

## **THE APPLICATIONS OF DATA SCIENCE AND BIG DATA ANALYTICS IN UNDERGROUND TRANSPORTATION INFRASTRUCTURE**

## **FINAL PROJECT REPORT**

by

Mohammad Pourhomayoun, Kabir Nagrecha, Mehran Mazari, Tonatiuh Rodriguez-Nikl

California State University Los Angeles

Sponsorship

UTC-UTI

For University Transportation Center for Underground Transportation Infrastructure (UTC-UTI)

March, 2020



#### **Disclaimer**

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.



objective is to develop advanced data mining and novel machine learning based methods for predicting or detecting ground conditions using the data collected before and during the TBM operations. The second objective is to design and develop data-driven predictive models that can predict the TBM state and status in real-time as well as adverse events and anomalies. The project includes 3 main phases: (I) large-scale UTI data collection, exploration, and pre-processing; (II) feature and knowledge extraction, and dimensionality reduction, (III) data analytics, and predictive analytics model using machine learning/deep learning methods and data visualizations.



# **Table of Contents**

<span id="page-3-0"></span>

## **3. List of Figures**

<span id="page-4-0"></span>Figure 1. (a) Structure and Components of a Tunnel Boring Machine, (b) Hierarchical structure of the tunnel rings composed of precast segments (Yi et al. 2019)

Figure 2. (a) Interpolation of geological data to generate tunneling operation data (after Sun et al. 2018), (b) 3D visualization of tunnel alignment with borehole geotechnical data (after Ozmutlu and Hack, 2003), (c) Integration of data-driven model for prediction of geological information (after Zhao et al. 2019)

Figure 3. Geological profile of the tunneling project

Figure 4. Input Data format for RNN model

Figure 5. Structure of a recurrent neural network (RNN)

Figure 6. A comparison of GRU and LSTM recurrent neurons.

Figure 7. General topographical difference between a) ANN, b) RNN

Figure 8. Comparisons of predictions against actual sensor outputs at future rings. The predictive model is trained and tested on separate datasets from the first tunnel.

Figure 9. Predictions versus actual values. The predictive model is trained and tested on separate datasets from the second tunnel.

Figure 10 Predictions of the model having been trained on the first tunnel plus less than 15% of the second tunnel.

Figure 12. Stack plot of component predictions from a) the proposed RNN-LSTM, and b) Kriging interpolation from borehole data.

Figure 13. Individual earth layer estimations from RNN-LSTM model compared to labeled data from Kriging interpolation of borehole data.

## **4. List of Tables**

<span id="page-5-0"></span>Table 1 – Sequence of time steps and labels

Table 2 – Results of the model trained and tested on the first tunnel alone

Table 3 – Results of training and testing only on the second tunnel

Table 4 – Results of training on the whole of the first tunnel and less than 15% of the second tunnel, predictions on the portion of the second tunnel used in past tests.

Table 5 – Results of training on the first tunnel and testing on the second.

# **5. List of Abbreviations**

<span id="page-6-0"></span>TBM: Tunnel Boring Machine ML: Machine Learning DL: Deep Learning AI: Artificial Intelligence ANN: Artificial Neural Networks RNN: Recurrent Neural Network LSTM: Long Short Term Memory

## **6. CHAPTER 1 – INTRODUCTION**

#### <span id="page-7-1"></span><span id="page-7-0"></span>**1.1. Overview**

Shield-driven tunneling using Tunnel Boring Machines (TBM) has become a popular underground construction technique in various geological conditions with minimal surface disruptions. The advantage of using such technology is that while excavating, the TBM is placing a tunnel liner of sequential rings composed of precast segments (Figure 1). The advancement of machine is powered by thrust cylinders while an erector places the interlocking segments. Thrust load depends on type of soil and other factors such as groundwater pressure. However, the typical thrust load for excavating in soft soil conditions is 10 to 20 percent of the total TBM thrust (Galvan et al. 2017). Another important component of the TBM is the cutterhead which is located at the frontal contact area of the machine during excavation process. Configuration of the cutterhead and its performance depends on many factors including cutter type, spacing of the cutters, cutterhead shape and balance of the head (Rostami and Chang, 2017). Cutter rotation speed and cutterhead speed are two of the notable TBM performance parameters that are collected continuously during the excavation. The later parameters as well as the other machine operational information such as advancement speed, thrust force and articulation force are recorded and stored for each tunneling project. However, only a portion of this data are utilized by the TBM operators mostly during the excavation to advance the machine along the designated alignment. Real-time sequential estimation of such information can significantly improve the operational performance and avoid any unpredicted encounters.



Figure  $1 - a$ ) Structure and Components of a Tunnel Boring Machine (www.railsystem.net), b) Hierarchical structure of the tunnel rings composed of precast segments (Yi et al. 2019)

Although a TBM is capable of excavating under different subsurface conditions, the complexity and uncertainty of geological conditions ahead of the machine can infer substantial construction delays as well as unforeseen damages to the cutter head and other TBM components. The geological profile and geotechnical conditions along the tunneling alignment are highly uncertain due to the limited sampling rates during site investigation, e.g., borehole spacings of 50-200 m. With pressure balance shield TBMs, project personnel are unable to see the encountered ground conditions while tunneling; there is no access to the face. The inability to identify and characterize the as-encountered ground makes it difficult to optimize the tunneling process. Further, differing site condition claims and disputes are common on tunnel projects; however, with no clear understanding of the actual ground conditions encountered, such claims and disputes become problematic to resolve. To this end, there is significant incentive to develop methodologies that can characterize the ground using the TBM data.

The first objective is to develop advanced data mining and novel machine learning based methods for predicting or detecting ground conditions using the data collected before and during the TBM operations. The second objective is to design and develop data-driven predictive models that can predict the TBM state and status in real-time as well as adverse events and anomalies. The system keeps updating the predictive model as new information are introduced to the model. A recurrent neural networks (RNN), which is a modified version of artificial neural network (ANN) was developed.

### <span id="page-8-0"></span>**1.2. Prediction of TBM Performance and State**

A number of studies in the literature were focused on predicting the operational parameters of TBM as well as information relevant to geological environment along the tunneling alignment (Mooney et al. 2012). Some of these efforts were limited to a specific ground conditions and environmental factors while some others showed limited ability to estimate the as-encountered TBM operational parameters. TBM performance prediction models are generally classified in three groups: theoretical models based on laboratory tests, empirical models based on field performance of TBM (Hasanpour et al. 2016) and data-driven models based on as-encountered operation parameters during excavation. Although there are several studies in the literature focused on developing and updating the first two types of models, only few studies were dedicated to develop data-driven algorithms with minimal dependency on empirical parameters.

