RESEARCH SERIES.

UNIVERSITY OF TEXAS AT ARLINGTON

DEPARTMENT OF CIVIL ENGINEERING

Hybrid Machine Learning Meta-Model for the Condition Assessment of Urban Underground Pipes

Mohsen Mohammadagha[1✉*], Mohammad Najafi[2],
Vinayak Kaushal[3], Ahmad Jibreen[4]

Abstract: Urban water infrastructure faces increasing deterioration, necessitating accurate, cost-effective condition assessment. Traditional inspection techniques are intrusive and inefficient, creating a demand for scalable machine learning (ML) solutions. This study develops a hybrid ML meta-model to predict underground pipe conditions using a comprehensive dataset of 11,544 records. The objective is to enhance multi-class classification performance while preserving interpretability. A stacked hybrid architecture was employed, integrating CatBoost, LightGBM, XGBoost, AdaBoost, TabNet, ANN, Logistic and Linear Regression, and Random Forest models. Following data preprocessing, feature engineering, and correlation analysis, the meta-model achieved 96.6% accuracy, outperforming individual models. Age emerged as the most influential feature, followed by material type and pipe length. ROC-AUC scores exceeded 0.95 across classes, confirming high discriminative capability. This work demonstrates the superiority of hybrid architectures for infrastructure diagnostics. Future research should incorporate real-time IoT sensor data and advanced models such as Graph Neural Networks or Transformers for dynamic, network-level condition forecasting.

Keywords: *Machine Learning, Meta-Learning Methods, Condition Assessment, Water Pipe, Underground Infrastructure.*

✉ **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: ahmadmahmoudahm.jibreen@mavs.uta.edu

1 INTRODUCTION

Urban underground pipe networks are fundamental to supporting the health, safety, and economic vitality of cities [1], providing essential services such as water supply, sewage management [2], and stormwater drainage. As the age and complexity of these buried infrastructures [3] increase, so too do the challenges associated with maintenance and timely rehabilitation. Asset managers frequently encounter difficulties in assessing the actual condition of underground pipes due to the limited accessibility and visibility inherent in these systems. Traditional assessment methods [4] often involve intrusive, time-consuming, and costly physical inspections, which may not adequately capture the nuanced interplay of factors influencing pipe deterioration. This growing need for efficient, accurate, and scalable assessment strategies has prompted researchers and practitioners to explore advanced data-driven approaches for pipe condition [4] evaluation.

Recent advances in machine learning have opened new avenues for condition assessment by leveraging the diverse and voluminous data generated from urban [5] utility networks. By systematically incorporating variables such as pipe age, material, diameter, length, soil properties, slope, and environmental indices, machine learning enables the extraction of hidden patterns and relationships beyond human capability. Conventional models [6], such as logistic regression [7] or decision trees , have demonstrated promise in classifying pipe condition [8] states when trained on large-scale datasets. However, the performance of any single model is often limited by dataset complexity [9], nonlinearity of the underlying processes, and the presence of noisy or missing data. To address these limitations, hybrid approaches [10] that combine multiple machine learning algorithms [11] into meta-models have emerged as powerful alternatives, offering robustness and improved generalizability.

Hybrid machine learning [10] meta-models integrate the strengths of individual learners by employing both integration strategies and meta-learning techniques, making them more advanced and efficient compared traditional ensemble approaches. This study explores a novel meta-model architecture [12] for the classification of water pipe conditions [13] an urban context. Diverse base models—such as Random Forest [14] [15], LightGBM [16], CatBoost [17], TabNet, and Artificial Neural Networks [18], are trained on a comprehensive data set featuring key attributes of urban pipes. Through stacking and meta-learning, their collective predictions are synthesized to achieve higher predictive accuracy and resilience to data imperfections. This approach enhances the representation of complex interactions [19] among features, mitigates the risks of overfitting, and provides interpretable insights to guide infrastructure [20] management decisions. Model evaluation is conducted via multifaceted metrics [21], including accuracy, precision, recall, and F1 score, supplemented by visualizations and feature importance analyses [22].

Overall, this research demonstrates the feasibility and advantages of a hybrid machine learning meta-model framework for the condition assessment of urban [23] underground pipes [24]. By incorporating heterogeneous algorithms [25] and focusing on both prediction quality and interpretability, the study addresses critical gaps in current assessment methodologies. The proposed approach fosters more reliable decision-making for urban asset management, potentially reducing maintenance costs, optimizing rehabilitation timing, and ultimately advancing the resilience of urban infrastructure systems. The findings highlight a scalable, data-driven path for municipalities and utility providers to modernize their assessment protocols and improve service continuity in rapidly evolving urban environments. The basic structure of this paper is as follows. Section 2 presents Related Works on machine learning approaches for infrastructure condition assessment; Section 3 introduces the main Methodology including the hybrid meta-model framework and evaluation metrics; Section 4 exhibits the experimental results from the stacking architecture and individual model comparisons; Section 5 discusses the results including feature importance analysis and model performance evaluation; and Section 6 concludes this paper and provides future research suggestions.

2 RELATED WORKS

Recent advancements in urban underground pipe condition assessment [26] research reveal a shift from traditional inspection techniques toward the integration of machine learning and data-driven strategies [27]. This evolution is propelled by the need to overcome the limitations of manual and technology-driven approaches, aiming for scalable solutions that ensure more accurate pipeline health monitoring [28]. To provide a comprehensive background, the following paragraphs critically compare pairs of significant studies, focusing on their methodologies, datasets, and results, and effectively illustrate the progression in this research domain.

To begin, the work of Zheng Liu and Yehuda Kleiner (2013) can be contrasted with the empirical modeling approach of Mosavi et al. (2020). Whereas Liu and Kleiner provided a qualitative evaluation of direct and indirect technologies—such as CCTV, ultrasound, and smart robotics—their review highlighted capability (e.g., SmartBall detecting leaks <0.026 L/h, LeakFinderRT location <10 cm) but did not employ a quantitative dataset [29]. In contrast, Mosavi et al. applied Recursive Feature Elimination and ensemble machine learning (Random Forest, AdaBoost, GamBoost, Bagged CART) to a dataset of 339 groundwater locations and 15 variables. The Random Forest achieved an accuracy of 0.86 and recall of 0.91, outperforming boosting models. Thus, this comparison demonstrates how transitioning from reviews of technical tools to integrated data-driven frameworks can yield higher predictive accuracy and actionable outcomes [30].

Similarly, a notable comparison can be drawn between Rakiba Rayhana et al. (2021) and Mohsen Mohammadagha et al. (2025), reflecting the progress from system-wide vision technology reviews to targeted machine learning implementations. While Rayhana et al. synthesized findings from datasets of 100 to over 2 million CCTV and SSET pipe images—demonstrating deep learning models such as DCNN and Faster R-CNN achieving defect detection accuracies up to 98% [31]—Mohammadagha et al. systematically modeled 612 cases of reinforced concrete sewer pipe inspections using both Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR). The ANN model yielded $R^2 = 0.9066$, outperforming MLR [32]. Both studies endorse data-intensive approaches but differ in scale, with Rayhana et al. emphasizing automated vision at a network level and Mohammadagha et al. focusing on feature-driven pipeline condition forecasting at the asset level.

Furthermore, when considering the broad landscape of pipeline monitoring, the reviews by Jawwad Latif et al. (2022) and by Liu & Kleiner (2013) illuminate the evolution of sensor integration and methodological sophistication. On the one hand, Latif et al. categorized monitoring into acoustic, electromagnetic, visual, and IoTenabled methods, discussing visual classifiers like YOLOv3 and acoustic detections (SmartBall <0.1 gal/hr) while emphasizing the need for robust, cost-effective, and real-time machine learning integration [33]. On the other hand, Liu & Kleiner mainly addressed the capabilities and cost limitations of traditional and semi-automated systems. This comparison highlights a shift from static technology evaluation to the advocacy for dynamic, adaptable, and intelligent monitoring platforms [29].

Equally important, the studies of Thikra Dawood et al. (2020) and Rakiba Rayhana et al. (2021) exemplify the harmonization of artificial intelligence theory with practical image-based inspection. Dawood et al. reviewed 66 studies across seven AI model categories, reporting that ANN models can reach R^2 of up to 0.9510 for failure prediction, though their findings relied on synthesizing published results rather than a singular dataset [34]. Conversely, Rayhana et al. demonstrated that vision-based deep learning (e.g., Faster R-CNN) achieved up to 98% accuracy in defect detection across enormous and diverse image collections [31]. Both works underscore the value of hybrid and data-driven frameworks, yet Dawood et al. foreground the potential of hybrid modeling logic, while Rayhana et al. stress the strengths of advanced computer vision in real-world application.

Ultimately, while previous studies have advanced pipe condition assessment [35] with machine learning, Hybrid models [36], and meta-analysis, existing research has not systematically evaluated nor benchmarked a comprehensive hybrid meta-model that integrates the latest tree-based mixture [37] methods (CatBoost [38], XGBoost [39], LightGBM [40]), deep learning (TabNet, ANN), and logistic/meta-learning for multi-class urban pipe condition prediction using real, multifaceted operational datasets [41]. In this research, this research address these gaps by designing and implementing a unified hybrid meta-learning framework, comparing algorithmic performance and feature importance, and providing interpretable model diagnostics using a large-scale, diverse urban pipe dataset.

