The *Next Motif*: Tapping into Recurrence Dynamics and Precursor Signals to Forecast Events of Interest

Miguel G. Silva[1,2], Sara C. Madeira[1], and Rui Henriques[2]

[1] LASIGE and Faculdade de Ciências, Universidade de Lisboa, Portugal
{mmgsilva,sacmadeira}@ciencias.ulisboa.pt
[2] INESC-ID and IST, Universidade de Lisboa, Portugal rmch@tecnico.ulisboa.pt

Abstract. Motifs in multivariate time series reveal critical non-periodic behaviors in biophysiological, geophysical, urban, and societal systems. This work takes motif analysis one step further by exploring the recurrence dynamics of multidimensional motifs to enhance the forecasting of events of interest, focusing on regressing the timing of a motif's next occurrence. This task is challenged by the inherent stochasticity of real-world system behaviors and heterogeneity of data inputs, combining the raw multivariate time series and available motif information. To address these challenges, two major hypotheses are established: i) that some irregular behaviors show indeed a form of less trivial temporal regularity, possibly described by a non-linear function; and ii) the occurrence of motifs in some systems can be anticipated by precursor signals, such as emerging traffic behaviors prior to congestion or subtle physiological patterns preceding health events. A novel methodology is proposed to expressively encode motif information and subsequently combine state-of-the-art neural processing principles to answer the target forecasting task. Experimental results demonstrate that it is feasible to accurately estimate the next occurrence of a given motif in arbitrarily complex tasks by leveraging the value embedded in the two proposed hypotheses. Furthermore, we show that augmenting the multivariate input series with motif-aware masking significantly enhances the predictive accuracy of recurrent and convolutional forecasters.

Keywords: Prediction · Multivariate Time series · Motif Discovery

1 Introduction

Motifs are recurring structures in data. Research on motif discovery has primarily focused on developing efficient and quasi-optimal search methods to identify motifs with unexpectedly high recurrence [48, 49], often overlooking the underlying regularities behind the recurrence phenomena. As recurrence is often non-periodic [21], understanding the latent patterns governing motif occurrences is crucial for knowledge acquisition, whether we are assessing the spatial distribution of motifs in network data or the temporal distribution of motifs in time series data. In this context, modeling the recurrence dynamics of a given motif can be used to establish a more comprehensive view of a system's behaviors and subsequently support descriptive and predictive analyses. In biophysiological systems, for instance, recognizing spatial regularities in brain activity or temporal patterns in cardiac events is crucial for predicting health states [4, 19].

This work focuses on motifs with irregular (non-periodic) occurrences along a multivariate time series and aims to forecast their next occurrence by learning recurrence dynamics. Specifically, given a multidimensional motif of interest, we address the rarely studied problem of predicting the timing of its next occurrence. This task is widely relevant, with applications ranging from forecasting social behaviors, organ activity, and traffic conditions to predicting stock movements, mechanical failures, and atmospheric or astrophysical events [24, 39].

To pursue the targeted task – predicting the next occurrence of a given motif – the work is grounded on two key premises that provide strong evidence for meaningful recurrence dynamics in irregular motifs. The first premise suggests that motif occurrences in many real-world systems are often preceded by precursor signals, which may be subtle but measurable [16]. This is commonly observed in various domains, where social, biophysiological, urban, machine, individual, ecosystemic, or financial behaviors are often preceded by identifiable interactions, symptoms, emerging mobility patterns, anomalies, habits, interconnected events, or market dynamics, respectively [15, 20, 26]. The second premise states that non-periodic behaviors in many real-world systems can be described by less trivial forms of recurrence, possibly by well-defined, yet non-linear, rules. Examples include emerging behaviors characterized by subexponential growth [5, 13, 27], as well as context-dependent motifs, such as mobility patterns during public events, utility consumption fluctuations on holidays, weather-driven user behaviors, and route-dependent vehicle dynamics [29, 31, 44].

A novel methodology for regressing the next occurrence time of a multidimensional motif is proposed by jointly leveraging these two key principles: modeling recurrence dynamics and capturing subtle precursor signals that may precede a motif. In this context, we present four major contributions:

1. formalization of the *next motif prediction* task in light of existing research, along with a comprehensive enumeration of its key requirements;
2. expressive mask-based encodings that consolidate raw time series data and motif information (e.g., pattern, occurrences) to aid the predictive learning;
3. a novel methodology grounded on neural processing principles to answer the targeted task, with a particular focus: i) on motif-aware segmentation strategies for predictive modeling; and ii) adequate architectural choices, encompassing feed-forward, recurrent, convolutional, and transformer paradigms;
4. empirical evaluation of the proposed methodology on both simulated and real-world datasets, assessing its effectiveness in capturing motif recurrence dynamics under varying stochastic behaviors, temporal misalignments, and precursor signal patterns.

The results confirm the feasibility of the next motif prediction task when either or both premises are present. Long short-term memory (LSTM) models excel in capturing non-trivial recurrence patterns governed by well-defined yet non-linear rules, while temporal convolutional networks (TCN) and Transformer architectures are more effective in leveraging subtle precursor signals for forecasting.

2 Problem formulation

Definition 1. *A time series* $X = < \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n >$ *is a sequence of observations of length* n*, where an observation at time step* t *is a feature vector* $\mathbf{x}_t = (x_t^1, \ldots, x_t^m)$ *with* m *denoting the number of variables (multivariate order). Each feature* x_t^j *is either numeric* $x_t^j \in \mathbb{R}$ *or symbolic* $x_t^j \in \Sigma_j$*, where* Σ_j *is a finite symbolic set.*

The discovery of local patterns in time series data, such as biclusters [6], anomalies [8], temporal rules [3], and motifs [42], is an active research field. Among these local regularities, motifs became one of the essential primitives in time series data mining, with several motif discovery algorithms proposed since their introduction [10, 41, 42] due to their key role for knowledge acquisition in diverse domains, such as biomedicine [19] and robotics [2].

Definition 2. *Given a time series* X *of length* n*, a* ***subsequence*** $X_{i,s,J}$ *is defined by its starting position* i*, length* s*, and a subset of variables* $J \subseteq \mathcal{Y}$*, where* $\mathcal{Y}$ *is the original set of* m *variables. If the subset* J *contains two or more variables (i.e.,* $|J| \geq 2$*), the subsequence is said to be* ***multidimensional****. Formally, a subsequence is defined as* $X_{i,s,J} = < \mathbf{x}_i, \ldots, \mathbf{x}_{i+s-1} >$*, where each observation* $\mathbf{x}_i = \left(x_i^j\right)_{Y_j \in J}$ *contains only the features from the* J *subset of variables.*

Definition 3. *The* ***motif discovery*** *task in a time series* X *aims to find motifs, where a motif* M *is a subsequence satisfying predefined criteria, classically:*

- *a minimum recurrence criterion* $r_{\min}$*, which specifies that the motif must appear at least* $r_{\min}$ *times in the time series.*
- *a maximum distance threshold* $d_{\max}$*, such that the dissimilarity between two motif subsequences must not exceed* $d_{\max}$ *in at least* q *of the* m *dimensions.*

