Distinct Regimes of Predictive Accessibility in Elementary Cellular Automata

Rodrigo A. Garrido
Faculty of Engineering & Sciences
Universidad Adolfo Ibáñez
Av. Diagonal Las Torres 2700, Peñalolén
Santiago, RM, Chile

Predicting the evolution of deterministic dynamical systems is often treated as a problem of model capacity, architecture design or training procedure. However, in local dynamical systems, long-term predictability may be constrained not only by the expressivity of the predictor, but also by the way dynamically relevant information propagates, disperses and remains accessible under finite learned representations. In this paper, we investigate this issue using elementary cellular automata (ECAs) as a minimal setting in which local deterministic rules generate qualitatively different global dynamics. Rather than framing prediction failure as a generic limitation of machine learning models, we study the emergence of distinct regimes of *predictive accessibility*: the extent to which information relevant for future states can be captured and exploited by finite learned representations under a given observational framework. Using neural predictors with controlled receptive fields, we show that different cellular automaton rules exhibit qualitatively different forms of predictive behavior. Rule 30 displays an abrupt loss of accessible predictive structure beyond a receptive-field-dependent horizon. Rule 45 exhibits a smoother degradation regime. Rule 90, despite being deterministic and linear over $\mathbb{F}_2$, remains near chance-level performance under the tested representations, consistent with distributed XOR-like dependencies that are difficult to compress statistically. Rule 110, by contrast, retains substantial predictive structure over longer horizons, suggesting that complex dynamics need not imply immediate loss of predictability. These results indicate that prediction limits in ECAs are not adequately described as a uniform consequence of error accumulation or insufficient model capacity. Instead, they reflect rule-dependent differences in information structure, causal accessibility and statistical compressibility. The proposed framework reframes predictability as an empirical property of the interaction between dynamics and representation, rather than as a purely algorithmic capability.

Keywords: elementary cellular automata; predictive accessibility; dynamical systems; machine learning; information propagation; statistical compressibility

https://doi.org/10.25088/ComplexSystems.35.3.267

1. Introduction

Determinism does not necessarily entail practical predictability. A dynamical rule may uniquely determine every future state while still offering no substantially shorter route to that state than explicitly following the system's evolution. Cellular automata provide a particularly clear setting in which to examine this distinction. Their microscopic dynamics are finite, local and exactly specified, yet repeated application of simple rules can generate spatially extended structures, apparent randomness and complex forms of information propagation.

This tension has long motivated research on the predictability of cellular automata. Wolfram's early studies showed that simple cellular automaton rules can generate qualitatively distinct forms of behavior, including irregular and computationally complex dynamics [1, 2]. This line of research later led to the notion of computational irreducibility: for some systems, determining a sufficiently detailed future state may require carrying out a computation comparable to the system's own step-by-step evolution [3]. The relevant difficulty is therefore not uncertainty about the governing rule or noise in the observations. Rather, it concerns the possible absence—or practical inaccessibility—of computational shortcuts capable of recovering future states without reproducing most of the intervening dynamics.

Computational irreducibility, however, does not imply that every aspect of a system is equally unpredictable. Israeli and Goldenfeld showed that cellular automata, including systems associated with universal computation, may admit coarse-grained descriptions whose large-scale evolution is simpler and more predictable than their microscopic dynamics [4]. Their results demonstrate that predictability depends partly on the level of description: a system may resist exact microscopic forecasting while retaining decidable or compact descriptions of selected macroscopic features.

A complementary line of research has approached cellular automaton dynamics through algorithmic complexity and compression [5]. Analyses based on compression treat regularities in spacetime evolutions as empirical indicators of descriptive simplicity and use them to distinguish dynamical classes, sensitivity to initial conditions and possible phase-transition behavior. From this perspective, prediction is related not only to the deterministic specification of the rule, but also to whether the resulting evolution contains exploitable regularities that can be represented more compactly than the raw trajectory itself.

These approaches frame predictability as a search for different kinds of computational reduction. Coarse-graining asks whether selected large-scale properties can be evolved by a simpler effective rule. Methods based on compression ask whether generated spacetime

patterns exhibit algorithmically exploitable regularity. Empirical investigations of computational irreducibility ask whether a predictor can outrun the original computation, at least for bounded systems, finite horizons or restricted observables [6]. None of these approaches establishes universal predictability or irreducibility from finite experiments. They instead provide operational ways of studying where predictive shortcuts appear, how far they extend and under which representational constraints they fail.

Moreover, elementary cellular automata (ECAs) have been extensively studied as models of causal structure, information propagation, emergent computation and algorithmic complexity. They therefore provide a natural experimental setting in which to investigate how different forms of dynamical organization affect empirical predictability under a common forecasting framework.

Because the governing rule is known exactly, the evolution is fully deterministic, and observational noise is absent, ECAs provide an unusually well-controlled environment for studying predictability. Differences in forecasting behavior can therefore be attributed more directly to differences in the organization of the underlying dynamics than to uncertainty about the system, noisy observations or model misspecification.

To study this question, we introduce the notion of *predictive accessibility*. Rather than asking whether future states are predictable in principle or whether a computational shortcut exists in the formal sense, predictive accessibility characterizes the extent to which future-relevant information remains empirically recoverable by learned predictors operating under a common experimental protocol. It is therefore not a property of a particular neural architecture, but an operational property of the interaction between a dynamical system and a controlled predictive framework.

Our experimental design uses neural predictors as measurement instruments rather than as the object of study. Representative ECAs provide deterministic systems exhibiting diverse forms of dynamical organization, while a common forecasting protocol fixes the training procedure, receptive field (RF), temporal history, model capacity and evaluation methodology across rules. This allows differences in forecasting behavior to be attributed primarily to differences in the organization of the underlying dynamics rather than to changes in the prediction method.

Because the experiments concern finite systems and finite horizons, they do not establish computational irreducibility in the formal sense. They instead reveal where predictive shortcuts are empirically accessible to particular finite representations and where this accessibility deteriorates.

https://doi.org/10.25088/ComplexSystems.35.3.267

Accordingly, we investigate whether representative ECAs exhibit distinct regimes of predictive accessibility rather than a single generic pattern of forecasting degradation under a common experimental protocol.

The results show that predictive accessibility does not evolve uniformly across representative ECAs. Different rules exhibit qualitatively distinct forecasting regimes: some show abrupt accessibility loss beyond a short prediction horizon, others deteriorate progressively, some remain near chance throughout the tested horizon range, whereas others preserve structured and recoverable dependencies over substantially longer temporal scales.

The central contribution of this paper is an empirical framework for studying predictive accessibility in deterministic dynamical systems, together with an operational characterization of distinct predictive regimes in representative ECAs. By combining controlled multi-step forecasting experiments with comparisons across rules exhibiting diverse dynamical organization, we show that predictability depends not only on deterministic evolution itself, but also on how future-relevant information is distributed, propagated and remains accessible under a common predictive protocol. The notion of predictive accessibility complements coarse-graining and compression-based approaches by focusing on the empirical recoverability of microscopic future states rather than on formal computational reductions.

The remainder of the paper is organized as follows. Section 2 reviews related work on cellular automata, dynamical prediction and information-theoretic approaches to complexity. Section 3 describes the cellular automata, forecasting task, model classes and evaluation metrics. Section 4 presents the empirical results and identifies distinct predictive regimes across representative rules. Section 5 discusses the implications of these findings for deterministic dynamical systems and finite predictive representations. Section 6 concludes the paper.

2. Related Work

2.1 Cellular Automata as Minimal Models of Dynamical Complexity

Cellular automata have long been used as minimal models for studying how complex macroscopic behavior can emerge from simple, deterministic and spatially local rules. In his foundational work, Wolfram showed that even one-dimensional binary cellular automata with nearest-neighbor interactions can generate a wide range of behaviors, from homogeneous and periodic patterns to apparent randomness and complex localized structures [1, 3]. This makes ECAs a particularly useful setting for studying prediction: the microscopic

rule is fully known, yet the resulting global dynamics may be difficult to summarize or anticipate.

