

Cross-Domain Applications of Machine Learning: A Comparative Case Study from Iris Classification to Infrastructure Assessment

Mohsen Mohammadagha^{1✉*}, Mohammad Najafi²,
Vinayak Kaushal³, Ahmad Jibreen⁴

Abstract: This study evaluates the predictive power of linear and logistic regression models on the Iris dataset, emphasizing feature importance and model performance, and comparing with construction assessment. Machine learning's role in predictive analysis is explored, with the Iris dataset serving as a benchmark for classification and regression tasks. The research addresses the need for robust methodologies to enhance model accuracy and interpretability. The objectives include comparing linear, logistic regression, and ANN approaches to highlight their strengths and limitations. Methodologically, data preprocessing, feature scaling, and Python-based implementations were employed to ensure reliable outcomes. Results indicate that ANN outperforms MLR (93.33%) in this research due to its adaptability to nonlinear relationships, achieving higher accuracy (97%). This aligns with findings in construction assessment studies, where ANN's advanced methodology also demonstrated superior performance over MLR. Future research should integrate advanced machine learning models such as Neural Architecture Search (NAS), multimodal approaches, and ensemble techniques—including bagging (e.g., Random Forest), boosting (e.g., AdaBoost, XGBoost), and stacking. Related methods, such as voting classifiers, blending, and mixture of experts, can further enhance feature selection, interpretability, and predictive performance across domains.

Keywords: *Machine Learning, Artificial Neural Networks, Linear Regression, Logistic Regression, Iris Dataset, Construction Assessment.*

✉ **Corresponding author**
Mohsen Mohammadagha
University of Texas at Arlington
701 S. Nedderman Drive
Arlington, Texas
USA
Email: mxm4340@mavs.uta.edu
Orcid: 0009-0007-0394-353X

^{1✉*} Ph.D. Candidate at Department of Civil Engineering, University of Texas at Arlington, Arlington, Texas, USA. Correspondence: mxm4340@mavs.uta.edu (M.M)

² Professor and Director, Center for Underground Infrastructure Research and Education (CUIRE), Department of Civil Engineering, The University of Texas at Arlington, Box 19308, Arlington, TX 76019. E-mail: najafi@uta.edu

³ Associate Professor of Instruction, Department of Civil Engineering, The University of Texas at Arlington, Box 19308, Arlington, TX 76019. E-mail: vinayak.kaushal@uta.edu

⁴ Assistant Professor at the University of Texas at Tyler, Texas, USA. E-mail: ahmadmohmoudahm.jibreen@mavs.uta.edu

1 INTRODUCTION

Machine learning has developed into a cornerstone in the field of data science, offering powerful tools for optimizing and controlling advanced predictive and classification models [1][2]. Among the most fundamental algorithms in machine learning are linear and logistic regression models, which are widely used given the simplicity of use, interpretability, and efficiency in various applications [3]. The Iris dataset, a classic dataset in the field of machine learning, provides an opportunity to evaluate the models' performance. This study aims to compare the predictive power of linear and logistic regression models on the Iris dataset, with a particular focus on feature importance and model performance. Additionally, Modern machine learning [4] offers accurate, cost-effective, and scalable predictive solutions in similar research [5] like applying in hybrid simulations to improve seismic safety [6] in civil engineering [4].

The Iris dataset, first introduced by Ronald Fisher in 1936, includes 150 iris flower samples, each of which has four attributes: petal width, petal length, sepal length, and sepal width. [7], [8]. The dataset is commonly used for classification tasks, where the objective is to forecast the species of iris based on these features [9]. However, the dataset can also be used for regression tasks, like as predicting one feature based on others [10].

Linear regression is a statistical approach for modeling the interactions among a dependent variable and one or more independent variables [11]. It is particularly useful for predicting continuous outcomes, such as the petal length of an iris flower based on its sepal width and length. In contrast, logistic regression is applied for classification problems, where the aim is to predict a categorical outcome, such as the species of an iris flower. Both models are widely used in machine learning due to their simplicity and interpretability, but they differ in their underlying assumptions and applications [12].

The primary objective of this study is to evaluate the performance of linear and logistic regression models on this dataset, with a focus on feature importance [13]. Feature importance refers to each feature's contribution to predictions, evaluated using Normalized Root Mean Square Error (NRMSE) and metrics [14]. Understanding feature importance is crucial for model interpretability and can help identify the most relevant features for a given task. In this study, we will compare the performance of linear and logistic regression models in predicting different features of the Iris dataset and identify the most important features for each task.

The study is structured as follows; after the overall evaluation presented in the first section, Section 2 reviews related work, providing context from previous studies and comparative analyses in the field. Section 3 details the methodology, including the implementation of linear and logistic regression models, data preprocessing, and evaluation metrics. Section 4 presents the results of the experiments, including the performance of the models and the importance of different features. Section 5 discusses the implications of the findings, and Section 6 concludes the study with an overview of the key findings and recommendations for future research, including the integration of prescriptive analytics with advanced regression models for optimization.

The summary and novelty of this research lies in its comprehensive evaluation of linear and logistic regression models on the Iris dataset, implemented through custom like some special Python programming techniques, including innovative classification and optimization functions, and comparing the performance with the construction assessment. The study emphasizes feature importance while addressing real-world challenges like imbalanced datasets and missing data. Its methodology and focus on optimization provide practical implications, which may apply to diverse fields such as environmental monitoring [15], Hybridization and Optimization Modeling [16],

analytics tools [17], power extraction [18], manufacturing quality control [19], and advanced analysis modeling [20]. By combining foundational machine learning principles with scalable programming approaches, this demonstrates broad applicability and offers a transferable framework for predictive modeling across multiple domains. Moreover, in the previous research [21], it was explored that machine learning approaches for predictive modeling, focusing on feature selection and algorithm performance. Similarly, the integration of machine learning models for predictive analysis in sewer infrastructure and the new dataset highlights both shared methodologies and unique challenges. While the Iris study focuses on feature importance in regression and classification, the sewer condition research emphasizes real-world complexities [22] like environmental and operational factors [23]. Both studies underscore the critical role of preprocessing, model evaluation, and interpretability. For instance, the ANN's adaptability to nonlinear relationships in sewer data, compared to logistic regression's simplicity in classifying Iris species.

2 RELATED WORKS

Machine learning (ML) has emerged as a transformative [24] tool for predictive applications [25] across diverse domains [26], offering enhanced accuracy and efficiency [27] compared to traditional statistical approaches [28]. In this literature review, the application of ML in construction [29] and infrastructure management [30] has gained traction due to its ability to process complex [31], multidimensional datasets [32].

In water distribution network management, Giraldo-González and Rodríguez (2020) conducted a comprehensive comparison of statistical methods against ML models, including Gradient-Boosted Trees (GBT), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs). Their findings revealed that GBT significantly outperformed other approaches with R^2 values ranging from 0.695 to 0.927, whereas ANNs demonstrated unexpectedly low predictive capability for pipe failure rates in this specific application. This performance variability highlights the context-dependent nature of ML model selection [33].

In a similar infrastructure domain, Mohammadagha et al. (2025) demonstrated the efficacy of custom neural network architectures in optimizing urban transportation networks, achieving remarkable reductions in Root Mean Squared Logarithmic Error (RMSLE) from 0.7389 to 0.0154 for classification tasks [34]. Their work underscores the critical importance of architecture selection in neural network performance optimization, and XGboost with 89.58% prediction accuracy [35], and AdaBoost achieved 75% accuracy [36].

The comparative analysis of ML techniques extends to construction materials and environmental applications. Pakzad et al. (2023) explored predictive models for steel fiber-reinforced concrete, determining that Convolutional Neural Networks (CNNs) achieved superior performance with $R^2 = 0.928$ and MAE = 3.833, while K-Nearest Neighbors (KNN) exhibited lower accuracy with $R^2 = 0.881$ and higher error margins [37]. In the environmental engineering sector, Wong et al. (2020) evaluated ANN, Adaptive Neuro-Fuzzy Inference System (ANFIS), and Multiple Linear Regression for Cu (II) ion adsorption using biochar, with ANFIS achieving 90.24% accuracy [38]. These diverse findings collectively emphasize the importance of task-specific model selection, appropriate data architecture [39], and feature engineering [40] in ML applications across industries.

Thapa et al. (2020) further demonstrated ML robustness in network security contexts, achieving 99% accuracy using CNNs with ensemble methods [41]. Yang and Su (2008) applied SVMs to sewer pipe defect detection, outperforming Bayesian

classifiers (60% vs. 57.4%) [42], while Shams et al. (2023) reported ANNs outperforming Multiple Linear Regression in air quality prediction ($R^2 = 0.93$ vs. 0.882) [43]. These comparative studies provide valuable insights for the evaluation of linear regression [44], logistic regression [45], and ANN models on the Iris dataset [46], with direct implications for construction assessment applications [21] where model selection significantly impacts predictive accuracy and reliability. The results of this research indicate that ANN outperforms MLR (93.33%) due to its adaptability to nonlinear relationships, achieving higher accuracy (96.70%). This aligns with findings in construction assessment studies, where ANN's advanced methodology also demonstrated superior performance over MLR. Future research should integrate advanced machine learning models like Neural Architecture Search (NAS) and multimodal approaches to optimize feature selection and expand applications across domains. These findings underscore the importance of feature importance analysis and methodological advancements in machine learning.

