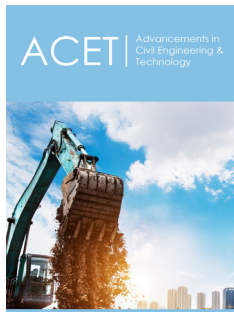


Toward Intelligent and Autonomous Tunnelling: A Review of Machine Learning Applications in Tunnel Boring Machines

ISSN: 2639-0574



Muhammad Shahroz Khalid^{1,2,3}, Syed Jamal Arbi^{1,2,3}, Muhammad Irsan Khalid^{1,2,3}, Muhammad Shahbaz^{1,2,3*}, Naveed Ahmad^{1,2,3} and Shujahat Anwar^{1,2,3}

¹State Key Laboratory of Intelligent Geotechnics and Tunnelling, Shenzhen University, China

²Key Laboratory of Geological Risk Prevention and Safety Control for Megacities (MONR), Shenzhen University, China

³College of Civil and Transportation Engineering, Shenzhen University, China

Abstract

Tunnel boring machines (TBMs) have become essential for modern underground infrastructure development; however, their operation remains challenging due to complex geological variability, uncertain ground conditions and strong interactions between machine behaviour and surrounding formations. The increasing digitalization of TBM systems has generated large volumes of operational and geological data, creating new opportunities for artificial intelligence (AI) and machine learning (ML)-based intelligent tunnelling. This mini-review provides a comprehensive overview of the recent progress of AI-driven TBM technologies, covering the data foundation of intelligent tunnelling, machine-learning paradigms, geological condition identification, TBM performance prediction, safety and anomaly detection, adaptive learning, operational optimization, and digital twin-enabled closed-loop control. The review highlights how ML has evolved from standalone predictive modelling toward integrated frameworks capable of connecting real-time sensing, state estimation, decision-making and operational control. Despite significant advances, practical deployment remains limited by challenges including data scarcity and quality, model transferability across different projects and geological environments, interpretability, uncertainty quantification, computational requirements and operational reliability. Future research directions are expected to focus on adaptive and physics-informed learning, multimodal data integration, transfer learning, human-supervised autonomous control and digital twin frameworks. These developments will be critical for transforming TBMs from data-rich excavation machines into intelligent systems capable of safer, more efficient and increasingly autonomous tunnelling operations.

Keywords: Tunnel boring machine; Artificial intelligence; Machine learning; Intelligent tunnelling; Digital twin; Autonomous excavation; TBM optimization

Introduction

Rapid urbanization and the increasing demand for underground infrastructure have accelerated the development of metro systems, transportation tunnels, and other subsurface facilities worldwide. Tunnel boring machines (TBMs) have consequently become a preferred excavation technology for many large-scale tunnelling projects because of their high construction efficiency, operational safety and relatively limited disturbance to the surrounding environment [1-3]. Despite these advantages, TBM excavation remains inherently challenging because geological conditions may vary substantially along a tunnel alignment, while machine performance is strongly influenced by the complex interaction between the excavation system and the surrounding ground. Unexpected geological transitions or inappropriate operating conditions can result in reduced excavation efficiency, prolonged stoppages, excessive equipment loading or damage, ground instability and in severe cases, tunnel-face collapse [4-6]. Therefore, reliable prediction of TBM performance and timely

***Corresponding author:** Muhammad Shahbaz, State Key Laboratory of Intelligent Geotechnics and Tunnelling, Key Laboratory of Geological Risk Prevention and Safety Control for Megacities (MONR), College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China

Submission: 📅 September 07, 2026

Published: 📅 September 29, 2026

Volume 7 - Issue 2

How to cite this article: Muhammad Shahroz Khalid, Syed Jamal Arbi, Muhammad Irsan Khalid, Muhammad Shahbaz*, Naveed Ahmad and Shujahat Anwar. Toward Intelligent and Autonomous Tunnelling: A Review of Machine Learning Applications in Tunnel Boring Machines. *Adv Civil Eng Tech.* 7(2). ACET.000660.2026.
DOI: [10.31031/ACET.2026.07.000660](https://doi.org/10.31031/ACET.2026.07.000660)

Copyright@ Muhammad Shahbaz, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

understanding of changing ground-machine interactions are essential for improving construction safety, maintaining excavation efficiency, optimizing resource allocation and reducing project delays and associated costs.

Traditionally, TBM performance prediction has relied on empirical and semi-empirical formulations, statistical regression models, and mechanistic or geo-mechanical approaches developed from geological properties, machine characteristics, and historical construction data [7,8]. Although these approaches have provided valuable engineering guidance, their applicability can be limited when tunnelling encounters heterogeneous geology, rapidly changing operating conditions, or complex nonlinear interactions between the machine and the surrounding ground [9]. With the increasing instrumentation of modern TBMs, research has therefore shifted toward exploiting continuously recorded operational data as an additional source of information on both machine behaviour and encountered ground conditions. In particular, parameters such as cutterhead torque, thrust, rotational speed and advance or penetration rate has been shown to contain meaningful signatures of changing geological conditions and can be used for data-driven ground classification and forecasting [10,11]. The combination of pre-construction geological information with increasingly abundant real-time TBM monitoring data consequently provides the fundamental data basis for developing machine-learning- and AI-enabled systems capable of supporting prediction, decision-

making, and ultimately more intelligent tunnelling operation [12].

Despite the increasing availability of monitoring data, TBM operation in practice still depends substantially on the experience and judgement of machine operators, who must continuously balance geological conditions, equipment response, construction efficiency, and operational safety when adjusting control parameters. Modern TBMs record numerous variables, including thrust, cutterhead torque, rotational speed, face pressure, and advance-related parameters, yet these data are not always fully exploited for proactive prediction and operational adjustment [13]. Inadequate or delayed adjustment of operating parameters under changing ground conditions may reduce excavation efficiency and increase the likelihood of excessive equipment loading, cutter wear, machine jamming, over-excavation, ground deformation, or other construction-related risks [14]. These limitations have stimulated increasing interest in artificial intelligence (AI) as a means of transforming continuously acquired construction data into actionable information for prediction, optimization, and decision support. Machine learning (ML), as a major branch of AI, is particularly suited to this task because it can learn complex and nonlinear relationships directly from historical and real-time data, thereby supporting ground-condition recognition, TBM performance prediction, operational parameter adjustment and progressively more automated tunnelling processes [15].

