Multi-Modal Classification of Extended 1D Cellular Automata

Zhaoyun Chen [*1] **Xianquan Yan** [*1] **Hakan Akgün** [1] **N. Duane Loh** [1] **Ching Hua Lee** [1]

[*]Equal contribution [1]*Department of Physics, National University of Singapore, Singapore 117551.* Correspondence to: Hakan Akgün hakan.akgun@u.nus.edu, N. Duane Loh duaneloh@nus.edu.sg, Ching Hua Lee phylch@nus.edu.sg.

1. Introduction

One-dimensional cellular automata (1D CA) exhibit diverse dynamical behaviors ranging from trivial fixed points to universal computation. Wolfram's seminal four-class taxonomy [1] provides a foundational framework for understanding these behaviors, yet practical classification of CA rules remains challenging—particularly as the neighborhood radius r increases. For $r = 1$, only $2^{2^{2\times1+1}} = 2^8 = 256$ elementary rules exist; for $r = 2$, the rule space explodes to $2^{2^{2\times2+1}} = 2^{32} \approx 4.3 \times 10^9$ possibilities, rendering exhaustive enumeration infeasible and demanding scalable classification methods. Indeed, certain r=2 rules support universal computation [2], making scalable classification useful for screening computationally interesting rules. Prior analyses and learning-based classifiers have focused mainly on ECA [3, 4, 5], leaving the far larger r=2 regime largely open.

We present a multi-modal neural network approach to classify $r = 2$ CA into five behavioral prototypes: *Homogeneous*, *Blinker*, *Propagator*, *Chaotic*, and *Complex*. Our five-class taxonomy refines Wolfram's original framework by splitting Class II into two visually and dynamically distinct categories (Table 1), motivated by the observation that, under periodic boundary conditions, locally oscillating patterns (*Blinker*) and unidirectional traveling structures (*Propagator*) exhibit fundamentally different spectral signatures. Fig. 1 illustrates representative spacetime diagrams for each class.

Our model fuses log-power spectrograms, which capture global periodic and propagating structure, with 24 physics statistics summarizing density, entropy, and compressibility. This helps separate classes that look locally similar in spacetime images, achieving **95.86%** test accuracy and generalizing across initial conditions without architectural changes.

Table 1: **Comparison between Wolfram's *four*-class taxonomy and our *five*-class refinement for 1D CA.** Wolfram's Class II is split based on the origin of periodicity.

Wolfram	Ours	Behavior
Class I	Homogeneous	Uniform fixed point
Class II	Blinker	Locally periodic
Class II	Propagator	Traveling periodic
Class III	Chaotic	Aperiodic
Class IV	Complex	Patterns

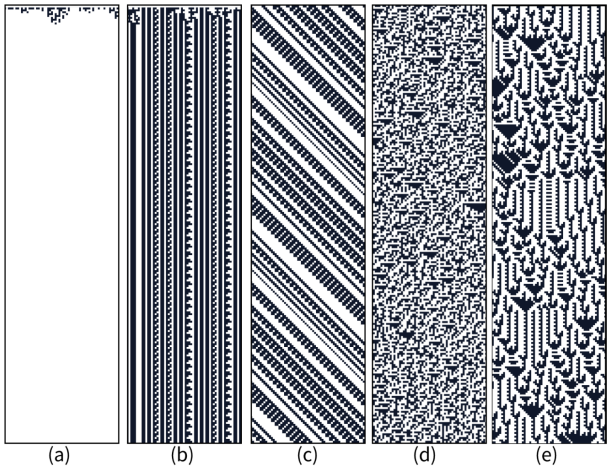

Fig. 1: **Representative spacetime diagrams for the five behavioral classes** Panels **(a–b)** show the first 300 steps from the initial condition (no burn-in), whereas **(c–e)** show a 300-step window *after* a burn-in period. **(a)** Homogeneous, **(b)** Blinker, **(c)** Propagator, **(d)** Chaotic, **(e)** Complex

2. Method

Data Generation. Given a rule R with radius r=2, we initialize a lattice of width W=300 with random density 0.5 and evolve under *periodic boundary conditions* for T_{burn}=10,000 transient steps, then record T=2,000 steps to obtain a sample of spatiotemporal patterns (spacetime matrix) $\mathbf{S} \in \{0,1\}^{T\times W}$.

