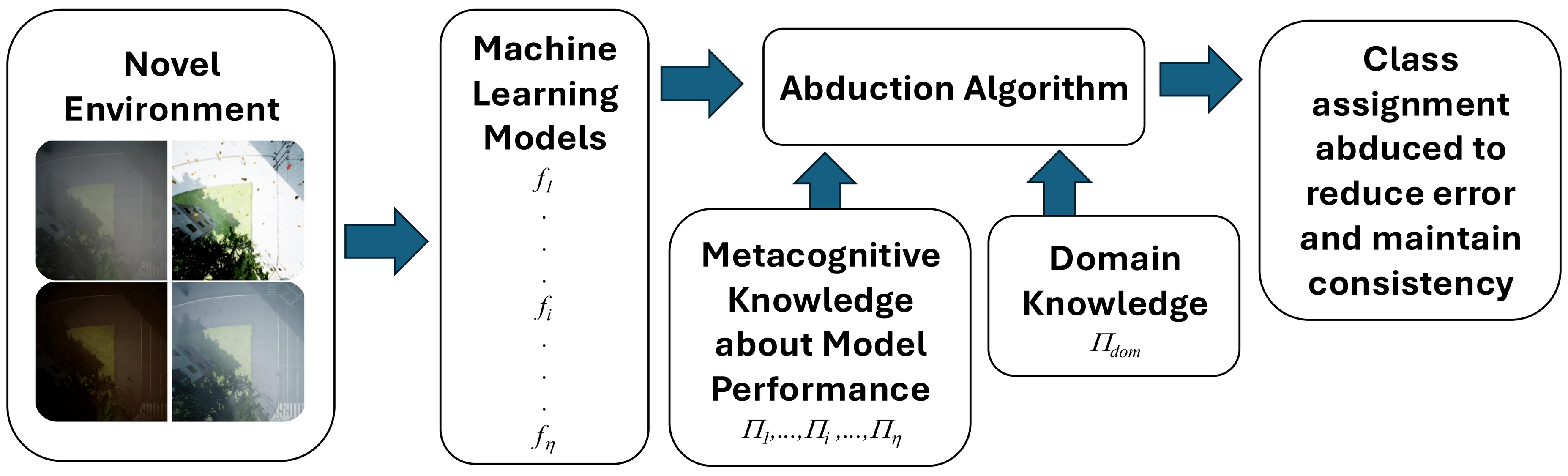
Consistency-based Abductive Reasoning over Perceptual Errors of Multiple Pre-trained Models in Novel Environments

Mario Leiva[1], Noel Ngu[2], Joshua Shay Kricheli[3], Aditya Taparia[2], Ransalu Senanayake[2], Paulo Shakarian[3], Nathaniel D. Bastian[4], John Corcoran[5] and Gerardo I. Simari[1]

[1]DCIC, Universidad Nacional del Sur (UNS) & ICIC (UNS-CONICET), Bahía Blanca, Argentina
[2]Arizona State University, Tempe, AZ USA
[3]Syracuse University, Syracuse, NY USA
[4]United States Military Academy, West Point, NY USA
[5]Systems Planning & Analysis, Alexandria, VA USA

Abstract

We address performance degradation in novel environments by integrating multiple pre-trained models via **consistency-based abduction** at test-time. Our approach encodes predictions and error detection rules into a logic program, utilizing **Integer Programming** and **Heuristic Search** to maximize prediction coverage while maintaining domain consistency. Extensive experiments on aerial datasets demonstrate a **13.6% F1-score** improvement and **16.6% accuracy gain** compared to standard baselines.



Abduction Problem

Hypothesis (H): A set of atoms $accept(i, c)$ indicating we trust model f_i for class c.

