CTIPS Center for Transformative Infrastructure Preservation and Sustainability
USDOT Region 8 University Transportation Center

Data-Driven Inspection Planning for Utah Culverts Using Federated Learning

CTIPS-005 – Full Project Description

Approved 5/6/2024

University

University of Utah

Principal Investigators

Abbas Rashidi, Ph.D.
Associate Professor
Dept. of Civil and Environmental Engineering
University of Utah
Phone: (801) 581-3155
Email: abbas.rashidi@utah.edu
ORCID: 0000-0002-4342-0588

Research Needs

Managing and maintaining critical transportation infrastructure components such as culverts are associated with significant challenges that directly impact public safety, environmental sustainability, and economic efficiency. Culverts, serving as conduits under roadways and railways for water flow (Figure 1), are susceptible to deterioration over time due to factors like weather conditions, material wear, and load stresses [1]. The failure to address the deteriorating condition of culverts in a timely manner can lead to severe consequences, including roadway collapses, flooding, environmental damage, and costly emergency repairs. To mitigate these risks and ensure the long-term functionality of the transportation network, a systematic approach to culvert condition assessment and maintenance is paramount [2]. Unfortunately, traditional culvert inspection and management practices are no longer appropriate to meet the growing demands of modern infrastructure networks due to their reactive nature, high labor costs, and the disruption they cause to traffic flow [3].

Figure 1. Culvert under a road

A strategy to enhance traditional culvert inspection and management practices lies in employing the power of artificial intelligence. Specifically, innovative machine learning (ML) algorithms have opened up fresh avenues for creating more efficient and powerful strategies for inspecting and managing culverts [4]. ML algorithms are capable of making accurate predictions. By analyzing historical inspection data, weather patterns, and environmental factors, ML models can predict the likelihood of culvert deterioration. This predictive capability enables a proactive approach to maintenance, allowing agencies to prioritize repairs before critical failures occur, minimizing disruptions, and optimizing resource allocation [5].

While data-driven approaches like condition prediction models offer the potential to transform infrastructure asset management, many DOTs face a major obstacle: data shortfalls. This lack of sufficient data can stem from several factors, including the high cost and resource demands of traditional data collection methods, as well as the absence of a centralized data management system [6]. For example, the Utah Department of Transportation (UDOT) currently lacks a comprehensive culvert management system. This results in sparse and inconsistent inspection records for its extensive culvert inventory with more than 47,000 culverts across the state, hindering the development of robust predictive models [7], [8].

Data scarcity poses a major obstacle for UDOT seeking to utilize predictive models in culvert asset management. Insufficient data limits the accuracy of models, prevents the use of advanced ML techniques, and can lead to overfitting. Tackling the issue of data scarcity for utilizing ML in managing culverts encompasses various strategic methods, including improving data collection techniques, partnering with other DOTs, and optimizing the use of available data through advanced analytics [9]. A common and practical method among DOTs is to merge inventory data from other DOTs into their existing dataset. Such collaborative data sharing significantly enhances the diversity of datasets for ML applications, offering a richer perspective on the demands and usage patterns of culverts. Nonetheless, this approach encounters three primary challenges. First, inconsistent data formats and protocols make combining datasets difficult. Second, and most significantly, data privacy regulations like the General Data Protection Regulation (GDPR) [10] and the California Consumer Privacy Act (CCPA) [11] impose strict limitations on data sharing. Finally, some DOTs hesitate to share due to competitive interests or administrative obstacles [12].

Researchers facing the challenges of limited infrastructure data and data-sharing obstacles have found a promising solution in federated learning (FL). This innovative ML technique, pioneered by Google in 2016 [13], was designed specifically to address data privacy and scarcity concerns. This was initially built to enhance text prediction across Android devices while complying with privacy regulations like GDPR and CCPA. The key to FL lies in its collaborative yet decentralized structure. Instead of directly exchanging sensitive raw data, a central server coordinates a network of local entities (like mobile devices or organizational servers). Each entity trains a model provided by the central server and sends back only the updated model parameters. The server then aggregates these updates to improve the overall model. This enables collective learning without compromising the privacy of individual data sources [14].

While previous studies have shown the promise of FL in various fields, its potential for infrastructure management with limited data remained underexplored [15]. UDOT faces the specific challenge of limited historical data, with only 272 rows in its culvert inventory. To enhance its culvert management system, UDOT requires predictive models. This research investigates the effectiveness of FL for predictive analysis in this context, specifically for UDOT's culvert inspection planning. The proposed FL method allows UDOT to benefit from data held by other state DOTs while maintaining strict data privacy and security. This research aims to showcase FL's potential for culvert condition prediction models, providing valuable insights for UDOT and other DOTs facing data scarcity and data privacy concerns. The developed model is shown in Figure 2.