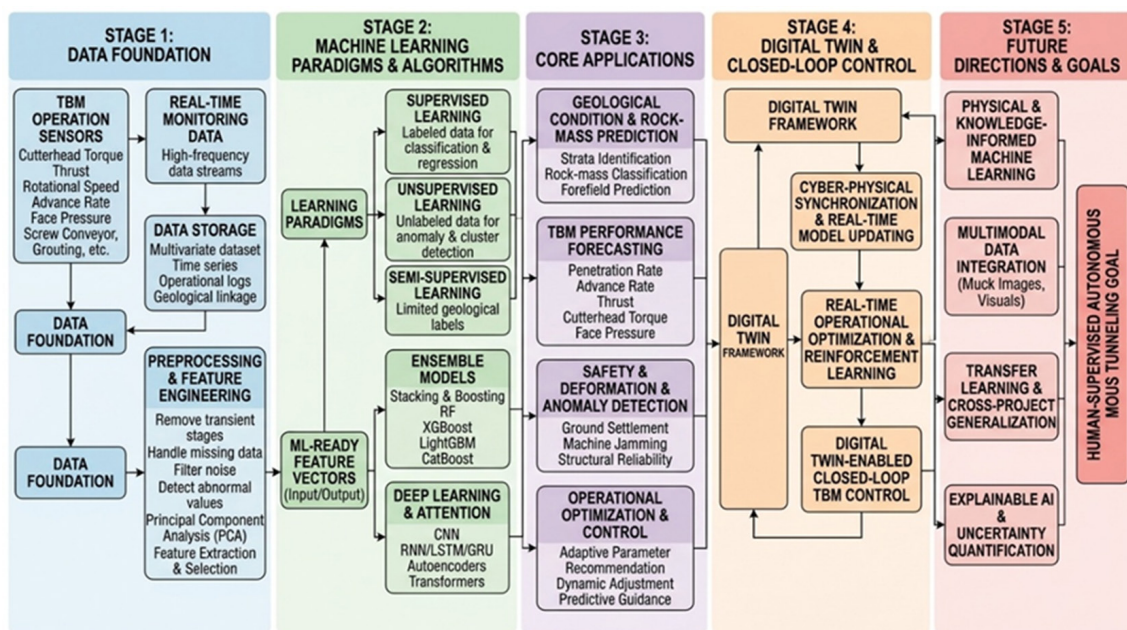


Figure 1: Framework illustrating the transition of TBM operations from raw monitoring data to AI-driven predictive modeling and autonomous decision support.

Accordingly, this mini-review examines the evolving role of artificial intelligence in advancing TBM tunnelling from conventional data-driven prediction toward intelligent, adaptive and increasingly autonomous operation. Unlike previous reviews that mainly summarize individual applications of AI and machine learning in TBM engineering, this review provides an integrated perspective

on the evolution of intelligent TBM systems by connecting data acquisition, machine-learning-based interpretation, prediction, operational decision-making, optimization, and autonomous control. The review first considers the data foundation of intelligent TBM systems and the principal machine-learning paradigms used to extract meaningful information from geological, operational

and monitoring data. It then synthesizes recent developments in geological condition and rock-mass prediction, TBM performance forecasting, safety and anomaly detection, adaptive and online learning, and AI-assisted operational optimization and control. Particular attention is given to the transition from standalone predictive models toward integrated frameworks in which real-time sensing, data processing, state estimation, decision-making, and machine control are interconnected. Furthermore, this review highlights the emerging role of digital twins, reinforcement learning, and physics-informed approaches as enabling technologies for developing closed-loop intelligent tunnelling systems. Finally, the review critically evaluates the major challenges limiting practical implementation, including data quality and scarcity, cross-project model transferability, interpretability, uncertainty quantification, computational requirements and operational reliability, and discusses future directions toward robust, adaptive, physics- and knowledge-informed, and human-supervised AI frameworks for safer and more efficient mechanized tunnelling (Figure 1).

Review Methodology

A structured literature search was conducted to identify and analyze research progress on artificial intelligence (AI) and machine learning (ML) applications in tunnel boring machine (TBM) operations. The literature search was performed using major scientific databases, including Web of Science Core Collection, Scopus, ScienceDirect and Google Scholar. The literature considered in this review covers publications from 2008 to 2026, based on the earliest and latest studies included in the reference collection. A total of 46 publications were analyzed, including foundational studies before 2020 and recent developments after 2020. The search focused on studies investigating the application of data-driven approaches for TBM-related tasks, including geological condition identification, TBM performance prediction, operational parameter forecasting, safety and anomaly detection, adaptive learning, optimization, and digital twin-based intelligent tunnelling.

The literature search was conducted using keywords associated with TBM automation and AI-based modelling, including “Tunnel Boring Machine,” “TBM,” “machine learning,” “artificial intelligence,” “deep learning,” “data-driven modelling,” “geological prediction,” “performance prediction,” “optimization,” “reinforcement learning,” “digital twin,” and “autonomous tunnelling.” The selected studies included research articles and review papers that applied AI/ML methods to TBM operational data, geological information, machine-ground interaction analysis, and intelligent decision-support systems. Studies were excluded when they focused solely on conventional empirical, analytical, or numerical approaches without AI/ML integration, or when they were unrelated to intelligent TBM applications.

Understanding of TBM Types

Tunnel boring machines (TBMs) are commonly classified according to the geological and groundwater conditions in which they operate, including hard rock, soft ground, and mixed or changing ground. For hard-rock excavation, Gripper TBMs are

generally used in competent rock where grippers provide reaction against the tunnel walls, whereas Single Shield TBMs provide continuous protection in weaker or fractured rock and advance by thrusting against installed lining segments. Double Shield TBMs combine these two operating principles, allowing efficient excavation under variable rock conditions. In soft ground, Earth Pressure Balance (EPB) TBMs stabilize the excavation face by maintaining controlled pressure with conditioned excavated soil inside the chamber, making them particularly suitable for cohesive and fine-grained soils, while Slurry TBMs use pressurized slurry for face support and hydraulic muck transportation and are especially effective in permeable granular soils and high-groundwater-pressure environments. For tunnel alignments characterized by substantial geological variability, Hybrid and Multi-mode TBMs can adapt or switch between different excavation and face-support modes, such as open, EPB, and slurry operation. Therefore, the selection of an appropriate TBM depends primarily on rock or soil characteristics, groundwater pressure, required face-support mechanism, excavation stability, and anticipated geological variability along the tunnel alignment (Figure 2).