3 METHODOLOGY

The methodology for this research is structured around a hybrid machine learning meta-model, designed for interpretable condition assessment of urban underground water pipes. The urban water pipe dataset comprises 11,544 records sourced from New Zealand's municipal infrastructure systems, with 3,297 records from the South Island and 8,247 records from the North Island, reflecting the comprehensive coverage of both the Waimakariri District Council (2022) [25] and Matamata-Piako District Council (2024) [26] networks [27]. While the original dataset contained more environmental and operational parameters, this study strategically selected seven core parameters that are most commonly available and represent the primary factors influencing pipe deterioration. Building on the Figure 1 workflow diagram, the process begins with extensive data acquisition, collecting features such as age, material, length, diameter, slope, soil properties, and thaw index from operational pipe inventories. Exploratory Data Analysis (EDA) then identifies outliers and distribution patterns, followed by data cleaning to address missing values and eliminate errors. Feature selection and scaling are performed to normalize variable influence, with categorical condition labels encoded for supervised learning. Correlation analysis and multidimensional visualizations guide further feature engineering, while summary plots (histograms, bar charts) illustrate the dominance of certain materials (e.g., Medium

Density Polyethylene, Unplasticized Polyvinyl Chloride) and the relationship between age, diameter, length, and deterioration, setting a transparent foundation for modeling.

Figure 1. Workflow for Underground Condition Assessment Using meta-learning approaches

A comprehensive suite of visual and statistical analyses is used to profile the dataset and guide model development. Feature correlation matrix heatmaps in Figure 2 reveal associations (e.g., slope-soil, r=0.69) and identify negative dependencies (slope-thaw index, r=-0.55), while 3D condition plots and kernel density estimations demonstrate how younger pipes are more likely to have better condition ratings. Material frequency bar charts, pair plots, and boxplots disaggregate key features—such as age, diameter, and length—across condition classes, highlighting that aging, as well as certain material types, are primary predictors of asset deterioration. These insights ensure that feature importance interpretation is both data-driven and aligned with real-world engineering knowledge, enhancing model transparency and domain relevance. Advanced visualizations, such as 3D health plots and pairwise feature analysis, underpin the data exploration stage and justify variable selection for the modeling pipeline.

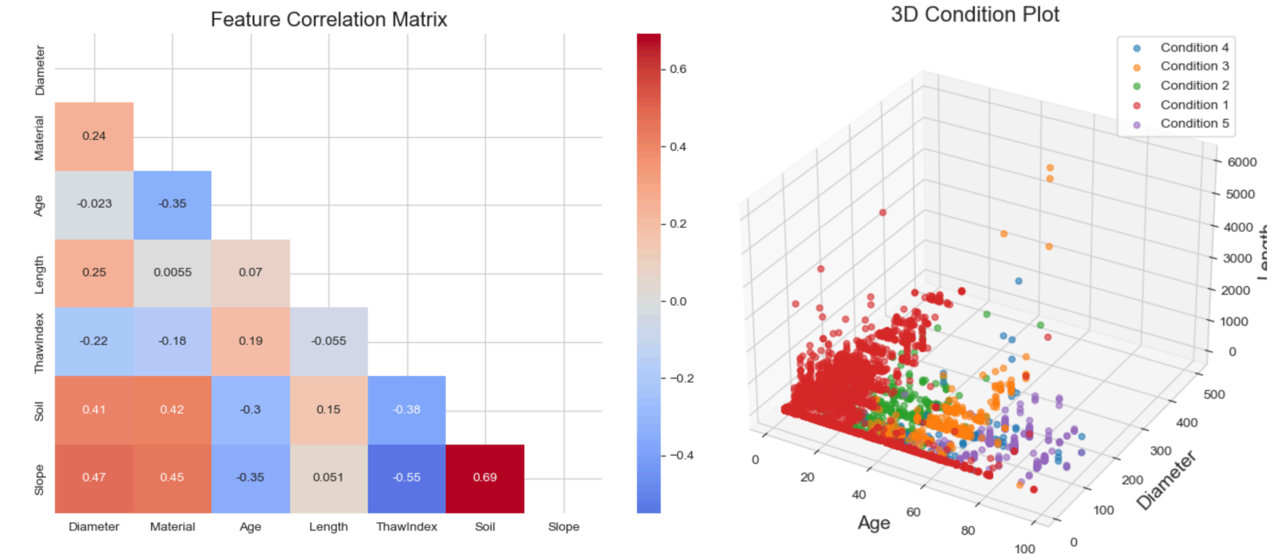


Figure 2. Feature Correlation Matrix and 3D condition Plot

Building upon these mathematical foundations, the combined visualization of age distribution and material frequency, **Error! Reference source not found.** provides c omprehensive dataset insights. The age histogram reveals a pronounced skew toward newer infrastructure, with most pipes under 50 years old and sharp frequency decline for older assets. Simultaneously, bar chart displays the frequency of urban water pipe materials used in the dataset. Medium Density Polyethylene (MDPE) dominates, followed by Unplasticized Polyvinyl Chloride (UPVC) and Polyvinyl Chloride (PVC).

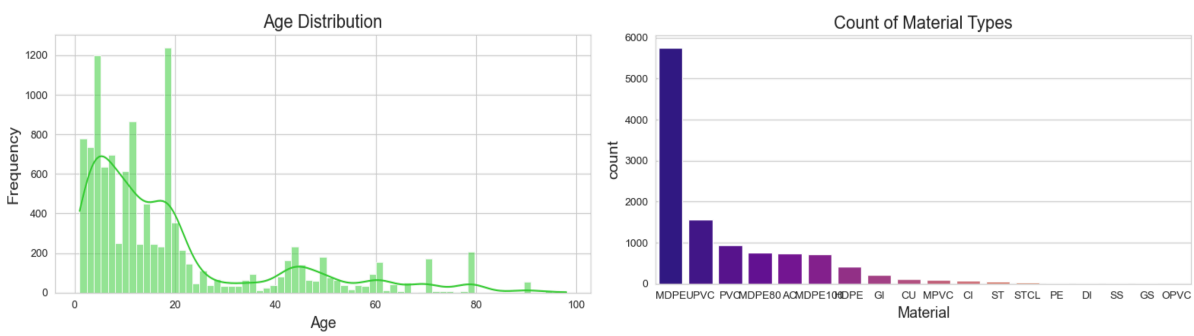


Figure 3. Urban water pipes: age-skewed asset with MDPE the most common

Others include Medium Density Polyethylene 80 (MDPE80), Asbestos Cement (AC), Medium Density Polyethylene 100 (MDPE100), High Density Polyethylene (HDPE), Galvanized Iron (GI), Copper (CU), Modified Polyvinyl Chloride (MPVC), Cast Iron (CI), Steel (ST), Steel Clad (STCL), Polyethylene (PE), Ductile Iron (DI), Stainless Steel (SS), Galvanized Steel (GS), and Oriented Polyvinyl Chloride (OPVC). MDPE pipes are by far the most common, followed by Unplasticized Polyvinyl Chloride (UPVC) and standard Polyvinyl Chloride (PVC), reflecting modern urban infrastructure trends and material evolution patterns.

Advancing from material composition analysis, the kernel density estimation in Figure 4 reveals critical age-condition relationships that validate deterioration patterns. Condition 1 pipes exhibit sharp density peaks at younger ages (below 20 years), while Conditions 2 and 3 demonstrate broader distributions centered around 40-60 years. Most significantly, Conditions 4 and 5 show pronounced density concentrations beyond 60 years, establishing age as the primary deterioration predictor and confirming time-dependent infrastructure degradation patterns essential for predictive modeling.

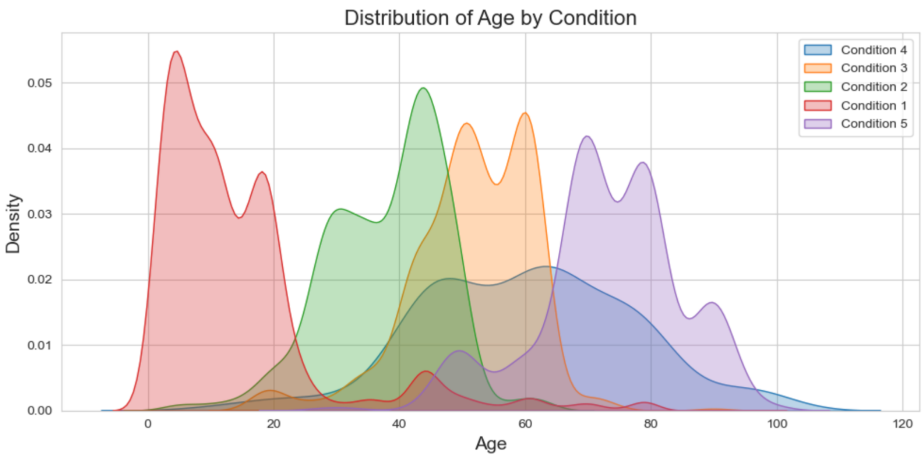


Figure 4. Age distribution reveals deterioration pattern of water pipes

The comprehensive pairplot in Figure 5 shows the matrix extends individual feature analysis by visualizing multivariate relationships across all seven input variables, colored by condition class. This analysis reveals distinct clustering patterns and separation boundaries between condition categories, particularly highlighting how combinations of diameter, age, material type, and length create identifiable condition signatures. The diagonal density plots and scatter plot matrices demonstrate clear feature interactions that justify the multi-model approach, showing complex non-linear relationships requiring sophisticated algorithmic treatment.

