0.973**
	RF	**0.983**	**0.983**	**0.983**	**0.983**	**0.999**	**0.984**
	SVM	**0.992**	**0.992**	**0.992**	**0.992**	**1.0**	**0.992**
	KNN	**0.988**	**0.988**	**0.988**	**0.988**	**0.996**	**0.989**
	XGB	**0.979**	**0.979**	**0.979**	**0.979**	**0.998**	**0.981**
[27]	**KNN**	**0.977**	**-**	**-**	**-**	**-**	**-**
	SVM	**0.955**	**-**	**-**	**-**	**-**	**-**
	LR	**0.955**	**-**	**-**	**-**	**-**	**-**
	DT	**0.933**	**-**	**-**	**-**	**-**	**-**
	RF	**0.977**	**-**	**-**	**-**	**-**	**-**
	MLP	**1.0**	**-**	**-**	**-**	**-**	**-**
[28]	**KNN**	**0.966**	**-**	**-**	**-**	**-**	**-**
	LINEAR SVM	**0.970**	**-**	**-**	**-**	**-**	**-**
	NL SVM	**0.991**	**-**	**-**	**-**	**-**	**-**
	LR	**0.966**	**-**	**-**	**-**	**-**	**-**
	RF	**0.970**	**-**	**-**	**-**	**-**	**-**
	K-MEANS 3	**0.893**	**-**	**-**	**-**	**-**	**-**
	K-MEANS 9	**0.966**	**-**	**-**	**-**	**-**	**-**
[29]	**DT**	**0.980**	**-**	**-**	**-**	**-**	**-**
	RF	**0.993**	**-**	**-**	**-**	**-**	**-**
	KNN	**1.0**	**-**	**-**	**-**	**-**	**-**
[30]	**ANN**	**0.986**	**0.980**	**0.963**	**-**	**-**	**-**
	SVM	**0.967**	**0.938**	**0.933**	**-**	**-**	**-**
[32]	**LR**	**0.980**	**-**	**-**	**-**	**-**	**-**
	KNN	**1.0**	**-**	**-**	**-**	**-**	**-**
	DT	**0.970**	**-**	**-**	**-**	**-**	**-**
	RF	**0.950**	**-**	**-**	**-**	**-**	**-**
[33]	**KNN N=5**	**0.966**	**-**	**-**	**-**	**-**	**-**
	KNN N=1	**1.0**	**-**	**-**	**-**	**-**	**-**
	LR	**0.960**	**-**	**-**	**-**	**-**	**-**
	LR (TRAIN AND SPLIT)	**0.950**	**-**	**-**	**-**	**-**	**-**
	KNN (TRAIN AND SPLIT N=5)	**0.950**	**-**	**-**	**-**	**-**	**-**
	KNN (TRAIN AND SPLIT N=1)	**0.966**	**-**	**-**	**-**	**-**	**-**
[34]	**NN**	**0.966**	**-**	**-**	**-**	**-**	**-**
	LR	**0.966**	**-**	**-**	**-**	**-**	**-**
	SVM	**0.980**	**-**	**-**	**-**	**-**	**-**
	KNN	**0.966**	**-**	**-**	**-**	**-**	**-**

H. Conclusion

The research illustrates the remarkable abilities of machine learning algorithms, especially XGBoost, in the accurate classification of different conditions of iris flowers using data on iris species. Utilizing a dataset comprising 1201 individual records across three categories - Iris Setosa, Iris Versicolor, and Iris Virginica, the study provided a solid foundation for assessing each model's performance. The XGBoost model excelled, surpassing others in crucial performance metrics like accuracy, precision, recall, and F1 score. This indicates its exceptional proficiency in accurately classifying the three iris flower species types. These outcomes underscore the transformative potential of sophisticated analytical methods in the field of iris flower species analysis, paving new paths for improved plant care and better monitoring outcomes. This study effectively demonstrates how machine learning models can adeptly navigate the complex and detailed aspects of plant diagnosis, particularly in the study of iris flower species.

References

[1] N. A. Kako and A. M. Abdulazeez, "Peripapillary Atrophy Segmentation and Classification Methodologies for Glaucoma Image Detection: A Review," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 18, no. 11, pp. 1140–1159, Nov. 2022, doi: 10.2174/1573405618666220308112732.

[2] F. I. Alarsan and M. Younes, "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms," *J Big Data*, vol. 6, no. 1, p. 81, Dec. 2019, doi: 10.1186/s40537-019-0244-x.

[3] A. O. Mohammed, H. I. Hussein, R. J. Mstafa, and A. M. Abdulazeez, "A blind and robust color image watermarking scheme based on DCT and DWT domains," *Multimed Tools Appl*, vol. 82, no. 21, pp. 32855–32881, Sep. 2023, doi: 10.1007/s11042-023-14797-0.

