Predicting the Transferability of Evidence Across Contexts: The Transferability Model

Presented by:

Miyaka Kinlabel

IT professional

ebase
effective basic services

Making Global Evidence Work Locally

- Most research is not designed for LMIC realities, specifically Sub Saharan Africa
- Context matters, but is often overlooked
- We built the **Transferability Model** to fix this
  → Predicts if an intervention will work in a new setting

Why this matters

- Saves time and resources
- Reduces research waste
- Considers the local perspective

What exactly is the transferability model?

- A scientifically robust tool that uses machine learning to decide whether to:
- adopt
- adapt
- contextualize and even
- reject evidence-based interventions.

How we make our decisions

% interventions transferable	Decision
80% or above	Evidence is transferable
50% - 70%	Transferable with modifications
Below 50%	New studies required

Understanding the predictors

1. **Relevance** of outcome	Does the outcome address a problem that exists and is urgent in the local context, is it policy-relevant, is it culturally fit, is it pertinent and feasible?	⇒ Yes or No (binary variable)
2. **Complexity** of intervention	How difficult/complicated will it be to implement the intervention in the local context? Consider all relevant stakeholders.	⇒ High/Medium/Low (categorical variable)

Understanding the predictors cont'd

3. **Cost** of intervention	How costly will it be to implement the intervention in the local context?	⇒ High/Medium/Low (categorical variable)
4. **Importance** of outcome	On a score of 0-10, how do stakeholders rate the importance of the outcome? Score differently for each type of stakeholder (policy makers, teachers, learners)	⇒ Continuous variable (mean of the different importance ratings)
5. **impact** of outcome	How effective was the intervention compared to the comparators? This is ascertained by looking at the effect size (e)	⇒ Low/Moderate/High - Low if $e < 0.2$ - Moderate if $0.2 \leq e \leq 0.45$ - High if $e > 0.45$

ebase effective basic services

evidence_review	int_relevance	int_complexity	int_cost	average_importance	effect_size	impact	transferabilit	transferablit
Arts Participation	Yes	Low	Low	7	0.48	High	Yes	1
Arts Participation	Yes	Low	Low	7	0.43	Moderate	Yes	1
Arts Participation	Yes	Low	Low	8	0.2	Moderate	Yes	1
Arts Participation	Yes	Medium	Low	8	0.45	High	Yes	1
Arts Participation	No	High	Low	4	0.09	Low	No	0
Arts Participation	No	High	High	5	0.32	Moderate	No	0
Arts Participation	No	Low	Low	4	0.1	Low	No	0
Aspiration Intervention	No	High	Low	3	0	Low	No	0
Block Scheduling	No	High	Low	3	0.1	Low	No	0
Block Scheduling	No	High	Low	4	-0.13	Low	No	0
Outdoor Adventure Learning	Yes	Low	Low	6	0.31	Moderate	Yes	1
Outdoor Adventure Learning	No	Medium	Low	3.6	0.43	Moderate	Yes	1
Outdoor Adventure Learning	Yes	Medium	Low	6	0.34	Moderate	Yes	1
Outdoor Adventure Learning	Yes	Medium	Low	6	0.17	Low	Yes	1
Parental Involvement	Yes	Medium	Medium	7.67	0.59	High	Yes	1
Parental Involvement	Yes	Medium	Medium	7.67	0.27	Moderate	Yes	1
Parental Involvement	Yes	Medium	Medium	7.67	0.25	Moderate	Yes	1
Parental Involvement	Yes	High	High	8	0.29	Moderate	Yes	1
Parental Involvement	Yes	High	High	8	0.43	Moderate	Yes	1
Parental Involvement	Yes	Medium	Medium	7.33	0.18	Low	Yes	1
Parental Involvement	Yes	Low	Low	8.3	0.65	High	Yes	1
Peer Tutoring	Yes	Medium	Medium	7.66	0.75	High	Yes	1
Peer Tutoring	Yes	Low	Medium	7	0.365	Moderate	Yes	1
Peer Tutoring	Yes	Medium	Medium	8	0.62	High	Yes	1
Peer Tutoring	Yes	High	High	7	0.35	Moderate	Yes	1
Peer Tutoring	Yes	Medium	Medium	6.67	1.05	High	Yes	1
Peer Tutoring	Yes	Medium	High	7.3	0.39	Moderate	Yes	1
Peer Tutoring	Yes	Low	Low	6.3	0.59	High	Yes	1
Peer Tutoring	Yes	Low	Low	6	0.43	Moderate	No	0
Performance Pay	No	High	High	8	-0.09	Low	No	0
Performance Pay	Yes	High	High	8	0.25	Moderate	No	0

How the model works

- Uses **Classification and Regression Trees (CART)**, a decision tree algorithm, specifically **Classification Trees** because the target variable (transferability) is categorical (Yes/No).