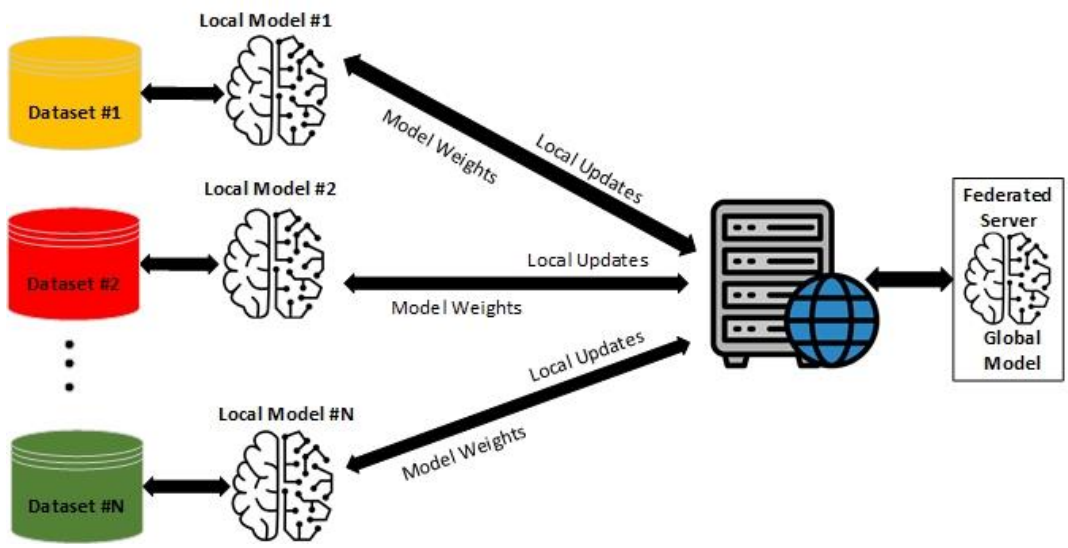

Figure 2. The proposed FL framework for predicting culvert conditions in Utah

Research Objectives

The objective of this research project is threefold:

1) Assess the feasibility of FL for culvert condition prediction with limited data.
2) Evaluate the performance of FL models compared to traditional centralized ML models.
3) Quantify the privacy benefits of FL for UDOT.

Research Methods

The project's backbone is to predict the condition of culverts with FL model to help UDOT prioritize culvert inspection. To estimate the deterioration of culverts, the authors will collect the spatial and temporal characteristics of culverts. To do so, based on the geographical location and culvert material type (e.g., corrugated steel, reinforced concrete, and plastic pipes), several factors such as soil physical and chemical characteristics (e.g., soil moisture, resistivity, pH, texture, and drainage class) and environmental conditions (e.g., flooding frequency), and culvert physical characteristics (e.g., length and span) will be considered to obtain from various DOTs' inventories, Web Soil Survey (WSS) [16], and Federal Emergency Management Agency's (FEMA) [17].

After collecting culvert data from multiple inventories, we will carefully preprocess it. During data processing, raw data is transformed into a usable format. Essential features and target labels are identified and extracted from raw data such as text or images. Various techniques, such as removing outliers, sampling the data, normalizing data for consistent scales, and thoughtfully combining features, will be employed for this procedure. To maintain the privacy-preserving nature of FL, we deliberately avoided fusing the datasets. However, for centralized learning we will use the fused datasets. The fused dataset is the global dataset made from the consolidation of preprocessed data from all sources.

In the next step, we will design the architecture of the FL model to predict the condition of culverts. FL offers a powerful approach to distributed learning where a global model is built by aggregating locally trained models across multiple entities. This process protects data privacy by avoiding the direct exchange of raw data. Instead, model parameters or gradients are shared to update the global model [18]. We will utilize the Flower framework to implement our FL approach. This open-source platform assists in the development and evaluation of FL models. Its versatility allows for the selection of various ML algorithms and optimization techniques [19]. We opted for an artificial neural network (ANN) as our core ML algorithm due to its strong predictive capabilities, suitability for multiclass classification, and seamless integration with Flower.

FL is divided into three separate categories: horizontal FL, vertical FL, and Federated Transfer Learning. Horizontal FL is applied when multiple entities have data with the same features but different data points, allowing them to collectively train a model while keeping their data private. Vertical FL is used when entities have the same data samples but different features, enabling collaborative model training by exchanging intermediate computations rather than raw data. Federated Transfer Learning is suitable for scenarios where entities possess both distinct features and data samples, facilitating knowledge sharing by adapting a model from one domain to another through fine-tuning [20]. This approach helps leverage collective insights without the need to share extensive local data. Our project employs horizontal FL, given the uniformity of features across different DOTs' inventories but variations in the data samples.