Formally, a motif M *is a subsequence* $X_{i,s,J}$ *that appears* $r \geq r_{\min}$ *times, with a pairwise distance between occurrences satisfying* $d(M_i, M_j) \leq d_{\max}$ *according to a distance function* d*, in at least* q *of* m *dimensions. Let* $\mathcal{O}_M = \{t_1, t_2, \ldots, t_r\}$ *be the set of starting positions where* M *occurs in* X*.*

Time series motifs are approximately recurring *subsequences* along a longer time series with unexpected frequency [41]. A match is recognized when the distance between two subsequences is below a predetermined distance threshold $d_{\max}$. Their recurrence can be concurrently observed across all time series variables or exclusively within a subset of variables. Figure 1 provides an illustrative

Multivariate Time Series with Shape, (4 x 15)

Integer	1	3	3	5	5	2	3	3	5	5	3	3	5	4	4
Continuous	4.3	4.5	2.6	3.0	3.0	1.7	4.9	2.9	3.3	1.9	4.9	2.5	3.1	1.8	0.3
Ordinal	A	D	B	D	A	A	A	C	C	B	D	D	C	A	A
Nominal	T	L	T	Z	Z	T	L	T	Z	T	L	T	Z	L	L

Fig. 1: A multivariate time series of length $n = 15$ and dimensionality $m = 4$, illustrating a motif of length $s = 3$ with $r = 3$ occurrences at positions $\mathcal{O}_M = \{2, 7, 11\}$. The motif spans the first, second, and fourth dimensions.

motif of length s=3 (highlighted in grey) in a time series of length n=15 and multivariate order m=4. The motif spans the first, second, and fourth dimensions, while the third is excluded for not meeting a predefined similarity threshold.

Depending on the application, different algorithms for multivariate motif discovery may be selected based on desirable types of patterns (e.g., fixed or variable motif lengths [18], exact or approximate motifs [10, 41]), search constraints [23, 41] (e.g., predefined minimum and maximum numbers of variables, and inclusion or exclusion of specific variables), incorporation of statistical tests to filter spurious motifs [7], or computational constraints (e.g., massive series [47], streaming data [25]). Definitions of motifs also vary, ranging from motif pairs to sets [24], with further differences in thresholds, recurrence criteria, and variable selection.

Definition 4. *Given a time series X of length n and a (multidimensional) motif of interest $M = X_{i,s,J}$ with past occurrences $\mathcal{O}_M$, the task of **next motif prediction** aims to forecast the starting position k of the next occurrence of M within a given future horizon of length h. The horizon is defined as the sequence of future observations $X' = \langle \mathbf{x}_{n+1}, \mathbf{x}_{n+2}, \ldots, \mathbf{x}_{n+h} \rangle$, where h defines the lookahead window within which the next occurrence of M is expected to occur.*

Formally, the objective is to learn a function f, such that: $\hat{k}=f(X_{n-w:n}, M, h)$, where $X_{n-w:n}$ is the history window and $\hat{k} \in \mathbb{N}^+$ is the predicted starting position of the next occurrence of motif M in the interval $[n+1, n+h]$. The predictor f is optimized to minimize a predefined loss function $\mathcal{L}(k, \hat{k})$ that penalizes deviations between $\hat{k}$ and the actual next true occurrence k.

Critical non-periodic behaviors are pervasive across real-world systems, including biological, urban, societal, physical, and digital systems [24, 33]. Motivated by the need to forecast occurrences of specific phenomena of interest, especially when past occurrences are irregular, the target task of predicting the next occurrence of a given motif is formalized in Def. 3. Unlike standard time series forecasting, which targets the projection of raw values, the goal is to estimate when a structured recurrence, represented by a motif M, will reappear.

The task of predicting next motif occurrences shares conceptual similarities with other well-established learning tasks for time series, particularly event prediction [46], where the goal is to estimate the timing of future events using methods such as survival models (e.g., Cox models [9]) and temporal point processes (e.g., Hawkes and Poisson processes [32, 45]). However, distinctions arise when considering how events and motifs are defined, the underlying recurrence assumptions, and the learning methodologies used to forecast them.

The formulated task can be straightforwardly extended to consider the prediction of *all* upcoming occurrences of a given motif along a horizon under a multi-output regression formulation as $\hat{\mathcal{O}}_M = f(X_{n-w:n}, M, h)$, where $\hat{\mathcal{O}}_M$ is the set of predicted motif start times. Additionally, the upper bound placed by the horizon of prediction can be removed for a more general formulation as a continuous time-to-motif prediction. The problem can be alternatively formulated as a binary classification task over a sliding window, which can be used to estimate the probability of a motif occurring within specific future intervals. The task formulation admits several natural extensions, including the possibility of inputting multiple motifs of interest $\mathcal{M}$, requiring the model to jointly predict multiple motifs' future occurrences to support cross-motif synergistic learning, resembling multi-task learning formulations.

3 Related work

Motif analysis has been previously explored to support classical time series forecasting tasks from multiple perspectives:

- to approximate periodic behaviors that fall outside seasonal patterning and their subsequent projection along the horizon of prediction [34];
- to discover temporal association rules that link recurring patterns to subsequent events, thereby capturing precursor–target relationships relevant to forecasting [11, 16];
- to engineer features from motif statistics (e.g., occurrences, frequencies) and incorporating them at the input level to enhance predictive performance [36, 38];
- to guide the segmentation of time series by selecting informative segments based on the presence of specific motifs of interest to improve instance construction during model training [40];
- to identify and isolate anomalous patterns to mitigate their distortive effects on predictive models [21];
- to undertake similarity-based forecasting by identifying segments that contain analogous behaviors to those close to the horizon of prediction [22];
- to reduce the complexity and overfitting risks of the forecasting task by using motifs for summarizing or decreasing the dimensionality of the given series, especially for series of high length or multivariate order [40]; or
- to integrate motif representations with raw time series inputs through dedicated encoders or hybrid forecasting architectures [30, 37].

Despite the relevance of these contributions, they are not designed to address the specific task of forecasting the next motif occurrence. In this context, five major challenges arise. First, the need to condition the prediction task on specific motifs of interest. Second, the challenge of uncovering potential regularities within non-periodic recurrence patterns, whether driven by non-linear dynamics or latent precursor events. Third, the importance of placing adequate segmentation criteria to guide the learning process of this specific task. Fourth, the need to handle the inherent stochasticity associated with temporal dynamics of real-world systems, generally associated with variable temporal misalignments. Finally, the necessity of modeling multivariate dependencies, where certain variables may serve as contextual signals for the occurrence of target behaviors.

