DOI: 10.3217/978-3-99161-057-1-044

'Machine Learning – Based Data Interpretation and Visualization for Tunnel Monitoring: A Case Study of Changshui Airport Tunnel'

Asif Ahmed Sourov[1], Yadong Xue[1], Yongfa Guo[2], Wenyun Ding[2], Jinjing Yang[2]

[1]Department of Geotechnical, College of Civil Engineering, Tongji University, 1239 Siping Road, Shanghai – 200092, China

[2]China Railway Eryuan Engineering Group Kunming Survey and Design Research Institute Co., Ltd, Kunming, Yunan, China

Email: asifahmedsourov@icloud.com, yadongxue@tongji.edu.cn, 524965753@qq.com, 592938576@qq.com, 357080499@qq.com

Abstract: Specially, when it comes to such risky construction projects as the Changshui Airport Railway Tunnel, the ground settlements surrounding the structure will be observed and evaluated. This case study is a combination of advanced machine learning algorithms, which are augmented with MATLAB to reinterpret, visualize and analyze settlement trends based on real – world tunnel monitoring data. The research starts with some of the time – series data that had been registered in the tunnel (settlement, factors that impact settlement, temperature, etc.). K – means clustering and hierarchical clustering are used to classify the settlement patterns and the clustering result indicates the difference in settlement monitoring points. Finally, we have used the feature importance analysis to explore the most significant factors that affect settlement decisions and to know more about the settlement processes in tunnels. The discussion of the Random Forests, Gradient Boosting and Artificial Neural Networks regression models is provided to predict settlement patterns to enable predictive risk. Heatmaps, time – series graphs and scatter plots are some of the comprehensive visualizations constructed to convey the discovery and help in decision making. The indicators to assess the model performance are R^2, RMSE, MAE and the findings show how to forecast settlements in the most optimal manner. Besides presenting the utility of machine learning tunnel surveillance, the case study also provides data driven decision making framework in underground engineering projects.

Keywords: Tunnel Monitoring, Ground Settlements, Machine Learning, Data Visualization, Changshui Airport Tunnel, Underground Infrastructure

1. Introduction

Tunnel monitoring is a highly significant element of safety and life of underground infrastructure, especially in high – stakes projects, like airport tunnels(Bao et al., 2018). The geotechnical and environmental forces that control the settlement behavior of these structures are complex and may undermine the structural integrity of these structures(Muhammed et al., 2019). Monitoring is thus essential in the realization of the impacts and of the excavation works on the surrounding ground and structural stability. The conventional approaches, however, despite their usefulness, are time – consuming and prone to errors because they involve a lot of manual analysis that may need human interpretation(Hua et al., 2021). Monitoring systems generate a lot of data in the course of a tunnel project life cycle because of the technology development(Wang et al., 2020). It is also difficult to analyze such datasets to draw conclusions, identify trends and discover impact factors. Nevertheless, the existing practices fail to capture this information in a manner that would allow proactive maintenance and safety decision – making.

The given Case study examines the application of machine learning approaches that have been adopted in the MATLAB environment to solve the above – mentioned issues. Using the data of Changshui Airport Tunnel Monitoring, it aims to interpret, visualize and analyze settlement behavior using clustering algorithms, feature importance analysis and regression models, thus determining patterns, whether to fill or settle and which factors have a significant impact on ground activity. The introduction of machine learning based systems into the tunnel monitoring process is a paradigm shift, since such methods enable quicker and more precise analyses due to automation of the extraction of complex datasets(Jin-miao et al., 2022). The results of the study can thus be used to make improved decisions, improve predictive maintenance and support the safety and efficiency of underground structures. The power of this study shows its ability to use the existing computational tools to produce engineering solutions that can have a significant real – life effect.

2. Settlement Monitoring and Prediction

Figure 1 is a summary of the settlement monitoring and prediction process, its challenges and developments. Settlement monitoring is an important aspect of safety and life of underground infrastructure in high-risk projects like airport tunnels(Jin-miao et al., 2022). The construction induced behaviors such as boring induced stresses, change in ground water level and geological heterogeneities are some of the causes of ground settlement(Ayasrah et al., 2020). These deformations can be hazardous and in case they are not identified or misinterpreted, they can result in structural instability, which can result in catastrophic failures(Ayasrah et al., 2020). Monitoring of landslides is achieved by surface and subsurface movements which are traditionally monitored by geodetic surveys, inclinometers and extensometers. Figure 1 represents the key areas of monitoring of settlements, including the traditional and contemporary approaches, and the utilization of the advanced instruments like machine learning and visualizations, to analyze large amount of data. It has been increasingly possible to obtain continuous deformation data, as newer forms of instrumentation have been developed well, including automated total stations, satellite-based InSAR (interferometric synthetic aperture radar) and fiber optic sensors, all modern(Karamvasis & Karathanassi, 2020).

DOI: 10.3217/978-3-99161-057-1-044

Monitoring therefore plays a role beyond detection to mechanisms of settlement(Martins et al., 2020).

2.1. Machine Learning in Civil Engineering

Machine learning (ML) is easy to use and understand complex data and predict structural behaviors in civil engineering(Marsella & Scaioni, 2018). Various clustering techniques (E.g.: K-means clustering, hierarchical clustering, etc.) are increasingly being used in settlement monitoring to classify settlement patterns as well as to detect spatially heterogeneous risks(Vadyala et al., 2021). To illustrate, showed that clustering had the potential to significantly increase the ability to distinguish settlement behavior at up to eight monitoring sites and could be applicable in the implementation of sector-specific mitigation measures(Zhou et al., 2020). The most important machine learning techniques of settlement monitoring are clustering and regression models, which allow identifying patterns and predicting trends in ground deformation, as illustrated in Figure 1. All regression models, including RF, GB, and ANN, had better predictive performance and modeled settlement trend than statistical methods(Jin-miao et al., 2022). This type of models is particularly applicable when there are non-linear relationships and high dimensional data as is the case in geotechnical applications. Other related fields where ML can be applied, i.e. SHM and anomaly detection(Mousavi & Beroza, 2022). The former includes clustering methods which have been applied to identify deformation anomalies in bridge structures and regression models are widely applied to predict loads in buildings(Jasmine & Arun, 2021).

2.2. Advanced Data Integration and Visualization

The capability to successfully integrate the heterogeneous data, which are position, deformation, temperature and environmental parameters in a single effective mechanism has been one of the most remarkable issues in tunnel monitoring(Zhao et al., 2021). MATLAB, which has a vast array of tools to conduct machine learning and visualization, is a perfect candidate to conduct such workflows(Ma et al., 2021). Settlement dynamics can be shown in time-series plots, and areas with high-risk levels can be shown in heatmaps(Yan et al., 2019). Together with clustering and regression analyses, these visualizations may give a complete image of the behavior of ground and in responsivity.

2.3. Research Gap and Addressing

Although the ML techniques have shown promising progress in geotechnical engineering, studies have been mainly focused on specific tasks(Marcher et al., 2020). As far as the authors are aware, few studies have attempted to develop a coherent framework to interpret and visualize settlement data that include clustering, feature importance analysis, and regression modeling(Chen et al., 2022). The literature also shows evidence of the need of domain-specific adaptations in ML algorithms. Or, feature engineering based on site condition, geological and environmental driven feature engineering could be used in settlement prediction models(Fan et al., 2019). These methods, however, offer more access and utility of data, yet there is no standard workflow to those adjustments, and thus, the overall use of these models becomes cumbersome(Merghadi et al., 2020).