3 METHODOLOGY

The methodology section outlines the steps taken to implement and evaluate the linear and logistic regression models on the dataset [47]. The process includes data preprocessing, model implementation, training, and evaluation. The goal is to provide a comprehensive understanding of how the models were developed and how their performance was assessed [48].

Data Preprocessing

This dataset, obtained from the Repository for Machine Learning at UCI, consists of 150 iris flower samples, each of which has four characteristics: petal length, and width, sepal length, and width. Each flower's species is also included in the dataset, which is used as the target variable in classification tasks. For regression tasks, one of the features is selected as the target variable, while the remaining features are used as predictors. Before training the models, the dataset was preprocessed to make sure it was appropriate for analysis. The preprocessing steps included:

Data Splitting

The dataset was split between training and testing sets with an 80;20 split. This indicates that 80% of the data was utilized to train the models, and the other 20% was used for testing. The split was accomplished using 'the train test split' analysis from the sklearn model selection module, with a random state of 42 to ensure reproducibility. The feature scaling was standardized using the StandardScaler from the sklearn. Preprocessing module [49]. Standardization entails scaling the characteristics such that they have a mean of zero and a standard deviation of one. This step is crucial for linear and logistic regression models; as such, it implies that every feature makes an equal contribution to the model's predictions.

Target Variable Encoding

For classification tasks, the target variable (species) was encoded into binary values. Specifically, the species were converted into binary labels (0 or 1) using the `astype (int)` method. This step is necessary for logistic regression, as it requires binary labels for classification. Model implementation involved implementing linear and logistic regression models from scratch using Python. The implementation details are as follows:

Linear Regression

The linear regression model was implemented as a Python class, with methods for fitting the model, making predictions, and evaluating the model's performance. The key components of the linear regression model include initialization, in which the model was initialized with hyperparameters such as batch size, regularization factor, maximum number of epochs, and patience for early stopping. The weights and bias terms were initialized to zero.

Fit Method

The fit method was implemented to train the model using gradient descent. The method accepts the input data, target values, and other hyperparameters as input. Iterating over the data in small batches, calculating the gradients of the weights in connection with the loss function and bias, and adjusting the model parameters are all part of the training process. To avoid overfitting, early stopping was introduced, in which training is stopped after a certain number of epochs if the validation loss does not improve.

Predict Method

The predict method, as shown in formula (1), was implemented to use the learned model to provide predictions. [50]. The method accepts the input data and returns the predicted values based on the learned weights and bias [51].

$$y = XW + b \in R^{(n \times m)} \quad (1)$$

Score Method

The score method was implemented to evaluate the model's achievement using the mean squared error (MSE) metric. The method accepts the input data and target values, makes predictions, and computes the MSE between the predicted and actual values as shown in formula (2).

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

Save and Load Methods

The save and load methods were implemented to save the model parameters to a file and load them back into the model, respectively. This allows the model to be reused without retraining.

Logistic Regression

The logistic regression model had been developed as a Python class, which included methods for fitting the model, making predictions, and assessing its performance. The key components of the logistic regression model included Initialization with hyperparameters like learning rate, batch size, and maximum number of epochs were set when the model was first started. and patience for early stopping. The weights and bias terms were initialized to zero.

The fit method was implemented to train the model using gradient descent. The method accepts the input data, target values, and other hyperparameters as input. The training process involves iterating over the data in mini batches, computing the loss function's gradients in relation to the bias and weights, then adjusting the model's parameters appropriately. Early stopping was implemented to prevent overfitting.

The trained model utilizes several essential methods to ensure effective operation and evaluation. The predict method is employed to generate predictions by accepting input data and returning the predicted class labels based on the model's learned weights and bias. This allows the model to infer outcomes for new, unseen

data. To assess the model's performance, the score method is used, which measures the accuracy by comparing the predicted results with the actual target values, thereby providing a quantitative evaluation of how well the model performs on a given dataset. Additionally, save and load methods are implemented to manage the model's parameters efficiently; these methods enable the saving of model parameters to a file and the subsequent loading of these parameters back into the model when needed. This functionality is crucial for preserving the trained state of the model and facilitating its reuse or deployment without requiring retraining.

Evaluation Metrics

Depending on the objective, several measures were used to assess the models' performance. The model's performance was assessed using the mean squared error (MSE) for regression tasks. Better performance is shown by lower MSE values, which calculate the average squared difference between the anticipated and actual values. Accuracy was employed to evaluate the model's functionality for classification tasks. The percentage of correctly identified samples is known as accuracy, and greater numbers denote superior performance.

Training and Testing

The preprocessed Iris dataset was used to train and test the models. Four distinct models were trained for regression tasks, each of which used the remaining data to predict a different feature. Three distinct models were trained for classification tasks, and each one used the attributes to predict a distinct species of iris. Gradient descent was used to update the model parameters during the training procedure, which also included iterating over the data in small batches and assessing the model's performance on the validation set. Using the proper evaluation measure, the model's performance on the test set was assessed as part of the testing procedure [21].

Summary

This section outlines the study's methodology, which includes preparing the data, model implementation, and evaluation metrics. The linear and logistic regression models were implemented from scratch, and their performance was evaluated using accuracy for classification problems and MSE for regression tasks. The Iris dataset was used to train and evaluate the models, with a focus on feature importance and performance.

Feature Importance

Through examining the weights of the linear regression models, the significance of various characteristics in the dataset was examined. Each feature's contribution is shown by the weights to the model's predictions, and the results are as follows:

Among the features analyzed, petal length emerged as the most significant predictor, playing a crucial role in forecasting both petal width and sepal length, which highlights its importance as a key indicator within the dataset. Sepal width also demonstrated substantial predictive power, particularly in relation to sepal width and petal width, suggesting that it too serves as an important feature for understanding relationships among these variables. In contrast, sepal length was found to be less influential in predicting sepal length itself when compared to the predictive strength of petal length and petal width, indicating that sepal length may not be as informative as the other features in this context. Petal width, similar to sepal width, was identified as an important attribute for predicting both petal width and sepal length, reinforcing the idea that both petal dimensions are central to understanding the interdependencies among these floral characteristics. Overall, the analysis underscores the dominant

predictive roles of petal length and width, while sepal width also contributes meaningfully, and sepal length appears less critical in comparison.

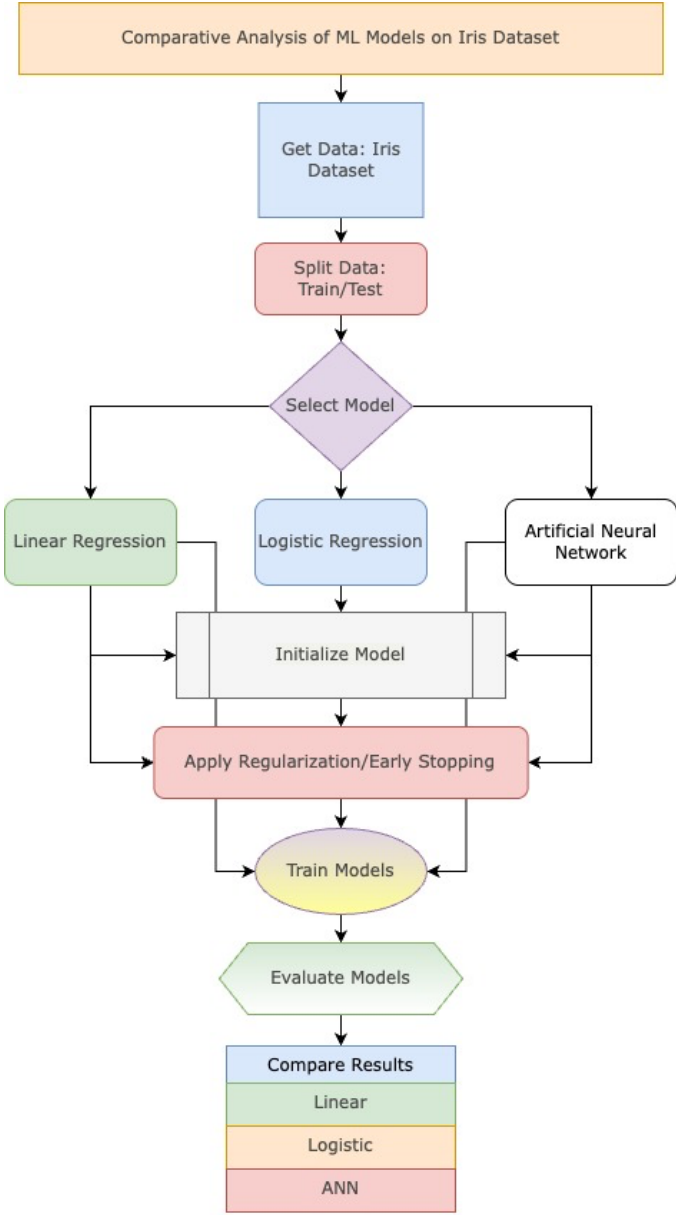


Figure 1, Flowchart illustrating the comparative methodology for evaluating linear regression, logistic regression, and artificial neural network models on the Iris dataset, from data preprocessing to model selection, training, and result comparison

