2025 ASHRAE Conference for Integrated Design, Construction & Operations

August 13-15, 2025 | Denver, CO

Machine Learning Based Metamodel for Faster Life Cycle Assessment of Large Portfolio of Buildings

Naveen Kumar Muthumanickam

National Renewable Energy Laboratory (NREL)

naveenkumar.muthumanickam@nrel.gov

This presentation was produced when the laboratory operated as the National Renewable Energy Laboratory (NREL). The laboratory is now the National Laboratory of the Rockies (NLR).

Learning Objectives

- Learn about how specific datasets were collected to train Machine Learning (ML) models for predicting specific LCA metrics

- Learn about different types of ML models, metamodel architecture, hyperparameters, and their performance in terms of accuracy during training and testing

- Learn about how to use the novel ML-based prediction tool (NREL RAPID-EC) developed by NREL researchers as part of this project.

Acknowledgments

- **Team:** Xin Wang, Sonja Adams, Jingying Hu, Julia Sullivan, Anna Neilsen, Sarah Abrahams, Jamie Cutlip-Gorman, Heather Goetsch, Michael Deru, Walter Tersch, David Leites, Beth Savage
- US General Services Administration (GSA)

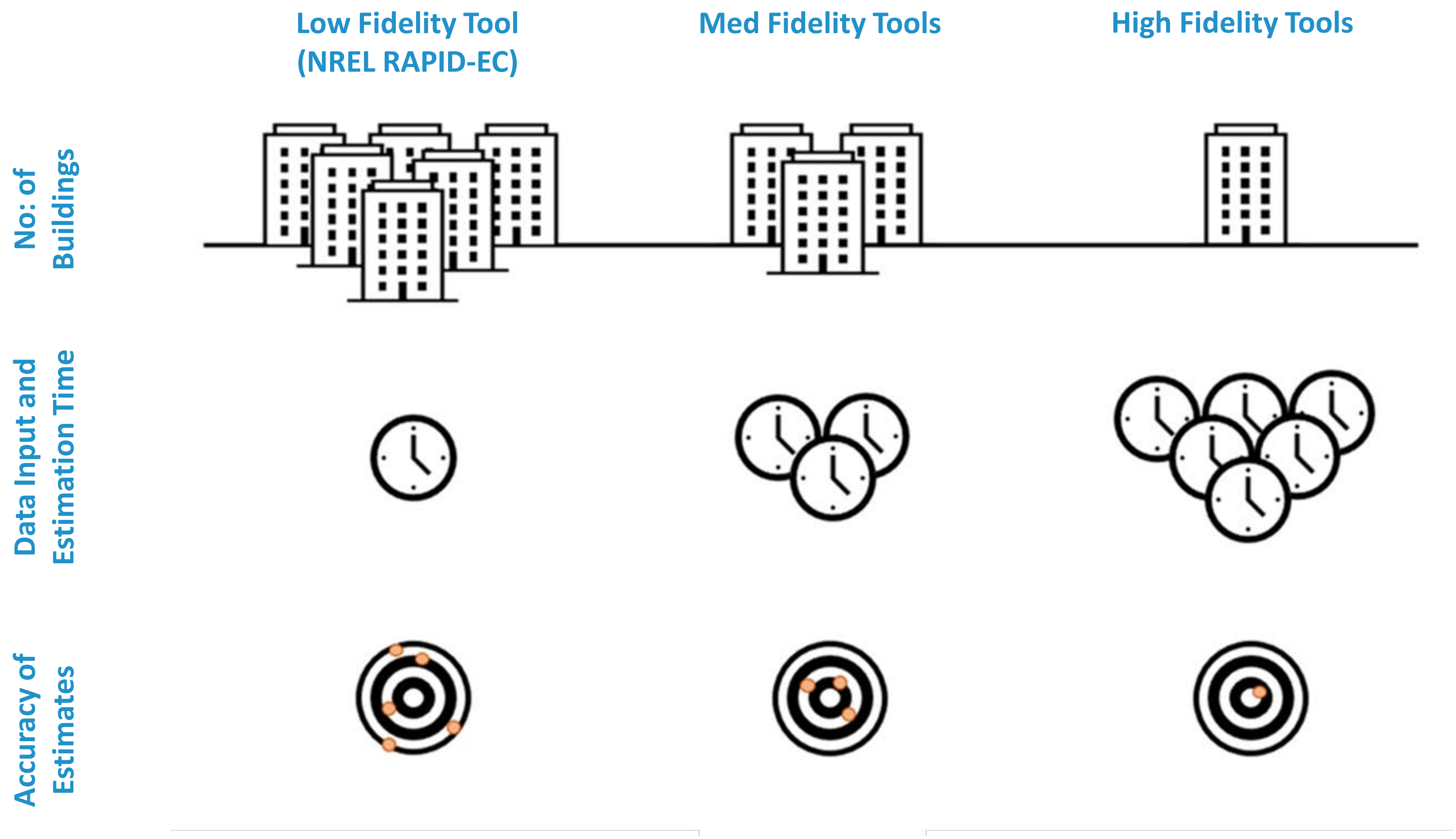
Outline/Agenda

- Background
