

## AI Applications in Pavement Management

James Bryce  
Pat Parsons Faculty Fellow in Asphalt Technology  
Assistant Professor, West Virginia University  
November 18<sup>th</sup>, 2025



West Virginia University

## Outline

- Ongoing work
  - Performance modeling
- Upcoming tasks
  - Applications for WV PMS
- High impact potential for AI applications
  - Brainstorming with FHWA

West Virginia University

## What is AI?

- ...any technology or machine that can perform complex tasks typically associated with human intelligence, such as problem-solving, planning, reasoning, and decision-making.



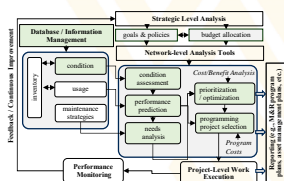
West Virginia University

## Categories of AI

- Machine learning
  - learn from data and improve system performance without being explicitly programmed
- Deep learning
  - Multi-layered NNs
    - Natural language processing
    - Computer vision

West Virginia University

## Pavement Management



West Virginia University

## Ongoing work...

### Performance Modeling and Unrecorded Maintenance

West Virginia University

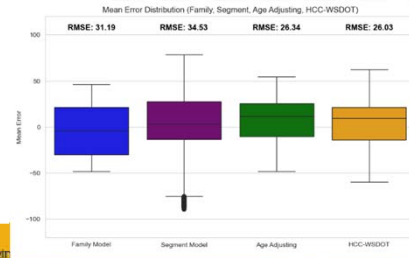
## FHWA Project Objective

Development of approaches and proof-testing of **adaptive project specific pavement performance models for PMS from “family” models** for developing prioritization projects.

**Leverage information from pavement segment and family to predict future segment condition**



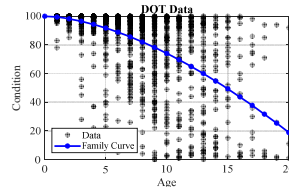
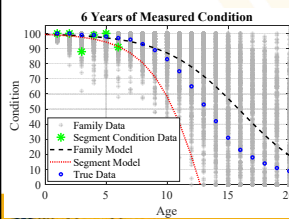
## Why not model segment data?



## Data used in project

### • Synthetic Data

### • State DOT Data



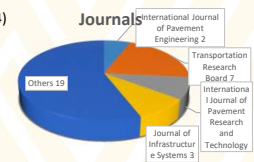
## Literature Review

- Articles were extracted between (2016-2024)

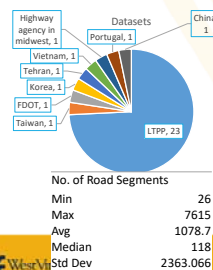
### • Keywords:

- 'Pavement performance prediction'
- 'Machine Learning'
- 'Neural networks'
- 'IRI'
- 'Performance index'

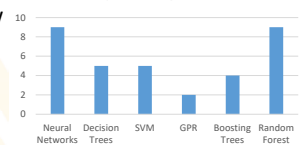
- 34 most relevant articles were studied in detail



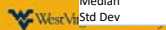
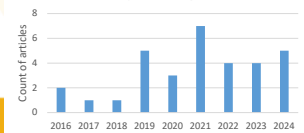
## Literature Review



### Top Used Algorithms



### Years of publishing article

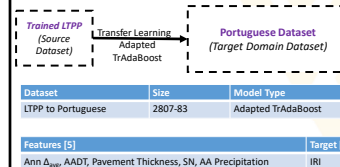


## Transfer learning for pavement performance prediction

Pedro Marcelino<sup>a\*</sup>, Maria de Lurdes Antunes<sup>a</sup>, Eduardo Fortunato<sup>a</sup>, Marta Castilho Gomes<sup>b</sup>

### Highlights

- Two stages: IRI from IRI0 and IRI from features.
- Transfer learning perform quite well in case of few data and at long-term forecasts.
- Boosting doesn't require same distribution and can perform well where available data is scarce.



## Synthetic Measurement Data

mm (actual PCI measurements)

|    | Y1  | Y2 | Y3 | Y4 | Y29 | Y30 |
|----|-----|----|----|----|-----|-----|
| R1 | 100 | 80 | 60 | 40 | 20  | 0   |
| R2 | 100 | 90 | 70 | 50 | 30  | 10  |
| R3 | 100 | 80 | 70 | 50 | 40  | 20  |
| R4 | 100 | 70 | 50 | 20 | 5   | 0   |
| R5 | 100 | 90 | 80 | 60 | 50  | 30  |

mm2 (PCI measurements+errors)

|    | Y1  | Y2 | Y3 | Y4 | Y29 | Y30 |
|----|-----|----|----|----|-----|-----|
| R1 | 100 | 82 | 42 | 58 | 18  | 3   |
| R2 | 98  | 89 | 68 | 54 | 29  | 12  |
| R3 | 99  | 83 | 58 | 69 | 41  | 27  |
| R4 | 99  | 76 | 52 | 28 | 7   | 1   |
| R5 | 99  | 94 | 82 | 63 | 48  | 36  |

