Erciyes Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 41, Sayı 2, 2025

Erciyes University
Journal of Institue Of Science and Technology
Volume 41, Issue 2, 2025

Artificial Intelligence-Driven Risk Assessment in Civil Engineering Projects: A Comparative Analysis of Machine Learning Models

Anıl UTKU[*1]

[*1] Munzur Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği, TUNCELİ

(Alınış / Received: 04.04.2025, Kabul / Accepted: 25.05.2025, Online Yayınlanma / Published Online: 30.08.2025)

Keywords
Risk prediction,
Construction management,
Artificial intelligence,
Machine learning,
Risk assessment model

Abstract: Successful completion of civil engineering projects involves complex interactions among various components such as managing costs, maintaining schedules, ensuring structural stability, and minimizing environmental impacts. Accurate risk prediction at the project level is important to support efficient resource planning and prevent potential delays or disruptions. Since traditional risk assessment methods based on expert judgments involve lengthy and subjective procedures, the development of data-driven and automated estimation systems is at the forefront. In this study, a comparative analysis of various machine learning methods such as RF, SVM, XGBoost, CatBoost, and LGBM for project risk assessments in civil engineering is presented. The dataset used in the study includes information on project type, cost change records and planning, and real-time history, as well as environmental data and safety documentation, and structural health measurements. Precision, accuracy, sensitivity, and F-score metrics were used to evaluate the models along with ROC-AUC measurements. Experimental studies showed that CatBoost and LGBM outperform the compared models with 95.5% accuracy.

İnşaat Mühendisliği Projelerinde Yapay Zekâ Odaklı Risk Değerlendirmesi: Makine Öğrenmesi Modellerinin Karşılaştırmalı Bir Analizi

Anahtar Kelimeler
Risk tahmini,
İnşaat yönetimi,
Yapay zekâ,
Makine öğrenmesi,
Risk değerlendirme modeli

Öz: İnşaat mühendisliği projelerinin başarılı bir şekilde tamamlanması, maliyetleri yönetmeye, programları sürdürmeye, yapısal kararlılığı sağlamaya ve çevresel etkileri en aza indirme gibi çeşitli bileşenler arasında karmaşık etkileşimleri içerir. Risk tahminlerinin proje düzeyinde doğru bir şekilde yapılması, verimli kaynak planlamasını desteklemek ve olası gecikmeleri veya kesintileri önlemek için önemlidir. Uzman değerlendirmelerine dayanan geleneksel risk değerlendirme yöntemleri, uzun ve öznel prosedürler içerdiği için veri odaklı ve otomatik tahmin sistemlerinin geliştirilmesi ön plana çıkmaktadır. Bu çalışmada, inşaat mühendisliğindeki proje risk değerlendirmelerine yönelik RF, SVM, XGBoost, CatBoost ve LGBM gibi çeşitli makine öğrenmesi yöntemlerinin karşılaştırmalı bir analizi sunulmuştur. Çalışmada kullanılan veri seti, proje türü hakkında bilgi, maliyet değişim kayıtları ve planlama ile gerçek süre geçmişinin yanı sıra çevresel veriler ve güvenlik belgeleriyle yapısal sağlık ölçümlerini içermektedir. Modellerin değerlendirilmesi amacıyla ROC-AUC ölçümleriyle birlikte kesinlik, doğruluk, duyarlılık ve F-skor metrikleri kullanılmıştır. Deneysel çalışmalar, CatBoost ve LGBM'nin %95,5'lik bir doğrulukla karşılaştırılan modellerden daha başarılı olduğunu göstermiştir.

*İlgili Yazar, email: anilutku@munzur.edu.tr

1. Introduction

Civil engineering development projects need significant financial backing and long-term planning sequences [1]. Different controlling variables including cost management and schedule planning and structural safety and environmental factors need to be controlled in these processes [2]. Success in project completion hinges on proper risk identification alongside professional risk management practices. The safety and sustainability standards of all projects depend heavily on risk management practices. Traditionally employed risk assessments use expert opinion as a basis but deliver ambiguous results because they do not effectively analyze changes in variables [3]. Artificial intelligence technologies enable civil engineering projects to obtain risk assessments that are data-driven and objective-based. AI-based models utilize vast datasets to predict the probability of every project component taking longer than expected alongside their potential to suffer damage or failure of security systems [4]. AI models support automatic learning abilities to enhance prediction accuracy since they learn and optimize prediction methods with each new data input. AI techniques have entered widespread use in civil engineering risk analysis activities since the last few years. Project processes receive full evaluations through financial analysis and assessment of structural elements and environmental aspects to enable decision-making according to [5]. Project managers achieve large projects benefits when they bring together BIM and IoT data using machine learning approaches [6].

This study provides an evaluation of multiple machine learning techniques that determine construction project risk levels. The dataset used consists of structural health status reports and environmental factors, as well as detailed information on financials along with resource allocation and safety protocols for construction projects. The main motivation of this study is to create an automated system that can perform precise risk assessments for civil engineering projects. The study aimed to evaluate machine learning methods in selecting the most suitable prediction model suitable for construction project applications. The data-driven method will enable project managers and decision makers to reduce future risks through their decisions.

The unique contribution of this study is the comprehensive comparison of advanced machine learning models such as RF, SVM, XGBoost, LGBM and CatBoost on real-world data (with BIM integration) for multi-class risk level prediction in civil engineering projects. Unlike previous studies, this study focuses on multi-class problem rather than binary classification and uses robust validation techniques such as Grid Search and cross-validation by comparing models under the same conditions. Furthermore, comprehensive analysis of the interpretability, efficiency and discriminative power of each model is presented, providing practical guidance to decision makers for risk management in construction projects.

In order to position the present study within the existing literature, the next section provides a comprehensive overview of relevant previous works in the domain of AI-based risk assessment in civil engineering.

2. Related Works

The implementation of AI methods in civil engineering during recent times revolutionized how engineers evaluate risks and direct their project activities. Research studies have examined how different ML algorithms can be used to forecast delays and safety and cost-related problems for construction projects. This section evaluates contemporary predictive models that analyze construction sector risks and perform building sector analytics by using relevant datasets and methodological approaches. The research examines proven methods while uncovering knowledge deficits and establishes the present work's location relative to AI-based decision systems for civil engineering projects.

Chattapadhyay et al. created a new dataset to assess the risks in mega construction projects [7]. A total of 63 risk factors were identified based on four basic project management criteria: cost, time, quality, and scope. These risk factors were assessed using a Likert scale (1–5), and the obtained data were analyzed using K-means clustering and Genetic Algorithm (GA) assisted K-means. Experimental results showed that the GA-K-means algorithm provides higher clustering accuracy compared to standard K-means.