Multi-Modal Classifier Architecture. Fig. 2 illustrates a two-branch pipeline designed to combine complementary global and statistical evidence:

Spectral Branch (Image). We compute the 2D Discrete Fourier Transform (DFT) of the spacetime diagram $\mathbf{S}$, apply log-power scaling, resize, and process through a ResNet [6]. This yields an embedding encoding global frequency landscape of the spatiotemporal pattern.

Physics Contextualization Branch (Feature Vector). We compute 24 established statistics. The most consequential 7 features are: final-state density, spatial standard deviation, temporal change rate, Langton's λ [7], Compressibility [8], Shannon entropy [9], and a Lyapunov-exponent proxy measuring divergence under 1-bit perturbation [10] (details in Appendix B). An MLP processes these features into a 64-dimensional embedding, modulated by a learned attention gate [11].

Fusion. The two embeddings are concatenated and passed through a fusion MLP, followed by a final linear classifier head. Training uses cross-entropy loss

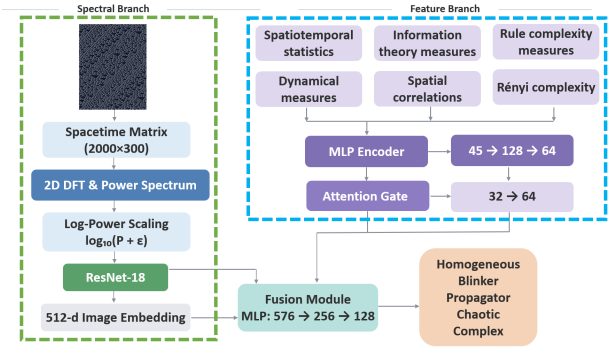

Fig. 2: **Multi-modal classifier architecture.** *Spectral Branch*: spacetime matrix transformed via 2D DFT and encoded by ResNet. *Feature Branch*: twenty-four physics statistics through MLP with attention gating. *Fusion*: concatenated embeddings processed by a final classifier head for 5-class logits.

Table 2: **Classification performance** on the rule-disjoint test split.

Class	Prec.	Recall	F_1
Homogeneous	0.993	1.000	0.996
Blinker	0.972	0.986	0.979
Propagator	0.970	0.936	0.953
Chaotic	0.992	0.893	0.940
Complex	0.878	0.979	0.926
Overall	**Accuracy: 95.86%**		

with inverse-frequency class weights to handle imbalance.

Dataset and Splits. We curated 3,500 unique CA rules, each **manually labeled** into one of five behavioral classes through visual inspection of spacetime diagrams. We split by *rule ID* and with ratio train/val/test = 70/10/20, ensuring *no rule appears in multiple splits* (rule-disjoint splits)[1]. The test set comprises 700 samples.

3. Results

Table 2 summarizes performance on the rule-disjoint test split. The model achieves **95.9%** accuracy, demonstrating strong generalization to unseen rules. *Homogeneous* rules are classified perfectly. The confusion matrix (row-normalized) in Fig. 3 indicates that errors concentrate at the *Chaotic–Complex* boundary—consistent with the observation that complex and chaotic rules can exhibit visually indistinguishable local textures.

4. Robustness to Initialization Randomness and Scalability

A key concern is whether predictions reflect intrinsic rule dynamics or seed-dependent artifacts. We evaluated 2,000 random r=2 rules, each evolved from 10 independent seeds (Table 3). Results show **90.8%** of rules receive consistent majority-class predictions, with mean intra-rule consistency of **94.7%**. The class distribution confirms that generic random rules rarely collapse to trivial dynamics.

[1]Rule-disjoint splitting prevents the same rule appearing across splits, so accuracy reflects generalization across rules.

