Article:
Huang, Yushan, Zhao, Yuchen, Capstick, Alexander et al. (2024) Analyzing entropy features in time-series data for pattern recognition in neurological conditions. Artificial intelligence in medicine. 102821. ISSN: 0933-3657

https://doi.org/10.1016/j.artmed.2024.102821

Artificial Intelligence In Medicine 150 (2024) 102821

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Research paper

Analyzing entropy features in time-series data for pattern recognition in neurological conditions

Yushan Huang [a,f], Yuchen Zhao [b], Alexander Capstick [d,f], Francesca Palermo [d,f], Hamed Haddadi [c], Payam Barnaghi [d,e,f,g,*]

[a] Dyson School of Design Engineering, Imperial College London, London, UK
[b] Department of Computer Science, University of York, York, UK
[c] Department of Computing, Imperial College London, London, UK
[d] Department of Brain Sciences, Imperial College London, London, UK
[e] The Great Ormond Street Institute of Child Health, University College London, London, UK
[f] Great Ormond Street Hospital for Children, London, UK
[g] Care Research and Technology Centre, The UK Dementia Research Institute, London, UK

ARTICLE INFO

Keywords:
Entropy
Time-series data
Pattern recognition
Neurological conditions

ABSTRACT

In the field of medical diagnosis and patient monitoring, effective pattern recognition in neurological time-series data is essential. Traditional methods predominantly based on statistical or probabilistic learning and inference often struggle with multivariate, multi-source, state-varying, and noisy data while also posing privacy risks due to excessive information collection and modeling. Furthermore, these methods often overlook critical statistical information, such as the distribution of data points and inherent uncertainties. To address these challenges, we introduce an information theory-based pipeline that leverages specialized features to identify patterns in neurological time-series data while minimizing privacy risks. We incorporate various entropy methods based on the characteristics of different scenarios and entropy. For stochastic state transition applications, we incorporate Shannon's entropy, entropy rates, entropy production, and the von Neumann entropy of Markov chains. When state modeling is impractical, we select and employ approximate entropy, increment entropy, dispersion entropy, phase entropy, and slope entropy. The pipeline's effectiveness and scalability are demonstrated through pattern analysis in a dementia care dataset and also an epileptic and a myocardial infarction dataset. The results indicate that our information theory-based pipeline can achieve average performance improvements across various models on the recall rate, F1 score, and accuracy by up to 13.08 percentage points, while enhancing inference efficiency by reducing the number of model parameters by an average of 3.10 times. Thus, our approach opens a promising avenue for improved, efficient, and critical statistical information-considered pattern recognition in medical time-series data.

1. Introduction

With the progress in digital health technologies, an ever-increasing amount of health-related data is being produced, creating unparalleled opportunities for biomedical research and the improvement of healthcare services [1,2]. The amalgamation of real-world patient data with these technologies allows for the analysis and extraction of critical health information, which in turn facilitates more precise clinical decision-making and enhances patient outcomes. However, the analysis of large, multi-sourced, and often noisy datasets introduces significant challenges, particularly regarding privacy issues and the effective recognition of patterns in neurological time-series data [3,4]. These obstacles become even more critical in areas such as dementia, heart disease, and epilepsy analysis, where swift and accurate interpretation of time-series data can greatly influence the quality of patient care.

The analysis of neural time-series data occupies a central role in fields such as neuroscience, medicine, and bioinformatics. This data is typically sourced from electroencephalography (EEG) [5], functional magnetic resonance imaging (fMRI) [6], or other neural recording technologies. It is essential for diagnostic decision-support and healthcare applications. Moreover, there are more generalized forms of time-series data, such as activity data, which are crucial for analyzing neurological conditions like dementia and traumatic brain injury [7–9]. These time-series datasets are characterized by their high dimensionality,

* Corresponding author at: Department of Brain Sciences, Imperial College London, London, UK.
E-mail address: p.barnaghi@imperial.ac.uk (P. Barnaghi).

https://doi.org/10.1016/j.artmed.2024.102821
Received 2 June 2023; Received in revised form 14 February 2024; Accepted 21 February 2024
Available online 22 February 2024

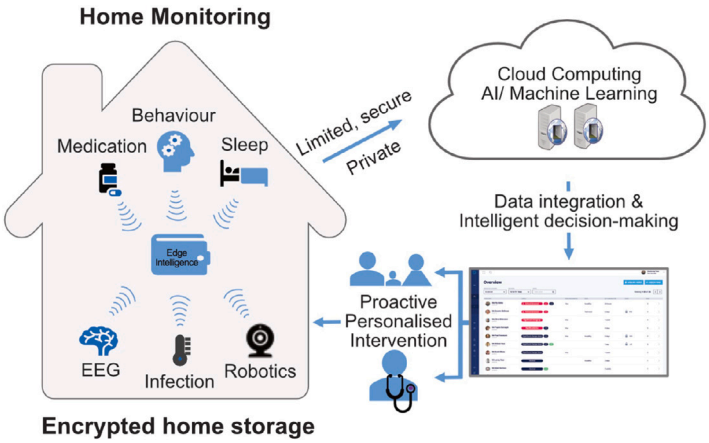


Fig. 1. An overview of the healthcare monitoring IoT platform.

non-linearity, and non-stationarity, which make their analysis and interpretation quite challenging. Traditional linear methods, such as the Fourier transform [10] and autoregressive models [11], can be effective in certain scenarios but often struggle to capture the complex dynamics and non-linear patterns inherent in the data. Consequently, more sophisticated methods, including neural network-based and information theory-based approaches, have become increasingly popular for analyzing neural time-series data. For instance, Recurrent Neural Networks (RNNs) [12] and Long Short-Term Memory networks (LSTMs) [13] have been utilized to handle the data's long-term dependencies and dynamic changes. Additionally, information theory-based methods like entropy and mutual information have been applied to quantify the complexity and uncertainty of the data, uncovering valuable insights hidden within [14,15].

Despite the progress made by information theory-based and neural network methods to analyze neural time series data, they still face significant limitations and challenges when dealing with complex and dynamic data. Most of the current studies are task-specific and lack generalization, requiring specific analysis and algorithm/model design when faced with new tasks or scenarios. On the other hand, some of the methods are highly sensitive to raw data and perform poorly when handling multivariate, multi-source, state-varying, and noisy data. In addition, the existing methods often overlook critical statistical information, such as the distribution of data points and inherent uncertainties in the raw data. The neural network-based methods often embed unnecessary non-task-related information, increasing privacy risks [16]. For example, in Federated Learning (FL), uploading raw time-series data to cloud servers can result in privacy breaches and impose significant transmission bandwidth constraints. Therefore, there is an unmet need to develop more general, robust, and efficient approaches to extract high-level features from multivariate, multi-source, state-varying, and noisy neural time series data. This is crucial to enhance model performance and accuracy, meeting the high requirements for data analysis precision and interpretability in clinical and research settings.

In previous work, we conducted a preliminary analysis of three Markov chain-based entropy features via heat maps [7], highlighting the potential of entropy in analyzing multivariate, multi-source, rapidly state-varying, and noisy neurological time-series data. However, it stopped short of presenting a complete pipeline for analyzing such data and did not validate the effectiveness of these methods with machine learning (ML) models. In this paper, we introduce an information theory-based method for neural time-series analysis within non-centralized medical decision systems, designed to maintain stable performance across multivariate, multi-source, state-varying, and noisy data contexts. The primary contributions of this paper are outlined as follows:

(1) We present an information theory-based method for analyzing neural time-series data, tailored for use within a decentralized healthcare monitoring IoT platform (as depicted in Fig. 1). This method is engineered to consistently perform well across multivariate, multi-source, state-varying, and noisy neural time-series datasets. We specifically differentiate neurological time series data into two categories: state-transition and non-state-transition data. Depending on the scenario and data characteristics, we apply and choose different entropy methods. In our decentralized healthcare monitoring IoT platform, data preprocessing and the extraction of entropy features are performed locally, whereas model execution and medical decision-making are centralized on cloud servers.

(2) Compared to traditional neural network-based methods, the proposed approach can select and extract the most informative, relevant, and representative high-level features from the data. These features are based on information theory, consider critical statistical information, and are easier to understand and interpret, thereby enhancing the interpretability of neural networks. This also mitigates privacy and security concerns because feature extraction is performed locally, eliminating the need to upload raw data to cloud servers. These entropy features enhance prediction accuracy.

(3) Our work showcases broad applicability and generalizability across a variety of datasets. We implement the proposed pipeline on a dementia care dataset derived from our clinical study on remote healthcare monitoring. To further validate its scalability, we test the pipeline on two publicly available datasets to demonstrate the effectiveness across different healthcare contexts: the Epileptic Seizure Recognition Dataset [17] and the PTB Diagnostic ECG Database [18,19].

(4) Our study highlights the performance and efficiency gains achievable with various models through the use of our method. Specifically, we evaluate the effectiveness of the extracted features using different models such as Logistic Regression (LR), Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). Our experimental results show that, for the three datasets, compared to the baseline methods, the information theory-based pipeline can significantly improve the accuracy, recall, and F1 scores of the models by an average of up to 13.08 percentage points (pp), and can simplify the model structure and enhance efficiency, reducing the number of model parameters by an average of 3.10 times.

The remainder of this paper is structured as follows: Section 2 delves into related works and examines the challenges within the scope of this research. Section 3 outlines the technical aspects of our proposed method, including an introduction to several entropy methods related to Markov chains and our feature selection methods. In Section 4, we detail the implementation and evaluation framework for our experiments, covering datasets, models, and performance metrics. Section 5

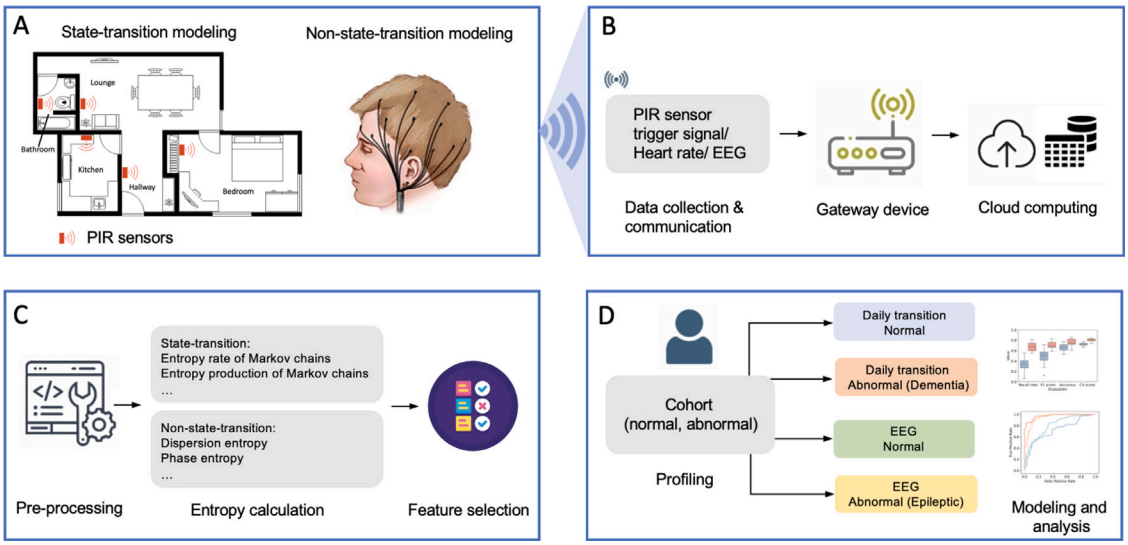


Fig. 2. Method overview.

provides a thorough analysis of the experimental outcomes. Discussion of the method and its limitations is presented in Section 6. The paper concludes in Section 7, summarizing our findings and suggesting directions for future research.