Anysz et al. used DT, Artificial Neural Networks (ANN), and association rules to support litigation decisions in contractual disputes experienced by a contractor in construction projects [8]. Data from 10 completed projects of a European construction company were used. The data included variables such as cost, profit, reserve, project type (design-build / build), duration, delays, additional work costs, whether litigation was filed, and whether the decision favored the contractor. Based on this small dataset, a simulation dataset of 100 projects was created. Experiments showed that the best accuracy rate of ANN was 93.3%.

Sanni-Anibire et al. presented a comparative analysis of K-Nearest Neighbor (KNN), ANN, SVM, and Ensemble methods for predicting delay risks in risky construction projects [9]. In the study, 36 delay risk factors were determined, and 48 responses were obtained from the questionnaire developed to evaluate the impact of these factors and converted into a dataset. Experimental results showed that ANN exhibited the best prediction performance with 93.75% classification accuracy.

Fitzsimmons et al. proposed a hybrid machine learning approach for time risk analysis in construction projects. Three hundred-two project files from 2007–2019 in the UK were used [10]. The project files extracted features such as description texts for each task planned and actual duration, weather, economic conditions, and schedule quality. Gaussian Mixture Model (GMM) was used to segment short-text task descriptions into topics. Empirical Bayesian Network (EBN) was used to model task duration deviation probabilities. SVM was used to estimate task duration deviations. Monte Carlo Simulation (MCS) was used to simulate the distribution of final project duration risk. Experiments showed that the proposed hybrid model achieved 54.4% higher accuracy than MCS.

Uddin et al. presented a comparative evaluation of SVM, RF, XGBoost, LGBM, CatBoost, and ANN to estimate the risk level of civil engineering projects [11]. The dataset comprises project characteristics, cost drivers, delay reasons, and workflow processes. Experiments showed that RF outperformed the compared models with 85% accuracy using SelectFromModel feature selection.

Ashtari et al. developed a Bayesian Network (BN) Classifier model to estimate and evaluate the cost overrun risks in construction projects. They compared it with Naïve Bayes (NB) and Decision Tree (DT) [12]. The dataset was created based on the survey responses provided by 48 experts. Thirty-six delay risk factors were determined, and a machine learning application was performed on the probabilities and consequences of these factors. The experiments showed that BN outperformed the compared models with 78.86% accuracy and 0.89 AUC.

Alshboul et al. used a hybrid mathematical model of LGBM and XGBoost to evaluate the impact of external support on the cost of sustainable construction projects [13]. The study collected data from 3578 green building projects in the North American region. The impact of external investment on construction costs was evaluated with macroeconomic modeling. The experiments showed that LGBM outperformed XGBoost with 0.985 R^2.

The reviewed literature shows that data-driven approaches are increasingly being adopted to predict and manage construction project risks. Although various models such as ANN, SVM, RF, and hybrid techniques have yielded successful results, their effectiveness largely depends on dataset quality, feature selection, and hyper-parameter optimization. Many existing studies are based on limited or domain-specific datasets, and their generalizability is limited.

Building on the insights from previous research, the following section outlines the methodology adopted in this study, including data characteristics and the applied machine learning models.

3. Material and Method

In this section, the dataset, data preprocessing steps, and applied machine learning models used to predict the risk level of civil engineering projects are explained in detail. First, the content of the dataset used in the study and its variables and data sources are introduced. Then, data preprocessing techniques such as missing data analysis, data transformations, and processing of categorical variables are discussed. The study applies powerful classification models such as SVM, RF, XGBoost, CatBoost, and LGBM to evaluate the performance of different machine-learning algorithms. The training of the models, hyper-parameter optimization, and evaluation metrics are detailed, and performance comparisons are made using accuracy, precision, recall, F-score, and ROC-AUC analyses.

3.1. Dataset

This study used a dataset created with the integration of BIM and AI techniques to manage and optimize the life cycle of civil engineering projects. The dataset used is publicly available via https://www.kaggle.com/datasets/ziya07/bim-ai-integrated-dataset/data. The dataset comprises project data, including cost management, scheduling, structural health monitoring, environmental condition assessment, resource allocation, safety risk analysis, and drone-based monitoring systems. The dataset consists of 1000 rows of data and 28 attributes related to the planning, execution, and control processes of construction projects. Project metadata includes general project information such as project identity, type, geographical location, and timeline.

Financial data provides detailed information about planned and actual costs, cost overrun amounts, and budget deviations. Time management data includes data for determining planned and actual duration information and possible schedule deviations related to the project.
Structural health monitoring includes data for maintaining structural integrity by monitoring critical engineering parameters such as vibration levels, crack width, and load-bearing capacity of civil structures. Environmental factors refer to environmental variables such as temperature, humidity, air quality, and weather conditions that affect the effectiveness of the construction process and structural safety. Resource management and safety risks refer to data such as material usage, labor hours, equipment usage, and accident records. Drone-based monitoring data allows the evaluation of construction project progress and completion percentage by analyzing drone images and provides anomaly detection. The target variable, risk level, refers to the project risk analysis variable classified as low, medium, and high-risk levels based on cost, schedule, safety, and structural health data.

The dataset used in this study consists of data with various features to predict risk levels in civil engineering projects. Various pre-processing steps were applied to the dataset to increase the model's performance and obtain reliable predictions. The data pre-processing process includes steps such as data loading, missing data analysis, determining variable types, feature engineering, and separating data into training and test sets. It was determined that some variables did not significantly contribute to direct prediction. Therefore, Project_ID, Start_Date, End_Date, and Location variables were removed from the dataset. Missing data is an important factor affecting the accuracy of the model. In particular, missing values were detected in the Risk_Level column, which is the target variable, and these rows were removed from the dataset so as not to affect the accuracy of the model. Categorical variables were converted to numerical values using the One-Hot Encoding method.

The distribution of risk levels is shown in Figure 1.

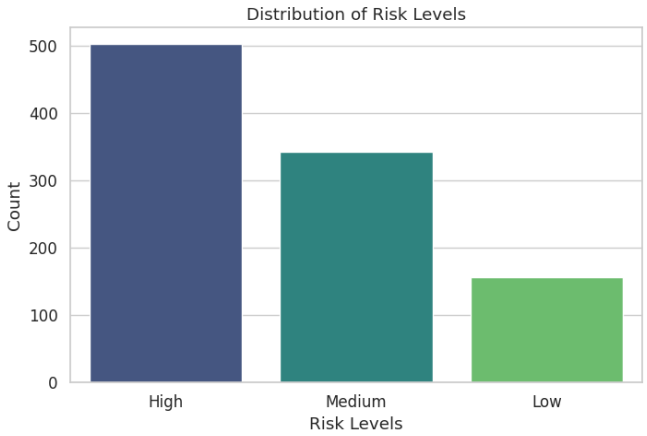

Figure 1. The distribution of risk levels

As seen in Figure 1, the high level has 502 samples, the medium level has 342 samples, and the low level has 156 samples. It is observed that the majority of the projects are classified as high risk, which indicates a critical need for accurate risk prediction models. High is the most common risk level in the dataset and indicates that a significant portion of the projects examined are high risk. Projects may have high-cost overruns, delays, structural problems, or safety risks. Medium is the second largest class. Medium-risk projects may have some risk factors but not be critical. Low is the class with the fewest observations and indicates that only a small portion of projects are considered low risk. Low-risk projects may be well-managed projects that progress according to plan.