A central insight of the cellular automata literature is that dynamical complexity cannot be inferred simply from the local rule table. Wolfram's classification into ordered, periodic, chaotic and complex regimes provided an influential qualitative taxonomy [1, 3]. However, later studies showed that classification is itself nontrivial. Wuensche [7], for example, proposed automated methods for classifying cellular automata using input-entropy variance, attractor-basin structure and the detection of gliders and related spacetime structures. Martínez [8] reviewed multiple elementary cellular automaton (ECA) classification schemes, including Wolfram's classes, Wuensche's classification, communication-complexity approaches, power spectra, compression-based measures and morphological diversity.

These works establish that different ECA rules organize information in fundamentally different ways. Some rules rapidly converge to simple attractors; others generate repetitive or nested structures; others produce apparent randomness; and some support persistent localized structures. However, most classification approaches focus on the intrinsic morphology or dynamical organization of the automaton itself. They do not directly ask how such organization constrains long-horizon forecasting under finite representations.

This distinction is important for the present paper. We do not aim to introduce a new global classification of ECA rules. Instead, we ask a more specific question: how does the information structure generated by a rule affect its *predictive accessibility* under finite learned representations? Addressing this question requires connecting the classification of cellular automaton dynamics with earlier work on computational irreducibility, coarse-graining, compression and predictive representation.

2.2 Computational Irreducibility, Coarse-Graining and Compression

The difficulty of predicting cellular automata has also been studied through the concepts of computational irreducibility, coarse-graining and algorithmic complexity. Wolfram's early analyses showed that simple deterministic rules can generate evolutions whose global behavior cannot be inferred directly from their local specification [2]. This line of research later motivated the notion of computational irreducibility: for some systems, no general procedure is known that can determine their future states substantially more efficiently than explicitly following the intermediate evolution.

Beyond its algorithmic formulation, computational irreducibility has also been discussed within broader debates on emergence and the limits of reductionist explanation. Hosseinie and Mahzoon [9] argue

https://doi.org/10.25088/ComplexSystems.35.3.267

that irreducibility reflects the insufficiency of purely reductionist descriptions for understanding complex systems, emphasizing that explanatory power may require multiple levels of organization. Although the present work remains entirely computational, this broader perspective is consistent with our view that predictive accessibility depends on how dynamical information is organized and represented, rather than on determinism alone.

Computational irreducibility should not, however, be interpreted as the claim that every property of an irreducible system is equally difficult to predict. A system may resist exact microscopic forecasting while still admitting simpler descriptions of selected macroscopic properties. Israeli and Goldenfeld [4] demonstrated that certain cellular automata can be coarse-grained into simpler effective automata that preserve relevant large-scale behavior.

This distinction is important because prediction always concerns a particular representation and a particular target. Exact prediction of every cell may require reproducing the complete microscopic evolution, whereas prediction of a coarse observable may admit an effective shortcut. Predictability therefore cannot be characterized solely as a binary distinction between reducible and irreducible systems. It also depends on which dynamical features are retained by the representation and which aspects of the future are being predicted.

A complementary line of research has approached cellular automaton dynamics through algorithmic complexity and compression [5]. Analyses based on compression treat regularities in spacetime evolutions as empirical indicators of descriptive simplicity and use them to distinguish dynamical classes, sensitivity to initial conditions and possible behavior near phase transitions. From this perspective, prediction is related not only to the deterministic specification of the rule, but also to whether the resulting evolution contains exploitable regularities that can be represented more compactly than the raw trajectory itself.

Zenil et al. [6] further investigated computational irreducibility empirically by examining whether the behavior generated by a system admits descriptions or computational procedures that are substantially shorter than explicit step-by-step evolution. Such analyses do not constitute formal proofs of irreducibility. Rather, they provide operational evidence about the extent to which regularities, symmetries or predictive reductions can be detected in generated trajectories.

These approaches frame predictability as a search for distinct forms of computational reduction. Coarse-graining asks whether selected large-scale properties can be evolved through a simpler effective rule. Methods based on compression ask whether generated spacetime patterns contain regularities permitting a shorter description.

Learned prediction, by contrast, asks whether a finite parametric representation can recover the dependencies required to estimate a future configuration from observed data.

The present paper is concerned primarily with the third form of reduction, while using the first two as theoretical context. Neural predictors are not treated here as formal demonstrations of computational reducibility or irreducibility. Instead, they serve as empirical probes of whether future-relevant dependencies remain accessible under controlled restrictions on RF, temporal history and representational capacity. This formulation connects predictive accessibility to earlier work on computational reduction without equating forecasting performance with a formal complexity-theoretic result.

2.3 Emergent Computation, Causal Structure and Information Processing

Whereas computational irreducibility and compression concern the possibility of simplifying or describing an evolution, another major line of research treats cellular automata as distributed information-processing systems. Crutchfield and Mitchell [10] showed that cellular automata can support emergent computation through localized particles, domain boundaries and their interactions, emphasizing that the relevant dynamical structures are not merely binary configurations, but coherent patterns through which information is stored, transmitted and transformed.

These ideas are closely connected to computational mechanics. Crutchfield and Young [11] introduced statistical complexity as a measure of the information required for optimal prediction, while Shalizi and Crutchfield [12] formalized causal states as equivalence classes of pasts yielding identical conditional futures. From this perspective, predictability depends not only on deterministic evolution, but on whether dynamically relevant information remains reconstructable from observable histories.

This perspective is particularly relevant for cellular automata because local update rules can generate qualitatively different forms of information propagation, including localized coherent structures, distributed parity relations and chaotic mixing dynamics. Consequently, forecasting degradation may arise not only from insufficient model capacity, but from the way predictive information becomes distributed across expanding causal regions.

The present paper adopts a complementary empirical approach. We do not attempt to reconstruct causal states or derive an optimal predictive representation in the formal sense of computational mechanics. Instead, we examine how different ECAs behave when prediction is attempted using finite neural representations with controlled receptive fields (RFs). Predictive accessibility therefore refers

https://doi.org/10.25088/ComplexSystems.35.3.267

here to an operational property of the interaction between dynamics, observation and representation, rather than to a claim that the learned models recover the minimal sufficient statistics of the underlying process.

2.4 Machine Learning and Prediction of Dynamical Systems

Machine learning has become a major tool for approximating dynamical systems from data. Recurrent neural networks, including long short-term memory networks [13], were designed to capture temporal dependencies in sequential data. Reservoir computing and echo-state networks (ESNs) provide another influential approach, in which a fixed high-dimensional dynamical reservoir is used to represent temporal history, and only the readout is trained [14]. More recently, machine learning methods have been used successfully for model-free prediction of chaotic and spatiotemporally chaotic systems. Pathak et al. [15], for example, showed that reservoir computing can predict large spatiotemporally chaotic systems from data over finite horizons.

These studies demonstrate that data-driven models can learn useful approximations of complex dynamics, especially at short and intermediate horizons. Yet they also highlight a persistent difficulty: accurate short-term prediction does not guarantee reliable long-term forecasting. In chaotic or high-dimensional systems, small errors may amplify, learned trajectories may leave the training distribution, and iterative prediction may eventually become unstable.

The machine-learning literature on multi-step forecasting distinguishes several strategies for dealing with this problem. Recursive methods train a one-step model and iterate it forward, direct methods train separate models for each forecast horizon, and multi-output or sequence-to-sequence methods predict several future steps jointly. Taieb et al. [16] reviewed and compared multi-step forecasting strategies, emphasizing that different methods involve different bias–variance and error-accumulation tradeoffs. Bontempi et al. [17] similarly discussed machine-learning strategies for time-series forecasting, focusing on how prediction strategy affects empirical performance.

This literature is essential for interpreting prediction failure carefully. A decline in long-horizon accuracy may reflect recursive error accumulation, insufficient training data, limited model capacity, distribution shift or a poorly chosen forecasting strategy. Such effects are properties of the learning and forecasting procedure and should not be confused automatically with intrinsic unpredictability of the underlying system.

The present paper addresses this distinction through a controlled deterministic setting. ECAs allow the observational framework, training protocol, model family and evaluation procedure to be held fixed

while the underlying dynamical rule is varied. We can therefore ask whether prediction loss occurs uniformly, as might be expected from a generic error-accumulation explanation, or whether different rules produce qualitatively distinct profiles of predictive accessibility under the same representational constraints.

2.5 Learning Cellular Automata with Neural Networks

Several studies have examined the relationship between cellular automata and neural-network architectures. Gilpin [18] showed that cellular automata can be represented using convolutional neural networks and that neural networks can learn cellular automaton rules from observed spacetime data. This work is especially relevant because it connects the locality of cellular automata with the local structure of convolutional architectures and demonstrates that one-step transition operators can be represented accurately by learned models.