Toth et al. (2013) analyzed TBM performance in various conditions to understand environmental impact on TBM performance. They focused on addressing the inability of models to predict penetration rate in mixed ground conditions by analyzing performance in homogeneous conditions. Prediction of penetration rate have been popular in recent TBM predictive analysis techniques, as the feature is an important part of understanding excavation performance. However, due to the limitation of prediction model to a specific geological composition, a lot more data is needed to generalize the model to be applicable to a variety of underground conditions. They found direct correlations between penetration rate and geological parameters but claim that additional research will be required to find out if the penetration rate was solely affected by the geological features or it relies on the experience of machine operators.

Avunduk et al. (2018) developed a process to predict the excavation performance based only on soil properties. They showed accuracy in predicting cutterhead performance and thrust force based on single-variate and multi-variate analysis of soil and clay composition with a simple regression model. Bilgin et al. (2012) analyzed TBM preformance based on rock and soil composition in fractured rock formations. They developed a model using a stochastic estimator and a Monte Carlo simulation for predicting performance in clay-heavy ground. However, they found that the applicability of the model in other geological conditions is limited. The developed model, although accurate in predicting the penetration rate, fails to estimate cutterhead torque and thrust with similar accuracy. These features are important to understand TBM performance as they reflect the machine ability to excavate.

Farrokh et al. (2012) reviewed the accuracy of models in predicting TBM performance with neural networks using basic mechanical data and found that the model does not propose reasonable predictions compared to traditional methodologies. They also concluded that most existing predictive models, both traditional and computer-aided, cannot offer accurate estimates of TBM performance on new excavations without significant re-training. They suggest that this is due to a lack of inclusion of important parameters, and that an accurate record of operational parameters from a variety of test sites could help in improving the reliability of prediction models.

Salimi et al. (2015) employed a number of artificial intelligence techniques to predict the advancement rate and other performance parameters of the TBM using the data extracted from two hard rock tunneling sites. Those techniques include principle component analysis (PCA) as a preprocessing approach and artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS) and support vector regression (SVR) to develop the prediction models. They evaluated the performance of prediction models using root mean square error (RMSE), variance account for (VAF), and mean absolute percentage error (MAPE). Their study found that although all prediction models showed acceptable performance, the SVM method outperformed the other two models. However, only one parameter known as field penetration index (FPI) was predicted under limited ground conditions.

One of the main objectives of this project was to evaluate the possibility of implementing a recurrent neural network (RNN), a machine learning technique, to predict some of the operational parameters of TBM, using earlier operating data during the excavation process. We extracted TBM data from a tunneling project in North America that involved construction of a double parallel set of tunnels. Our study also evaluates the possibility of applying training data from one tunnel to another with both minimal and nonexistent re-training.

### <span id="page-9-0"></span>**1.3. Prediction of Geological Composition**

UTC-UTI 10 Due to the uncertainties involved with tunneling and unknown properties of earth layers that could delay the tunnel construction process and impose extra cost in terms of cutterhead replacement or repairs, TBM monitoring has been the focus of several studies over the past few decades. Several studies have focused on developing in-situ geophysical and imaging techniques to estimate the unanticipated geological conditions along the tunneling path. As one of the early stages of such attempts, Kneib et al. (2000) developed a methodology for automatic seismic prediction ahead of the tunnel boring machine. Both sources and receivers were mounted on the cutterhead for optimal spatial coverage. The sonic soft-ground probing (SSP) system excites a high frequency P-wave that is recorded by mounted accelerometers. The setup yields a three-dimensional reflection image of the ground condition ahead of cutting wheel. However, at the time of that study, several challenges such as the need for real time signal processing, limited computational power and relatively high noise levels due to construction process hindered the implementation of such system. Kaus and Boening (2008) introduced a non-intrusive electrical induced polarization technique that predicts the ground conditions while TBM is operating. The Bore-Tunneling Electrical Ahead Monitoring (BEAM) allows for prediction of earth layers about three times the diameter ahead of TBM cutter head. It is capable of early detection and warning of geological and geotechnical ground conditions as well as real-time visualization of earth layer classifications. The system minimizes the need for geotechnical baseline report borehole data to estimate and reconstruct the geological layer combinations. However, the additional components of the data acquisition system should be mounted on the cutter head as well as the TBM operating center to visualize the process.

Mooney et al. (2012) reviewed the state of the art in real-time TBM monitoring. They include a comprehensive list of different methods and approaches that are mostly focused on in-situ and geophysical techniques implemented at the tunnel face. Those methods include passive monitoring of TBM interaction with tunnel face, acoustic reflection, electrical resistivity, cutterhead monitoring, backfill grout monitoring, and muck monitoring. The emphasis on look-ahead techniques with seismic, acoustic and electrical methods is highlighted in that review. Schaeffer and Mooney (2016) performed an experimental and computational investigation of electrical resistivity imaging for prediction ahead of TBM. That study presented real-time and continuous imaging solution to extract more information along the tunneling direction. Such tools can help detecting the unpredicted changes in earth layer properties ahead of cutterhead that could impose construction delays and excessive costs of maintenance and repairs. The study was focused on laboratory scale experiments and showed the potential of detecting most changes ahead of TBM even in high electrical noise.

Although experimental and geophysical prediction of earth conditions ahead of TBM cutterhead have been investigated in several studies in the literature, the use of data-driven approaches is relatively more recent. A few of those studies are employing traditional statistical methods while the most recent ones are employing advanced prediction models such as machine learning and deep neural networks. Zhao et al. (2019) proposed a data-driven framework for tunnel geologicaltype prediction based on TBM operating data. They proposed a real-time process that first converts the discontinuous operating information to continuous displacement data and then augments TBM features using first and second order difference method. The developed artificial neural network predicts multiple geological layer properties using physical and mechanical indices. Those indices include natural severity, internal friction angle, deformation modulus, Poisson's ratio, coefficient of lateral pressure, permeability coefficient, and cohesive strength between rock mass and anchors. The authors show that the developed algorithm outperforms conventional statistical prediction models such as random forest, support vector regression and K-nearest neighbors. The main source of geological data exploration in this study was from boreholes along the tunnel length and use the ring sections corresponding to the location of drilling for training the algorithm. The authors later conclude that use of more advanced outlier detection methods as well as using other training algorithms could help improve the accuracy of the prediction model.