The distribution of project types is shown in Figure 2.

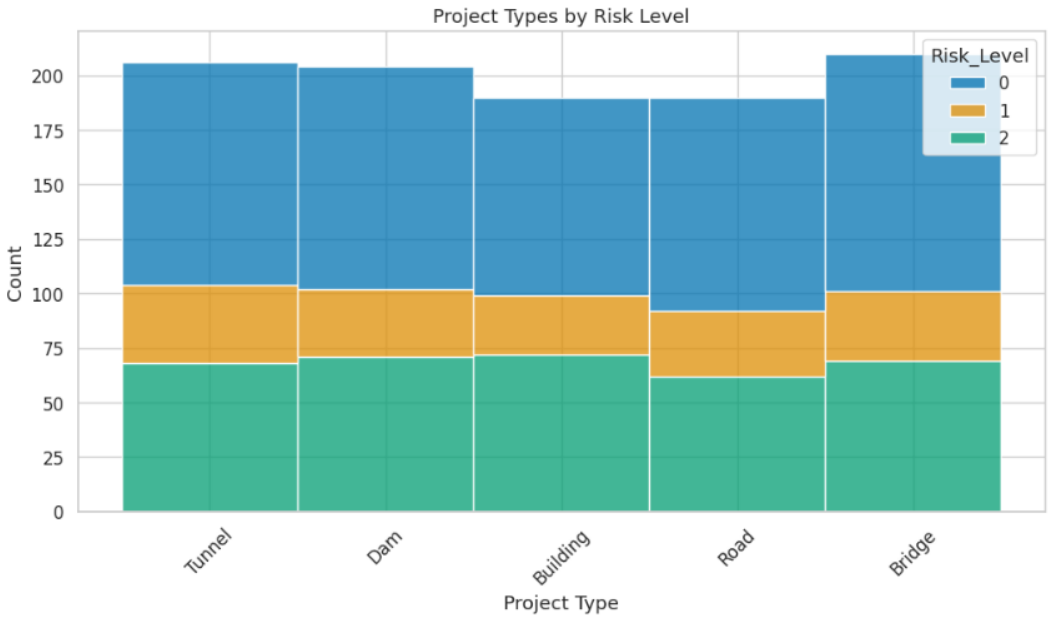

Figure 2. The distribution of project types

As seen in Figure 2, the bridge project has 210 examples: the tunnel project 206, the roof project 204, the building project 190, and the road project 190. As seen in Figure 2, risk management is critical in infrastructure projects. Bridge and tunnel projects constitute a significant portion of the dataset, suggesting that high-risk infrastructure projects are predominant. Since the high-risk rate is higher in large engineering projects such as bridges, tunnels, and dams, such projects require detailed analysis and planning. Low-risk projects are fewer, showing that completely risk-free projects are impossible in the construction sector. Risk analyses of projects vary according to project type. For example, building projects may be more controlled, while bridge and tunnel projects contain higher risk levels due to their more complex structures.

Figure 3 shows the heat map of the correlation, showing the relationship between the numerical variables in the dataset. The correlation of each variable with other variables is expressed by color intensity. Red represents positive correlation; blue represents negative or low correlation.

