Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

AI-driven anomaly detection for fraud prevention in project monitoring

Mehmet Tahir SANDIKKAYA[1*], Onur Behzat TOKDEMİR[2, 3]

[1] Istanbul Technical University, Department of Computer Engineering, sandikkaya@itu.edu.tr, Orcid No: 0000-0002-9756-603X
[2] Istanbul Technical University, Department of Civil Engineering, otokdemir@itu.edu.tr, Orcid No: 0000-0002-4101-8560
[3] Istanbul Technical University, AI Center, otokdemir@itu.edu.tr, Orcid No: 0000-0002-4101-8560

ARTICLE INFO

Article history:

Received 13 March 2025
Received in revised form 26 October 2025
Accepted 29 October 2025
Available online 30 December 2025

Keywords:

Cybersecurity, Anomaly Detection, Data Poisoning Attacks, Fraudulent Reporting Detection, Insider Threat Detection, Project Management

Doi: : 10.24012/dumf.1656802

* Corresponding author

ABSTRACT

Accurate tracking of project progress is crucial for timely delivery, cost control, and fraud prevention. Issues in progress reporting, whether due to real mistakes, employee inefficiencies, or internal threats, present considerable risks to major projects. This study aims to examine statistical and machine learning techniques to identify data inconsistencies, fraudulent reporting, and other anomalies in project tracking. Utilizing a dataset of 118 weekly snapshots, including genuine and tainted data, this research assesses the effectiveness of the interquartile range, isolation forest, and an ensemble approach in detecting anomalies. The results underscore the strengths and weaknesses of statistical and machine learning models while proposing an optimal detection framework for effective project management.

Introduction

Large-scale construction projects require thorough coordination from several workflows involving numerous subcontractors and vast data reports [1–3]. Proper progress monitoring is important, as both accidental and intentional discrepancies as well as weak reporting, require workarounds and adjustments that lead to delays, increase likelihood of going over budget, and cause conflicts over contracts. Traditional project management tools like Critical Path Method (CPM) and Program Evaluation and Review Technique (PERT) are effective for scheduling; however, they lack real-time fraud detection and anomaly identification capabilities. The benefits of anomaly detection are often ignored in between numerous tasks to be accomplished by limited technical man-hours. This study examines the use of statistical approaches and machine learning algorithms for anomaly detection in construction project management and their use for decision support systems.

Machine learning (ML) methods have been increasingly used for construction progress monitoring and control [4, 5]. The construction industry can benefit if anomalies can be detected and addressed early in large projects and tight timeframes. The construction industry suffers from anomalies like demurrage, breaches, misappropriation of resources, and unintentional data breaches. Making conscious performance decisions via ML methods in projects make project stakeholders more risk-free.

Traditional methods for monitoring construction progress rely on manual inspections, analysis of past data, and auxiliary rule-based systems. These approaches have varying degrees of success, but generally lag behind the complexity and variability associated with mega construction projects. These widely applied techniques include CPM and PERT. Hybrid techniques are more adaptive in construction [6] while predictive features are mostly missing. However, these techniques are often lengthy, error-prone and cannot handle uncertainties. Moreover, classical methods might overlook hints, which, could be an early indicator of the emergence of anomalous behavior [7].

ML, therefore, is promising in monitoring progress in construction by providing automated, data-driven solutions. Supervised and unsupervised learning algorithms have a long history of successfully identifying anomalies in data. Anomaly detection can be more accurate if supervised models such as decision trees, random forests, and SVMs on labeled data are utilized. They are trained on past data to learn normal behavior, and then can identify deviations from the learned behaviors as anomalies [8, 9]. Some studies that utilize random forest have been used to create dashboards to predict the construction delay, and their recall values reached above 90% in some cases [9]. However, accessing project-specific data history where the number of normal instances is large enough to correctly identify anomalies is far from a practical reality.

Algorithms like autoencoders and K-means clustering are used under unsupervised learning, a standard approach to locating outliers. Autoencoders have attracted attention for investigating complex and high-dimensional data in construction. They are able to recover when they encounter small but significant outliers, i.e., anomalous or rare events in the data set featuring an increased activity in the transaction rate in the project, in addition to the addition of non-viable project timelines to the data set [10]. Another example is the one where K-means clustering is used to find out the major risk factors that contribute to the delay of construction processes, which aids the risk-management process [11].

Many other works propose case studies of using ML for fault detection to enhance construction progress tracking. For example, the use of autoencoder neural networks in identifying irregular investment trends and project durations in the Kuwait Construction Market demonstrates the deviations from typical market activities that lead to additional cost [10]. The predictive analysis utilizing ML algorithms has achieved wide acceptance for designating potential VOs and issuing early alerts such that construction delays can be avoided [9]. Such applications demonstrate the merits of ML in streamlining operations and defining more effective temporary organizational structures.

The integration of ML in construction progress monitoring provides numerous benefits. Especially in construction projects where early detection of anomalies can save time and reduce costs, ML can process extensive datasets quickly and accurately, identifying anomalies and patterns that may be missed by traditional techniques [7]. In addition, ML models can learn iteratively, which means that ML models can adapt their learning in response to changes during the project, therefore increasing their predictive accuracy over time [12]. Predicting delays and cost overages with machine learning models [13] allows project managers and stakeholders to prevent and react to possible complaints early on. It is an advantage in complex infrastructural projects where it is critical to avoid any delays. This might greatly impact the timings and project costs [14]. Machine Learning also enables effective resource allocation and scheduling, which ensures the end project delivery on time while sticking to the budget [7].