4 Predicting the next motif occurrence

To address the task of predicting the next occurrence of a non-periodic behavior in a high-dimensional time series, we explore two key premises. First, the recurrence dynamics of non-periodic motifs may follow complex, non-linear recurrence dynamics, influenced by latent interactions, hidden regularities, or emerging behaviors that do not conform to strict periodicity. Second, motifs may be preceded by subtle precursor signals embedded within the time series data, acting as predictive proxies that inform the timing of the next occurrence. Based on these premises, this work proposes multiple-input neural network models to regress the next occurrence of a motif subsequence in a multivariate time series. In particular, this methodology is positioned to answer the following technical questions:

- Which features can be extracted to guide the forecasting task? How to optimally encode them?
- Which neural architectures are best suited to learn non-periodic motif recurrence dynamics? How can temporal dependencies be effectively captured?
- How to segment the series and organize the resulting segments to support the learning and assessment of the target predictors?

Encoding Motif Recurrence. It is often overly optimistic to expect learners to predict occurrences of a motif using only the raw time series data as input and the occurrence time as output. A simple regression model lacks critical information about which motif is being analyzed, its prior occurrences, and how these relate to future recurrence behavior.

To support neural models in learning motif recurrence patterns, we propose encoding strategies that jointly represent the multivariate time series and the structural recurrence of a given motif. Let $X = \langle \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \rangle$ be the multivariate time series, $M = X_{i,s,J}$ the motif of interest, and $\mathcal{O}_M = \{t_1, t_2, \ldots, t_r\} \subseteq \{1, \ldots, n\}$ the set of known start positions of past occurrences of M. Let $w \in \mathbb{N}^+$ be the length of the historical window and define the set of window indices as $\mathcal{W} = \{n - w + 1, \ldots, n\} \subseteq \{1, \ldots, n\}$. We define the following encodings over the window $X_{n-w:n}$:

- *Binary Mask Encoding:* A sequence $B \in \{0,1\}^w$, where $B[t] = 1$ if time index $t \in \mathcal{W}$ falls within any occurrence of motif M, i.e., if there exists $t_j \in \mathcal{O}_M$ such that $t \in \{t_j, \ldots, t_j + s - 1\}$; otherwise, $B[t] = 0$.
- *Historical Index Encoding:* A sequence $\mathbf{I} = \langle t_1, \ldots, t_k \rangle$, where each $t_j \in \mathcal{O}_M \cap \mathcal{W}$ represents a past motif occurrence within the historical window.

These two forms of encoding can be used separately or complementarily, and should provide sufficient information for the learners to perform the task, even when the recurrence dynamics of a given motif are non-linear in nature.

While motif recurrence encodings provide structural recurrence information, they may not be sufficient if the recurrence lacks strong underlying regularities. In such cases, the raw (normalized) multivariate time series can be included as an additional input source to capture precursor signals or latent dependencies that influence motif timing. In this context, multimodal stances, defined by the input of multiple time series with non-identical length and order can be considered, including those combining the raw time series, the target motif sequence, and/or its historical indices.

Complementary encodings can be further envisioned, including numerical time series representations that estimate the probability of a motif occurring at each time point t in the historical window. This probability can be derived from predefined (elastic) distance ratios, such as multivariate dynamic time warping.

Neural Network Architectures for Motif Prediction. Considering the proposed representations, time-aware deep learning models for next motif occurrence prediction are explored. We focus on neural architectures that leverage well-established principles in multivariate time series learning, including:

- *Long Short-Term Memory* Networks [43]. Designed to capture long-range temporal dependencies within multivariate time series, LSTMs aim to model complex, non-periodic recurrence dynamics by maintaining and updating memory states over time.
- *Temporal Convolutional Neural Networks* [17]. By employing stacked convolutional and pooling layers, TCNs effectively extract medium-range recurrence patterns in motifs, capturing local structures within multivariate dependencies.
- *Transformer-Based Models* [35]. Leveraging self-attention mechanisms, Transformers dynamically weigh relevant past information, enhancing their ability to model long-range dependencies and interactions between precursor events.

Predictive Modeling via Motif-Aware Segmentation. To create the necessary supervision to answer the target predictive task, state-of-the-art principles for motif-aware segmentation of the input multivariate time series are suggested [28]. In its simplest form, a sliding-window method can be used to generate training, validation, and testing input-output pairs. The size of the input window should be sufficient to contain multiple instances of the motif – to facilitate the learning phase –, with the output indicating the position of the subsequent motif occurrence, either using as reference the start of the forecasting period or the last occurrence of the motif. To assess the model's performance, time-aware partitioning of data instances should be employed to carefully preserve the temporal order during the model's training and testing phases, preventing data leakage. The proposed setting can also be adapted to streaming scenarios, where predictions are made online by applying incremental motif detection and continuously evaluating the prediction function f over incoming data.

5 Results

Three distinct experimental settings were created to assess the role of the proposed methodology in predicting motif occurrences. In particular, the following three major research questions are tackled:

- **RQ1.** To what degree can the regression approaches predict the emergence of motif recurrence patterns?
- **RQ2.** To what extent can the proposed approaches leverage subtle signals preceding motifs while maintaining robustness to stochastic fluctuations in the motif's occurrences?
- **RQ3.** How does the encoding of motif information impact forecasting behavior?

5.1 Datasets

We evaluate our method on synthetic and real-world multivariate time series.

Synthetic. Each dataset is a discrete-valued time series $X = \langle \mathbf{x}_1, \ldots, \mathbf{x}_n \rangle$ of length $n = 100{,}000$, dimension $m = 3$, with $x_t^j \in \{1, \ldots, 5\}$.

Synthetic 1 embeds a fixed bivariate motif $M = X_{i,5,\{1,3\}}$ at positions following a repeating cycle of prime-numbered intervals $\{5, 7, 11, \ldots, 47\}$, generating structured, non-uniform recurrence.

Synthetic 2 inserts the same motif at random intervals (20–45 steps apart), each preceded at 13–15 steps by a 5-length precursor in variables $\{1, 3\}$, with all values set to 5, forming a weak, yet informative, temporal cue with bounded noise.

Real-world. *Household Power.* Minute-level electricity data from a French household [12], from Dec 2006–Nov 2010. We use Sept 1–Oct 24, 2008, aggregated to 5-minute intervals ($n = 15{,}552$, $m = 2$: active/reactive power in kW).

Lisbon Population. Hourly population density estimates via cellphone triangulation (Vodafone Portugal), Sept 15–Nov 30, 2021. We use a univariate series ($n = 1{,}847$) from the `Avenidas Novas` district.

5.2 Experimental setting

Evaluation. The evaluation differs between synthetic and real-world datasets. For synthetic experiments, a rolling five-fold cross-validation is used. Each fold consists of input–output pairs: 100-point input windows and 50-point prediction horizons. The target is the index of the next motif occurrence, relative to the prediction start. A sliding window with step size 5 ensures chronological integrity and sufficient training instances (Figure 2). We use a 64%/16%/20% train-validation-test split, consistently across models.

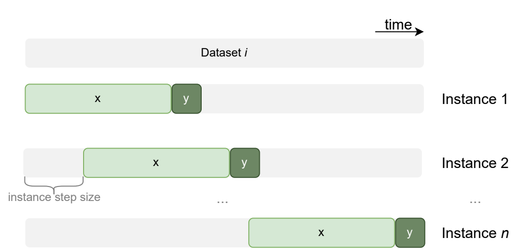


Fig. 2: Input–output pair creation for training the regression models.