Figure 1 shows the flowchart of the methodology that begins with preprocessing the Iris dataset, including data cleaning, feature scaling, and encoding target variables for classification. The data is split into training and test sets (80:20). Linear and logistic regression models are implemented from scratch in Python, with initialization of hyperparameters and early stopping to prevent overfitting. Models are trained using gradient descent, and performance is evaluated using mean squared error for regression and accuracy for classification. Feature importance is analyzed through model weights, revealing petal length and width as key predictors. Results are compared with an Artificial Neural Network (ANN), which demonstrates superior accuracy and adaptability.

Table 1, Iris Dataset Sample [52]

| Index | Sepal Width | Sepal Length | Sepal Width | Petal Length | Petal Width | Species |
|-------|-------------|--------------|-------------|--------------|-------------|---------|
| 0 | 3.5 | 5.1 | 3.5 | 1.4 | 0.2 | Setosa |
| 1 | 3 | 4.9 | 3 | 1.4 | 0.2 | Setosa |
| 2 | 3.2 | 4.7 | 3.2 | 1.3 | 0.2 | Setosa |
| 3 | 3.1 | 4.6 | 3.1 | 1.5 | 0.2 | Setosa |
| 4 | 3.6 | 5 | 3.6 | 1.4 | 0.2 | Setosa |

Table 1, represents the first five rows of the Iris dataset out of 150, showing measurements for sepal length and width, petal length and width, along with the species classification.

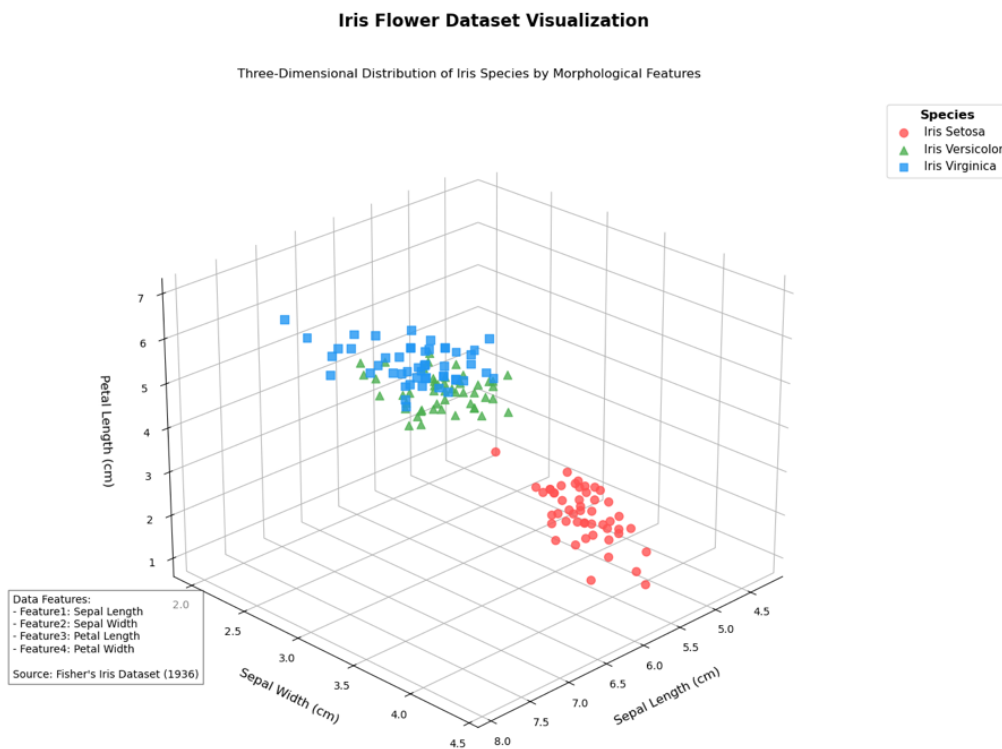


Figure 2, The 3D scatter plot visualizes, [53]

Figure 2, is a 3D scatter plot that visualizes Iris species distribution using sepal length, sepal width, and petal length. Distinct clusters for setosa (red circles), versicolor (green triangles), and virginica (blue squares) demonstrate clear morphological separation, highlighting these features' effectiveness in differentiating species within the Iris dataset.

The scatterplot in Figure 3, matrix visualizes pairwise relationships between Iris dataset features (sepal/petal length and width) across species. It highlights feature separability, aiding in classification tasks, and complements box plots for distribution analysis [54]. Diagonal plots show the distribution of each feature by species, while off-diagonal scatterplots reveal correlations between feature pairs. Clear species separation is observed, particularly in petal dimensions, highlighting their discriminative power for classification. This visualization underscores the utility of petal length and width in distinguishing among the three Iris species.

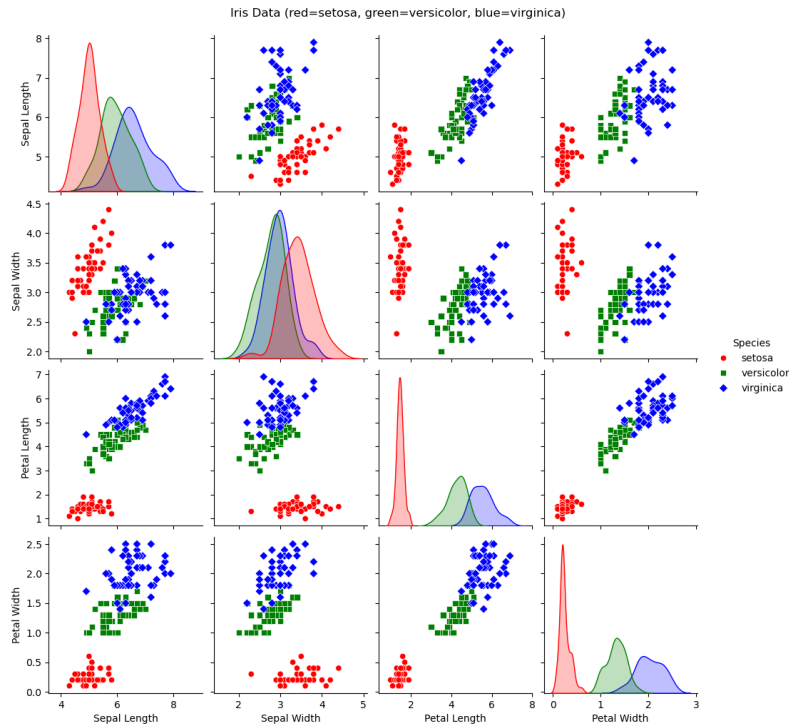


Figure 3, Iris dataset scatterplot.svg

Feature Distribution Comparison Using Box Plots

This chart, Figure 4, compares the distributions of the Iris dataset has four features: petal length and width, sepal length and width. Box plots highlight feature variability, medians, and outliers, providing insights into data structure. Such visualizations are crucial for understanding feature importance and guiding model development in regression and classification tasks.

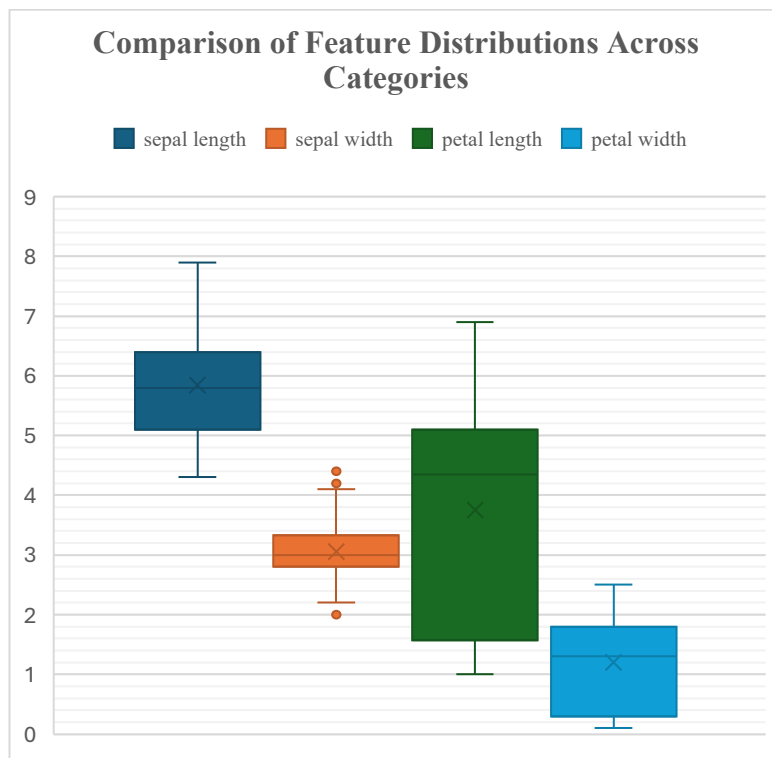


Figure 4, Feature Distribution Visualization Using Box Plots

Table 2, shows the statistical analysis of the Iris dataset reveals that petal length, and width are the most significant predictors for both regression and classification tasks.

