Evidence-Based Machine Learning Algorithm Selection for Construction Data Analytics: A Systematic Review

Stuti Garg, Vivek Sharma, Dhaval Gajjar

CITC- 15 | November 10 - 14, 2025
Hosted by The International University of Rabat
Rabat, Morocco

CITC GLOBAL
Construction in the 21st Century

Overview

1. Research Problem
2. Research Objectives
3. Methodology
4. Key Findings
5. Problem Illustrated
6. Why This Matters (Implications)
7. Recommendations for Future Research

CITC GLOBAL
Construction in the 21st Century

Research Problem

The Challenge:

> **ML applications in construction lack structured decision-making frameworks for algorithm selection**

- Construction projects generate massive*, diverse datasets*
- Traditional analysis methods are insufficient
- No systematic approach to match algorithms with dataset characteristics
- Current practice: trial-and-error approach

CITC GLOBAL
Construction in the 21st Century

Research Objectives

What We Investigated?

Primary Question:

Is evidence-based guidance being used to select ML algorithms in construction research?

We Analyzed:

- ✓ Frequency of ML algorithm usage
- ✓ Model objectives (prediction vs. classification)
- ✓ Reasoning behind algorithm selection
- ✓ Correlations with dataset characteristics

CITC GLOBAL
Construction in the 21st Century

Methodology

Systematic Literature Review (PRISMA)

Data Sources:

- Web of Science
- IEEE Xplore
- ICONDA
- ScienceDirect
- ASCE

450	70	30	115
Initial Articles	After Screening	Final Articles	ML Methods

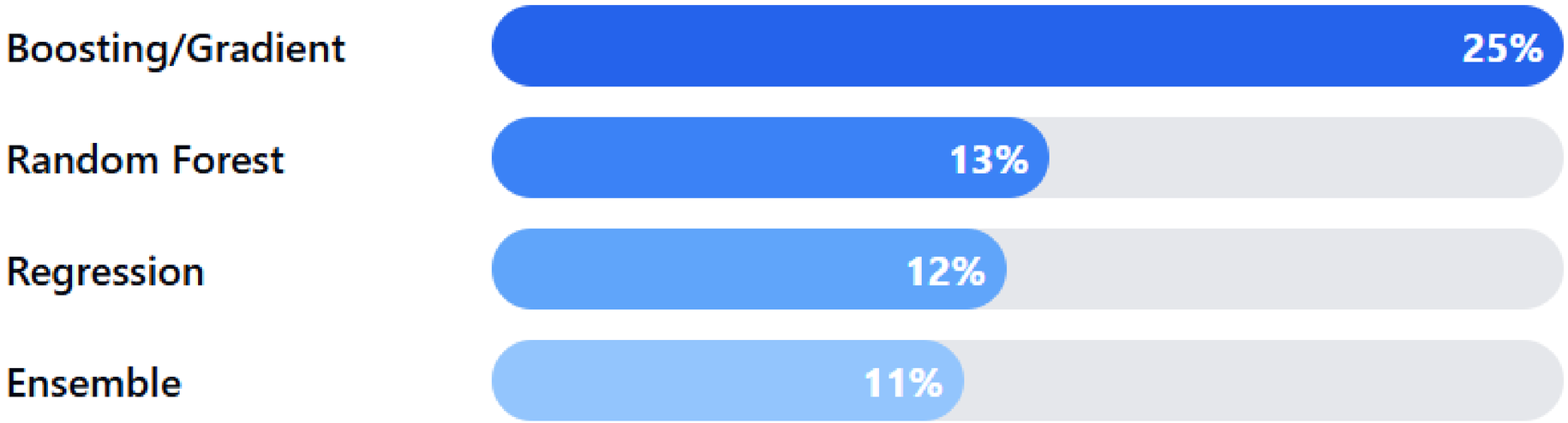
Search: "machine learning" AND "construction industry"

CITC GLOBAL
Construction in the 21st Century

KEY FINDINGS

1. Algorithm Frequency

Most Frequently Used ML Algorithms



Key Insight:

Advanced ensemble methods dominate, suggesting construction datasets require sophisticated algorithms to handle complexity.