The ability to learn a local cellular automaton update rule, however, does not settle the question of multi-step predictability. Learning the one-step operator and estimating a configuration several steps into the future are distinct predictive tasks. In a radius-one cellular automaton, the causal region affecting a target cell expands with the prediction horizon. A predictor may therefore approximate the local update rule accurately, while lacking access to all the variables on which a future-state prediction depends.

This observation motivates the RF analysis developed in the present paper. In convolutional predictors, RF provides an operational measure of the spatial context available to the model. If relevant dependencies extend beyond that context, prediction cannot in general remain exact. Nevertheless, RF coverage alone does not determine empirical predictability. Dependencies inside the accessible region may differ substantially in their algebraic structure, spatial distribution, compressibility and sensitivity to local perturbations.

The unresolved issue is therefore not whether neural networks can learn cellular automata at short horizons. Prior work has established that they can. The issue is how different ECA rules organize future-relevant dependencies and how that organization appears in multi-step prediction under finite representations. The present paper addresses this question by comparing rules with ordered, periodic, chaotic, additive and complex dynamics and by examining whether forecasting performance remains stable, declines gradually, collapses abruptly or remains near chance within the tested horizon range.

2.6 Gap Addressed by This Paper

The literature reviewed provides several complementary foundations. Cellular automaton studies show that simple local rules can generate

qualitatively different forms of dynamical organization [1, 2, 7, 8]. Research on computational irreducibility, coarse-graining and compression has shown that predictability depends on the possibility of constructing effective computational reductions or compact representations of the dynamics [4–6]. Computational mechanics further characterizes prediction in terms of the information about the future retained in observable histories [11, 12]. Finally, machine-learning studies demonstrate that neural models can learn local transition operators and accurately reproduce short-term dynamics, while long-horizon forecasting remains substantially more difficult [15–18].

Taken together, these studies offer three complementary perspectives on predictability. One emphasizes the computational structure of dynamical systems through concepts such as computational irreducibility, coarse-graining and algorithmic complexity. Another characterizes prediction in terms of the organization of information, causal states and computational mechanics. A third investigates the empirical forecasting capabilities of finite learning systems. Although these perspectives illuminate different aspects of predictability, their empirical relationship remains only partially understood.

What remains insufficiently understood is how these perspectives are connected empirically under finite learned representations. Existing classification schemes describe cellular automata in terms of morphology, attractor structure or computational organization [7, 8], but they do not directly characterize how these differences manifest themselves across increasing prediction horizons. Conversely, studies of learned dynamical prediction usually emphasize forecasting accuracy or architectural performance [15, 18], without isolating how the organization of the underlying dynamics affects the accessibility of future-relevant dependencies.

Existing work on neural representations of cellular automata further shows that local transition rules can be learned effectively [18]. It does not, however, determine whether this local success translates into similar long-horizon behavior across rules with fundamentally different dynamical organization. In particular, generic explanations based only on recursive error accumulation or RF size do not establish whether prediction degradation should be abrupt, gradual, persistent or largely absent under otherwise comparable experimental conditions.

The present paper addresses this gap by introducing a common empirical framework for studying predictive accessibility. Within this framework, neural predictors serve as experimental probes, while representative ECAs provide a family of deterministic systems with diverse dynamical organization. Rather than asking whether a particular neural architecture achieves the highest forecasting accuracy, we

ask whether different forms of dynamical organization give rise to qualitatively distinct forecasting regimes under a common experimental protocol. By keeping the observational and learning protocol fixed across rules—including the training procedure, RF, temporal history, model capacity and evaluation method—differences in forecasting behavior can be attributed to the organization of the underlying dynamics rather than to changes in the prediction method.

Predictive accessibility is introduced as an operational concept that characterizes the extent to which future-relevant dependencies remain empirically recoverable by neural predictors operating under this common experimental framework. It does not measure whether prediction is possible in principle, nor whether a formal computational shortcut exists. Instead, it characterizes how much predictive information remains accessible to a learned predictor under identical experimental conditions.

The goal is therefore neither to prove computational irreducibility [2, 6] nor to propose a new classification of cellular automata. Rather, the contribution is to connect established perspectives on cellular automaton dynamics, computational reduction and learned prediction within a unified experimental framework. Under identical observational and learning conditions, we show that representative ECAs exhibit qualitatively distinct regimes of predictive accessibility, including persistent predictability, stable predictability, abrupt accessibility loss, gradual degradation, persistent near-chance performance and sustained structured accessibility.

3. Methods

3.1 Experimental Rationale

The objective of the experimental design is not to compare neural architectures or maximize forecasting performance. Rather, neural predictors are used as experimental instruments for investigating predictive accessibility under a common forecasting protocol.

ECAs provide an unusually well-controlled environment for this purpose. Because the governing rule is known exactly, the dynamics are fully deterministic, and observational noise is absent, differences in forecasting behavior can be attributed more directly to differences in the organization of the underlying dynamics than to uncertainty about the system, noisy observations or model misspecification.

Accordingly, the experimental protocol is designed so that the principal source of variation is the underlying ECA rule, while the forecasting task, training procedure, evaluation metrics and model configurations remain comparable across experiments.

https://doi.org/10.25088/ComplexSystems.35.3.267

3.2 Elementary Cellular Automata

Representative ECAs are used as a family of minimal deterministic dynamical systems exhibiting qualitatively different forms of information organization.

We consider ECAs, which are one-dimensional binary dynamical systems defined on a lattice of N cells. Each cell $s_i(t) \in \{0, 1\}$ evolves synchronously according to a local update rule that depends on its nearest neighbors:

$$s_i(t+1) = f(s_{i-1}(t), s_i(t), s_{i+1}(t)), \tag{1}$$

where f is a Boolean function defined by one of the 256 possible ECA rules. Periodic boundary conditions are used throughout.

We analyze representative ECAs spanning multiple dynamical regimes commonly discussed in the cellular automata literature, including ordered, periodic, chaotic, additive and complex behavior.

Rule	Binary Code	Wolfram Class	Dynamical Type	Predictive Regime
0	00000000	I	fixed-point dynamics with trivial contraction	persistent predictability
4	00000100	II	simple periodic and locally repetitive structure	stable predictability
30	00011110	III	chaotic nonlinear local mixing	abrupt accessibility loss
45	00101101	III	distributed nonlinear dynamics	gradual degradation
90	01011010	III	additive XOR dynamics over $\mathbb{F}_2$	near-chance prediction
110	01101110	IV	complex dynamics with persistent coherent structures	sustained accessibility

Table 1. Representative ECAs analyzed in this paper.

The objective is not to propose a new classification of ECAs, but to investigate how different forms of dynamical organization affect predictive accessibility under a common experimental protocol.

Among these cases, rule 90 plays a particularly important role because, despite its deterministic and fully additive structure over $\mathbb{F}_2$, its dependencies become distributed across expanding causal cones under iteration. Rule 90 is defined by

$$s_i(t+1) = s_{i-1}(t) \oplus s_{i+1}(t), \tag{2}$$

where $\oplus$ denotes XOR over $\mathbb{F}_2$.

3.3 Data Generation

For each rule, we generate independent trajectories by sampling random initial conditions. Each trajectory consists of a binary configuration of length $N = 256$ evolved for $T = 120$ time steps. Initial states are sampled i.i.d. from a Bernoulli distribution with parameter $p = 0.5$.

For each experiment, we generate 1000 trajectories and split them into training, validation and test sets with sizes (700, 150, 150), respectively. In some experiments, we use smaller subsets (e.g., 200 trajectories with splits (140, 30, 30)) to reduce computational cost while preserving qualitative behavior.

To assess robustness with respect to stochastic effects, all experiments are repeated across multiple random seeds. Each seed defines both the sampling of initial conditions and the stochastic components of model training. Reported results correspond to either a representative seed or averages over multiple seeds, depending on the experiment.

3.4 Forecasting Task

We consider a supervised multi-step prediction task. Given a sequence of past configurations, the goal is to predict the configuration at a future time horizon τ.

The forecasting task is designed to operationalize predictive accessibility. Rather than evaluating one-step rule learning, the goal is to determine how much information about a future configuration remains recoverable from a bounded observation of the past as the prediction horizon increases.