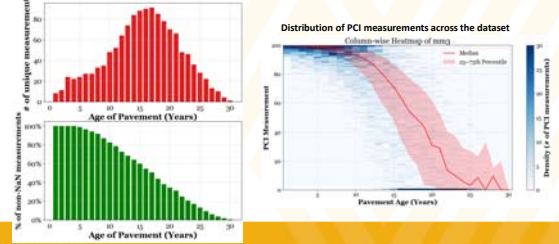
mm3 (PCI measurements + errors + NaNs)

|    | Y1  | Y2 | Y3 | Y4 | Y29 | Y30 |
|----|-----|----|----|----|-----|-----|
| R1 | 100 | 82 | 42 | 58 | 18  | 3   |
| R2 | 98  |    |    |    |     |     |
| R3 | 99  | 83 | 58 | 69 |     |     |
| R4 | 99  | 76 | 52 |    |     |     |
| R5 | 99  | 94 | 82 | 63 | 48  | 36  |

- 'mm' is generated measurements of 500 segments for 30 years.
- 'mm2' we added errors to it
- 'mm3' we hid some measurements.



## DATA ANALYSIS



## The challenge...

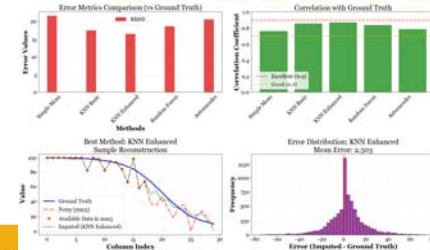
- ML methods are not traditionally designed to deal with missing data, so how to select features?

Imputation  
algorithm + ANN  
or LSTM

Train ANN on  
features of the  
data

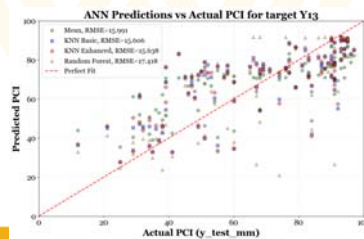


## Imputation (NOT PREDICTION) results



## Prediction Model FROM IMPUTED DATA

Features: Y6,Y7,Y8,Y9,Y10  
Target: Y13

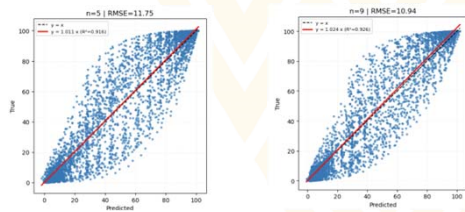


## Example Feature definitions (20+)

Dennoised segment statistics

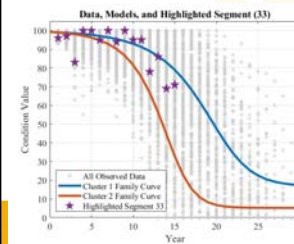
| Feature            | Definition / How it's calculated  | Role   |
|--------------------|---|--|
| seg_mean           | mean(y_den) where y_den is Huber-smoothed fit on (t_seg, y_meas[t_seg]) | Central level of observed history (noise-reduced). |
| slope_first        | Local slope between first two points: $\Delta y / \Delta t$ .           | Early-rate indicator.                              |
| slope_last         | Local slope between last two points: $\Delta y / \Delta t$ .            | Recent-rate indicator.                             |
| slope_mid_theilsen | Median of all pairwise slopes on (t_seg, y_den) (Theil-Sen).            | Robust overall trend.                              |
| last_age           | Latest observed age in t_seg.   | Defines window end                                 |

## Low-noise synthetic data



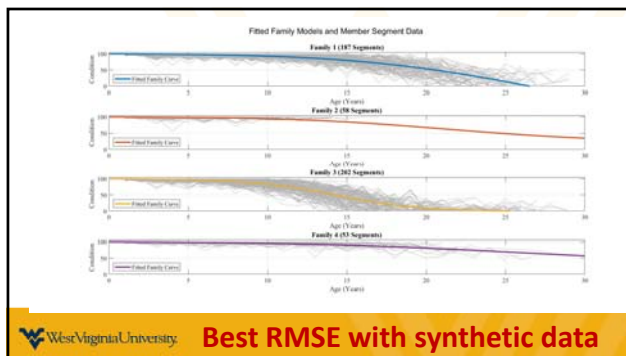
West Virginia University

## Clustering?

Article  
Deep Time-Series Clustering: A ReviewAli Alqahtani<sup>1,2,3,\*</sup>, Mohammed Ali<sup>2,3</sup>, Xianghua Xie<sup>3</sup> and Mark W. Jones<sup>3</sup>