Assignment Rule: We assign class c to object ω only if predicted by an accepted model with no detected errors:

$$assign(c, \omega) \leftarrow \neg error(i, c, \omega) \wedge (f_i(\omega) = c) \wedge accept(i, c)$$

Optimization: We find H to maximize valid assignments ($Pred(H)$) while keeping inconsistencies ($Inc(H)$) below a threshold δ:

$$\max_{H \in \mathcal{H}} Pred(H)$$

subject to:

$$Inc(H) \leq \delta, \quad \delta \in [0, 1]$$

and

$$(H \cup O \cup \Pi) \setminus \Pi_{dom} \quad \text{is consistent}$$

Algorithms

Exact Approach
Integer Program (IP)

$$\max \sum_{\omega \in \Omega} \sum_{c \in \mathcal{C}} A_{c,\omega},$$

subject to:

$$\sum_{\omega \in \Omega} \sum_{(c,c') \in IC} Con_{\omega,(c,c')} \leq \delta,$$

$$X_{\omega,f,c} \leq 1 - Elim_{f,c}$$

$$X_{\omega,f,c} \cdot pred_{f,c,\omega} \leq A_{c,\omega}$$

Next, for each c, ω we have:

$$A_{c,\omega} \leq \sum_f X_{\omega,f,c} \cdot pred_{f,c,\omega}$$

For each $\omega \in \Omega, (c, c') \in IC$:

$$A_{c,\omega} + A_{c',\omega} - 1 \leq Con_{\omega,(c,c')}$$

for each ω, we have:

$$\sum_{c \in \mathcal{C}} A_{c,\omega} \geq 1$$

Finally:

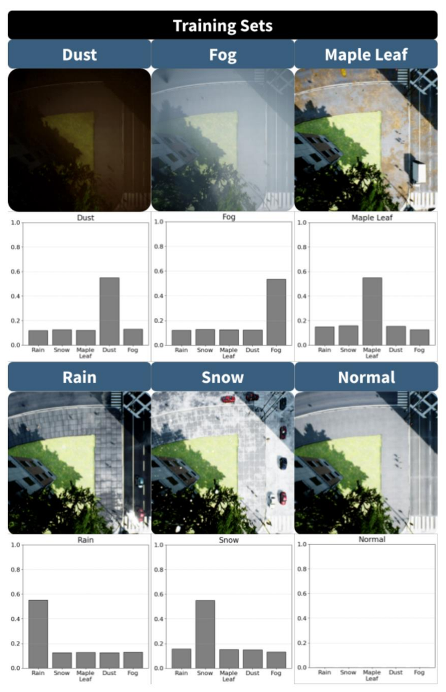
$$\sum_{\omega \in \Omega} \sum_{(c,c') \in IC} Con_{\omega,(c,c')} \leq \delta$$

Approximate Approach
Heuristic Search (HS)