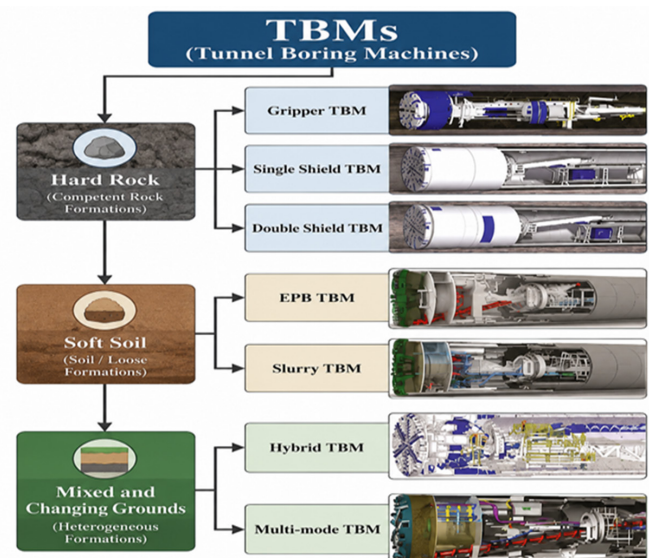


Figure 2: Overview of TBM types categorized by hard rock, soft ground, and mixed ground conditions.

Data Foundation for Intelligent TBM Systems

Data acquisition

Modern TBMs are highly instrumented excavation systems in which numerous sensors continuously record the interaction between the machine, the excavation process and the surrounding ground. Depending on the TBM type and monitoring configuration, commonly acquired variables include cutterhead torque and rotational speed, total thrust, penetration or advance rate, chamber or face pressure, screw-conveyor parameters, hydraulic pressures, grouting variables, machine position and attitude, and other subsystem operating states. These measurements are generally collected through the TBM control system and associated monitoring devices at high temporal resolution, generating large multivariate datasets throughout excavation. For shield TBMs using

segmental lining, the data are commonly associated with individual rings or excavation cycles, allowing operational responses to be linked with tunnel position and encountered geological conditions. More importantly, TBM operational data contain indirect information on machine-ground interaction because changes in geological conditions alter excavation resistance and are reflected in variations in thrust, torque, rotational speed, penetration rate and face pressure. This strong coupling between machine response

and geological conditions provides the fundamental data basis for ground-condition identification, performance prediction and intelligent TBM operation. However, raw monitoring data often include transient operating stages, abnormal values, missing records, sensor noise, and redundant variables; therefore, systematic data processing and feature engineering are required before these data can be reliably used for machine-learning model development (Figure 3).

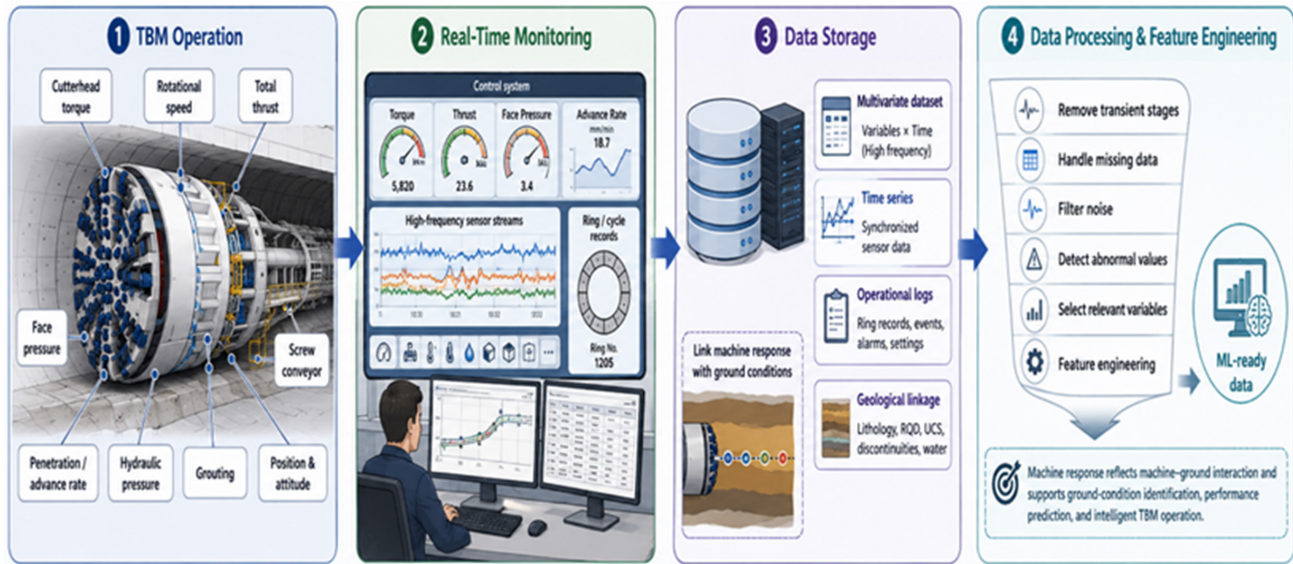


Figure 3: Workflow of TBM data acquisition and preprocessing.

Data processing and feature engineering

Extracting the enormous amount of information contained in a dataset into a condensed and clean feature vector as the input of a machine learning algorithm is a strategic endeavor that determines the success of predictions [16]. From raw data to feature data. The research's main aim is to remove irrelevant and redundant data from the dataset obtained from TBM. The dimensions of the input and output were considerably reduced after the process because unnecessary parameters were removed. There are two dimensionality-reduction techniques: feature extraction and feature selection [17,18]. The researchers use a certain data-mining method, whereas the latter only selects useful information based on human judgment and knowledge. From the feature data to the processed data. The extracted information still contains flaws requiring a process called pre-processing. This process involves (1) abandoning the incomplete information not suitable for learning, (2) amending some irregular data, and (3) denoising the measured data using filters [19]. From the processed data to the feature vectors. The processed data may be directly used as the input and output of a machine learning algorithm. However, many contributors wished to further reduce the dimensions and possible noise in the data. After these steps, we obtained the data used as input and output in the machine learning algorithms, which are called feature vectors (Figure 4).

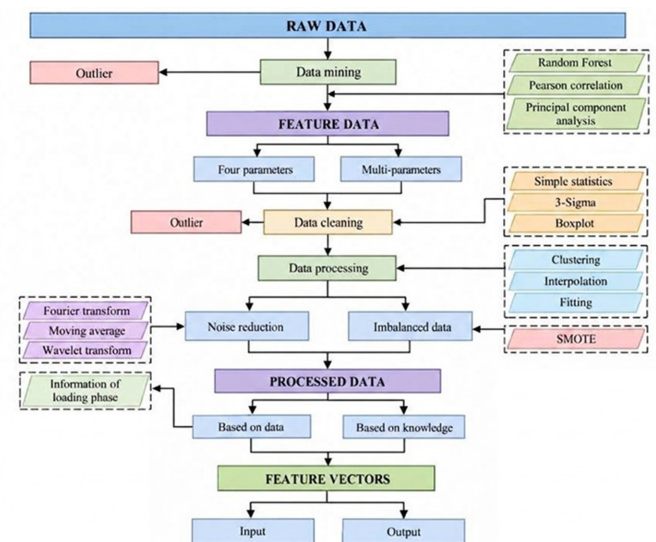


Figure 4: Multi-stage data preparation framework incorporating feature selection, outlier removal, noise reduction, and data balancing.