Maher (2013) presented a machine learning approach to predict penetration rate in earth pressure balance (EPB) tunnel boring machines. Two methods used for feature selection were Guided Regularized Random Forests (GRRFs) and Stepwise Forward Feature Selection (SFFS). However, a combination of multiple linear regression and support vector regression was employed to perform the predictions using the features selected with SFFS and GRRF. The application of these methods showed that some of the selected features based on data-driven approach were not previously considered as identified by lab experiments in the literature.

Shi et al. (2018) compared the statistical learning methods with deep neural networks (DNN) to predict the geological conditions based on TBM operating data. They applied a DNN model to a set of TBM data with 53 attributes that were measured continuously at a 1 Hz frequency. The developed model was then employed to predict 7 geological layers along the tunneling direction. The reported accuracy of the developed technique was optimal for some layers and relatively poor for the other geological layers. However, the performance of the model was superior compared to conventional statistical methods such as random forest, k-nearest neighbor and linear regression. The authors employed a combination of algorithms to avoid over-fitting, minimizing the loss function and fixing the unbalanced data layers.

The state of practice in tunnel boring process is to estimate subsurface geological and geotechnical layer information by drilling several boreholes along the estimated tunnel path. The data extracted from boreholes reflect precise information about the soil type at different depths and the location of transition layers as documented in the geotechnical data report (GDR). However, the profile of earth layers between borehole locations is still unknown. Therefore, the geological profile produced and typically provided in the geotechnical baseline report (GBR) carries significant uncertainty. Several geospatial analysis methods have been developed in the literature for threedimensional visualization of subsurface geological layers. However, all of the aforementioned models are associated with some level of uncertainty that impacts the risk involved with underground tunneling. An example of interpolating geological data at borehole locations is using kriging algorithm (Oliver and Webster, 1990). Figure 2 shows an example of interpolating borehole data and three-dimensional visualization of tunnel alignment through geological layers estimated from borehole data.

Several studies in the literature have focused on geospatial analysis of geological earth layers from drilled borehole data. Kavoura et al. (2016) studied three-dimensional geological modelling from borehole data using geographic information system (GIS) and remote sensing. A digital surface model was developed to represent the geological layer properties between borehole locations. Xiong et al. (2017) proposed a three-dimensional multi-scale geology modelling methodology for risk assessment of tunneling. They used the hermit radial basis function and Monte Carlo simulation for dynamic risk evaluation during tunneling operations. Their model includes several scales including regional sub-model for preliminary evaluation and outcrop scale for dynamic evaluation. Although their model showed significant success in risk management of a tunneling case study, they recommend that there is a need for advanced geological data prediction.

Among methods that are mostly implemented for geospatial correlation of georeferenced data (i.e. borehole drilling location) kriging is an efficient and accurate geostatistical algorithm. It generates an estimated surface from scatter data points. This interpolation method weights the neighboring measurements to predict an unmeasured location as follows:

$$
\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)
$$
 (1)

where  $Z(s_i)$  is the measured values at location *i*,  $s_0$  is the prediction location, *n* is the number of measured locations, and  $\lambda i$  is the unknown weight for the measured value at location  $i$ . The accuracy of the predicted values is measured with semivariance defined as the squared difference between the values of paired locations.

One of the main objectives of this project was to develop a real-time prediction algorithm that can predict the geological and geotechnical properties of earth layers ahead of TBM cutter-head during a tunneling operation using information from boreholes and operational TBM data. The process of data extraction, pre and post-processing, model development and testing of the developed model is presented in the following sections.



UTC-UTI AND A 14 Figure 2 – (a) Interpolation of geological data to generate tunneling operation data (after Sun et al. 2018), (b) 3D visualization of tunnel alignment with borehole geotechnical data (after Ozmutlu and Hack, 2003), (c) Integration of data-driven model for prediction of geological information (after Zhao et al. 2019)

## **7. CHAPTER 2 – DATA PROCESSING**

#### <span id="page-14-1"></span><span id="page-14-0"></span>**7.1. Datasets**

The data used in this project was extracted from the Seattle Northgate Link Extension tunneling project in North America. The dataset includes 30 GB of data samples collected from a Hitachi Zosen TBM during the boring process. The Northgate Link Extension will extend service north from the University of Washington to the University District, Roosevelt, and Northgate neighborhoods by 2021, and is expected to cost approximately \$2.1 Billion. Most of this 6.9 km extension will be underground and includes the construction of 5.6 km of twin Earth pressure balance (EPB) tunnels. Also included are the excavations of the Maple Leaf Portal (MLP) where the light rail will transition from tunnels to elevated guide-way and two large underground station boxes, one for the University District Station (UDS) and one for the Roosevelt Station (RVS). The N125 tunnels are excavated through glacial and non-glacial sediments of the Puget Trough deposited during the Quaternary and Holocene periods. The Quaternary sediments are generally overconsolidated due to several glaciations, while the recent Holocene sediments are normally consolidated. The Engineering Soil Units (ESU) defined for this study are Engineering and Non-Engineered Fill (ENF), Recent Granular Deposits (RGD), Recent Clays and Silts (RCS), Till and Till-Like Deposits (TLD), Cohesionless Sand and Gravel (CSG), Cohesionless Silt and Fine Sand (CSF), and Cohesive Clay and Silt (CCS). ENF, RGD, and RCS are recent, normally consolidated sediments, whereas TLD, CSG, CSF, and CCS are glacial, overconsolidated sediments (Northlink Tunnel Partners, 2009). Figure 3 shows the geological profile of the project.

The prediction targets (output labels) for this dataset are the percentage of each soil component within the TBM tunnel envelope during excavation. The total composition includes the following four geomaterial types: Cohesive Clay and Silt (CCS), Cohesionless Silt and Fine Sand (CSF), Cohesionless Sand and Gravel (CSG), and Till-Like Deposits (TLD). The sum of CCS, CSF, CSG and TLD layer percentages at all times is assumed to be 100 percent. Since the construction sequence is evaluated by each tunnel ring, one set of labels is generated per tunnel ring. Only a few tunnel rings intersect with the location of drilled boreholes where the geological composition is accurately measured. Those rings were used for training the model. The following sections explains the process of selecting features and extracting the required data from the TBM dataset.