For real-world case studies, the datasets are partitioned using a 70%/15%/15% train-validation-test split. Five datasets are created, where each dataset corresponds to one of the top 5 (multidimensional) motifs with regards to the number of repetitions as identified by the motif discovery algorithm proposed in Yeh et al. [41]. Each dataset consists of instances derived from its respective motif, which serves as a reference pattern for the models to learn and predict. In the population case, motifs repeat every 12 hours; inputs span 3 weeks (504h), outputs 2 days (48h), step 1. For electricity data, motifs are 3-hour sequences; inputs span 2 days (576 min), with 1-day horizons (288 min), step 5.

Model performance is evaluated using *Mean Absolute Error (MAE)* and *Root Mean Squared Error (RMSE)* on test sets. Statistical significance is assessed via paired t-tests and Wilcoxon signed-rank ($\alpha \leq 0.05$).

Hyperparameterization. Hyperparameter tuning is done using Optuna [1], with models trained using *mean squared error (MSE)* loss and the Adam optimizer [14]. Each model is optimized over 100 trials, tuning learning rates (log-uniform between 10^{-5} and 10^{-3}), batch sizes from $\{4, 8, 16, 32, 64, 128\}$, and architecture-specific settings. Training runs up to 500 epochs, with early stopping and a 15-minute time limit per trial.

We assess five architectures for the *next motif* task: Feed-Forward Neural Networks (FNNs), Long Short-Term Memory (LSTMs), 1D Convolutional Neural Networks (1D-CNNs), Temporal Convolutional Networks (TCNs), and Transformers. Table 1 details each model, including the baselines used for comparison.

FNNs use 1–4 layers, with hidden units sampled from $\{16, 32, 64, 128, 256\}$. LSTMs have 1–3 layers with hidden sizes from the same set. 1D-CNNs use 1–3 convolutional layers with filters from $\{16, 32, 64\}$, kernel sizes $\{3, 5, 7\}$, and optional pooling ($\{2, 3\}$). TCNs use kernel sizes from $\{3, 5, 7\}$, with block counts set by receptive field size and channel sizes sampled from $\{16, 32\}$. Transformers vary over model dimensions $\{64, 128, 256, 512\}$, attention heads $\{2, 4, 8, 16\}$, 1–3 encoder layers, feedforward sizes $\{128, 256, 512\}$, and dropout in $[0.0, 0.5]$. Optimal configurations are selected based on validation loss.

Model	**Description**
FNN	A fully connected neural network with configurable input, hidden, and output layers, with a final output layer for regression.
LSTM	An LSTM-based sequence model with configurable layers, processing optional auxiliary inputs, and a fully connected regression output layer.
1D-CNN	A 1D convolutional neural network with configurable convolutional and pooling layers, followed by a fully connected regression output layer.
TCN	A temporal convolutional network using dilated convolutions and residual connections, with a fully connected regression output layer.
Transformer	A time series transformer model using attention mechanisms for sequence forecasting, with configurable layers, heads, and a regression output layer.
Naive	Naive predictive models as baselines. The time of next repetition is predicted using: *Average*: average difference between consecutive time points. *Last*: last observed difference between consecutive time points.

Table 1: Baseline and deep learning models for time series regression, including details of the regression output layers.

5.3 Synthetic data results

We assess model performance under two scenarios: *Synthetic 1* with structured, non-trivial motif recurrence, and *Synthetic 2* with irregular motifs preceded by weak precursors.

Non-trivial recurrence. *Synthetic 1* simulates motifs occurring at intervals defined by a cycle of prime-numbered lags. Three input encodings are compared:

- **TS**: Raw time series, normalized across features.
- **TS+B**: TS augmented with a binary mask marking motif occurrences. This encoding results in an augmented multivariate time series representation with $m+1$ variables, where m is the number of original variables.
- **I**: Normalized indices of past motif occurrences within the input window. For example, <4, 18, 36> represents three occurrences of the target motif at the specified time points.

Table 2 shows clear advantages for encodings that explicitly represent motif positioning. Using raw time series data alone led to the poorest performance across all models, as evidenced by the higher MAE and RMSE scores. The FNN model showed the weakest performance (MAE = 8.937, RMSE = 11.175) significantly worse than both baselines, including the average distance between motifs

Model	Input	Configuration	MAE	RMSE
Naive Average	I		7.061 ± 0.013	9.467 ± 0.009
Naive Last	I		4.363 ± 0.008	7.538 ± 0.004
FNN	TS	Units=256, LR=3.0e-04, BS=16	8.937 ± 0.299	11.175 ± 0.288
	TS+B	Units=256, 256, 256, 128, LR=8.6e-04, BS=16	1.012 ± 0.084	1.687 ± 0.186
	I	Units=32, 256, 128, 256, LR=6.4e-04, BS=16	0.550 ± 0.299	1.043 ± 0.812
LSTM	TS	Units=32, 16, 128, LR=8.7e-04, BS=32	4.428 ± 1.535	7.396 ± 1.348
	TS+B	Units=16, 128, 256, LR=3.8e-04, BS=16	1.100 ± 0.319	2.905 ± 0.883
	I	Units=16, 128, 256, LR=2.0e-04, BS=64	$\mathbf{0.149 \pm 0.056}$	$\mathbf{0.187 \pm 0.058}$
1D-CNN	TS	Filters=16, 64, 64, KS=7, PS=3, LR=7.5e-04, BS=32	6.418 ± 0.392	8.569 ± 0.501
	TS+B	Filters=32, 32, 64, KS=7, PS=3, LR=2.9e-04, BS=64	0.701 ± 0.085	0.968 ± 0.125
	I	Filters=64, 64, 64, 64, KS=7, PS=None LR=8.6e-04, BS=32	0.345 ± 0.172	0.555 ± 0.281
TCN	TS†	Filters=16, 16, 16, 16, 32, KS=3, Dropout=0.026, LR=9.8e-04, BS=16	$\mathbf{1.088 \pm 0.049}$	$\mathbf{1.752 \pm 0.347}$
	TS+B†	Filters=32, 32, 32, 16, KS=7, Dropout=7e-05, LR=6.1e-04, BS=32	$\mathbf{0.521 \pm 0.301}$	$\mathbf{0.708 \pm 0.441}$
	I	Filters=32, 16, KS=5, Dropout=0.041, LR=8.1e-04, BS=16	0.817 ± 0.094	1.084 ± 0.198
Transformer	TS	Emb.=64, Attention Heads=16, Enc. Layers=3, FF=128, Dropout=0.093, LR=4.2e-04, BS=16	1.750 ± 0.242	3.790 ± 0.317
	TS+B	Emb.=256, Attention Heads=8, Enc. Layers=3, FF=256, Dropout=4e-04, LR=3.5e-04, BS=64	0.553 ± 0.769	0.791 ± 1.137
	I	Emb.=128, Attention Heads=8, Enc. Layers=1, FF=128, Dropout=2e-04, LR=6.13e-04, BS=64	0.404 ± 0.130	0.533 ± 0.191

Table 2: Results on Synthetic 1 (non-trivial recurrence), reported as mean ± std. †: early stopping due to time limits. Inputs: TS = raw series; TS+B = binary mask; I = index encoding of prior motifs. Bold: best per input with stat. significant improvement over baselines.

for both MAE (p-value = 1e-04) and RMSE (p-value = 1e-04), as well as the last distance between motifs for MAE (p-value = 4e-06) and RMSE (p-value = 7e-06). Likewise, the 1D-CNN (MAE = 6.418, RMSE = 8.569) and LSTM (MAE = 4.428, RMSE = 7.396) models encountered difficulty in effectively capturing motif patterns, further highlighting the limitations of this encoding scheme when applied under the assumption of periodic recurrence. Among all architectures, the TCN model achieved the best performance using raw time series input, notwithstanding early stopping constraints during training (MAE = 1.088, RMSE = 1.752). However, even this superior performance was, on average, inferior to the scores obtained with the **TS+B** encoding (statistically significant for MAE with p-value = 3e-04) and Index Encoding, both directly encoding the positional information of the motif.