The existing anomaly detection methods cannot directly embrace the natural progress of the stages of a construction project. From a project management perspective, each stage is tied either loosely or tightly to previous stages; the progress of a stage has a warm-up, peak, and tailing time, and the progress might get stuck occasionally. To be able to handle these, the data from progress reports are preprocessed to reveal if the sane and natural expectations of any progress (e.g., the amount of reinforced concrete and reinforcement used have to be proportional to the design documents) are met.

This study evaluates the ability of statistical and ML-based techniques to identify anomalies in construction progress tracking. This research uses 118 weekly dataset snapshots with artificially introduced anomalies to mimic fraudulent reporting and insider threats. IQR, isolation forest, and an ensemble method were compared to evaluate their performance. This reinforces the conclusion that proposed approaches outperform standard methods in terms of accuracy and ease of adaptation and provide a more robust framework for anomaly detection. Such insights also shed light on the merits of integrating these ideas into project management software solutions, in a manner to promote visibility and reduce exposure to data manipulation risks that can plague large construction projects.

The examination is conducted systematically, incorporating data collection, preprocessing, anomaly detection, and evaluation (Figure 1). By utilizing this structured framework, we can methodically identify inconsistencies that might be used to qualitatively compare traditional statistical analysis methods with more advanced ML-based methods.

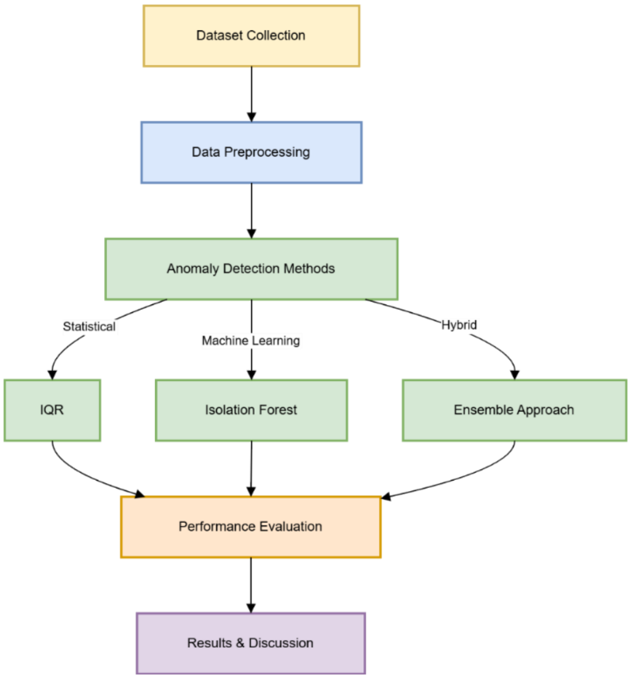

Figure 1 Methodology of the study

Related Work

The construction industry is adopting the digital transformation, emphasizing Industry 4.0, and is increasingly critical in implementing cybersecurity measures against insider threats or intrusions on construction sites and mega projects. However, this shift has incorporated digital technologies, exposing construction sites to risks. Comprehensive cybersecurity spans multiple strategies and technologies. A key aspect of this approach is creating a cyber risk assessment framework leveraging machine learning methodologies, enabling continuous monitoring and updates for scenario parameters to undertake proactive risk mitigation tailored to industry-relevant specifications [15]. Moreover, the collaboration of cyber-physical systems (CPSs) improves the resilience and authenticity of the smart building initiatives, helping with real-time data gathering, tracking, and appropriate risk assessment and decision-making [16]. RFID-based systems for intrusion detection enable real-time monitoring and alerts of unauthorized entry into restricted areas, enhancing safety and efficiency on construction sites [17].

In addition, the high precision of potential hazard identification and safety improvement of computer vision and 4D Building Information Modeling (BIM)-based spatio-temporal analysis are applied to intrusion detection in hazardous areas [18]. The construction industry can also learn from critical infrastructure protection efforts and use well-established processes to promote better cybersecurity practices [19]. In addition, data security and privacy management can safeguard sensitive data against unauthorized access and attacks, maximizing risk assessment and resource allocation [20]. Using a vulnerability assessment framework based on the common vulnerability scoring system (CVSS) to identify and address vulnerabilities in autonomous site monitoring systems before deploying cyber-physical systems on-site enables decision-makers to prioritize critical metrics [21]. These steps help ensure a safer construction site while minimizing digitization risk, making buildings safer and more efficient.

Few machine learning techniques have acted as catalysts for anomaly detection to the forefront of cybersecurity, i.e., isolation forest, one-class SVM, autoencoders, Gaussian mixture model (GMM), and interquartile range (IQR). As an unsupervised learning approach, isolation forest identifies steep changes tied to cyberattacks with high accuracy and few false positives in the smart grid cybersecurity domain [22]. It has also demonstrated strong web traffic anomaly detection results with high precision and recall [23, 24]. One-class SVM is another unsupervised model that can help detect smooth changes and general clustering, but is less effective for catching abrupt anomalous events [22].