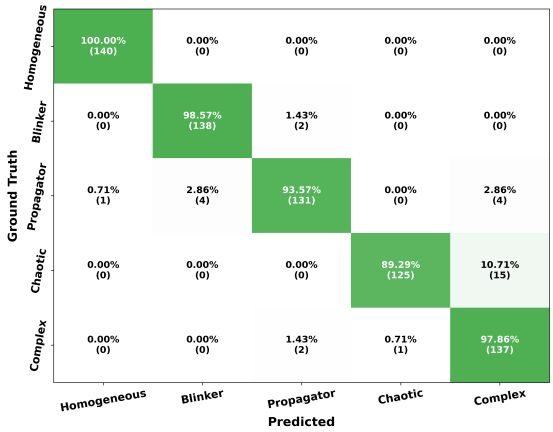

Fig. 3: **Confusion matrix (row-normalized) on the test set.** The primary off-diagonal mass is *Chaotic→Complex*, reflecting the subtle boundary between high-entropy dynamics and edge-of-chaos nuances.

Table 3: **Robustness to Initialization Randomness** (2000 random r=2 rules × 10 seeds each). *Consistency* = fraction of seeds agreeing with the rule's majority predicted class; *Count* = number of rules in each consistency bin. *Pct.* = fraction of all 20,000 (rule,seed) runs predicted as each class. High agreement (90.8% rules with ≥80% consistency; mean 94.7%) indicates seed-insensitive predictions.

Consistency	Count	Class	Pct.
≥80%	1815	Homogeneous	1.7%
50–79%	183	Blinker	12.9%
<50%	2	Propagator	37.4%
		Chaotic	24.5%
Overall Mean:	**94.7%**	Complex	23.5%

5. Conclusion and Outlook

We presented a multi-modal neural network approach that is scalable and generalizable in classifying 1D cellular automata with extended neighborhood. We refined Wolfram's classification on ECA and classify extended CA into *five* behaviorally distinct classes (Table 1, Fig. 1) with 95.86% accuracy and strong robustness to initialization randomness. The architecture naturally applies to larger neighborhood radii without modification, requiring only re-training on appropriately labeled data.

We are optimizing our pipeline to accelerate data generation and curating a *large-scale* dataset via semi-supervised learning. Future directions include: (1) active learning to efficiently label the vast rule space with various neighborhood radii, (2) extension to higher-dimensional CA, and (3) using the classifier to identify candidate rules in Complex class at scale and analyze what rule structure underlies "complexity".

References

[1] Stephen Wolfram. Universality and complexity in cellular automata. *Physica D: Nonlinear Phenomena*, 10(1-2):1–35, 1984.

[2] Matthew Cook. Universality in elementary cellular automata. *Complex Systems*, 15(1):1–40, 2004.

[3] Wentian Li, Norman Packard, et al. The structure of the elementary cellular automata rule space. *Complex systems*, 4(3):281–297, 1990.

[4] J. Machicao, L. C. Ribas, L. F. S. Scabini, and O. M. Bruno. Cellular automata rule characterization and classification using texture descriptors. *Physica A: Statistical Mechanics and its Applications*, 497:109–117, 2018.

[5] T. Comelli, F. Pinel, and P. Bouvry. Comparing elementary cellular automata classifications with a convolutional neural network. In *Proceedings of ICAART 2021*, pages 467–474, 2021.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Chris G Langton. Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D: nonlinear phenomena*, 42(1-3):12–37, 1990.

[8] Hector Zenil and Jean-Paul Delahaye. On the algorithmic nature of the world. *Information and computation*, 2:477–496, 2010.

[9] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[10] Franco Bagnoli, R Rechtman, and Stefano Ruffo. Damage spreading and lyapunov exponents in cellular automata. *Physics Letters A*, 172(1-2):34–38, 1992.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[12] Howard A Gutowitz, Jonathan D Victor, and Bruce W Knight. Local structure theory for cellular automata. *Physica D: Nonlinear Phenomena*, 28(1-2):18–48, 1987.