Adding a mask encoding to the time series data consistently produced statistically significant reductions in MAE and RMSE scores across all models, particularly for the FNN and 1D-CNN architectures (p-values of 4e-07 and 4e-06 for MAE, respectively). This suggests that explicitly marking motif positions within the input enhances model predictive accuracy. Nevertheless, while **TS+B** encoding significantly outperformed the **TS** scheme, it did not consistently surpass the performance of Index Encoding, as confirmed by statistical tests.

Notably, the use of Index Encoding markedly enhanced model performance, especially for architectures designed to capture sequential dependencies. The LSTM model, under Index Encoding, achieved the lowest error rates (MAE = 0.149, RMSE = 0.187), outperforming all other configurations. Similarly, other models, such as the 1D-CNN and Transformer, exhibited notable improvements with this encoding, underscoring its effectiveness in representing motif occurrences as a compact yet informative input format when occurrences conform to well-defined yet non-linear rules.

These results confirm that under deterministic recurrence, explicitly encoding prior motif positions is essential for accurate prediction, and that most architectures cannot reliably infer this structure from raw input alone.

Precursors and stochastic recurrence. *Synthetic 2* evaluates models under irregular recurrence, where motifs are weakly preceded by literal precursor signals with bounded stochasticity. An LSTM with Index input (MAE = 5.780, RMSE = 6.921) is used as a baseline, since it relies solely on motif positioning and disregards precursor signals.

As shown in Table 3, all models outperformed naive baselines with TS input, demonstrating that they capture some level of the underlying structures that discriminate a motif occurrence. Specifically, they appear to account for the temporal dependency that governs motif recurrence—when a motif has not appeared for an extended period, its likelihood of occurring increases, whereas if it has occurred recently, the probability of an immediate repetition decreases.

Using the LSTM with Index Encoding as a baseline, the LSTM, TCN, and Transformer models exhibited statistically significant reductions in MAE (p-values of 2.9e-04, 6.8e-04, and 3.8e-03, respectively). These results indicate that these models effectively learn to detect and leverage the precursor signal under

Model	Input	Configuration	MAE	RMSE
Naive Average	I		7.128 ± 0.268	8.855 ± 0.239
Naive Last	I		7.911 ± 0.267	9.941 ± 0.223
LSTM	I	Layers=2, Units=32, 32, LR=9.2e-04, BS=16	5.780 ± 0.278	6.921 ± 0.172
FNN	TS	Units=32, 32, 32, 256, LR=5.6e-04, BS=16	7.174 ± 0.241	10.316 ± 0.418
	TS+B	Units=16, 256, LR=8.4e-04, BS=32	6.292 ± 0.108	8.903 ± 0.406
LSTM	TS	Units=64, 64, LR=9.8e-04, BS=16	4.894 ± 0.212	7.320 ± 0.338
	TS+B	Units=128, 64, LR=9.0e-04, BS=16	4.976 ± 0.387	7.499 ± 0.792
1D-CNN	TS	Filters=32, 64, 32, KS=3, PS=None LR=5.4e-04, BS=32	6.271 ± 0.191	8.282 ± 0.251
	TS+B	Filters=64, 64, 32, KS=3, PS=None LR=1.6e-04, BS=16	5.526 ± 0.189	7.429 ± 0.210
TCN	TS†	Filters=16, 32, 16, 32, 32, KS=3, Dropout=0.196, LR=4.6e-04, BS=16	**4.795 ± 0.227**	7.132 ± 0.414
	TS+B†	Filters=16, 16, 32, 16, 32, KS=3, Dropout=0.389, LR=8.9e-04, BS=16	**4.871 ± 0.142**	7.067 ± 0.261
Transformer	TS	Emb.=64, Attention Heads=16, Enc. Layers=2, FF=256, Dropout=0.176, LR=3.8e-04, BS=16	5.033 ± 0.206	7.743 ± 0.442
	TS+B	Emb.=128, Attention Heads=2, Enc. Layers=2, FF=256, Dropout=0.116, LR=2.9e-04, BS=32	4.960 ± 0.385	7.727 ± 0.712

Table 3: Results on Synthetic 2 (irregular recurrence with precursors), reported as mean ± std. †: early stopping due to time limits. Inputs: TS = raw series; TS+B = binary mask; I = index encoding of prior motifs. Bold indicates best per input with statistically significant improvement over baselines.

this setting, despite the stochastic variability of the signals with regard to their time occurrence. However, none of the models outperformed the LSTM baseline according to RMSE, suggesting that while average prediction errors decreased, occasional large deviations still occurred. This highlights the challenge of predicting motif recurrence in stochastic contexts.

The introduction of explicit motif-positioning information via the **TS+B** encoding further improved the performance of the FNN and 1D-CNN architectures, likely by aiding temporal extrapolation between the last observed motif and the output value. For models already capable of leveraging temporal structure (e.g., LSTM, TCN), the mask added little or no benefit, suggesting these architectures could infer motif context without explicit labels.

5.4 Real data results

We evaluate model robustness on two real-world case studies involving complex, naturally recurring motif structures.

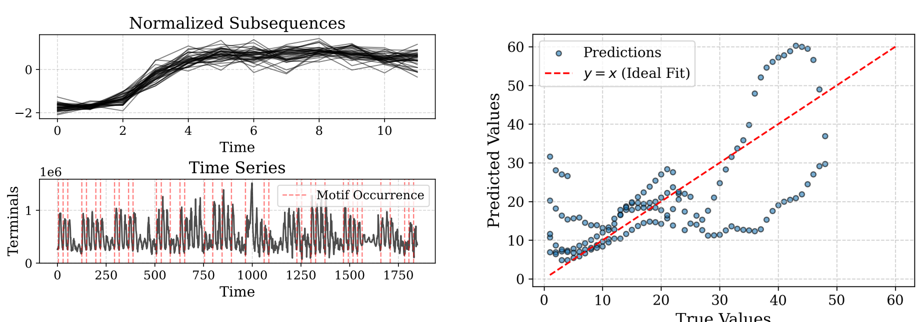
Population density. In this case study, we forecast 12-hour population density motifs in Lisbon's `Avenidas Novas` district. From 27 motifs (avg. 20.61 occurrences each), the five most frequent were selected for evaluation. Table 4 summarizes the results, with naive models (average and last distance between contiguous motifs) serving as baselines. The models significantly outperformed naive baselines in MAE, confirming their ability to capture meaningful population density trends.