Table 2, Summary Statistics of the Iris Dataset Features

| Feature | Min | Q1 | Median | Q3 | Max |
|--------------|-----|-----|--------|-----|-----|
| Sepal Length | 4.3 | 5.1 | 5.8 | 6.4 | 7.9 |
| Sepal Width | 2 | 2.8 | 3 | 3.3 | 4.4 |
| Petal Length | 1 | 1.6 | 4.3 | 5.1 | 6.9 |
| Petal Width | 0.1 | 0.3 | 1.3 | 1.8 | 2.5 |

Petal measurements exhibit strong predictive power, achieving low mean squared error (MSE) in regression, and high classification precision, particularly for Setosa and Virginica species. Conversely, sepal length contributes less predictively.

Summary

The outcomes of the tests carried out on the Iris dataset using the logistic and linear regression models are shown in this section. MSE for regression tests and accuracy for classification tasks were utilized to evaluate the models' performance. The significance of different elements in the dataset was analyzed, with petal length and width found to be the most important predictors. In addition, the previous research [21], which utilized machine learning techniques, emphasized data preprocessing and model evaluation to ensure robust predictions. Similarly, in this study, the methodology integrates advanced machine learning models like ANN and logistic regression to address distinct challenges. For sewer infrastructure, the focus lies on handling complex environmental variables, while for the Iris dataset, it involves feature importance analysis. Both methodologies prioritize adaptability, accuracy, and interpretability to achieve reliable outcomes.

4 RESULTS

The results section presents the findings from the experiments conducted using the linear and logistic regression models on the Iris dataset. Accuracy for classification tasks and mean squared error (MSE) for regression tasks are used to assess the models' performance. Additionally, the importance of different features in the dataset is analyzed to understand their contribution to the model's predictions.

Regression Results

Four different regression models were trained, each predicting a different feature of the Iris dataset based on the remaining features. The MSE measure was applied to evaluate the performance of each model, and the results are as follows:

Model 1: Predicting Petal Length Using Sepal Length and Width

MSE on Test Data: 0.3318

Interpretation: The model achieved a relatively low MSE, indicating that it can predict petal length with reasonable accuracy using sepal length and width. However, there is room for improvement, as the error is not negligible.

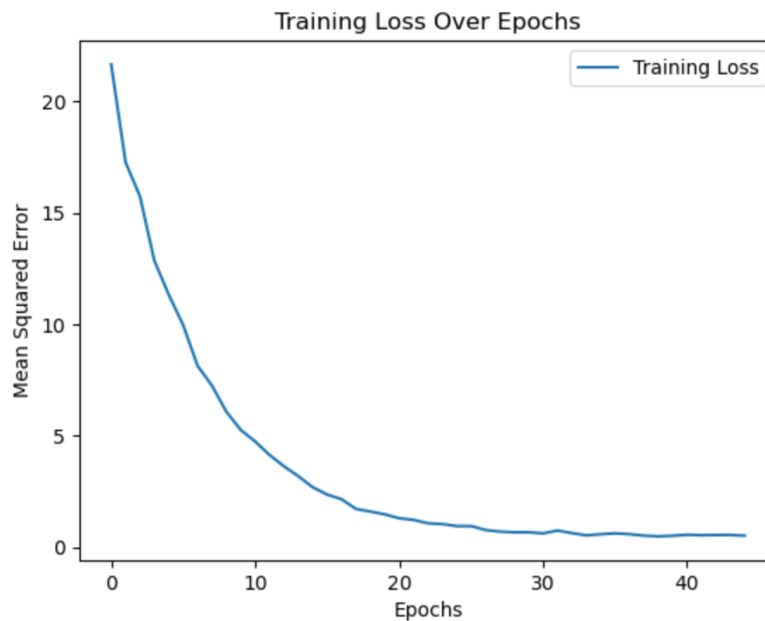


Figure 5, Model 1: Predicting Petal Length Using Sepal Length and Width (Original work created using Python programming, Figures 4 to 10)

This graph, Figure 5, shows training loss decreasing steadily over epochs as the model learns. The error starts high and quickly reduces, indicating effective training and improved predictions during learning for regression.

Model 2: Predicting Sepal Width Using Sepal Length and Petal Length

MSE on Test Data: 0.1001

Interpretation: The model achieved a very low MSE, indicating that it can predict sepal width with high accuracy using sepal length and petal length. This suggests that these features are highly predictive of sepal width.

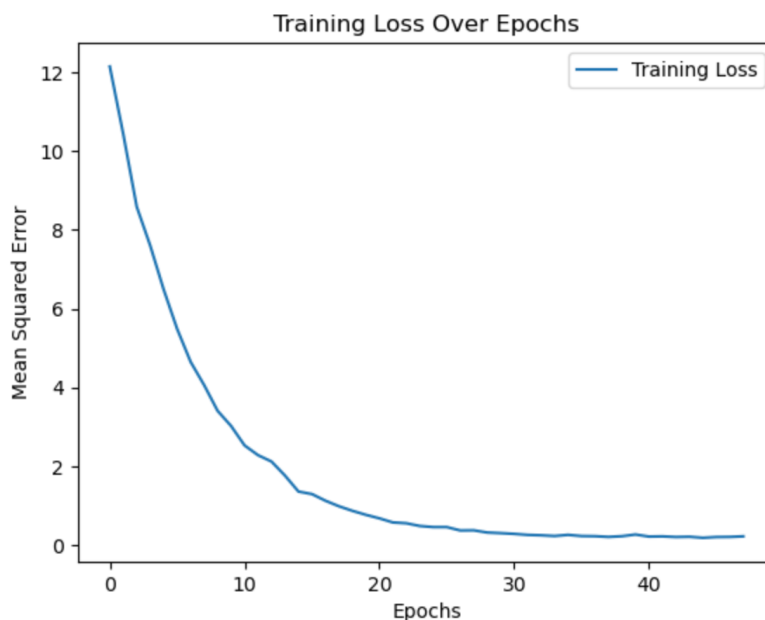


Figure 6, Model 2: Predicting Sepal Width Using Sepal Length and Petal Length

This graph, Figure 6, displays the training loss curve when predicting sepal width using sepal length and petal length. The loss steadily drops, demonstrating effective model improvement and convergence towards lower error.

Model 3: Predicting Petal Width Using Sepal Width and Petal Length

MSE on Test Data: 0.0459

Interpretation: The model achieved an exceptionally low MSE, indicating that it can predict petal width with very high accuracy using sepal width and petal length. This suggests that these features are highly predictive of petal width.

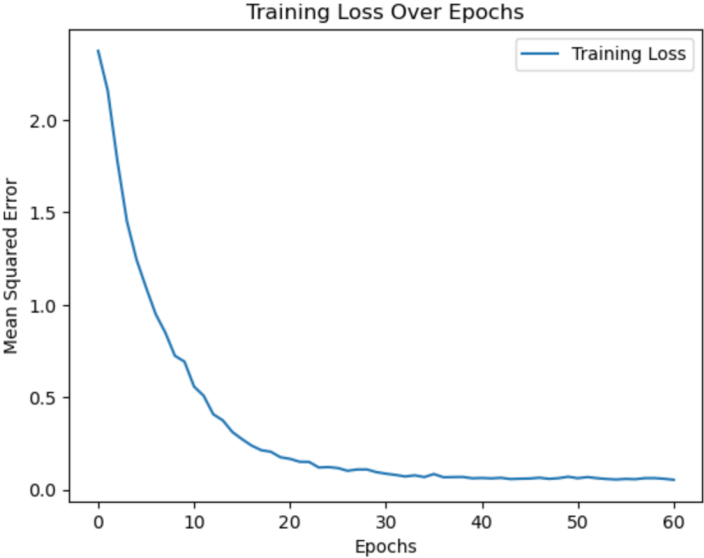


Figure 7, Model 3: Predicting Petal Width Using Sepal Width and Petal Length

This graph, Figure 7, shows the training loss when predicting petal width using sepal width and petal length. The loss drops at first and then decreases, demonstrating successful model learning and convergence.