[13] Pooya Hatami, Raghav Kulkarni, and Denis Pankratov. Variations on the sensitivity conjecture. *arXiv preprint arXiv:1011.0354*, 2010.

[14] Stephen Wolfram. Statistical mechanics of cellular automata. *Reviews of modern physics*, 55(3):601, 1983.

[15] Ming Li, Paul Vitányi, et al. *An introduction to Kolmogorov complexity and its applications*, volume 3. Springer, 2008.

[16] Andrew Wuensche. *Complexity in one-D cellular automata: Gliders, basins of attraction and the Z parameter*. University of Sussex, School of Cognitive and Computing Sciences, 1994.

[17] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, volume 4, pages 547–562. University of California Press, 1961.

[18] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

Appendix A. Embedding Visualizations (UMAP)

To better interpret what the model learns, we visualize the embeddings (latent representation) at several stages of the pipeline. Specifically, we apply UMAP to (i) the raw (input) spectral features, (ii) the raw (input) physics feature vector, and (iii) the penultimate-layer embedding after multimodal fusion and before mapping to logits in the last layer. These dimension reductions provide a qualitative view of how data distribution evolves from the input space to the learned, semantically organized embedding space.

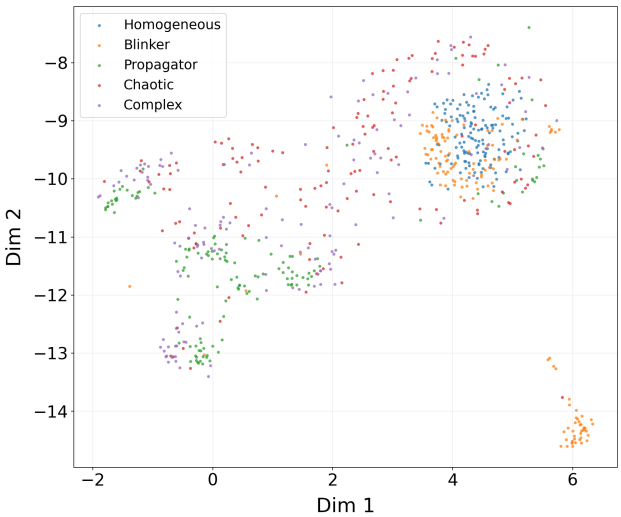

Fig. A1: **Raw spectral representation (UMAP-2D).** UMAP projection of the *raw flattened log-power spectrogram* features (before the spectral encoder), with points colored by class. This view reflects class separability available from the frequency-domain representation alone: some classes form partially distinct regions, while substantial overlap remains, indicating that raw spectral cues are informative but not fully discriminative without learned feature extraction.

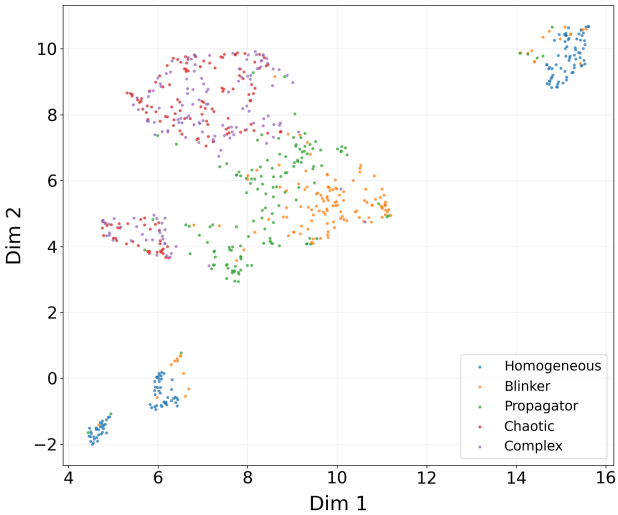

Fig. A2: **Raw physics features (UMAP-2D).** We project the 24 handcrafted physics statistics (before the feature-branch MLP) into 2D using UMAP, with points colored by class. Compared to the raw spectral features, these physics features make several classes cluster more clearly, suggesting they provide complementary dynamical information. Some overlap still remains, especially between the Chaotic and Complex classes, which are inherently harder to separate.