These deep learning techniques through autoencoders enable handling very complex and multidimensional data, thereby creating a stable framework to manage the anomaly detection process in dynamic environments like cybersecurity, removing false positives, and significantly improving the detection process [25, 26]. Statistical GMM-based outlier detection is used by modeling the typical behavior of the system that could serve as a pattern to detect the cyber threats [25]. A well-known statistical method for outlier identification is the Interquartile Range, which gives the spread of the middle 50% of data. But it struggles in more complex, high-dimensional cybersecurity data scenarios.

Integrating these models into cybersecurity paradigms has proven to enhance detection rates, reduce false positives, accelerate response times, and improve the overall efficiency of the framework in comparison to legacy rule-based systems [26]. When appropriately tailored to specific new cybersecurity applications, they can provide robust and tunable protection against sophisticated intrusions, showing the importance of model selection and tuning in practical use cases [27, 28].

Helping to prevent fake reporting and discrepancies in data, project delays, workforce allocation, and false progress reporting in large-scale construction projects using advanced technologies and methodologies. The Kuwait construction market was analyzed by autoencoder neural networks, which provide a generic approach in that they train historical data to learn common patterns, so that deviations from normal patterns can be detected that could indicate inefficiency or fraudulent behavior [10].

Another study proposes rule-based forward chaining-based techniques and states that the techniques have improved the systems for project monitoring, thereby optimizing fraud detection and control mechanisms, which were left unaddressed mainly by traditional project management software, which utilized planning and scheduling to gain control over the project [29]. As such, Computer Vision and Internet of Things (IoT) systems for automated construction progress monitoring present a potential route for real-time data collection and processing to reduce divergences between anticipated and actual progress and avoid delays and budget overruns [30, 31].

Another study [32] argues that advancements in statistical pattern recognition techniques allow for the classification of the construction activity monitoring model based on multiple scenarios that may arise. This is particularly crucial given the apparent information technology usage gap in the construction industry, as information technology has shown promising potential in its application to progress measurement, with potential for automation of progress information generation, making it timelier with respect to decision making and mitigation of risk in fraud, and ultimately enhancing project control [33]. Incorporating these new analytical frameworks and technology can create a better strategy, improve efficiency, and detect fraud in mega construction projects.

Relevant studies specific to financial fraud detection in the construction domain include many perspectives. Most of the time, these studies focus on embracing the root of the fraudulent activity, then preventing it. A study that incorporates a questionnaire to model the influence factors on external financial auditors, so that more accurate and independent audits could be possible [34]. Another study, which also conducts a questionnaire with the experts, focuses on the protection of the project data [35]. The study compiles best practices for security against external threats that could be conducted in the organization. Another study, which expands fraudulent data manipulation beyond the construction domain, utilizes Bayesian Nash Equilibrium to predict better strategies to protect data [36]. The study considers not only outsiders but also misuse or insider attacks. In contrast to those, the proposed approach focuses on the detection of the fraudulent reporting of project data.

Methodology and Dataset

A large-scale construction project dataset was analyzed over 193 weeks, with reliable data available from the 62nd week onward. At the beginning of the project, the project suffered from budget-related issues and could not progress as planned. Therefore, the dataset consists of 118 healthy weekly snapshots, covering planned, acquired, and realized progress metrics for over 350 tasks. These tasks were categorized into six major groups: concrete works, electrical works, finishing works, façade, mechanical works, and insulation. Additionally, the total progress is included.

Planned progress stands for the expected completion percentage based on the project schedule, acquired progress stands for the self-reported completion percentage provided by subcontractors or project managers, and realized progress stands for the validated completion percentage after independent verification.

The six categories of tasks include:

- Concrete Works: Foundations, structural elements, and reinforced concrete components.
- Electrical Works: Power distribution, panel installations, and wiring.
- Finishing Works: Interior detailing, painting, tiling, and carpentry.
- Façade: Cladding, external insulation, and waterproofing.
- Mechanical Work: Heating, ventilation, and air conditioning (HVAC) systems, plumbing, and mechanical piping.
- Insulation: Thermal and acoustic insulation layers.

Even though the original dataset includes higher resolution data, progress tracking at weekly intervals was selected to tradeoff between data granularity and computation load. Daily tracking would result in too much noise that could prevent meaningful outlier detection, while monthly tracking might average out abnormal patterns. So that, it could be too late to prevent project delays. By processing the data points weekly, we can quickly compare the progress of our anomalies while maintaining a helpful anomaly detection framework. The proposed methods are applied to the cleaned and preprocessed dataset. This preprocessing step was required to increase the robustness of statistical and ML-based anomaly detection. Figure 2 plots the total acquired value per week in each major group to sketch a sample from the dataset.

As there were no recorded fraudulent activities in the dataset, a simulation approach is taken: a practical, real-world, adversarial, data poisoning attack. The poisoned data is injected into the dataset to assess the robustness of relevant anomaly detection. This attack consisted of seven manipulated snapshots, each of which was carefully designed to insert anomalies that would only be latent in a typical project management system. Six of these snapshots were manipulated so that only a single task group was affected, which simulated slightly deceitful acts or system-wide inconsistencies that could occur due to purposeful falsification or human error during data entry. The seventh snapshot, however, was constructed to be more aggressive, changing all task groups at once to mimic massive reporting fraud or a systemic failure in project monitoring.