Machine Learning in TBM Tunnelling Overview

Machine learning comprises a large family of individual methods. Generally, it performs classification and regression tasks.

Classification learning normally involves the investigation of objects into several distinct categories, such as geological type, rock-mass class, or abnormal operating state. Regression learning predicts future performance in terms of magnitude, such as penetration rate, advance rate, cutterhead torque, thrust, face pressure, settlement, or deformation. Many algorithms work for both classification and regression purposes, depending on the goal and settings adopted by researchers [20]. Reliable model development, however, depends not only on the selected algorithm. But also, the adoption of various support techniques for reliable and effective machine learning is an important part of this study. Various researchers have discussed the details that may be summarized as follows: (1) selection of hyperparameters, (2) sensitivity study by the K-fold cross-validation method, (3) model validation by the Precision-recall-F1 evaluation, (4) treatment for imbalanced data, and (5) evaluation of important parameters.

Role of ML in intelligent TBM systems

Recent TBM research increasingly positions machine learning not simply as a prediction technique, but as an intermediate intelligence layer that connects sensing, interpretation, decision-making and machine operation. In this role, ML first converts heterogeneous geological, operational, and monitoring data into models describing the evolving TBM ground interaction; it then supports continuous assessment of machine and ground states and provides predictive information that can guide operational decisions. A recent review of AI-enabled TBM management organizes this functionality into three closely related components: modeling, monitoring, and control, which together represent the progression from understanding system behaviour to acting upon it [21]. At the modeling level, ML establishes data-driven relationships among geological conditions, machine settings, and excavation responses; at the monitoring level, these relationships can be used to identify changing ground conditions, abnormal behaviour, or emerging risks during construction; and at the decision and control level, predicted states can support the selection or adjustment of operating parameters. Earlier tunnelling studies similarly demonstrated the potential of ML for automatic rock-mass classification, geological prognosis updating, monitoring interpretation, and automated tunnel inspection [22]. More recent research extends this concept toward automation, where reliable ground prediction and anomaly recognition become inputs to real-time operational decision support rather than isolated analytical outputs [23]. Thus, the principal role of ML in intelligent TBM systems is to progressively transform raw construction data into system knowledge, system knowledge into predictions, and predictions into informed operational actions, providing the link between data acquisition and increasingly autonomous tunnelling.

Learning paradigms and ML algorithms

Within TBM engineering, the choice of a machine-learning paradigm is determined primarily by the availability of reliable labels, the structure of the monitoring data and the intended engineering task. Supervised learning remains the dominant approach because operational records can often be paired with geological classifications, measured machine responses,

or monitored ground behaviour to establish input-output relationships. By contrast, unsupervised learning is useful when reliable labels are unavailable or incomplete, allowing latent operating states, geological groupings, and anomalous observations to be identified directly from similarities within multivariate TBM data. Reviews of TBM-related machine learning confirm that both supervised and unsupervised methods have been widely investigated, although supervised approaches remain particularly common for ground-condition and performance prediction [24]. Semi-supervised learning provides an important intermediate strategy because TBMs generate abundant unlabelled operational data, whereas reliable geological labels are often available only near boreholes or mapped sections. Yu and Mooney [24] demonstrated this advantage by combining sparsely labelled borehole-associated TBM data with much larger quantities of unlabelled excavation records, showing that semi-supervised learning can outperform conventional supervised models when labelled geological information is limited.

At the model level, TBM research has progressively expanded from relatively simple statistical and individual machine-learning algorithms toward ensemble [25], deep-learning and more recently, attention-based architectures [26]. Commonly adopted conventional models include linear and Bayesian regression, k-nearest neighbours, support vector machines, decision trees, Gaussian-process methods, artificial neural networks, and neuro-fuzzy approaches [27]. Ensemble techniques such as random forests, AdaBoost, gradient boosting, XGBoost, LightGBM, CatBoost and stacking combine multiple learners to better represent nonlinear relationships and have become particularly attractive for structured TBM datasets [28]. A review of 45 studies developed from a shared TBM dataset found extensive use of both ensemble and deep-learning techniques, confirming a broader methodological shift toward more flexible data-driven architectures [29]. Deep-learning models, including CNNs, RNNs, LSTMs, GRUs, and autoencoders, further enable automatic representation learning from high-dimensional and sequential monitoring data. More recently, attention and Transformer-based architectures have begun to appear in TBM forecasting, reflecting growing interest in modelling both long-range temporal dependencies and interactions among multiple operational variables. For example, recent Transformer-based studies have targeted multivariate long-term excavation-parameter forecasting and simultaneous temporal-feature representation in TBM performance prediction [30]. The suitability of these model families, however, depends strongly on the characteristics of the available dataset rather than on algorithm complexity alone. TBM monitoring data contain pronounced nonlinear, multivariate and sequential characteristics, making tree-based ensembles attractive for modelling nonlinear relationships among structured geological and operational variables, while recurrent and attention-based architectures are better suited to exploiting temporal dependencies between successive excavation states. Xu et al. [31] compared Bayesian ridge regression, nearest-neighbour regression, random forest, gradient tree boosting (GTB), SVM, CNN, and LSTM for the prediction of advance rate, rotational speed, thrust and cutterhead torque. GTB

achieved the highest prediction accuracy among the statistical and ensemble approaches, while random forest provided the most favourable balance between predictive accuracy and computational time; LSTM also outperformed CNN for the sequential data examined. Importantly, the authors found that ensemble models remained highly competitive for the relatively limited dataset and emphasized that the advantages of deep neural networks depend on the availability of sufficient training data. Consequently, the evolution of ML in TBM engineering should not be interpreted as a simple replacement of conventional models by increasingly complex architectures. Rather, current research indicates that model selection must balance prediction accuracy, data availability, temporal structure, computational demand, interpretability, and generalization, with no single learning algorithm being universally optimal across different TBM types, geological environments, and engineering objectives. This conclusion is also consistent with cross-project ground-classification studies, which show that model performance is strongly dependent on-site conditions, input variables and the adopted geological classification system [32].