We suggested a machine learning pipeline that included clustering, feature importance, regression, and application of that regression with clustering to provide a general analysis framework (as illustrated in Figure 1) to be used in tunnel monitoring.

3. Data Preprocessing

Monitoring of settlement is an important process in the safety of high-risk infrastructure projects like tunnels during construction and long-term prevention and remedial maintenance of the infrastructure after construction(Tan et al., 2019). These techniques offer optimal solutions to capture high-resolution spatial-temporal trends but the traditional settlement monitoring methods (geodetic surveys, inclinometers) are time-consuming, subject to human error, and do not capture such trends(Wang et al., 2020). In this section, a step-by-step data-driven framework of the analysis of the settlement behavior in the Changshui Airport Tunnel will be discussed. Figure 2 gives a general description of the proposed framework.

3.1. Data Preprocessing

The pre-processing of data is performed to clean, standardize and prepare the data to be utilized in future machine learning activities. Using MATLAB workspace, we can see that our original data has 13 columns and 10,704 rows. The notable variables are dated Measurement Time, Cumulative Settlement, Relative Settlement, Settlement Rate, Geological Grade, and Distance from Start. The selection of these variables reflects the time and space aspects of the settlement monitoring which constitutes the basis analysis of any type of settlement monitoring.

a. Standardization of Time Intervals: The time interval of data collection is not of regular nature (Time Interval Days) and is recorded in Measurement Time. Time matching functions in MATLAB where timestamps are converted to fixed equidistant time (e.g. 0.5 days, 1 day) to provide uniform time-based monitoring of settlement patterns.

b. Missing Values: Handling Missing Values in Missing Values: Interpolated Settlement Rate and Relative Settlement with the fill missing () function in MATLAB. Linear interpolation maintained the trends of time in the data to avoid bias.

c. Normalization of Variables: Cumulative Settlement, Settlement Rate, and Geological grade were normalized by the min-max scaling by using the normalize () function of MATLAB. This process minimizes the differences between the applications of different systems that improve the performance of the machine learning models.

d. Detect and Remove Outliers: The sensor errors and environmental disturbances that caused the sudden anomalies in the settlement rates were detected by using the Z-scores to detect statistical outliers and examine sudden spikes by plotting time-series data.

DOI: 10.3217/978-3-99161-057-1-044

3.2. Clustering Analysis

Settlement behaviors were used to mine, which was done through clustering analysis. This step exposes spatial settlement patterns and points out areas of intervention. Clustering Techniques:

a. K-Means Clustering: The variables such as Settlement Rate, Cumulative Settlement, Geological Grade were used as inputs. Elbow Method was used to determine the best number of clusters using within-cluster sum of squares (WCSS). The MATLAB assigned the monitoring points to various clusters using the K – means () function.

b. Hierarchical Clustering: The results of K-means were confirmed through the hierarchical clustering which provided a hierarchical view of settlement behaviors. The hierarchical relationships between the settlement points were visualized with the help of the linkage () function in MATLAB that produced a dendrogram.

c. Clustering Results: Cluster 1: Monitoring points outside structural elements (i.e. tunnel border), i.e. the address which was above the maximum settlement value; Cluster 2: intermediate settlement rates -> transitional zones; Cluster 3: Stable regions with thin layers.

3.3. Feature Importance Analysis

The feature importance analysis shows the most important factors in the determination of settlement behavior(Oh et al., 2021). This procedure increases the interpretability of the predictive models and follows the geotechnical principles. The following is the outline of feature importance analysis:

a. Supervised Learning Models: To rank predictor variables Random Forest (RF) and Gradient Boosting (GB) models were applied. Dependent Variables: Settlement Rate, Cumulative Settlement Predictor variables: Geological Grade, Distance from Start, Monitoring Point Elevation and Time Interval Days

b. Feature Importance Ranking: The RF and GB models provided feature importance scores of all the predictors. The strongest features were: Geological Grade: The main factor that affects settlement, which means the effect of soil properties; Distance from Start: Settlement is not evenly distributed along the length of the tunnel; Time Interval Days: Settlement behavior varied greatly depending on the time of year.

3.4. Predictive Modeling

Predictive modeling was also carried out to forecast settlement patterns in order to carry out proactive maintenance. It entailed a split of the data into 80 percent training and 20 percent test and the models were ranked based on their capacity to predict Cumulative Settlement.

a. Training and Testing: Single supervised machine learning models were trained and tested: Random Forest (RF): It is interpretable and robust; Gradient Boosting (GB): Very accurate iterative model; ANNs: Can capture potential non-linear relationships and required tuning up; The robustness was achieved by cross-validation.

b. Evaluation Metrics: Models were tested on R^2: R^2 is a measure that compares the predicted values with the actual values in the training data; RMSE (Root Mean Square Error): How large is the error of the prediction; MAE (Mean Absolute Error): A measure to normalize the error of the prediction.

4. Graphical Analysis and Data Interpretation

Some of the visual outputs and the interpretation applied in this section include Feature Importance, Predictive Modeling, and other statistics that are used to monitor and predict on settlements. Dataset Summary: The target variable Cumulative Settlement was initially analyzed statistically and then feature importance analysis was done. The most important statistics are:

Rows in Dataset: 10,704; Missing Values: 841 (preprocessing step); Minimum Settlement Value: 0; Median Vela Value: 14.1412; Settlement Value Max: 2,064.6; Average Settlement Value: 940.1619; Standard Deviation: 1,019.7

Following data cleaning, 9863 rows remained to be used in train and evaluation. Such preprocessing steps were followed to ensure data reliability in the execution of the following model.

4.1. Analysis of Feature Importance

The results are displayed graphically with measures Table 1. Numerical Insights from ANN Feature Importance.. Figure 3 below is a bar chart that indicates the importance of each feature (e.g. Time Interval Days, Distance from Start, Monitoring Point Elevation, Relative Settlement) to the prediction of settlement rates using Random Forest (RF) model. As we observe, Monitoring Point Elevation is the most influential predictor with an importance of 50, which dwarfs the other features. Monitoring Point Elevation remains the most influential predictor in the List of Features in Gradient Boosting with an importance by value of more than 10,000 as in Figure 4. With the permutation-based importance, the ANN indicates that the most important predictor of settlement rates is Monitoring Point Elevation with an importance score of approximately 1,000 as indicated in Figure 5. The importance score of Monitoring Point Elevation is overwhelmingly high (1,028.8246). The architecture of ANN employed in the prediction of settlement in Changshui Airport Tunnel is shown in Figure 6. The model has an input layer that has the four main predictors of interest, which are Time Interval Days, Distance from Start, Monitoring Point Elevation, and Relative Settlement, which cover the effects of time, distance, and elevation on settlement, and the effects of relative settlement.

Random Forest Figure 3 – Monitoring Point Elevation is the most important, and all others (Time Interval Days, Distance from Start, Relative Settlement) are practically 0. Gradient Boosting Figure 4 - The importance score of Monitoring Point Elevation is about 10,000, which again confirms its importance. ANN Figure 5 The permutation-based analysis of ANN shows that Monitoring Point Elevation has an extremely high importance score of 1,028.8246, which is far higher than any other feature. Temporal (Time Interval Days) and spatial (Distance from Start) features were of moderate

DOI: 10.3217/978-3-99161-057-1-044

importance in all models: Random Forest Figure 3 – Time Interval Days and Distance from Start are rather insignificantly contributing yet still distinguishable compared to Relative Settlement. Gradient boosting Figure 4 in this case, these features have small but distinctly visible impact.

