HorNets: Learning from Discrete and Continuous Signals with Routing Neural Networks

Boshko Koloski[1,2] Nada Lavrač[1] Blaž Škrlj[1]

[1]Jožef Stefan Institute, Ljubljana, Slovenia [2]Jožef Stefan Postgraduate School, Ljubljana, Slovenia

Motivation

- **Complex Data Modeling:** Neural networks struggle with modelling logic in high dimensions.
- **High-Dimensional Challenges:** Limited instances in high-dimensional tabular data demand data-efficient models.
- **Neuro-Symbolic Integration:** Combining neural networks with propositional logic enhances reasoning capabilities.
- **Current Model Limitations:** Existing models often fail to handle both discrete and continuous data effectively.

PolyClip Activation Function

$$\text{polyClip}(x,k) = \text{clip}(x^{2\cdot k+1}, -1, 1) = \frac{x^{2k+1}}{\max\{|x^{2k+1}|, 1\}} = \begin{cases} -1; x \leq -1 \\ 1; x \geq 1 \\ x^{2k+1}; \text{else.} \end{cases}$$

- **Output Range:** Maps inputs to {-1, 0, 1}, enabling logical interpretations.
- **Logical Expressiveness:** Capable of modeling negation and conjunction (AND, NOT).
- **Probabilistic Link:** Connects to zero-order (propositional) logic.

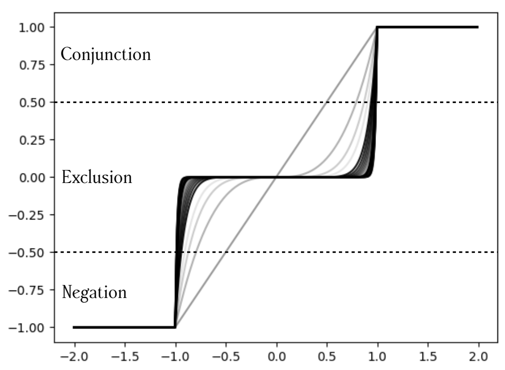
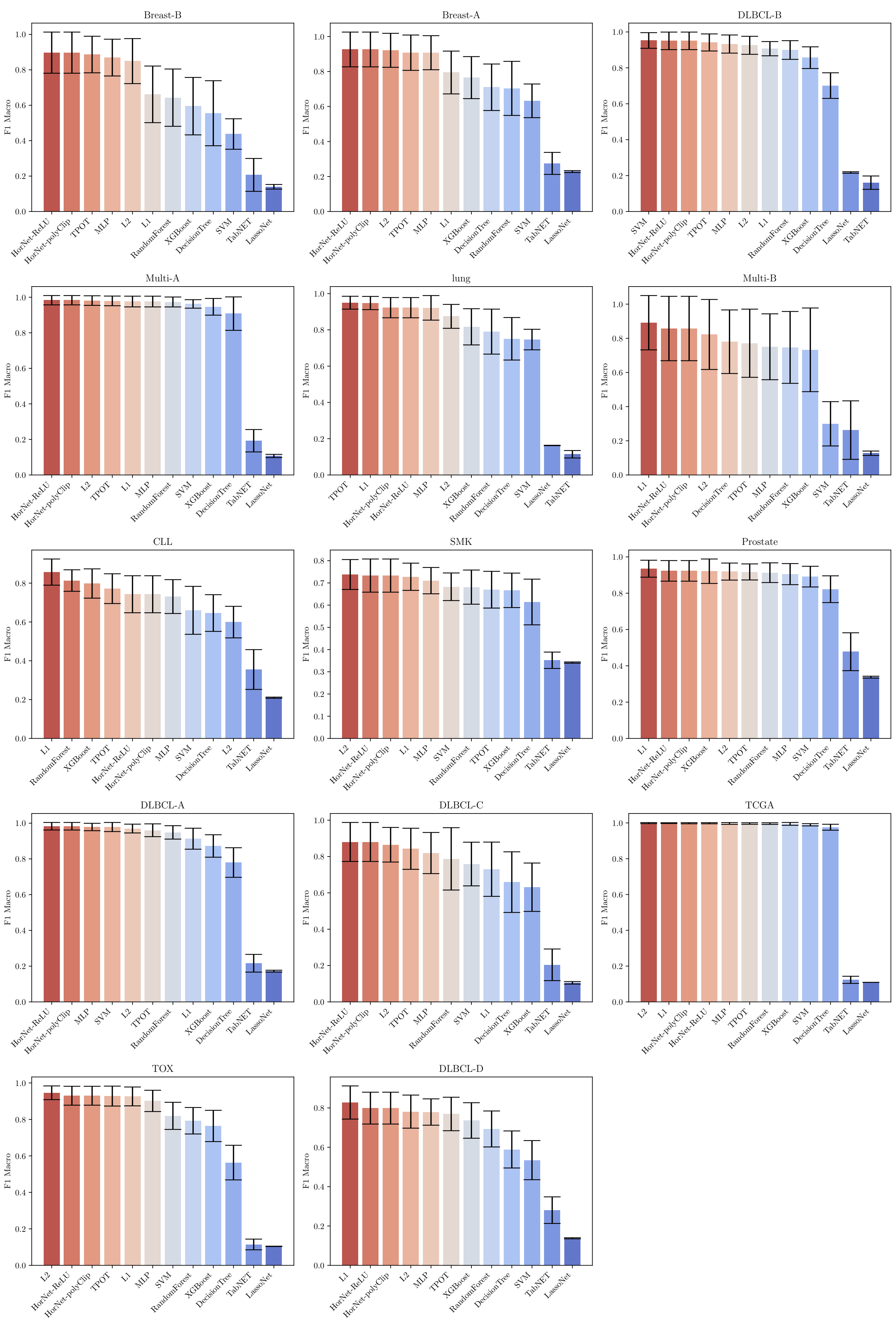