- ML-based Rapid LCA of Buildings
- ML-based Rapid LCA of Buildings - Demo
- Conclusions

Background

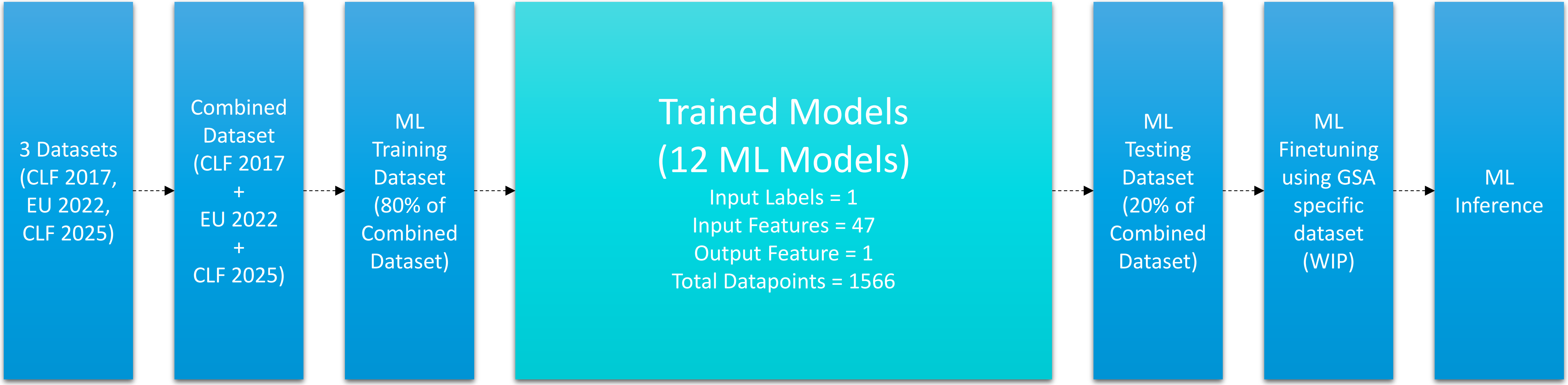
- Traditional life cycle assessment tools are labor- and compute-intensive.
- Machine Learning (ML)-based metamodels (surrogate models) offer faster alternatives but are limited by scarce life cycle data.
- Propose a ML-based metamodel that leverages few-shot learning and zero-shot learning techniques for rapid estimation of life cycle metric – embodied carbon.
- The presentation will cover the model's architecture, and potential for rapidly evaluating large portfolio of buildings.

Background



Comparison of NREL RAPID-EC with medium and high fidelity LCA tool capabilities

ML-based Rapid LCA of Buildings



Overall architecture of ML-based LCA tool development

ML-based Rapid LCA of Buildings

Data Collection → Data Preparation → ML Training → ML Testing → ML Finetuning → ML Inference

Dataset Features	CLF 2017	EU 2022	CLF 2025
No: of Input Columns	15	65	15
No: of Output Columns	2	120	2
No: of Data Rows (No: of Buildings)	1191	355	20

Raw training data

ML-based Rapid LCA of Buildings

Data Collection → **Data Preparation** → ML Training → ML Testing → ML Finetuning → ML Inference