4.2. Analysis of Model Performance Metrics

Figure 7 shows that the random forest (RF) model possesses a good prediction power (R^2 close to 0.90). Similarly, the Root Mean Square Error (RMSE) is also very small, implying that the difference between the actual value and the predicted value is less. Thus, we can say that RF is very efficient when it comes to this dataset. The MAE (Mean Absolute Error) indicates the average absolute error between the predicted and actual values and as it can be seen, it is very low and hence RF can make accurate predictions. In Figure 8, Gradient Boosting (GB) has similar R^2 values (~0.90) to RF, which means that it performs well in predicting.

In Figure 9 it reflects the data non-linearities more than RF and GB since the third ANN model has the greatest degree of accuracy in the terms of R^2. RMSE of ANN is slightly greater than GB but in general ANN is more able to capture the patterns but that can also be a sign of ANN overfitting or high variance in localized predictions. The grouped bar chart Figure 10 in Figure 7, Figure 8, Figure 9 gives a summary of the performances of the model in terms of R^2, RMSE, and MAE of the RF, GB, and ANN. The three models are remarkable in accuracy (R^2 close to 0.90). However: ANN does get a bit better in R^2, so it is now the most successful model in the sense of understanding complex interactions. The RMSE of GB is lower than other models, which means that the predictions of GB are better in magnitude, especially in the pleasant regions. These results in Figure 10 showed that the choice of the model depends on the application, ANN is better to work with highly non-linear data, GB is the preferred method to minimize catastrophic errors, and that RF is a universal and interpretable solution.

4.3. Residual Analysis

The residual plot of RF Figure 11 indicates that the residuals are between -60 and +60. The majority of them are clustered at the zero line, which shows that the accuracy of many of the predictions was reasonable, but the dispersion of residuals, particularly at higher values, provides a clue as to how this model performed poorly on some of the cases where the predictions were relatively too distant to the actual value. Figure 12 The range of residuals is -2 to +2 in Gradient Boosting, which is much narrower, showing much higher precision compared to RF. The Figure 13 ANN shows the least range of residuals, -1.2 to +0.2 residuals that are tightly clustered around zero. This means that it is the most accurate in predicting among the models, and it has the fewest errors and excellent generalization abilities. Based on Figure 14 these findings are pointing out that ANN is the most effective in capturing complex relationships in the data and therefore it would emerge as the most accurate model in predicting settlement. Combined Residual Plot - RF, GB and ANN The combined residual plot displays residuals of RF, GB, and ANN. RF has the broadest scope of residuals between -60 and +60, thus performing the worst in the minimization of prediction error. ANN residuals are the closest to each other, and they are well within the range of -1.2 to +0.2, which is exceptionally precise.

4.4. Settlement Risk Analysis

The Actual Settlement and the RF, GB, and ANN model predictions of spatial settlement risks are shown in Figure 15 below. Actual Settlement (Top Left) The baseline comparison is provided by this heat map because it indicates the settlement values that actually took place. It demonstrates the actual pattern of settlement risk by subterrains as reflected by the data recorded in Figure 15. However, at the higher risk regions at Figure 15, there are minute differences. This means that RF will only capture the overall trend and will not be in a position to follow the minor details, particularly in cases where the values of settlements are extreme. GB Predicted Settlement (Bottom Left) The gradient boosting heat map would see it to be in good fitness to real settlement. ANN Predicted Settlement (Bottom Right): ANN is the most accurate model among RF and GB and it gives the closest resemblance to the actual settlement. All the heatmaps show that ANN gives the most accurate approximation of spatial settlement risks, then Gradient Boosting, and then Random Forest. It also agrees with the results that were obtained previously that ANN is more accurate overall, especially on complex data, GB is moderately accurate; RF is less accurate in extreme cases as indicated in Figure 15.

4.5. Error Distribution Analysis

Figure 16 Error Histogram (RF, GB, ANN) The histograms of RF, GB, and ANN give a detailed statistic of the accuracy errors of the three models. The Errors variance is very broad in RF & a range of approximately -60 ~ +60 is essentially eminent. When the value of minimum is high, it means that the model is not as precise as other models. The GB histogram range is lower, with a range of mostly between -2 and + 2, which means that it minimizes errors and is more stable in performance. The distribution of errors is the most concentrated in ANN and the errors are near -0.01 and +0.01 and this indicates that ANN predicts settlement values more accurately than other networks. Figure 17 Grid density plot of Errors RF, GB and ANN The mixed density plot indicates the variations in the distribution of the various errors of these three models to indicate the difference in performance. This Figure 17 graph therefore reinstates the ability of ANN to generalize complex patterns and to be highly accurate in the predictions.

4.6. Clustering Analysis

The Elbow Method to validate the number of the optimal clusters in the K-Means clustering analysis(Schubert, 2023). The y-axis indicates the within-cluster sum of squares that is an indicator of the compactness of the clusters and the x-axis indicates the number of clusters. Figure 18, The decrease in WCSS as k = 1 to k = 2 means that the variance can be explained by two clusters only to a reasonable extent. WCSS begins to flatten out after k=2, the returns to adding more clusters are getting small. K mean clustering k=2, frequencies of prevailing settlement patterns in the data set. X: The following scatter plot presents the outcome of K-Means clustering of the settlement data. The data are shown as color-

DOI: 10.3217/978-3-99161-057-1-044

coded clusters: red = Cluster 1, green = Cluster 2, blue = Cluster 3. In these two axes, the x-axis (Feature 1) would show something like Time Interval Days and the y-axis (Feature 2) could show Distance from Start. The clusters indicate spatial settlement patterns: Cluster 1 (red): The regions of maximum movement, most probably in the area of structural features or ground disturbance. 4465 points, Cluster 2 (green): The regions of transition with average settlement rates and medium stability. 1341 points, Cluster 3 (blue): Stable areas, where there are few changes in the settlement, corresponding to less vulnerable areas. 4057 points. The Euclidean distances between clusters are plotted on the y-axis and the data points (e.g., monitoring locations) on the x-axis Figure 20. The vertical lines are also branched vertically on the height of the dendrogram where the significant differences are found and this is another evidence that there are three main groups that are identified in the hierarchy. Table 4. Feature Importance of Each Cluster Model.

5. Results

A detailed analysis of the Changshui Airport Tunnel based on various state-of-the-art machine learning algorithms- Random Forest (RF), Gradient Boosting (GB), and Artificial Neural Networks (ANN) was conducted to analyze and predict the settlement behaviors. we are presenting the results with performance metrics and feature importance of all the models. Model Performance Metrics Table 2. Model Performance Metrics for Each Model summarizes quantitative performance metrics of each model, presented in Figure 7. Performance Metrics – RF.. These results indicate that ANN model is more accurate in predicting the settlement behavior, which proves its strength and reliability in predicting the settlement behavior more accurately than the other models. Table 3. Feature Importance of Each Model below shows the importance of each feature to the respective models. The importance scores enable us to know the values of features that are most closely associated with model predictions. In addition, the significance scores are also graphically illustrated in Figure 3, Figure 4, Figure 5, which indicates that Feature 3 is dominant in all models, which is a significant environmental or geotechnical factor. The total residuals are presented in Figure 14, which indicates that ANN has predicted very little wrong. The heatmap plots of the actual and predicted settlement were very similar, and this fact proved that the ANN predictions are the closest to the actual ones, and the predictions of GB and RF were the second. This spatial disaggregation plays an important role in the proper depiction of risk distribution. The error distribution in each model is described and shown in Figure 16, which shows the frequency of the prediction errors that have been observed in some ranges. Elbow Method determines the number of clusters as illustrated in Figure 18. In order to get a meaningful segmentation of settlement patterns the elbow plot indicates that the optimal number of clusters is three. The outcome of the clustering can still be visualized, which is K-means clustering of data points in Figure 19 and the dendrogram of hierarchical clustering in Figure 20, which confirm the segmentation results based on hierarchical relationships between individual points of settlement.