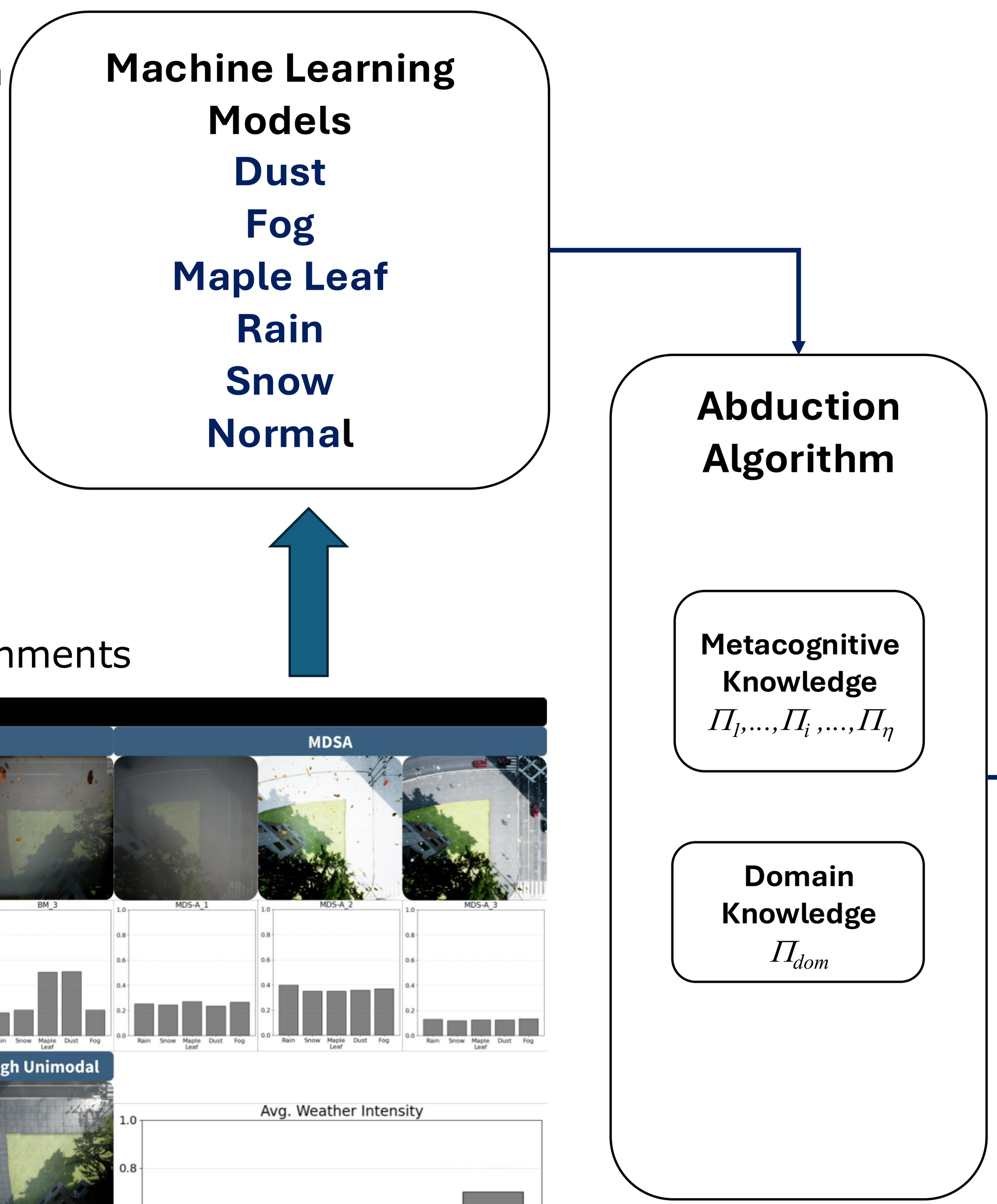
Algorithm 1: Heuristic Search (HS) for Prediction Optimization

Input:
  P_{raw} (Set of all raw prediction tuples (o, l, f, c))
  δ (Maximum allowed inconsistency for S_{final})
  E_{set} (Set of EDR ϵ thresholds to evaluate)
  {Implicit: Sets $\mathcal{F}$ (models), $\mathcal{C}$ (classes); Functions $GetFilteredPreds(f, c, \epsilon, P_{raw})$ and $CalcIncon(S)$.}
Output: S_{final} (Optimized set of prediction tuples (o, l))
$S_{final} \leftarrow \emptyset$
for each model $f \in \mathcal{F}$ and class $c \in \mathcal{C}$ **do**
  $P_{best\_add} \leftarrow \emptyset$ {Best predictions from current (f, c) to add}
  $n_{current\_max} \leftarrow |S_{final}|$ {Max size of $S_{final} \cup P_{new}$}
  for each $\epsilon \in E_{set}$ **do**
    $P_{new} \leftarrow GetFilteredPreds(f, c, \epsilon, P_{raw})$
    $S_{cand} \leftarrow S_{final} \cup P_{new}$
    if $CalcIncon(S_{cand}) \leq \delta$ **and** $|S_{cand}| > n_{current\_max}$ **then**
      $P_{best\_add} \leftarrow P_{new}$
      $n_{current\_max} \leftarrow |S_{cand}|$
    end if
  end for
  if $P_{best\_add} \neq \emptyset$ **then**
    $S_{final} \leftarrow S_{final} \cup P_{best\_add}$
  end if
end for
return S_{final}