#### <span id="page-14-2"></span>**7.2. Feature Extraction and Selection for Geological Composition**

The TBM dataset includes both operational parameters and sensor measurements. In total, about 1000 data elements are recorded in 5-second time intervals. We extracted and selected a set of features from the raw data to be used in the predictive model. The selected features include, cutterhead torque normalized by average excavation chamber pressure, total thrust force normalized by average chamber pressure, screw conveyor torque normalized by average screw conveyor pressure, foam injection volume, additive injection volume, cutterhead revolution speed, screw conveyor revolution speed, TBM advance rate, average chamber pressure, apparent muck unit weight, screw conveyor pressure, average shield pressure, and front body rolling rate normalized by shield pressure. The data samples collected during excavation of each ring can be aggregated and represented by a new set of statistical features. In this approach, the statistical changes in sensor measurements will be reflected as new features while allowing for the interval to change from a time-based interval to ring-based interval. The statistically derived features are kurtosis, skewness, maximum, minimum, mean, median, standard deviation, quartiles, and pairwise approximation aggregate.



Figure 3. Geological profile of the tunneling project

 The drilled boreholes contain the actual geological composition. Chainage was used to match rings with boreholes, so that the TBM features could be associated with the accurate geological borehole sampling at the appropriate ring location. To keep the sampling as relevant as possible to the tunnel, only boreholes within 200 feet of the tunnel were selected. Ordering those borehole-associated rings allowed the construction of a continuous series of rings as a training set. The rings not associated with boreholes were then used as an evaluation dataset. The following section include the steps taken for preprocessing of the input feature before training the model.

Principal Component Analysis (PCA**)** algorithm is widely used in machine learning processes to reduce the dimensions of the large datasets. PCA uses an orthogonal transformation to convert a set of possibly correlated variables into a set of linearly uncorrelated values called principal components (Ding and He, 2004). The feature matrix  $(X_{train})$  is first normalized using min-max scaling with the range  $(R)$  from -1 to 1. The transformation used to scale  $(X_{train})$  is then applied to ( $X_{test}$ ) matrix. PCA is used to further reduce the dimensionality of the scaled ( $X_{train}$ ), and the transformation is then applied on the scaled matrix  $(X_{test})$ .

$$
X_{std} = \frac{X - X_{min}}{X_{max} - X_{min}}\tag{2}
$$

$$
X_{scaled} = X_{std} * (R_{max} - R_{min}) + R_{min}
$$
\n(3)

where  $X_{std}$  = standard deviation of the feature,  $X_{max}$  = maximum value for the feature,  $X_{min}$  = minimum value for the feature, and *Xscaled* = scaled feature.

The transformation derived and applied to the training matrix is also applied to the testing matrix. The input shape passed into the model is:

$$
S \times T \times F \tag{4}
$$

where *S* represents the number of samples, *T* is the number of time steps of model, and *F* is the number of features (the details are provided in the following section). The features matrices  $X_{train}$  and  $X_{test}$  only have  $S \times F$  dimensionality and must be further modified by adding the time dimension. This is accomplished by using the features from previous samples as the time steps. However, the labels  $(y)$  are not part of the feature space (see Table 1).

<b>Time Step</b>	Label
$F_{t-2}$ $F_{t-1}$ , $F_t$	$y_t$
$F_{t-1}$ $F_t$ , $F_{t+1}$	$y_{t+1}$
$F_t$ $F_{t+1}$ , $F_{t+2}$	$y_{t+2}$
	÷
$F_{n-2}$ $F_{n-1}$ , $F_n$	$\mathcal{V}_n$

**Table 1 – Sequence of time steps and labels**

#### <span id="page-16-0"></span>**7.3. Feature Extraction and Selection for TBM Performance and State**

To prevent tautological bias, a correlation heatmap was created to identify features which were directly mapped to one another. Certain features which were being derived directly from other sources were removed to prevent any overlap or bias between features and labels. Some TBM sensors collected data per ring rather than collecting continous operation data similar to other TBM sensors, which generated an irregularity in sampling rate. To address this issue, several statically derived features were calculated based on the collected data within each ring to aggregate and compress their information and append it to the ring they were associated with. The statistical

features derived from the intra-ring samples were mean, median, range, max, min, kurtosis, skewness, standard deviation, and quartiles.

Addition of each of these features per TBM sensor compensated for the loss of highdefinition that had been provided by the intra-ring samples, while matching the sampling rate to that of the lower-rate sensors. The feature data is then normalized to avoid any scaling concerns (1).

The TBM operation data processed through a Recurrent Neural Network (RNN) to predict performance parameters. The key characteristic of the RNN is its ability to take samples in previous timesteps and utilize that data to predict the samples in the future. In order to do this, the data is reshaped into a three dimensional matrix, with dimensions: features, samples, and time steps. Figure 4 shows the schematic of data format.



Figure 4. Input Data format for RNN model

The prediction for time *t* is made only based on previous timesteps *t*-3, *t*-2, and *t*-1. This approach enables the model to predict features denoted by labels ahead of the machine at the future timestep *t*. In this study, we utilized 3 previous timesteps, since passed that threshold, noise will be added to the data. This approach was applied to both datasets. The availability of two datasets for the adjacent tunnels can be used to verify the results of trained model. By utilizing the same neural network structure on both tunnels, and verifying the same levels of accuracy, we were able to determine if the model could be applicable to a different excavation environment.

UTC-UTI 18 Another application is the ability of the network to derive generalized inferences which can be applied to other tunnels with minimal, or even zero, training. If the model can train on one tunnel and apply the training to the other tunnel without even having been exposed to the other tunnel conditions, it would have derived intra-sensor relations in a manner such that it is entirely environment independent. Such a result would imply that this model can be used to improve and predict TBM performance on new tunnels with near zero prior data collection from the new

worksite. The model could predict accurately having only trained on prior excavations, without being re-trained on the new environment. Eliminating the need for prior analysis, or at least reducing it, greatly advances current construction techniques.

To evaluate these scenarios, the model undergoes 4 experiments. First, it is trained and tested only on a data split of the first tunnel. Second, it is trained and tested on a split of the second tunnel to verify the adaptability of the model structure. Third, it is trained on the first tunnel and tested on the second tunnel without ever having trained on any data from the second tunnel. Finally, the model utilizes all the data from the first tunnel and a minimal amount of data from the second tunnel (<15%) to attempt to achieve results comparable to having been trained only on that tunnel.