6. Discussion

The study demonstrates that geotechnical monitoring can be enhanced considerably with the assistance of deep analysis based on ML to process complicated data sets(Ritter & Frauenfelder, 2021). The concentration of monitor location to some clusters that depict characteristics assists in the development of particular place of resident behaviors. In general, Cluster 1 had 4,465 points (marked as RESILIENT / STABLE on the map) and we can use these areas as reference areas, or even a baseline of comparison of how we can improve and others areas came out as Cluster 2, which had 1,341 points and marked as THREATENED on the maps, which are transitional or Undeveloped areas, which need more care and attention with interventions more proactive. These clusters are consistent with the predictive data of machine learning models (ANN, GB, etc.) that showed Monitoring Point Elevation as the most correlated parameter, which confirms that the higher settlement was observed in the sites of well locations that were in Cluster 2. The clustering results give the necessary information and confirmations to help in maintenance planning and resources allocation besides confirming the model predictions.

6.1. Analysis of Improved Model Effectiveness, Interpretation and Consequences

The ANN is the solution to risk prediction and mitigation in tunnel scenarios because of its excellent capacity to comprehend complex, non-linear relationships in geotechnical data(Ramezanshirazi et al., 2019). On the other hand, GB and RF were good alternatives, and GB reduced prediction errors by iterative refinement and RF was easy to interpret. The elevation of the monitoring point was also identified as a critical factor in all the models since the feature significance analysis revealed the critical importance of the elevation changes in influencing the tunnel settlements(Apoji et al., 2022). According to this conclusion, the influence of the elevation changes on the stress distribution and the following settlement patterns is considerable, which is consistent with the geotechnical principles (Figure 3, Figure 4, and Figure 5).

Both models possessed some advantages within the context of geotechnical data, Random Forest: It was very robust and provided interpretable results and therefore it can be applied to problems where the impact of a specific feature is of interest. Gradient boosting was very good in minimizing the prediction error and managing the interaction between features. This meant that it was suitable in the modeling of the non-linear dynamics of the settlement data. Artificial Neural Networks: Have proven to be very accurate in handling complex data structure, but they are very demanding in terms of processing power and they need to be fine-tuned to avoid overfitting.

6.2. Prospective Research Directions

The proactive management of infrastructure is being embraced by integrating machine learning (ML) in tunnel monitoring(Plevris & Papazafeiropoulos, 2024). Machine learning enhances tunnel safety and life by identifying potential danger zones early enough due to its ability to process and analyze large volumes of data. Nevertheless, the

DOI: 10.3217/978-3-99161-057-1-044

indictment also points to areas that should be further investigated, to enhance the effectiveness of machine learning in the area of tunnel monitoring: These are the incorporation of real-time data inputs to enable adaptive responses to environmental deviations, and a data-driven dynamic update of predicted models as part of a real-time processing solution(Zhang et al., 2020). Cross-Project Validation: experiments are conducted on a number of tunnel projects to determine the generalizability of the models to standardize ML applications in civil engineering.

7. Conclusion

Further use of machine learning methods in this field may assist us in getting nearer to the realization of the role of different parameters in the monitoring of tunnel (as a part of Geotechnical Engineering) and the qualitative side of Geotechnical engineering as perceived through the eyes of a geotechnical engineer. This paper was restricted to the monitoring of tunnel in Changshui and the monitoring system is comprehensive to monitor the deformation patterns of Changshui. Security aspects of Geotechnical Engineering Which are likely to be messages of improved predicting performances and functional profitability of the Geologist economical relevant hazard mapping linked with geological changes(Wu et al., 2021). Therefore, Random Forests, Gradient Boosting Machines, and Artificial Neural Networks have been applied effectively in enhancing the accuracy and reliability of settlement predictions, which is essential to the stability, as well as sustainability, of tunnel infrastructures(Yan et al., 2019). This was the main area/work in the investigation of the possibilities of these models, where, as they have a good understanding of complex and nonlinear data interactions, in the sense of the higher R^2 and lower RMSE and MAE on the Artificial Neural Network, the best results were obtained. These excellent results demonstrate the potential of sophisticated ML models to become a disruptive technology to traditional geotechnical monitoring practices that allow the risk assessment and mitigation plans to go much deeper than ever before. The effective use of these technologies can also be extended to other geotechnical events that are high risk(Wang et al., 2021). Moreover, the use of machine learning methods must be applied to various project scenarios in other studies, which will allow a better idea of whether/what algorithms can be used in different construction conditions in real time.

8. Mind Maps, Figures and Tables

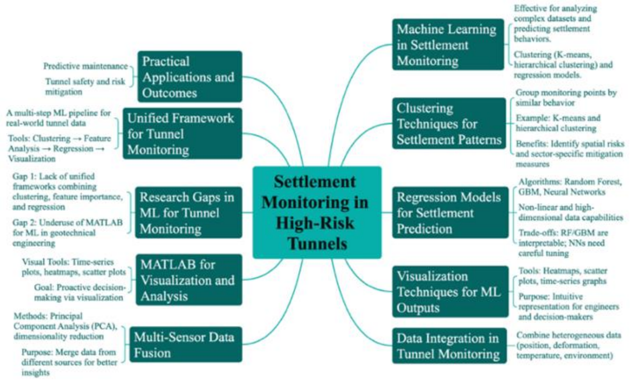

Figure 1. Mind map of settlement monitoring in high-risk tunnels.

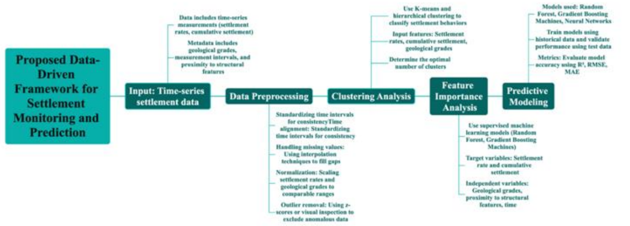

Figure 2. Mind Map of Proposed Data-Driven Framework for Settlement Monitoring and Prediction

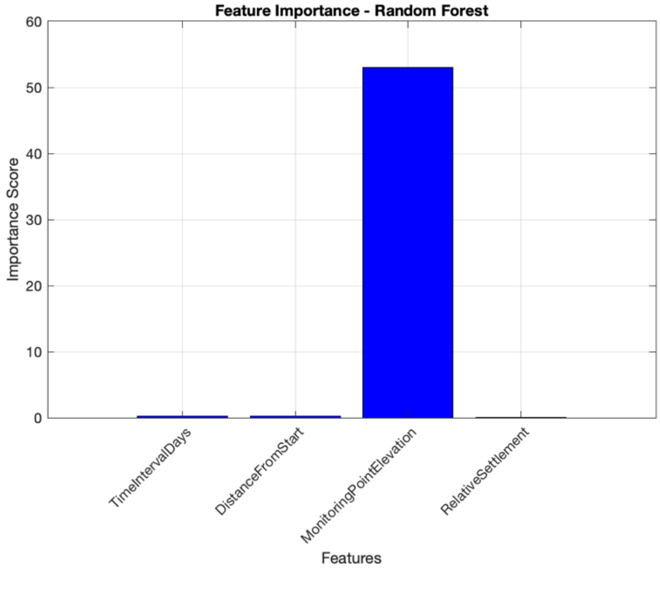

Figure 3. Feature Importance – RF.