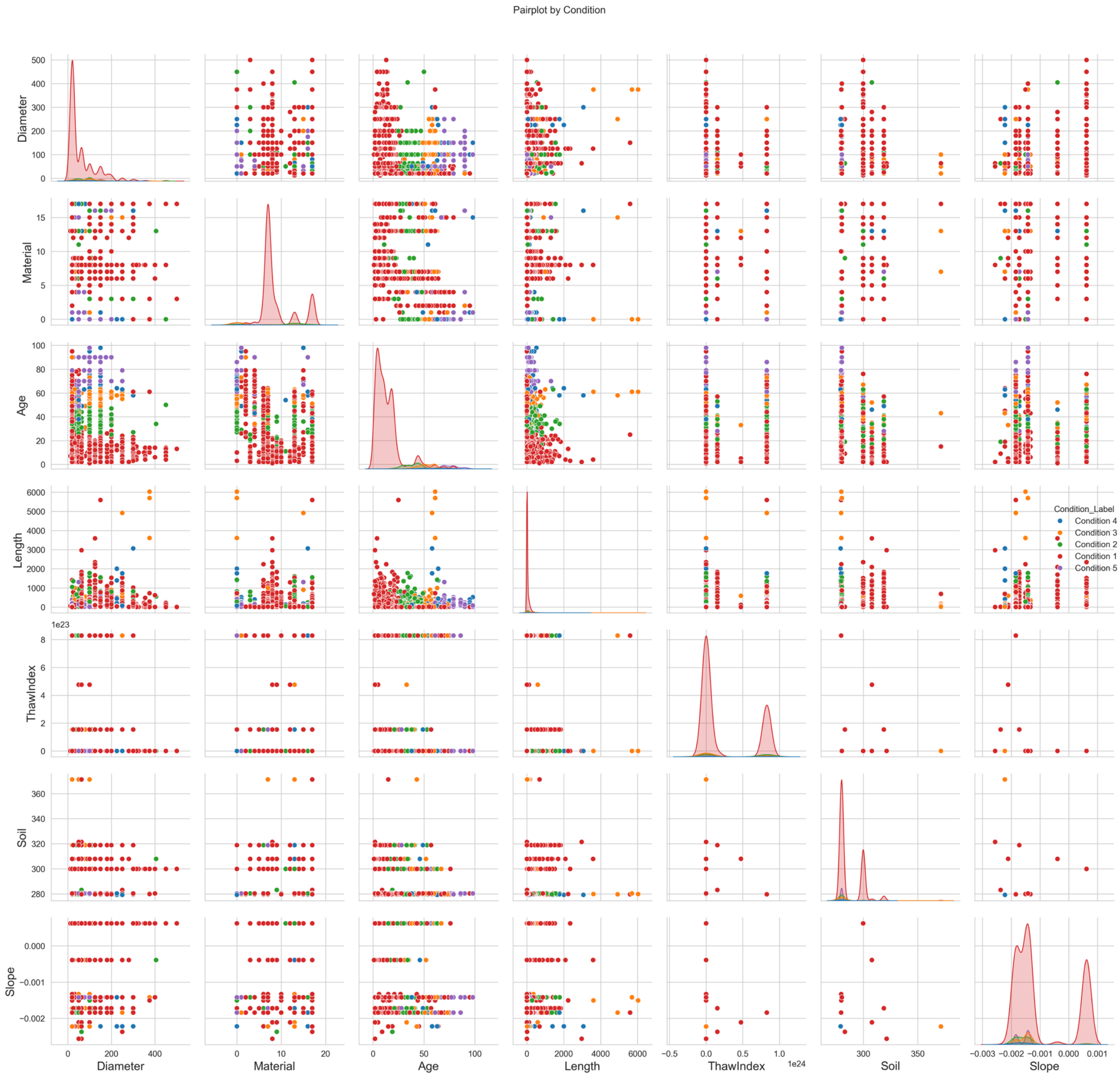


Figure 5. Clusters of pipe condition classes by age, material, diameter and length

Finally, the boxplot analysis in Figure 6, quantifies distributional differences across condition classes for each feature, providing statistical validation of predictive variable importance. Age exhibits the most dramatic progression from young pipes in Condition 1 to significantly older assets in Condition 5, while diameter and length show notable variability within condition categories. Material, soil, thaw index, and slope demonstrate varying degrees of separation, with age emerging as the most discriminative feature, directly supporting the feature importance rankings obtained from the machine learning models.

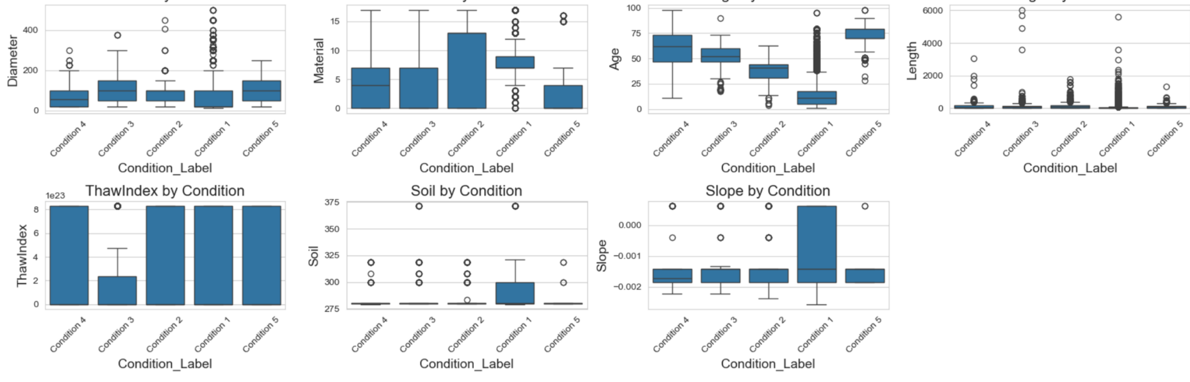


Figure 6. Boxplots of Key Pipeline Features by Condition Category

A crucial component of this methodology is the use of mathematical formulas and evaluation metrics in the background of the models, ensuring best-practice model comparison and interpretability by calculating the accuracy, Precision, Recall, and F1 score. Seven common formulas drive the assessment: (1) Min-Max Normalization for feature scaling [42]; (2) Accuracy for overall performance; (3) Precision and (4) Recall (Sensitivity) for quantifying the correctness and completeness of class predictions; (5) F1 Score to balance precision and recall, especially under class imbalance; (6) Feature Importance, typically the mean decrease in impurity in tree-based models, to prioritize influential variables; and (7) Stacking Prediction [43], which mathematically combines the predictions of diverse base models via a meta-learner for improved overall robustness. This methodology by using of well-established mathematical formulas [44] and evaluation metrics that underpin the modeling and assessment processes. Some common formulas central to this study include Min-Max Normalization for Purpose Scales, a feature (variable) to a range between 0 and 1. This ensures all features contribute equally to the model and prevents features with larger ranges from dominating, which is shown in equation 1 [45]:

$$\mathrm{x}' = \frac{\mathrm{x} - \mathrm{x}_{\min}}{\mathrm{x}_{\max} - \mathrm{x}_{\min}} \tag{1}$$

Accuracy is a fundamental metric for evaluating classification models, representing the proportion of correctly predicted instances among all predictions. Calculated as shown in equation 2, it provides a straightforward measure of overall model performance, especially when the dataset is balanced between classes [28].

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{2}$$

Precision quantifies the correctness of positive predictions made by a model. Defined as shown in equation 3, it measures the proportion of true positives among all instances predicted as positive. High precision indicates that the model makes very few false positive errors, which is vital in many applications [46].

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

Recall, also known as sensitivity, assesses a model's ability to identify all relevant positive cases. The formula as shown in equation 4, calculates the proportion of actual positives correctly predicted. High recall is essential when missing positive instances has significant consequences, such as in medical diagnoses or fraud detection.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

The F1 score is the harmonic mean of precision and recall, providing a balanced metric for model evaluation, particularly with imbalanced datasets. Calculated as shown in equation 5, it penalizes extreme values and offers a single, interpretable measure of model effectiveness.