For visual comparison, the same data in Figure 2 after insider modification is sketched in Figure 3. One can notice that the difference is faint and could hardly be noticed by the naked eye.

This approach indicates that the statistical and ML models could differentiate between the actual project variance and the artifacts added afterwards. The situations in practice at a construction site can cause anomalies such as fraudulent reporting, mismanagement, or even cyber threats. This study aims to achieve an integrated evaluation of the practical and useful anomaly detection methods by integrating localized data manipulations. The proposed method is evaluated whether the automated systems can identify unusual behaviors by determining whether the variations are irrelevant or indicative of a fraud.

Statistical approach: interquartile range (IQR)

The interquartile range is a statistical method used to detect anomalies by identifying values that deviate significantly from the mean of a distribution. IQR detects outliers in numerical data. Therefore, it is appropriate to identify progress deviations in construction reporting. The IQR works as follows:

The first and third quartiles ($Q1, Q3$) are computed. They represent 25th and 75th percentiles of the values, where they are the lower and upper boundaries of normal values. Then, the IQR is calculated based on the difference between the first and third quartiles as $IQR = Q3 - Q1$. Any value outside the range is determined as outliers after the boundaries are defined as follows:

$$Lower\ Bound = Q1 - (k \times IQR)$$

$$Upper\ Bound = Q3 + (k \times IQR)$$

k is a scaling factor that sets the sensitivity of outlier detection. Even though the typical value of $k = 1.5$, the natural fluctuations of construction projects require a higher value. Therefore, $k = 4.8$ is selected based on the construction projects' erraticism.

The IQR method was selected as a baseline anomaly detection technique due to its simplicity and interpretability. IQR provides transparent anomaly detection without requiring training data. Additionally, IQR is effective when the data is skewed. Construction progress data often contains nonlinear distributions, making mean-based methods unreliable. IQR is not sensitive to extreme values. Moreover, it is non-parametric and can detect significant deviations without requiring extensive computation.