The source code, constructed models, and links to the public datasets are made available on a GitHub repository [20].

2. Related works

Among the solutions developed to tackle these challenges, deep neural networks (DNNs) have gained popularity for their proficiency in learning the spatio-temporal characteristics of data and autonomously extracting features for pattern recognition and prediction outcomes [21]. These models often employ feature extraction techniques such as convolutional neural networks (CNN) and long short-term memory (LSTM) units. Notable innovations include the CNN-LSTM architecture analyzed by Hussain et al. [22], the CNN-LSTM framework augmented with a self-attention mechanism by Park et al. [23], and the integration of transformers and generative adversarial networks (GANs) by Shankar et al. [24]. However, despite their benefits, DNNs can be cumbersome, inefficient, and may not fully leverage critical statistical data insights, such as point distributions and the inherent uncertainties within these distributions, potentially rich in useful, high-level features [25]. Additionally, these models risk privacy breaches by often requiring the input of all raw data for processing, which collects more information than necessary. A further issue is the opacity and limited interpretability of neural network-based methods, hindering scientists' and engineers' deep comprehension of the models and restricting their application in crucial decision-making scenarios, particularly in medicine and healthcare. In these fields, model interpretability is essential, as it significantly influences the confidence and reliance that healthcare professionals and patients have on the model's predictions. This underscores the necessity for methodologies that can adeptly manage complex data, safeguard privacy, and elucidate high-level features, thus facilitating improved decision-making in health-related contexts.

Information theory-based methods are increasingly utilized in analyzing neural time-series data, particularly for feature extraction using mathematical principles combined with machine learning models for data analysis. These approaches enable the identification and extraction of the most informative, relevant, and representative features, which are often more interpretable due to their direct relationship with the data's inherent properties and structure. By simplifying the original data through the extraction of key features, the complexity of the data and the model's "black-box" nature are reduced. This reduction not only streamlines model input but also addresses privacy and security concerns by minimizing the amount of information processed by the network, thereby focusing the model on specific goals. In decentralized machine learning systems, this aspect of privacy protection is accentuated, as only the essential extracted features, rather than raw data, are transmitted to the server side. The concept of entropy, introduced by Shannon to quantify information uncertainty, lays the foundational scientific theory for modern information theory [26]. Following Shannon's entropy, various entropy variants like spectral entropy [27] and sample entropy [28] have been developed. Studies such as those by Nurwulan et al., comparing traditional features with multi-scale entropy (MSE) features extracted from 3-axis acceleration data, have demonstrated MSE's superior performance in KNN and random forest (RF) classifications [29]. Similarly, Bao et al. have shown the effectiveness of frequency-domain entropy features combined with statistical measures like mean, energy, and correlation for building predictive models [30]. These approaches not only enhance learning model performance and safeguard privacy but also often outperform traditional deep neural network models, which is crucial for informed clinical decision-making. However, many existing studies either focus solely on a single entropy feature or use entropy features alongside traditional ones without tailoring the selection to specific data characteristics, such as Markovian systems and stochastic state transitions, or justify the choice of entropy features. Moreover, these methods tend to be task-specific and lack a comprehensive, scalable entropy-based pipeline applicable to various tasks.

A study closely related to ours is the entropy measurement model developed by Howedi et al. [31], which employs approximate entropy (ApEn), sample entropy (SampEn), and fuzzy entropy (FuzzyEn) for detecting visitors in a home environment. However, this study does not tailor the selection of entropy features to the data characteristics, such as Markovian systems and stochastic state transitions, nor does it provide a rationale for the chosen entropy features.

3. Methodology

The process proposed in this paper consists of four primary components: data collection, data pre-processing, feature construction, and modeling, as illustrated in Fig. 2. The initial three stages—data collection, pre-processing, and feature construction are executed locally on an edge device (such as a gateway), whereas the modeling phase takes

Index	Time	Location
0	2021-04-30 06:20:45	bedroom
1	2021-04-30 06:23:12	lounge
2	2021-04-30 06:23:13	hallway
3	2021-04-30 06:23:27	bedroom
4	2021-04-30 06:23:39	kitchen
…	…	…
13	2021-04-30 06:27:19	kitchen

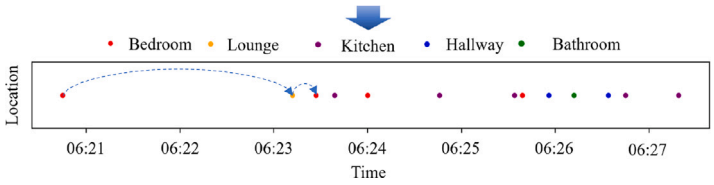


Fig. 3. An example of the state-transition data and Markov chains.

place in the cloud. In detail, sensors gather raw data, which is then forwarded to the edge device. One advantage of this approach is that the processing of the raw data can be conducted locally at the edge, and only the extracted features are transmitted to the cloud, enhancing the privacy protection of medical information. This also maximizes the computational capabilities of both the edge and cloud, improving system efficiency.

A key benefit of this methodology is its local processing of raw data at the edge, permitting only the distilled features to be uploaded to the cloud. This strategy significantly bolsters the privacy protection of medical data. Furthermore, it leverages the computational strengths of both edge and cloud platforms, thereby enhancing the overall efficiency of the system.

3.1. Data collection and pre-processing

Data and activity collection are achieved through various sensors installed in the home. We classify the collected time-series signals into state-transition and non-state-transition based on whether the time-series signals can constitute a state transition. State-transition time-series signals can form state transitions, for example, PIR sensors, as shown in Fig. 2.A, which are placed at different locations in the home and are triggered as someone passes by, record both the location and time simultaneously. As Fig. 3 shows, these points are discrete. Through data pre-processing, the discrete points can be connected to represent transition signals. Therefore, this kind of time-series data is referred to as state-transition signals. In contrast, non-state-transition signals cannot form signals representing state transitions, such as EEG signals. These signals are continuous and challenging to transform into signals that depict state transitions.

In the data pre-processing phase, addressing missing values is our initial step. To handle these, we utilize forward-fill or backward-fill techniques, contingent on their contextual placement and distribution within the dataset. Predominantly, forward-fill is applied, but backward-fill comes into play for missing values situated at the dataset's commencement. This strategy is adopted to maintain the continuity and integrity of the data. Further, we implement label encoding to convert categorical data into a machine-readable format and apply Z-score standardization. The latter process normalizes the data, setting the mean to 0 and standard deviation to 1, thereby optimizing the data for improved model performance. The final step involves aligning the data to ensure uniformity in scale across all features, facilitating smoother subsequent analysis.

3.2. Feature construction

We utilize entropy and its variants to extract features from the raw data. Specifically, for state-transition data, after pre-processing it, we construct entropy features based on the Markov chain. For non-state-transition data, we first construct various entropy features (such as dispersion entropy and phase entropy) and then proceed with feature selection. Next, we provide a detailed introduction to the entropy features based on the Markov chain and the feature selection methods. The construction methods for other entropy features can be found in the Appendix.

3.2.1. State-transition data and Markov chains

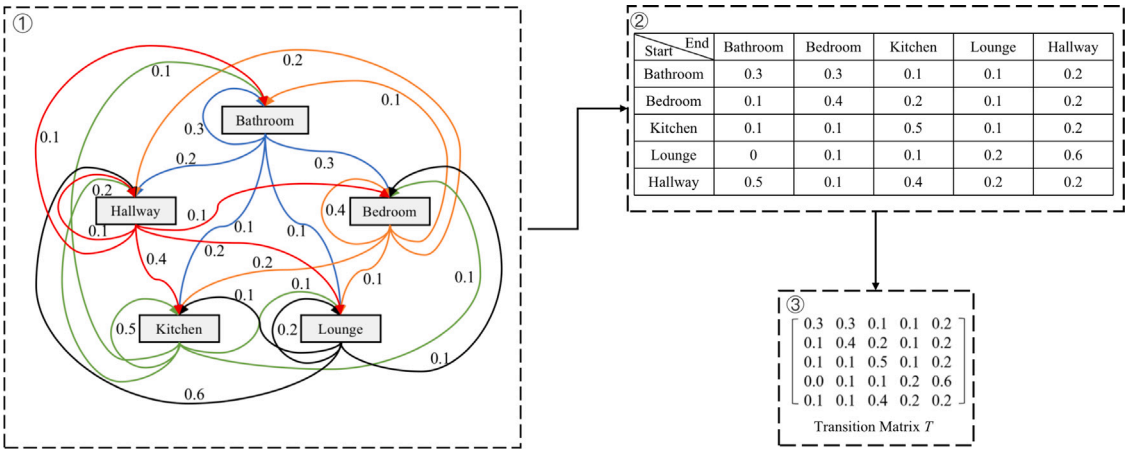
State-transition data, exemplified by the output from PIR sensors, encapsulates both the time and location of sensor activation, thereby providing rich spatial and temporal insights. Spatially, each activation location is considered an individual node, and nodes are sequentially linked according to the order of sensor triggers to construct a pathway, representing a first-order Markov chain, as depicted in Fig. 3. This arrangement allows for the computation of how frequently and how many times a specific location or path is activated. Temporally, the analysis extends to calculating the duration of occupancy at a given location and the transit time between two points, location A to location B. Extracting these high-level features, which encapsulate both spatial and temporal dimensions, presents a challenge for conventional deep learning models due to their complexity. Hence, our feature extraction process leverages the unique properties of state-transition data through the application of Markov chains and entropy techniques.