A critical comparison of these model families shows that increasing model complexity does not automatically lead to better TBM engineering performance. Conventional ML models such as regression, SVM, KNN and decision trees are useful when datasets are small, variables are well selected and interpretability or rapid deployment is important, but they may underrepresent strong nonlinear coupling among geology, machine settings and excavation response. Ensemble models such as random forest, gradient boosting, XGBoost, LightGBM, and CatBoost generally provide stronger nonlinear representation and feature-importance analysis for structured TBM records, although they can still overfit site-specific operating regimes if validation is not chronology- or project-aware. Deep-learning models are better suited to high-frequency sequential data because recurrent, convolutional, and hybrid architectures can learn temporal dependencies and hidden operating states, but their advantages depend on sufficiently large and clean datasets and they are often less transparent. Transformer-based approaches further improve long-range temporal and cross-variable representation, which is valuable for multivariate excavation-parameter forecasting, but they require careful data volume, computational and uncertainty considerations before they can be used as reliable operational decision tools.

Feature selection, model development and validation

The reliability of machine-learning models in TBM tunnelling depends not only on the choice of algorithm but also on the quality of data preparation, feature representation, and validation strategy. TBM monitoring systems generate large-scale heterogeneous datasets containing numerous operational, geological and machine-condition variables; however, these raw datasets commonly include machine stoppages, start-up and shutdown periods, missing observations, sensor noise, abnormal measurements, redundant variables, and data collected under different excavation states. Therefore, systematic preprocessing is required before model training, including identification of valid excavation periods, removal of unreliable records, filtering or smoothing of noisy

signals, normalization and construction of representative datasets. The selection of informative features is particularly important because incorporating excessive or highly correlated variables may increase computational complexity, reduce model interpretability and promote overfitting. Consequently, TBM studies have adopted both engineering-based and data-driven feature-selection strategies, including expert knowledge, correlation analysis, principal component analysis (PCA), statistical tests, regularization methods, and feature-importance evaluation from ensemble models. For example, TBM geological-recognition studies have demonstrated that dimensionality reduction and sensitive-feature selection can improve prediction efficiency while maintaining high classification accuracy, whereas large-scale TBM datasets have shown that appropriate feature extraction and selection are essential for developing reliable machine-learning frameworks [33]. More recently, researchers have investigated advanced feature-combination optimization approaches because using all available TBM variables does not necessarily improve prediction performance and may increase computational cost [34]. Following feature preparation, robust model development requires careful separation between model fitting and model generalization through appropriate training, validation and testing procedures. Training datasets are used for parameter estimation, validation datasets support model selection and hyperparameter optimization, while independent test datasets provide an unbiased assessment of predictive performance. This consideration is especially critical for TBM applications because excavation data are sequential and spatially correlated along the tunnel alignment; therefore, randomly dividing adjacent excavation cycles into training and testing sets may result in information leakage and overly optimistic accuracy estimates. For real-time prediction tasks, validation strategies should reproduce the actual deployment scenario by training models using previously available excavation data and evaluating predictions on subsequent rings or excavation sections. Recent TBM studies increasingly emphasize evaluation using unseen excavation data, independent sections, or chronology-preserving validation approaches to better assess model transferability and practical applicability [35]. Therefore, effective ML-based TBM systems require an integrated framework combining reliable preprocessing, meaningful feature selection, appropriate model development, and realistic validation to ensure that predictions remain accurate and applicable under changing geological and operational conditions.

ML Applications in TBM Tunnelling

ML for geological condition and rock-mass prediction

One of the most important applications of ML in mechanized tunnelling is the recognition and forecasting of geological conditions ahead of or around the TBM face. This application is based on the principle that different ground types generate distinguishable patterns in machine response. Parameters such as cutterhead thrust, torque, rotational speed, penetration rate, specific energy, field penetration index, torque penetration index, chamber pressure, and muck-related measurements have therefore been employed as indirect indicators of geological conditions. Ground-condition prediction is generally formulated as a classification problem in

which operational observations are mapped to geological classes obtained from boreholes, tunnel mapping, muck inspection, or rock-mass classification systems.

A wide variety of ML models have been explored for this purpose. ANN, KNN, SVM, CART, RF, AdaBoost-CART, XGBoost, LightGBM, CNN, LSTM, stacking ensembles, clustering algorithms and hybrid unsupervised-supervised frameworks have all been applied to different soil, rock and mixed-ground environments. Ensemble learning has attracted particular attention. For example, stacking classifiers have been used to combine predictions from several learners, whereas boosting methods sequentially improve weak learners and have frequently shown competitive geological-classification performance. More recent work has also moved from simple geological classification toward automatic identification coupled with construction decision support; Fu et al. developed an automatic strata-identification framework linked to decision support for multiple working conditions, illustrating this transition from recognition toward actionable TBM intelligence.

Nevertheless, the assumption that one highly accurate model can be transferred universally among projects is not supported by current evidence. Ayawah et al. [32] evaluated several previously proposed ML models using an independent tunnelling dataset and found that their performance generally decreased relative to the accuracy reported in their original applications. Their results indicate that model suitability depends on the ground conditions, TBM configuration, available variables and classification system, making site-specific model development and validation particularly important. This finding represents a critical issue for practical TBM automation: high accuracy within one project does not necessarily imply reliable performance when the model encounters another machine or geological environment.

ML for TBM performance and operational parameter prediction

ML for TBM performance and operational-parameter prediction focuses on estimating excavation quantities that directly affect productivity, machine loading and controllability rather than only identifying geological classes. Typical prediction targets include penetration rate, advance rate, field penetration index, cutterhead torque, thrust force, cutterhead rotational speed, chamber or face pressure, shield attitude, energy consumption, cutter wear indicators, and recommended parameter ranges for safe operation. The corresponding input variables usually combine geological descriptors such as rock-mass class, RMR, UCS, weathering, overburden, and groundwater condition with TBM operational signals such as thrust, torque, rotation speed, penetration or advance rate, specific energy, chamber pressure, screw-conveyor variables, grouting parameters and lagged time-series features. For example, Xu et al. [31] compared Bayesian ridge regression, nearest-neighbour regression, random forest, gradient tree boosting, SVM, CNN and LSTM for predicting advance rate, rotational speed, thrust and cutterhead torque, showing that ensemble methods remained highly competitive while LSTM was advantageous for sequential patterns. Recent systematic and generalization-focused studies further indicate that robust performance prediction depends not only on the selected algorithm but also on preprocessing, feature selection, data balance, project similarity, and validation design [36,37]. Therefore, operational prediction should be understood as a deployment-oriented task: a model must predict machine response accurately enough to support parameter adjustment while remaining interpretable, transferable, and stable under changing ground and operating conditions (Table 1).

Table 1: Comparative summary of representative ML and DL approaches for TBM applications.