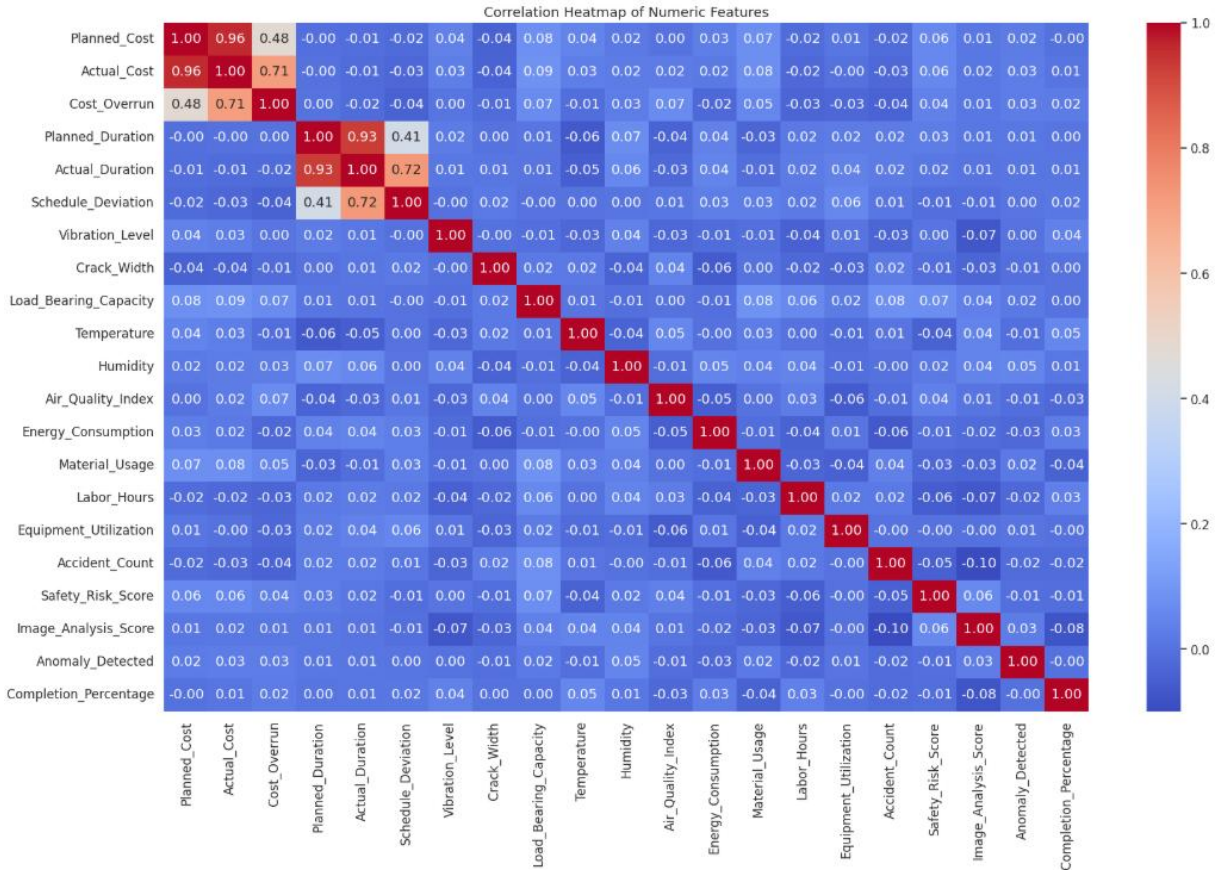

Figure 3. Correlation heatmap of numeric features

As seen in Figure 3, there are strong relationships between planned and actual values. There is a strong correlation between planned and actual values, while environmental variables show weaker relationships with risk indicators. Cost overrun and time deviation show a moderate correlation. It is seen that delays, although they may create additional costs, do not have a direct and strong relationship. Environmental factors show a low correlation with project duration and cost variables. Some structural health variables, such as load-carrying capacity, are not directly related to project cost and schedule.

3.2. Classification models

RF is an ensemble learning algorithm with many decision trees [14]. Each tree is trained with a randomly sampled subset of the training data, and decisions are made by a majority vote of these trees [15]. This structure can effectively model interactions between variables while reducing overfitting. RF provides stable and reliable results, especially on complex, multivariate datasets, but is limited in interpretability [16]. Compared to boosting-based models like XGBoost and CatBoost, RF offers more interpretability and robustness but falls short in terms of precision and accuracy, especially in imbalanced and complex datasets.

SVM is a supervised learning algorithm that aims to find a hyperplane that provides maximum separation between classes [17]. SVM exhibits strong performance, especially in small and high-dimensional datasets. SVM can create non-linear decision boundaries with kernel functions for non-linearly separable data [18]. However, computational costs may increase in multi-class problems and large datasets, and hyper-parameter adjustments are critical for accurate results. While SVM performs reasonably well on smaller and high-dimensional datasets, it is generally outperformed by ensemble tree-based models in terms of overall accuracy and handling of multiclass classification tasks in large datasets.

XGBoost is an optimized and regularized ensemble learning method using gradient-boosting decision trees [19]. It includes both L1 and L2 regularization to prevent overfitting and can effectively deal with missing data [20]. Thanks to its parallel processing capabilities, XGBoost provides fast and highly accurate results on large datasets. It is especially successful in analyzing feature importance and is frequently preferred in competitions [21]. XGBoost strikes a good balance between accuracy and efficiency and performs slightly below CatBoost and LGBM, making it a strong yet slightly less dominant choice among boosting algorithms.

LGBM is a gradient-boosting decision tree algorithm that works fast and efficiently on large datasets [22]. It preserves model accuracy while reducing training time, especially with Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) techniques [23]. Thanks to its deep tree structure, LGBM can learn complex data patterns. LGBM stands out in large-scale projects with its low memory usage and high-speed advantages [24]. LGBM exhibits a highly competitive performance and matches CatBoost in most metrics, especially excelling in large-scale datasets with limited computational resources.

CatBoost is a gradient-boosting tree-based machine-learning algorithm that works especially well with categorical variables [25]. Thanks to the Categorical Boosting technique, CatBoost processes categorical data directly, reducing preprocessing and preventing information loss [26]. CatBoost also includes unique regularization mechanisms that prevent overfitting without creating a ranking difference [27]. CatBoost offers high accuracy, especially in complex data sets, while attracting attention with its ease of use and automatic feature processing. CatBoost demonstrates superior overall accuracy and ease of use, particularly in handling categorical variables, slightly outperforming other boosting models such as XGBoost and LGBM in this study.

After detailing the dataset and model structures, the next section presents the experimental results and performance comparisons of the applied machine learning models.

4. The Experimental Results

This study comprehensively compared RF, SVM, XGBoost, LGBM, and CatBoost algorithms for risk-level classification in construction projects. The evaluation metrics used to compare the models are accuracy, precision, recall, F-score and ROC (Receiver Operating Characteristic), and AUC (Area Under the Curve), which comprehensively assess the classification performance. The results are analyzed in detail, and the reasons behind the performance differences among the models are discussed. This section aims to highlight the effectiveness of different algorithms in predicting project risk levels and provide insights into the most suitable approach for risk assessment in civil engineering projects. The hyper-parameters of each model were optimized using Grid Search to provide the best performance, as seen in Table 1.

Table 1. The hyper-parameters of each model

Model	Hyper-parameters
RF	max_depth: None, min_samples_leaf: 1, min_samples_split: 10, n_estimators: 200
SVM	C=10, kernel='rbf', gamma='auto'
XGBoost	n_estimators=200, learning_rate=0.1, max_depth=5, subsample=0.9, colsample_bytree=0.9, gamma=0.1, reg_alpha=0.01, reg_lambda=1
LGBM	n_estimators=300, learning_rate=0.05, max_depth=6, subsample=0.8, colsample_bytree=0.8, reg_alpha=0.1, reg_lambda=1
CatBoost	iterations=500, learning_rate=0.05, depth=6, l2_leaf_reg=3

Grid Search is a systematic method used to optimize hyper-parameter in machine learning models. It determines the parameter set that gives the best result by trying different hyper-parameter combinations to maximize the model's performance. First, the hyper-parameters are to be optimized, and the values they can take are determined. Grid Search trains and tests the model for each specified hyper-parameter combination. 80% of the dataset was used for training, 20% for testing, and 10% of the training data was used for validation. Models were tested by performing 5-fold cross-validation.

Figure 4 shows the confusion matrix of RF.

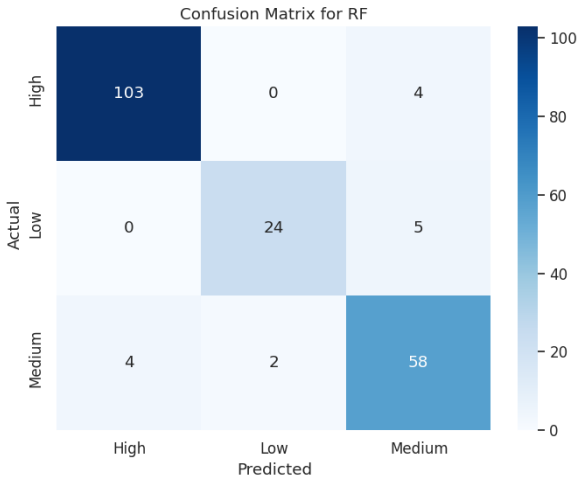

Figure 4. The confusion matrix of RF

As seen in Figure 4, RF correctly classified 185 out of 200 samples and incorrectly classified 15. As seen in Figure 4, RF has a successful performance, especially in the High and Medium classes, but there are misclassifications in the Low class. RF has difficulty distinguishing low-risk level projects. RF model performs well in identifying high and medium risk levels but has difficulty distinguishing the low-risk class. Table 2 shows the experimental results of RF.