Feature Engineering Techniques	Data conversion
Label Encoders	Converted text to integer
Column Data Type Casting	Strings into numbers
Scaling	Numeric values to a standard range (like 0 to 1)
Imputation	Missing values to mean or median
Feature Binning	Grouping continuous values into discrete values
Reverse Feature Binning	Converting ranges to discrete values

Techniques used to clean and combine raw training data

ML-based Rapid LCA of Buildings

Data Collection → **Data Preparation** → ML Training → ML Testing → ML Finetuning → ML Inference

Dataset Features	Combined and Cleaned Dataset (CLF 2017 + EU 2022 + CLF 2025)
No: of Input Labels	1
No: of Input Columns	47
No: of Output Columns	1
No: of Data Rows (No: of Buildings)	1566

Cleaned and combined master dataset for training after feature engineering

ML-based Rapid LCA of Buildings

Data Collection → Data Preparation → **ML Training** → **ML Testing** → ML Finetuning → ML Inference

12 ML models that were trained and tested using combined dataset:

1. Linear Regression
2. Ridge
3. Lasso
4. ElasticNet
5. DecisionTree
6. RandomForest
7. GradientBoosting
8. Adaptive Boosting (AdaBoost)
9. K-Nearest Neighbor (KNN)
10. Support Vector Regression (SVR)
11. eXtreme Gradient Boosting (XGBoost)
12. Light Gradient Boosting Machine (LightGBM)

ML models trained and tested using dataset

ML-based Rapid LCA of Buildings

Data Collection → Data Preparation → ML Training → ML Testing → ML Finetuning → ML Inference

ML modes ranked based on accuracy (high to low)

Validation metrics used to rank: Mean Absolute Error (MAE), Mean Square Root Error (MSE), Root Mean Square Error (RMSE), Combined Score

1. K-Nearest Neighbor (KNN)
2. DecisionTree
3. eXtreme Gradient Boosting (XGBoost)
4. GradientBoosting
5. Ridge
6. ElasticNet
7. RandomForest
8. Light Gradient Boosting Machine (LightGBM)
9. Lasso
10. Adaptive Boosting (AdaBoost)
11. Support Vector Regression (SVR)
12. Linear Regression

ML models ranked based on 4 validation metrics

ML-based Rapid LCA of Buildings

Data Collection → Data Preparation → ML Training → ML Testing → ML Finetuning → ML Inference

Major Input Features of Importance (across all 12 ML models)

1. Building Sub-typology
2. Building Location
3. Total Mass of the Building
4. Structural Material Type
5. Gross Floor Area
6. Heated Floor Area/Conditioned Floor Area
7. No: of Floors above ground
8. No: of features below ground

Important input features identified via sensitivity analysis

ML-based Rapid LCA of Buildings

Data Collection → Data Preparation → ML Training → ML Testing → **ML Finetuning** → ML Inference