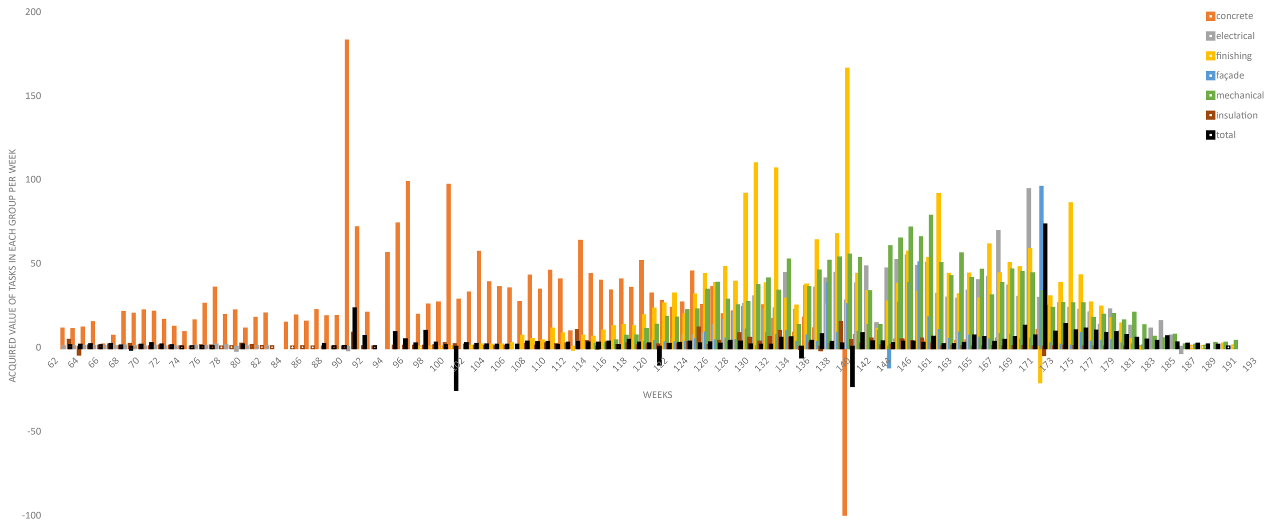

Figure 2 Actual acquired values per week

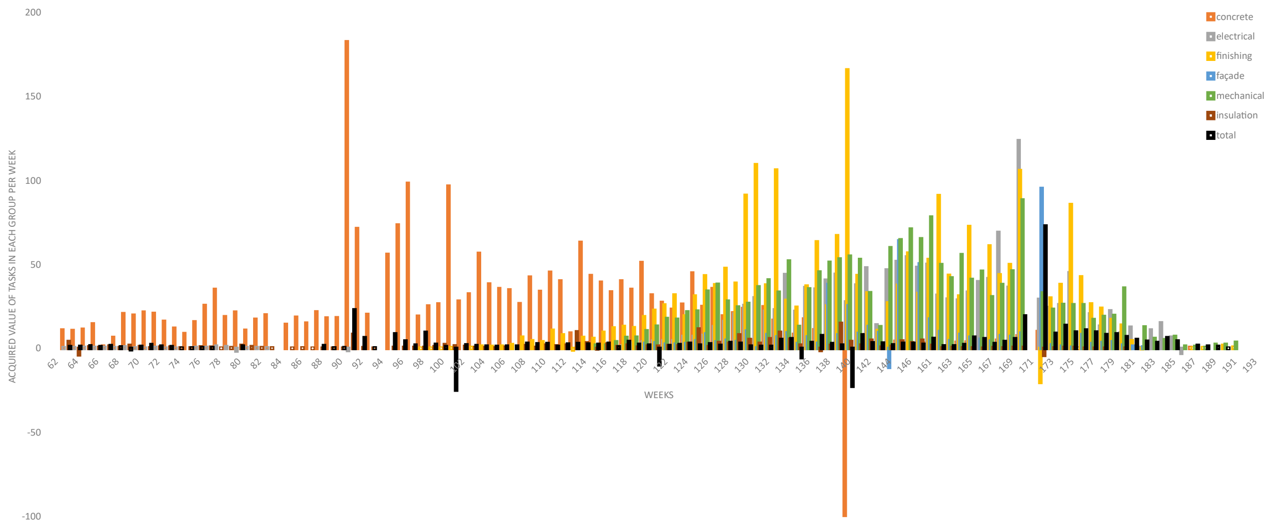

Figure 3 Modified acquired values per week

Table 1 below presents the confusion matrix and standard metrics during the detecting of fake instances.

Table I Confusion matrix and relevant metrics of fake instances for IQR

		Detected	
		Real	Fake
Actual	Real	92	19
	Fake	1	6
Accuracy			0.8305
Precision			0.2400
Recall			0.8571
F1 score			0.3750

It is seen that IQR successfully flags large deviations. However, its high false positive rate indicates that not all flagged anomalies are genuine. This suggests that, even though statistical approach could build a handy baseline detection system, contextual variations in construction progress (e.g., milestone-based completions) might be misclassified as anomalies.

Machine learning approach: isolation forest (IF)

The isolation forest (IF) algorithm is an unsupervised ML method for anomaly detection in high-dimensional data. It randomly partitions data and measures how quickly an instance becomes isolated. Anomalies tend to be isolated in fewer partitions than normal data points as they differ from normal data. In other words, an ensemble of randomly generated decision trees is built and the anomalies are identified based on their isolation depth.

The algorithm is as follows. The dataset is randomly split into multiple subgroups. Then a decision tree is built by selecting a random feature and a random split value at each node for each subgroup. Anomalous instances require fewer splits to be isolated because they do not stay together with the others. Anomaly scoring is computed based on the average depth required to isolate a point. A lower depth (indicates fewer splits) indicates a higher likelihood of an anomaly. The formula for the anomaly score for a data point x is calculated as follows.

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$$

$E(h(x))$ is the average depth of x across all trees, and $c(n)$ is the normalization factor based on tree depth for a given dataset size n.

Isolation forest is utilized for anomaly detection as it is efficient and scalable. $O(n.logn)$ time complexity makes it useful for large datasets. As it is an unsupervised model, labelling is not a necessity. This is suitable for real-world construction progress data as labelled anomaly prediction data is not realistic. Again, akin to IQR, isolation forest can also handle nonlinear data. This is desirable due to the fact that the data may include sudden variations. Its robustness against high-dimensionality is beneficial since construction tracking involves multiple parameters (planned, acquired, and realized progress). Furthermore, isolation forest is not being affected by feature scaling.

The confusion matrix and standard metrics for fake instance detection are provided in Table 2, below. It is observed that isolation forest has higher precision and fewer false positives than IQR. It has similar recall to IQR, so that, determining actual anomalies around the same rate. Finally, F1-score has improved, indicating a more balanced detection capability. One could conclude that it can effectively distinguish irregular yet valid milestones from genuinely fraudulent instances. However, some categories showed higher misclassification rates because of project-specific fluctuations in progress, again as in the IQR approach. For instance, milestone-based tasks (e.g., façade completion) sometimes seemed unusual when legitimate progress jumps occurred.

The limitations of both methods yield to an ensemble approach that combined statistical and machine learning methods to enhance detection accuracy.

Ensemble approach

The interquartile range method is significant and isolation forest provides enhanced results. IQR performs better in discovering extreme deviations but results in a large number of false positives. On the other hand, isolation forest yields a refined detection mechanism; yet, it presents sudden jumps based on milestones as anomalies. An ensemble approach could solve the problems while providing a trade-off between the two levels of accuracy and sensitivity owed to the complementary characteristics of both approaches.

Table II Confusion matrix and relevant metrics of fake instances for IF

		Detected	
		Real	Fake
Actual	Real	100	11
	Fake	1	6
Accuracy		0.8983	
Precision		0.3529	
Recall		0.8571	
F1 score		0.5000	

Using an ensemble of multiple models lessens the deficiencies of singular models. In this study, the ensemble method is proposed with the intention of reducing false positives, enhancing recall, and improving decision confidence. Cross-validating flagged anomalies reduces false positives where legitimate milestone completions were misclassified. Recall enhances when significant irregularities could be consistently detected. Finally, providing a layered detection mechanism, where multiple models confirm an anomaly classification enhances confidence.