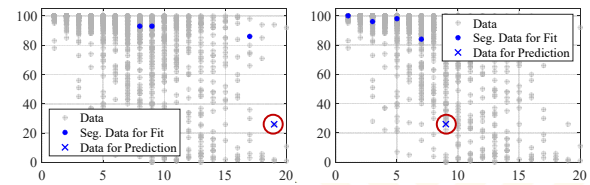
$$L_{ik} = \prod_{t \in S_i} \mathcal{L}(x_i(t) | \Theta_k).$$

$$u_{ik}^{\text{lik}} = \frac{L_{ik}}{\sum_{j=1}^K L_{ij}} \quad \text{with} \quad \sum_{k=1}^K u_{ik}^{\text{lik}} = 1.$$



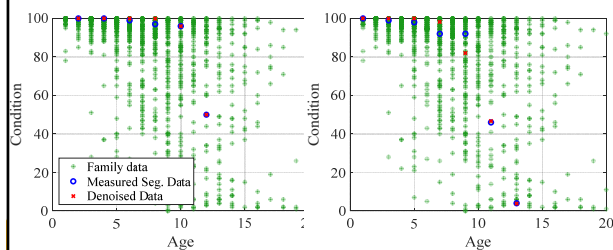
West Virginia University

Best RMSE with synthetic data

Challenge moving to DOT Data  
...How to Select Target???

West Virginia University

## Huber Loss to Denoise?



West Virginia University

## Metric to evaluate quality of fit?

## Three criteria:

- How well does predicted data fit with measured segment data from prior years in model  
 $\mathcal{M} = \log \mathcal{L}(x | \mathcal{F}(x, \text{age}))$
- How likely is the predicted data to be observed in the family?  $\mathcal{P}(y | \phi)$ , where  $y$  is predicted data and  $\phi$  is the distribution from the family model
- How likely is predicted data to be part of segment given the measured data?  $\mathcal{P}(y | y_{\text{meas}})$

West Virginia University

## Upcoming...

### Applications of AI in WV PMS



## Objectives and Tasks

- **Objective** – identify areas of pavement management that can be enhanced by AI and **build prototype tools** to demonstrate their benefits for a **subset** of those areas.
  1. Comprehensively map the pavement management process in WV and identify key areas that AI can be integrated.
  2. Gather data & models for those areas where AI will be investigated.
  3. Build prototype AI models and evaluate their performance compared to traditional pavement management processes
  4. Develop a final report and user guides
  5. Implementation Support.



## Holistic Asset Management and Accurate Cost Estimation

- Network-level cost estimates for paving projects are often inaccurate
  - neglect the condition of ancillary assets like guardrails, signage, or drainage, which are only assessed during costly site visits.

Computer Vision  
to Evaluate  
Existing Images

+

Deep Learning to  
Categorize  
Condition & Cost

*generate a comprehensive  
project scope and a more  
accurate, all-inclusive cost  
estimate*



## Promising applications...

### Brainstorming with FHWA & other colleagues



## Physics-Informed Language Models (PILM) for Advanced Mechanistic-Empirical (ME) Prediction

- **The Challenge:** Current ME models are constrained by pre-defined transfer functions that limit the complexity of the inputs they can consider from mechanistic models (e.g., only peak strain).
- **The Solution:** move beyond traditional PINNs to **Physics-Informed Language Models**. Instead of just replacing a transfer function with a neural network, use LLM's ability to process complex sequences and relationships. The model could take the *entire distribution of stresses and strains* from a mechanistic simulation, along with material properties and environmental data, as a complex input sequence.



## Are we losing information?

### PM

Condition Images  
Distresses  
Condition Index  
Performance Model  
Needs Assessment





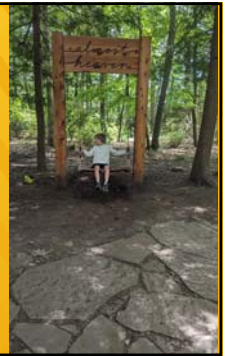
## Generative AI for Multi-Modal Pavement Analysis and Data Augmentation

- **The Challenge:** Pavement management involves a significant *loss of information*. Rich visual data from images is condensed into simplified distress codes (e.g., "medium severity alligator cracking"), which are then further abstracted into a single condition index. This process loses critical context about the *causes* of deterioration and the relationships between different distress types.
- **The LLM Solution: Multi-Modal Large Language Models.** These models can process and understand diverse data types—images, sensor readings (e.g., GPR), historical maintenance text logs, and structured data—simultaneously. Instead of reducing images to codes, an LLM can generate rich, descriptive text summaries of pavement condition.



## Conclusions...

Under development.



## Questions and Discussion?

**James Bryce, Ph.D.**  
Pat Parsons Faculty Fellow in Asphalt Technology  
Wadsworth Department of Civil and  
Environmental Engineering  
West Virginia University  
[James.Bryce@wvu.edu](mailto:James.Bryce@wvu.edu)