Formally, for each trajectory, we construct input–output pairs of the form:

$$\mathbf{x} = (s(t - L + 1), \ldots, s(t)), \quad \mathbf{y} = s(t + \tau), \tag{3}$$

where L is the input length and τ is the prediction horizon. The input $\mathbf{x}$ is a tensor of shape (L, N) and the target $\mathbf{y}$ is a binary vector of length N.

The dataset is constructed by sliding a window over each trajectory, generating all valid pairs such that both the input window and the prediction target are contained within the trajectory.

3.5 Model Architectures

Multiple predictor families are included to assess whether the observed forecasting regimes primarily reflect differences in the underlying dynamics rather than properties specific to a particular neural architecture.

The selected predictor families span complementary inductive biases for learning deterministic dynamics. A convolutional neural

network (CNN) exploits local spatial structure through finite RFs, multilayer perceptrons (MLPs) provide global but unstructured representations, recurrent networks model temporal dependencies explicitly, and ESNs offer a reservoir-based approach to sequence processing. Together, these architectures allow us to assess whether the observed predictive regimes are robust across substantially different representational mechanisms.

We use the CNN as the primary architecture because its RF can be controlled explicitly through network depth, allowing us to investigate how spatial access to information influences predictive accessibility.

The CNN operates on inputs of shape (L, N) interpreted as L input channels over a spatial dimension of size N. It consists of a stack of one-dimensional convolutional layers with kernel size $k = 5$, stride 1 and same padding. The number of layers (depth) is varied across experiments, and each hidden layer has 64 channels with ReLU activations. A final 1×1 convolution produces the output. The effective RF is given by

$$\mathrm{RF} = 1 + d \cdot (k - 1), \tag{4}$$

where d is the network depth.

The multilayer perceptron (MLP) takes as input a flattened vector of size $L \cdot N$ and consists of three hidden layers of dimension 512, each followed by ReLU activation, and a final linear layer producing an output of size N. This architecture has global access to the input but no explicit spatial structure.

The long short-term memory (LSTM) processes the input sequence of length L with input size N. It consists of two stacked LSTM layers with hidden size 256, followed by a linear projection to an output of size N based on the final hidden state.

The echo state network (ESN) consists of a fixed random recurrent reservoir with tanh activation and a linear readout trained via ridge regression. The reservoir is driven by the input sequence, and the readout produces the predicted configuration.

Model sizes are kept fixed across rules to enable fair comparison, while the CNN depth is varied to control the RF.

3.6 Training Procedure

Gradient-based models (CNN, MLP and LSTM) were trained using standard stochastic optimization procedures with binary cross-entropy loss. Training budgets and model sizes were kept comparable across rules to enable qualitative comparison between dynamical regimes. Additional implementation details and the code used in the experiments are publicly available at gitlab.com/rodrigogarridohidalgo/eca-predictive-accessibility; see the code availability statement in the Acknowledgments.

For the ESN, the reservoir remained fixed and only the linear readout was trained using ridge regression.

3.7 Evaluation Metrics

Forecasting performance provides the primary empirical indicator of predictive accessibility. We therefore evaluate forecasting performance using two complementary prediction metrics:

- *Binary cross-entropy* (BCE): measures probabilistic prediction quality.
- *Hamming accuracy*: fraction of correctly predicted cells.

For reference, a random predictor yields BCE $\log 2 \approx 0.693$ and accuracy 0.5.

3.8 Structural Descriptors of Predictive Accessibility

Predictive accessibility is treated here as a latent property of the interaction between a dynamical system and a forecasting protocol. Because no single observable fully characterizes this property, we complement forecasting performance with several structural descriptors that quantify how predictive information propagates, disperses and remains empirically recoverable across increasing prediction horizons.

Each metric captures a different aspect of predictive accessibility: mutual information characterizes the persistence of predictive information, causal cone accessibility quantifies geometric access to causally relevant variables, compressibility measures the presence of algorithmically exploitable regularities, and parity rank growth characterizes the emergence of distributed algebraic dependencies.

Specifically, we analyze:

1. Mutual information decay, which measures how much information about future configurations remains recoverable from present states.
2. Causal cone accessibility, which compares the RF of the predictor with the spatial extent of the causal cone.
3. Compressibility, which quantifies the degree of effective structural regularity in the generated configurations.
4. Parity rank growth, which characterizes the expansion of distributed parity dependencies in additive rules such as rule 90.

These metrics are intended as structural descriptors of the organization of the dynamics rather than as alternative forecasting scores.

3.8.1 Mutual Information Decay

Let X_t denote the configuration of the automaton at time t, and let $X_{t+\tau}$ denote the configuration after a prediction horizon τ.

https://doi.org/10.25088/ComplexSystems.35.3.267

The mutual information between present and future states is defined as

$$I(X_t; X_{t+\tau}) = H(X_{t+\tau}) - H(X_{t+\tau} \mid X_t), \tag{5}$$

where

$$H(X) = -\sum_x p(x) \log p(x) \tag{6}$$

is the Shannon entropy of the random variable X, and

$$H(X \mid Y) = -\sum_{x,y} p(x, y) \log p(x \mid y) \tag{7}$$

is the conditional entropy.

Here:

- X_t and $X_{t+\tau}$ are random variables representing automaton configurations,
- x and y denote particular realizations of those variables,
- $p(x)$ denotes the probability of observing configuration x,
- $p(x, y)$ denotes the joint probability distribution.

Mutual information quantifies how much uncertainty about the future state is reduced by knowledge of the present state. As τ increases, a decay in $I(X_t; X_{t+\tau})$ indicates that predictive information becomes progressively less accessible.

3.8.2 Causal Cone Accessibility

In ECAs with neighborhood radius $r = 1$, the state of a cell after τ time steps depends only on cells contained within a causal cone of width

$$C(\tau) = 2\tau + 1. \tag{8}$$

This quantity defines the minimal spatial support required for exact prediction at horizon τ.

For a convolutional neural network predictor with depth d and kernel size k, the RF is approximately

$$\mathrm{RF} = 1 + d(k - 1). \tag{9}$$

We define the causal accessibility ratio as

$$\rho(\tau) = \frac{\mathrm{RF}}{C(\tau)}. \tag{10}$$

Values $\rho(\tau) < 1$ indicate that the predictor does not have access to the full causal support required for prediction.

Access to the full causal cone is not necessarily sufficient for accurate prediction, since the relevant dependencies may remain difficult to represent under finite predictors.

3.8.3 Compressibility

To quantify the effective structural organization of the generated configurations, we measure their compressibility using lossless compression algorithms.

Let S denote a binary spacetime configuration generated by the automaton. Its compressibility is defined as

$$\kappa(S) = \frac{|C(S)|}{|S|}, \tag{11}$$

where $|S|$ denotes the size of the original representation and $|C(S)|$ denotes the size after lossless compression.

Low values of $\kappa(S)$ indicate the presence of regularities, redundancies or repeated structures, whereas values approaching unity indicate low compressibility and effective statistical complexity.

Compressibility captures higher-order and distributed regularities that may not be visible through local correlation measures alone.

3.8.4 Parity Rank Growth

Certain ECAs, such as rule 90, are additive systems that evolve linearly over the finite field $\mathbb{F}_2$.

Rule 90 evolves according to

$$x_i^{t+1} = x_{i-1}^t \oplus x_{i+1}^t, \tag{12}$$

where $\oplus$ denotes addition modulo 2 (XOR).

After τ iterations, the state of a cell can be expressed as a parity function over an expanding subset of cells within the causal cone:

$$x_i^{t+\tau} = \bigoplus_{j \in A_\tau} x_{i+j}^t, \tag{13}$$

where A_τ denotes the set of contributing positions.

As the prediction horizon increases, future states depend on increasingly distributed XOR combinations of initial cells. To characterize this algebraic expansion, we analyze the binary linear operator associated with rule 90 over the finite field $\mathbb{F}_2$.

Let M_τ denote the linear transformation mapping the initial configuration to the configuration at horizon τ:

$$x^{t+\tau} = M_\tau x^t. \tag{14}$$

Each row of M_τ specifies which initial cells contribute, through XOR operations, to the value of a future cell.