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{5}$$

Feature importance [47] quantifies the contribution of each input variable to a model's predictive power. In tree-based models, it is often computed as the mean decrease in impurity as shown in equation 6, when a feature is used for splitting. This metric aids in model interpretation, enabling practitioners to identify and prioritize influential variables in decision-making.

$$\text{Feature Importance}_j = \sum_{t \in \text{Splits on } j} \frac{N_t}{N} \cdot \Delta I_t \tag{6}$$

Where t is indexes the tree nodes where feature j is used for splitting. N_t is the number of samples reaching node t . N is the total number of samples. ΔI_t is the decrease in impurity (such as Gini impurity or entropy) caused by the split at node t. Stacking prediction combines multiple base models' outputs using a meta-learner to improve predictive accuracy and robustness. The final prediction is expressed as shown in equation 7. This approach leverages the strengths of diverse models, often outperforming individual learners in complex tasks.

$$\hat{\mathrm{y}} = \mathrm{f}_{\mathrm{meta}}\left(\mathrm{f}_1(\mathrm{X}), \mathrm{f}_2(\mathrm{X}), \ldots, \mathrm{f}_\mathrm{n}(\mathrm{X})\right) \tag{7}$$

By applicable of all the above formula in the model, the pipeline deploys and benchmarks Random Forest, LightGBM, CatBoost, AdaBoost, XGBoost, TabNet, Logistic Regression, Linear Regression, and Artificial Neural Network (ANN) models—each trained on stratified training/test splits, and their performance compared using accuracy, precision, recall, and F1. Individual feature importances are computed for transparent, actionable interpretation; stacking meta-models aggregate these predictions to further elevate performance, achieving a stacking accuracy of ~0.966—superior to any single method. The integration of TabNet represents methodological novelty, allowing enhanced representation learning directly from tabular infrastructure data.

In summary, this research advances the condition assessment of urban pipes by integrating a diverse stack of machine learning models—uniquely combining tree-based, neural, and attention-driven deep tabular learning approaches within a rigorous meta-learning framework. Some prior literature that focused on classic or singly-ensembled models, our workflow consistently benchmarks combined Meta-Learning models, which find better operational water pipe data, revealing domain-driven feature patterns and reporting interpretable decision rules. The principal novelty lies in the introduction of TabNet within the meta-model learning, directly addressing prior gaps in feature interaction learning and transparency. This architecture delivers state-of-the-art prediction and interpretability, enabling urban utilities and asset managers to implement data-driven, generalizable, and actionable assessments for maintenance prioritization and long-term infrastructure resilience

4 RESULTS

The hybrid machine learning meta-model developed for urban underground pipe condition assessment was rigorously evaluated using a comprehensive dataset of 11,544 water pipe records. The dataset encompasses seven key features: diameter, material, age, length, thaw index, soil properties, and slope, with condition ratings ranging from 1 (excellent) to 5 (poor). This section presents the detailed analysis results, including feature correlations, distribution patterns, model performance comparisons, and the effectiveness of the proposed stacking meta-model architecture. The results demonstrate the superiority of hybrid methods over individual algorithms and provide insights into the most influential factors affecting pipe deterioration.

Figure 7 presents a comprehensive analysis dashboard displaying the performance and characteristics of a Stacking Model developed for urban water pipe condition assessment. This machine learning framework represents a meta-learning approach that combines multiple base algorithms to achieve superior predictive accuracy. In the Performance Metrics Overview, the stacking model demonstrates exceptional results across all key indicators, achieving 96.6% accuracy, 96.3% precision, 96.6% recall, and a 96.3% F1-score. These metrics, displayed in the colorful bar chart, indicate that the meta-model approach significantly outperforms individual models. The 1-Loss metric (shown in purple) represents the error rate, confirming the model's robust performance with minimal prediction errors. This level of accuracy is particularly impressive given the complexity of multi-class pipe condition classification, where conditions range from 1 (excellent) to 5 (poor). The confusion matrix reveals the model's classification accuracy across different condition categories. The diagonal pattern shows strong true positive predictions, with 1,985 correct predictions for Condition 1 and 76 correct predictions for Condition 5. The matrix demonstrates minimal misclassification errors, which is particularly notable in the clear distinction between excellent (Condition 1) and poor (Condition 5) pipe conditions. The color intensity gradient from blue to red effectively visualizes prediction density, with darker blue regions indicating higher prediction accuracy. The Age-Based Deterioration Patterns, illustrated by a percentage distribution chart, explore the relationship between pipe age and condition ratings over time. The area plot reveals that newer pipes (0–15 years) predominantly remain in Condition 1, while pipes aged 40–65 years exhibit increased deterioration patterns with higher percentages falling into Conditions 3, 4, and 5. This visualization supports the age-dependent deterioration hypothesis, which is central to infrastructure management, where material degradation notably accelerates after the 40-year threshold.

In Neural Network Training Dynamics, the ANN accuracy plot demonstrates the training convergence pattern over 50 epochs, with both training and validation accuracy curves stabilizing at around 95%. This convergence without significant overfitting indicates strong model generalization capability, which is essential for real-world deployment in urban infrastructure assessment. In the Decision Boundary Visualization, the bottom panel displays PCA-reduced 2D decision boundaries for four algorithms: LightGBM, Random Forest, AdaBoost, and Decision Tree. These visualizations illustrate how each algorithm approaches the classification task differently, showcasing unique boundary patterns and cluster separations. The colored regions represent different condition classes, while the scattered points reflect the actual data distribution in the reduced feature space. The Meta-Learning Architecture utilizes a stacking approach that integrates predictions from Random Forest, LightGBM, CatBoost, and Logistic Regression as base learners, with a final meta-learner synthesizing their outputs. This meta-model methodology addresses the limitations of individual models while leveraging their collective strengths, leading to improved robustness and enhanced accuracy in urban pipe condition assessment.

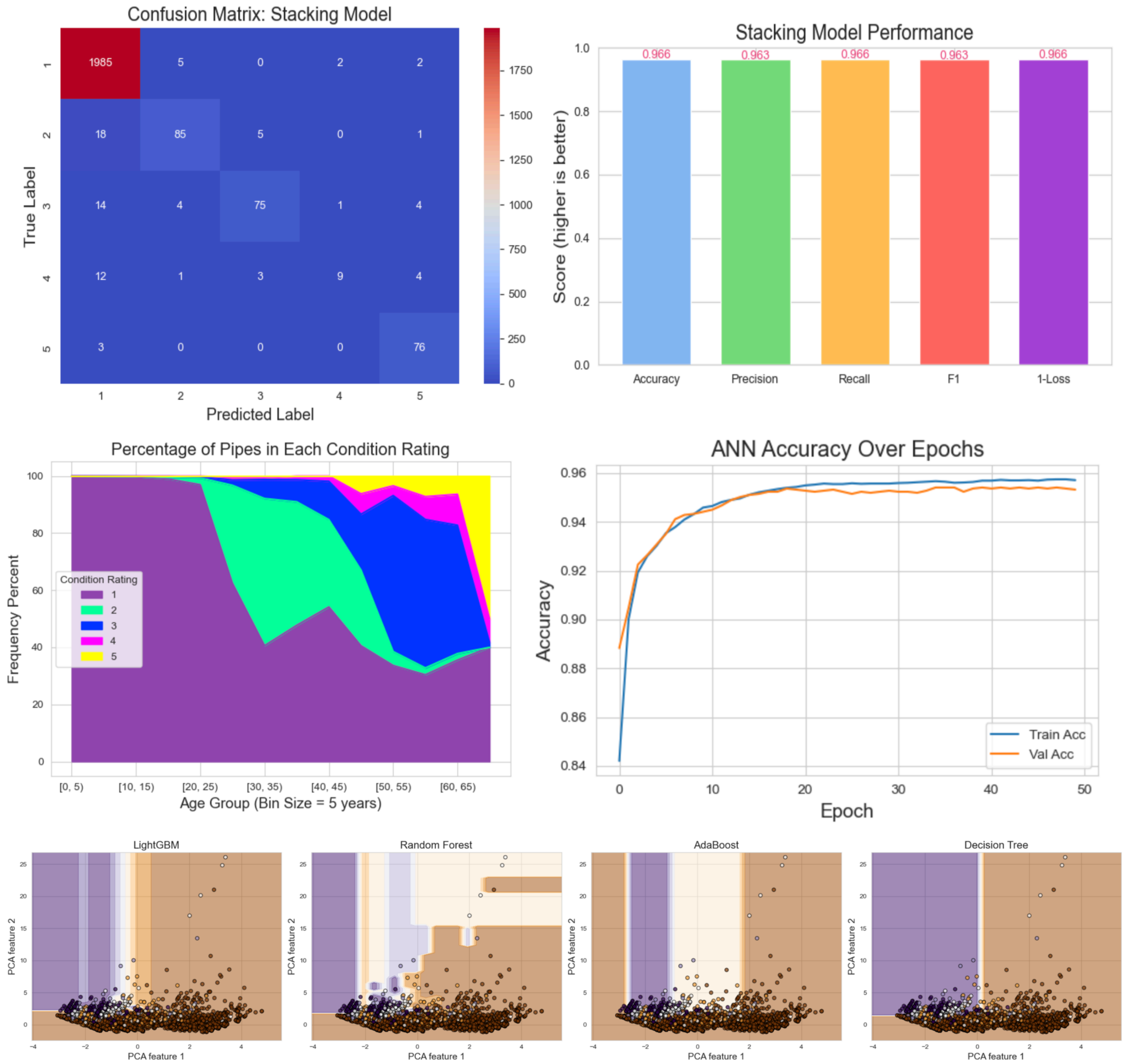


Figure 7. Pipe Cndition Classification, ANN Accuracy and Confusion Matrix in Stacking Meta-Model

Figure 8 presents feature importance plots for four machine learning models—TabNet, CatBoost, XGBoost, and AdaBoost—applied to the assessment of urban water pipe conditions. Each plot highlights which input features most strongly influence the model's predictions, revealing significant variation across algorithms: TabNet identifies "Age," "Diameter," and "Slope" as the dominant predictors, with "Age" standing out as most influential. CatBoost emphasizes "Age" heavily, followed by "Diameter," "Length," and "Material"—showing a notable drop in importance for soil, slope, and thaw index. XGBoost also puts "Age" and "Diameter" at the top but assigns relatively higher importance to "Material," while "Length" is least impactful.