Table 2. The experimental results of RF

Class	Accuracy	Precision	Recall	F-score
High	0.925	0.963	0.963	0.963
Low		0.923	0.828	0.873
Medium		0.866	0.906	0.885
Macro average		0.917	0.899	0.907
Weighted average		0.926	0.925	0.925

As seen in Table 2, the overall accuracy rate of RF is 92.5%. The precision, recall, and F-score values of 0.963 for the high-risk level indicate that the model successfully classifies high-risk projects and rarely makes false predictions. The lower recall value for the low-risk level indicates that RF incorrectly predicts some low-risk projects and can especially mix them with other classes. The high recall value for the medium-risk level indicates that RF predicts medium-risk projects well. However, the relatively lower precision value indicates that there may be some false positive predictions. Figure 5 shows the confusion matrix of SVM.

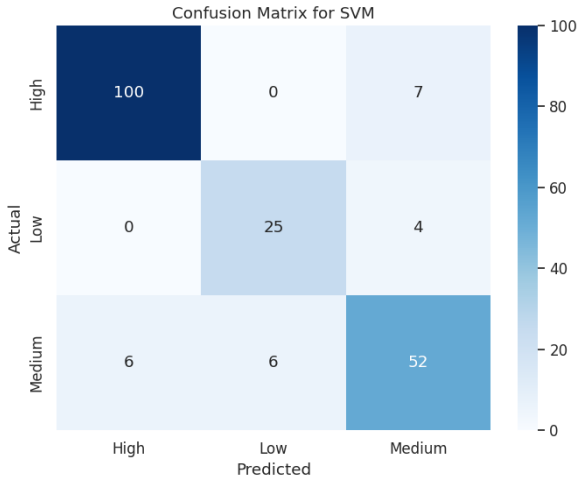

Figure 5. The confusion matrix of SVM

As seen in Figure 5, SVM correctly classified 177 out of 200 samples and incorrectly classified 23. SVM predicts the High-risk level quite successfully, with low errors. The low-risk level is also generally successful, but some samples are classified as Medium. The most incorrect classifications were made in the Medium class. Table 3 shows the experimental results of SVM.

Table 3. The experimental results of SVM

Class	Accuracy	Precision	Recall	F-score
High	0.885	0.943	0.935	0.939
Low		0.806	0.862	0.833
Medium		0.825	0.812	0.819
Macro average		0.858	0.870	0.864
Weighted average		0.886	0.885	0.885

As seen in Table 3, the overall accuracy of SVM is 88.5%. SVM is relatively successful at the High-risk level. At the Low-risk level, SVM's recall value is at a successful level of 86.2%. However, since SVM predicts other classes as Low, its precision value is lower. At the medium risk level, the model's success is lower than that of other risk levels.

Figure 6 shows the confusion matrix of XGBoost.

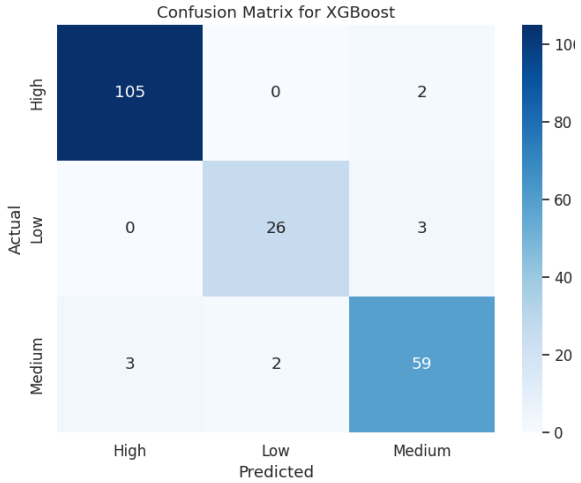

Figure 6. The confusion matrix of XGBoost

As seen in Figure 6, XGBoost correctly classified 190 out of 200 samples and incorrectly classified 10. XGBoost is quite successful in the High-risk class. Generally, correct predictions were made in the Low-risk class, but some projects were incorrectly classified as Medium risk. XGBoost predicted some medium-risk projects to be high or low in the medium-risk class.

Table 4 shows the experimental results of XGBoost.

Table 4. The experimental results of XGBoost

Class	Accuracy	Precision	Recall	F-score
High	0.950	0.972	0.981	0.977
Low		0.929	0.897	0.912
Medium		0.922	0.922	0.922
Macro average		0.941	0.933	0.937
Weighted average		0.950	0.950	0.950

As seen in Table 4, XGBoost's overall accuracy is 95%, and very high precision, recall, and F-score values were obtained in all classes. It performed particularly well in the High-risk class, with 97% precision and 98% recall. Over 90% of the results were also obtained at low and medium risk levels. XGBoost generally shows a balanced and strong prediction performance in all classes.

Figure 7 shows the confusion matrix of CatBoost.

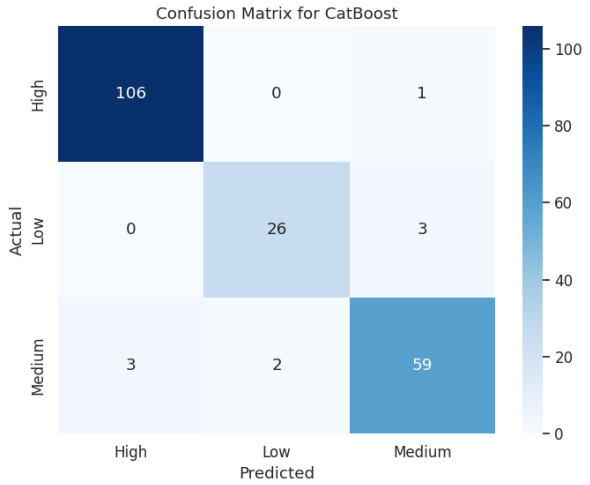

Figure 7. The confusion matrix of CatBoost

As seen in Figure 7, CatBoost correctly classified 191 out of 200 samples and incorrectly classified 9. CatBoost classified the High-risk level with the highest accuracy rate, with 106 correct predictions. The error rate is also relatively low at the Low and Medium risk levels. The low number of errors, especially at the low-risk level, shows that CatBoost's overall predictive power is relatively high. CatBoost generally predicted the risk levels in a balanced and successful manner. Table 5 shows the experimental results of CatBoost.