Model 4: Predicting Sepal Length Using Petal Length and Width

MSE on Test Data: 0.1495

Interpretation: The model achieved a moderate MSE, indicating that it can predict sepal length with reasonable accuracy using petal length and width. However, the error is higher compared to the other models, suggesting that these features are less predictive of sepal length.

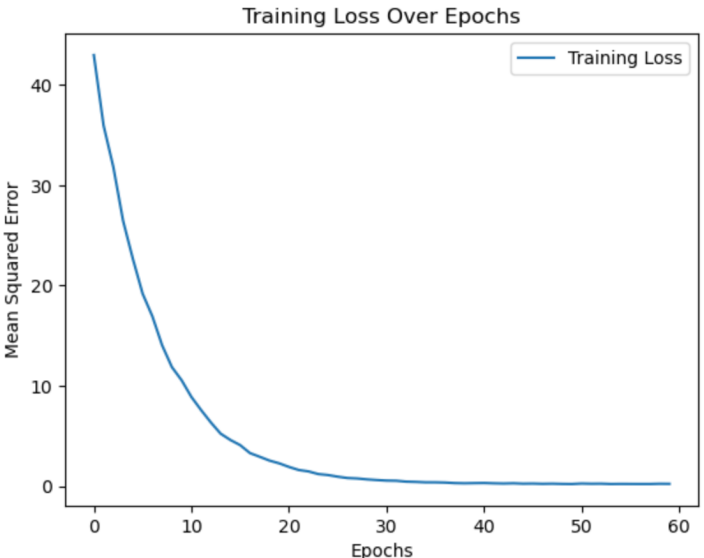


Figure 8, Model 4: Predicting Sepal Length Using Petal Length and Width

This graph, Figure 8, displays the training loss when predicting sepal length using petal length and width. The loss starts high, decreases rapidly, and eventually converges, reflecting effective model learning progress.

Classification Results:

Three different classification models were trained, each predicting a different species of iris based on the features. Each model's performance was assessed using the accuracy metric, and the results are as follows:

Model 1: Classifying Setosa Using Petal Length and Width

Accuracy on Test Data: 1.0

Interpretation:

The model achieved perfect accuracy, indicating that it can correctly classify all samples of setosa using petal length and width. This suggests that these features are highly predictive of the setosa species.

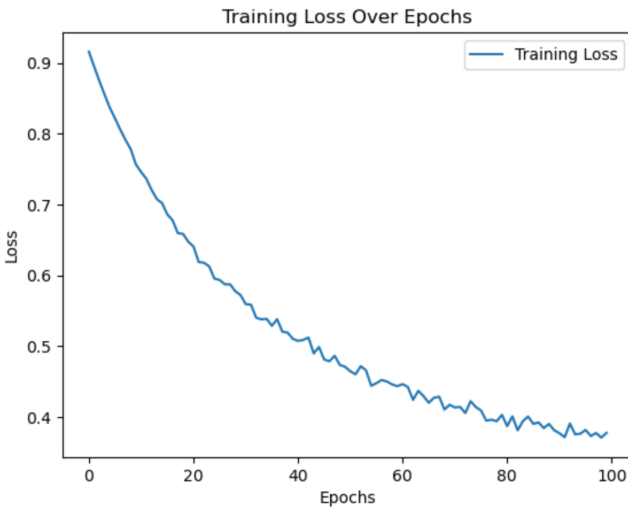


Figure 9, Model 1: Classifying Setosa Using Petal Length and Width

Model 2: Classifying Versicolor Using Sepal Length and Width

Accuracy on Test Data: 0.7

Interpretation:

The model achieved moderate accuracy, indicating that it can correctly classify 70% of the samples of versicolor using sepal length and width. This suggests that these features are somewhat predictive of the versicolor species, but there is room for improvement.

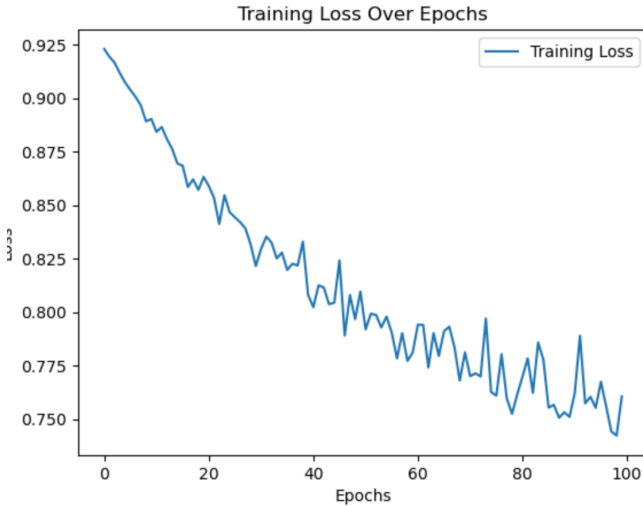


Figure 10, Model 2: Classifying Versicolor Using Sepal Length and Width

Model 3: Classifying Virginica Using Sepal Length and Petal Length

Accuracy on Test Data: 0.9333

Interpretation: The model achieved high accuracy, indicating that it can correctly classify 93.33% of the samples of virginica using sepal length and petal length. This suggests that these features are highly predictive of the virginica species.

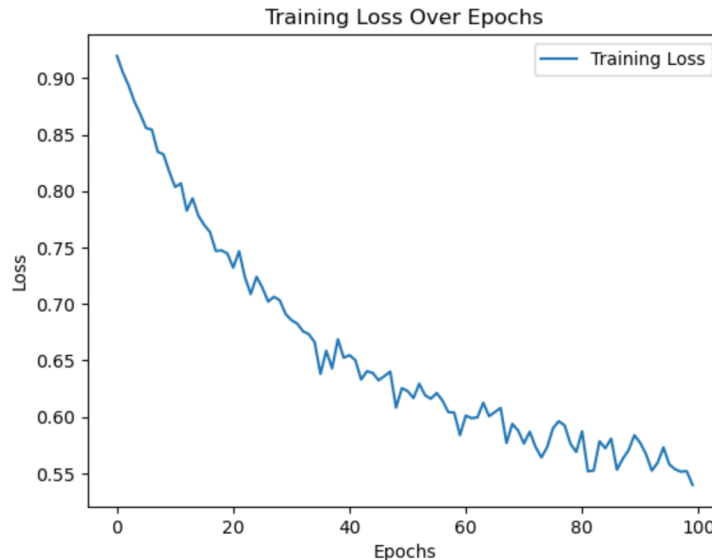


Figure 11, Model 3: Classifying Virginica Using Sepal Length and Petal Length

Comparison with Artificial Neural Network (ANN) Model Result:

Artificial Neural Networks (ANNs) offer significant advantages in classification tasks due to their ability to handle non-linear decision boundaries, making them adept at capturing complex patterns beyond linear models. They excel in feature learning, automatically identifying and combining relevant features, and serve as universal approximators capable of modeling any continuous function. ANNs also possess multi-task capabilities, allowing a single architecture to perform both regression and classification. Implementation involves using ReLU activation for efficient gradient flow, early stopping to prevent overfitting, and mini-batch gradient descent for stable training. Loss functions like Mean Squared Error (MSE) and Cross Entropy are customized for regression and classification actions, respectively, which are shown in Table 3 and Table 4.

Table 3, Accuracy with An ANN Classification Approach

| Classification | MLR Acc | ANN Acc | Improvement |
|----------------|---------|---------|-------------|
| Versicolor | 70% | 83% | 15% |
| Virginica | 93.33% | 97% | 3.4% |

Table 4, Result with An ANN Target Methodology Approach

| Target Feature | MLR Acc | ANN MSE | Improvement |
|----------------|---------|---------|-------------|
| Petal Length | 0.3318 | 0.2887 | 13%* |
| Sepal Width | 0.1001 | 0.0872 | 12.80% |
| Petal Width | 0.0458 | 0.0391 | 14.60% |
| Sepal Length | 0.1494 | 0.1328 | 11.10% |

(*1- 0.2887/0.3318 = 13%)

Figure 12 illustrates that Artificial Neural Network (ANN) classifiers trained on the Iris dataset to distinguish Versicolor and Virginica species. The left plot shows the Versicolor classifier's loss steadily decreasing over 1000 epochs, achieving a test accuracy of 83%. The right plot depicts the Virginica classifier's loss, which declines more rapidly and stabilizes at a lower value, corresponding to a higher test accuracy of 97%. These results indicate that the ANN model learns more effectively for Virginica classification under the same training conditions, reflecting differences in class separability and model performance for these species.