AdaBoost differs by ranking "Length" as the most important feature, then "Age," shifting traditional expectations. These divergences arise because each model processes variable interactions and handles categorical, imbalanced, or noisy data differently. For instance, neural models like TabNet learn deep feature representations, potentially prioritizing statistical patterns distinct from boosting methods. Tree-meta models (CatBoost, XGBoost) are sensitive to how features are split at each decision node and can inflate importance for variables that best reduce impurity on particular data samples. Algorithm sensitivity plays a key role: tree-based models might prioritize features that split the data early or frequently in the meta trees, while TabNet's sequential attention can focus on combinations not as apparent in trees. Additionally, feature correlations influence rankings—if two features are correlated, one model may split importance between them, while another may favor just one. The way each model handles data—such as dealing with missing values, feature scaling, or encoding—also affects how certain variables appear more or less important, especially for categorical inputs like "Material." Moreover, differences in dataset characteristics and research goals further shape feature importance outcomes.

Urban utility networks can vary significantly in terms of construction era, pipe materials, maintenance records, and documentation practices, which introduces variability in available features and their predictive power across different studies. Studies may also differ in core tasks, such as binary failure prediction versus multi-class condition rating, which naturally shifts the relative weight of influential variables. Furthermore, the choice of machine learning model itself influences the findings: some models are designed for maximum predictive performance, while others prioritize simplicity or interpretability, making the importance of specific features context-dependent. In summary, differences in feature importance reflect a dynamic interplay of data characteristics, model architecture, and research objectives. This variation explains why no universal ranking of features exists across predictive studies of water pipe condition and underscores the value of meta, hybrid, and meta-learning strategies—as explored in the attached research—for generating more robust, generalizable insights.

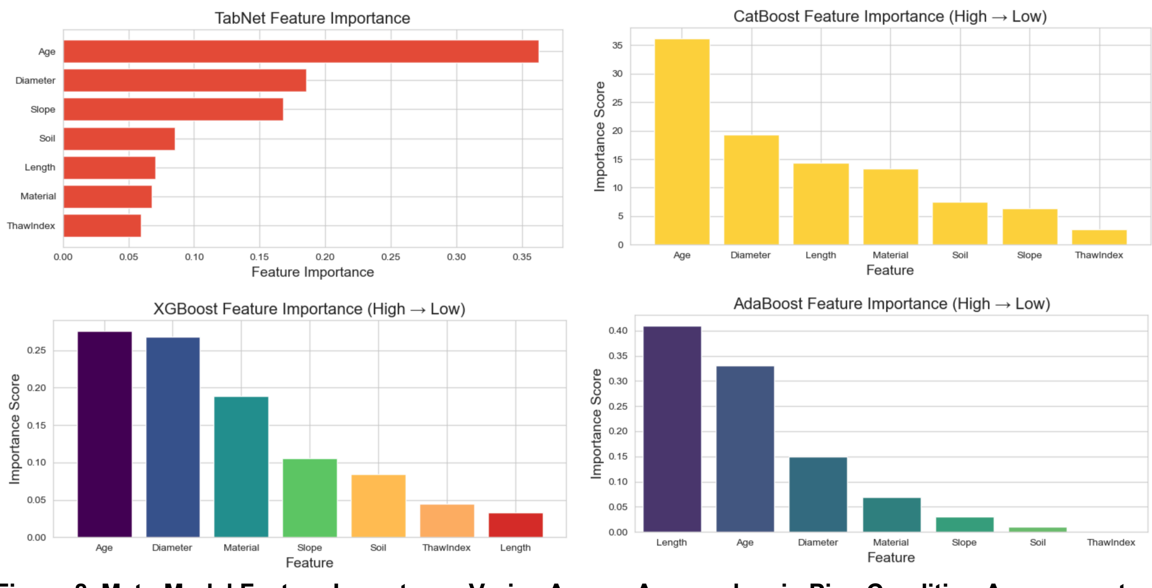


Figure 8. Meta-Model Feature Importance Varies Across Approaches in Pipe Condition Assessment

The performance analysis of various machine learning models for urban water pipe condition assessment, as illustrated in Figure 9 through receiver operating characteristic (ROC) curves, paints a nuanced picture. hybrid algorithms—particularly stacking, CatBoost, LightGBM, and XGBoost—consistently achieve high overall accuracy and AUC values, frequently surpassing 0.96 for both macro and micro averages. These models demonstrate strong discriminative capabilities across all pipe condition categories. Their strength lies in combining predictions from diverse base models, which enables them to capture complex variable interactions and nonlinear relationships that simpler models, such as logistic regression and AdaBoost, often miss. Visualizations confirm that advanced models like stacking, CatBoost, XGBoost, TabNet, and artificial neural networks (ANNs) produce ROC curves that closely hug the top-left corner, indicating high true positive rates with low false positive rates.

Despite a strong overall performance, a closer inspection at the class level reveals significant disparities in predictive accuracy. Class 1 (best condition) and Class 5 (worst condition) are classified with almost perfect precision, achieving AUC scores between 0.96 and 1.00—a reflection seen in their close alignment with the ideal ROC diagonal. In contrast, the intermediate categories—especially Class 2 and Class 3—pose greater challenges. For instance, AdaBoost registers notably poor results for these classes, with AUCs as low as 0.28 and 0.41. Even top-performing meta models can experience a decline in AUC for these middle conditions, underscoring a broader issue in multiclass classification. These inconsistencies are largely attributed to class imbalance and overlapping feature distributions, which hinder model differentiation between moderately degraded pipe conditions. Additionally, real-world data uncertainty, including measurement variability and observational ambiguity, further complicates accurate classification for these intermediate states.

Model architecture plays a critical role in managing such complexity. Tree-based meta-learners and deep learning models like TabNet and ANN can uncover nuanced, high-dimensional patterns, thereby improving classification accuracy for the more ambiguous classes. The stacking meta-model, in particular, excels due to its hybrid-based design that aggregates outputs from multiple learners, mitigating individual model weaknesses and offering a form of model-level error correction. Nonetheless, even the most advanced models are not immune to data-related limitations. The presence of feature overlap and class imbalance ensures that some degree of class-level divergence will persist. Therefore, while overall metrics like macro AUC provide a reassuring snapshot of model performance, they can mask deficiencies at the class level. These findings emphasize the value of meta-learning and mixture models in advancing reliable, multiclass ROC performance—so long as practitioners remain focused on evaluating and improving classification outcomes for the more challenging categories during real-world deployment.