[4] R. Rajab Asaad and R. Masoud Abdulhakim, "The Concept of Data Mining and Knowledge Extraction Techniques," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 17–20, Mar. 2021, doi: 10.48161/qaj.v1n2a43.

[5] R. D. Ravipati and M. Abualkibash, "Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets - A Review Paper," *SSRN Electronic Journal*, 2019, doi: 10.2139/ssrn.3428211.

[6] R. Rajab Asaad, "Review on Deep Learning and Neural Network Implementation for Emotions Recognition," *Qubahan Academic Journal*, vol. 1, no. 1, pp. 1–4, Feb. 2021, doi: 10.48161/qaj.v1n1a25.

[7] H. T. Sadeeq and A. M. Abdulazeez, "Giant Trevally Optimizer (GTO): A Novel Metaheuristic Algorithm for Global Optimization and Challenging Engineering Problems," *IEEE Access*, vol. 10, pp. 121615–121640, 2022, doi: 10.1109/ACCESS.2022.3223388.

[8] M. Binkhonain and L. Zhao, "A review of machine learning algorithms for identification and classification of non-functional requirements," *Expert Systems with Applications: X*, vol. 1, p. 100001, Apr. 2019, doi: 10.1016/j.eswax.2019.100001.

[9] D. Mustafa Abdullah, A. Mohsin Abdulazeez, and A. Bibo Sallow, "Lung cancer Prediction and Classification based on Correlation Selection method Using Machine Learning Techniques," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 141–149, May 2021, doi: 10.48161/qaj.v1n2a58.

[10] M. N. Al-Mhiqani *et al.*, "A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations," *Applied Sciences*, vol. 10, no. 15, p. 5208, Jul. 2020, doi: 10.3390/app10155208.

[11] A. A. H. Alkurdi, "Enhancing Heart Disease Diagnosis Using Machine Learning Classifiers," *Fusion: Practice and Applications*, vol. 13, no. 1, pp. 08–18, 2023, doi: 10.54216/FPA.130101.

[12] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," 2020, pp. 99–111. doi: 10.1007/978-981-13-7403-6_11.

[13] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "Automatic Facial Aesthetic Prediction Based on Deep Learning with Loss Ensembles," *Applied Sciences (Switzerland)*, vol. 13, no. 17, Sep. 2023, doi: 10.3390/app13179728.

[14] T. Tiay, P. Benyaphaichit, and P. Riyamongkol, "Flower recognition system based on image processing," in *2014 Third ICT International Student Project Conference (ICT-ISPC)*, IEEE, Mar. 2014, pp. 99–102. doi: 10.1109/ICT-ISPC.2014.6923227.

[15] F. Didi, H. Pallathadka, S. Abdullaev, R. R. Asaad, S. Askar, and N. H. Haroon, "Probing the impact of process variables in laser-welded aluminum alloys: A machine learning study," *Mater Today Commun*, vol. 38, p. 107660, Mar. 2024, doi: 10.1016/j.mtcomm.2023.107660.

[16] M. Cıbuk, U. Budak, Y. Guo, M. Cevdet Ince, and A. Sengur, "Efficient deep features selections and classification for flower species recognition," *Measurement*, vol. 137, pp. 7–13, Apr. 2019, doi: 10.1016/j.measurement.2019.01.041.

[17] S. Talukdar *et al.*, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sens (Basel)*, vol. 12, no. 7, p. 1135, Apr. 2020, doi: 10.3390/rs12071135.

[18] R. Aruna *et al.*, "Coalescing novel QoS routing with fault tolerance for improving QoS parameters in wireless Ad-Hoc network using craft protocol," *Wireless Networks*, Oct. 2023, doi: 10.1007/s11276-023-03515-1.

[19] S. Bondre and U. Yadav, "Automated Flower Species Identification by Using Deep Convolution Neural Network," 2022, pp. 1–10. doi: 10.1007/978-981-16-6624-7_1.

[20] M. Toğaçar, B. Ergen, and Z. Cömert, "Classification of flower species by using features extracted from the intersection of feature selection methods in convolutional neural network models," *Measurement*, vol. 158, p. 107703, Jul. 2020, doi: 10.1016/j.measurement.2020.107703.

[21] I. Gogul and V. S. Kumar, "Flower species recognition system using convolution neural networks and transfer learning," in *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, IEEE, Mar. 2017, pp. 1–6. doi: 10.1109/ICSCN.2017.8085675.