Figure 1. PolyClip Activation Function

Modelling High-Dimensional Real-Life Biomedical Data

- Consistently ranks among the top three performers across 14 datasets, outperforming strong baselines such as AutoML-based **TPOT**.



HorNets Architecture

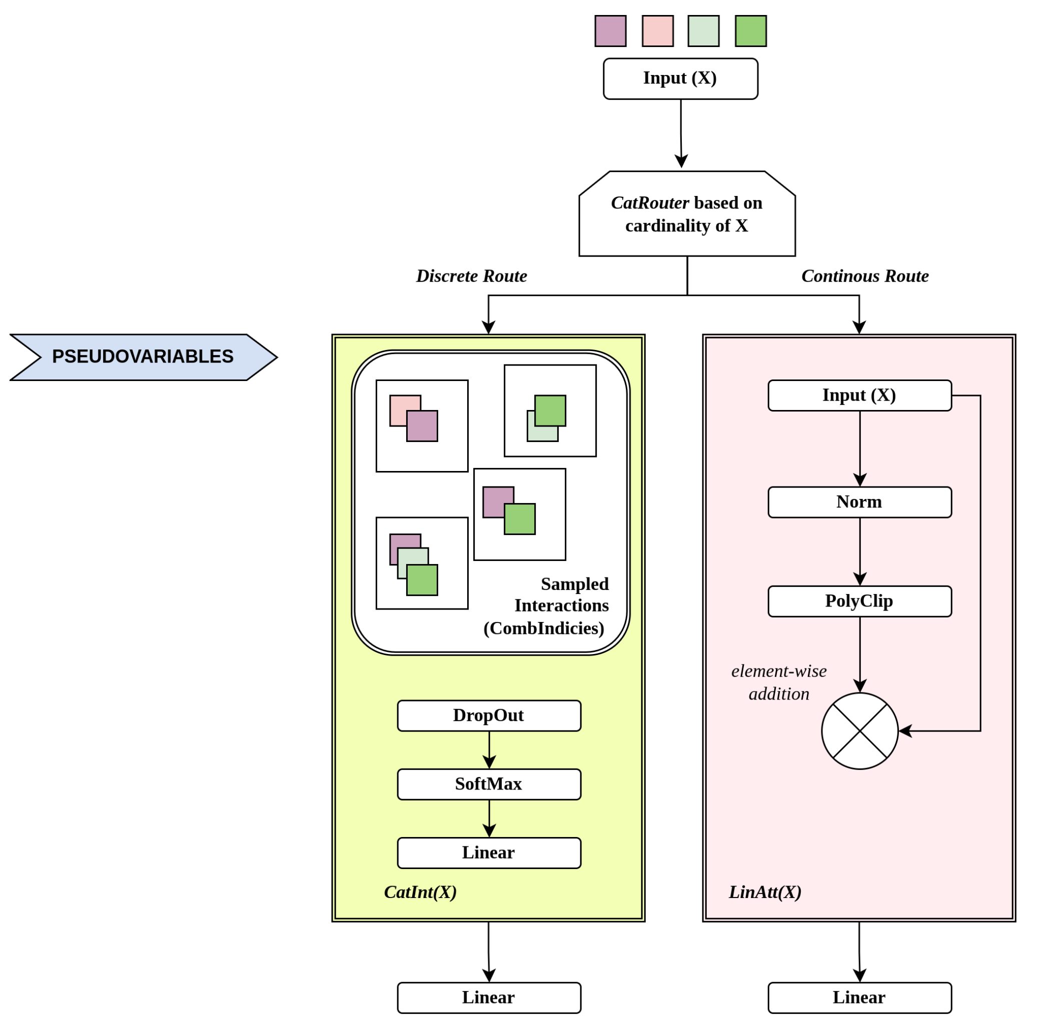


Figure 2. HorNets Architecture Overview

- **Routing Operator (CatRouter):** Distinguishes between discrete (CatInt) and continuous (LinAtt) data based on input cardinality (binary vs. high cardinality).
- **Categorical Interaction Module (CatInt):** Models feature interactions using factorization and incorporates **pseudovariables** for efficient handling of higher-order interactions. The key component is combActOp. This operator enables the traversal of considered feature combination space with incremental factorization (and activation).

$$\mathbf{F} = \text{combActOp}(\mathbf{M}, x, \text{CombIndices}), \quad x_0 = \text{dropout}(\mathbf{F}),$$

$$\text{x}_1 = \text{softmax}(x_0), \quad x_2 = x_1 \cdot \mathbf{w} + b.$$

$$\text{combActOp}(\mathbf{M}, \mathbf{x}, \text{CombIndices}) = \\ \mathbf{F}[:, \text{combIndex}] = \text{polyClip}(x[:, \text{CombIndices}] \cdot \mathbf{M}[:, \text{CombIndices}]),$$

- **Linear Attention Module (LinAtt):** Handles continuous data through linear attention mechanisms and utilizes PolyClip activation for enhanced interpretability.

$$x_0 = \frac{x}{\max(\|x\|_p, \epsilon)}, \quad x_1 = \text{polyClip}(x_0) \otimes x, \quad x_2 = x_1 \cdot \mathbf{w} + b.$$

Modelling Binary Logic

We aim to model arbitrary binary logical clauses of the form $\text{OP}(\{0,1\}^{a \times b}) \to \{0,1\}^a$, where $b > 2$ and $\text{OP} \in$ {AND, XOR, NOT, XNOR}.

- Achieves perfect F1 scores on logical operations; and outperforms Logistic Regression, Random Forest, MLP, and TabNET.

Problem	TabNet	Random Forest	MLP	Logistic Regression	HorNets -ReLU	**HorNets -PolyClip**