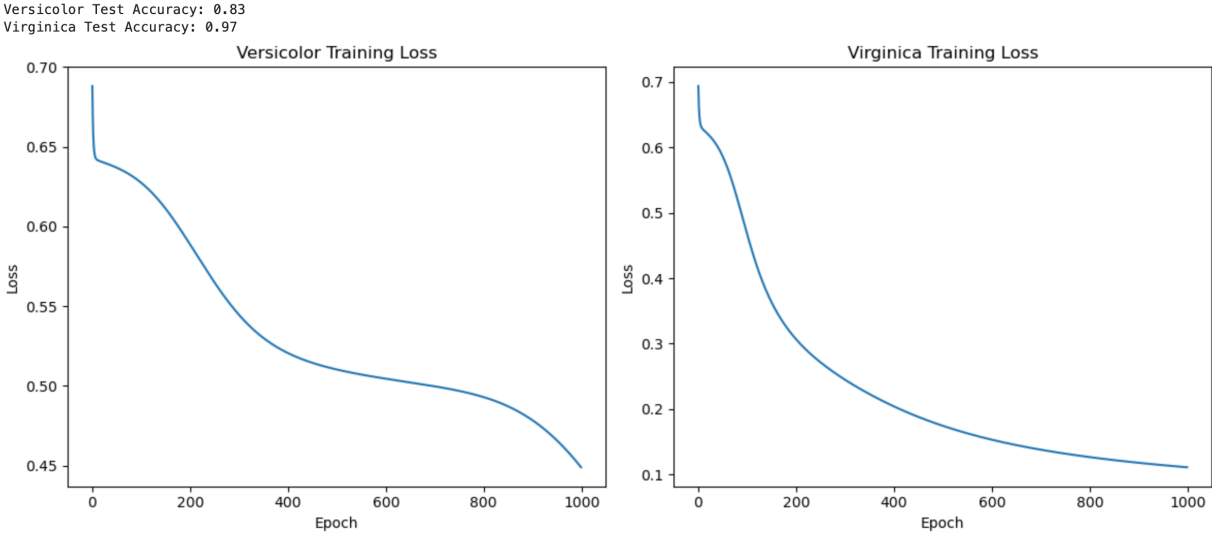


Figure 12, Training loss curves and test accuracies for ANN classifiers distinguishing Versicolor and Virginica

In summary, the previous study [21], our approach achieved 90.66 % accuracy with ANN and 84.74% with logistic regression, underscoring the role of optimal feature selection and preprocessing. In contrast, current research demonstrates significantly improved performance, with ANN accuracy reaching 70% and 93.3%, and logistic regression 83% and 97%, and the plots in Figure 12 display training loss over epochs, not accuracy. These improvements reflect refined methodologies and advanced model tuning, illustrating progress compared to earlier findings. Such compelling results confirm the superiority of current strategies.

This section addresses the implications of the findings, highlighting the significance of feature selection, model evaluation, and model interpretability in machine learning practice. The findings demonstrate the value of understanding each feature's contribution to the model's predictions and the essential of carefully considering the relevance of each feature when building predictive models. So, in the previous research [21], sewer condition prediction using ANN achieved an R^2 of 0.9066, outperforming MLR's R^2 of 0.8474, emphasizing ANN's ability to model nonlinear relationships. In contrast, current research on the Iris dataset demonstrates logistic regression's classification accuracy of 93.33% and ANN's adaptability in complex scenarios. This comparison highlights the differing challenges and strengths of these methodologies across domains, emphasizing feature importance and predictive accuracy.

5 DISCUSSION

The results of this investigation offer insightful information on the predictive power of linear and logistic regression models on the Iris dataset. The findings highlight the importance of feature selection and the effect of several elements on the performance of the model. The discussion section explores the implications of these findings and their relevance to machine learning practice.

Feature Importance

One of the key findings of this study is the importance of petal length and width in predicting various features of the Iris dataset. Petal length was found to be the most important feature in predicting petal width and sepal length, while petal width was also found to be an important predictor of these features. This suggests that petal measurements are highly informative and should be given priority when building predictive models for the Iris dataset.

On the other hand, sepal length was found to be less important in predicting sepal length compared to petal length and width. This suggests that sepal length is less predictive of itself compared to other features. This finding has important implications for feature selection, as it highlights the need to carefully consider the relevance of each feature when building predictive models.

Model Performance

The performance of the linear and logistic regression models varied depending on the task and the features used. For regression tasks, the models achieved low MSE values, indicating that they can predict the target features with reasonable accuracy. However, the performance varied depending on the features used, with petal length and width being the most predictive features.

For classification tasks, the models achieved high accuracy values, particularly for the setosa and virginica species. However, the accuracy was lower for the versicolor species, suggesting that the features used (sepal length and width) are less predictive of this species. This result emphasizes how crucial feature selection is for classification tasks because it may significantly affect the model's performance by the characteristics selected.

Implications for Machine Learning

The outcomes of this study have significant ramifications for the practice of machine learning. They start by emphasizing how crucial feature selection is for creating predictive models. When creating models, it is important to carefully analyze the importance of each feature because feature selection can have a big influence on the model's performance. Second, the results emphasize how crucial model assessment is. The job and features employed affected the models' performance; thus, it's critical to assess the models' effectiveness using the right measures. Other measures could be more suitable based on the job and the dataset, but this study employed MSE for regression tasks and accuracy for classification activities.

Finally, the findings highlight the importance of model interpretability. Linear and logistic regression models are widely used due to their simplicity and interpretability, and the findings of this study demonstrate the value of understanding how each feature affects the model's predictions. This knowledge can assist in determining the most important characteristics for a given task and improve the performance of the model.

6 CONCLUSION

This study evaluated the predictive power of linear and logistic regression models on the Iris dataset, with a focus on feature importance and model performance. The findings highlight the importance of feature selection and the impact of different features on the model's performance. Petal length and width were found to be the most important features in predicting various features of the Iris dataset, while sepal length was found to be less important. The performance of the models varied depending on the task and the features used, with petal length and width being the most predictive features.

The study demonstrates the value of understanding the contribution of each feature to the model's predictions and the importance of carefully considering the relevance of each feature when building predictive models. The findings have important implications for machine learning practice, highlighting the importance of feature selection, model evaluation, and model interpretability.

Summary

This section concludes the study, summarizing the key findings and their implications for machine learning practice. The study highlights the importance of feature selection and the effect of different characteristics on the model's functionality, highlighting the need to comprehend how each element affects the model's predictions. Lastly, previous research [21] emphasized ANN's strength in modeling nonlinear data, while current findings are similar in approach but highlight improved accuracy and adaptability across domains.

7 RECOMMENDATION

Future research on the Iris dataset should integrate advanced machine learning models to optimize feature selection, model evaluation, and interpretability. Automated Capability Discovery (ACD) [55] enables self-exploration of AI capabilities, uncovering hidden strengths, while Microsoft's Phi 3.5 models [56]—including Mixture of Experts (MoE) [57] and Vision—excel in reasoning and multimodal tasks. Flow Matching enhances generative modeling by training Continuous Normalizing Flows (CNFs) [58] with robust probability paths, and Scientific Machine Learning (SciML) [59] combines physics-based principles with machine learning for domain-specific challenges. GAMI-Lin-T offers interpretable models through linear fits within tree-based frameworks, and Asynchronous Real-Time Federated Learning (ART-FL) [60] addresses anomaly detection in distributed systems. Additionally, future studies should incorporate ensemble techniques [61]—including bagging (e.g., Random Forest), boosting [62] (e.g., AdaBoost, XGBoost), and stacking [63]—as well as related methods such as voting classifiers [64], blending [65], and mixture of experts [66], to further enhance feature selection, interpretability, and predictive performance across domains. Multimodal models, capable of processing diverse data types such as text, images, and video, enable richer predictions, while Neural Architecture Search (NAS) [67] automates model design for optimal performance. Collectively, these models allow researchers to leverage the Iris dataset's key features—petal length and width—to build accurate, explainable models and to explore novel applications in generative AI, anomaly detection, and multimodal learning.