Method Family	Typical Inputs	Typical Task/Dataset	Validation Strategy	Main Findings and Limitations
Conventional ML: regression, KNN, SVM, decision tree, Gaussian process, ANN	Geological descriptors, selected operational parameters, manually engineered features	Penetration or advance rate, torque, thrust, ground or rock-mass class	Train/validation/test split, k-fold CV, independent test sections when available	Efficient and interpretable for small to medium datasets, but performance is sensitive to feature design and nonlinear coupling.
Ensemble learning: RF, AdaBoost, GTB, XGBoost, LightGBM, CatBoost, stacking	Structured TBM monitoring variables, geological labels, feature-importance scores	Performance prediction, geological recognition, jamming or risk classification	Independent section testing, chronology-preserving splits, cross-project comparison	Often strong for nonlinear tabular TBM data and supports interpretability, but may learn project-specific operating regimes.
Deep learning: CNN, RNN, LSTM, GRU, autoencoder, hybrid CNN-LSTM	High-frequency time series, image or multimodal data, lagged operational sequences	Sequential forecasting, muck or face-image interpretation, anomaly detection	Hold-out temporal sequences, rolling prediction, unseen rings or tunnel sections	Captures temporal and high-dimensional patterns, but needs large clean datasets and careful explanation.
Transformer and attention-based models	Multivariate long-term excavation-parameter sequences and temporal-feature embeddings	Long-range forecasting of TBM operational parameters and multi-output prediction	Temporal hold-out, long-horizon forecasting, comparison with recurrent baselines	Useful for long-range dependencies and cross-variable interaction, but computationally demanding and data hungry.

Semi-supervised, unsupervised, and adaptive learning	Sparse geological labels plus abundant unlabelled TBM data; streaming operational data	Ground-condition characterization, latent-state discovery, drift-aware prediction	Label-scarcity tests, incremental updating, project-transfer assessment	Promising for real construction data where labels are limited, but requires reliable quality control and drift monitoring.
Optimization, reinforcement learning, and digital-twin-linked models	Predicted states, machine constraints, safety thresholds, energy and wear indicators	Parameter recommendation, multi-objective control, closed-loop decision support	Simulation, field case evaluation, constraint checking, human-supervised validation	Moves beyond prediction toward decision support, but practical use depends on safe constraints and engineering supervision.

Safety, deformation and anomaly detection

Beyond productivity improvement, ML is increasingly being applied to risk-sensitive TBM applications, including settlement and deformation prediction, tunnel crown convergence, abnormal operating-state detection, collapse warning, and jamming-risk assessment. Recent deformation studies have coupled TBM-generated variables with optimized tree-based models; for example, PSO-XGBoost variants have been used to predict tunnel deformation and support TBM jamming response, showing how deformation forecasting can be linked to practical risk mitigation [38]. Interpretable ML approaches for TBM tunnel crown convergence further demonstrate the value of combining Bayesian optimization or gradient-boosting models with SHAP analysis so that predicted deformation can be related to influential geological and construction variables [39]. For jamming and complex geological risk, cross-project databases and ensemble models have been used to predict TBM performance and combined jamming risk in Himalayan geology, while SHAP analysis identifies the contribution of thrust, torque, penetration response and rock-mass conditions to risk prediction [40]. These studies indicate that safety-oriented TBM ML should not be evaluated only by accuracy; it also requires early-warning capability, interpretability, uncertainty awareness, and validation on unseen geological sections.

From Offline Prediction to Adaptive Machine Learning

Most existing TBM machine-learning studies are developed using offline datasets collected from completed projects; however, real tunnelling environments are dynamic, with continuously changing geological conditions, evolving machine responses, and potential data distribution shifts. These limitations highlight the need for adaptive and online learning strategies that can update models as new excavation data become available. Mostafa et al. emphasized data drift and unseen geological conditions as major barriers to real-time ML deployment and identified incremental learning as a promising solution for adaptive TBM prediction [13]. Semi-supervised learning has also attracted increasing attention because TBMs generate abundant unlabelled operational data but limited geological labels, enabling models to exploit both labelled and unlabelled information for improved generalization [24]. Recent systematic and generalization-focused literature further shows that preprocessing, smoothing, data balance, recurrent time-series modelling, cross-project validation, and incremental updating are central to robust TBM performance prediction

[36,37,41]. Accordingly, adaptive ML should be developed as a monitored updating process rather than a one-time offline modelling exercise, with explicit checks for data drift, prediction confidence, and engineering plausibility.

AI-Based Operational Optimization and Control

After predicting geological conditions and TBM responses, the next stage of intelligent tunnelling is to use AI for operational optimization and control, where model outputs are transformed into recommended machine actions. Traditional TBM parameter adjustment mainly depends on operator experience, whereas AI-based approaches aim to optimize variables such as thrust, torque, cutterhead speed, penetration rate, and chamber pressure by considering excavation efficiency, machine constraints and ground stability simultaneously [21]. Early studies developed ML-based decision-support systems that learned relationships between TBM measurements and optimal operating states to provide adaptive parameter recommendations [42]. More recent approaches have combined predictive models with optimization algorithms, multi-objective strategies, and physics-informed learning to achieve safer and more efficient parameter selection under changing geological conditions [43]. Reinforcement-learning-based frameworks have further advanced this concept by enabling agents to learn optimal operational policies through interaction with simulated or data-driven environments, showing potential for adaptive control of excavation performance, cutter maintenance and machine stability [44]. However, practical implementation still requires reliable constraints, sufficient training data and integration of engineering knowledge to ensure that optimized decisions remain physically feasible and operationally safe.

Digital Twins and Closed-Loop Intelligent Tunnelling

Digital twin (DT) technology represents a further evolution of intelligent TBM systems by integrating real-time sensing, data analytics, virtual modelling, and decision-making within a unified cyber-physical framework. Unlike standalone ML models that provide predictions, DTs enable continuous synchronization between the physical TBM and its virtual representation, allowing monitoring, simulation, optimization, and feedback-based control throughout the excavation process [45]. Recent studies have combined DTs with ML models to predict TBM performance, visualize construction states and support real-time operational optimization using continuously updated machine data. Zhang et al. [46] developed a DT-enabled TBM control framework integrating

deep learning and online multi-objective optimization to improve excavation performance while considering efficiency, safety, energy consumption, and tool wear. More advanced frameworks are now incorporating reinforcement learning and physics-informed models within DT environments to achieve adaptive control under changing geological conditions [47]. Despite these advances, practical implementation remains challenging due to difficulties in multi-source data integration, real-time model updating, virtual-physical synchronization, uncertainty management, and ensuring reliable closed-loop operation in complex underground environments [48].