Table 5. The experimental results of CatBoost

Class	Accuracy	Precision	Recall	F-score
High	0.955	0.972	0.991	0.981
Low		0.929	0.897	0.912
Medium		0.937	0.922	0.929
Macro average		0.946	0.936	0.941
Weighted average		0.955	0.955	0.955

As seen in Table 5, the overall accuracy of CatBoost is relatively high at 95.5%, and the F-score values for all classes are balanced. The high precision and recall values for the High class show that CatBoost successfully detected the High class. In general, the model successfully predicted all classes and presented one of the best results in terms of classification performance.

Figure 8 shows the confusion matrix of LGBM.

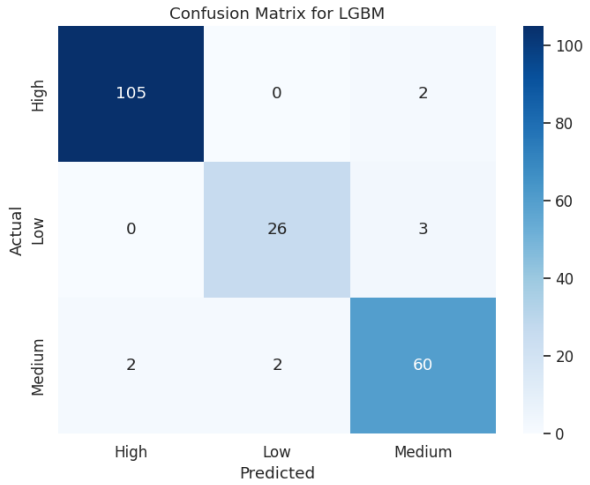

Figure 8. The confusion matrix of LGBM

As seen in Figure 8, LGBM correctly classified 191 out of 200 samples and incorrectly classified 9. LGBM's very accurate prediction of the High-risk level shows that LGBM is quite successful in recognizing high-risk projects. There are 26 correct predictions in the Low class and three incorrect predictions. It has an acceptable error rate in low-risk projects. There are 60 correct predictions and two incorrect classifications in the Medium class, which shows that it can also successfully distinguish medium risk levels. LightGBM generally performs well in distinguishing risk levels with a low error rate.

Table 6 shows the experimental results of LGBM.

Table 6. The experimental results of LGBM

Class	Accuracy	Precision	Recall	F-score
High	0.955	0.981	0.981	0.981
Low		0.929	0.897	0.912
Medium		0.923	0.938	0.930
Macro average		0.944	0.938	0.941
Weighted average		0.955	0.955	0.955

LGBM's overall 95.5% accuracy shows that it performs successful classification for all classes. For the High class, precision, recall, and F-score values are at very high levels, with 98.1%, indicating that LGBM is highly successful in distinguishing high-risk projects. For the Low and Medium risk levels, F-score values of 91.2% and 93.0%, respectively, indicate that LGBM successfully detects low and medium risk levels. In general, LGBM exhibits a consistent and balanced performance, contributing to the high-accuracy prediction of risk levels.

Table 7 and Figure 9 show comparative experimental results according to weighted average.

Table 7. Comparative experimental results according to weighted average

Model	Accuracy	Precision	Recall	F-score
SVM	0.885	0.886	0.885	0.885
RF	0.925	0.926	0.925	0.925
XGBoost	0.950	0.950	0.950	0.950
CatBoost	0.955	0.955	0.955	0.955
LGBM	0.955	0.955	0.955	0.955

Table 7 and Figure 9 compare the weighted average accuracy, precision, recall and F-score performances of SVM, RF, XGBoost, CatBoost and LGBM models.

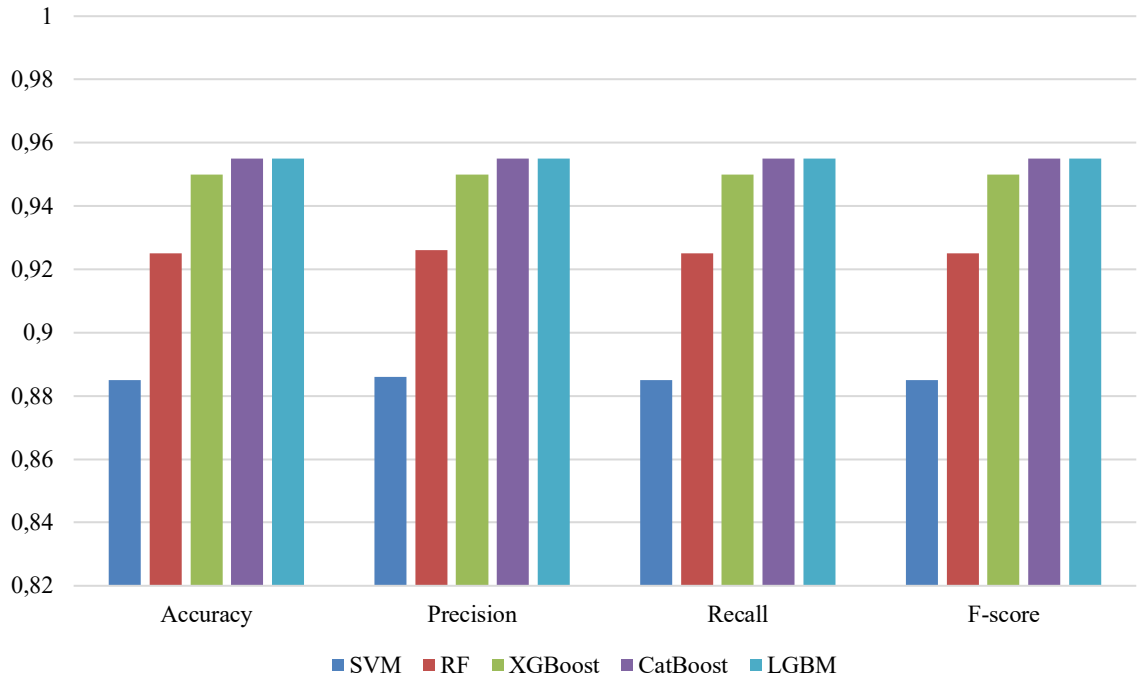

Figure 9. Comparative experimental results according to weighted average

LGBM and CatBoost achieved 95.5% accuracy, 95.5% precision, 95.5% recall and 95.5% F1-score. XGBoost ranks third with 95% accuracy. XGBoost showed a performance close to CatBoost and LGBM. RF is behind XGBoost with 92.5% accuracy. Similarly, RF's precision, recall, and F1-score values are 92.5%. The lowest-performing model is SVM. SVM has the lowest success rate with 88.5% accuracy and is the last in all other metrics. Figure 10 shows the ROC curves of the compared models. CatBoost and LGBM outperform other models across all evaluation metrics, indicating their robustness in risk classification tasks.