[22] B. R. Mete and T. Ensari, "Flower Classification with Deep CNN and Machine Learning Algorithms," in *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, Oct. 2019, pp. 1–5. doi: 10.1109/ISMSIT.2019.8932908.

[23] N. Asaad Zebari, R. B. Marqas, M. Shamal Salih, and A. A. H. Alkurdi, "Enhancing Brain Tumor Classification with Data Augmentation and

DenseNet121," *Academic Journal of Nawroz University*, vol. 12, no. 4, pp. 323–334, Oct. 2023, doi: 10.25007/ajnu.v12n4a1985.

[24] M. M. Jaber *et al.*, "Future smart grids creation and dimensionality reduction with signal handling on smart grid using targeted projection," *Sustainable Computing: Informatics and Systems*, vol. 39, p. 100897, Sep. 2023, doi: 10.1016/j.suscom.2023.100897.

[25] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIsci Remote Sens*, vol. 57, no. 1, pp. 1–20, Jan. 2020, doi: 10.1080/15481603.2019.1650447.

[26] E. Poornima *et al.*, "Fog robotics-based intelligence transportation system using line-of-sight intelligent transportation," *Multimed Tools Appl*, Mar. 2023, doi: 10.1007/s11042-023-15086-6.

[27] Y. Zhou, "Study for Iris Classification Based on Multiple Machine Learning Models," 2022.

[28] S. Akter, S. A. Mithy, S. Hossain, S. Akter, U. Honey, and S. B. Sogir, "Classification of Iris Flower Dataset using Different Algorithms," *International Journal of Scientific Research in _____ Research Paper. Mathematical and Statistical Sciences*, vol. 9, no. 6, pp. 1–10, 2022, doi: 10.26438/ijsrmss/v9i6.110.

[29] B. Taha Chicho, A. Mohsin Abdulazeez, D. Qader Zeebaree, and D. Assad Zebari, "Machine Learning Classifiers Based Classification For IRIS Recognition," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 106–118, May 2021, doi: 10.48161/qaj.v1n2a48.

[30] A. Yehya Hussien, "Flower Species Recognition Using Machine Learning Classifiers," *Academic Journal of Nawroz University*, vol. 11, no. 4, pp. 469–475, Dec. 2022, doi: 10.25007/ajnu.v11n4a1636.

[31] M. Swain, "An Approach for IRIS Plant Classification Using Neural Network," *International Journal on Soft Computing*, vol. 3, no. 1, pp. 79–89, Feb. 2012, doi: 10.5121/ijsc.2012.3107.

[32] C. F. Aziz and B. J. Awrahman, "Prediction Model based on Iris Dataset Via Some Machine Learning Algorithms," *Journal of Kufa for Mathematics and Computer*, vol. 10, no. 2, pp. 64–69, Aug. 2023, doi: 10.31642/jokmc/2018/100210.

[33] T. Srinivasarao *et al.*, "Iris Flower Classification Using Machine Learning," 2021.

[34] A. Shukla, A. Agarwal, M. Lucknow, I. H. Pant, and P. Mishra, "Flower Classification using Supervised Learning." [Online]. Available: www.ijert.org

[35] S. Vatshayan, "Performance Evaluation of Supervised Learning for Iris Flower Species," 2019. [Online]. Available: https://www.researchgate.net/publication/332332854

[36] D. Rana, S. Prabha Jena, and S. Kumar Pradhan, "Performance Comparison of PCA and LDA with Linear Regression and Random Forest for IRIS Flower Classification PJAEE, 17 (9) (2020) Performance Comparison of PCA and LDA with Linear Regression and Random Forest for IRIS Flower Classification Performance Comparison of PCA and LDA with Linear Regression and Random Forest for IRIS Flower Classification--Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(9)."

[37] T. Gupta, P. Arora, R. Rani, G. Jaiswal, P. Bansal, and A. Dev, "Classification of Flower Dataset using Machine Learning Models," in *2022 4th International*

Conference on Artificial Intelligence and Speech Technology (AIST), IEEE, Dec. 2022, pp. 1–6. doi: 10.1109/AIST55798.2022.10065178.

[38] P. Agarwal, Gurjot, K. K. Sharma, Reshma, and A. Bindle, "Optimised Ensemble Machine Learning Model for IRIS Classification," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, Apr. 2022, pp. 1188–1193. doi: 10.1109/ICACITE53722.2022.9823519.

[39] Y. Pachipala, H. C. Maddipati, P. S. Udimudi, V. Thota, V. N. Sree, and L. R. Burra, "Iris Flower Classification by using Random Forest in AWS," in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, May 2022, pp. 1772–1777. doi: 10.1109/ICICCS53718.2022.9788415.