3.2.2. Shannon's entropy of a Markov chain

We assume that a specific human activity, such as a sequence of locations, could be modeled as a Markov chain. In this model, the occurrences of these activities are treated as random events, and their frequency and pattern could be quantitatively measured. Shannon's entropy is used to effectively represent the complexity and variability in human activity patterns. This choice is motivated by Shannon's entropy's ability to quantify the uncertainty and complexity of information in a data-driven manner, eliminating the need for pre-established assumptions and models. Suppose that there are n locations $X = x_1, x_2, \ldots, x_n$ in a participant's activity, then Shannon's entropy of a Markov chain $H(x)$ can be described as:

$$H(X) = -\sum_{i=1}^{n} P\left(x_i\right) \log P\left(x_i\right) \tag{1}$$

In which $P(X_i)$ is the probability of activity x_i. When the frequency of a participant's activity changes, $H(x)$ will change accordingly to represent the change in activity pattern. A higher value indicates increased uncertainty and complexity, reflecting a more diverse and unpredictable human activity pattern. Conversely, a lower value signifies reduced uncertainty, indicative of a more regular and predictable pattern of behavior.



Start \ End	Bathroom	Bedroom	Kitchen	Lounge	Hallway
Bathroom	0.3	0.3	0.1	0.1	0.2
Bedroom	0.1	0.4	0.2	0.1	0.2
Kitchen	0.1	0.1	0.5	0.1	0.2
Lounge	0	0.1	0.1	0.2	0.6
Hallway	0.5	0.1	0.4	0.2	0.2

Fig. 4. An example of the Entropy rate of a Markov chain. In ①, the rectangular boxes represent the locations (states) in the Markov chain, and the arrows represent the routes between locations in the house. Different colors represent different start locations (blue: bathroom, orange: bedroom, green: kitchen, black: lounge, and red: hallway). The numbers next to the lines represent route probabilities which correspond to the table ② and Transition Matrix T ③. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Algorithm 1 Entropy rate of a Markov chain

Define: $S = \{s_1, s_2, ..., s_L\}$ is a Markov chain trajectory, where L is the length of the trajectory, and $s \in X, X = \{x_1, x_2, ..., x_n\}$, n is the number of states in the Markov chain. TW_1 is the time window required for the stationary distribution, $TW_1 \leq L$. TW_2 is the time window required for the target task, $TW_2 \leq L$. P_{ij} is the probability from state x_i to state x_j. SP is the start point;
Input: Markov chain trajectory S;
Output: Entropy rate ξ of the Markov chain;
Set TW_1 and TW_2;
// Stationary Distribution Function
Function StationaryDistribution(S, TW_1)
  $S_{TW_1} = S[0 : TW_1]$;
  $T = P\left(l'_b = x'_j \mid l'_a = x'_i\right)$, where $l'_a, l'_b \in X$, represent the previous state and the current state, $a' \in [2, TW_1], b' \in [1, TW_1 - 1], x'_i \in X, x'_j \in X$;
  $\pi = \pi T$;
  return π;
// Entropy Rate Function
Function EntropyRate(π, S, TW_2)
for $SP = 0$; $SP + TW_2 \leq L$; $SP = SP + TW_2$ **do**
    $S_{TW_2} = S[SP : TW_2]$;
    $P_{ij} = P\left(l_b = x_j \mid l_a = x_i\right)$, where $l_a, l_b \in X$, represent the previous state and the current state, $a \in [2, TW_2], b \in [1, TW_2 - 1], x_i \in X, x_j \in X$;
    $\xi_m = -\sum_{x_i, x_j \in X}^{n} \pi_i P_{ij} \log P_{ij}$;
end for
  return $\xi = \{\xi_1, \xi_2, ..., \xi_m\}$;
end

3.2.3. Entropy rate of a Markov chain

The adoption of the entropy rate of a first-order Markov chain in our study stems from the inherent limitations associated with Shannon's entropy when applied to human activity data. While Shannon's entropy is instrumental in quantifying the uncertainty or randomness of individual events, it falls short of capturing the sequential and dependent nature of human activities. In scenarios where activities are interconnected, and one event potentially influences the next, a more detailed measure is required to encapsulate the complexity and variability inherent in such data. This motivates our approach to the entropy rate of a Markov chain, a metric that not only accounts for the uncertainty of individual states but also integrates the probabilistic dependencies between consecutive states, offering a holistic view of the dynamism in human activity patterns [32].

Suppose that $X = \{x_1, x_2, \ldots, x_n\}$ represents n states in a Markov chain. Let $x_i, x_j \in X$ represent the previous state and the current state, respectively. Then the probability P_{ij} of the route from x_i to x_j can be represented as:

$$P_{ij} = P\left(x_j \mid x_i\right) \tag{2}$$

Where x_i and $x_j \in X$. Suppose that there are n states in a Markov chain, then the Markov chain can be represented as $n \times n$ matrix $P_{ij\,i,j \in X}$, called Transition Matrix T, an example is shown in Fig. 4. From Markov chains, stationary distributions π can be calculated, which represents:

$$\pi = \pi T \tag{3}$$

In which, π is an n-dimension vector associated with a Markov chain with n states. Using this, the entropy rate of a Markov chain can be expressed as [33]:

$$\xi = -\sum_{ij}^{n} \pi_i P_{ij} \log P_{ij} \tag{4}$$

In which, π_i is the probability in the stationary distribution associated with activity $x_i \in X$ in a Markov chain with the stationary distribution. When calculating the entropy rate of a Markov chain, there are two time-windows that need to be set, one time-window is used to calculate P_{ij} for target time-series data, and the other is used to calculate π_i to represent the characteristics of time-series data. The time window to calculate P_{ij} is set by the task objective. It should be noted that the time window to calculate the stationary distribution π_i is important, as it should reflect the stationary pattern of the participant. For example, participants' routines may be affected by the seasons, thus we need to avoid the possible effects of the seasons when setting up the time windows to calculate the stationary distribution, such as setting the time windows to override the seasonal variations. The dual time-window approach for estimating ensures that the entropy rate is sensitive to both immediate and long-term behavioral patterns, offering a robust metric for analyzing and interpreting the multifaceted nature of human activities. The complete procedure for calculating the Entropy Rate of a Markov Chain is shown in the Algorithm 1.

3.2.4. Entropy production of a Markov chain

In dealing with complex dynamic time-series data, Shannon's entropy and entropy rate of Markov chains still have limitations. Shannon's entropy, while effective in quantifying the uncertainty of information, often overlooks the connectivity and associativity of information. On the other hand, although the entropy rate of Markov chains considers the probability of state transitions, it primarily describes the system's uncertainty and complexity at a statistical level, potentially

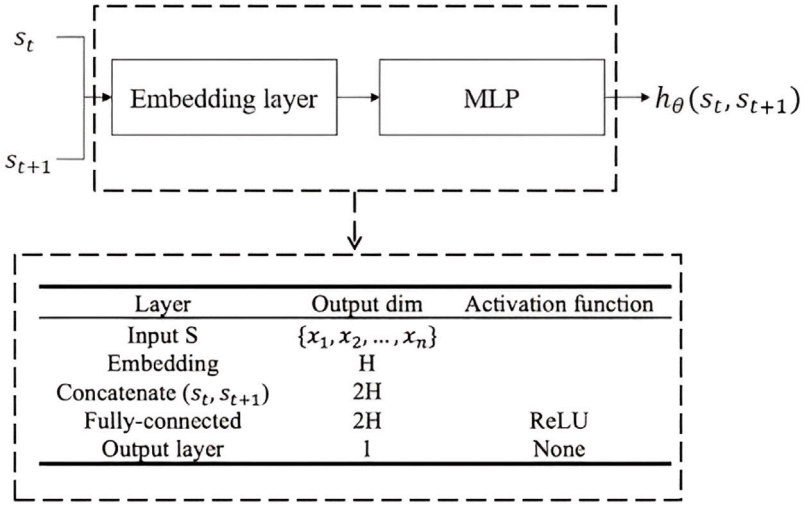


Fig. 5. The model structure of NEEP of Markovian systems. H is the size of the embedding dimension.

failing to unveil the system's internal dynamic evolution and complex interactions in depth. Considering the activity patterns obtained from PIR sensors as an example, dynamic evolution refers to how the behavior and activity patterns change over time and under different conditions. For example, an individual might exhibit different activity patterns during various health conditions.

To overcome these limitations, we consider using entropy production rate (EP) [34]. Originating in the field of physics, particularly thermodynamics and statistical physics, EP not only quantifies the system's uncertainty but also delves into the system's instability and dynamic changes. This makes entropy production a powerful tool capable of revealing the intrinsic dynamics and complex interactions of complicated dynamic systems. Although entropy production has its roots in physics, its core idea focuses on quantifying and analyzing the instability and dynamic changes of a system. In pattern recognition, each person's activity patterns can be viewed as a complex dynamic system [35,36]. By calculating the entropy production of these systems, we can unveil the complex dynamic patterns hidden in spatiotemporal data and deeply understand the driving factors behind these patterns.

EP can be estimated by machine learning models such as the Neural Estimator for Entropy Production (NEEP), which can estimate EP of Markovian systems [37]. Given a Markov chain trajectory $S = \{s_1, s_2, \ldots, s_L\}$ and a function h_θ acting over previous state s_t and the current state s_{t+1} in the Markov chain, where θ denotes the trainable neural network parameters, then the output of NEEP can be defined as [37]:

$$\hat{J}(\theta) = \sum_{t \in L} \left[\Delta S_\theta \left(s_t, s_{t+1} \right) - e^{-\Delta S_\theta (s_t, s_{t+1})} \right] \tag{5}$$

Where ΔS_θ is:

$$\Delta S_\theta \left(s_t, s_{t+1} \right) \equiv h_\theta \left(s_t, s_{t+1} \right) - h_\theta \left(s_{t+1}, s_t \right) \tag{6}$$

The model structure of NEEP is shown in Fig. 5 and the procedure for training NEEP is shown in Algorithm 2. In NEEP, an embedding layer is used to transform the discrete state into a trainable continuous vector [37], and then the embedded data is input into a hidden MLP layer. It has to be noted that, the length of the time series data is very important when training NEEP, as we need to ensure that the data for this period of time is sufficient for training and can reflect the participant's characteristics.