CITC GLOBAL
Construction in the 21st Century

KEY FINDINGS

2. Algorithm by Analysis Objective

Prediction Models (55.7%)

1. Regression	**73%**
2. ANN	**62%**
3. Boosting/Gradient	**54%**

Classification Models (44.3%)

1. KNN	**67%**
2. Decision Tree	**58%**
3. SVM	**56%**

Note: Some algorithms like Random Forest and Ensemble show versatility across both objectives.

CITC GLOBAL
Construction in the 21st Century

KEY FINDINGS

3. Selection Reasoning

Why Researchers Choose Algorithms

Past use for similar data	23%
Better accuracy	17%
Past use for similar objective	16%
NO REASON provided	16%
Literature review	15%

No systematic methodology!

CRITICAL FINDING:
39% rely on precedent alone (past use) - not systematic analysis
16% provide NO justification whatsoever

CIC GLOBAL
Construction in the 21st Century

The Problem Illustrated

Current Selection Approach

❌ Current Practice

- Trial-and-error testing
- Following precedent
- "It worked before"
- No systematic rationale
- Time-consuming
- May miss optimal solutions

✓ Needed Approach

- Evidence-based selection
- Data-driven decisions
- Match algorithm to data characteristics
- Consider analysis objectives
- Efficient selection process
- Optimal performance

CITC GLOBAL
Construction in the 21st Century

The Problem Illustrated

3. Current Selection Approach

The Gap:
Researchers intuitively recognize algorithm-data relationships but lack a systematic framework to guide their decisions.

CITC GLOBAL
Construction in the 21st Century

Algorithm Characteristics Summary

Boosting/Gradient ⭐

+ High accuracy, handles non-linear data, reduces overfitting
- Complex, computationally intensive

Random Forest

+ Robust, versatile, reduces overfitting
- Scalability issues with large datasets

Regression

+ Simple, interpretable, fast
- Assumes linear relationships, limited with complex data

ANN

+ Handles complex patterns, high accuracy
- "Black box", prone to overfitting

KNN

+ Simple, captures local patterns
- Poor with high-dimensional or imbalanced data

Decision Tree

+ Easy to interpret, handles mixed data types
- Prone to overfitting, sensitive to noise

Full details available in paper Table 1 and Section 4.1

CITC GLOBAL
Construction in the 21st Century

Implication for Construction

Why This Matters?

For Researchers:

✓ Save time in algorithm selection
✓ Improve model performance
✓ Justify methodological choices
✓ Avoid trial-and-error approaches

For Industry Practitioners:

✓ Better prediction of costs, schedules, safety
✓ More reliable risk assessments
✓ Evidence-based decision making
✓ Efficient use of data analytics resources

For the Field:

✓ Standardization of ML practices
✓ Foundation for best practices
✓ Improved reproducibility

CITC GLOBAL
Construction in the 21st Century

Recommendation for Future Research

What Next?

1. Selection Matrix Development

Create evidence-based norms mapping algorithms to dataset characteristics and objectives

2. Quadrant Framework

Categorize algorithms as traditional/regular/advanced based on complexity and timeline

CITC GLOBAL
Construction in the 21st Century

Recommendation for Future Research

What Next?

3. Cross-Industry Analysis

Compare with retail, manufacturing, finance to identify algorithmic versatility

4. Empirical Validation

Test algorithm-dataset-objective combinations to validate optimal performance

CITC GLOBAL
Construction in the 21st Century

Conclusions

Key Takeaways

Main Findings:

✓ Analyzed 115 ML methods from 30 construction studies

✓ Boosting/gradient methods most common (25%)

✓ Clear patterns by objective (Regression for prediction, KNN for classification)

✓ 39% rely on precedent, 16% provide no justification

Confirmed Gap:

No systematic selection approach exists that maps algorithms to dataset characteristics and analysis objectives

CIC GLOBAL
Construction in the 21st Century

Thank you!

Questions?

CITC GLOBAL
Construction in the 21st Century

Acknowledgements

- Name - Affiliation
- Name - Affiliation
- ETC.

CITC GLOBAL
Construction in the 21st Century

Thank you

For any quesitions, please contact

.....

CITC GLOBAL
Construction in the 21st Century