Additionally, we will apply two different model aggregation algorithms, federated averaging (FedAvg) and Federated Proximal (FedProx), which are the most prevalent aggregation techniques. In FL, model aggregation, or model fusion, combines models trained on decentralized entities to construct robust global models [21]. To ensure accurate prediction of the culvert condition rating for unseen data using the ML models developed in this study, we will utilize the hold-out cross-validation technique. However, some adjustments need to be made when using this method in the FL framework. Instead of a centralized evaluation, each participant entity independently assesses the global model based on its hold-out data. The evaluation metrics, such as accuracy or loss, will then be aggregated across agents to derive an overall performance measure for the global model.

In summary, in this project, we will develop a robust FL model to estimate the culverts' condition, and we will provide an FL framework for complying with privacy issues. Culverts can be tracked using the condition of culverts in other states. UDOT staff can repair/replace the culvert in advance and prevent high costs.

Relevance to Strategic Goals

Primary strategic goal: Economic Strength and Global Competitiveness

UDOT is tasked with compiling a list of capital improvement projects to secure the necessary funding, likely within the two or three years, to ensure investments are made at the most opportune time. In other words, recommendations for culvert inspection or maintenance actions (repair, rehabilitation, and replacement) need to be assessed and prioritized while adhering to budgetary allocations and minimizing risks and costs associated with failure. The proposed approach aims to optimize the yearly allocation of maintenance budgets by identifying culverts that require inspection, rehabilitation, or replacement. At the network level, the allocation of funds is determined based upon an initial budget. Furthermore, the optimum sequential path in the annual decision-making process may be determined using a combination of operations research tools.

The system offers a cost-efficient method for assessing culvert conditions and prioritizing maintenance tasks in situations with limited highway infrastructure data. The system also can help in providing an estimate of the minimum annual budget needed over a specified planning period to maintain or enhance the total value of the assets. Furthermore, it supports project-level prioritization regarding inspections, rehabilitations, replacements, or inaction when resources are scarce. Efficient transportation infrastructure keeps workers and goods moving, fostering economic activity and job creation. Furthermore, by prioritizing preventative maintenance, UDOT reduces the need for costly emergency repairs and minimizes environmental disruptions often associated with culvert failure.

Secondary strategic goal: Safety

In the United States, traffic accidents claim approximately 40,000 lives annually, a toll that surpasses the fatalities from many wars, diseases, and natural disasters. Notably, around one-third of these highway fatalities occur off the main road, with incidents involving hitting a culvert or a ditch accounting for over 10% of fatal run-off-road crashes. Culverts, which are essential for crossing streams and diverting water away from roads, are often in disrepair, posing significant risks. A failure in these structures can lead to traffic disruptions, environmental damage, property loss, and loss of life tragically.

To mitigate these risks, a new system is being developed to actively monitor culverts, specifically targeting those that are aged and deteriorating beneath roadways and in need of urgent repair. This proactive approach aims to significantly reduce the potential for accidents and fatalities caused by culvert failures on main roads, ensuring safer travel for all.

Educational Benefits

The PI of this project is currently teaching two relevant undergraduate and graduate-level classes called “CVEEN 6790: Advanced Computer-Aided Construction” and “CVEEN 5740: Horizontal Construction”. It is expected that the developed algorithms, methods, and case studies in this project will be directly converted into new course materials for these courses. In addition, a number of selected undergraduate and graduate students will be participating in different steps of this project, including data collection, processing, and validating the obtained results.

Outputs through Technology Transfer

The technology transfer process for this project will take place through three major channels: 1) publishing (presenting) research results in scholarly journals (peer-reviewed journal articles or conference papers); 2) direct interactions and with UDOT personnel through training sessions and workshops as the potential end-users for the results of this study, and 3) developing a prototype platform for potential commercialization and integrating it into Atom software as a final product offering.

Expected Outcomes and Impacts

The expected outcomes for this project will include the following items: an FL-based predictive model, which increases the performance of a centralized model in predicting the future condition of culverts while complying with privacy concerns. In addition, we will provide a data-driven prioritization approach for inspecting culverts with limited data availability. This framework will be implemented on a GIS platform or Atom software.

It is also necessary to mention that the outcomes of this project will be discussed and evaluated by UDOT personnel as the practitioners who will benefit from this project.