The ROC curves seen in Figure 10 compare the classification performance of the machine learning algorithms in risk-level classification. The higher the AUC value, the better the model's classification performance. RF, XGBoost,

CatBoost, and LGBM have AUC values above 0.99 for all classes and exhibit successful classification performance. It is seen that the models perform classification with high accuracy with a low false positive rate. On the other hand, SVM has a lower AUC value of 0.723, especially for the Medium class compared to other models, indicating that SVM has difficulty detecting the Medium class. However, SVM's AUC values are high for the High and Low classes. LGBM, XGBoost, and CatBoost had the highest AUC values in all classes and exhibited the best performance. LGBM and CatBoost strengthen weak classifiers using boosting-based learning and minimize error rates. The gradient boosting mechanism works well, especially with categorical and large-scale datasets. CatBoost can process categorical data naturally and has a regularization mechanism that prevents overfitting. LGBM processes large datasets quickly and provides very high accuracy thanks to its deep tree structure. It is especially effective in using feature importance (feature importance ranking) and ensures efficient model operation by eliminating unnecessary features. CatBoost, LGBM, and XGBoost achieved near-perfect ROC curves with AUC values exceeding 0.99 across all risk levels. This indicates that these models can distinguish between the risk classes with high confidence and minimal false positives. In particular, CatBoost's AUC curve is slightly more stable and closer to the top-left corner, suggesting its superior discriminative ability. LGBM also demonstrated a highly consistent ROC curve, reflecting its strong generalization capability on the dataset. XGBoost, while slightly behind, still offers highly competitive classification performance.

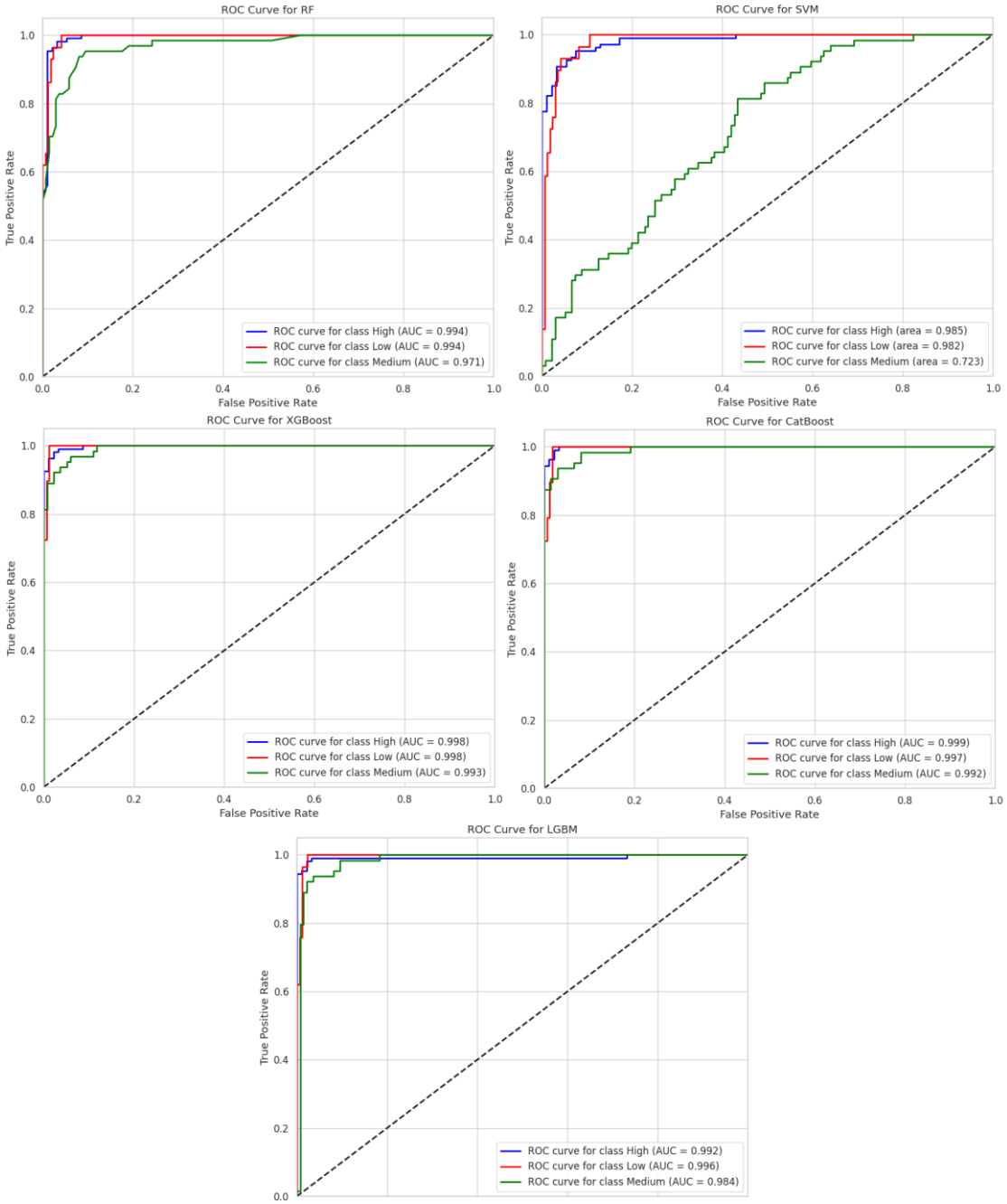

Figure 10. The ROC curves of the compared models

Using the modified gradient boosting algorithm, XGBoost is more potent than RF and SVM. XGBoost has a regularization mechanism to prevent overfitting. It effectively analyzes feature importance and makes it easier to understand which features the model gives more weight. XGBoost minimizes error rates using deeper trees and better optimization techniques than the RF model. Since RF works by creating independent trees, it cannot thoroughly learn decision boundaries. However, boosting models such as XGBoost, CatBoost, and LGBM have more powerful learning mechanisms. The computational cost increases when more trees are used, but boosting algorithms make this process more efficient. SVM can influence high-dimensional and non-linear datasets, but it does not offer the advantages that tree-based models provide on this dataset. While tree-based models learn better the relationship between variables in the dataset, SVM falls short in non-linear classifications.

Based on the findings obtained through extensive experimentation, the conclusion section summarizes the outcomes and provides future directions for enhancing AI-based risk prediction systems.