Machine Learning in Tunnelling: Capabilities, Challenges, and Future Research Directions

Despite substantial progress, the transition from accurate ML models to dependable engineering systems remains incomplete. Three interrelated limitations are particularly important. First, data scarcity and imbalance remain widespread because abnormal geological or hazardous conditions occur much less frequently than normal excavation states, while project databases are rarely standardized or openly shared. Second, cross-project generalization remains difficult because TBM dimensions, sensor configurations, operator strategies, geological conditions and data distributions vary substantially among projects. Third, many high-performing ML and deep-learning approaches remain effectively “black-box” models and provide deterministic point predictions without explicitly quantifying the uncertainty associated with highly variable underground conditions. These issues are particularly consequential in tunnelling, where an inaccurate but apparently confident prediction may have direct implications for machine safety and ground stability.

Accordingly, recent research is moving toward transfer learning, online learning, interpretable ML, uncertainty-aware prediction, physics- or knowledge-informed learning and hybrid data-driven/ mechanistic models. The literature already demonstrates knowledge-driven transfer learning for cross-project TBM performance prediction, while recent research identifies interpretability, data integration and computational efficiency as central requirements for intelligent TBM systems. The longer-term objective is therefore not simply to obtain a model with a higher R^2 or classification accuracy, but to establish a reliable perception-prediction framework capable of learning from continuously arriving TBM data, recognizing changes in geological and operational states, quantifying predictive confidence, and providing trustworthy information to subsequent optimization and control modules. This evolution forms the conceptual bridge from machine learning-based TBM prediction to intelligent and increasingly autonomous tunnelling, including reinforcement learning and digital-twin-enabled closed-loop control systems. Recent developments in digital-twin-assisted deep learning and reinforcement learning already demonstrate this transition from offline prediction toward real-time operational control.

A broader lesson from recent civil- and structural-engineering modelling studies is that data-driven prediction must be coupled

with mechanics-based constraints, soil-structure interaction, and robustness considerations when it is used to support safety-critical infrastructure decisions [49,50]. Although these studies are not TBM-specific, they reinforce the need for future TBM AI systems to combine operational data with physical constraints and resilience-oriented decision rules rather than relying on black-box prediction alone.

Conclusion

This mini-review establishes that AI and ML have already become effective tools for TBM geological interpretation, performance forecasting, anomaly recognition, safety assessment, operational optimization and digital-twin-enabled decision support. The established evidence shows that TBM monitoring data can reveal machine-ground interaction and that conventional ML, ensemble learning, deep learning, and attention-based models each have useful but context-dependent roles. At the same time, practical deployment remains limited by data quality, label scarcity, site-specific generalization, interpretability, uncertainty quantification and safe integration with operator decisions. Future research should therefore prioritize adaptive and physics-informed models, cross-project validation, uncertainty-aware prediction, multimodal sensing, and human-supervised closed-loop control. Progress toward autonomous tunnelling should be gradual and safety constrained, using AI to strengthen expert decision-making rather than replacing engineering judgement.

References

1. Nie Y, Zhang Q, Hou L, Du L, Zhao X, et al. (2025) TBM rock mass classification using XGBoost and interpretable machine learning. *Advanced Engineering Informatics* 66: 103459.
2. Nie Y, Zhang Q, Du Y, Du L (2025) TBM jamming risk prediction method based on fuzzy theory and Bi-LSTM. *Tunnelling and Underground Space Technology* 157: 106329.
3. Zhang Q, Nie Y, Zhao W, Du L (2025) Research on TBM parameter optimization based on failure probability. *Engineering Failure Analysis* 167: 109036.
4. Hou S, Liu Y, Yu J, Zhang R, Cheng L, et al. (2025) Intelligent decision-making for TBM tunnelling control parameters using multi-objective optimization. *Journal of Rock Mechanics and Geotechnical Engineering* 17(5): 2943-2963.
5. Qian Zhang, Yan Ma, Yanliang Du, Lijie Du, Minyuan Wang, et al. (2025) Investigation into the collapse height of TBM jamming machinery induced by fault fracture zones and the verification of applicability. *Tunnelling and Underground Space Technology* 155(Part 1): 106196.
6. Jinhui Li, Dong Guo, Zuyu Chen, Xu Li, Zhaofeng Li, et al. (2024) Transfer learning for collapse warning in TBM tunneling using databases in China. *Comput Geotech* 166: 105968.
7. Yagiz S (2008) Utilizing rock mass properties for predicting TBM performance in hard rock condition. *Tunnelling and Underground Space Technology* 23(3): 326-339.
8. Farrokh E, Rostami J, Laughton C (2012) Study of various models for estimation of penetration rate of hard rock TBMs. *Tunnelling and Underground Space Technology* 30: 110-123.
9. Qianli Zhang, Zhenyu Liu, Jianrong Tan (2019) Prediction of geological conditions for a tunnel boring machine using big operational data. *Automation in construction* 100: 73-83.