Work Plan

The project will include the following major tasks:

1) Literate review and initial evaluation of the existing predictive modelling methods; Expected completion date: end of 2nd month
2) Collecting culvert inventories from different DOTs to enrich the existing dataset; Expected completion date: end of 5nd month
3) Developing centralized learning models for condition prediction of culverts; Expected completion date: end of 6th month

4) Developing federated learning models for condition prediction of culverts; Expected completion date: end of 8th month
5) Analyzing a comparing the results; Expected completion date: end of 9th month
6) Preparing the final report; Expected completion date: end of 10th month

Project Cost

Total Project Costs:	$100,000
CTIPS Funds Requested:	$ 50,000
Matching Funds:	$ 50,000
Source of Matching Funds:	Utah Department of Transportation

References

[1] O. Salem, B. Salman, and M. Najafi, "Culvert asset management practices and deterioration modeling," *Transp. Res. Rec.*, no. 2285, pp. 1–7, 2012, doi: 10.3141/2285-01.

[2] P. Mohammadi, A. Rashidi, M. Malekzadeh, and S. Tiwari, "Evaluating various machine learning algorithms for automated inspection of culverts," *Eng. Anal. Bound. Elem.*, vol. 148, pp. 366–375, 2023, doi: 10.1016/j.enganabound.2023.01.007.

[3] K. R. Piratla, H. Jin, and S. Yazdekhasti, "A Failure Risk-Based Culvert Renewal Prioritization Framework," *Infrastructures* , vol. 4, no. 3. 2019. doi: 10.3390/infrastructures4030043.

[4] M. Stoner, W. Pang, and K. Piratla, "Predicting Culvert Deterioration Using Physical and Environmental Time-Independent Variables," *J. Pipeline Syst. Eng. Pract.*, vol. 10, no. 4, p. 04019035, 2019, doi: 10.1061/(asce)ps.1949-1204.0000409.

[5] J. N. Meegoda, T. M. Juliano, L. Potts, C. Tang, and T. Marhaba, "Implementation of a drainage information, analysis and management system," *J. Traffic Transp. Eng. (English Ed.*, vol. 4, no. 2, pp. 165–177, 2017, doi: 10.1016/j.jtte.2017.03.002.

[6] K. C. Sinha, S. Labi, and B. R. D. K. Agbelie, "Transportation infrastructure asset management in the new millennium: continuing issues, and emerging challenges and opportunities," *Transp. A Transp. Sci.*, vol. 13, no. 7, pp. 591–606, 2017.

[7] P. Mohammadi, B. Sherafat, and A. Rashidi. "A Risk-Based Framework for Optimizing Inspection Planning of Utah Culverts," Mountain-Plains Consortium, 2023.

[8] J. L. Beaver and T. J. McGrath, "Management of Utah highway culverts," *Transp. Res. Rec.*, no. 1904, pp. 113–123, 2005, doi: 10.3141/1904-12.

[9] M. Gillani and H. A. Niaz, "Machine learning based data collection protocol for intelligent transport systems: a real-time implementation on Dublin M50, Ireland," *Complex Intell. Syst.*, pp. 1–19, 2023.

[10] Regulation of the European Parliament and of the council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, "EU," Dec. 2018. https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en (accessed Apr. 22, 2023).

[11] California Legislative Information, *California Consumer Privacy Act*. 2018, pp. 1–59. Accessed: Apr. 22, 2023. [Online]. Available: https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5

[12] A. S. Figueiredo, “Data sharing: convert challenges into opportunities,” *Front. public Heal.*, vol. 5, p. 327, 2017.

[13] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” *arXiv Prepr. arXiv1610.05492*, 2016.

[14] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Artificial intelligence and statistics*, 2017, pp. 1273–1282.

[15] P. Mohammadi, A. Rashidi, and S. Asgari, “Improving Culvert Condition Prediction Models Using Federated Learning: The Case Study of Utah,” 2024.

[16] “Web Soil Survey.” https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx (accessed May 28, 2022).

[17] “Flood Zones | FEMA.gov.” https://www.fema.gov/glossary/flood-zones (accessed Oct. 23, 2022).

[18] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated Learning: Challenges, Methods, and Future Directions,” *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020, doi: 10.1109/MSP.2020.2975749.

[19] D. J. Beutel *et al.*, “Flower: A Friendly Federated Learning Research Framework,” 2020, [Online]. Available: http://arxiv.org/abs/2007.14390

[20] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, “A review of applications in federated learning,” *Comput. Ind. Eng.*, vol. 149, p. 106854, 2020.

[21] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” *Proc. Mach. Learn. Syst.*, vol. 2, pp. 429–450, 2020.