We define the parity rank as

$$R(\tau) = \operatorname{rank}_{\mathbb{F}_2}(M_\tau). \tag{15}$$

Growth in $R(\tau)$ indicates that prediction increasingly depends on independent high-order parity relations distributed across the causal cone, rather than on simple local correlations.

https://doi.org/10.25088/ComplexSystems.35.3.267

Such distributed parity dependencies are known to be difficult for local statistical predictors and standard neural architectures to represent efficiently.

3.9 Experimental Design

We study prediction across multiple horizons $\tau \in \{1, 2, 5, 6, 7, 8, 9, 10, 11, 20\}$ and vary model capacity, particularly the RF in CNNs. Horizons were selected to cover early, intermediate and long-term regimes while limiting computational cost.

For key experiments, we analyze the dependence of predictive performance on RF and horizon, as well as robustness across random seeds.

To investigate whether prediction failure is due to insufficient temporal information, we also vary the input length L in selected experiments. Figures 5–10 show results for a single representative seed; Table 2 reports means and standard deviations over three seeds for some selected rules, RF and horizon configurations.

4. Results

4.1 Distinct Regimes of Predictive Accessibility

The experiments show that predictive accessibility is not a uniform property of deterministic cellular automata. Under a common forecasting protocol, representative rules exhibit qualitatively distinct patterns of forecasting behavior that cannot be explained solely by whether the underlying dynamics are deterministic.

Ordered and periodic rules provide a useful baseline for interpreting the remaining cases. Rule 0 remains trivially predictable across all tested horizons, while rule 4 preserves stable forecasting performance under relatively small RFs. In contrast, chaotic, additive and complex rules exhibit markedly different regimes of predictive accessibility, including abrupt accessibility loss, progressive degradation, persistently low accessibility and sustained accessibility over substantially longer prediction horizons.

These observations indicate that deterministic evolution alone is insufficient to characterize predictive accessibility. Instead, the structural organization of the underlying dynamics strongly influences whether predictive dependencies remain recoverable within a finite observational window. The following sections examine these distinct accessibility regimes in representative rules and relate them to complementary structural descriptors, including mutual information, causal accessibility, compressibility and parity structure.

4.2 Information-Structural Mechanisms

The information-structural analyses indicate that reductions in predictive accessibility are not simply associated with a loss of deterministic information. Rather, future-relevant information becomes progressively distributed across larger causal structures and increasingly difficult to recover under finite predictive representations.

4.2.1 Scaling of Predictive Information Accessibility

Mutual information provides a complementary perspective on predictive accessibility by quantifying how much information contained within local spatial neighborhoods remains relevant for predicting future cell states as the prediction horizon increases.

For a spatial block of radius r, we evaluate

$$I\left(X_t^{(r)}; x_i^{t+\tau}\right), \tag{16}$$

where $X_t^{(r)}$ denotes the local neighborhood centered at cell i with width $2r + 1$.

Figure 1 shows that predictive information does not simply disappear as the prediction horizon increases. Instead, the spatial support required to recover that information grows systematically with τ.

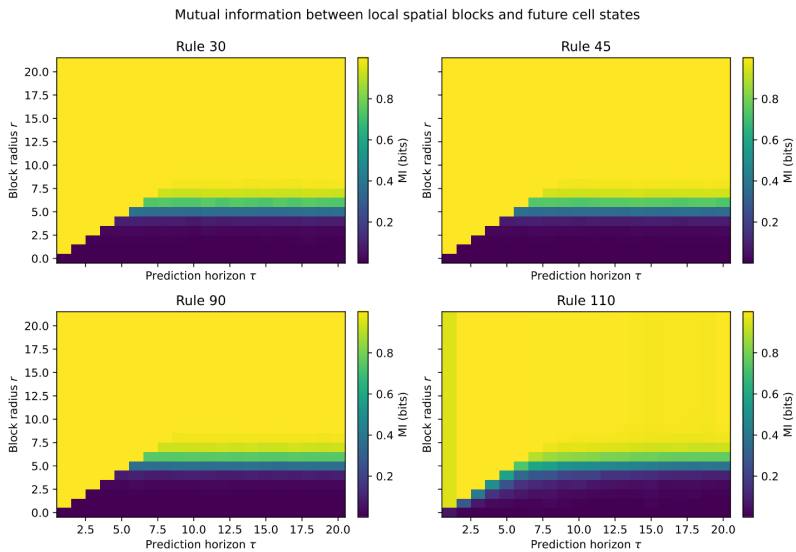


Figure 1. Mutual information between local spatial blocks and future cell states as a function of prediction horizon and block radius. Predictive information remains near zero when the observed block is smaller than the causal support required for prediction, but rises sharply once the block reaches the relevant causal scale.

For rules 30, 45 and 90, small local neighborhoods contain almost no predictive information beyond short horizons. However, once the observed block reaches the corresponding causal scale, mutual information rapidly approaches unity, indicating that the future state remains deterministically recoverable.

Rule 110 exhibits a qualitatively different profile, retaining partial predictive accessibility even for smaller neighborhoods. This behavior is consistent with the smoother degradation and sustained predictability observed in the forecasting experiments.

These results indicate that the degradation of forecasting performance is not necessarily associated with a loss of deterministic information. Rather, future-relevant information becomes progressively distributed across larger spatial supports, making it increasingly difficult to recover under predictors with finite RFs and representational capacity.

4.2.2 Causal Cone Accessibility and Receptive Field Limitations

For radius-1 cellular automata, the spatial support required for exact prediction grows linearly with the prediction horizon. To quantify how much of this causal support remains accessible to the predictor, we define the causal accessibility ratio

$$\rho(\tau) = \frac{\mathrm{RF}}{2\tau + 1}. \tag{17}$$

Figure 2 illustrates the causal accessibility ratio as a function of prediction horizon for different RFs.

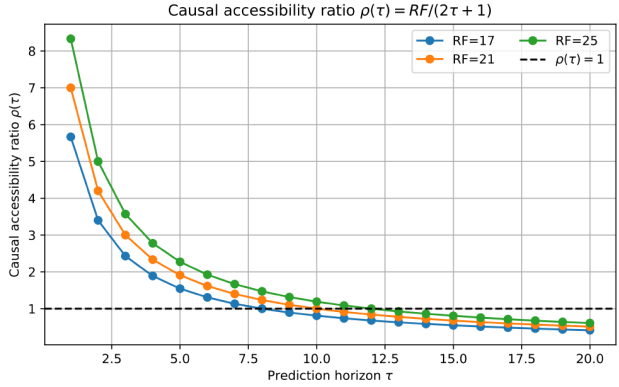


Figure 2. Causal accessibility ratio $\rho(\tau) = \mathrm{RF}/(2\tau + 1)$ as a function of prediction horizon τ for different RFs. The dashed line indicates the critical regime $\rho(\tau) = 1$, where the RF approximately matches the width of the causal cone. As the horizon increases, the fraction of causally accessible information decreases systematically for fixed local predictors.

As the horizon increases, the causal support required for exact prediction expands linearly, while the RF of the predictor remains fixed. Consequently, the fraction of causally accessible information decreases systematically with τ.

The regime $\rho(\tau) \approx 1$ corresponds to the point at which the RF approximately matches the width of the causal cone. Beyond this regime, part of the dynamically relevant information necessarily lies outside the spatial context available to the predictor.

Importantly, however, causal accessibility alone is not sufficient to guarantee accurate prediction. While rules 30 and 45 exhibit strong sensitivity to the accessibility ratio, rule 90 remains difficult even when the RF fully covers the causal cone. This indicates that accessibility and representability constitute distinct constraints on predictability.

4.2.3 Compressibility and Effective Structural Complexity

Compressibility provides a complementary characterization of predictive accessibility by quantifying the degree of effective structural organization present in the generated spacetime configurations.

To estimate compressibility, we apply lossless compression to the binary spacetime trajectories generated by each rule and compute the compression ratio

$$\kappa(S) = \frac{|C(S)|}{|S|}, \tag{18}$$

where $|S|$ denotes the size of the original configuration and $|C(S)|$ the size after compression.

Figure 3 reveals substantial differences in effective structural complexity across dynamical regimes.

Ordered dynamics such as rule 0 remain extremely compressible throughout the evolution, reflecting the rapid convergence toward homogeneous configurations. Periodic dynamics such as rule 4 also retain high compressibility due to the persistence of repetitive spatial structure.