## **8. CHAPTER 3 – PREDICTIVE MODEL**

#### <span id="page-19-1"></span><span id="page-19-0"></span>**8.1. Predictive Model for TBM Performance and State Prediction**

The data is used to train an Artificial Neural Network (ANN). An ANN is generally used to model complex relationships between variables, using a multilayer system with weighted connections. The model is trained on the provided data to adjust both the layers as well as the weights between layers. Standard ANNs are incapable of handling temporal connections. Thus, in this study we use Recurrent Neural Networks (RNN), a subcategory of ANN, for their ability to model time-based connections using a memory system. In this case, the temporal connection is between the past samples and the current one, as they are directly related to one another. Figure 5 shows the structure of an RNN.



Figure 5. Structure of a recurrent neural network (RNN)

Each of the recurrent neurons within the recurrent layer can be structured differently depending on the type of RNN. In this case, a Gated Recurrent Unit (GRU) is used. GRU is a relatively simple recurrent neuron, and one of the more recent developments in this subcategory. It acts as a set of memory cells, each with an input gate and a "forget" gate. The cell remembers information to train the network. The input gate filters the information to be added to the cell while the forget gate chooses information to drop. This allows the cell to derive long-term relations between values while avoiding overly specific short-term relations. This is especially critical in our application (tunneling data), where simple performance relations may be present in a certain subset of geological environment, but not present in the overall pattern. To avoid allowing the predictive system to fall into these sub-relations, rather than finding an overall pattern, the forget system is important to the model.

Another model that was employed was the Long Short-Term Memory Network (LSTM). LSTM adds an additional "output" gate to the GRU cell, to filter the information outflow.

#### UTC-UTI 20

However, the additional complexity of the LSTM necessitates a larger amount of training data. The GRU handles the reduced quantity of data better, as it is a simple system with less mechanisms to train. Figure 6 visualizes the difference between an LSTM and GRU unit.



Figure 6. A comparison of GRU and LSTM recurrent neurons.

To assess the model's performance, we split the data into training and validation sets. This separation is defined prior to the reshaping to prevent any overlap or relationship between the two sets, ensuring the model is only predicting between features, rather than connecting the training data directly to testing. Separating prior to reshaping prevents any of the past timesteps in the early samples of the validation data overlapping with the late samples in the training set. Therefore, some of the early samples in the validation set must simply be dropped as they do not have enough past samples from their own validation set to be reshaped. To avoid a lack of testing data, the separation point must be adjusted to favor the validation set more than the standard split.

A search through various optimizers and models showed that the ideal optimizer and error loss for this particular system was the Adam optimizer with an MAE error loss. To handle the relative lack of data, a Gaussian noise filter is introduced during the training stage. The validation set, however, only uses the actual values from the machine. The predictions are then evaluated using Root Mean Squared Error (RMSE).

### <span id="page-20-0"></span>**8.2. Predictive Model for Geological Composition Prediction**

The sequential estimation of geological composition are performed using an Artificial Neural Network (ANN). ANNs are able to model complex non-linear relationships between several input features and output labels using a set of pseudo variables in middle hidden layers. In this study, the output labels are percentages of each soil type in the tunneling profile. In a basic feedforward ANN, the layers are connected using weighted links. During the training phase, the network updates the weights to find the best fit with the training data.

UTC-UTI 21

Since our dataset contains both spatial and temporal correlations, a regular feedforward ANN cannot properly trained to perform predictions. To address this concern, the TBM data was treated as a time series dataset. Then, a modified type of ANN, known as Recurrent Neural Networks (RNN) that is ideal for learning and predicting patterns in time series data, was developed. RNNs are a type of ANN that hold an internal state. The weights are not updated past the training phase but the states change with every prediction and then both are used to make the next prediction. This allows RNNs to make predictions on sequential data with greater accuracy than traditional neural networks.

To accommodate for sequential nature of the tunneling data, a specific type of RNNs known as Long Short-Term Memory (LSTM) was employed in this study. LSTMs are able to hold onto much longer temporal relationships due to the ability to forget irrelevant information and use parts of its internal state to make predictions. By discarding irrelevant information, the RNN-LSTM is able to use more important parts of the internal state to make predictions without additional noise from patterns existing in other temporal segments that do not apply over the long term. RNN-LSTM use mechanisms that help decide what information is relevant to store and to use in making future predictions. The information that is not important does not alter the internal state as much as it would in a normal RNN. However, the relevant parts of the state are used to make the prediction, similar to a normal RNN. Figure 7 illustrates a schematic structures of a regular feedforward ANN compared to RNN.



Figure 7. General topographical difference between a) ANN, b) RNN

UTC-UTI 22 Walk-forward validation is used to assess the performance of the model for the testing dataset. It takes a part of dataset to optimize the system and then use another part to validate. As noted in the previous section, the predicted labels  $(y)$  will not be included in the feature space. Instead we will depend on the stateful nature of an LSTM and use previous states to help make future predictions. The state of the LSTM is built up using the training data as a starting point and is maintained between predictions to make future predictions. The training tunnel rings are located within various chainage from each other but are treated sequentially for model training purposes. However, treating all of the training rings as continuous adjacent samples introduces variance to the model. This means that the training data is noisier than the testing data.

A portion of the testing data is used for validation and adjusting the neural network parameters. Selection of validation data better reflects how the model will perform on the testing data. We used Grid Search algorithm, also known as hyperparameter optimization or tuning, on a set of preliminary models to select the best optimizer and loss function for the RNN as well as other hyperparameters (Claesen and Moor, 2015). A hyperparameter is a value used for controlling the learning process only. Based on initial evaluations, the Adam Optimizer algorithm (Kingma and Ba, 2014), which is used to update network weights in training process, was selected for the optimization process and Mean Squared Error (MSE) was selected as the loss function. In every subsequent model we trained, the Adam optimizer is used during the backpropagation process. MSE is defined as follows:

$$
MSE = \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2
$$
 (5)

where y is the actual value of the label and  $\hat{y}$  is the predicted value. Root Mean Squared Error (RMSE) is used to check the performance of the neural network defined as follows:

$$
RMSE = \sqrt{MSE} \tag{6}
$$

The noise introduced to the model during the training phase is present during the predictions on the testing data. To evaluate the results on testing data, Moving Average (MA) was employed as a low pass filter to accommodate for the noise that is expressed in the raw predictions. MA of ring *n* and ring  $n+1$  are defined as follows:

$$
\overline{R}_n = \frac{R_{n-1} + R_{n-2} + R_{n-3} + \dots + R_{n-l}}{l} = \frac{1}{l} \sum_{i=0}^l R_{n-i}
$$
(7)