10. Hu Shi, Huayong Yang, Guofang Gong, Lintao Wang (2011) Determination of the cutterhead torque for EPB shield tunneling machine. *Automation in Construction* 20(8): 1087-1095.
11. Samadi H, Hassanpour J (2024) EPB-TBM cutterhead torque and thrust modelling in rock tunnels through an analytical method and TSFS model. *Heliyon* 10(11): e31773.
12. Shahrouz I, Zhang W (2021) Use of soft computing techniques for tunneling optimization of tunnel boring machines. *Undergr Space* 6(3): 233-239.
13. Mostafa S, Sousa RL, Einstein HH (2024) Toward the automation of mechanized tunneling "exploring the use of big data analytics for ground forecast in TBM tunnels". *Tunnelling and Underground Space Technology* 146: 105643.
14. Rostami J (2016) Performance prediction of hard rock Tunnel Boring Machines (TBMs) in difficult ground. *Tunnelling and Underground Space Technology* 57: 173-182.
15. Paul R, Mishra S, Khatti J (2025) Role of Artificial Intelligence (AI) techniques in tunnel engineering-a scientific review. *Indian Geotechnical Journal* 56: 4760-4790.
16. Khalid S, Khalil T, Nasreen S (2014) A survey of feature selection and feature extraction techniques in machine learning. *Science and Information Conference*, pp. 372-378.
17. Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J, et al. (2020) A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends* 1: 56-70.
18. Arbi SJ, Rehman ZU, Hassan W, Khalid U, Ijaz N, et al. (2025) Optimized machine learning-based enhanced modeling of pile bearing capacity in layered soils using random and grid search techniques. *Earth Science Informatics* 18: 332.
19. Xu X, Liang T, Zhu J, Zheng D, Sun T, et al. (2019) Review of classical dimensionality reduction and sample selection methods for large-scale data processing. *Neurocomputing* 328: 5-15.
20. Khalid MS, Rehman ZU, Alshameri B, Maqsood Z, Hussain F, et al. (2025) Application of multiple machine learning algorithms for intelligent prediction of the strength of fine-grained natural soils. *Arabian Journal of Geosciences* 18: 115.
21. Loy Benitez J, Song MK, Choi YH, Lee JK, Lee SS, et al. (2024) Breaking new ground: Opportunities and challenges in tunnel boring machine operations with integrated management systems and artificial intelligence. *Automation in Construction* 158: 105199.
22. Marcher T, Erhartner GH, Winkler M (2020) Machine learning in tunnelling-capabilities and challenges. *Geomechanics and Tunnelling* 13(20): 191-198.
23. Ahuchogu MC (2025) Real-time image-based data processing and its applications in managerial decision-making and risk analysis. *Eksplorium: Bulletin of the Center for Nuclear Mineral Materials Technology* 46: 1552-1565.
24. Yu H, Mooney M (2023) Characterizing the as-encountered ground condition with tunnel boring machine data using semi-supervised learning. *Comput Geotech* 154: 105159.
25. Samie MA (2026) Ensemble deep learning: A State-of-the-Art comprehensive review. *Journal of Information System Exploration and Research* 4(1): 9-30.
26. Hafiz AM, Parah SA, Bhat RUA (2021) Attention mechanisms and deep learning for machine vision: A survey of the state of the art. *arXiv* 2106-07550.
27. Xuan Nam Bui, Hoang Nguyen, Hai An Le, Hoang Bac Bui, Ngoc Hoan Do, et al. (2020) Prediction of blast-induced air over-pressure in open-pit mine: Assessment of different artificial intelligence techniques. *Natural Resources Research* 29: 571-591.
28. Nugroho S (2026) A stacking ensemble learning framework for network intrusion detection using XGboost, LightGBM, CatBoost and Adaboost with random forest meta model. 2026 International Conference on Communication Networks and Machine Learning (CNML) IEEE, pp. 20-23.
29. Jian Bin Li, Zu Yu Chen, Xu Li, Liu Jie Jing, Yun Pei Zhang, et al. (2023) Feedback on a shared big dataset for intelligent TBM Part I: Feature extraction and machine learning methods. *Undergr Space* 11: 1-25.
30. Mingjun Liu, Jianqin Liu, Wei Guo, Hongxu Liu, Xiao Guo, et al. (2025) Multi-patch attention Transformer for multivariate long-term time series forecasting of TBM excavation parameters *Undergr Space* 23: 285-306.
31. Xu C, Liu X, Wang E, Wang S (2021) Prediction of tunnel boring machine operating parameters using various machine learning algorithms. *Tunnelling and Underground Space Technology* 109: 103699.
32. Prosper EA Ayawah, Sylvanus SN, Jessica WA Azure, Azupuri GA Kaba, et al. (2022) A review and case study of artificial intelligence and machine learning methods used for ground condition prediction ahead of tunnel boring Machines. *Tunnelling and Underground Space Technology* 125: 104497.
33. Zhang Q, Yang K, Wang L, Zhou S (2020) Geological type recognition by machine learning on in-situ data of EPB tunnel boring machines. *Mathematical Problems in Engineering* 2020: 1-10.
34. Zhang CH, Wang Y, Li X (2026) Active learning for optimal feature combination selection in machine learning of TBM operational data. *Automation in Construction* 184: 106826.
35. Morteza Abbasi, Amir Hossein Namadchi, Mehdi Abbasi, Mohsen Abbasi (2024) Evaluation of machine learning algorithms in tunnel boring machine applications: A case study in Mashhad metro line 3. *International Journal of Geo-Engineering* 15(28).
36. Shan F, He X, Xu H, Armaghani DJ, Sheng D, et al. (2023) Applications of machine learning in mechanised tunnel construction: A systematic review. *Eng* 4(2): 1516-1535.
37. Huang S, Korfiatis G, Sousa R (2025) Generalization challenges and strategies in tunnel boring machine performance prediction. *Civil Engineering Design* 7(2): 63-84.
38. Yin Bo, Xiaogang Guo, Quansheng Liu, Yucong Pan, Lei Zhang, et al. (2024) Prediction of tunnel deformation using PSO variant integrated with XGBoost and its TBM jamming application. *Tunnelling and Underground Space Technology* 150: 105842.
39. Hu J, Li Y, Wang H, Zhang Y, (2025) Interpretable machine learning approach for TBM tunnel crown convergence prediction with Bayesian optimization. *Frontiers in Earth Science* 13: 1608468.
40. Katuwal TB, Panthi KK, Basnet CB (2026) Machine learning approach for prediction of TBM performance and risk of jamming in Himalayan geology using a cross-project tunnelling database. *Scientific Reports* 16(1): 4094.
41. Shan F, He X, Armaghani DJ, Sheng D (2024) Effects of data smoothing and recurrent neural network (RNN) algorithms for real-time forecasting of tunnel boring machine (TBM) performance. *Journal of Rock Mechanics and Geotechnical Engineering* 16(5): 1538-1551.
42. Garcia GR, Michau G, Einstein HH, Fink O (2021) Decision support system for an intelligent operator of utility tunnel boring machines. *Automation in Construction* 131: 103880.
43. Lin P, Wu M, Xiao Z, Tiong RL, Zhang L (2024) Physics-informed deep reinforcement learning for enhancement on tunnel boring machine's advance speed and stability. *Automation in Construction* 158: 105234.

-
44. Hansen TF, Erharder T Marcher (2024) Towards reinforcement learning-driven TBM cutter changing policies. *Automation in Construction* 165: 105505.
 45. Li T, Li X, Rui Y, Ling J, Zhao S, et al. (2024) Digital twin for intelligent tunnel construction. *Automation in Construction* 158: 105210.
 46. Latif K, Sharafat A, Seo J (2023) Digital twin-driven framework for TBM performance prediction, visualization, and monitoring through machine learning. *Appl Sci* 13(20): 11435.
 47. Zhang L, Guo J, Fu X, Tiong RLK, Zhang P (2024) Digital twin enabled real-time advanced control of TBM operation using deep learning methods. *Automation in Construction* 158: 105240.
 48. Liu G, Tang Y, Zhang H, Li R, Wang H, et al. (2025) Advances in artificial intelligence and digital twin for tunnel boring machines. *Artificial Intelligence Review* 58(2025): 257.
 49. Kılıçer S, Özgan K (2026) Evaluation of progressive collapse risk in reinforced concrete structures considering soil-structure interaction via the Winkler model. *Structural Engineering International*, pp. 1-11.
 50. Kılıçer S, Kılıçer ST (2026) Progressive-collapse robustness assessment of existing reinforced concrete buildings in diverse geographical regions using the tie-force method. *Buildings* 16(5): 1090.