In contrast, rules 30 and 45 exhibit substantially lower compressibility, indicating that the generated configurations contain fewer statistical redundancies and less easily compressible large-scale organization. These results are consistent with the rapid degradation of predictive accessibility observed in the forecasting experiments.

Rule 110 occupies an intermediate regime. Although its dynamics are visually complex, the resulting trajectories remain significantly more compressible than those of rules 30 and 45. This suggests that dynamically relevant structure persists over longer temporal scales, helping explain the sustained predictability observed even at large horizons.

https://doi.org/10.25088/ComplexSystems.35.3.267

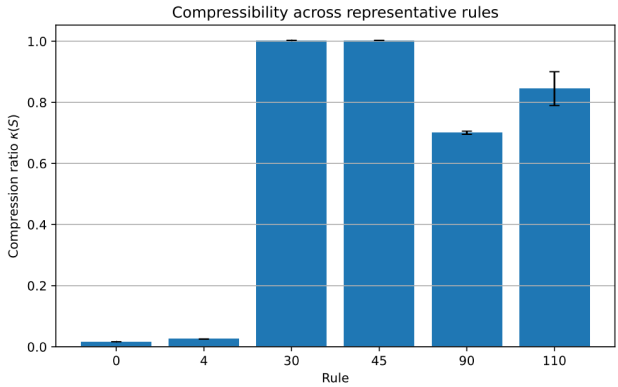


Figure 3. Compression ratios for different ECA rules. Ordered and periodic rules remain highly compressible, whereas chaotic dynamics exhibit substantially lower compressibility. Rule 110 retains intermediate compressibility, indicating the persistence of structured organization despite its complex behavior.

Importantly, compressibility captures forms of higher-order and distributed organization that are not necessarily visible through local correlation measures alone. The results therefore support the interpretation that predictability depends not only on local causal accessibility, but also on the existence of compact and persistent structural representations within the evolving dynamics.

4.2.4 Distributed Parity Dependencies in Rule 90

Rule 90 exhibits a qualitatively distinct form of predictive difficulty associated with the algebraic organization of its causal dependencies.

Unlike rules 30 and 45, rule 90 is additive over the finite field $\mathbb{F}_2$, evolving according to XOR operations distributed across the causal cone. Although the rule is fully deterministic and algebraically simple, iterative evolution progressively redistributes predictive dependencies across increasingly nonlocal parity combinations.

To characterize this effect, we analyze the parity support generated under temporal iteration. For each prediction horizon τ, we compute the number of initial cells contributing through XOR operations to the future state of a target cell.

Figure 4 shows that the parity support exhibits highly non-monotonic behavior across prediction horizons. Rather than increasing smoothly, the number of contributing XOR terms fluctuates due to repeated cancellations over $\mathbb{F}_2$. These cancellations depend on the binary structure of τ and are closely related to the Sierpinski-like fractal organization characteristic of rule 90.

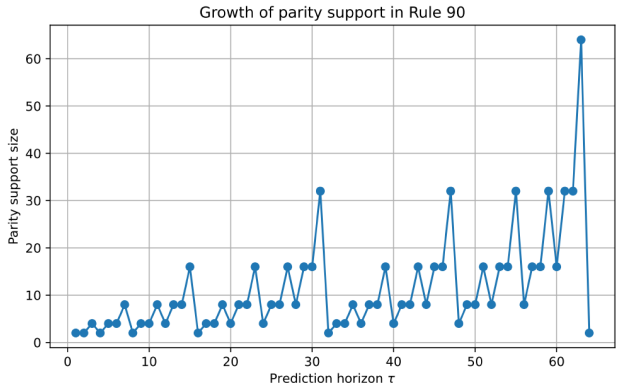


Figure 4. Growth of parity support in rule 90 as a function of prediction horizon. Future states depend on increasingly distributed XOR combinations of initial cells, generating expanding nonlocal parity dependencies across the causal cone.

Despite these fluctuations, the predictive dependencies remain distributed across increasingly nonlocal regions of the causal cone, generating parity relations that are difficult to represent through finite local statistical models.

This behavior is particularly important because XOR-like parity relations are known to be difficult for many statistical and neural predictors to represent efficiently, especially when the dependencies become increasingly distributed and high order.

Unlike chaotic rules, therefore, the failure on rule 90 is not associated with the disappearance of predictive information. Indeed, the mutual information analysis shows that the future state remains fully determined once the relevant causal support is observed. Instead, the difficulty arises because the relevant information is organized through distributed parity relations that are poorly captured by finite local representations and standard learning architectures.

This result highlights an important distinction between causal accessibility and representational accessibility. Even when the RF fully covers the causal cone, predictive performance may remain near chance level if the underlying dependency structure is algebraically distributed rather than locally compressible.

4.3 Rule 30: Abrupt Predictability Collapse

Rule 30 exhibits the clearest example of an abrupt architecture-dependent transition (Figure 5). Using convolutional predictors with kernel size five and varying depth, we observe a sharp boundary between highly accurate forecasting and complete collapse to chance level.

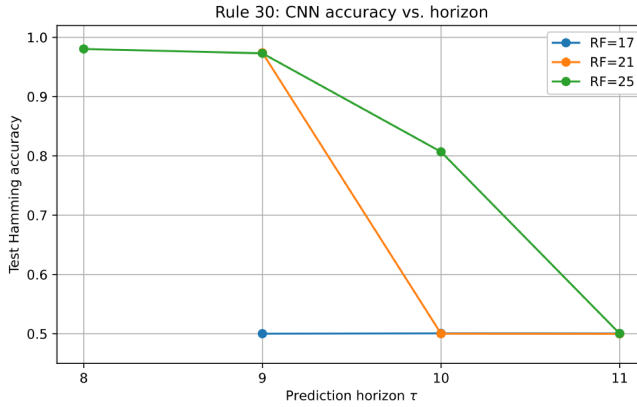


Figure 5. CNN forecasting accuracy for rule 30 as a function of prediction horizon τ for different RFs. Prediction remains highly accurate up to an RF-dependent threshold, after which performance collapses abruptly to chance level. This rule provides the clearest example of a sharp RF-dependent predictability transition.

For a fixed prediction horizon τ, increasing the RF extends the range of accurate forecasting. However, once the relevant causal dependencies exceed the accessible spatial context, predictive performance drops rapidly to chance-level accuracy, with binary cross-entropy approaching log2.

The sharpness of this degradation suggests a regime in which predictive accessibility is strongly constrained by the relation between horizon and RF. In this setting, accurate local rule learning does not guarantee sustained multi-step predictability once dynamically relevant information propagates beyond the effective representation available to the model.

4.4 Rule 45: Smoother Receptive-Field Dependence

Rule 45 also shows strong sensitivity to RF, but with a qualitatively different transition profile (Figure 6). Increasing RF systematically extends the prediction horizon, but the degradation in performance is smoother than in rule 30.

In contrast to rule 30, the degradation in predictive performance is more progressive. Larger RFs still improve forecasting accuracy, but the loss of performance occurs over a broader range of horizons, producing an extended intermediate regime of partial predictability.

This behavior indicates that predictive accessibility degrades more gradually in rule 45, suggesting a dynamical organization in which relevant information remains partially recoverable even as the horizon increases.

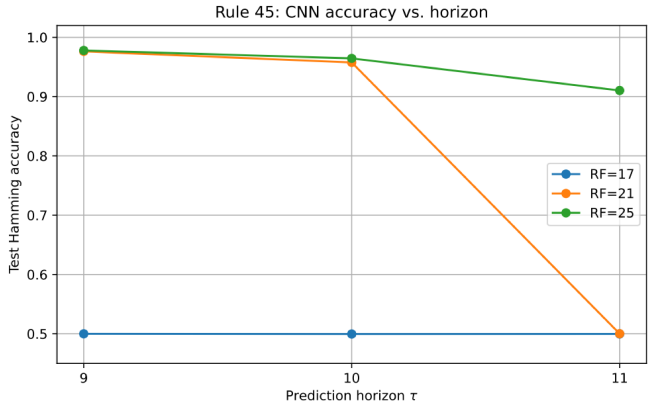


Figure 6. CNN forecasting accuracy for rule 45 as a function of prediction horizon τ for different RFs. Increasing RF extends the prediction horizon, but the transition is smoother than in rule 30. In particular, the largest RF retains substantial predictive power even at longer horizons, indicating gradual rather than abrupt degradation.