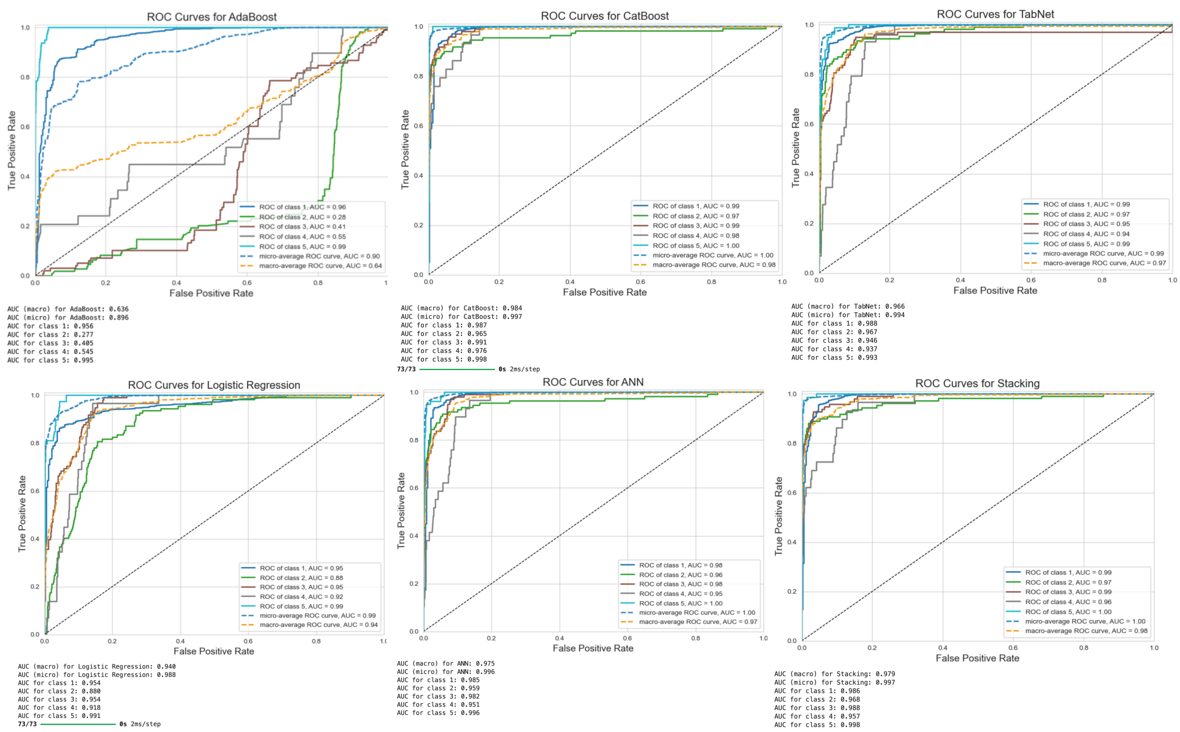
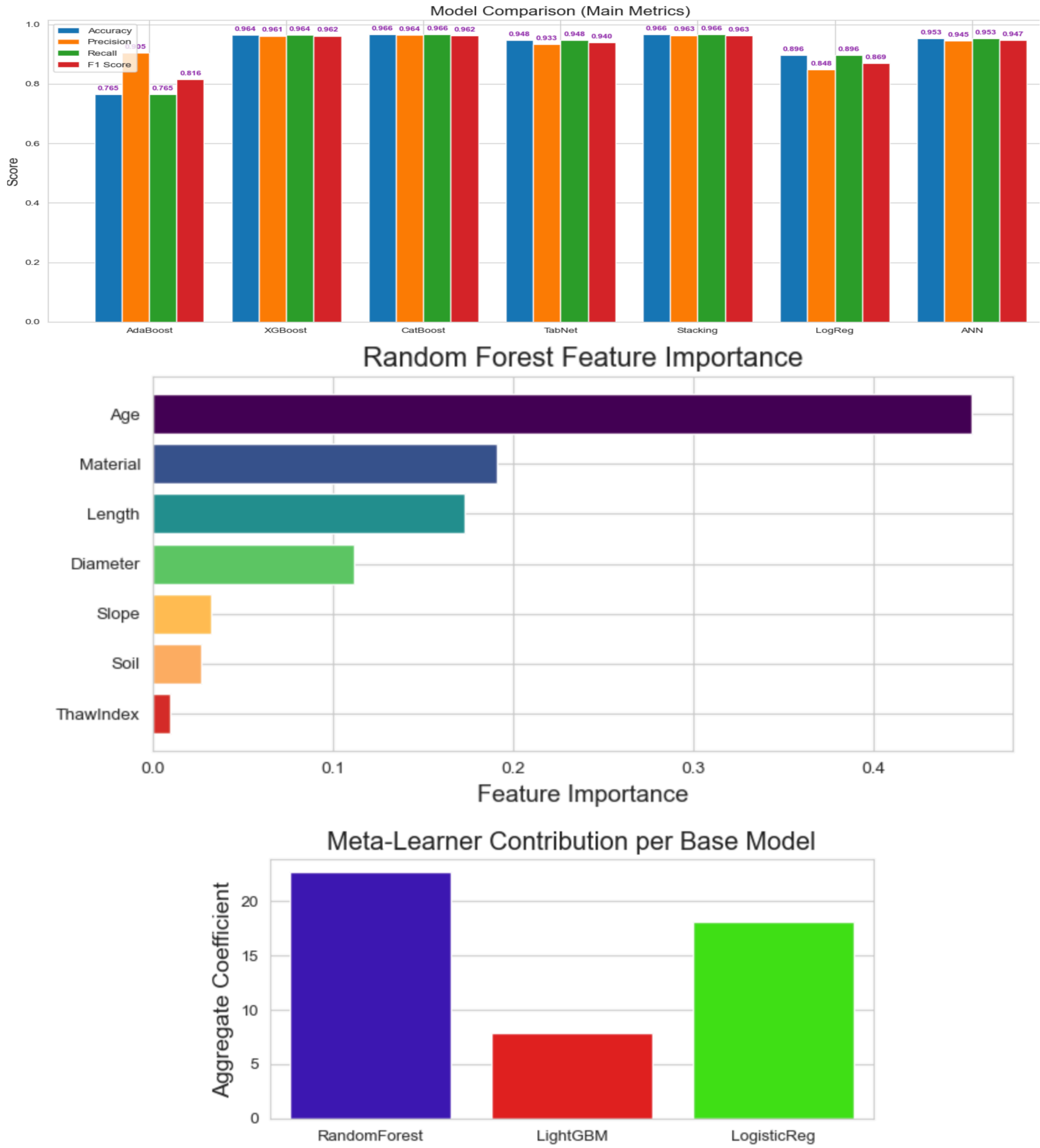


Figure 9. Stacking Meta-Model Delivers Superior ROC, Class Performance Varies Significantly

A comprehensive evaluation of predictive model performance for urban water pipe condition assessment demonstrates the potential of hybrid meta-learning strategies in complex real-world environments. As visualized in Figure 10, the bar plot identifies how different base learners contribute to the final decision-making process within a stacking meta-model, while the accompanying metric summary highlights the relative success of each major model tested. In the contribution chart, the Random Forest model exhibits the greatest influence on the stacking hybrid (tallest bar, blue), followed by Logistic Regression (green) and LightGBM (red). These contributions are quantified based on the absolute magnitude of the logistic regression coefficients used in the meta-learner, summing across all five multiclass targets. This aggregation reflects the extent to which the final model relies on each base learner's output, showing Random Forest as the most statistically influential component of the hybrid.

Beneath this bar plot, the summary performance table underscores the advantage of complex and hybrid models over simpler baselines. With a dataset consisting of 11,544 labeled pipe segments and seven engineered features (Diameter, Material, Age, Length, Thaw Index, Soil, Slope), models like CatBoost, XGBoost, and the stacking meta-learner all achieved high performance across accuracy, recall, and F1-score metrics—often near or above 0.96. For instance, CatBoost achieved an accuracy of 0.9658 and an F1-score of 0.9625, while the stacking meta-model slightly outperformed with 0.9658 and 0.9629, respectively. TabNet and neural networks (ANNs) also maintained strong results, with TabNet at 0.9476 accuracy and ANN at 0.9532. In sharp contrast, simpler models like Logistic Regression (Accuracy = 0.8965, F1 = 0.8691) and AdaBoost (Accuracy = 0.7653, F1 = 0.8159) performed considerably worse, illustrating the difficulty of modeling non-linear, imbalanced, and multi-class datasets with shallow learners. These differences largely arise from the inability of simpler models to capture complex interactions between features such as material composition, geometry, and environmental factors.

The stacking framework effectively mitigates the shortcomings of individual learners by leveraging their complementary strengths—tree-based models capture non-linear patterns, boosting methods resist overfitting, and linear models provide interpretability and generalization. Notably, the prominent influence of the Random Forest model within the stacking component correlates with its standalone classification success and frequent application in hybrid learning literature, underscoring its compatibility with this dataset. While not explicitly shown in Figure 10, broader analytical results repeatedly identify 'Age' as the most predictive feature, aligned with known deterioration trends in civil infrastructure. Additional features such as diameter, material, and length also contribute substantially to model decisions, reflecting domain-specific reliability indicators. These findings align well with previous studies: while traditional or stand-alone models can reveal broad performance trends, only hybrid aggregate and meta-learning-based approaches simultaneously deliver strong accuracy, resilience to class overlap, and meaningful interpretability. In this context, Figure 10 encapsulates the critical balance between model complexity, diversity, and practical application, highlighting the value of Random Forest and mixture meta-learning systems as powerful tools for transforming sensor, material, and operational data into actionable urban infrastructure insights




```
✅📊📝Final Model Metrics:
Logistic Regression:  Accuracy=0.8965, Precision=0.8483, Recall=0.8965, F1=0.8691
Linear Regression:    R²=0.5845, MSE=0.3255
CatBoost:               Accuracy=0.9658, Precision=0.9639, Recall=0.9658, F1=0.9625
AdaBoost:             Accuracy=0.7653, Precision=0.9050, Recall=0.7653, F1=0.8159
XGBoost:              Accuracy=0.9645, Precision=0.9615, Recall=0.9645, F1=0.9616
TabNet:               Accuracy=0.9476, Precision=0.9334, Recall=0.9476, F1=0.9398
ANN:                  Accuracy=0.9532
Stacking Meta-Model:  Accuracy=0.9658, Precision=0.9635, Recall=0.9658, F1=0.9629
```

Figure 10. Meta-Learner Model Results

The results demonstrate that the proposed hybrid machine learning meta-model represents a significant advancement in urban underground pipe condition assessment. The stacking-based hybrid architecture achieves superior performance

(96.6% accuracy) compared to individual algorithms by effectively integrating the strengths of tree-based methods (CatBoost, XGBoost, LightGBM), deep learning approaches (TabNet, ANN), and traditional statistical models. The Random Forest feature importance analysis reveals age as the dominant predictor, followed by material, length, and diameter. While other factors contribute smaller. Feature correlation and distribution [48] studies further validate the multi-dimensional nature of pipe degradation and reinforce the need for sophisticated machine learning approaches. The model's performance across all condition classes—confirmed through detailed confusion matrices and ROC curve analyses [49]—highlights its practical applicability for data-driven infrastructure management. These findings form a foundation for implementing intelligent maintenance strategies and optimizing resource allocation. The successful fusion of diverse machine learning [50] paradigms within the meta-model architecture marks a methodological contribution to the field. By combining complementary strengths from different algorithmic families, this mixture-based approach delivers both accuracy and interpretability—critical qualities for real-world deployment in urban infrastructure systems. Beyond water pipe assessment, this hybrid framework demonstrates strong potential for adaptation to other infrastructure evaluation challenges, reinforcing its value as a versatile and scalable data-driven decision-support tool.