Two ensemble scenarios are tested:

1. Inclusive Ensemble: A data point is flagged as an anomaly if IQR **or** IF detected it as an outlier.

2. Consensus Ensemble: A data point is flagged only if IQR **and** IF detected it as an anomaly.

Scenario 1

Any snapshot detected as an anomaly by either method is flagged. The confusion matrix and standard metrics for fake instance detection are provided in Table 3, below.

Table III Confusion matrix and relevant metrics of fake instances for the inclusive ensemble scenario

		Detected	
		Real	Fake
Actual	Real	89	22
	Fake	0	7
Accuracy		0.8136	
Precision		0.2414	
Recall		1.0000	
F1 score		0.3889	

Scenario 2

A snapshot was flagged only if both methods detected an anomaly. The confusion matrix and standard metrics for fake instance detection is provided in Table 4, below.

Table IV Confusion matrix and relevant metrics of fake instances for the consensus ensemble scenario

		Detected	
		Real	Fake
Actual	Real	103	8
	Fake	2	5
Accuracy		0.9153	
Precision		0.3846	
Recall		0.7143	
F1 score		0.5000	

It is observed that Scenario 1 ("Inclusive") detected all actual anomalies (100% recall) but suffered from more false positives, lowering precision. Scenario 2 ("Consensus") achieved the best balance, reducing false positives while maintaining high accuracy and F1-score. As a matter of fact, two scenarios could be put together where an inclusive approach warns human actors as a decision support system, and a consensus approach is considered as a more serious alert that requires investigation.

By combining statistical and machine learning techniques, the ensemble approach provides a robust and scalable framework for anomaly detection in construction project tracking. Early fraud detection is one of the main advantages, as this reduces the reliance on manual inspections and creates a more automated, data-driven system to detect and scrutinize anomalies in progress reporting. Also, given its ability to correlate the suspicious behavior with the project, it can easily work with the project management packages like Primavera, MS Project, or in-house project management packages to ensure timely alerts are triggered. The method is also extremely scalable, so it is computationally cheap to apply across construction sites in real time, which allows companies to control all their large-scale operations and make better decisions.

Discussion and Practical Implications

Interquartile range and other statistical techniques are effective at detecting large anomalies but tend to generate false positives. This leads to false investigations and decreased trust in the system. Machine learning based isolation forest, on the other hand, finely detects anomaly with a focus on localized patterns and distance from the average.

Since both scenarios have their advantages and disadvantages, ensemble methods bring the best of both worlds. The combination of statistical approach with machine learning approach brings high recall with low false positive rates. The output of the ensemble approach, which remarkably decreases likely false alarms, could be a primary information during project decision making.

The findings on the early detection of anomalies let project managers to take early action. In addition, real-time identification of workforce inefficiencies, delays or disruptions make room for proactive reorganization of the project tasks. The proposed approaches could be easily integrated into existing project management tools, like Primavera or MS Project. In this case, the need for manual monitoring and missing unnoticed anomalies would be reduced. Moreover, insider manipulations, either to report artificial progress or to gain financial benefit, could be noticed; therefore, ends up with more trustworthy project reporting.

The potential of machine learning could grow larger in anomaly detection in construction project management. Static rule-based monitoring could be replaced with automated tools and present a steadily approach learning from the fluctuating conditions of the projects. This ensures that any outliers, whether due to misreporting, or fraudulent activity, could be noticed. Thus, project managers could shift from problem solving afterwards to averting problems beforehand. The proposed approach might have optimizing effects on other industrial projects. Moreover, financial accountability of projects could be enhanced. In the future, the proposed tool integrated to other project management tools together with real-time monitoring data could produce better predictions and lead to more intelligent, digitized project management.

The proposed approach is probably one of the many possibilities of the application of artificial intelligence in the construction domain. Even though the proposed approach utilizes two fundamental methods in an ensemble approach, in the availability of more data, combining autoencoders with other machine learning methods may improve the output of anomaly detection [9, 10]. Future developments may provide continuous information flow where potential anomalies could be identified before they develop into critical issues, thereby their effects could be timely mitigated [7, 14]. Further enhancements might be enabled by integrating additional contextual data, such as productivity measures of the subcontractors and their financial history. This could improve the model output by being aware of not only the project-related data but also other affecting factors. To provide validity to the proposed approach, presented historical study could be implemented in large construction sites and compared hand in hand with conventional methods. Practical deployment could also point several constraints that cannot be foreseen in this paper.

Conclusion

This paper present that anomaly detection could be realized via statistical and machine learning methods in the construction industry. The interquartile range offers a basic approach for defining an outlier; however, this leads to many false positives. Isolation forest produces more data-driven approach so that it decreases the number of false positives. The first ensemble method provides perfect detection at the cost of more false positives. The second ensemble method, on the other hand, provides the least number of false positives. Therefore, hybrid ensemble methods seem to improve accuracy, reduces false positives and presents a desirable solution for practical construction monitoring.

From project managers perspective, utilizing AI-based anomaly detection is useful for resource optimization and timely risk management. Early anomaly detection improves proactiveness in decision-making, leaves room for taking action before delays become costly. So that, keeping projects on track. AI tools could be more valuable as the construction projects grow bigger and become more sophisticated.