and(dim=4)	0.204 ± 0.136	1.0 ± 0.0	0.905 ± 0.134	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
and(dim=32)	0.331 ± 0.147	0.901 ± 0.122	0.899 ± 0.131	0.912 ± 0.107	0.821 ± 0.21	1.0 ± 0.0
and(dim=128)	0.463 ± 0.085	0.484 ± 0.093	0.655 ± 0.178	0.627 ± 0.185	0.811 ± 0.209	1.0 ± 0.0
not(dim=4)	0.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
not(dim=32)	0.146 ± 0.094	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
not(dim=128)	0.521 ± 0.169	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
or(dim=4)	0.574 ± 0.211	1.0 ± 0.0	0.961 ± 0.125	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
or(dim=32)	0.48 ± 0.175	0.91 ± 0.123	0.87 ± 0.168	0.92 ± 0.078	1.0 ± 0.0	1.0 ± 0.0
or(dim=128)	0.285 ± 0.073	0.558 ± 0.244	0.764 ± 0.199	0.699 ± 0.227	1.0 ± 0.0	1.0 ± 0.0
xnor(dim=4)	0.093 ± 0.067	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
xnor(dim=32)	0.19 ± 0.086	1.0 ± 0.0	0.948 ± 0.164	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
xnor(dim=128)	0.623 ± 0.261	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
xor(dim=4)	0.511 ± 0.18	1.0 ± 0.0	0.747 ± 0.15	0.497 ± 0.189	1.0 ± 0.0	1.0 ± 0.0
xor(dim=32)	0.436 ± 0.125	0.678 ± 0.158	0.647 ± 0.237	0.554 ± 0.188	0.911 ± 0.172	1.0 ± 0.0
xor(dim=128)	0.45 ± 0.097	0.473 ± 0.146	0.505 ± 0.143	0.584 ± 0.107	0.966 ± 0.082	1.0 ± 0.0

Table 1. Modelling of high-dimensional binary logic.

Key Takeaways

- **State-of-the-Art Performance:** HorNets excel on both synthetic and real-life datasets.
- **Novel Activation Function:** Introduction of PolyClip with a link to propositional logic.
- **Routing Mechanism:** Efficiently distinguishes between discrete and continuous data.
- **Logical Clause Retrieval:** Capable of extracting interpretable logical expressions, including noisy XNOR.
- **Superior to Baselines:** Outperforms gradient-boosted trees and AutoML classifiers.