5 DISCUSSION

This study evaluated that the proposed hybrid machine learning meta-model architecture achieves superior performance compared to individual algorithms and aligns with the progressive evolution observed in recent urban pipe assessment research. The stacking meta-model's accuracy of 96.6% significantly surpasses the performance reported in comparable studies, such as Mohammadagha et al. (2025), who achieved $R^2 = 0.9066$ with ANN models on reinforced concrete sewer pipes [32], and Mosavi et al. (2020) whose Random Forest model reached 86% accuracy for groundwater potential prediction [30]. This performance enhancement validates the effectiveness of integrating diverse algorithmic paradigms—tree-based methods (CatBoost, XGBoost, LightGBM), deep learning approaches (TabNet, ANN), and traditional statistical models—within a unified meta-learning framework. The consistent feature importance rankings across multiple algorithms, with age emerging as the dominant predictor (importance ~0.4), material type (~0.15), and environmental factors contributing secondary but meaningful influences, corroborate findings from Dawood et al. (2020) who reported similar hierarchical importance patterns in their comprehensive AI review [34]. However, unlike previous studies that focused on single algorithmic approaches or limited collective methods, this research demonstrates that systematic meta-model integration can overcome individual model limitations while maintaining interpretability. The ROC curve analyses reveal macro-averaged AUC values exceeding 0.95 for all condition classes, indicating robust discriminative capability across the multi-class classification task. Compared to the vision-based approaches reviewed by Rayhana et al. (2021), which achieved up to 98% accuracy but required massive image datasets (up to 2 million images) [31], the proposed framework achieves comparable performance using structured operational data with significantly reduced computational requirements.

6 CONCLUSION

This research successfully demonstrates the development and implementation of a hybrid machine learning meta-model framework for urban underground pipe condition assessment, achieving state-of-the-art predictive performance while addressing critical gaps in current assessment methodologies. The comprehensive analysis of 11,544 water pipe records reveals that the proposed stacking architecture, integrating Random Forest, LightGBM, CatBoost, TabNet, and Artificial Neural Networks through meta-learning, consistently outperforms individual algorithms with an accuracy of 96.6%, precision of 96.4%, and F1-score of 96.3%. The study establishes age as the primary deterioration predictor, followed by material type, length, and diameter, providing actionable insights for infrastructure management prioritization. The robust feature correlation analysis and multi-dimensional visualizations offer unprecedented transparency in understanding pipe deterioration patterns, enabling data-driven maintenance strategies. The integration of TabNet within the meta-model composite represents a methodological advancement, allowing enhanced feature interaction learning directly from tabular infrastructure data while maintaining interpretability through comprehensive feature importance analyses. The consistent performance across all condition classes, validated through confusion matrices and ROC curves with macro-averaged AUC values exceeding 0.95, demonstrates the framework's practical applicability.

7 RECOMMENDATION

Future research should explore the temporal dynamics of pipe deterioration by incorporating longitudinal data collection protocols, enabling the development of dynamic condition prediction models that account for different variations and long-term degradation trends. The successful integration of TabNet suggests opportunities for investigating other advanced machine learning architectures, including Transformer-based models, Vision Transformers (ViTs) [51], Graph Neural Networks (GNNs) [52], Deep Reinforcement Learning agents, Variational Autoencoders (VAEs) [53], Generative Adversarial Networks (GANs) [54], Bayesian Neural Networks, Multi-task Learning frameworks, Attention Mechanisms, Capsule Networks, Neural Architecture Search (NAS) [55], Federated Learning approaches, and Quantum Machine Learning [56] algorithms. Research institutions should develop real-time condition monitoring systems that integrate IoT sensors with the proposed meta-model framework.

8 REFERENCES

[1] K. F. Tee, L. R. Khan, H. P. Chen, and A. M. Alani, “Reliability based life cycle cost optimization for underground pipeline networks,” *Tunnelling and Underground Space Technology*, vol. 43, pp. 32–40, Jul. 2014, doi: 10.1016/J.TUST.2014.04.007.

[2] M. Najafi and S. B. Gokhale, “Trenchless technology : pipeline and utility design, construction, and renewal,” *McGraw-Hill*, 2022, Accessed: Jul. 19, 2025. [Online]. Available: https://cir.nii.ac.jp/crid/1130000794815910144

[3] C. D. F. Rogers *et al.*, “Condition assessment of the surface and buried infrastructure – A proposal for integration,” *Tunnelling and Underground Space Technology*, vol. 28, no. 1, pp. 202–211, Mar. 2012, doi: 10.1016/J.TUST.2011.10.012.

[4] T. Hao *et al.*, “Condition assessment of the buried utility service infrastructure,” *Tunnelling and Underground Space Technology*, vol. 28, no. 1, pp. 331–344, Mar. 2012, doi: 10.1016/J.TUST.2011.10.011.

[5] 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,” Apr. 2025, doi: 10.36227/TECHRXIV.174239957.76698574/V1.
[6] N. K. Chauhan and K. Singh, “A review on conventional machine learning vs deep learning,” *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018*, pp. 347–352, Mar. 2019, doi: 10.1109/GUCON.2018.8675097.
[7] E. Christodoulou, J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel, and B. Van Calster, “A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models,” *J Clin Epidemiol*, vol. 110, pp. 12–22, Jun. 2019, doi: 10.1016/J.JCLINEPI.2019.02.004.
[8] S. K. Sinha and P. W. Fieguth, “Neuro-fuzzy network for the classification of buried pipe defects,” *Autom Constr*, vol. 15, no. 1, pp. 73–83, Jan. 2006, doi: 10.1016/J.AUTCON.2005.02.005.
[9] M. R. Smith, T. Martinez, and C. Giraud-Carrier, “An instance level analysis of data complexity,” *Mach Learn*, vol. 95, no. 2, pp. 225–256, Nov. 2014, doi: 10.1007/S10994-013-5422-Z/TABLES/17.
[10] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, “Hybrid approaches to optimization and machine learning methods: a systematic literature review,” *Mach Learn*, vol. 113, no. 7, pp. 4055–4097, Jul. 2024, doi: 10.1007/S10994-023-06467-X/TABLES/3.
[11] M. Mohammadagha, “A Comprehensive Analysis of Classical Sorting Algorithms with Diverse Input Conditions ,” *Int. J. of Complexity in Applied Science and Technology*, 2025.
[12] P. Zhang, H. N. Wu, R. P. Chen, and T. H. T. Chan, “Hybrid meta-heuristic and machine learning algorithms for tunneling-induced settlement prediction: A comparative study,” *Tunnelling and Underground Space Technology*, vol. 99, p. 103383, May 2020, doi: 10.1016/J.TUST.2020.103383.
[13] B. Aslam, A. Maqsoom, A. H. Cheema, F. Ullah, A. Alharbi, and M. Imran, “Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach,” *IEEE Access*, vol. 10, pp. 119692–119705, 2022, doi: 10.1109/ACCESS.2022.3221430.
[14] M. Mohammadagha, “Hyperparameter Optimization Strategies for Tree-Based Machine Learning Models Prediction: A Comparative Study of AdaBoost, Decision Trees, and Random Forest,” Apr. 2025, doi: 10.31219/OSF.IO/XBKR5_V1.
[15] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, “Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines,” *Ore Geol Rev*, vol. 71, pp. 804–818, Dec. 2015, doi: 10.1016/J.OREGEOREV.2015.01.001.
[16] Q. Meng *et al.*, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” 2017.
[17] A. A. Ibrahim, R. L. Ridwan, M. M. Muhammed, R. O. Abdulaziz, and G. A. Saheed, “Comparison of the CatBoost Classifier with other Machine Learning Methods,” *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, 2020, Accessed: Jul. 19, 2025. [Online]. Available: www.ijacsa.thesai.org
[18] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, “Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial,” *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3039–3071, Oct. 2019, doi: 10.1109/COMST.2019.2926625.

[19] Y. Tang, J. Kurths, W. Lin, E. Ott, and L. Kocarev, “Introduction to Focus Issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics,” *Chaos*, vol. 30, no. 6, Jun. 2020, doi: 10.1063/5.0016505/1030253.

[20] M. Mohammadagha, M. Najafi, V. Kaushal, and A. Jibreen, “Cross-Domain Applications of Machine Learning: A Comparative Case Study from Iris Classification to Infrastructure Assessment,” p. 1, Jun. 2025, doi: 10.31224/4727.