DOI: 10.3217/978-3-99161-057-1-044

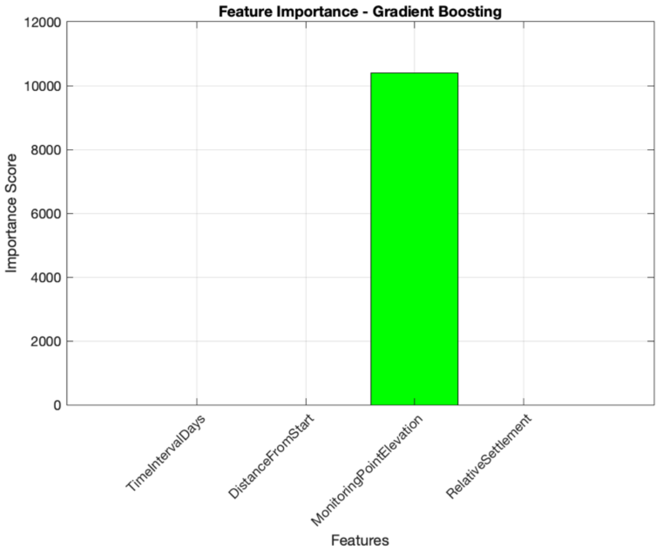

Figure 4. Feature Importance– GB.

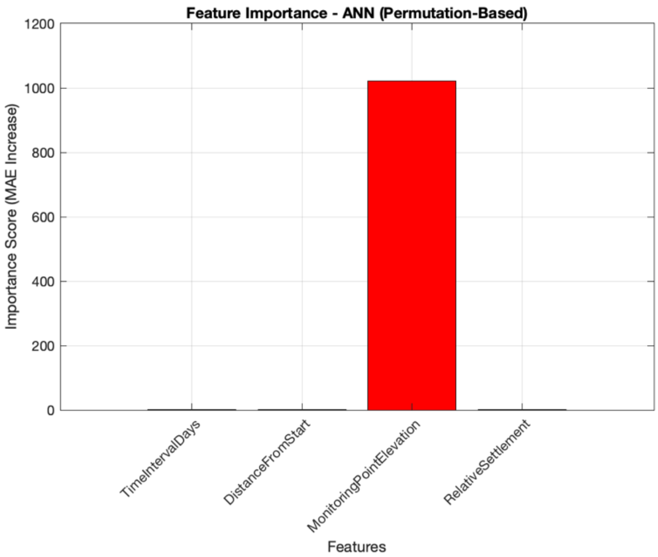

Figure 5. Feature Importance – ANN.

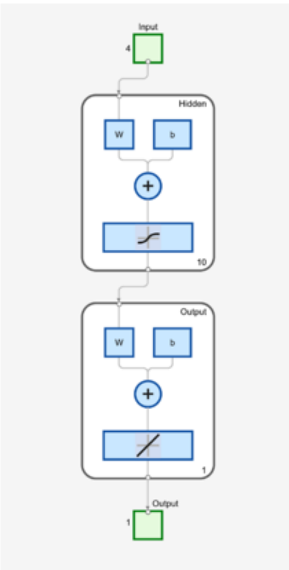

Figure 6. Architecture of ANN.

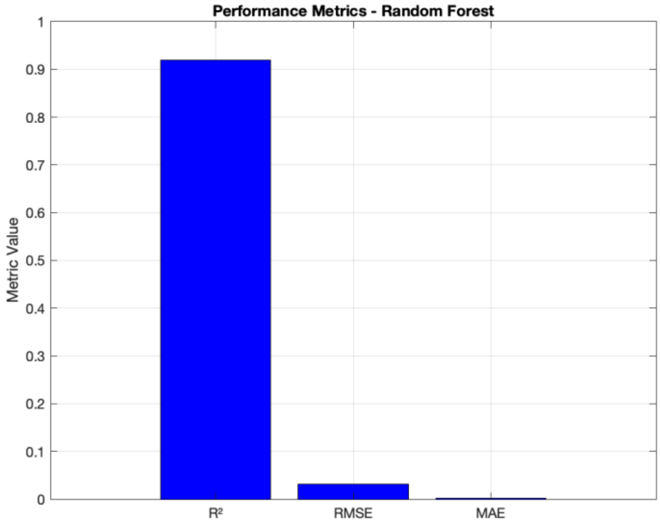

Figure 7. Performance Metrics – RF.

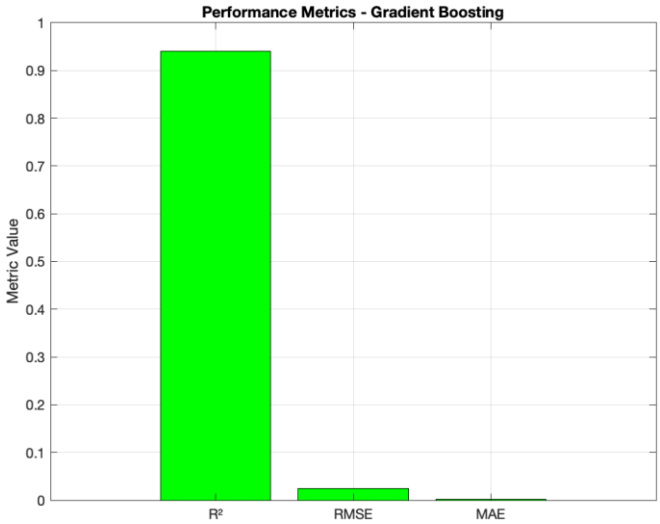

Figure 8. Performance Metrics – GB.

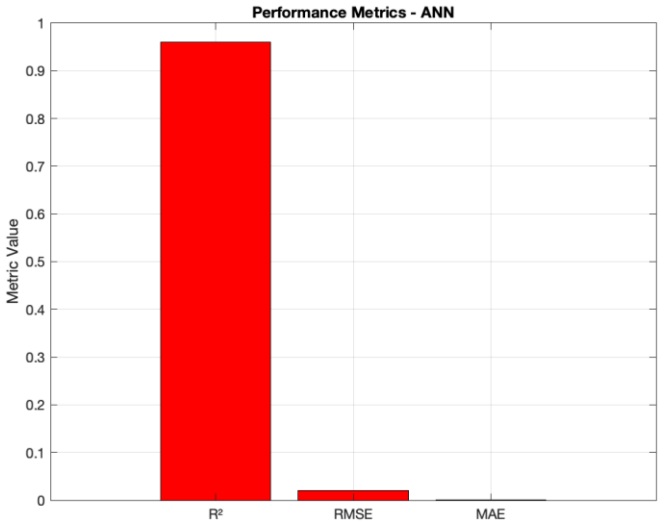

Figure 9. Performance Metrics – ANN.