Work in Progress with GSA specific Data

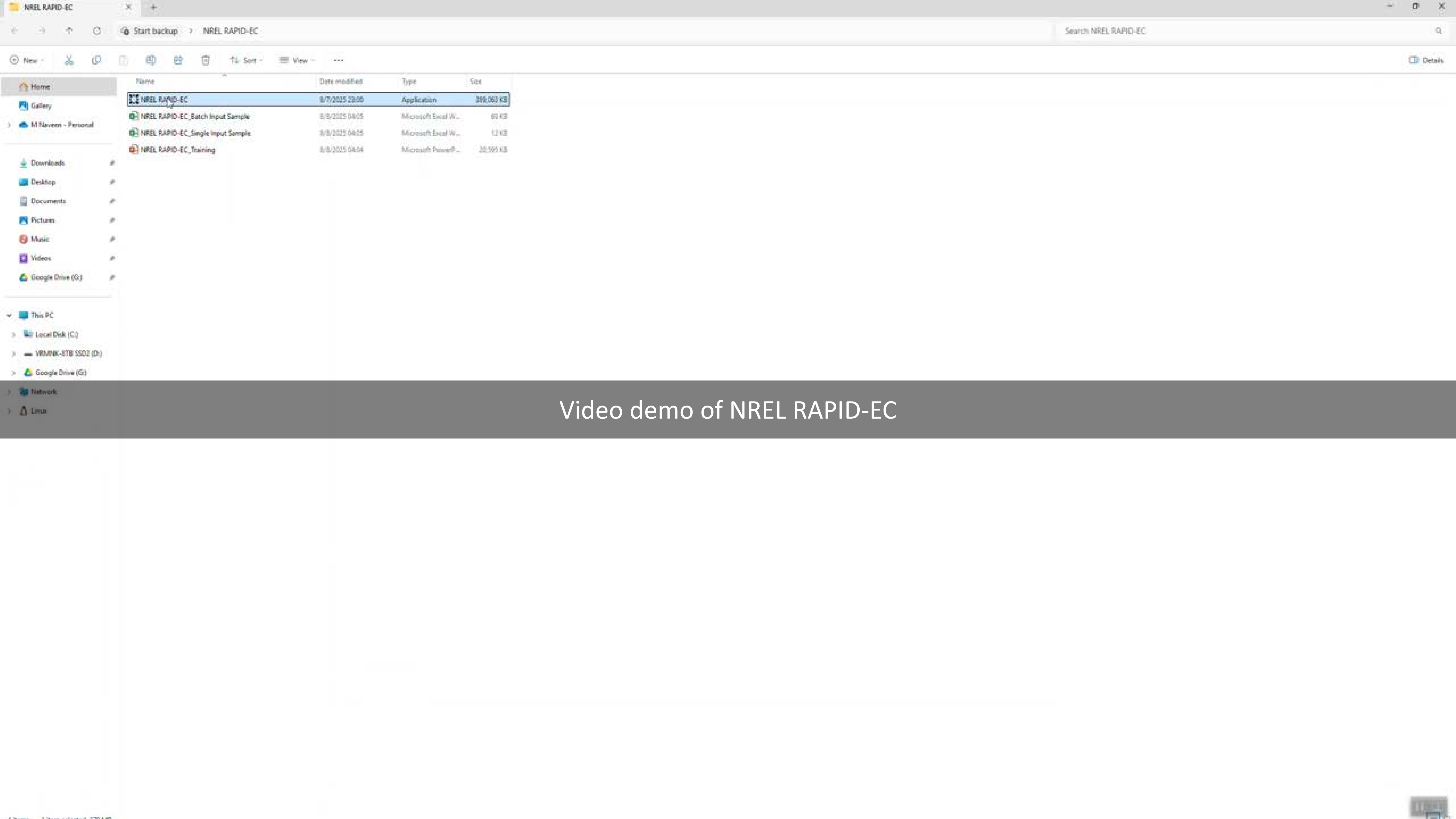
ML finetuning based on GSA specific data

ML-based Rapid LCA of Buildings - Demo

Data Collection → Data Preparation → ML Training → ML Testing → ML Finetuning → **ML Inference**