[21] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, “Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics,” *Electronics 2021, Vol. 10, Page 593*, vol. 10, no. 5, p. 593, Mar. 2021, doi: 10.3390/ELECTRONICS10050593.

[22] K. O. Asare, Y. Terhorst, J. Vega, E. Peltonen, E. Lagerspetz, and D. Ferreira, “Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: Exploratory study,” *JMIR Mhealth Uhealth*, vol. 9, no. 7, p. e26540, Jul. 2021, doi: 10.2196/26540.

[23] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, “Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection,” *Journal of Structural Engineering*, vol. 146, no. 5, p. 04020073, May 2020, doi: 10.1061/(ASCE)ST.1943-541X.0002535/ASSET/0AF76EF2-44EE-4A82-8C73-A08F115F5D6D/ASSETS/IMAGES/LARGE/FIGURE4.JPG.

[24] M. Najafi and G. Kulandaivel, “Pipeline Condition Prediction Using Neural Network Models,” *Proceedings of the Pipeline Division Specialty Conference*, pp. 767–781, 2005, doi: 10.1061/40800(180)61.

[25] 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.

[26] M. Wang, H. Luo, and J. C. P. Cheng, “Towards an automated condition assessment framework of underground sewer pipes based on closed-circuit television (CCTV) images,” *Tunnelling and Underground Space Technology*, vol. 110, p. 103840, Apr. 2021, doi: 10.1016/J.TUST.2021.103840.

[27] R. Pollice *et al.*, “Data-Driven Strategies for Accelerated Materials Design,” *Acc Chem Res*, vol. 54, no. 4, pp. 849–860, Feb. 2021, doi: 10.1021/ACS.ACCOUNTS.0C00785/ASSET/IMAGES/LARGE/AR0C00785_0006.JPEG.

[28] F.-G. Yuan, S. A. Zargar, Q. Chen, and S. Wang, “Machine learning for structural health monitoring: challenges and opportunities,” *https://doi.org/10.1117/12.2561610*, vol. 11379, p. 1137903, Apr. 2020, doi: 10.1117/12.2561610.

[29] Z. Liu and Y. Kleiner, “State of the art review of inspection technologies for condition assessment of water pipes,” *Measurement*, vol. 46, no. 1, pp. 1–15, Jan. 2013, doi: 10.1016/J.MEASUREMENT.2012.05.032.

[30] 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.

[31] R. Rayhana, Y. Jiao, A. Zaji, and Z. Liu, “Automated Vision Systems for Condition Assessment of Sewer and Water Pipelines,” *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 4, pp. 1861–1878, Oct. 2021, doi: 10.1109/TASE.2020.3022402.

[32] M. Mohammadagha, M. Najafi, V. Kaushal, and 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, Accessed: Jul. 19, 2025. [Online]. Available: https://arxiv.org/pdf/2502.00363

[33] J. Latif, M. Z. Shakir, N. Edwards, M. Jaszczykowski, N. Ramzan, and V. Edwards, "Review on condition monitoring techniques for water pipelines," *Measurement*, vol. 193, p. 110895, Apr. 2022, doi: 10.1016/J.MEASUREMENT.2022.110895.

[34] T. Dawood, E. Elwakil, H. M. Novoa, and J. F. G. Delgado, "Artificial intelligence for the modeling of water pipes deterioration mechanisms," *Autom Constr*, vol. 120, p. 103398, Dec. 2020, doi: 10.1016/J.AUTCON.2020.103398.

[35] Y. Liu and Y. Bao, "Review on automated condition assessment of pipelines with machine learning," *Advanced Engineering Informatics*, vol. 53, p. 101687, Aug. 2022, doi: 10.1016/J.AEI.2022.101687.

[36] Y. Mao, Y. Li, F. Teng, A. K. S. Sabonchi, M. Azarafza, and M. Zhang, "Utilizing Hybrid Machine Learning and Soft Computing Techniques for Landslide Susceptibility Mapping in a Drainage Basin," *Water 2024, Vol. 16, Page 380*, vol. 16, no. 3, p. 380, Jan. 2024, doi: 10.3390/W16030380.

[37] I. Nunez, A. Marani, and M. L. Nehdi, "Mixture Optimization of Recycled Aggregate Concrete Using Hybrid Machine Learning Model," *Materials 2020, Vol. 13, Page 4331*, vol. 13, no. 19, p. 4331, Sep. 2020, doi: 10.3390/MA13194331.

[38] A. N. Beskopylny *et al.*, "Concrete Strength Prediction Using Machine Learning Methods CatBoost, k-Nearest Neighbors, Support Vector Regression," *Applied Sciences (Switzerland)*, vol. 12, no. 21, p. 10864, Nov. 2022, doi: 10.3390/APP122110864/S1.

[39] Mohsen Mohammadagha, Hajar Kazemi Naeini, Saeed Asadi, Mohammad Najafi, and Vinayak Kasuhal., "Machine Learning Model for Condition Assessment of Trenchless Vitrified Clay Pipes," in *North American Society for Trenchless Technology (NASTT) 2025 No-Dig Show, Mar 2025, Denver, United States.*, Mar. 2025. Accessed: Apr. 18, 2025. [Online]. Available: https://hal.science/hal-05019707v1

[40] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *Adv Neural Inf Process Syst*, vol. 30, 2017, Accessed: Jul. 19, 2025. [Online]. Available: https://github.com/Microsoft/LightGBM.

[41] K. Chauhan *et al.*, "Automated Machine Learning: The New Wave of Machine Learning," *2nd International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2020 - Conference Proceedings*, pp. 205–212, Mar. 2020, doi: 10.1109/ICIMIA48430.2020.9074859.

[42] 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.

[43] T. Wang, K. Zhang, J. Thé, and H. Yu, "Accurate prediction of band gap of materials using stacking machine learning model," *Comput Mater Sci*, vol. 201, p. 110899, Jan. 2022, doi: 10.1016/J.COMMATSCI.2021.110899.

[44] M. Mohammadagha, "Hyperparameter Optimization Strategies for Tree-Based Machine Learning Models Prediction: A Comparative Study of AdaBoost, Decision Trees, and Random Forest," *SSRN Electronic Journal*, Apr. 2025, doi: 10.2139/SSRN.5226457.

[45] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *International Journal of Informatics and*

Information Systems, vol. 4, no. 1, pp. 13–20, Mar. 2021, Accessed: Jul. 11, 2025. [Online]. Available: https://ijiis.org/index.php/IJIIS/article/view/73

[46] 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.

[47] A. Zien, N. Krämer, S. Sonnenburg, and G. Rätsch, “The Feature Importance Ranking Measure,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5782 LNAI, no. PART 2, pp. 694–709, 2009, doi: 10.1007/978-3-642-04174-7_45.

[48] M. A. Hall, “Correlation-based feature selection for machine learning,” 1999. Accessed: Jul. 19, 2025. [Online]. Available: https://hdl.handle.net/10289/15043

[49] A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognit*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997, doi: 10.1016/S0031-3203(96)00142-2.

[50] M. Mohammadagha, I. Tshitenge, and I. Adebambo, “Multimodal Sentiment Analysis: Fusion of Emojis, Images, and Text in State-of-the-Art Machine Learning Techniques for Social Media Using Deep Learning,” *Int J Med Net*, 2025.

[51] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in Vision: A Survey,” *ACM Comput Surv*, vol. 54, no. 10, Jan. 2022, doi: 10.1145/3505244/ASSET/97278908-BA90-453A-B2F3-54B479FEAEE8/ASSETS/IMAGES/LARGE/CSUR-2021-0095-F02.JPG.

[52] S. Munikoti, D. Agarwal, L. Das, M. Halappanavar, and B. Natarajan, “Challenges and Opportunities in Deep Reinforcement Learning With Graph Neural Networks: A Comprehensive Review of Algorithms and Applications,” *IEEE Trans Neural Netw Learn Syst*, vol. 35, no. 11, pp. 15051–15071, 2024, doi: 10.1109/TNNLS.2023.3283523.

[53] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li, and Y. Xu, “Autoencoders and their applications in machine learning: a survey,” *Artificial Intelligence Review 2024 57:2*, vol. 57, no. 2, pp. 1–52, Feb. 2024, doi: 10.1007/S10462-023-10662-6.

[54] J. Cheng, Y. Yang, X. Tang, N. Xiong, Y. Zhang, and F. Lei, “Generative Adversarial Networks: A Literature Review,” *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 12, pp. 4625–4647, Dec. 2020, doi: 10.3837/TIIS.2020.12.001.

[55] Y. Liu, Y. Sun, B. Xue, M. Zhang, G. G. Yen, and K. C. Tan, “A Survey on Evolutionary Neural Architecture Search,” *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 2, pp. 550–570, Feb. 2023, doi: 10.1109/TNNLS.2021.3100554.

[56] L. Banchi, J. Pereira, and S. Pirandola, “Generalization in Quantum Machine Learning: A Quantum Information Standpoint,” *PRX Quantum*, vol. 2, no. 4, p. 040321, Dec. 2021, doi: 10.1103/PRXQUANTUM.2.040321/FIGURES/8/MEDIUM.