Algorithm 2 Training process of NEEP

1: Define: $S = \{s_1, s_2, \ldots, s_L\}$ is a Markov chain trajectory, where L is the length of the trajectory, and $s \in X, X = \{x_1, x_2, \ldots, x_n\}$, n is the number of states in the Markov chain.
Input: Markov chain trajectory S;
Output: The values calculated by the loss function $\hat{J}(\theta)$;
2: **while** termination condition not met **do**
3: Embedding layer;
4: Objective function

$$\hat{J}(\theta) = \sum_{t \in L} \left[\Delta S_\theta \left(s_t, s_{t+1} \right) - e^{-\Delta S_\theta (s_t, s_{t+1})} \right] \tag{7}$$

5: Compute gradients $\nabla_\theta \hat{J}(\theta)$;
6: Update parameters θ;
7: **end while**

3.2.5. Von Neumann entropy of a Markov chain

The von Neumann entropy (VNE) originates from quantum mechanics. In quantum mechanics, VNE is employed to quantitatively depict the uncertainty of a quantum state [38]. Compared to other entropy measures (such as Shannon's entropy), von Neumann entropy is based on a matrix form. This means that it can not only capture the characteristics of a single variable in the time series but also deeply grasp the interactions and relationships between different variables. This integrative nature displays its strength in capturing complex data structures.

Moreover, VNE possesses robustness, especially when dealing with continuous values and limited data samples. The robust nature of von Neumann entropy primarily stems from the density matrix in its computational methodology. By building an appropriate density matrix, we can capture the relationships and interactions between these states. This allows VNE to consider the continuity and overall characteristics of the data. Even if the data is noisy or incomplete, it can provide a suitable approximation of the system's overall state. This approach ensures that, even when faced with minor perturbations or incomplete data, VNE can deliver stable and consistent results.

As previously mentioned, we can regard each person's activity patterns as a complex dynamic system. Therefore, VNE offers us a new perspective to analyze time-series data. Taking PIR sensors as an example, for the complex time-series data collected in household settings, we can approach from both the perspectives of activity frequency and activity duration, as shown in Fig. 6. This allows us to analyze time-series data from both spatial and temporal perspectives. In addition, due to the presence of the density matrix, VNE can also conduct a comprehensive analysis of the data from both horizontal and vertical perspectives.

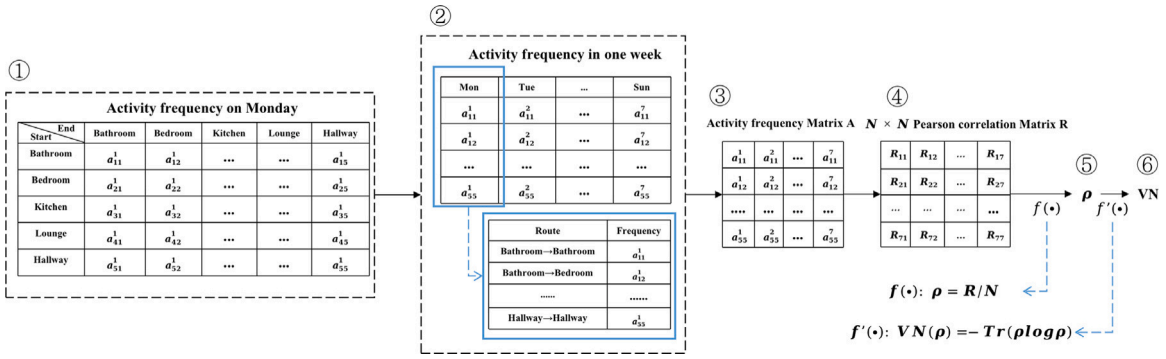


Fig. 6. An example for the von Neumann entropy of a Markov chain. Suppose that there are five locations (states) in a Markov chain, and we plan to calculate the von Neumann entropy of one week. From the perspective of spatial, ①: count the frequency $a_{i,j}^{d}$ of different routes in a Markov chain for each day of the week, in which i and j represent the previous location and the current location, respectively; ②: aggregate weekly activity frequency; ③: transfer the weekly activity frequency to the activity frequency matrix A; ④: calculate the Pearson correlation Matrix R between each day; ⑤: calculate the density matrix ρ by $f(\bullet)$; ⑥: calculate the von Neumann entropy by $f'(\bullet)$. And from the perspective of temporal, the only difference is changing the activity frequency to activity duration.

Given a density operator ρ with N eigenvalues $\lambda_{1,\dots,n}$, VNE is defined as follows:

$$S(\rho) = -\operatorname{tr}(\boldsymbol{\rho} \log \boldsymbol{\rho}) = -\sum_{j=1}^{N} \lambda_j \log \lambda_j \tag{8}$$

One of the key points to calculate VNE is to obtain the density operator ρ, which must satisfy (i) be Hermitian, (ii) have unit trace, and (iii) be positive semi-definite. Given $\boldsymbol{R} \in \mathbb{R}^N$, an N-dimension Pearson correlation matrix of the human activity data, then the density operator $\boldsymbol{\rho}$ can be defined as [39]:

$$\boldsymbol{\rho} = \boldsymbol{R}/N \tag{9}$$

The density operator ρ, calculated by Eq. (9) satisfies all the requirements. However, it has to be noted that the density operator ρ, which is calculated from real-world data, may be sparse, and thus, there may be anomalies in the calculation of $log\rho$ using standard classical mathematical methods. Therefore, we calculate $\log \rho$ by Mercator's Series. Suppose B is a matrix and sufficiently close to the identity matrix I, and $\|B - I\| < 1$, then a logarithm of B can be computed by means of the following k-power series [40]:

$$\log(B) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(B-I)^k}{k} \tag{10}$$

This means we can obtain $\log \rho$ by:

$$\log(\boldsymbol{\rho}) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(\boldsymbol{\rho}-I)^k}{k} \tag{11}$$

Integrating Eq. (8), Eqs. (9) and (10), the VNE can be obtained. The complete procedure for calculating VNE is shown in the Algorithm 3.

3.3. Feature selection

Feature selection has a crucial role in modeling by identifying relevant variables that contribute to the predictive power of the model while reducing dimensionality and computational cost. When the dataset allows for state-transition modeling, we prioritize the entropy features associated with Markov chains, including Shannon's entropy, entropy rate, entropy production, and von Neumann entropy of Markov chains. Conversely, for non-state-transition modeling, we deploy a feature selection methodology utilizing mutual information (MI) and the Pearson correlation coefficient (PCC) to select appropriate entropy features from several entropy methods. The detailed steps of the feature selection process are shown in Algorithm 4.

Initially, features are ranked based on the MI between each feature f_i and the target label T. This ranking reflects the mutual dependency of variables, which is essential for unraveling complex and non-linear relationships. The MI for each feature is computed as:

$$I(f_i; T) = \sum_{t \in T} \sum_{f_i \in F} p(f_i, t) \log \left(\frac{p(f_i, t)}{p(f_i)p(t)} \right), \tag{12}$$

Algorithm 3 von Neumann entropy of a Markov chain

Define: $S = \{s_1, s_2, ..., s_L\}$ is a Markov chain trajectory, where L is the length of the trajectory, and $s \in X, X = \{x_1, x_2, ..., x_n\}$, n is the number of states in the Markov chain. VN is the von Neumann entropy of a Markov chain. TW_2 is the time window required for the target task, where $TW_2 <= L$. SP is the start point;
Input: Markov chain trajectory S;
Output: The VN;
Set TW_2;
for $SP = 0$; $SP + TW_2 <= L$; $SP = SP + TW_2$ **do**
  Calculate original matrix A (e.g., activity frequency matrix);
  $N \times N$ Pearson correlation Matrix R of A;
  Density operator $\rho \leftarrow \rho = R/N$;
  von Neumann entropy $VN(\rho) \leftarrow VN(\rho) = \mathrm{Tr}(\rho \log \rho), \log \rho = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(B-I)^k}{k}$;
end for

Algorithm 4 Feature Selection Using MI and PCC

Define: Dataset D with features F and target variable T. Mutual information I, Pearson correlation coefficient ρ, and selection threshold $\theta = 0.8$.
Input: Dataset D, Number of features to select N;
Output: Set of selected features F_s;
Calculate MI for each feature f_i in F with the target T using:
$I(f_i; T) = \sum_{t \in T} \sum_{f_i \in F} p(f_i, t) \log \left(\frac{p(f_i, t)}{p(f_i)p(t)} \right)$
Rank features in F based on calculated MI, resulting in ranked list F_{ranked}
$F_s \leftarrow \emptyset$, $F_{selected} \leftarrow$ first N features from F_{ranked}
while the number of features in F_s is less than N **do**
  $F_{temp} \leftarrow F_{selected}$
  for each pair (f_i, f_j) in $F_{selected}$ where $i \neq j$ **do**
    Calculate Pearson correlation coefficient ρ for f_i and f_j using:
    $\rho(f_i, f_j) = \frac{\sum(f_{ik} - \bar{f}_i)(f_{jk} - \bar{f}_j)}{\sqrt{\sum(f_{ik} - \bar{f}_i)^2} \sqrt{\sum(f_{jk} - \bar{f}_j)^2}}$
    if $\rho(f_i, f_j) > \theta$ **then**
      Remove the feature with lower MI from F_{temp}
    end if
  end for
  $F_s \leftarrow F_{temp}$
  $F_{selected} \leftarrow$ next (N − number of features in F_s) features from F_{ranked}
end while
Return F_s;

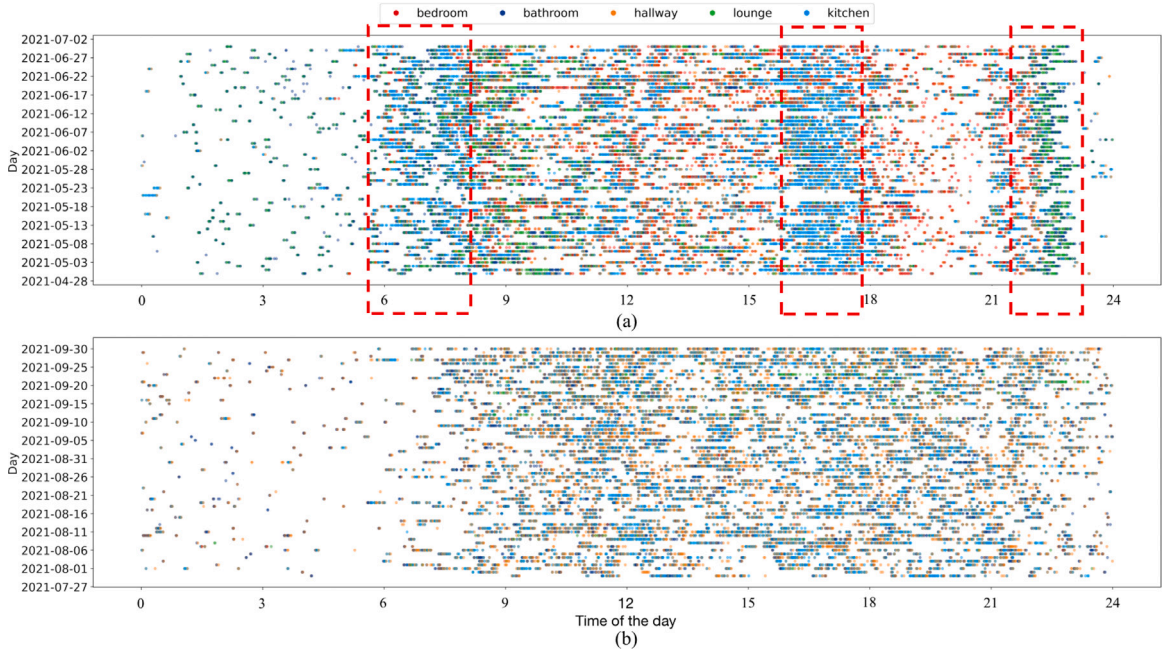


Fig. 7. An example of a PLWD with clear routine activities (a) and another PLWD with fewer routine activities (b). The participant with more routine activities tends to have a more consistent daily activity pattern at the same time each day, as shown in the red boxes. The x-axis shows the time of the day, the y-axis shows different days, and the different colors represent different locations in the house. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Where f_i represents the ith feature and T is the target label. This ranking process is fundamental as it guides the subsequent selection of features by highlighting their respective importance.