Model	Input	MAE	RMSE
Naive Average	I	12.951 ± 2.935	15.588 ± 3.456
Naive Last	I	18.528 ± 9.754	29.364±13.468
FNN	TS	9.336 ± 1.536	11.408 ± 1.353
	TS+B	**9.164 ± 0.995**	11.874 ± 1.460
	I	11.138 ± 2.185	13.243 ± 1.844
LSTM	TS	9.793 ± 0.991	12.219 ± 1.060
	TS+B	9.472 ± 1.478	11.842 ± 1.423
	I	**9.783 ± 0.996**	12.221 ± 1.044
1D-CNN	TS	9.595 ± 1.534	11.600 ± 1.172
	TS+B	11.498 ± 1.678	13.672 ± 1.338
	I	10.262 ± 1.301	12.556 ± 0.866
TCN	TS	**9.066 ± 1.954**	11.506 ± 1.843
	TS+B	9.448 ± 0.981	11.722 ± 0.649
	I	9.896 ± 1.249	12.023 ± 1.253
Transformer	TS	11.466 ± 4.332	14.569 ± 3.769
	TS+B	9.979 ± 1.524	12.899 ± 1.955
	I	10.910 ± 1.309	12.919 ± 1.190

Table 4: MAE and RMSE on population density data, reported as mean ± std. Inputs: TS = raw series; TS+B = binary mask; I = index encoding of prior motifs. Bold indicates best per input with statistically significant improvement over baselines.

Notably, the TCN utilizing time series **TS** input (p-value = 0.041 vs. Naive Avg. for MAE), the FNN with the **TS+B** encoding (p-value = 0.038 vs. Naive Avg. for MAE), and the LSTM applied to sequences of indices (p-value = 0.047 vs. Naive Avg. for MAE) were identified as the most effective architectures for their respective input types. These differences suggest that models exploit different input representations depending on architecture.

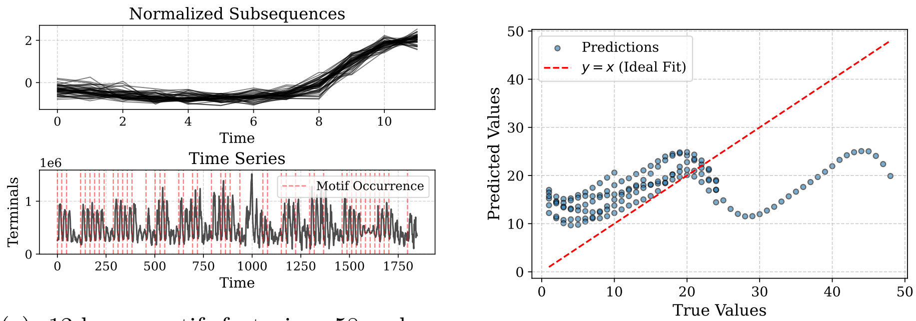
Figures 3 and 4 provide a comparison of motif recurrence predictions from two models: the FNN model with **TS+B** input (MAE = 8.090, RMSE = 11.080) and the TCN model using **TS** as input (MAE = 5.966, RMSE = 9.553). The first mo-



(a) 12-hour motif featuring 38 subsequences.

(b) Predictions vs. True Values

Fig. 3: (a) Visualization of a 12-hour motif found on the population density data from `Avenidas Novas, Lisbon`. The red dashed lines indicate motif occurrences. (b) Predictions of the FNN regression model with **TS+B** as input vs. the true values, where the red dashed line represents the ideal $y = x$ fit.



(a) 12-hour motif featuring 58 subsequences.

(b) Predictions vs. True Values

Fig. 4: (a) Visualization of a 12-hour motif found on the population density data from `Avenidas Novas, Lisbon`. The red dashed lines indicate motif occurrences. (b) Predictions of the TCN regression model with **TS** as input vs. the true values, where the red dashed line represents the ideal $y = x$ fit.

tif consists of 38 subsequences, exhibiting a sharp increase in population density in `Avenidas Novas` after the second hour, followed by a period of stabilization at a high density for the remainder of the 12-hour cycle. In contrast, the second motif is characterized by low population density during the first eight hours, followed by a rapid increase in the remaining hours. The predictive behaviors indicate that both models perform more accurately when forecasting short-term motif recurrences within a 24-hour window, as evidenced by their alignment with the ideal y=x reference line. However, the models tend to underestimate motif occurrences farther into the forecast horizon, suggesting difficulties in capturing long-term dependencies. While self-attention mechanisms have the potential to improve performance, the transformer-based architecture struggled in this context, likely due to the relatively small dataset size, which may have hindered its ability to effectively learn meaningful long-range dependencies.

Electricity consumption. This study analyzes motifs in household power usage, with 19 recurring 2-hour patterns (avg. 219.47 occurrences each). Table 5 presents the predictive performance of various models across different input encoding schemes, using the same naive distance criteria as baselines.

Regarding predictive errors, all models outperformed the naive baselines in both MAE and RMSE, demonstrating their ability to capture meaningful patterns in household electricity usage. The LSTM trained with index-based encoding achieved the lowest MAE (30.8) and one of the lowest RMSE values (41.0). Among models using **TS** input, the TCN exhibited the lowest prediction errors (MAE = 32.2, RMSE = 41.6); showing statistically significant improvements on RMSE against the Naive Average (p-value=0.036 for RMSE, p-value=0.114 for MAE). The Transformer model trained with index encoding achieved the lowest RMSE overall (40.9), but its results were not statistically significant compared to other non-baseline models using the same input.

Model	Input	MAE	RMSE
Naive Average	I	37.435±14.656	48.206±17.560
Naive Last	I	40.447 ± 9.781	53.754±11.480
FNN	TS	34.523 ± 9.854	44.126±12.094
	TS+B	33.268 ± 8.641	45.107±13.095
	I	31.844 ± 7.135	42.039±10.101
LSTM	TS	32.419 ± 8.289	41.870±13.081
	TS+B	**31.152 ± 7.897**	40.939±12.982
	I	**30.800 ± 8.553**	40.969±13.619
1D-CNN	TS	32.804 ± 8.355	42.631±12.451
	TS+B	35.722±10.876	45.666±12.941
	I	31.926 ± 7.438	42.335±11.360
TCN	TS	32.181 ± 8.359	**41.624±12.402**
	TS+B	32.558 ± 8.320	46.219±12.811
	I	31.660 ± 7.808	42.333±13.360
Transformer	TS	33.431 ± 8.797	43.784±12.701
	TS+B	32.243 ± 9.170	42.731±12.872
	I	31.987 ± 6.727	**40.922 ± 8.333**

Table 5: MAE and RMSE on household electricity data, reported as mean ± std. Inputs: TS = raw series; TS+B = binary mask; I = index encoding of prior motifs. Bold indicates best per input with statistically significant improvement over baselines.