$$
\overline{R}_{n+1} = \overline{R}_n + \frac{R_{n+1}}{l} - \frac{R_{n-l}}{l} \tag{8}
$$

where *l* is the moving average size, and  $R_n$  is the  $n<sup>th</sup>$  ring. The normalized predications are then converted back to the original scale. For prediction *n*:

$$
\frac{cSS_n}{cSS_n + cSF_n + cSG_n + TLD_n} \times 100 = CSS_n \tag{9}
$$

$$
\frac{c_{SF_n}}{c_{SS_n} + c_{SF_n} + c_{SG_n} + TLD_n} \times 100 = CSF_n
$$
\n(10)

$$
\frac{csc_n}{csc_n + csc_n + rL_n} \times 100 = CSG_n \tag{11}
$$

$$
\frac{TLD_n}{\text{CSS}_n + \text{CSS}_n + \text{CSG}_n + TLD_n} \times 100 = TLD_n \tag{12}
$$

where  $CSS_n + CSF_n + CSG_n + TLD_n$  is equal to 100.

### **9. CHAPTER 4 – RESULTS AND DISCUSSION**

#### <span id="page-23-1"></span><span id="page-23-0"></span>**9.1. Prediction Results for TBM Performance and State**

The model underwent four scenarios. Training and testing on the first tunnel, training and testing on the second tunnel, training only on the first tunnel data and then testing on the second tunnel, and finally training on the first tunnel and a small (<15%) portion of the second tunnel data and then testing on the remainder of the second tunnel dataset.

The results of the model performance in predicting key features of the first tunnel are displayed in Figure 8, compared against the actual machine outputs. The RMSE values for each of the variables are displayed in Table 2, along with the normalized RMSE (RMSE divided by the mean of the parameter). This is necessary since the machine outputs are measured at different scales and need a scale independent metric.



Figure 8. Comparisons of predictions against actual sensor outputs at future rings. The predictive model is trained and tested on separate datasets from the first tunnel.

<b>Predicted Feature</b>	<b>RMSE</b>	<b>NRMSE</b>
Advance Speed (mm/min)	10.78	0.159
Articulation Force (kN)	1201.99	0.171
Cutterhead Speed (rpm)	0.026	0.226
Rotation Cutter Speed (rpm)	0.472	0.22
Thrust Force (kN)	3385.84	0.366

**Table 2 – Results of the model trained and tested on the first tunnel alone**

Prediction of critical features such as advance speed, thrust force, and articulation force showed noticeable results. Estimation of these features could be used to inform and improve the operation of the TBM. Since each of these predictors is running on the same dataset consisting of the past inputs, with no change being made per label, these predictions can all be run simultaneously to the same degree of accuracy, meaning that the predictor can provide an entire view of all the parameters as they will appear at the next ring. The model shows exceptional abilities in terms of predicting sensor features, particularly considering that it is doing so without the knowledge of the other sensor data at the same ring. The model holds its performance quite consistently throughout several runs.

The same tests were performed on the model being trained and tested only on the second tunnel. Figure 9 and Table 3 provide the results, which are quite similar in scope and accuracy to the results of the first scenario.



Figure 9. Predictions versus actual values. The predictive model is trained and tested on separate datasets from the second tunnel.

<b>Predicted Feature</b>	<b>RMSE</b>	<b>NRMSE</b>
Advance Speed (mm/min)	9.34	0.166
Articulation Force (kN)	2144.4	0.283
Cutterhead Speed (rpm)	0.209	0.259
<b>Cutter Rotation Speed (rpm)</b>	0.457	0.212
Thrust Force (kN)	2606.23	0.248

**Table 3 – Results of training and testing only on the second tunnel**

In the third experiment, the model was trained on the entirety of the first tunnel and approximately 150 rings of the second tunnel, leaving about 900 rings for validation. In order to give a balanced comparison to the model evaluated in the second scenario, this model is not tested on all 900 rings, but only on the last 100 rings, the same ones that second would have evaluated upon. Figure 10 and Table 4 shows the prediction results and accuracy.



Figure 10. Predictions of the model having been trained on the first tunnel plus less than 15% of the second tunnel.





Finally, the model was trained only on the first tunnel data and tested on the second tunnel dataset. In order to give a balanced comparison to the model evaluated in test 2 and 3, this model is tested on the last 100 rings, the same ones that scenario 2 and 3 would have evaluated upon. Figure 11 provides the results of several key features, and Table 5 provides key metrics on these results.



Figure 11. Predictions based on model training only on the first tunnel and testing only on the selected portion of the second tunnel used in Figure 8.

<b>Predicted Feature</b>	<b>RMSE</b>	<b>NRMSE</b>
Advance Speed (mm/min)	8.10	0.144
Articulation Force (kN)	1366.66	0.180
Cutterhead Speed (rpm)	0.248	0.308
Rotation Cutter Speed (rpm)	0.56	0.263
Thrust Force (kN)	2073.474	0.197

**Table 5 – Results of training on the first tunnel and testing on the second.**

#### <span id="page-28-0"></span>**9.2. Prediction Results for Geological and Soil Composition**

Figure 12a shows the prediction results for soil composition using the proposed model with a trained RNN-LSTM containing three hidden layers. It is important and interesting to notice that the proposed method (RNN-LSTM method) does not use the bore-hole data as input information to make prediction for the soil composition. It only uses the Shield Tunnel Boring Machine Data as input features. However, the predictions are comparable to the Kriging interpolation values, which are derived based on bore-hole samples (Figure 12b). In other word, the machine learning model has discovered a correlation between TBM data and soil composition, and learned to use it to predict the soil composition a head of TBM, even without any prior information about the geological composition (i.e. without using prior borehole data). Figure 13 shows the RNN-LSTM prediction results during the boring process for each soil composition (i.e., CCS, CSF, CSG, TLD). In Figures 12 and 13, the vertical axis shows the percent of the soil composition, and the horizontal axis shows the ring number. It is very important to notice that this figure compares the RNN-LSTM prediction method versus the Kriging interpolation. However, none of them are the ground truth (in this case, the actual ground truth soil conditions are unknown).



Figure 12. Stack plot of component predictions from a) the proposed RNN-LSTM, and b) Kriging interpolation from borehole data.

The ability of the RNN-LSTM to make predictions on individual components is notable (Figure 13). Even when the magnitude of the changes is different the model captures the pattern changes very well.





Figure 13. Individual earth layer estimations from RNN-LSTM model compared to labeled data from Kriging interpolation of borehole data.