4.5 Rule 90: Failure across Architectures

Rule 90 behaves fundamentally differently from rules 30 and 45 (Figure 7). Across all tested RFs and prediction horizons, convolutional models fail to achieve performance above chance level.

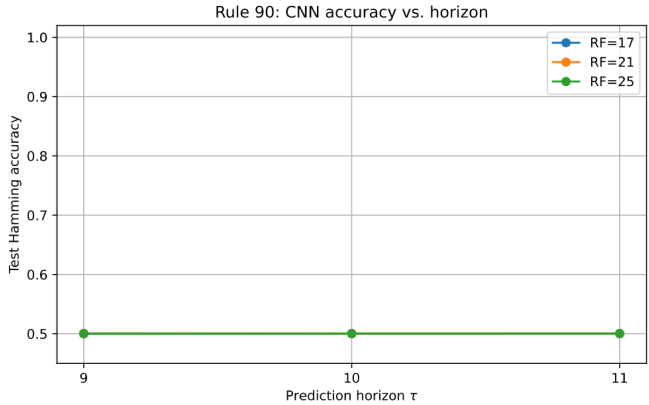


Figure 7. CNN forecasting accuracy for rule 90 as a function of prediction horizon τ for different RFs. Rules 30 and 45 exhibit measurable predictive structure, whereas performance here remains at chance level across all tested horizons and RFs. This indicates that increasing local spatial context alone does not recover predictability for this rule.

To determine whether this failure is specific to convolutional locality, we evaluated additional model classes on rule 90, including a

multilayer perceptron (MLP), an LSTM and an echo state network (ESN) (Figure 8). All models remain at chance-level accuracy, indicating that the failure is not architecture specific.

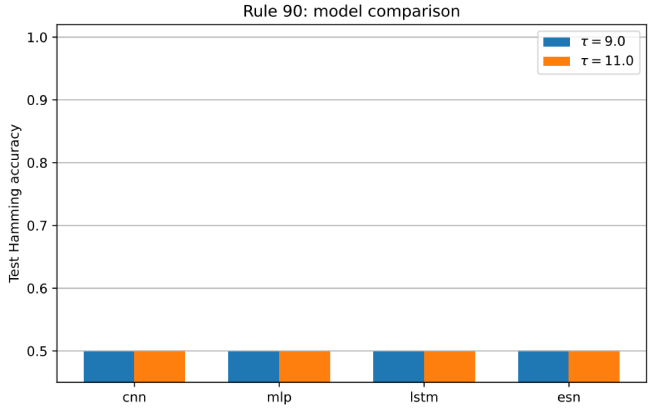


Figure 8. Comparison of forecasting performance across model classes for rule 90 at horizons $\tau = 9$ and $\tau = 11$. Convolutional, fully connected, recurrent and reservoir-based predictors all remain at chance level, indicating that failure on rule 90 is not specific to convolutional locality.

We further tested whether increasing the temporal input length could improve performance. As shown in Figure 9, providing longer input sequences does not lead to any measurable improvement. Across all tested architectures, RFs and input lengths, predictive performance remains near chance level. Increasing spatial context, temporal history or model capacity does not recover substantial predictive structure.

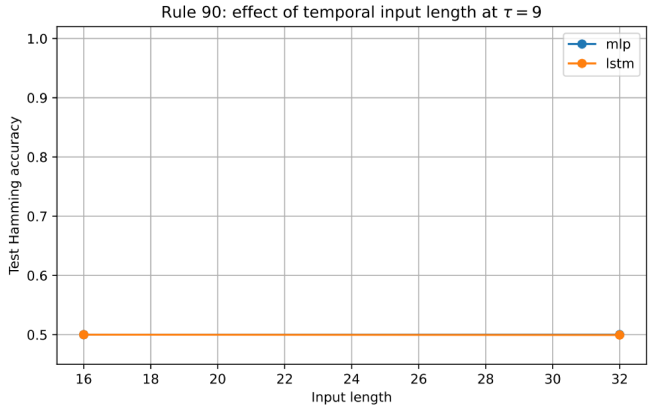


Figure 9. Effect of temporal input length on forecasting performance for rule 90 at $\tau = 9$. Increasing the input length does not improve performance for either MLP or LSTM predictors, indicating that the failure on rule 90 is not rescued by providing additional temporal context.

This behavior is particularly notable because rule 90 is fully deterministic and additive over $\mathbb{F}_2$. Under iteration, however, dependencies become distributed across an expanding causal cone through XOR-like combinations, generating parity relations that are difficult to compress into finite learned representations. The results therefore suggest a regime of limited predictive accessibility in which relevant information remains dynamically distributed despite the simplicity of the underlying rule.

4.6 Rule 110: Sustained Predictability

Rule 110 exhibits yet another qualitatively distinct regime (Figure 10). Unlike the previous cases, prediction remains highly accurate across all tested horizons and RFs.

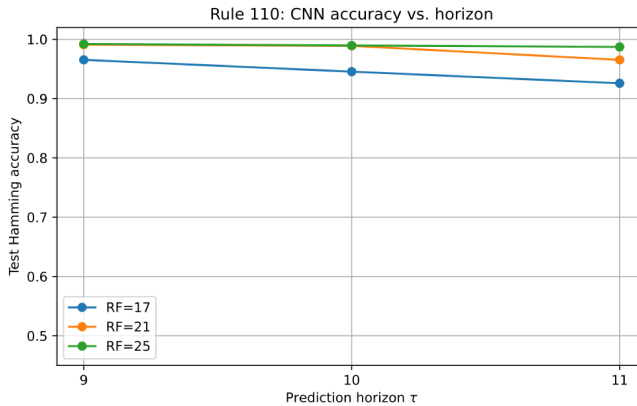


Figure 10. CNN forecasting accuracy for rule 110 as a function of prediction horizon τ for different RFs. Unlike rules 30, 45 and 90, rule 110 remains highly predictable even at the longest tested horizons. Larger RFs improve performance, but no abrupt predictability boundary is observed.

Rule 110 exhibits sustained predictive performance across the explored horizons. While larger RFs improve accuracy, no abrupt degradation is observed within the tested range.

Unlike the previous cases, the dynamics retain accessible structured information over extended temporal scales. This indicates that complex behavior does not necessarily imply immediate degradation of forecasting performance under finite observational windows, but depends on how dynamically relevant information is organized and propagated.

Importantly, the present results concern statistical prediction of local configurations rather than exact prediction of emergent computational structures or particle interactions in the full rule 110

dynamics. Consequently, sustained forecasting performance within the explored horizons should not be interpreted as evidence against the known computational complexity or universality properties of the rule.

4.7 Qualitative Consistency across Representative Random Seeds

To verify that the observed behaviors are not artifacts of a particular initialization, we repeated representative experiments across a small number of random seeds (Table 2). The main qualitative regimes remain consistent across the explored seeds, including high-performance regimes, abrupt degradation and near-chance prediction behavior.

Rule	RF	τ	Accuracy (mean ± std)	BCE (mean ± std)
30	17	9	0.5002 ± 0.0004	0.6932 ± 0.0000
30	17	11	0.5005 ± 0.0004	0.6931 ± 0.0000
30	25	9	0.9733 ± 0.0028	0.0540 ± 0.0135
30	25	11	0.5004 ± 0.0002	0.6931 ± 0.0000
45	17	9	0.4999 ± 0.0003	0.6932 ± 0.0000
45	17	11	0.4997 ± 0.0005	0.6932 ± 0.0000
45	25	9	0.9779 ± 0.0000	0.0371 ± 0.0011
45	25	11	0.9104 ± 0.0127	0.2049 ± 0.0266
90	17	9	0.5006 ± 0.0003	0.6931 ± 0.0000
90	17	11	0.5004 ± 0.0002	0.6932 ± 0.0000
90	25	9	0.4999 ± 0.0005	0.6931 ± 0.0000
90	25	11	0.5003 ± 0.0003	0.6931 ± 0.0000
110	17	9	0.9654 ± 0.0009	0.0760 ± 0.0018
110	17	11	0.9259 ± 0.0027	0.1689 ± 0.0066
110	25	9	0.9921 ± 0.0004	0.0181 ± 0.0008
110	25	11	0.9872 ± 0.0003	0.0315 ± 0.0005

Table 2. Robustness of CNN forecasting performance across random seeds. Mean and standard deviation are computed over three seeds.