6 Conclusions and future work

This work introduces and formalizes the task of predicting the next occurrence of a (non-periodic) multidimensional motif, and outlines methodological principles to address it. The approach leverages neural models to uncover complex recurrence dynamics and identify precursor signals in the data that anticipate future motif instances. Mask-based encodings are proposed to integrate motif recurrence and pattern information into the raw multivariate time series, serving as structured inputs for supervised regression.

To explore the three guiding research questions, we conducted experiments on both synthetic and real-world datasets, showing that different architectures are better suited for different conditions: LSTM models excel under non-trivial recurrence, while TCNs and Transformers perform better when precursor signals are predictive. These findings highlight the importance of architectural alignment with the temporal nature of motif recurrence.

This work lays a foundation for forecasting critical events in real-world systems, with applications ranging from organ failure and financial volatility to atmospheric risks and traffic anomalies. Future directions include: i) incorporating both historical and forward-looking context to enhance event forecasting [29]; ii) deeper analysis of the recurrence dynamics behind irregular motifs; iii) evaluate the robustness of the proposed methods under varying levels of noise, precursor signal strengths, and missing rates; iv) evaluating the added predictive value of explicitly modeling upcoming motif occurrences; and v) exploring multi-input learning, where raw time series, recurrence masks, and index sequences are processed as separate input branches rather than merged into a single multivariate series.

Acknowledgments. This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through FRAIL project, ref. 2024.07266.IACDC; LAIfeBlood+ project, ref. 2024.07475.IACDC; LASIGE Research Unit, ref. UID/000408/2025 - LASIGE; INESC-ID pluriannual, ref. UIDB/50021/2020; and the PhD research scholarship with ref. UIBD/153086/2022 to Miguel G. Silva.

Disclosure of Interests. The authors have no competing interests to declare.

Code Availability The code used in this study is available at `https://github.com/MiguelGarcaoSilva/motifpred`.

References

[1] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2623–2631. ACM (2019). https://doi.org/10.1145/3292500.3330701

[2] Allred, C., Pusey, J.L., Harper, M.: Detecting ballistic motions in quadruped robots: A boosted tree motif classifier for understanding reinforcement learning. In: International Conference on Robotic Computing. pp. 143–151. IEEE (2023). https://doi.org/10.1109/IRC59093.2023.00032

[3] Bahri, O., Li, P., Boubrahimi, S.F., Hamdi, S.M.: Shapelet-based temporal association rule mining for multivariate time series classification. In: IEEE International Conference on Big Data. pp. 242–251. IEEE (2022). https://doi.org/10.1109/BIGDATA55660.2022.10020478

[4] Battiston, F., Nicosia, V., Chavez, M., Latora, V.: Multilayer motif analysis of brain networks. Chaos: An Interdisciplinary Journal of Nonlinear Science **27**(4) (2017). https://doi.org/10.1063/1.4979282

[5] Bürger, R., Chowell, G., Lara-Díaz, L.Y.: Measuring differences between phenomenological growth models applied to epidemiology. Mathematical Biosciences **334**, 108558 (2021). https://doi.org/10.1016/j.mbs.2021.108558

[6] Castanho, E.N., Aidos, H., Madeira, S.C.: Biclustering fmri time series: a comparative study. BMC bioinformatics **23**(1), 192 (2022). https://doi.org/10.1186/S12859-022-04733-8

[7] Castro, N.F., Azevedo, P.J.: Significant motifs in time series. Statistical Analysis and Data Mining: The ASA Data Science Journal **5**(1), 35–53 (2012). https://doi.org/10.1002/SAM.11134

[8] Cheng, H., Tan, P., Potter, C., Klooster, S.A.: Detection and characterization of anomalies in multivariate time series. In: SIAM International Conference on Data Mining, SDM. pp. 413–424. SIAM (2009). https://doi.org/10.1137/1.9781611972795.36

[9] Deep, A., Veeramani, D., Zhou, S.: Event prediction for individual unit based on recurrent event data collected in teleservice systems. IEEE Transactions on Industrial Informatics **69**(1), 216–227 (2020). https://doi.org/10.1109/TR.2019.2909471

[10] Ferreira, P.G., Azevedo, P.J., Silva, C.G., Brito, R.M.M.: Mining approximate motifs in time series. In: Discovery Science. Lecture Notes in Computer Science, vol. 4265, pp. 89–101. Springer (2006). https://doi.org/10.1007/11893318_12

[11] He, Y., Chu, X., Peng, J., Gao, J., Wang, Y.: Motif-based rule discovery for predicting real-valued time series. CoRR **abs/1709.04763** (2017), http://arxiv.org/abs/1709.04763

[12] Hébrail, G., Berard, A.: Individual household electric power consumption (08 2012). https://doi.org/10.24432/C58K54

[13] Jin, C., Song, C., Bjelland, J., Canright, G., Wang, D.: Emergence of scaling in complex substitutive systems. Nature human behaviour **3**(8), 837–846 (2019). https://doi.org/10.1038/s41562-019-0638-y

[14] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

[15] La Rovere, M.T., Bigger, J.T., Marcus, F.I., Mortara, A., Schwartz, P.J.: Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction. The Lancet **351**(9101), 478–484 (1998)

[16] Lamp, J., Derdzinski, M., Hannemann, C., Hatfield, S., van der Linden, J.: Motifdisco: Motif causal discovery for time series motifs. CoRR **abs/2409.15219** (2024). https://doi.org/10.48550/ARXIV.2409.15219

[17] Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017. pp. 1003–1012. IEEE Computer Society (2017). https://doi.org/10.1109/CVPR.2017.113

[18] Li, Y., Lin, J., Oates, T.: Visualizing variable-length time series motifs. In: International Conference on Data Mining. pp. 895–906. SIAM / Omnipress (2012). https://doi.org/10.1137/1.9781611972825.77

[19] Liu, B., Li, J., Chen, C., Tan, W., Chen, Q., Zhou, M.: Efficient motif discovery for large-scale time series in healthcare. IEEE Transactions on Industrial Informatics **11**(3), 583–590 (2015). https://doi.org/10.1109/TII.2015.2411226

[20] Lu, W., Li, J., Wang, J., Qin, L.: A cnn-bilstm-am method for stock price prediction. Neural Computing and Applications **33**(10), 4741–4753 (2021). https://doi.org/10.1007/S00521-020-05532-Z

[21] Lu, Y., Wu, R., Mueen, A., Zuluaga, M.A., Keogh, E.J.: Matrix profile XXIV: scaling time series anomaly detection to trillions of datapoints and ultra-fast arriving data streams. In: SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 1173–1182. ACM (2022). https://doi.org/10.1145/3534678.3539271

[22] Martínez-Álvarez, F., Lora, A.T., Riquelme, J.C., Aguilar-Ruiz, J.S.: Discovery of motifs to forecast outlier occurrence in time series. Pattern Recognition Letters **32**(12), 1652–1665 (2011). https://doi.org/10.1016/J.PATREC.2011.05.002