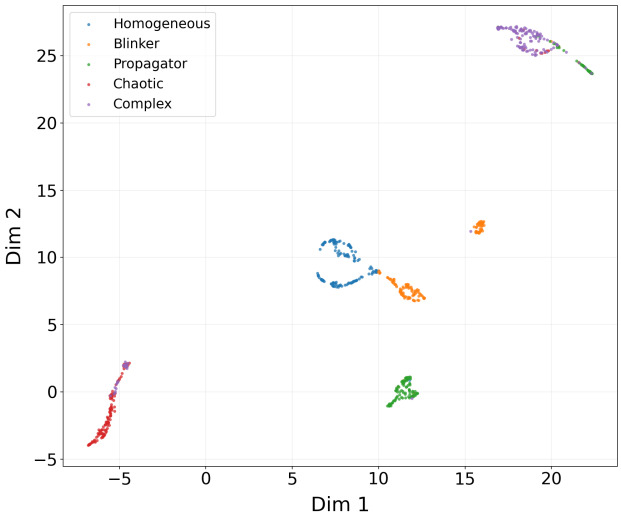

Fig. A3: **Penultimate-layer embedding (UMAP-2D).** UMAP projection of the penultimate-layer (pre-logit) embedding of the trained multimodal classifier, colored by class. The five classes form well-separated and compact clusters with minimal overlap, indicating that the fused representation is highly discriminative and aligns closely with the behavioral taxonomy. Only a few clusters remain close to each other, which marks where the model is most likely to confuse classes.

Appendix B. Feature Definitions

We define all physics-informed features used in the multi-modal classifier. Let $\mathbf{S} \in \{0,1\}^{T\times W}$ denote the spacetime matrix with T=2000 time steps and width W=300, where $S_{t,x}$ is the cell state at position x and time t. Let R denote the rule table mapping neighborhood configurations to output states, with radius r=2 and neighborhood size n=2r+1=5.

2.1 Basic Spatiotemporal Statistics

(1) Final-State Density. The fraction of active cells in the final configuration:

$$\rho_{\text{final}} = \frac{1}{W}\sum_{x=1}^{W} S_{T,x} \tag{A1}$$

(2) Spatial Standard Deviation. The spatial variation of the final state:

$$\sigma_{\text{spatial}} = \sqrt{\frac{1}{W}\sum_{x=1}^{W}\left(S_{T,x} - \rho_{\text{final}}\right)^2} \tag{A2}$$

(3) Temporal Change Rate. The mean absolute difference between consecutive time steps over the final 100 steps:

$$\Delta_{\text{temporal}} = \frac{1}{99W}\sum_{t=T-99}^{T-1}\sum_{x=1}^{W}\left|S_{t+1,x} - S_{t,x}\right| \tag{A3}$$

2.2 Information-Theoretic Measures

(4) Shannon Entropy [9]. The average information content per cell:

$$H = -p_0 \log_2 p_0 - p_1 \log_2 p_1 \tag{A4}$$

where p_0 and p_1 are the empirical frequencies of states 0 and 1 in $\mathbf{S}$.

(5) Compressibility [8]. The zlib compression ratio:

$$\gamma = \frac{\text{compressed size}(\mathbf{S})}{\text{original size}(\mathbf{S})} \tag{A5}$$

2.3 Rule-Based Complexity Measures

(6) Langton's λ Parameter [7]. The fraction of non-quiescent configurations:

$$\lambda = \frac{1}{2^n}\sum_{c=0}^{2^n-1} R(c) \tag{A6}$$

where $R(c) \in \{0,1\}$ is the rule output for configuration c.

(7) Mean-Field Cluster Statistics [12]. We first compute the conditional transition probabilities for each cluster size k:

$$\mu_k = \frac{1}{|C_k|}\sum_{c\in C_k} R(c), \quad k = 0,1,\ldots,n \tag{A7}$$

where C_k is the set of all n-bit configurations with exactly k ones. For r=2 (n=5), the vector $\{\mu_k\}_{k=0}^{n}$ has

length 6; we convert it into **four** scalar features by taking its mean, standard deviation, minimum, and maximum.