Upon ranking the features, we select the top N features for further analysis. To address potential multicollinearity, we calculate the pairwise PCC among these features. If the correlation coefficient between any two features f_i and f_j exceeds the empirically determined threshold, we retain only the feature with the higher MI score and select the next feature in the rank until the top N features are distinct and less correlated. The PCC between two features f_i and f_j can be calculated as:

$$\rho(f_i, f_j) = \frac{\sum(f_i - \bar{f}_i)(f_j - \bar{f}_j)}{\sqrt{\sum(f_i - \bar{f}_i)^2}\sqrt{\sum(f_j - \bar{f}_j)^2}}, i \neq j \tag{13}$$

Where ρ denotes the PCC between features f_i and f_j.

In this paper, the default threshold is 0.8, following the convention established in prior research where it is often considered a cutoff point for high correlation [41,42]. This value is widely recognized as a practical balance between including informative features and excluding redundant ones. Our approach ensures that the retained features are not only relevant to the target but also provide unique information, thereby mitigating the risk of multicollinearity.

4. Implementation & evaluation setup

We have implemented our entropy-based analysis framework on three distinct datasets: a dementia care dataset from the ongoing UK Dementia Research Institute's Minder study, and two publicly available datasets on epileptic seizure (ESRD) and heart disease (PTBDB) focusing on EEG and ECG signals, respectively. In terms of model selection, our analysis encompasses Logistic Regression (LR), Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a combined CNN-LSTM architecture (further details are provided in Section 4.2). The datasets are partitioned into training, validation, and test sets, following a 70%, 15%, 15% split ratio, respectively. To ensure the reliability of our findings, all experiments are conducted 30 times, with both the mean and standard deviation of the results being computed.

4.1. Datasets

The datasets we used in our experiment are: the Minder dataset, collecting activity data of the people living with dementia (PLWD) [33]; the epileptic seizure recognition dataset (ESRD), collecting EEG data [17]; and the PTB diagnostic ECG database (PTBDB), collecting ECG data [18,19].

4.1.1. Minder dataset

We use data from an in-home monitoring study (illustrated in Fig. 1), called Minder, to support PLWD [33]. The Minder platform collects various digital markers, including activity data, home device usage, and clinical information. The Minder study protocol received ethical approval from the London-Surrey Borders Research Ethics Committee and South West London Ethics Committee (linkhere) and we obtained informed written consent from all study participants.

The dataset is labeled by our monitoring team in response to alerts generated by the Minder platform, which operates 24/7. These alerts are confirmed with either the person living with dementia (PLWD) or their caregivers, ensuring the accuracy of information on potential health-related events such as falls, abnormal motor behavior, hospital admissions, urinary tract infections, and symptoms of anxiety, depression, agitation, confusion, and disturbed sleep patterns. Data corresponding to participants who have experienced such events are specifically labeled to indicate these adverse health occurrences.

In our research, we concentrate on analyzing Minder's activity data, which falls under the category of state-transition data. This dataset encompasses 3,762 person-weeks of information, gathered from December 2020 to March 2022. The average age of the study participants is 79 years. It is important to note that all data used in this study has been anonymized to protect participant privacy.

The activity data in the Minder platform is collected using PIR sensors installed in various locations, including the kitchen, bathroom, bedroom, lounge, and hallway, as shown in Fig. 2.A. The PIR sensor logs an event with seconds precision when a person passes by. The recorded data shows the household's life patterns over time; an example of raw data is shown in Fig. 7, which compares the routine activities of two PLWDs.

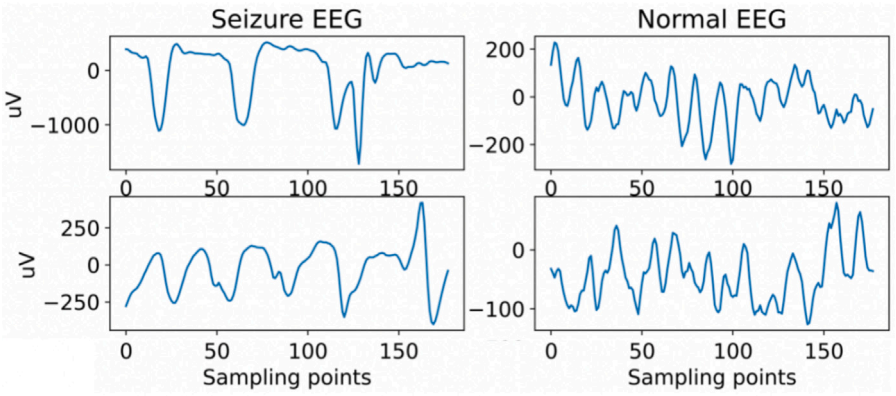


Fig. 8. Visualization of ESRD. The x-axis represents sampling points, and the y-axis represents EEG signals (μV).

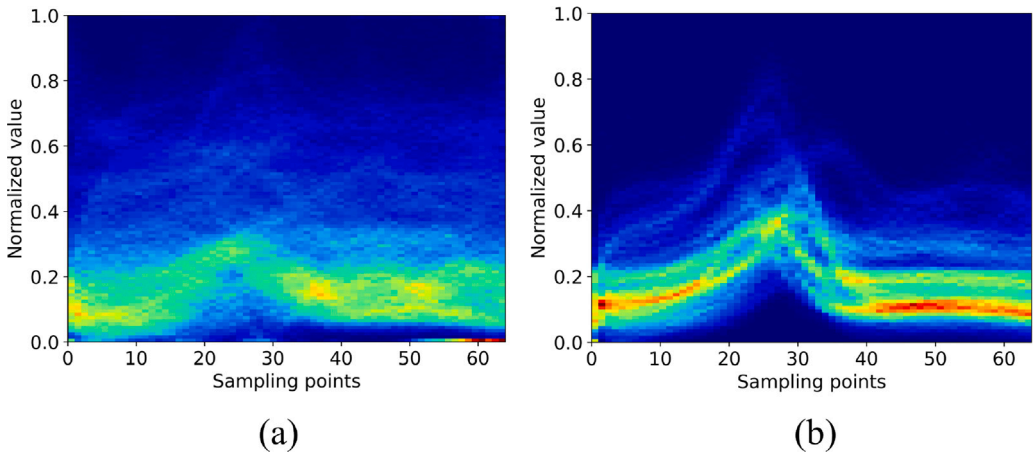


Fig. 9. The histogram color maps for PTBDB are marked as abnormal (a) and normal (b). The x-axis represents sampling points, and the y-axis represents the normalized value of the heartbeat.

4.1.2. Epileptic seizure recognition dataset

The Epileptic Seizure Recognition Dataset (ESRD) comprises 11,500 time-series EEG signal samples from 500 subjects, designed for analyzing EEG signal variations during seizures, categorizing it as non-state-transition data [17]. Each sample is divided into 23 segments, with each segment containing 178 data points spanning one second. The UCI has processed the original dataset, shuffling the segments to create the 11,500 time-series EEG signal samples. An illustrative visualization of ESRD is presented in Fig. 8. This dataset encapsulates five distinct health states, one of which is specific to epileptic seizures, while the remaining four represent normal conditions without epilepsy symptoms. Despite the dataset's potential for multi-class analysis, many researchers opt for binary classification, focusing on differentiating class 1 (epileptic seizures) from the other classes. Similarly, our research aims to distinguish between individuals with and without epileptic seizures.

4.1.3. PTB diagnostic ECG database

To further assess the applicability and scalability of our analysis pipeline, we evaluated it using the PTB Diagnostic ECG Database (PTBDB), which consists of 549 records from 290 subjects (n = 209 male, and 81 female), classified as non-state-transition data [18,19]. The participants' ages range from 17 to 87 years, with an average age of 57.2 years. The data is recorded at a sampling frequency of 125 Hz. The Diagnostic class encompasses a variety of conditions including myocardial infarction, cardiomyopathy/heart failure, bundle branch block, dysrhythmia, myocardial hypertrophy, valvular heart disease, myocarditis, among others, as well as healthy controls. For this study, we specifically extract heartbeat signals from ECG lead 2 [43], concentrating on the myocardial infarction and healthy control groups. The dataset comprises a total of 14,552 samples. Color maps in histograms depicting the PTB data as abnormal and normal are illustrated in Fig. 9.

4.2. Models and performance metrics

In this section, we introduce models of the state-transition data, models of the non-state-transition data, and the performance metrics.

4.2.1. Models of the state-transition data

For the model of the state-transition data, after feature extraction, we utilize LR, SVM, MLP, and LSTM as classifiers, following the standard procedure in the domain of medical feature extraction [44–46].

To evaluate the performance of our method on both linear and non-linear classifiers, we utilize LR and SVM (with a linear kernel) as linear classifiers and MLP and LSTM as non-linear classifiers. To optimize the parameter configurations for the LR and SVM models, we employ a grid search strategy. This strategy is implemented using the $GridSearchCV$ tool from the $scikit-learn$ library in $Python$. For LR, the grid search covers multiple candidate values for the regularization strength parameter C (0.001, 0.01, 0.1, 1, 10, 100), regularization methods (L1 and L2), and solver (liblinear, newton-cg, lbfgs, sag, saga). For SVM, we consider different C values (0.001, 0.01, 0.1, 1, 10, 100).

To optimize MLP and LSTM, we used a strategy that combined temperature warm-up and cosine annealing to optimize the learning rate. For the network structures of MLP and LSTM, to ensure a fair comparison, we set the same number of hidden layers (two hidden

layers). For the number of nodes, we scale them proportionally to the input size to ensure a fair comparison. For example, if the input size of the baseline is a, the hidden size of the baseline is b, the input size of the entropy-based method is c, and the hidden size of the entropy-based method is d, we ensure that the ratio of a/c is consistent with b/d. We utilize Binary Cross-Entropy loss function, and SGD optimizer. Considering the effect of the month, we set the time step of LSTM to 4 weeks of data.

