New Study Establishes Supremacy of Complexity Science and Randomness in Machine Learning for Neurosymbolic Computation

Unsupervised, model-free method preserves key data better than traditional statistical techniques for next generation cognitive ML for multi-modal data.

MILS method

LONDON, UNITED KINGDOM, August 6, 2025 /EINPresswire.com/ -- A landmark study published in the international journal Information Sciences from Elsevier introduced a comprehensive framework for dimensionality reduction, feature selection, and network sparsification, establishing the first principled approach to neurosymbolic computation grounded in the formal theory of algorithmic probability and algorithmic randomness. The study provides both the theoretical foundations and empirical evidence for a new family of machine learning (ML) methods that not only compete with but outperform state-of-the-art statistical techniques across multiple domains and against a multiplicity of traditional but top performing methods currently used in ML.

Neurosymbolic computation combines the pattern-recognition power of neural networks with the logic and reasoning abilities of symbolic systems. It’s like giving a machine both the intuition to recognise apples and oranges from pictures and the reasoning skills to understand a recipe or solve a riddle about them.

The work, authored by an international consortium of researchers from top universities, presents the Minimal Information Loss Selection (MILS) algorithm—a model-free, unsupervised method that preserves the most structurally and causally significant elements of data while reducing its size. Unlike traditional statistical approaches, which rely on assumptions about data distribution or require supervision, MILS is based on algorithmic probability and perturbation analysis, allowing it to detect and retain meaningful patterns—including nonlinear and recursive structures—often invisible to conventional methods.


MILS supremacy

Algorithmic complexity is a way of measuring how simple or complicated something really is, based not on how it looks, but on how short a set of instructions you'd need to recreate it. Imagine trying to describe an apple: if it's a perfect red sphere with a short stem, you can describe it quickly—it has low algorithmic complexity. But a strangely shaped, multi-coloured apple with odd bumps would take longer to describe—it's more complex. Algorithmic probability builds on this by asking: what's the chance that something (like an apple or an image) could be produced by a short random model or computer program? If it's common or patterned, the chance is higher.

> "These methods prove the utility of algorithmic complexity over statistical measures like entropy and compression for real-world ML tasks. It is no longer a theoretical curiosity but a powerful tool."
>
> *Dr. Hector Zenil, Associate Professor, Kings' College London*

Meanwhile, algorithmic randomness refers to data that cannot be simplified at all—it looks truly random, like a pile of mixed apples and oranges with no obvious order. This differs from traditional compression techniques like ZIP or LZW, or measures like Shannon entropy, which only look at repeated patterns or frequency—like counting how many red apples versus green ones you've seen. Those methods can miss deeper structure. Algorithmic methods go further, detecting order even in things that look random on the surface but are actually generated by hidden rules—like discovering all your apples and oranges were laid out by a clever robot rather than by chance.

This new approach offers a general-purpose solution to core challenges in machine learning, demonstrating exceptional performance in preserving topological features in graphs and networks, as well as delivering the highest classification accuracy per bit in compressed image data. Tests conducted on standard benchmarks such as the MNIST dataset show that MILS surpasses leading dimensionality reduction algorithms including PCA, kernel methods, and spectral sparsification, especially in memory-constrained and quantised environments. These are common methods used widely in the ML community to characterise real-world data. Crucially, this new method does so by identifying the elements of data that contribute most to its algorithmic description, using no statistical shortcuts or training labels.

Beyond performance, the framework breaks new ground in its unification of symbolic reasoning and statistical learning. The authors show that MILS can serve as a foundation for neurosymbolic computation—providing a pathway to intelligent systems that combine the strengths of data-driven learning with the interpretability and generalisation capabilities of symbolic approaches. Grounded in algorithmic information theory, this work represents the first scalable and computable implementation of ideas long considered mathematically intractable.

The findings carry implications for fields ranging from biology and neuroscience to network science, image analysis, and next-generation AI architectures. The suite of tools led by MILS and other variations introduced in the paper, are particularly well-suited for applications requiring robust generalisation, such as edge computing, retrieval-augmented generation systems, and model- and causal-driven AI.

The full paper is available under open access.

PR Team
OxfordIA
email us here