DOI: 10.3217/978-3-99161-057-1-044

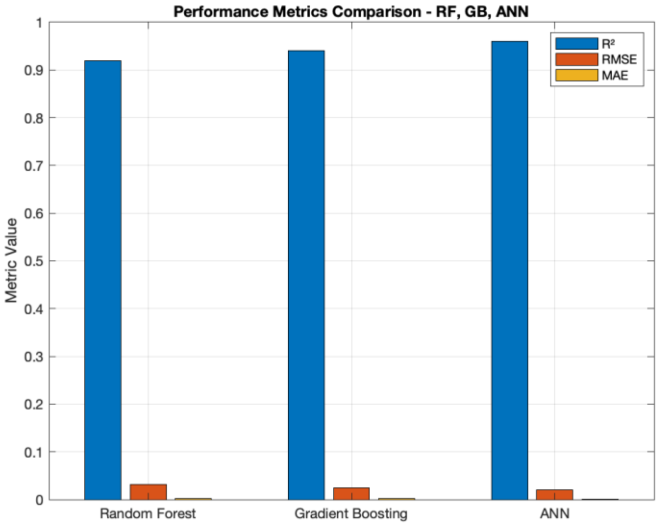

Figure 10. Performance Metrics – RF, GB, ANN.

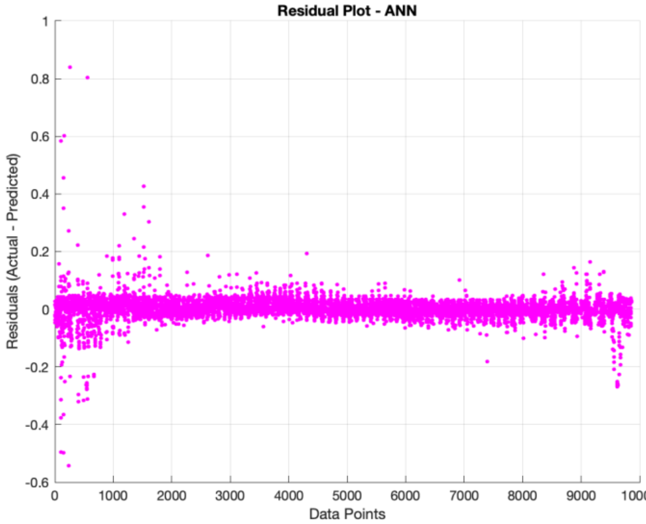

Figure 13. Residual Plot – ANN.

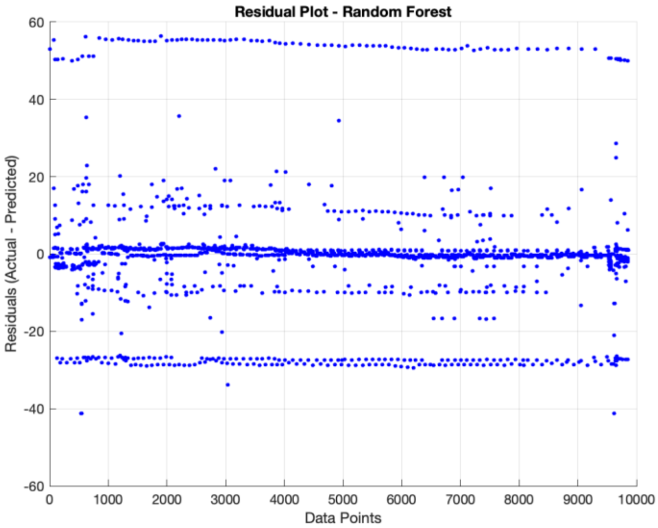

Figure 11. Residual Plot – RF.

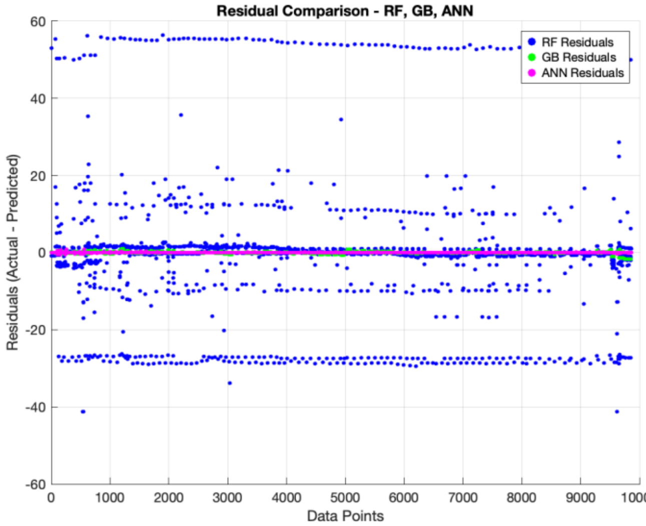

Figure 14. Residual Plot – Rf, GB, ANN.

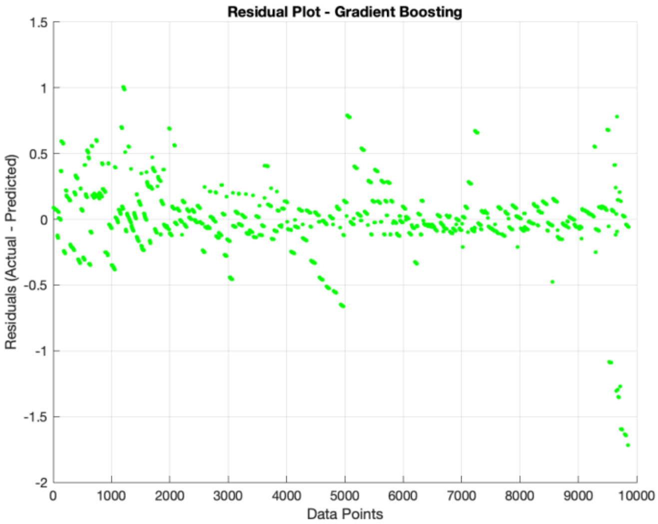

Figure 12. Residual Plot – GB.

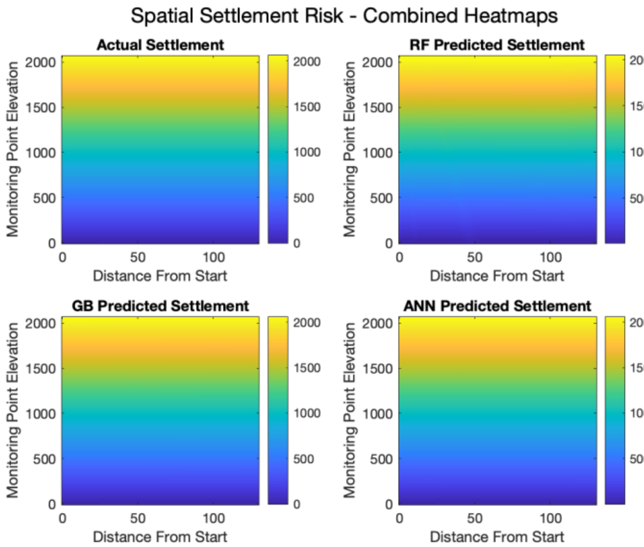

Figure 15. Spatial Settlement Risk – Combined Heatmaps of RF, GB, ANN.