A promising future focus could be introducing interpretability to AI-driven tools for the construction sector. Even though the automated detection mostly provides useful output, the models are imperfect. The untrustworthy black boxes could be replaced by more transparent tools by developing explainable methods with the same functionality. Finally, developed methods could be integrated into building information modeling systems and integrated sensor data flow from the project site could be used to further optimize a digital information system of construction project management. Broadening the use of AI in construction progress tracking could pave the way for the industry to make a major leap towards improved efficiency, and risk mitigation. Project monitoring might shift the construction sector to an intelligent and robust state by embracing data-driven automation.

References

[1] Jin, S. (2024). Measuring complexity in mega construction projects: fuzzy comprehensive evaluation and grey relational analysis. Engineering, Construction and Architectural Management. https://doi.org/10.1108/ecam-07-2024-0951

[2] Zhao, X., Liu, Y., Lang, X., Liu, K., Yang, X., & Liu, L. (2024). Study on the Characteristics and Operational Mechanisms of Industry–University–Research Collaborative Innovation in Megaprojects: The Case from China. Systems. https://doi.org/10.3390/systems12120553

[3] Xiang, P., Yang, Y., Yan, K., & Jin, L. (2024). Identification of Key Safety Risk Factors and Coupling Paths in Mega Construction Projects. Journal of Management in Engineering. https://doi.org/10.1061/jmenea.meeng-5926

[4] Mostofi, F., Toğan, V., & Tokdemir, O. B. (2025). A cost estimation recommendation system for improved contingency management in construction projects. Neural Computing & Applications 37, 3521–3538. https://doi.org/10.1007/s00521-024-10740-y

[5] Mostofi, F., Tokdemir, O. B., & Toğan, V. (2024). Generating synthetic data with variational autoencoder to address class imbalance of graph attention network prediction model for construction management. (2024). https://doi.org/10.1016/j.aei.2024.102606

[6] Shyam, R., & Tiwari, S. (2025). A Comprehensive Review of Machine Learning and Multi-Criteria Decision Analysis in Construction Delay Management. International Research Journal on Advanced Science Hub. https://doi.org/10.47392/irjash.2025.002

[7] Manu, B. A. (2024). Leveraging Artificial Intelligence for optimized project management and risk mitigation in construction industry. World Journal of Advanced Research and Reviews. https://doi.org/10.30574/wjarr.2024.24.3.4026

[8] Zapata-Cortes, O., Serna, M. D. A., Zapata-Cortés, J. A., & Restrepo-Carmona, J. A. (2024). Machine Learning Models and Applications for Early Detection. Sensors. https://doi.org/10.3390/s24144678

[9] Nishat, M. M., Neraas, S. M., Marsov, A., & Olsson, N. O. E. (2024). Prediction of project activity delays caused by variation orders: a machine-learning approach. https://doi.org/10.1088/1755-1315/1389/1/012038

[10] Al-Sabah, B., & Anbarjafari, G. (2024). Anomaly Detection in Kuwait Construction Market Data Using Autoencoder Neural Networks. Information. https://doi.org/10.3390/info15080424

[11] Al-Bataineh, F., Khatatbeh, A. A., & Alzubi, Y. (2024). Unsupervised machine learning for identifying key risk factors contributing to construction delays. Organization, Technology and Management in Construction: An International Journal. https://doi.org/10.2478/otmcj-2024-0014

[12] Alhasan, A. M. A., & Alawadhi, E. K. E. (2024). Evaluating the Impact of Artificial Intelligence in Managing Construction Engineering Projects. The Journal of Engineering Sciences and Information Technology. https://doi.org/10.26389/ajsrp.k090724

[13] Mostofi, F., Toğan, V., Ayözen, Y., & Tokdemir, O. B. (2022). Predicting the Impact of Construction Rework Cost Using an Ensemble Classifier. Sustainability, 14 (22), 14800. https://doi.org/10.3390/su142214800

[14] Gajera, R. (2024). The Impact of SmartPM's Ai-Driven Analytics on Predicting and Mitigating Schedule Delays in Complex Infrastructure Projects. International Journal of Scientific Research in Science, Engineering and Technology. https://doi.org/10.32628/ijsrset24115101

[15] Yao, D., & Soto, B. G. de. (2024). Cyber Risk Assessment Framework for the Construction Industry Using Machine Learning Techniques. Buildings. https://doi.org/10.3390/buildings14061561

[16] Li, S., & Zhu, A. (2024). Integrating Wireless Sensor Networks in Construction Project Risk Management: A Cyber-Physical Approach to Enhancing Robustness and Security in Smart Building Developments. https://doi.org/10.1109/wrcsara64167.2024.10685736

[17] Yang Z., Xiaocheng Y. & Yongmei Y., "Design and implementation of RFID-based construction restricted area intrusion detection system," Proc. SPIE 13160, Fourth International Conference on Smart City Engineering and Public Transportation (SCEPT 2024), 1316002 (16 May 2024); https://doi.org/10.1117/12.3030396

[18] Tran, S. V.-T., Lee, D., Bao, Q. L., Yoo, T., & Jo, J. (2023). A Human Detection Approach for Intrusion in Hazardous Areas Using 4D-BIM-Based Spatial-Temporal Analysis and Computer Vision. Buildings. https://doi.org/10.3390/buildings13092313