UTC-UTI 31

## **10. CHAPTER 5 – SUMMARY AND CONCLUSION**

<span id="page-31-0"></span>This research project focuse on the applications of data science, machine learning, and big data analytics in the construction, maintenance and performance of the underground transportation infrastructure. The first objective of this project is to develop advanced data mining and novel machine learning based methods for predicting or detecting ground conditions and geological composition using the data collected before and during the TBM operations. The second objective was to design and develop data-driven predictive models that can predict the TBM state, performance, and status in real-time.

The inability to identify and characterize the as-encountered ground for excavations using Pressure balance shield tunnel boring machines (TBM) makes it difficult to optimize the tunneling process. There is significant incentive to develop methodologies that can characterize the ground using the large volume of data collected during TBM operation. With the recent advancements in artificial intelligence and machine learning for solving complex problems, there is a potential for applications in tunneling using TBM data. The first objective of this study was to develop an advanced machine learning algorithm that is capable of sequentially estimating the geological composition of earth layers encountered by the TBM during tunneling. The data used in this project was extracted from the Seattle Northgate Link Extension tunneling project in North America. The prediction targets for this dataset were the percentage of each soil component within the TBM tunnel envelope during excavation. The prediction model was developed using an Artificial Neural Network (ANN). Due to the sequential nature of the TBM operation, the collected data was treated as a time series. Therefore, a specific type of ANN known as Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), was employed in this study.

It is important to notice that the proposed RNN-LSTM method only uses the TBM data as input features to make prediction for the soil composition (it does not directly use bore-hole data as input information). However, the comparison of model-estimated as-encountered geological composition of the earth layers with interpolated data from actual borehole samples shows an agreement on the soil composition and pattern. In other word, the proposed RNN-LSTM model has discovered a correlation between TBM data and soil composition, and learned to use it to predict the soil composition, even without any prior information about the geological composition (i.e. without using prior borehole data). The performance of the model in terms of mean squared error was comparable with interpolated data. Since the common practice is only relying on interpolated borehole data during TBM operation, it cannot provide the unpredicted ground conditions between borehole locations that can infer additional cost due to project delays and equipment maintenance. Comparing the model estimations for each individual soil type with provided labels show even a higher accuracy in most cases. The prediction performance of the model can be further improved by providing more training data with higher level of details regarding the geological and geotechnical parameters of the tunneling alignment.

The second objective of this project was to design and develop data-driven predictive models that can predict the TBM state, performance, and status in real-time. The results of this project suggest that utilizing recurrent neural networks would allow for prediction of critical excavation performance evaluation features at future states. The structure of the model appears to be applicable to different tunneling zones, as verified by the application of the model to two different

tunnels. In addition, the model proved to be accurate in applying learning from one tunnel to another despite the two having different soil composition. Providing light re-training even allowed it to outperform the model fully trained on the tunnel from where the testing data came. Prediction of critical features such as advance speed, thrust force, and articulation force showed noticeable results. The model shows exceptional abilities in terms of predicting sensor features, particularly considering that it is doing so without the knowledge of the other sensor data at the same ring. The model holds its performance quite consistently throughout several runs. The success of the model implies a transferability of training of the RNN predictive model. Such generalization is a feature that will be very useful in optimizing the TBM operation.

## **11. REFERENCES**

- <span id="page-33-0"></span>- Avunduk, E., and Copur, H. Empirical modeling for predicting excavation performance of EPB TBM based on soil properties. *Tunnelling and Underground Space Technology,* Vol. 71, 2018, pp. 340-353.
- Bilgin, N., Copur, H., Balci, C. Effect of replacing disc cutters with chisel tools on performance of a TBM in difficult ground conditions. *Tunnelling and Underground Space Technology,* Vol. 27, 2012, pp. 41-51.
- Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*.
- Ding, C., & He, X. (2004, July). K-means clustering via principal component analysis. In Proceedings of the twenty-first international conference on Machine learning (p. 29). *ACM*.
- Farrokh, E., Rostami, J., Laughton, C. Study of various models for estimation of penetration rate of hard rock TBMs. *Tunnelling and Underground Space Technology,* Vol. 30, 2012, pp. 110- 123.
- Galván, A., Peña, F., & Moreno-Martínez, J. Y. (2017). Effect of TBM advance in the structural response of segmental tunnel lining. *International Journal of Geomechanics*, *17*(9), 04017056.
- Hassanpour, J., Vanani, A. G., Rostami, J., & Cheshomi, A. (2016). Evaluation of common TBM performance prediction models based on field data from the second lot of Zagros water conveyance tunnel (ZWCT2). *Tunnelling and Underground Space Technology*, *52*, 147-156.
- Kavoura, K., Kordouli, M., Nikolakopoulos, K., Elias, P., Sykioti, O., Tsagaris, V. & Anastasopoulos, V. (2014, August). Subsurface geological modeling using GIS and remote sensing data: a case study from Platanos landslide, Western Greece. *Second International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2014)* (Vol. 9229, p. 92290C). International Society for Optics and Photonics.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Maher, J. (2012). A Machine Learning Approach to Predicting and Maximizing Penetration Rates in Earth Pressure Balance Tunnel Boring Machines, *Association for Computing Machinery, ACM*.
- Mooney, M. A., Walter, B., & Frenzel, C. (2012). Real-time tunnel boring machine monitoring: A state of the art review. *North American Tunnelling*, 2012 proceedings, 73-81.
- Nagrecha, K. L. Fisher, M. Mooney, E. Alavi, T. Rodriguez-Nikl, M. Mazari, M. Pourhomayoun, "As-Encountered Prediction of Tunnel Boring Machine Performance Parameters Using Recurrent Neural Networks," Transportation Research Board Annual Conference (TRB 2020).
- Northlink Tunnel Partners (2009). Geotechnical Baseline Report Univerity Link 230. Seattle, WA, *Northlink Tunnel Partners*. 6: 66.
- Ozmutlu, S., & Hack, R. (2003). 3D modelling system for ground engineering. *New Paradigms in subsurface prediction* pp. 253-26). Springer, Berlin, Heidelberg.
- Rostami, J., & Chang, S. H. (2017). A Closer Look at the Design of Cutterheads for Hard Rock Tunnel-Boring Machines. *Engineering*, *3*(6), 892-904.
- Salimi, A., Rostami, J., Moormann, C., & Delisio, A. (2016). Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs. *Tunnelling and Underground Space Technology*, *58*, 236-246.
- Schaeffer, K., & Mooney, M. A. (2016). Examining the influence of TBM-ground interaction on electrical resistivity imaging ahead of the TBM. *Tunnelling and Underground Space Technology*, 58, 82-98.
- Shi, M., Song, X., & Sun, W. (2018). Geology prediction based on operation data of TBM: comparison between deep neural network and statistical learning methods. *arXiv preprint arXiv:1809.06688*.
- Sun, J., Li, T., Yu, M., & Zhang, H. M. (2018). Exploring the congestion pattern at long-queued tunnel sag and increasing the efficiency by control. *IEEE Transactions on Intelligent Transportation Systems*, 19(12), 3765-3774.
- Swinnen, G., Thorbecke, J.W., and Drijkoningen, G.G. Seismic Imaging from a TBM. *Rock Mechanics and Rock Engineering*, Vol. 40, No. 6, 2007, pp. 577–590.
- Toth, A., Gong. Q., and Zhao, J. Case Studies of TBM tunneling performance in rock-soil interface mixed ground. *Tunnelling and Underground Space Technology*, Vol. 38, 2013, pp. 140–150.
- Xiong, Z., Guo, J., Xia, Y., Lu, H., Wang, M., & Shi, S. (2018). A 3D Multi-scale geology modeling method for tunnel engineering risk assessment. *Tunnelling and Underground Space Technology*, 73, 71-81.
- Yi, C., Lu, D., Xie, Q., Liu, S., Li, H., Wei, M., & Wang, J. (2019). Hierarchical tunnel modeling from 3D raw LiDAR point cloud. *Computer-Aided Design*, *114*, 143-154
- Zhao, J., Shi, M., Hu, G., Song, X., Zhang, C., Tao, D., & Wu, W. (2019). A Data-Driven Framework for Tunnel Geological-Type Prediction Based on TBM Operating Data. *IEEE Access*.