5. Conclusions

Risk management in civil engineering projects is important in ensuring project continuity and optimizing costs. Accurate prediction of risk levels can be used as a decision support mechanism in critical processes such as resource management, scheduling and budget control, and increasing project safety. This study aimed to predict the risk levels of construction projects using different machine learning methods and compared the performances of various models. Experimental results showed that CatBoost, XGBoost, and LGBM models accurately predict the risk level. Although RF and SVM models also showed relatively successful performance, it was observed that tree-based deep learning algorithms such as XGBoost, CatBoost, and LGBM provided higher success in risk classification tasks. In evaluating the models, comprehensive analyses were performed using performance evaluation metrics such as accuracy, precision, recall, F-score, and ROC-AUC. However, one of the limitations of this study is the imbalance in the class distribution, which may have affected the prediction performance, particularly for the minority (low-risk) class. Future studies can benefit from applying resampling techniques or class weighting strategies to improve model robustness across all risk levels.

This study aims to contribute to data-driven decision-making processes in construction projects and demonstrates the effectiveness of AI-based risk prediction models. In addition, the analyses performed on this unique dataset, which has not been previously addressed in academic studies, contribute to the literature. Future studies aim to make improvements such as using deep learning-based models, investigating different hyper-parameter optimization strategies, and enhancing the generalization ability of the model with larger datasets. In this way, more sensitive and reliable risk prediction systems can be developed, and more efficient management of civil engineering projects can be achieved.

References

[1] Lovallo, D., Cristofaro, M., Flyvbjerg, B. 2023. Governing large projects: a three-stage process to get it right. Academy of Management Perspectives, 37(2), 138-156.

[2] Parsamehr, M., Perera, U. S., Dodanwala, T. C., Perera, P., Ruparathna, R. 2023. A review of construction management challenges and BIM-based solutions: perspectives from the schedule, cost, quality, and safety management. Asian Journal of Civil Engineering, 24(1), 353-389.

[3] Khodabakhshian, A., Puolitaival, T., Kestle, L. 2023. Deterministic and probabilistic risk management approaches in construction projects: A systematic literature review and comparative analysis. Buildings, 13(5), 1312.

[4] Akeiber, H. J. 2025. Artificial Intelligence in Engineering Management: Revolutionizing Decision-Making and Automation. Al-Rafidain Journal of Engineering Sciences, 317-349.

[5] Afzal, F., Yunfei, S., Nazir, M., Bhatti, S. M. 2021. A review of artificial intelligence based risk assessment methods for capturing complexity-risk interdependencies: Cost overrun in construction projects. International Journal of Managing Projects in Business, 14(2), 300-328.

[6] Chen, Y., Wang, X., Liu, Z., Cui, J., Osmani, M., Demian, P. 2023. Exploring building information modeling (bim) and internet of things (IoT) integration for sustainable building. Buildings, 13(2), 288.

[7] Banerjee Chattapadhyay, D., Putta, J., Rao P, R. M. 2021. Risk identification, assessments, and prediction for mega construction projects: A risk prediction paradigm based on cross analytical-machine learning model. Buildings, 11(4), 172.

[8] Anysz, H., Apollo, M., Grzyl, B. 2021. Quantitative risk assessment in construction disputes based on machine learning tools. Symmetry, 13(5), 744.

[9] Sanni-Anibire, M. O., Zin, R. M., Olatunji, S. O. 2022. Machine learning model for delay risk assessment in tall building projects. International Journal of Construction Management, 22(11), 2134-2143.

[10] Fitzsimmons, J. P., Lu, R., Hong, Y., Brilakis, I. 2022. Construction Schedule Risk Analysis-a Hybrid Machine Learning Approach, Journal of Information Technology in Construction (ITcon), 27, 70-93.

[11] Uddin, S., Ong, S., Lu, H. 2022. Machine learning in project analytics: a data-driven framework and case study. Scientific Reports, 12(1), 15252.

[12] Ashtari, M. A., Ansari, R., Hassannayebi, E., Jeong, J. 2022. Cost overrun risk assessment and prediction in construction projects: A Bayesian network classifier approach. Buildings, 12(10), 1660.

[13] Alshboul, O., Shehadeh, A., Almasabha, G., Mamlook, R. E. A., Almuflih, A. S. 2022. Evaluating the impact of external support on green building construction cost: A hybrid mathematical and machine learning prediction approach. Buildings, 12(8), 1256.

[14] Ghiasi, M. M., Zendehboudi, S. 2021. Application of decision tree-based ensemble learning in the classification of breast cancer. Computers in biology and medicine, 128, 104089.

[15] Salman, H. A., Kalakech, A., Steiti, A. 2024. Random forest algorithm overview. Babylonian Journal of Machine Learning, 2024, 69-79.

[16] Eskandari, H., Saadatmand, H., Ramzan, M., Mousapour, M. 2024. Innovative framework for accurate and transparent forecasting of energy consumption: A fusion of feature selection and interpretable machine learning. Applied Energy, 366, 123314.

[17] Abdullah, D. M., Abdulazeez, A. M. 2021. Machine learning applications based on SVM classification a review. Qubahan Academic Journal, 1(2), 81-90.

[18] Hamzah, M. A., Othman, S. H. 2021. A review of support vector machine-based intrusion detection system for wireless sensor network with different kernel functions. International Journal of Innovative Computing, 11(1), 59-67.

[19] Natras, R., Soja, B., Schmidt, M. 2022. Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting. Remote Sensing, 14(15), 3547.

[20] Asselman, A., Khaldi, M., Aammou, S. 2023. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), 3360-3379.

[21] Mehdary, A., Chehri, A., Jakimi, A., Saadane, R. 2024. Hyperparameter optimization with genetic algorithms and xgboost: A step forward in smart grid fraud detection. sensors, 24(4), 1230.

[22] Cai, W., Wei, R., Xu, L., Ding, X. 2022. A method for modelling greenhouse temperature using gradient boost decision tree. Information Processing in Agriculture, 9(3), 343-354.

[23] Abu-Salih, B., Alotaibi, S., Abukhurma, R., Almiani, M., Aljaafari, M. 2024. DAO-LGBM: dual annealing optimization with light gradient boosting machine for advocates prediction in online customer engagement. Cluster Computing, 27(4), 5047-5073.

[24] Kedam, N., Tiwari, D. K., Kumar, V., Khedher, K. M., Salem, M. A. 2024. River stream flow prediction through advanced machine learning models for enhanced accuracy. Results in Engineering, 22, 102215.

[25] Nwafor, E. O., Akintayo, F. O. 2024. Predicting Trip Purposes of Households in Makurdi Using Machine Learning: A Comparative Analysis of Decision Tree, CatBoost, and XGBoost Algorithms. Engineering Applications, 3(3), 260-274.

[26] Alsulamy, S. 2025. Predicting construction delay risks in Saudi Arabian projects: A comparative analysis of CatBoost, XGBoost, and LGBM. Expert Systems with Applications, 268, 126268.

[27] Zhang, S., Lu, X., Lu, Z. 2023. Improved CNN-based CatBoost model for license plate remote sensing image classification. Signal Processing, 213, 109196.