8 REFERENCES

- [1] R. Gupta, A. Sharma, and T. Alam, "Building Predictive Models with Machine Learning," *Studies in Big Data*, vol. 145, pp. 39–59, 2024, doi: 10.1007/978-981-97-0448-4_3/FIGURES/3.
- [2] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *J Big Data*, vol. 6, no. 1, pp. 1–28, Dec. 2019, doi: 10.1186/S40537-019-0219-Y/FIGURES/5.
- [3] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics 2019, Vol. 8, Page 832*, vol. 8, no. 8, p. 832, Jul. 2019, doi: 10.3390/ELECTRONICS8080832.
- [4] P. K. Kanti, P. Sharma, M. P. Maiya, and K. V. Sharma, "The stability and thermophysical properties of Al₂O₃-graphene oxide hybrid nanofluids for solar energy applications: Application of robust autoregressive modern machine learning technique," *Solar Energy Materials and Solar Cells*, vol. 253, p. 112207, May 2023, doi: 10.1016/J.SOLMAT.2023.112207.
- [5] A. Aminimajd, J. Maia, and A. Singh, "Scalability of Graph Neural Network in Accurate Prediction of Frictional Contact Network in Suspensions," *Soft Matter*, 2025, doi: 10.1039/D4SM01391C.
- [6] Y. Zhang, H. V. Burton, H. Sun, and M. Shokrabadi, "A machine learning framework for assessing post-earthquake structural safety," *Structural Safety*, vol. 72, pp. 1–16, May 2018, doi: 10.1016/J.STRUSAFE.2017.12.001.
- [7] A. Gorban and A. Zinovyev, "Principal Manifolds and Graphs in Practice," *Int J Neural Syst*, vol. 20, no. 03, p. 219, Jan. 2010, [Online]. Available: <http://arxiv.org/abs/1001.1122>
- [8] R. A. FISHER, "THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS," *Ann Eugen*, vol. 7, no. 2, pp. 179–188, Sep. 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.
- [9] S. Hossain *et al.*, "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/ijssrmss/v9i6.110.
- [10] G. Sateesh Babu, P. Zhao, and X.-L. Li, "Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9642, Springer Verlag, 2016, pp. 214–228. doi: 10.1007/978-3-319-32025-0_14.

- [11] K. J. Kim, H. Moskowitz, and M. Koksalan, "Fuzzy versus statistical linear regression," *Eur J Oper Res*, vol. 92, no. 2, pp. 417–434, Jul. 1996, doi: 10.1016/0377-2217(94)00352-1.
- [12] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nature Reviews Genetics* 2015 16:6, vol. 16, no. 6, pp. 321–332, May 2015, doi: 10.1038/nrg3920.
- [13] Z. Sun and T. Tan, "Ordinal measures for iris recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 31, no. 12, pp. 2211–2226, 2009, doi: 10.1109/TPAMI.2008.240.
- [14] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, and A. J. Tallón-Ballesteros, "The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis," *Lecture Notes in Networks and Systems*, vol. 750 LNNS, pp. 344–353, 2023, doi: 10.1007/978-3-031-42536-3_33/FIGURES/3.
- [15] M. Y. Heravi, Y. Jang, and H. Chauhan, "Comparative Analysis of Deep-Learning Approaches for Automatic Recognition of Awkward Postures in Construction Environments through Wrist-Worn Biosensors," *J Constr Eng Manag*, vol. 151, no. 8, p. 04025108, Jun. 2025, doi: 10.1061/JCEMD4.COENG-15888.
- [16] M. Mohammadagha, "Hybridization and Optimization Modeling, Analysis, and Comparative Study of Sorting Algorithms: Adaptive Techniques, Parallelization, for Mergesort, Heapsort, Quicksort, Insertion Sort, Selection Sort, and Bubble Sort," Apr. 2025, doi: 10.31224/4537.
- [17] C. Abdollahi, M. Mollajafari, A. Golroo, S. Moridpour, and H. Wang, "A review on pavement data acquisition and analytics tools using autonomous vehicles," *Road Materials and Pavement Design*, vol. 25, no. 5, pp. 914–940, May 2024, doi: 10.1080/14680629.2023.2237601.
- [18] Y. Mousavi, G. Bevan, I. B. Küçükdemiral, and A. Fekih, "Maximum Power Extraction from Wind Turbines Using a Fault-Tolerant Fractional-Order Nonsingular Terminal Sliding Mode Controller," *Energies* 2021, Vol. 14, Page 5887, vol. 14, no. 18, p. 5887, Sep. 2021, doi: 10.3390/EN14185887.
- [19] S. R. Boroujeni, H. Abedi, and T. Bush, "Enhancing Glass Defect Detection with Diffusion Models: Addressing Imbalanced Datasets in Manufacturing Quality Control," *Computer and Decision Making: An International Journal*, vol. 2, pp. 687–707, May 2025, doi: 10.59543/comdem.v2i.14391.
- [20] M. Mohammadagha, I. Tshitenge, and I. Adebambo, "State-of-the-Art Machine Learning Techniques in Sentiment Analysis for Social Media: L'État de l'Art des Techniques d'Apprentissage Automatique en Analyse de Sentiment pour les Médias Sociaux," p. 1, May 2025, doi: 10.31224/4590.

- [21] M. Mohammadagha, M. Najafi, V. Kaushal, and A. M. A. Jibreen, "Machine Learning Models for Reinforced Concrete Pipes Condition Prediction: The State-of-the-Art Using Artificial Neural Networks and Multiple Linear Regression in a Wisconsin Case Study," Feb. 2025, [Online]. Available: <http://arxiv.org/abs/2502.00363>
- [22] J. Mashford, D. Marlow, D. Tran, and R. May, "Prediction of Sewer Condition Grade Using Support Vector Machines," *Journal of Computing in Civil Engineering*, vol. 25, no. 4, pp. 283–290, Jul. 2011, doi: 10.1061/(ASCE)CP.1943-5487.0000089/ASSET/AD324E48-8037-446F-A8C2-AE86257315BB/ASSETS/IMAGES/LARGE/1.JPG.
- [23] J. J. Zhu, M. Yang, and Z. J. Ren, "Machine Learning in Environmental Research: Common Pitfalls and Best Practices," *Environ Sci Technol*, vol. 57, no. 46, pp. 17671–17689, Nov. 2023, doi: 10.1021/ACS.EST.3C00026/ASSET/IMAGES/LARGE/ES3C00026_0007.JPEG.
- [24] D. Khuong Nguyen, G. Sermpinis, C. Stasinakis, A. Smith Business School, and E. Financ, "Big data, artificial intelligence and machine learning: A transformative symbiosis in favour of financial technology," *European Financial Management*, vol. 29, no. 2, pp. 517–548, Mar. 2023, doi: 10.1111/EUFM.12365.
- [25] S. Zhong *et al.*, "Machine Learning: New Ideas and Tools in Environmental Science and Engineering," *Environ Sci Technol*, vol. 55, no. 19, pp. 12741–12754, Oct. 2021, doi: 10.1021/ACS.EST.1C01339/ASSET/IMAGES/LARGE/ES1C01339_0001.JPEG.
- [26] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach Learn*, vol. 79, no. 1–2, pp. 151–175, Oct. 2010, doi: 10.1007/S10994-009-5152-4/METRICS.
- [27] B. Manavalan, S. Subramaniyam, T. H. Shin, M. O. Kim, and G. Lee, "Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy," *J Proteome Res*, vol. 17, no. 8, pp. 2715–2726, Aug. 2018, doi: 10.1021/ACS.JPROTEOME.8B00148/ASSET/IMAGES/LARGE/PR-2018-00148H_0006.JPEG.
- [28] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS One*, vol. 13, no. 3, p. e0194889, Mar. 2018, doi: 10.1371/JOURNAL.PONE.0194889.
- [29] S. D. Datta, M. Islam, M. H. Rahman Sobuz, S. Ahmed, and M. Kar, "Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review," *Heliyon*, vol. 10, no. 5, p. e26888, Mar. 2024, doi: 10.1016/J.HELİYON.2024.E26888/ASSET/3C33DBF1-44EB-4325-8187-2762E356BE7C/MAIN.ASSETS/GR4.JPG.
- [30] H. Inam, N. U. Islam, M. U. Akram, and F. Ullah, "Smart and Automated Infrastructure Management: A Deep Learning Approach for Crack Detection in Bridge Images,"