# <span id="page-35-0"></span>**12. APPENDIX A – TECHNOLOGY TRANSFER ACTIVITIES**

### **1 Accomplishments**

- Developed novel prediction techniques based on advanced machine learning for predicting or detecting ground conditions and soil composition using the data collected before and during the TBM operations.
- Designed and developed data-driven predictive models that can predict the TBM state and performance in real-time (during the boring process) as well as adverse events in UTI such as structural defects, and anomalies.
- We bublished a journal paper in The Journal of the Transportation Research Board 2020
- One UTC-UTI student Presented his research in TRB conference 2020
- One UTC-UTI student Presented her research in Women in Data Science Conference 2020
- Two UTC-UTI students Presented her research in Cal State LA DIRECT STEM workshop
- We have been reaching out to LA Metro Rail, collected new big datasets for training modeling, and validating the developed algorithms, and also discussed potential collaborations, discussed potential project detail.
- Mohammad Pourhomayoun received new grants from NASA for projects focusing on various applications of AI in Urban Sustainability.
- Mohammad Pourhomayoun received new grants from Sikand Foundation focusing on various applications of AI in Urban Sustainability.

### **1.1 What was done? What was learned?**

In this project, we have developed and applied Data Science, Artificial Intelligence, and big data analytics techniques for construction and maintenance of the underground transportation infrastructure. The first objective of this project was to develop data mining and machine learning methods for predicting or detecting ground conditions. The second objective was to design and develop data-driven predictive models and Artificial Intelligence (AI) techniques to predict the TBM state and status in real-time. The third objective was to design and develop predictive models and AI techniques to predict future adverse events in UTI such as structural defects and anomalies, and defect progression and consequences over time. We learned that data science and AI can be beneficial tools and techniques to improve the performance, quality, and efficiency of construction and maintenance of the underground transportation infrastructure. They can let us predict important parameters and states, and also predict unexpected adverse events during the construction or afterwards.

#### **1.2 How have the results been disseminated?**

• We published a journal paper in The Journal of the Transportation Research Board 2020

- One UTC-UTI student Presented his research in TRB conference 2020
- One UTC-UTI student Presented her research in Women in Data Science Conference 2020
- Two UTC-UTI students Presented her research in Cal State LA DIRECT STEM workshop

### **2 Participants and Collaborating Organizations**

Name: LA Metro

Location: Los Angeles, CA

Contribution: Collaborations with LA Metro Rail, collected new big datasets for training modeling, and validating the developed algorithms, and also discussed potential collaborations, discussed potential project detail.

### **3 Outputs**

#### *Journal publications*

K. Nagrecha, L. Fisher, M. Mooney, E. Alavi, T. Rodriguez-Nikl, M. Mazari, M. Pourhomayoun, "As-Encountered Prediction of Tunnel Boring Machine Performance Parameters Using Recurrent Neural Networks," The Journal of the Transportation Research Board, July 2020. [\(https://doi.org/10.1177/0361198120934796\)](https://doi.org/10.1177/0361198120934796)

#### *Presentations*

K. Nagrecha, L. Fisher, M. Mooney, E. Alavi, T. Rodriguez-Nikl, M. Mazari, M. Pourhomayoun, "As-Encountered Prediction of Tunnel Boring Machine Performance Parameters Using Recurrent Neural Networks," Transportation Research Board Annual Conference (TRB 2020), July 2020.

#### *Workshops*

L. Fisher, K. Nagrecha, T. Rodriguez-Nikl, M. Mazari, M. Pourhomayoun, "Real-Time Prediction of Geological Composition using Recurrent Neural Networks and Shield Tunnel Boring Machine Data," Cal State LA DIRECT STEM workshop 2020.

E. Estrada Medina, K. Nagrecha, T. Rodriguez-Nikl, M. Mazari, M. Pourhomayoun, "Prediction of Soil Composition using Artificial Neural Networks," Women in Data Science workshop 2020.

#### **4 Outcomes**

We have developed novel prediction techniques for predicting or detecting ground conditions and soil composition using the data collected before and during the TBM operations. We have also designed and developed data-driven predictive models that can predict the TBM state and performance in real-time (during the boring process) as well as adverse events in UTI such as structural defects, and anomalies. They can let us predict important parameters and states, and also predict unexpected adverse events during the construction or afterwards.

### **5 Impacts**

The developed systems and methods are significantly effective in improving the performance, quality, and efficiency of construction and maintenance of the underground transportation infrastructure.

# **13. APPENDIX B - DATA FROM THE PROJECT**

<span id="page-38-0"></span>

### **Table1: Sample sensor data used to develop and train machine learning models**





