[40] L. Pawar, J. Singh, R. Bajaj, G. Singh, and S. Rana, "Optimized Ensembled Machine Learning Model for IRIS Plant Classification," in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, Apr. 2022, pp. 1442–1446. doi: 10.1109/ICOEI53556.2022.9776724.

[41] J. P. Pinto, S. Kelur, and J. Shetty, "Iris Flower Species Identification Using Machine Learning Approach," in *2018 4th International Conference for Convergence in Technology (I2CT)*, IEEE, Oct. 2018, pp. 1–4. doi: 10.1109/I2CT42659.2018.9057891.

[42] R. Ferinia *et al.*, "Factors determining customers desire to analyse supply chain management in intelligent IoT," *J Comb Optim*, vol. 45, no. 2, p. 72, Mar. 2023, doi: 10.1007/s10878-023-01007-8.

[43] H. T. Sadeeq and A. M. Abdulazeez, "Improved Northern Goshawk Optimization Algorithm for Global Optimization," in *ICOASE 2022 - 4th International Conference on Advanced Science and Engineering*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 89–94. doi: 10.1109/ICOASE56293.2022.10075576.

[44] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "An Ensemble DCNNs-Based Regression Model for Automatic Facial Beauty Prediction and Analyzation," *Traitement du Signal*, vol. 40, no. 1, pp. 55–63, Feb. 2023, doi: 10.18280/ts.400105.

[45] H. T. Sadeeq and A. M. Abdulazeez, "Car side impact design optimization problem using giant trevally optimizer," *Structures*, vol. 55, pp. 39–45, Sep. 2023, doi: 10.1016/j.istruc.2023.06.016.

[46] J. N. Saeed, A. M. Abdulazeez, and D. A. Ibrahim, "FIAC-Net: Facial Image Attractiveness Classification Based on Light Deep Convolutional Neural Network," in *2022 2nd International Conference on Computer Science, Engineering and Applications, ICCSEA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICCSEA54677.2022.9936582.

[47] J.-S. Lee, "AUC4.5: AUC-Based C4.5 Decision Tree Algorithm for Imbalanced Data Classification," *IEEE Access*, vol. 7, pp. 106034–106042, 2019, doi: 10.1109/ACCESS.2019.2931865.

[48] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling," *J Pet Sci Eng*, vol. 174, pp. 776–789, Mar. 2019, doi: 10.1016/j.petrol.2018.11.067.

[49] R. Ghezelbash, A. Maghsoudi, M. Shamekhi, B. Pradhan, and M. Daviran, "Genetic algorithm to optimize the SVM and K-means algorithms for mapping of

mineral prospectivity," *Neural Comput Appl*, vol. 35, no. 1, pp. 719–733, Jan. 2023, doi: 10.1007/s00521-022-07766-5.

[50] M. Mohammadi *et al.*, "A comprehensive survey and taxonomy of the SVM-based intrusion detection systems," *Journal of Network and Computer Applications*, vol. 178, p. 102983, Mar. 2021, doi: 10.1016/j.jnca.2021.102983.

[51] H. T. Sadeeq and A. M. Abdulazeez, "Metaheuristics: A Review of Algorithms," *International journal of online and biomedical engineering*, vol. 19, no. 9. International Association of Online Engineering, pp. 142–164, 2023. doi: 10.3991/ijoe.v19i09.39683.

[52] C. Ma, X. Du, and L. Cao, "Improved KNN Algorithm for Fine-Grained Classification of Encrypted Network Flow," *Electronics (Basel)*, vol. 9, no. 2, p. 324, Feb. 2020, doi: 10.3390/electronics9020324.

[53] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, IEEE, May 2019, pp. 1255–1260. doi: 10.1109/ICCS45141.2019.9065747.

[54] X. Zhu, J. Chu, K. Wang, S. Wu, W. Yan, and K. Chiam, "Prediction of rockhead using a hybrid N-XGBoost machine learning framework," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 13, no. 6, pp. 1231–1245, Dec. 2021, doi: 10.1016/j.jrmge.2021.06.012.

[55] P. Carmona, A. Dwekat, and Z. Mardawi, "No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure," *Res Int Bus Finance*, vol. 61, p. 101649, Oct. 2022, doi: 10.1016/j.ribaf.2022.101649.

[56] Y.-A. Hsieh and Y. J. Tsai, "Machine Learning for Crack Detection: Review and Model Performance Comparison," *Journal of Computing in Civil Engineering*, vol. 34, no. 5, Sep. 2020, doi: 10.1061/(ASCE)CP.1943-5487.0000918.

[57] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics (Basel)*, vol. 8, no. 8, p. 832, Jul. 2019, doi: 10.3390/electronics8080832.

[58] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-Label Confusion Matrix," *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.