How It Works

Gini Index Calculation

- Measures how "pure" a node is
- The algorithm selects the variable that best splits the data to improve prediction

Table 1. Factors influencing transferability

Attribute Feature	Gini Impurity Index
Relevance of Intervention	0.332
Cost of Intervention	0.41*
Average Importance	0.344*
Impact of Intervention	0.52*

How the model works cont'd

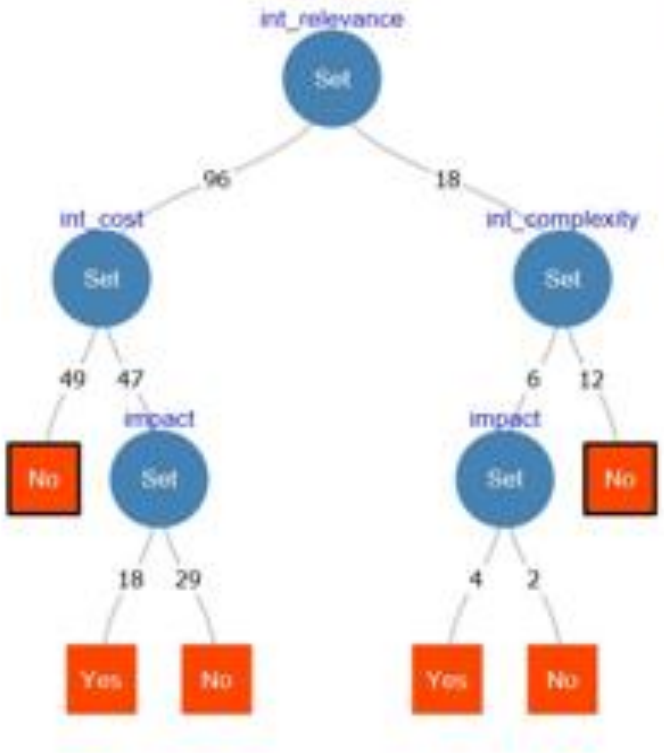
Tree Generation

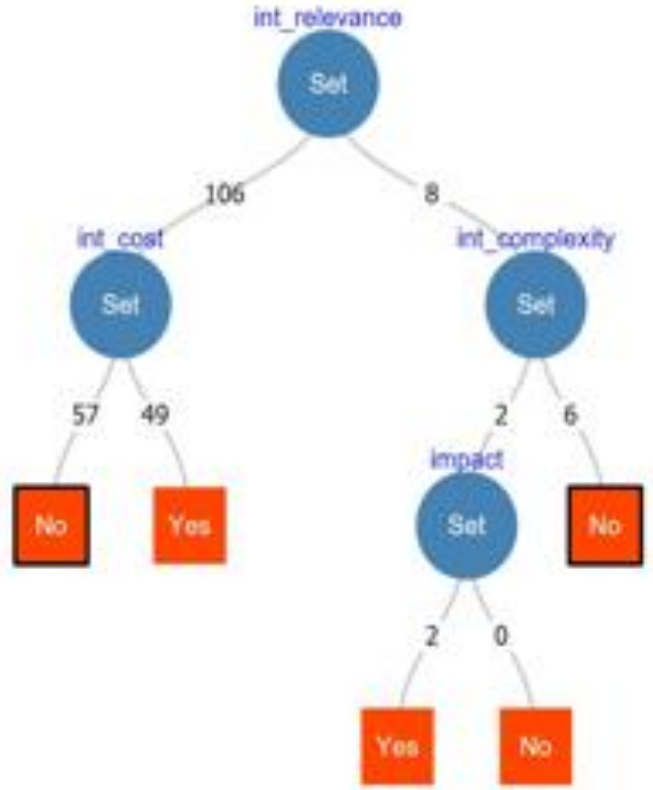
- The dataset is split based on variable thresholds (e.g. cost = high or low)
- Splits continue until the best classification is achieved