[23] Minnen, D., Jr., C.L.I., Essa, I.A., Starner, T.: Detecting subdimensional motifs: An efficient algorithm for generalized multivariate pattern discovery. In: International Conference on Data Mining. pp. 601–606. IEEE Computer Society (2007). https://doi.org/10.1109/ICDM.2007.52

[24] Mueen, A.: Time series motif discovery: dimensions and applications. WIREs Data Mining and Knowledge Discovery **4**(2), 152–159 (2014). https://doi.org/10.1002/WIDM.1119

[25] Mueen, A., Keogh, E.J.: Online discovery and maintenance of time series motifs. In: International Conference on Knowledge Discovery and Data Mining. pp. 1089–1098. ACM (2010). https://doi.org/10.1145/1835804.1835941

[26] Nassif, A.B., Talib, M.A., Nasir, Q., Dakalbab, F.M.: Machine learning for anomaly detection: A systematic review. Ieee Access **9**, 78658–78700 (2021). https://doi.org/10.1109/ACCESS.2021.3083060

[27] Neves, F., Finamore, A.C., Henriques, R.: Efficient discovery of emerging patternsin heterogeneous spatiotemporal data from mobile sensors. In: MobiQuitous 2020-17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. pp. 158–167. ACM (2020). https://doi.org/10.1145/3448891.3448949

[28] Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-BEATS: neural basis expansion analysis for interpretable time series forecasting. In: International Conference on Learning Representations, ICLR. OpenReview.net (2020). https://doi.org/10.48550/arXiv.1905.10437

[29] Palet, J., Manquinho, V., Henriques, R.: Multiple-input neural networks for time series forecasting incorporating historical and prospective context. Data Mining and Knowledge Discovery **38**(1), 315–341 (2024). https://doi.org/10.1007/S10618-023-00984-Y

[30] Pradeepkumar, D., Ravi, V.: Financial time series prediction: An approach using motif information and neural networks. International Journal of Data Science **5**(1), 79–109 (2020). https://doi.org/10.1504/IJDS.2020.109489

[31] Silva, M.G., Madeira, S.C., Henriques, R.: Actionable descriptors of spatiotemporal urban dynamics from large-scale mobile data: A case study in lisbon city. Environment and Planning B: Urban Analytics and City Science **51**(8), 1725–1741 (2024)

[32] Simma, A., Jordan, M.I.: Modeling events with cascades of poisson processes. In: UAI 2010, Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence. pp. 546–555. AUAI Press (2010), https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=2139&proceeding_id=26

[33] Torkamani, S., Lohweg, V.: Survey on time series motif discovery. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **7**(2) (2017). https://doi.org/10.1002/WIDM.1199

[34] Truong, C.D., Tin, H.N., Anh, D.T.: Combining motif information and neural network for time series prediction. International Journal of Business Intelligence and Data Mining **7**(4), 318–339 (2012). https://doi.org/10.1504/IJBIDM.2012.051734

[35] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems. pp. 5998–6008 (2017), https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[36] Wang, L., Ren, J., Xu, B., Li, J., Luo, W., Xia, F.: MODEL: motif-based deep feature learning for link prediction. Transactions on Computational Social Systems **7**(2), 503–516 (2020). https://doi.org/10.1109/TCSS.2019.2962819

[37] Wong, Y.S., Lee, N.K., Omar, N.: GMFR-CNN: an integration of gapped motif feature representation and deep learning approach for enhancer prediction. In: International Conference on Computational Systems-Biology and Bioinformatics. pp. 41–47. ACM (2016). https://doi.org/10.1145/3029375.3029380

[38] Xiong, H., Capurso, D., Sen, Ś., Segal, M.R.: Sequence-based classification using discriminatory motif feature selection. PloS one **6**(11), e27382 (2011). https://doi.org/10.1371/journal.pone.0027382

[39] Yankov, D., Keogh, E.J., Medina, J., Chiu, B.Y., Zordan, V.B.: Detecting time series motifs under uniform scaling. In: International Conference on Knowledge Discovery and Data Mining. pp. 844–853. ACM (2007). https://doi.org/10.1145/1281192.1281282

[40] Ye, L., Keogh, E.J.: Time series shapelets: a new primitive for data mining. In: SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 947–956. ACM (2009). https://doi.org/10.1145/1557019.1557122

[41] Yeh, C.M., Kavantzas, N., Keogh, E.J.: Matrix profile VI: meaningful multidimensional motif discovery. In: International Conference on Data Mining. pp. 565–574. IEEE Computer Society (2017). https://doi.org/10.1109/ICDM.2017.66

[42] Yeh, C.M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H.A., Silva, D.F., Mueen, A., Keogh, E.J.: Matrix profile I: all pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In: International Conference on Data Mining. pp. 1317–1322. IEEE Computer Society (2016). https://doi.org/10.1109/ICDM.2016.0179

[43] Yu, Y., Si, X., Hu, C., Zhang, J.: A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput. **31**(7), 1235–1270 (2019). https://doi.org/10.1162/NECO_A_01199

[44] Zhang, D., Kabuka, M.R.: Combining weather condition data to predict traffic flow: A GRU based deep learning approach. In: Intl Conference on Dependable, Autonomic and Secure Computing. pp. 1216–1219. IEEE Computer Society (2017). https://doi.org/10.1109/DASC-PICOM-DATACOM-CYBERSCITEC.2017.194

[45] Zhang, Q., Lipani, A., Kirnap, Ö., Yilmaz, E.: Self-attentive hawkes process. In: Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event. Proceedings of Machine Learning Research, vol. 119, pp. 11183–11193. PMLR (2020), http://proceedings.mlr.press/v119/zhang20q.html

[46] Zhao, L.: Event prediction in the big data era: A systematic survey. ACM Computing Surveys **54**(5), 94:1–94:37 (2022). https://doi.org/10.1145/3450287

[47] Zhu, Y., Yeh, C.M., Zimmerman, Z., Kamgar, K., Keogh, E.J.: Matrix profile XI: SCRIMP++: time series motif discovery at interactive speeds. In: International Conference on Data Mining. pp. 837–846. IEEE Computer Society (2018). https://doi.org/10.1109/ICDM.2018.00099

[48] Zhu, Y., Zimmerman, Z., Senobari, N.S., Yeh, C.M., Funning, G.J., Mueen, A., Brisk, P., Keogh, E.J.: Matrix profile II: exploiting a novel algorithm and gpus to break the one hundred million barrier for time series motifs and joins. In: International Conference on Data Mining. pp. 739–748. IEEE Computer Society (2016). https://doi.org/10.1109/ICDM.2016.0085

[49] Zimmerman, Z., Kamgar, K., Senobari, N.S., Crites, B., Funning, G.J., Brisk, P., Keogh, E.J.: Matrix profile XIV: scaling time series motif discovery with gpus to break a quintillion pairwise comparisons a day and beyond. In: Proceedings of the ACM Symposium on Cloud Computing. pp. 74–86. ACM (2019). https://doi.org/10.1145/3357223.3362721