[19] Soto, B. G. de, Georgescu, A., Mantha, B. R. K., Turk, Z., Maciel, A., & Sonkor, M. S. (2022). Construction cybersecurity and critical infrastructure protection: new horizons for Construction 4.0. Journal of Information Technology in Construction. https://doi.org/10.36680/j.itcon.2022.028

[20] Saah, A. E. N., Yee, J.-J., & Choi, J.-H. (2023). Securing Construction Workers' Data Security and Privacy with Blockchain Technology. Applied Sciences. https://doi.org/10.3390/app132413339

[21] Sonkor, M. S., Xu, X., Prieto, S. A., & Soto, B. G. de. (2022). Vulnerability Assessment of Construction Equipment: An Example for an Autonomous Site Monitoring System. Proceedings of the 39th ISARC, Bogotá, Colombia, https://doi.org/10.22260/isarc2022/0040

[22] Kochergin, S. V., Artemova, S., Bakaev, A. A., Mityakov, E. S., Vegera, Z. G., & Maksimova, E. (2024). Cybersecurity of smart grids: Comparison of machine learning approaches training for anomaly detection. Russian Technological Journal. https://doi.org/10.32362/2500-316x-2024-12-6-7-19

[23] Chua, W., Pajas, A. L. D., Castro, C. L. B. de, Panganiban, S. P., Pasuquin, A. J., Purganan, M. J., Malupeng, R., Pingad, D. J., Orolfo, J. P., Lua, H. H., & Velasco, L. C. (2024). Web Traffic Anomaly Detection Using Isolation Forest. Informatics (Basel). https://doi.org/10.3390/informatics11040083

[24] Yılmazer Demirel, D. and Sandikkaya, M.T. (2023) ACUM: An Approach to Combining Unsupervised Methods for Detecting Malicious Web Sessions. 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, pp. 288-293, https://doi.org/10.1109/UBMK59864.2023.10286727

[25] Manavadaria, M. S., Srinivas, T. A. S., Ahamed, S. K., Amshavalli, M., Nadaf, A. B., & Varadharajan, B. (2024). Anomaly Detection Algorithms in Cybersecurity. Advances in Computational Intelligence and Robotics Book Series. https://doi.org/10.4018/979-8-3693-7540-2.ch013

[26] Sharma, R., & Grover, M. (2024). Enhancing Cybersecurity with Machine Learning: Evaluating the Efficacy of Isolation Forests and Autoencoders in Anomaly Detection. https://doi.org/10.1109/iccpct61902.2024.10673338

[27] Ness, S., Eswarakrishnan, V., Sridharan, H., Shinde, V., Janapareddy, N. V. P., & Dhanawat, V. (2025). Anomaly Detection in Network Traffic using Advanced Machine Learning Techniques. IEEE Access. https://doi.org/10.1109/access.2025.3526988

[28] Bhuiyan, M. R. I., Faraji, M. R., Tabassum, M. N., Ghose, P., Sarbabidya, S., & Akter, R. (2024). Leveraging Machine Learning for Cybersecurity: Techniques, Challenges, and Future Directions. Edelweiss Applied Science and Technology. https://doi.org/10.55214/25768484.v8i6.2930

[29] Abas, A., Oshoiribhor, O. E., & John-Otumu M. A. (2017). Rule Based Forward Chaining Technique for Detecting and Controlling Fraud in Project Monitoring System. International Journal of Software Engineering & Applications. https://doi.org/10.5121/IJSEA.2017.8503

[30] Wang, K., Zhao, Z., Deng, R., & Mao, Y. (2023). Automated Construction Progress Monitoring and Controlling Framework Based on Computer Vision and Internet of Things System. 2023 6th International Conference on Electronics Technology (ICET). https://doi.org/10.1109/icet58434.2023.10211573

[31] Shamsollahia, D., Moselhib, O., & Khorasanic, K. (2022). Construction Progress Monitoring and Reporting using Computer Vision Techniques – A Review. 39th International Symposium on Automation and Robotics in Construction (ISARC 2022).

[32] Abdel-Wahhab, O., & Elazouni, A. (2010). Progress Monitoring of Construction Projects Using Statistical Pattern Recognition. https://doi.org/10.1061/41109(373)121

[33] Jin, X., & Le, Y. (2014). Monitoring Construction Projects Using Information Technologies. Proceedings of the 17th International Symposium on Advancement of Construction Management and Real Estate, pp. 1011-1020. https://doi.org/10.1007/978-3-642-35548-6_104

[34] Hansen, S. Setiono, B., Handayani, S., & Dewobroto, W. S. (2025). Influence Factor Model for Fraud Detection in Construction Companies' Financial Statements. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction. https://doi.org/10.1061/JLADAH.LADR-1301

[35] Tambwe, O. T., Aigbavboa, C. O., Akinradewo, O. I., & Adekunle, P. A. (2025). Measures to Address Cyber-Attacks in Construction Project Data Management Processes: A Cybersecurity Perspective. IET Information Security. https://doi.org/10.1049/ise2/7398742

[36] Lalropuia, K. C., Goyal, S., de Soto, B. G., Yao, D., Sonkor, M. S. (2025). Mitigating Malicious Insider Threats to Common Data Environments in the Architecture, Engineering, and Construction Industry: An Incomplete Information Game Approach. Journal of Cybersecurity and Privacy. https://doi.org/10.3390/jcp5010005