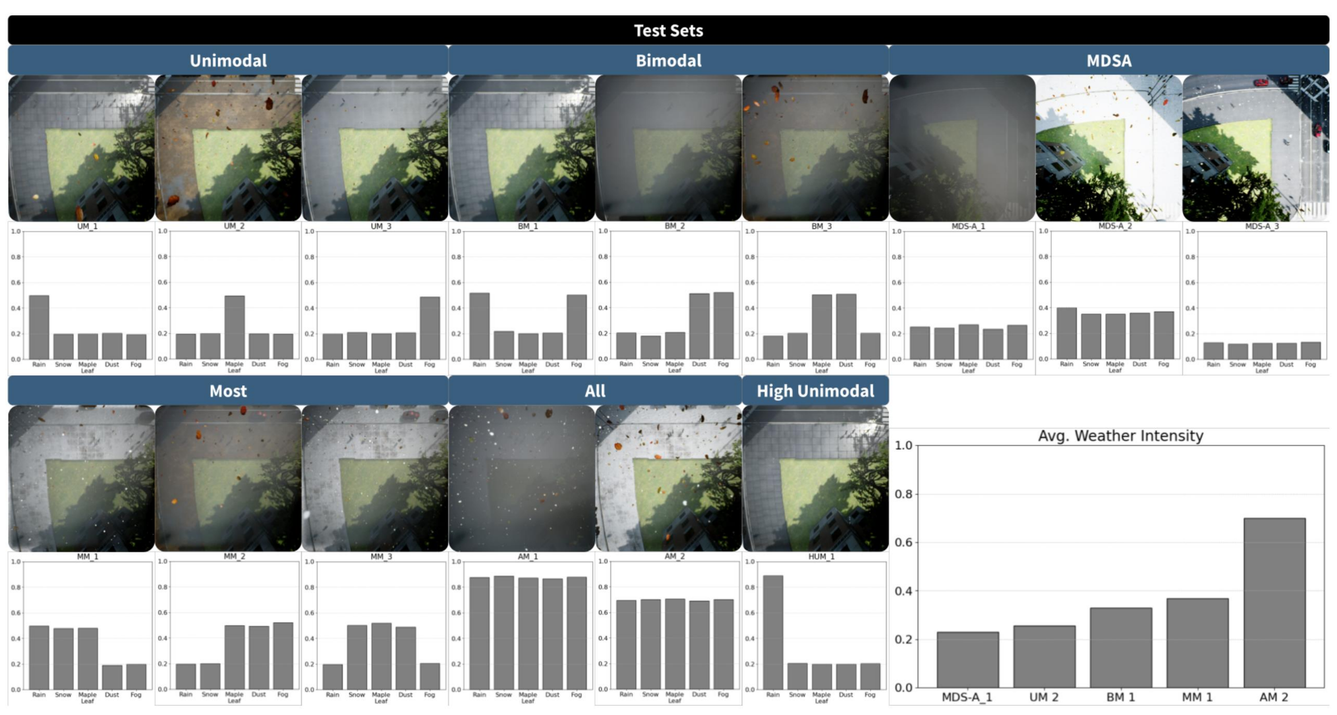
Experimental Setup and Results



Six training sets, each simulating a specific weather condition



15 Test environments



Test Set	Best		Avg.		MV		IP+TB		HS+TB	
	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc
MDS-A_1	0.57	0.40	0.52	0.36	0.28	0.34	**0.58**	**0.41**	**0.58**	**0.41**
MDS-A_2	0.33	0.20	0.29	0.17	0.26	**0.22**	**0.37**	**0.22**	0.32	0.19
MDS-A_3	0.54	0.37	0.49	0.33	0.39	0.29	**0.56**	**0.39**	0.55	0.38
UM_1	0.54	0.37	0.47	0.31	0.26	0.23	**0.64**	**0.47**	0.61	0.44
UM_2	0.56	0.38	0.46	0.31	0.25	0.22	**0.64**	**0.47**	0.61	0.44
UM_3	0.54	0.37	0.43	0.28	0.22	0.19	**0.63**	**0.46**	0.59	0.42
BM_1	0.42	0.27	0.33	0.20	0.19	0.16	**0.45**	**0.29**	0.39	0.24
BM_2	0.33	0.20	0.25	0.15	0.14	0.12	**0.37**	**0.23**	0.36	0.22
BM_3	0.37	0.23	0.31	0.19	0.18	0.16	**0.43**	**0.27**	0.40	0.25
MM_1	0.46	0.30	0.40	0.25	0.22	0.21	**0.51**	**0.34**	0.46	0.30
MM_2	0.32	0.19	0.24	0.14	0.13	0.10	**0.36**	**0.22**	0.29	0.17
MM_3	0.41	0.26	0.35	0.22	0.18	0.16	**0.46**	**0.30**	0.39	0.24
AM_1	0.18	0.10	0.12	0.07	0.05	0.04	**0.21**	**0.11**	0.18	0.10
AM_2	0.23	0.13	0.18	0.10	0.07	0.06	**0.28**	**0.16**	0.23	0.13
HUM_1	0.45	0.29	0.40	0.25	0.18	0.17	**0.57**	**0.40**	0.55	0.38

Test Set	IP (No TB)		HS (No TB)	
	F1 (% Diff)	Acc (% Diff)	F1 (% Diff)	Acc (% Diff)