DOI: 10.3217/978-3-99161-057-1-044

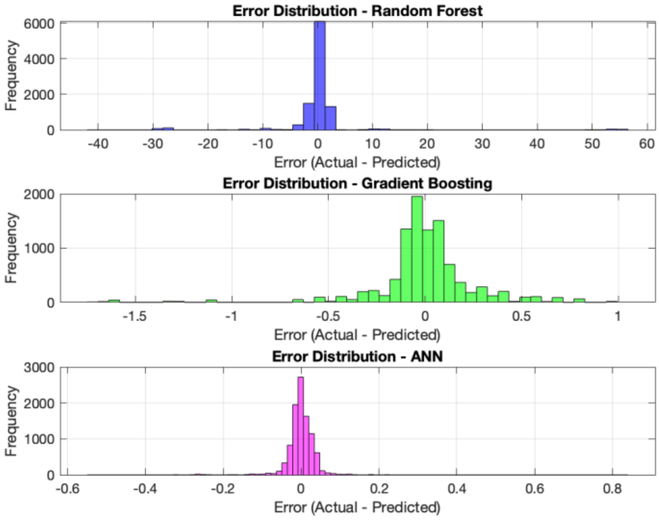

Figure 16. Error (Actual – Predicted) RF, GB, ANN.

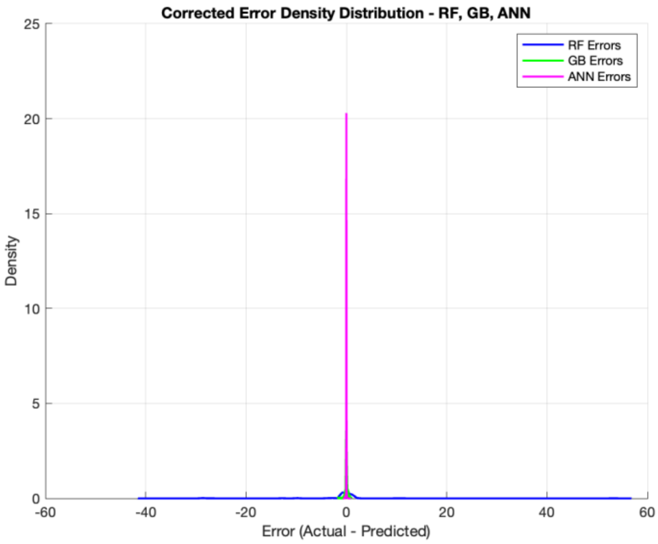

Figure 17. Density Combined Error (Actual – Predicted) RF, GB, ANN.

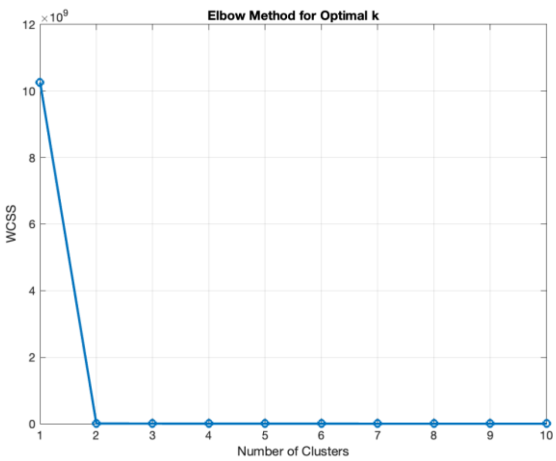

Figure 18. Elbow Method for Optimal k.

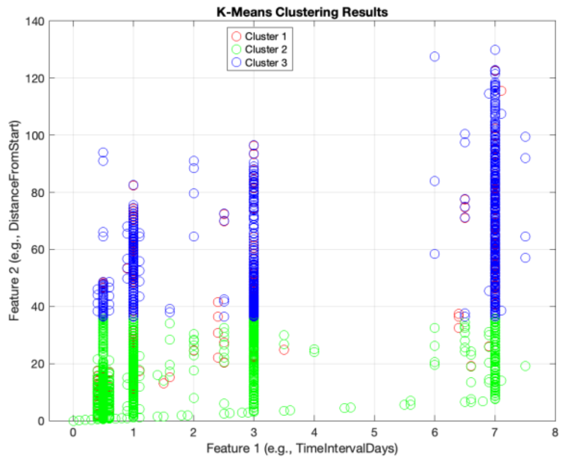

Figure 19. K-Means Clustering Results.

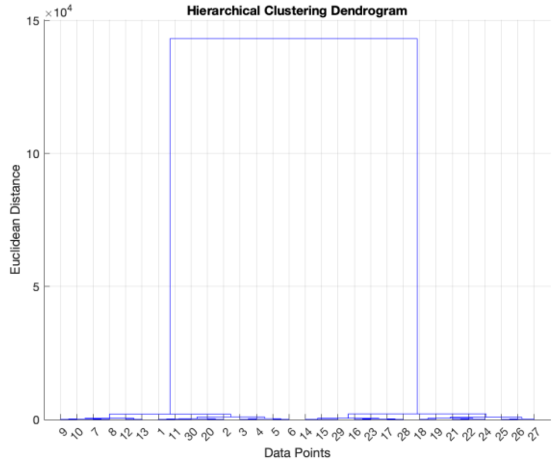

Figure 20. Hierarchical Clustering Dendrogram.

Table 1. Numerical Insights from ANN Feature Importance.

Feature	Permuted MAE	Importance Score	Interpretation
Time Interval Days	0.0038008	0.0016423	Temporal variations moderately affect settlement.
Distance From Start	0.003671	0.0015125	Distance influences settlement slightly
Monitoring Point Elevation	1028.8267	1028.8246	Elevation dominates as the most critical factor
Relative Settlement	0.0021976	0.00003908	Minor impact on model predictions

Table 2. Model Performance Metrics for Each Model

Model	R^2	MAE	RMSE
RF	0.92	0.002	0.031
GB	0.94	0.0015	0.025

DOI: 10.3217/978-3-99161-057-1-044

ANN	0.96	0.001	0.02

Table 3. Feature Importance of Each Model

Feature	RF Importance	GB Importance	ANN Importance
1	0.23432	0	0.0078671
2	0.24303	0	0.004688
3	50.255	10396	987.70
4	0.085561	0	0.00029713

Table 4. Feature Importance of Each Cluster Model

Cluster	Points	Feature 1	Feature 2	Feature 3	Feature 4
Cluster 1	4465	0.0018	0.0249	2.0613	0.0004
Cluster 2	1341	4.6088	60.3270	12.8568	0.3447
Cluster 3	4057	1.4209	11.9267	12.8346	0.4621

Acknowledgement

The financial support of the Key Technology R&D Plan is appreciated (Grant No. 202303AA080003) by the Yunnan Provincial Department of Science and Technology. This grant enabled the research that is the subject of this paper. We would also like to express our gratitude to all the colleagues and technical personnel who participated in the model and implementation of this project in particular to their contribution and efforts.