We divide the entire dataset into three parts: a training set, a validation set, and a test set. The specific split ratio is 70% for the training set, 15% for the validation set, and 15% for the test set. The training set is used for model training, the validation set is used for tuning model parameters, and the test set is used to evaluate the final performance of the models.

• Baseline. For the state-transition data (Minder dataset), there is limited research on feature extraction for PIR sensors. We follow the research of Chimamiwa et al. [47], utilize the average frequency and average duration of different activities as the baseline, and also split one day into daytime (06:00–18:00) and nighttime (18:00–24:00 and 00:00–06:00). The extracted feature is then fed into the classifier including LR, SVM, MLP, and LSTM.

• Entropy-Based. For the state-transition data, we prioritize the entropy features associated with Markov chains, including Shannon's entropy of Markov chains, entropy rate of Markov chains, EP of Markov chains, VNE of Markov chains (activity frequency), VNE of Markov chains (activity duration), and activity duration difference of Markov chains in each week (daytime and nighttime).

4.2.2. Models of the non-state-transition data

For non-state transition data (ESRD and PTBDB dataset), since deep model-based feature extraction methods are popular recently, we utilize CNN-based feature, LSTM-based feature, and CNN-LSTM-based feature as the baseline. Considering that these three approaches utilize deep learning models and to ensure experimental consistency, fairness, and effective comparisons, our entropy-based method employs MLP as the classifier. Additionally, MLP offers advantages in terms of computational efficiency and model simplification. All these deep learning models utilize the strategy of combining temperature warm-up and cosine annealing to optimize the learning rate.

• Baseline. For the CNN-based features, when dealing with the ESRD and PTBDB datasets, we employ a 1D-CNN model. This model design is followed by Khalilpour et al.'s research [48], and we optimize its structure to suit our data better. The raw data is used as input to this model. Similarly, for LSTM-based features, we reference the LSTM model proposed by Farisi et al. [49]. This model also takes the raw data as input, and we make adjustments to it to better align with our data. For the CNN-LSTM-based feature analysis, we follow the model proposed by Hussain et al. [22]. Specifically, they propose a CNN-LSTM model, in which the CNN is used for front-end feature extraction, while the LSTM is employed for learning temporal patterns in the back end. To better extract features and capture non-stationary and time-varying information, they transform the original signal into time–frequency domain signals through a five-level decomposition and a one-level approximate discrete wavelet transform (DWT). The loss function and optimizer of the above models are kept consistent with the state-transition model.

• Entropy-Based (Ours). For non-state-transition data, we first calculate several entropy features such as increment entropy, approximate entropy, and slope entropy; we then utilize MI and PCC for feature selection (as illustrated in Section. 3.3). After feature extraction, we add an MLP classifier. The model structure of the MLP is designed to be as simple as possible while maintaining model performance. The loss function and optimizer are kept consistent with the baseline model.

4.2.3. Performance metrics

To evaluate the performance of these models comprehensively, we utilize accuracy, F1 score, recall, and Area Under the Curve (AUC) as the evaluation metrics. Accuracy intuitively reflects the overall classification capability of the models, while F1 score and recall provide a deeper understanding of the model performance in binary classification, especially important in medical scenarios. Additionally, AUC, as an important metric, can reflect the model's ability to differentiate between positive and negative classes at different thresholds. It takes into account the model's sensitivity and specificity and serves as an effective tool for assessing the accuracy of model predictions in terms of probabilities. The combined use of these metrics allows us to comprehensively assess the performance of the models from multiple dimensions, ensuring a comprehensive understanding of model effectiveness. In addition, for the evaluation of the non-state-transition data, since we compare the entropy-based method to the deep models, we also calculate the model parameter count by the $summary()$ function from $TensorFlow$ or by iterating through the model in $PyTorch$.

5. Evaluation results

We evaluate our method on three datasets: Minder, ESRD, and PTBDB. Minder is a state-transition dataset, ESRD and PTBDB are non-state-transition datasets. We utilize accuracy, F1 score, recall, and ROC-AUC as the evaluation metrics.

5.1. Minder database

Minder data represents a state-transition model in which each location is a node/state, and movements between locations are represented as transitions. We prioritize the entropy features associated with Markov chains. Additionally, we consider the effect of sundowning and circadian rhythms in people living with dementia (PLWD) [50] by dividing one day into two time periods: daytime (06:00–18:00) and night (18:00–24:00 and 00:00–6:00). Following these configurations, the baseline features are the average frequency and average duration of different activities in each week (daytime and nighttime) [47]. The output of the models is healthcare-related events (True or False).

The entropy features are Shannon's entropy of Markov chains, Entropy rate of Markov chains, EP of Markov chains, VNE of Markov chains (activity frequency), VNE of Markov chains (activity duration), and activity duration difference of Markov chains in each week (daytime and nighttime). Since the frequency of our dataset labels is one week (indicating whether there are any anomalies within the week), we calculate a series of entropy values by the week. For Shannon's entropy, we first calculate the entropy of each day (distinguishing between daytime and nighttime) and then compute the weekly average value. For EP, as it is calculated by machine learning models, for each participant, we first train the model using the training data and then calculate the EP for each week. For the entropy rate, to represent the stationary pattern and avoid the impact of seasonal changes, we set the time window to four months to calculate the stationary distribution π, with the target time window being one day, and then calculate the weekly average value. For VNE, we calculate the frequency (or duration) of various activities within a day and then compute the VNE for each week. The output of the models is healthcare-related events (True or False).

Fig. 10, Fig. 11, and Table 1 show the evaluation results of Minder. The entropy-based method exhibits effectiveness across different models. Through comprehensive analysis, we find that the entropy-based method can significantly improve performance for both linear models, such as LR and SVM (linear kernel), as well as non-linear models like MLP and LSTM. Specifically, we observe an average increase of 13.08pp in recall rate, 10.80pp in F1 score, and 7.88pp in accuracy.

This performance improvement can be attributed to the enhanced ability of the entropy-based method to capture the complexity of patterns and temporal dependencies. Traditional baseline features, such

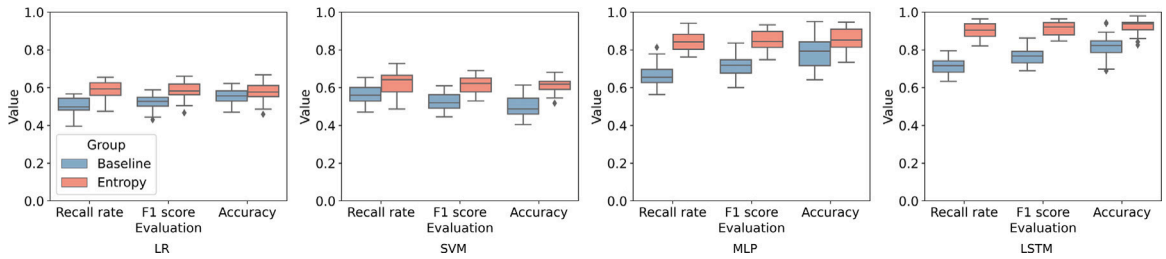


Fig. 10. The evaluation results of the Minder database. The x-axis represents different evaluation metrics. We can find that, for four different models, compared with the baseline features, modeling by the entropy features can improve the recall rate, F1 score, and accuracy.

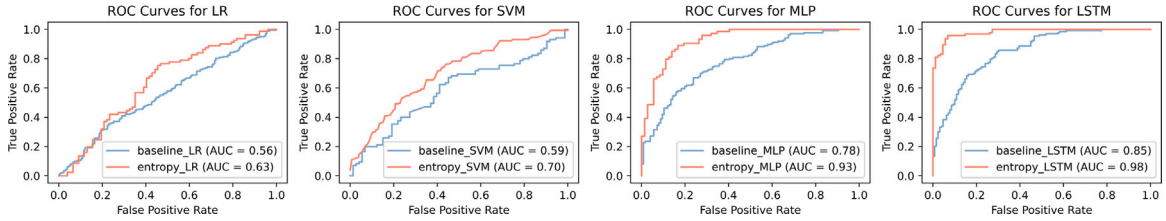


Fig. 11. The ROC-AUC curves of the Minder database.

as average activity frequency and duration, may have limitations in revealing complex data structures with temporal dependencies. In contrast, the entropy-based method provides a powerful way to deepen the understanding of intrinsic dynamic variations in data by quantifying the uncertainty and randomness of state transitions in Markov chains. Measures of entropy such as Shannon's entropy, entropy rate, EP, and VNE not only capture the probability distribution of activities but also reflect the diversity and uncertainty of system states, helping the model to uncover nonlinear relationships and complex interactions in the data more effectively.

The entropy-based method performed better on non-linear models such as LSTM and MLP. For example, after adopting the entropy-based method, the LSTM model shows an 18.78pp increase in recall rate, 15.12pp in F1 score, and 10.69pp in accuracy. Similarly, the MLP model also exhibits an 18.02pp increase in recall rate, 13.42pp in F1 score, and 7.51pp in accuracy. This is because these non-linear models have multi-layer processing and memory capabilities, significantly improving the performance of complex pattern recognition and the utilization of high-dimensional features. This enhancement is particularly important when dealing with multi-variate, multi-source, state-varying, and noisy time-series data, highlighting the potential of the entropy-based method in time-series analysis.

Furthermore, we evaluate the model performance using ROC-AUC curves. As shown in Fig. 11, all models exhibit an improvement in AUC values after utilizing entropy-based features. In the LR model, the introduction of entropy-based features increases the AUC value from 0.56 to 0.63, while the SVM model's AUC value improves from 0.59 to 0.70. In the MLP model, the entropy method improves the AUC value from 0.78 to 0.93, indicating its sufficient performance in distinguishing between positive and negative classes. The LSTM model also achieves an AUC value of 0.98 after adopting entropy features. These improvements underscore the effectiveness of the entropy-based method in processing complex time-series data. Consistent with our previous analysis, the entropy-based method significantly enhances the model's ability to analyze time-series data.

5.2. Epileptic seizure recognition dataset (ESRD)

ESRD is a non-state-transition dataset; we aim to differentiate between the normal participants and those with epileptic seizures. Due to the popularity of deep model-based feature extraction methods in the analysis of EEG and ECG data, we use CNN, LSTM, and CNN-LSTM models as baselines. All of these baseline models are derived from state-of-the-art (SOTA) research [22,48,49] and are adjusted to better fit the characteristics of our dataset. We perform Z-score normalization and alignment on the raw data. For CNN-based and LSTM-based models, the input to the models is the pre-processed data. For the CNN-LSTM-based model, in addition to data pre-processing, we also conduct DWT transformations and utilize the transformed signals as model inputs. The output of the models is participants with epileptic seizures (True or False).