Importantly, while the figures suggest relatively flat behavior within each regime, the numerical values in Table 2 reveal small but systematic differences across RFs and prediction horizons. In particular, configurations near the transition boundary (e.g., rule 45 at larger RFs) exhibit measurable degradation in accuracy and increases in loss, even when this variation is not visually prominent.

In the collapsed regime, accuracy tightly concentrates around chance level (≈ 0.5) and loss around $\log 2$, reinforcing the interpretation of an effectively unpredictable phase. Conversely, in the

predictable regime, accuracy remains close to unity with low variance, indicating stable and reproducible performance.

This separation between visually flat regimes and quantitatively distinct values is consistent with a phase-transition-like behavior, where fluctuations within each regime are small compared to the discontinuity at the transition.

5. Discussion

The results suggest that prediction in local deterministic systems depends not only on model expressivity, but also on how dynamically relevant information remains accessible under finite representations.

In convolutional predictors, RF provides an operational notion of accessible causal context. As the prediction horizon increases, the effective dependencies required for accurate forecasting expand spatially. Prediction failure emerges when these dependencies are no longer sufficiently captured within the available representation.

However, the experiments also show that RF alone does not fully determine predictive behavior. Different rules exhibit qualitatively distinct forms of degradation despite being governed by equally local deterministic dynamics. This indicates that predictability depends not only on the size of the causal region, but also on how information is distributed within it.

Rule 90 is particularly illustrative in this respect. Although its evolution is fully deterministic and additive over $\mathbb{F}_2$, the relevant dependencies become distributed through parity-like combinations across expanding causal cones. Under the tested representations, these dependencies remain difficult to compress into stable predictive mappings, leading to persistent near-chance forecasting performance.

The observed differences between rules are broadly consistent with the view that distinct cellular automata organize predictive information in fundamentally different ways. Ordered and periodic rules preserve compact local dependencies, and additive rules distribute information through nonlocal parity-like structures, while complex rules may support persistent coherent structures that remain partially accessible over long temporal scales.

This interpretation aligns with broader ideas from computational mechanics and information-theoretic approaches to dynamical systems, in which prediction depends on the accessibility and reconstruction of dynamically relevant information rather than on determinism alone.

By contrast, rule 110 retains accessible structured information over substantially longer horizons. This suggests that complex dynamics do not necessarily imply immediate loss of predictability, but rather

that predictability depends on whether dynamically relevant information remains sufficiently organized to be exploited by finite learned representations.

These observations motivate interpreting predictability as an empirical property of the interaction between dynamics and representation. From this perspective, prediction limits arise not solely from stochasticity or model insufficiency, but from constraints on information accessibility and statistical compressibility under finite observational frameworks.

More broadly, the results are consistent with the idea that deterministic systems may exhibit practical limits of prediction even when their microscopic rules are fully known. Rather than establishing formal computational irreducibility, the present paper provides an empirical characterization of how different local dynamical rules generate distinct predictive regimes.

6. Conclusion

This paper investigated multi-step prediction in elementary cellular automata (ECAs) using neural predictors with controlled receptive fields (RFs). Rather than exhibiting a single universal form of prediction degradation, different rules generated qualitatively distinct regimes of predictive accessibility.

The results show that accurate local rule learning does not necessarily imply sustained long-horizon predictability. Instead, forecasting performance depends on how dynamically relevant information propagates and remains accessible under finite learned representations.

These findings suggest that predictability in deterministic systems should be understood not only as a property of the underlying dynamics, but also as a property of the interaction between dynamics, observation and representation. More generally, the study highlights how simple local systems can generate substantially different forms of predictive accessibility despite comparable microscopic definitions.

Acknowledgments

Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (CCSS210001). All data used in this study was generated synthetically from ECAs and is available upon request. The code used to generate the datasets, train the models and reproduce the figures is publicly available at gitlab.com/rodrigogarridohidalgo/eca-predictive-accessibility.

During the preparation of this work, the author used ChatGPT in order to translate from Spanish to English. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

[1] S. Wolfram, "Statistical Mechanics of Cellular Automata," *Reviews of Modern Physics*, **55**(3), 1983 pp. 601–644. doi:10.1103/RevModPhys.55.601.

[2] S. Wolfram, "Universality and Complexity in Cellular Automata," *Physica D: Nonlinear Phenomena*, **10**(1), 1984 pp. 1–35. doi:10.1016/0167-2789(84)90245-8.

[3] S. Wolfram, *A New Kind of Science*, Champaign, IL: Wolfram Media, Inc., 2002.

[4] N. Israeli and N. Goldenfeld, "Coarse-Graining of Cellular Automata, Emergence, and the Predictability of Complex Systems," *Physical Review E*, **73**(2), 2006 026203. doi:10.1103/PhysRevE.73.026203.

[5] H. Zenil, "Compression-Based Investigation of the Dynamical Properties of Cellular Automata and Other Systems," *Complex Systems*, **19**(1), 2010 pp. 1–28. doi:10.25088/ComplexSystems.19.1.1.

[6] H. Zenil, F. Soler-Toscano and J. J. Joosten, "Empirical Encounters with Computational Irreducibility and Unpredictability," *Minds and Machines*, **22**(3), 2012 pp. 149–165. doi:10.1007/s11023-011-9262-y.

[7] A. Wuensche, "Classifying Cellular Automata Automatically: Finding Gliders, Filtering, and Relating Space-Time Patterns, Attractor Basins, and the Z Parameter," *Complexity*, **4**(3), 1999 pp. 47–66. doi:10.1002/(SICI)1099-0526(199901/02)4:3%3C47::AID-CPLX9%3E3.0.CO;2-V.

[8] G. J. Martínez, "A Note on Elementary Cellular Automata Classification," *Journal of Cellular Automata*, **8**(3–4), 2013 pp. 233–259.

[9] R. Hosseinie and M. Mahzoon, "Irreducibility and Emergence in Complex Systems and the Quest for Alternative Insights," *Complexity*, **17**(2), 2011 pp. 10–18. doi:10.1002/cplx.20377.

[10] J. P. Crutchfield and M. Mitchell, "The Evolution of Emergent Computation," *Proceedings of the National Academy of Sciences*, **92**(23), 1995 pp. 10742–10746. doi:10.1073/pnas.92.23.10742.

[11] J. P. Crutchfield and K. Young, "Inferring Statistical Complexity," *Physical Review Letters*, **63**(2), 1989 pp. 105–108. doi:10.1103/PhysRevLett.63.105.

[12] C. R. Shalizi and J. P. Crutchfield, "Computational Mechanics: Pattern and Prediction, Structure and Simplicity," *Journal of Statistical Physics*, **104**(3), 2001 pp. 817–879. doi:10.1023/A:1010388907793.

[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, **9**(8), 1997 pp. 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[14] H. Jaeger, "The "Echo State" Approach to Analysing and Training Recurrent Neural Networks," Technical Report GMD Report 148, Bonn, Germany: German National Research Center for Information Technology, 2001. publica.fraunhofer.de/entities/publication/7d4a7eec-a22c-4df0-903d-93f9cd5aca02.

[15] J. Pathak, B. Hunt, M. Girvan, Z. Lu and E. Ott, "Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach," *Physical Review Letters*, **120**, 2018 024102. doi:10.1103/PhysRevLett.120.024102.

[16] S. B. Taieb, G. Bontempi, A. F. Atiya and A. Sorjamaa, "A Review and Comparison of Strategies for Multi-step Ahead Time Series Forecasting Based on the NN5 Forecasting Competition," *Expert Systems with Applications*, **39**(8), 2012 pp. 7067–7083. doi:10.1016/j.eswa.2012.01.039.

[17] G. Bontempi, S. B. Taieb and Y.-A. Le Borgne, "Machine Learning Strategies for Time Series Forecasting," in *Business Intelligence: Second European Summer School (eBISS 2012)*, Brussels, Belgium (M.-A. Aufaure and E. Zimányi, eds.), Berlin, Heidelberg: Springer, 2013 pp. 62–77. doi:10.1007/978-3-642-36318-4_3.

[18] W. Gilpin, "Cellular Automata as Convolutional Neural Networks," *Physical Review E*, **100**(3), 2019 032402. doi:10.1103/PhysRevE.100.032402.