- Sustainability* 2023, Vol. 15, Page 1866, vol. 15, no. 3, p. 1866, Jan. 2023, doi: 10.3390/SU15031866.
- [31] Y. A. Ali, E. M. Awwad, M. Al-Razgan, and A. Maarouf, "Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity," *Processes* 2023, Vol. 11, Page 349, vol. 11, no. 2, p. 349, Jan. 2023, doi: 10.3390/PR11020349.
- [32] K. Abbas *et al.*, "Measurements and determinants of extreme multidimensional energy poverty using machine learning," *Energy*, vol. 251, p. 123977, Jul. 2022, doi: 10.1016/J.ENERGY.2022.123977.
- [33] M. M. Giraldo-González and J. P. Rodríguez, "Comparison of Statistical and Machine Learning Models for Pipe Failure Modeling in Water Distribution Networks," *Water* 2020, Vol. 12, Page 1153, vol. 12, no. 4, p. 1153, Apr. 2020, doi: 10.3390/W12041153.
- [34] M. Mohammadagha, S. Asadi, and H. K. Naeini, "Evaluating Machine Learning Performance Using Python for Neural Network Models in Urban Transportation in New York City Case Study," *Authorea Preprints*, Mar. 2025, doi: 10.36227/TECHRXIV.174239957.76698574/V1.
- [35] Mohsen Mohammadagha, Hajar Kazemi Naeini, Saeed Asadi, Mohammad Najafi, and Vinayak Kasuhal, "Machine Learning Model for Condition Assessment of Trenchless Vitrified Clay Pipes," Denver: North American Society for Trenchless Technology (NASTT) 2025 No-Dig Show, Mar. 2025. Accessed: Apr. 06, 2025. [Online]. Available: <https://hal.science/hal-05019707v1>
- [36] M. Mohammadagha, "Hyperparameter Optimization Strategies for Tree-Based Machine Learning Models Prediction: A Comparative Study of AdaBoost, Decision Trees, and Random Forest," *OSF*, Apr. 2025, doi: 10.31219/OSF.IO/XBKR5_V1.
- [37] S. S. Pakzad, N. Roshan, and M. Ghalehnovi, "Comparison of various machine learning algorithms used for compressive strength prediction of steel fiber-reinforced concrete," *Scientific Reports* 2023 13:1, vol. 13, no. 1, pp. 1–15, Mar. 2023, doi: 10.1038/s41598-023-30606-y.
- [38] Y. J. Wong, S. K. Arumugasamy, C. H. Chung, A. Selvarajoo, and V. Sethu, "Comparative study of artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) for modeling of Cu (II) adsorption from aqueous solution using biochar derived from rambutan (*Nephelium lappaceum*) peel," *Environ Monit Assess*, vol. 192, no. 7, pp. 1–20, Jun. 2020, doi: 10.1007/S10661-020-08268-4/TABLES/7.
- [39] D. Kreuzberger, N. Kuhl, and S. Hirschl, "Machine Learning Operations (MLOps): Overview, Definition, and Architecture," *IEEE Access*, vol. 11, pp. 31866–31879, 2023, doi: 10.1109/ACCESS.2023.3262138.

- [40] Y. P. Zhang *et al.*, “Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling,” *Military Medical Research* 2023 10:1, vol. 10, no. 1, pp. 1–33, May 2023, doi: 10.1186/S40779-023-00458-8.
- [41] N. Thapa, Z. Liu, D. B. Kc, B. Gokaraju, and K. Roy, “Comparison of Machine Learning and Deep Learning Models for Network Intrusion Detection Systems,” *Future Internet* 2020, Vol. 12, Page 167, vol. 12, no. 10, p. 167, Sep. 2020, doi: 10.3390/FI12100167.
- [42] M. Der Yang and T. C. Su, “Automated diagnosis of sewer pipe defects based on machine learning approaches,” *Expert Syst Appl*, vol. 35, no. 3, pp. 1327–1337, Oct. 2008, doi: 10.1016/J.ESWA.2007.08.013.
- [43] S. R. Shams *et al.*, “Assessing the effectiveness of artificial neural networks (ANN) and multiple linear regressions (MLR) in forecasting AQI and PM10 and evaluating health impacts through AirQ+ (case study: Tehran),” *Environmental Pollution*, vol. 338, p. 122623, Dec. 2023, doi: 10.1016/J.ENVPOL.2023.122623.
- [44] G. Forkuor, O. K. L. Hounkpatin, G. Welp, and M. Thiel, “High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models,” *PLoS One*, vol. 12, no. 1, p. e0170478, Jan. 2017, doi: 10.1371/JOURNAL.PONE.0170478.
- [45] J. A. J. Singh, J. Gnanasoundharam, M. Birunda, G. Sudha, S. P. Maniraj, and C. Srinivasan, “Wearable Sepsis Early Warning Using Cloud Computing and Logistic Regression Predictive Analytics,” *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2024*, 2024, doi: 10.1109/ICRITO61523.2024.10522250.
- [46] Y. N. TAHIR, U. AMINU, H. S. DJINGARE, and A. S. AJIBOLA, “UTILIZING ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS) FOR COMPLEX PATTERN RECOGNITION AND SPECIES PREDICTION FOR IRIS DATASET,” *International Journal of Engineering Processing and Safety Research*, vol. 5, no. 5, Sep. 2024, Accessed: Mar. 22, 2025. [Online]. Available: <https://cambridgeresearchpub.com/ijepsr/article/view/312>
- [47] A. S. Al-Waisy, R. Qahwaji, S. Ipson, S. Al-Fahdawi, and T. A. M. Nagem, “A multi-biometric iris recognition system based on a deep learning approach,” *Pattern Analysis and Applications*, vol. 21, no. 3, pp. 783–802, Aug. 2018, doi: 10.1007/s10044-017-0656-1.
- [48] Alex Dillhoff, “Alex Dillhoff.” [Online]. Available: <https://ajdillhoff.github.io/>
- [49] A. Zollanvari, “Supervised Learning in Practice: the First Application Using Scikit-Learn,” *Machine Learning with Python*, pp. 111–131, 2023, doi: 10.1007/978-3-031-33342-2_4.

- [50] L. Liu and R. C. Chen, "A novel passenger flow prediction model using deep learning methods," *Transp Res Part C Emerg Technol*, vol. 84, pp. 74–91, Nov. 2017, doi: 10.1016/J.TRC.2017.08.001.
- [51] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Syst Appl*, vol. 42, no. 4, pp. 2162–2172, Mar. 2015, doi: 10.1016/J.ESWA.2014.10.031.
- [52] Viveksinght, "Exploratory data analysis(EDA) on iris data set by python. | by Viveksinght | EDA iris dataset | Medium." Accessed: Feb. 12, 2025. [Online]. Available: <https://medium.com/eda-iris-dataset/exploratory-data-analysis-eda-on-iris-data-set-by-python-fbaae921654a>
- [53] statistician and Ronald Fisher, "Iris flower data set - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Iris_flower_data_set
- [54] Nicoguardo, "File:Iris dataset scatterplot.svg - Wikimedia Commons." [Online]. Available: <https://commons.wikimedia.org/w/index.php?curid=46257808>
- [55] A. Wayne McCord and W. Creek, "Method and system for enabling automated and real-time discovery of skills available to agents and systems in a multimedia communications network," Jun. 2003, Accessed: Jun. 21, 2025. [Online]. Available: <http://computer.org/internet?>,
- [56] M. Abdin *et al.*, "Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone," Apr. 2024, Accessed: Jun. 21, 2025. [Online]. Available: <https://arxiv.org/pdf/2404.14219>
- [57] N. Du *et al.*, "GLaM: Efficient Scaling of Language Models with Mixture-of-Experts," PMLR, Jun. 2022. Accessed: Jun. 21, 2025. [Online]. Available: <https://proceedings.mlr.press/v162/du22c.html>
- [58] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan, "PointFlow: 3D Point Cloud Generation With Continuous Normalizing Flows," 2019.
- [59] J. Thiyagalingam, M. Shankar, G. Fox, and T. Hey, "Scientific machine learning benchmarks," *Nature Reviews Physics*, vol. 4, no. 6, pp. 413–420, Jun. 2022, doi: 10.1038/S42254-022-00441-7;SUBJMETA=1046,172,639,704,705;KWRD=ENVIRONMENTAL+SCIENCES,SCIENTIFIC+DATA.
- [60] J. Liu *et al.*, "Adaptive Asynchronous Federated Learning in Resource-Constrained Edge Computing," *IEEE Trans Mob Comput*, vol. 22, no. 2, pp. 674–690, Feb. 2023, doi: 10.1109/TMC.2021.3096846.
- [61] T. G. Dietterich, "Ensemble Methods in Machine Learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1.

- [62] A. Mosavi, F. Sajedi Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva, and E. Rafiei Sardooi, "Ensemble Boosting and Bagging Based Machine Learning Models for Groundwater Potential Prediction," *Water Resources Management*, vol. 35, no. 1, pp. 23–37, Jan. 2021, doi: 10.1007/S11269-020-02704-3/TABLES/3.
- [63] B. Pavlyshenko, "Using Stacking Approaches for Machine Learning Models," *Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018*, pp. 255–258, Oct. 2018, doi: 10.1109/DSMP.2018.8478522.
- [64] E. Bauer and R. Kohavi, "Empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Mach Learn*, vol. 36, no. 1, pp. 105–139, 1999, doi: 10.1023/A:1007515423169/METRICS.
- [65] R. Maulik, O. San, J. D. Jacob, and C. Crick, "Sub-grid scale model classification and blending through deep learning," *J Fluid Mech*, vol. 870, pp. 784–812, Jul. 2019, doi: 10.1017/JFM.2019.254.
- [66] Y. Zhou *et al.*, "Mixture-of-Experts with Expert Choice Routing," *Adv Neural Inf Process Syst*, vol. 35, pp. 7103–7114, Dec. 2022.
- [67] B. Zoph and Q. V. Le, "Neural Architecture Search with Reinforcement Learning," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, Nov. 2016, Accessed: Jun. 21, 2025. [Online]. Available: <https://arxiv.org/pdf/1611.01578>