Table 1
The average performance of the models for the Minder.

	Evaluation	Baseline	Entropy	Improvement
LR	Recall rate	50.13 ± 4.68%	58.97 ± 4.61%	8.84pp
	F1 score	52.63 ± 4.11%	58.46 ± 4.42%	5.83pp
	Accuracy	55.52 ± 4.13%	58.03 ± 4.71%	2.51pp
SVM	Recall rate	56.05 ± 4.60%	62.72 ± 6.56%	6.67pp
	F1 score	52.82 ± 4.75%	61.63 ± 4.30%	8.81pp
	Accuracy	50.12 ± 5.75%	60.94 ± 4.03%	10.82pp
MLP	Recall rate	66.14 ± 5.70%	84.16 ± 5.21%	18.02pp
	F1 score	71.55 ± 5.61%	84.97 ± 4.97%	13.42pp
	Accuracy	78.37 ± 8.11%	85.88 ± 5.51%	7.51pp
LSTM	Recall rate	71.51 ± 4.04%	90.29 ± 4.41%	18.78pp
	F1 score	76.17 ± 4.27%	91.29 ± 3.72%	15.12pp
	Accuracy	81.72 ± 6.41%	92.41 ± 4.18%	10.69pp
Average	Recall rate	–	–	**13.08**pp
	F1 score	–	–	**10.80**pp
	Accuracy	–	–	**7.88**pp

For the entropy-based method, we first calculate a series of entropy features, then employ feature selection by MI and PCC. The feature selection result is shown in Fig. 12. Based on the MI, we first select the top four features, including increment entropy, approximate entropy, slope entropy, and sample entropy. Subsequently, to avoid feature collinearity and redundancy, we conduct a secondary selection of the selected features based on the PCC. We observe that the sample entropy and approximate entropy exhibit a relatively high PCC (higher than 0.8), with approximate entropy having a higher MI score. Therefore, we retain approximate entropy and, in descending order according to MI ranking, select phase entropy as the next feature to keep. Afterwards, by checking the PCC again, we find that all the selected features have no significant collinearity or redundancy and have sufficient MI scores. Therefore, the selected entropy-based features for ESRD are increment entropy, approximate entropy, slope entropy, and phase entropy. Since

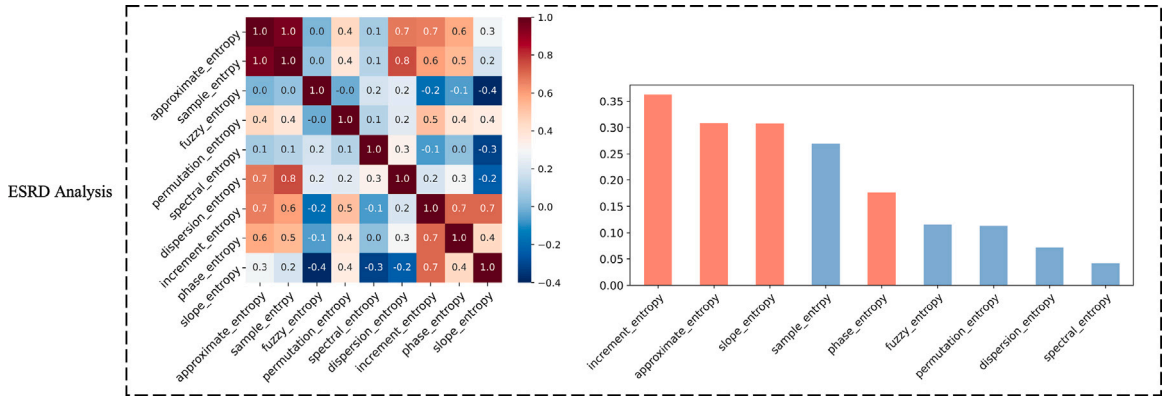


Fig. 12. The feature selection result of ESRD, including Pearson relationship matrices (the left) and mutual information (the right). In the right figure, the selected features are in red while the unselected features are in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

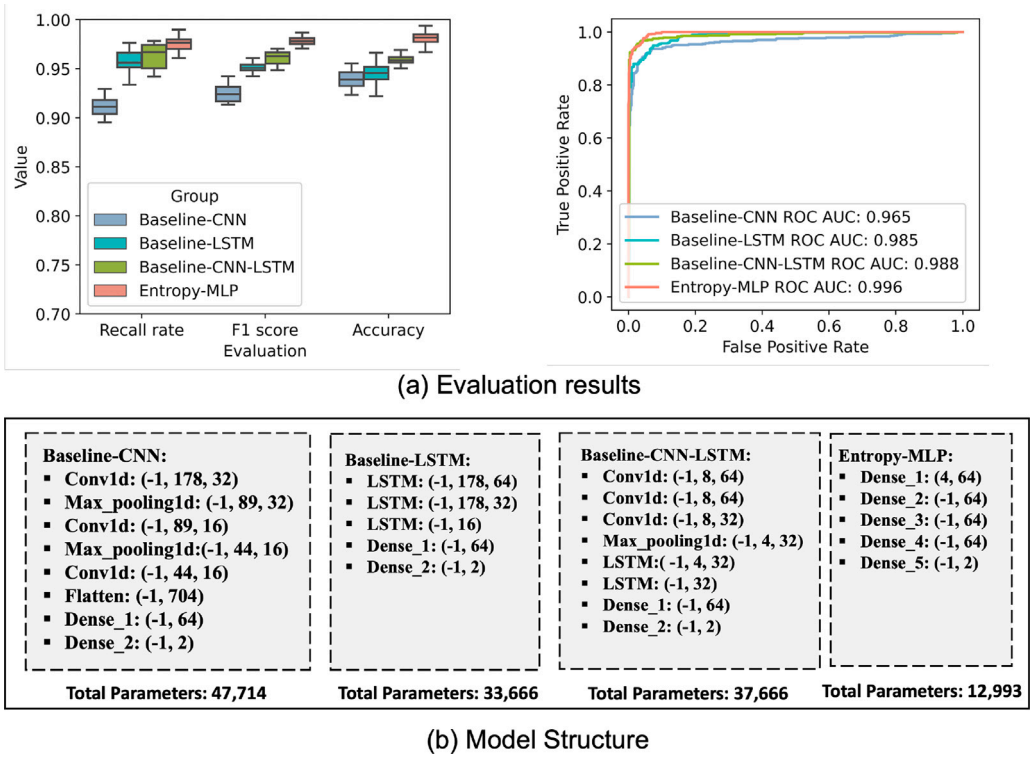


Fig. 13. The results of ESRD. (a) is the evaluation results, including recall rate, F1 score, accuracy, and ROC-AUC. (b) is the comparison of the model structure between the baseline and entropy models. As we utilize pre-processed data with less data noise, the AUC-ROC performances of all the models are ideal.

Table 2
Comparison of ESRD classification results.

	Recall rate	F1 score	Accuracy
Baseline-CNN	91.18 ± 1.04%	92.51 ± 0.98%	93.90 ± 1.07%
Baseline-LSTM	95.74 ± 1.09%	95.12 ± 0.54%	94.52 ± 0.93%
Baseline-CNN-LSTM	96.30 ± 1.35%	96.08 ± 0.77%	95.88 ± 0.53%
Entropy-MLP	97.53 ± 0.74%	97.82 ± 0.43%	98.12 ± 0.67%
Avg Improvement	**3.12pp**	**3.25pp**	**3.35pp**

the baseline utilizes deep learning models, we employ MLP as the classifier for the entropy-based method, which can also offer advantages in terms of computational efficiency and model simplification.

The experimental results for ESRD are shown in Fig. 13 and Table 2, revealing improvements in both model performance and efficiency with the entropy-based method. In the task of epileptic seizure recognition, compared to the baselines, the entropy-based MLP method shows an average improvement of 3.12pp in recall, 3.25pp in F1 score, and 3.35pp in accuracy. The ROC-AUC curve also shows a sufficient model performance of the entropy-based model. Furthermore, the model's parameter count is significantly reduced, with an average reduction of 3.05 times. This improvement shows the effectiveness and efficiency of the entropy-based method.

The effectiveness of the entropy-based method can be explained from several perspectives. Firstly, entropy can measure signal uncertainty and complexity, which is particularly crucial in the analysis of EEG and ECG signals. By combining and integrating series entropy features, we can capture the nonlinear characteristics and underlying dynamic variations from different aspects, which are vital for recognizing atypical pattern changes such as epileptic seizures. Secondly, feature selection by MI and PCC can effectively remove redundant features while retaining the most informative ones. This reduces model complexity, prevents overfitting, and enhances model generalization.

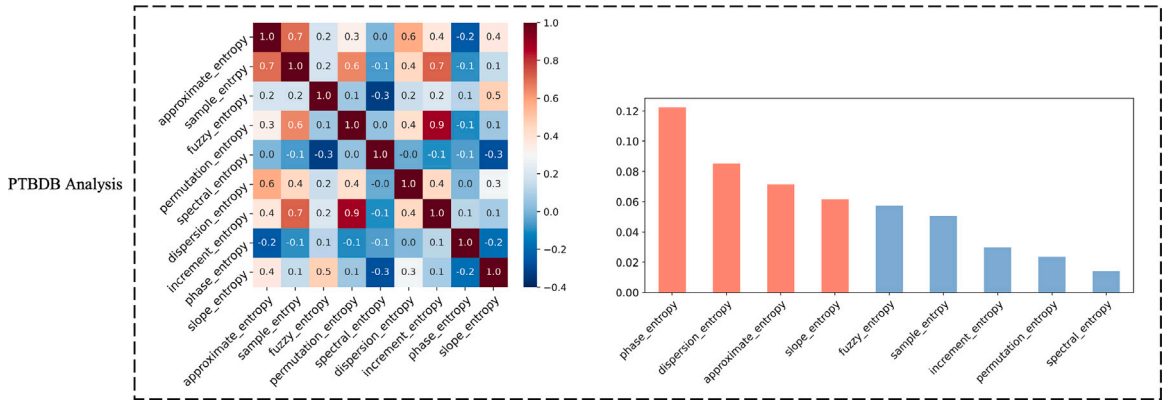


Fig. 14. The feature selection result of PTBDB, including Pearson relationship matrices (the left) and mutual information (the right). In the right figure, the selected features are in red and the unselected features are in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The efficiency improvement of the entropy-based method stems from the fewer model parameters compared to the baselines, reducing computational complexity. This not only lowers the risk of overfitting but also makes the model more suitable for real-time analysis and large-scale data processing. Additionally, parameter reduction implies that the model is more efficient during both training and inference stages, which is particularly important for medical applications requiring rapid responses.