(8) Sensitivity Parameter μ [13]. The average change in rule output when a single bit in the neighborhood is flipped:

$$\mu_{\text{sens}} = \frac{1}{n \cdot 2^n} \sum_{c=0}^{2^n-1} \sum_{i=0}^{n-1} |R(c) - R(c \oplus 2^i)| \tag{A8}$$

where $\oplus$ denotes bitwise XOR.

(9) Obstruction Parameter θ [14]. Measures the degree to which the rule output depends on specific bit positions:

$$\theta = \frac{1}{n} \sum_{i=0}^{n-1} |\Pr[R(c) = 1 \mid c_i = 1] - \Pr[R(c) = 1 \mid c_i = 0]| \tag{A9}$$

(10) Kolmogorov Rule Complexity [15]. Approximated by the zlib compression ratio of the rule table itself:

$$K(R) = \frac{\text{compressed size}(R)}{\text{original size}(R)} \tag{A10}$$

(11) Z-Parameter (Reversibility) [16]. The fraction of evolved states that can be uniquely inverted:

$$z = \frac{1}{N \cdot T_{\text{trials}}} \sum_{j=1}^{T_{\text{trials}}} \sum_{t=1}^{N} \mathbf{1}[\text{preimage of } s_t^{(j)} \text{ is unique}] \tag{A11}$$

computed over T_{trials}=64 random initial conditions of length N=64.

2.4 Dynamical Measure

(12) Lyapunov Exponent Proxy [10]. Measures sensitivity to single-bit perturbations. Let $\mathbf{S}'$ be the perturbed trajectory:

$$\ell = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \log \left(1 + \sum_{x=1}^{W} |S_{t,x} - S'_{t,x}| \right) \tag{A12}$$

2.5 Spatial Correlation Functions

(13) Two-Point Correlation [14]. The spatial autocorrelation at distance d:

$$C(d) = \frac{1}{W} \sum_{x=1}^{W} S_{T,x} \cdot S_{T,(x+d \bmod W)} \tag{A13}$$

We compute $C(1)$ and $C(5)$ as separate features.

2.6 Rényi Complexity Measures

(14) Rényi spatial set measure [17]. (Shannon limit, α=1)

$$I_{1,b} = -\frac{1}{b} \sum_{p \in \mathcal{P}_b} p_p \log_2 p_p \tag{A14}$$

(15) Rényi temporal entropy [17]. (Shannon limit, α=1)

$$H_{1,T_b} = -\frac{1}{T_b} \sum_{q \in Q_{T_b}} p_q \log_2 p_q \tag{A15}$$

2.7 Spectral Analysis

(16) Power Spectrum Summary Features [18]. We compute the site-averaged temporal power spectrum by demeaning each site over time and applying an (R)FFT along the time axis:

$$X_{t,x} = S_{t,x} - \frac{1}{T} \sum_{t'=1}^{T} S_{t',x} \tag{A16}$$

$$F_{k,x} = \frac{1}{T} \sum_{t=1}^{T} X_{t,x}\, e^{-2\pi i k t/T} \tag{A17}$$

$$S_k = \langle |F_{k,x}|^2 \rangle_x, \quad k = 0, \ldots, \lfloor T/2 \rfloor. \tag{A18}$$

From $\{S_k\}$ we compute a small dictionary of scalar descriptors (e.g., mean, std, min, max, median, band-averages, peak frequency), and apply a generic `scalarize` operator to obtain **five** features.

2.8 Feature Preprocessing

All features are standardized to zero mean and unit variance before feeding into the MLP branch:

$$\tilde{f}_i = \frac{f_i - \mu_i}{\sigma_i} \tag{A19}$$

where μ_i and σ_i are computed from the training set. NaN and infinite values are replaced with zero before standardization.