Pruning

- Shapes overfitting branches that don't improve accuracy
- Keeps the model simple, fast, and generalizable

Decision making with CART algorithm

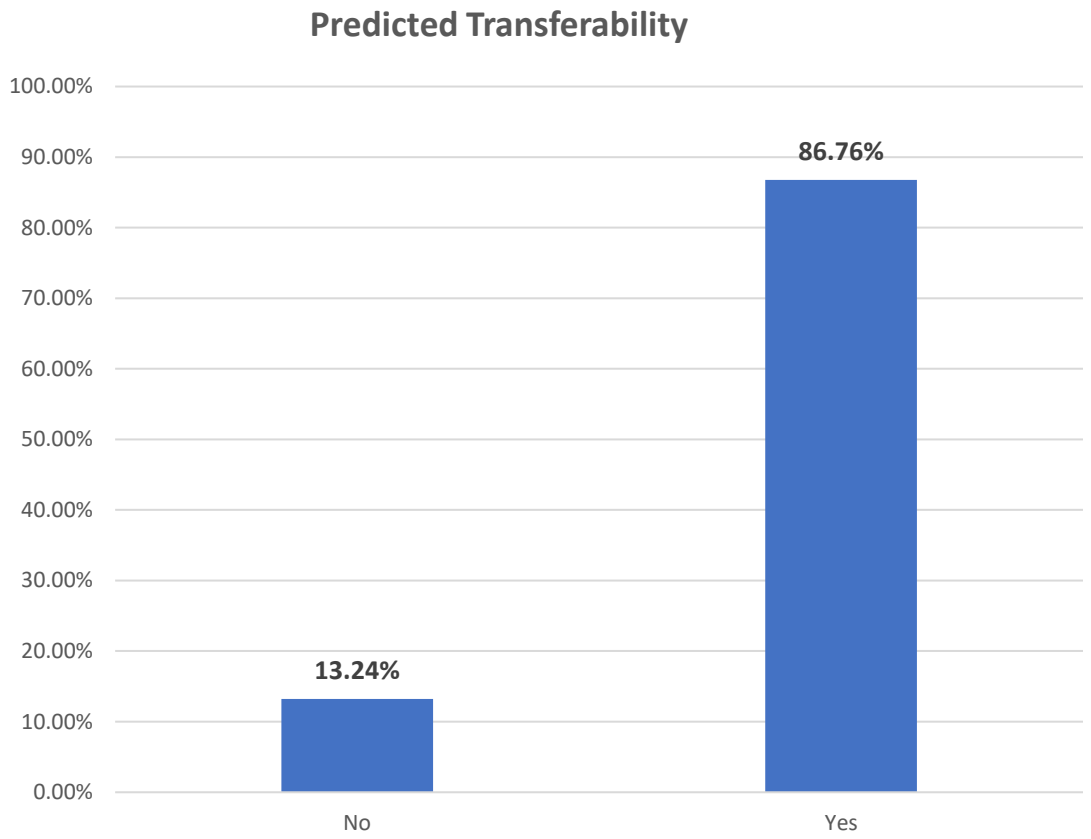




Applications of the transferability model

EEF Toolkit Pilot

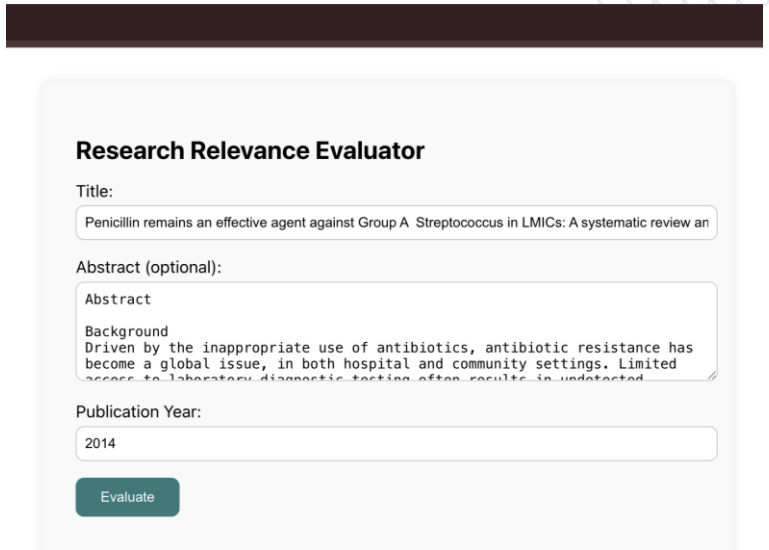
- Analysed 35 education strands
- Across 228 interventions from the Education Endowment Foundation (EEF)
- Goal: Predict which interventions are transferable to Sub-Saharan Africa (SSA)

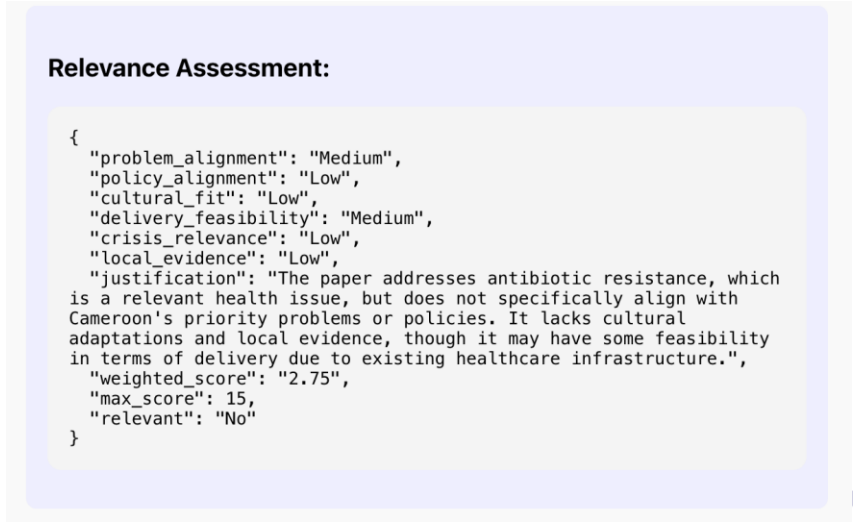


Applications of the transferability model cont'd

Transferability model within DEST tools

- Applied directly to research papers
- Uses large language models (e.g., GPT-4) to analyze full texts
- Automatically predicts whether a paper's recommendations are transferable to your context





Applications of the transferability model cont'd

Relevance Indicators

- Problem Alignment
- Policy Alignment
- Delivery Feasibility
- Cultural Fit
- Crisis relevance
- Local Evidence
- Timeliness

```
function getTransferabilityVerdictWeighted(json) {
  const weights = {
    problem_alignment: 3,
    policy_alignment: 3,
    delivery_feasibility: 2.5,
    cultural_fit: 2,
    crisis_relevance: 2,
    local_evidence: 1,
    timeliness:1.5
  };

  const scoreMap = { High: 1, Medium: 0.7, Low: 0 };

  let total = 0;
  let max = 0;

  for (const key in weights) {
    const relevance = json[key] || "Low";
    const weight = weights[key];
    total += scoreMap[relevance] * weight;
    max += weight;
  }

  let verdict = "No";
  if (total >= max * 0.75) {
    verdict = "Yes";
  } else if (total >= max * 0.55) {
    verdict = "Moderate";
  }


  return {
    ...json,
    weighted_score: total.toFixed(2),
    max_score: max,
    relevant: verdict,
  };
}
```

Relevance Checker Performance & Validation

Testing Summary(pilot):

- Target sample size: 100 abstracts
- Reviewed so far: 15 (25% of sample)
- Early tool accuracy: 80% agreement with human raters
- Strongest performance in: problem alignment, feasibility
- Weakest performance in: cultural fit, policy citation detection

Limitations

1. Multi-Country Pipelines Add Complexity
2. Abstract-Level LLM Inference Is Costly & Slow
3. Local Evidence Is Hard to Extract

Thank you for your kind Attention