MDS-A_1	0.58 (0.0)	0.41 (0.0)	0.52 (-10.3%)	0.35 (-14.6%)
MDS-A_2	0.37 (0.0)	0.22 (0.0)	0.27 (-15.6%)	0.16 (-16.7%)
MDS-A_3	0.56 (0.0)	0.39 (0.0)	0.49 (-10.9%)	0.32 (-15.8%)
UM_1	0.64 (0.0)	0.47 (0.0)	0.53 (-13.1%)	0.36 (-18.2%)
UM_2	0.64 (0.0)	0.47 (0.0)	0.52 (-14.1%)	0.35 (-18.8%)
UM_3	0.63 (0.0)	0.46 (0.0)	0.52 (-11.9%)	0.35 (-16.7%)
BM_1	0.45 (0.0)	0.29 (0.0)	0.34 (-11.1%)	0.20 (-16.7%)
BM_2	0.37 (0.0)	0.23 (0.0)	0.31 (-13.5%)	0.19 (-13.6%)
BM_3	0.43 (0.0)	0.27 (0.0)	0.34 (-15.0%)	0.20 (-20.0%)
MM_1	0.51 (0.0)	0.34 (0.0)	0.38 (-15.7%)	0.24 (-20.0%)
MM_2	0.36 (0.0)	0.22 (0.0)	0.25 (-13.8%)	0.14 (-17.6%)
MM_3	0.46 (0.0)	0.30 (0.0)	0.33 (-15.4%)	0.20 (-16.7%)
AM_1	0.21 (0.0)	0.11 (0.0)	0.15 (-16.7%)	0.08 (-20.0%)
AM_2	0.28 (0.0)	0.16 (0.0)	0.19 (-17.4%)	0.11 (-15.4%)
HUM_1	0.57 (0.0)	0.40 (0.0)	0.48 (-12.7%)	0.32 (-15.8%)

- **Best**: Best Individual Model
- **Avg**: Average of Models
- **MV**: Majority Vote
- **IP + TB**: Integer Programming + Tie Breaker
- **HS + TB**: Heuristic Search + Tie Breaker
- **IP (No TB)**: IP with Tie Breaker Ablation
- **HS (No TB)**: HS with Tie Breaker Ablation

Code — github.com/lab-v2/EDCR PyReason AirSim
Extended version — https://arxiv.org/abs/2505.19361

arXiv

AAAI-26 / IAAI-26 / EAAI-26
JANUARY 20–27, 2026 | SINGAPORE

Visit https://leibniz.syracuse.edu for more information.