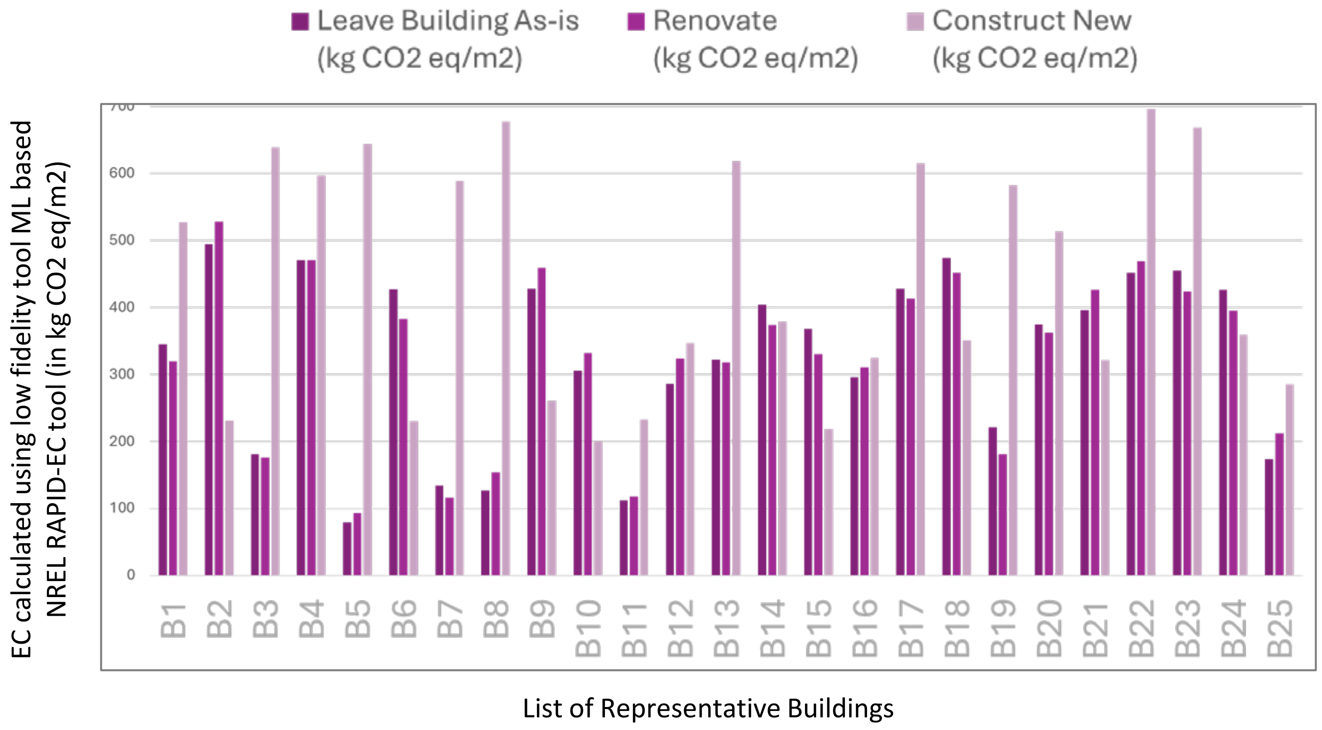
Full-screen video demo of NREL RAPID-EC in next slide

Deployment/Inference of actual ML-based LCA tool (NREL RAPID-EC)


NREL RAPID-EC
Start backup
NREL RAPID-EC
Search NREL RAPID-EC
New
Sort
View
Details
Home
Gallery
M Naveem - Personal
Downloads
Desktop
Documents
Pictures
Music
Videos
Google Drive (G:)
This PC
Local Disk (C:)
Google Drive (G:)
Network
Linux
Name
Date modified
Type
Size
NREL RAPID-EC
8/7/2025 23:00
Application
NREL RAPID-EC_Batch Input Sample
NREL RAPID-EC_Single Input Sample
NREL RAPID-EC_Training
Video demo of NREL RAPID-EC

ML-based Rapid LCA of Buildings - Usecase

Data Collection → Data Preparation → ML Training → ML Testing → ML Finetuning → **ML Inference**



Possible Portfolio level EC Visualization using the NREL RAPID-EC Tool

Conclusion

Pros

- Rapid prediction of embodied carbon
- KNN, XGBoost, ElasticNet predictions with reasonable accuracy

Cons

- Limited to prediction of embodied carbon among other life cycle metrics
- Input fields are still many in number

Future Recommendations

- Expand training and prediction to other LCA metrics
- Limit no: of inputs based on sensitivity analysis

Bibliography

Simonen, K., Rodriguez, B., Barrera, S., Huang, M., McDade, E., & Strain, L. (2017). Embodied Carbon Benchmark Study: LCA for Low Carbon Construction (Version 1.0). University of Washington.

Benke, B., Chafart, M., Shen, Y., Ashtiani, M., Carlisle, S., & Simonen, K. (2025). A Harmonized Dataset of High-Resolution Whole Building Life Cycle Assessment Results in North America (Version X) [Dataset]. Figshare.

Röck, M., & Sørensen, A. (2022). *Embodied Carbon of European Buildings Database (EU-ECB-DB)* [Data set and code]. Zenodo.

Questions

Naveen Kumar Muthumanickam
National Renewable Energy Laboratory (NREL)
naveenkumar.muthumanickam@nrel.gov

NLR/PR-5500-94241

This work was authored in part by NREL for the U.S. Department of Energy (DOE), operated under Contract No. DE-AC36-08GO28308. Funding provided by the General Services Administration. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.