References

[1] Apoji, D., Fujita, Y., & Soga, K. (2022). Soil Classification and Feature Importance of EPBM Data Using Random Forests. Geo-Congress 2019, 520. https://doi.org/10.1061/9780784484029.052

[2] Ayasrah, M., Qiu, H., Zhang, X., & Daddow, M. (2020). Prediction of Ground Settlement Induced by Slurry Shield Tunnelling in Granular Soils. Civil Engineering Journal, 6(12), 2273. https://doi.org/10.28991/cej-2020-03091617

[3] Bao, Y., Tang, Z., Li, H., & Zhang, Y. (2018). Computer vision and deep learning–based data anomaly detection method for structural health monitoring. Structural Health Monitoring, 18(2), 401. https://doi.org/10.1177/1475921718757405

[4] Chen, Q., Yang, Z., Xue, B., Wang, Z., & Xie, D. (2022). Monitoring and Prediction Analysis of Settlement for the Substation on Soft Clay Foundation. Advances in Civil Engineering, 2022(1). https://doi.org/10.1155/2022/1350443

[5] Fan, C., Sun, Y., Zhao, Y., Song, M., & Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. Applied Energy, 240, 35. https://doi.org/10.1016/j.apenergy.2019.02.052

[6] Hua, J., Cheng, J., Zhang, J., Zhang, J., Su, Y., & Zheng, Y. (2021). Principle and application of in-situ monitoring system for ground displacement induced by shield tunnelling. Tunnelling and Underground Space Technology, 112, 103905. https://doi.org/10.1016/j.tust.2021.103905

[7] Jasmine, P. H., & Arun, S. P. (2021). Machine learning applications in structural engineering - a review [Review of Machine learning applications in structural engineering - a review]. IOP Conference Series Materials Science and Engineering, 1114(1), 12012. IOP Publishing. https://doi.org/10.1088/1757-899x/1114/1/012012

[8] Jin-miao, C., Shen, X., & Chen, Q. (2022). Prediction of Maximum Surface Settlements of Bai~Hua Tunnel Section based on Machine Learning. Journal of Physics Conference Series, 2185(1), 12042. https://doi.org/10.1088/1742-6596/2185/1/012042

[9] Karamvasis, K., & Karathanassi, V. (2020). Performance Analysis of Open Source Time Series InSAR Methods for Deformation Monitoring over a Broader Mining Region. Remote Sensing, 12(9), 1380. https://doi.org/10.3390/rs12091380

[10] Ma, G., Zhang, S., Li, H., Li, P., & Liu, Y. (2021). A Method of Monitoring the Cross-Section Deformation of Tunnels Using the Strain Data from the Fully Distributed Optical Fibre Sensors. IOP Conference Series Materials Science and Engineering, 1203(2), 22125. https://doi.org/10.1088/1757-899x/1203/2/022125

[11] Marcher, T., Erharter, G. H., & Winkler, M. (2020). Machine Learning in tunnelling – Capabilities and challenges. Geomechanics and Tunnelling, 13(2), 191. https://doi.org/10.1002/geot.202000001

[12] Marsella, M., & Scaioni, M. (2018). Sensors for Deformation Monitoring of Large Civil Infrastructures. Sensors, 18(11), 3941. https://doi.org/10.3390/s18113941

[13] Martins, B. H., Suzuki, M., Yastika, P. E., & Shimizu, N. (2020). Ground Surface Deformation Detection in Complex Landslide Area—Bobonaro, Timor-Leste—Using SBAS DInSAR, UAV Photogrammetry, and Field Observations. Geosciences, 10(6), 245. https://doi.org/10.3390/geosciences10060245

[14] Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., Pham, B. T., Bui, D. T., Avtar, R., & Boumezbeur, A. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225. https://doi.org/10.1016/j.earscirev.2020.103225

[15] Mousavi, S. M., & Beroza, G. C. (2022). Machine Learning in Earthquake Seismology. Annual Review of Earth and Planetary Sciences, 51(1), 105. https://doi.org/10.1146/annurev-earth-071822-100323

[16] Muhammed, H. H., Zakaria, W. A., & Mahdi, Q. A. (2019). Prediction of Stresses and Settlement for TBM Tunnel Surrounding Soil. Diyala Journal of Engineering Sciences, 12(4), 50. https://doi.org/10.24237/djes.2019.12405

[17] Oh, D.-W., Kong, S.-M., Lee, Y.-J., & Park, H. (2021). Prediction of Change Rate of Settlement for Piled Raft Due to Adjacent Tunneling Using Machine Learning.

DOI: 10.3217/978-3-99161-057-1-044

Applied Sciences, 11(13), 6009. https://doi.org/10.3390/app11136009

[18] Plevris, V., & Papazafeiropoulos, G. (2024). AI in Structural Health Monitoring for Infrastructure Maintenance and Safety. Infrastructures, 9(12), 225. https://doi.org/10.3390/infrastructures9120225

[19] Ramezanshirazi, M., Sebastiani, D., & Miliziano, S. (2019). Artificial Intelligence to Predict Maximum Surface Settlements Induced by Mechanized Tunnelling. In Lecture notes in civil engineering (p. 490). Springer Nature. https://doi.org/10.1007/978-3-030-21359-6_52

[20] Ritter, S., & Frauenfelder, R. (2021). InSAR monitoring data to assess building response to deep excavations. IOP Conference Series Earth and Environmental Science, 710(1), 12036. https://doi.org/10.1088/1755-1315/710/1/012036

[21] Schubert, E. (2023). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. ACM SIGKDD Explorations Newsletter, 25(1), 36. https://doi.org/10.1145/3606274.3606278

[22] Tan, X., Chen, W., Wang, L., Tan, X., & Yang, J. (2019). Integrated Approach for Structural Stability Evaluation Using Real-Time Monitoring and Statistical Analysis: Underwater Shield Tunnel Case Study. Journal of Performance of Constructed Facilities, 34(2). https://doi.org/10.1061/(asce)cf.1943-5509.0001391

[23] Vadyala, S. R., Betgeri, S. N., Matthews, J. C., & Matthews, E. (2021). A review of physics-based machine learning in civil engineering [Review of A review of physics-based machine learning in civil engineering]. Results in Engineering, 13, 100316. Elsevier BV. https://doi.org/10.1016/j.rineng.2021.100316

[24] Wang, J., Tang, S., Heap, M. J., Tang, C., & Tang, L. (2021). An auto-detection network to provide an automated real-time early warning of rock engineering hazards using microseismic monitoring. International Journal of Rock Mechanics and Mining Sciences, 140, 104685. https://doi.org/10.1016/j.ijrmms.2021.104685

[25] Wang, L., Xu, S., Qiu, J., Wang, K., Ma, E., Li, C., & Guo, C. (2020). Automatic Monitoring System in Underground Engineering Construction: Review and Prospect. Advances in Civil Engineering, 2020(1). https://doi.org/10.1155/2020/3697253

[26] Wu, J., Chen, J., Chen, G., Wu, Z., Yu, Z., Chen, B., Ke, W., & Huang, J. (2021). Development of Data Integration and Sharing for Geotechnical Engineering Information Modeling Based on IFC. Advances in Civil Engineering, 2021(1). https://doi.org/10.1155/2021/8884864

[27] Yan, K., Dai, Y., Xu, M., & Mo, Y. (2019). Tunnel Surface Settlement Forecasting with Ensemble Learning. Sustainability, 12(1), 232. https://doi.org/10.3390/su12010232

[28] Zhang, P., Chen, R., Tian, D., Wang, Z.-T., & Wu, K. (2020). An AIoT-based system for real-time monitoring of tunnel construction. Tunnelling and Underground Space Technology, 109, 103766. https://doi.org/10.1016/j.tust.2020.103766

[29] Zhao, Y., Han, Y., Kou, Y., Li, L., & Du, J.-H. (2021). Three-Dimensional, Real-Time, and Intelligent Data Acquisition of Large Deformation in Deep Tunnels. Advances in Civil Engineering, 2021(1). https://doi.org/10.1155/2021/6671118

[30] Zhou, C., Kong, T., Jiang, S., Chen, S., Zhou, Y., & Ding, L. (2020). Quantifying the evolution of settlement risk for surrounding environments in underground construction via complex network analysis. Tunnelling and Underground Space Technology, 103, 103490. https://doi.org/10.1016/j.tust.2020.103490