Table 3
Comparison of PTBDB classification results.

	Recall rate	F1 score	Accuracy
Baseline-CNN	92.16 ± 1.83%	92.92 ± 1.22%	93.70 ± 0.88%
Baseline-LSTM	93.11 ± 2.02%	93.61 ± 1.27%	94.14 ± 0.60%
Baseline-CNN-LSTM	94.07 ± 1.57%	94.32 ± 1.28%	94.59 ± 1.58%
Entropy-MLP	95.02 ± 0.76%	95.21 ± 0.64%	95.42 ± 0.99%
Avg Improvement	**1.91pp**	**1.59pp**	**1.28pp**

5.3. PTBDB

PTBDB is a non-state-transition dataset. We aim to distinguish the ordinary participants and the participants with any heart disease. Following the configurations of the experiments on the ESRD dataset, we also employ CNN-based, LSTM-based, and CNN-LSTM-based models as baselines [22,48,49]. The preprocessing steps are the same as the ESRD experiments. The output of the models is the participants with any heart disease (True or False).

The feature selection result is shown in Fig. 14. Based on the MI, we first select the top four features, including phase entropy, dispersion entropy, approximate entropy, and slope entropy. Then we conduct a secondary selection of the selected features based on the PCC, and all the selected features exhibit a relatively low Pearson correlation coefficient (lower than 0.8). Therefore, we select the above entropy features as the entropy-based features. Similar to the experiment configuration with the ESRD dataset, we employ MLP as the classifier for the entropy-based method.

The experimental results for PTBDB are shown in Fig. 15 Table 3, showing improvements in both model performance and efficiency with the entropy-based method. The entropy-based method achieves similar model performance to the SOTA, showing an average improvement of 1.91pp in recall rate, 1.59pp in F1 score, and 1.28pp in accuracy. The ROC-AUC curves also demonstrate that we achieve similar model performance to the SOTA methods. For the model's parameter count, compared to the baselines, the entropy-based model can achieve an average reduction of 3.14 times.

The entropy-based approach has shown consistent effectiveness across both state-transition and non-state-transition datasets, marked by both robust model performance and operational efficiency. This success is largely due to the utilization of various entropy features for feature extraction and the implementation of strategic feature selection processes. Incorporating multiple entropy features facilitates a comprehensive analysis of time-series data by harnessing a broader spectrum of information, which traditional methods might overlook. Simultaneously, the process of feature selection meticulously eliminates features that exhibit high redundancy or collinearity, thereby streamlining the dataset and improving the efficiency and accuracy of the model. The strategic choice and use of entropy features significantly enhance the models' efficiency, allowing for the achievement of desired performance levels without resorting to overly complex models. This aspect is particularly beneficial in healthcare settings, where quick and accurate feedback is paramount.

6. Discussion and future work

To evaluate our method, we conduct experiments on three different datasets. The results show that, compared to the baseline, our method can improve performance on various evaluation metrics and different types of datasets. This can help to develop more robust decision-support tools for applications that use neural time-series data or applications in other fields that use similar data. The first dataset is the daily activity data of PLWD, which is state-transition data collected by PIR sensors. Information theory-based methods can effectively identify unusual patterns in this type of data. The second and third datasets are the seizure dataset and the electrocardiography dataset, which are non-state-transition data. Information theory-based methods also accurately identify patients with epilepsy and heart disease, achieving an effective diagnosis. We attribute the improvements mainly to:

• *Effective high-level entropy features.* Entropy features consider critical statistical information, aiding in the extraction of advanced features from raw data that traditional neural networks struggle to automatically extract. We design, apply, and select different entropy features for various time-series data. For state-transition data, besides using Shannon entropy, we also preprocess the original data into a first-order Markov chain and apply the entropy rate, entropy production, and von Neumann entropy. For non-state-transition data, we employ different variants of entropy, such as increment entropy and slope entropy.

• *Multi-dimensional Entropy Feature Analysis.* Our methodology integrates a variety of entropy features, each designed to uncover the inherent complexity of the data from distinct viewpoints [51,52]. By adopting this multi-dimensional strategy, we achieve a deeper and more nuanced comprehension of the dynamics and structures present in neural time-series data. Whereas conventional analysis techniques might concentrate on a singular data characteristic or statistic, thus risking the omission of other crucial data attributes, our approach to

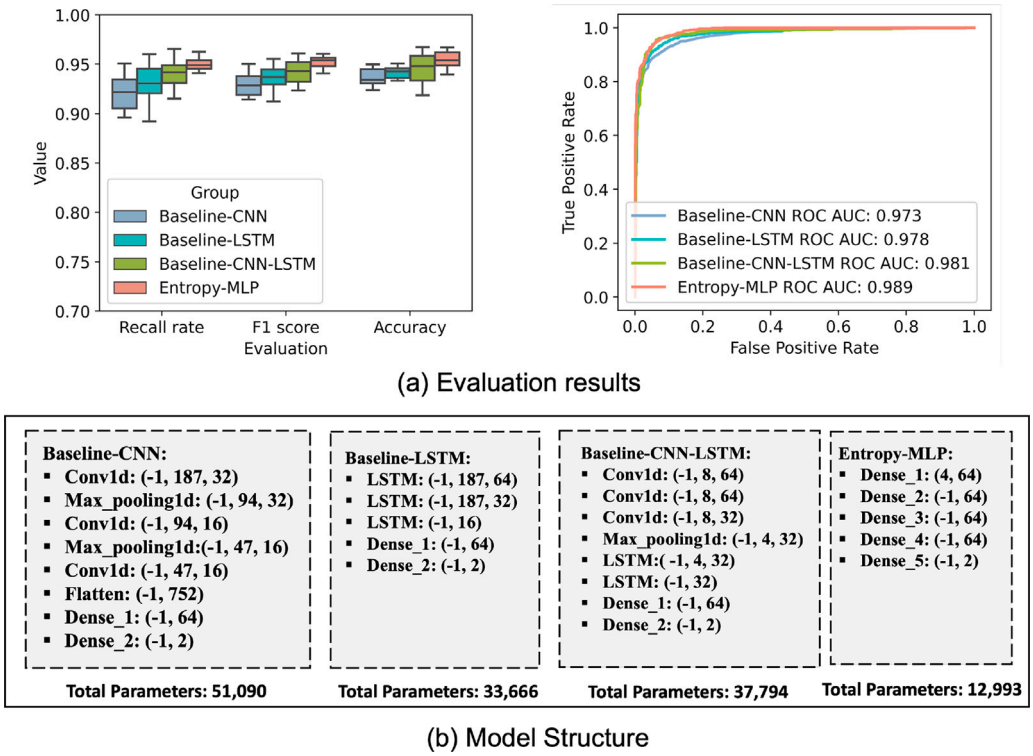


Fig. 15. The results of PTBDB database. (a) is the evaluation results. (b) is the comparison of the model structure between the baseline and entropy models. As we utilize pre-processed data with less data noise, the AUC-ROC performances of all the models are ideal.

multi-dimensional entropy feature analysis ensures a broad capture of the data's facets, providing a more detailed and precise representation

• *Efficient feature selection.* To mitigate the potential decline in model performance caused by redundant and highly collinear features, we implement a feature selection process grounded in information theory. This process employs both mutual information and the Pearson correlation coefficient matrix to discern and select the optimal k entropy features for our modeling efforts.

• *Model simplicity enhanced by efficient entropy features.* Traditional deep learning models often necessitate extensive computational resources and time to analyze complex time-series data effectively. Yet, by leveraging entropy features, our approach enables the attainment of favorable outcomes using simpler model architectures, like LR and MLP. This advancement not only boosts the efficiency of the model but also aids in minimizing model complexity and enhancing interpretability.

However, there are also some limitations in the proposed solutions. We plan to investigate and address the limitations of the solution in future work. The key limitations are listed below.

• *Dynamic Feature Selection.* We introduce a feature selection methodology centered around Mutual Information (MI) and Pearson Correlation Coefficient (PCC), designed for static and comprehensive medical time-series data. This selection approach is initially formulated in the Cloud and subsequently implemented locally. Nonetheless, the dynamic nature and frequent incompleteness of medical data pose significant challenges in real-world settings [53–55]. Liu et al. [56] developed a technique for dynamic mutual information feature selection that caters to evolving datasets over time, proving adept at real-time adjustments and responsiveness to data fluctuations. Similarly, Luo et al. [57] have concentrated on feature selection for datasets with incomplete information, addressing the critical issue of handling missing or partial data. In our subsequent efforts, we aim to incorporate these dynamic and incomplete data-focused techniques to broaden and refine our method's effectiveness across diverse medical data contexts. This enhancement will significantly extend the applicability and robustness of our approach in various medical scenarios.

• *Model Interpretation.* While incorporating diverse entropy features has improved both the performance and stability of our models, it has also somewhat complicated the interpretation of these models. In our forthcoming research endeavors, we intend to investigate and devise more streamlined approaches and tools specifically aimed at facilitating the interpretation and comprehension of complex entropy features, as well as elucidating their impact on model performance.

•*Integration of Multimodal Datasets.* We also recognize the complexity and diversity of neural time series data and plan to explore how to effectively incorporate data from different sources and modalities into our analysis framework. This will involve developing data fusion techniques and algorithms to enhance the model's generalizability and accuracy.

• *Expansion of Application Scope.* The proposed method has the potential to be extended to other types of time series data. In future work, we plan to explore how this methodology can be applied to fields such as finance, meteorology, and energy, aiming to unlock new insights and enhance analytical precision in these areas.

• *Personalization and Customization.* We will continue to explore how to customize and optimize our method to meet the needs of specific applications and groups. This may include developing personalized models that take into account individual differences and specifications.

7. Conclusions

We propose a highly general pipeline that uses information theory and entropy to extract high-level features and analyze pattern recognition from inherent low-level neurological time-series data, which can reduce privacy risks, improve model performance and enhance

efficiency. This pipeline's effectiveness and scalability have been validated through pattern analysis in datasets concerning dementia care, epilepsy, and myocardial infarction. For instances involving stochastic state transitions, our approach utilizes features derived from Shannon's entropy, entropy rates, entropy production rate, and von Neumann entropy of Markov chains. In situations where state transition modeling does not apply, we employ approximate entropy, increment entropy, dispersion entropy, phase entropy, and slope entropy. Furthermore, we introduce an entropy feature selection method based on mutual information and the Pearson correlation matrix. The results show that, compared with the baseline, the entropy-based method improves the recall rate, F1 score, and accuracy on average by up to 13.08pp. We also compared the pipeline with deep learning models on ESRD and PTBDB. The results show that the pipeline can also enhance efficiency with an average reduction of 3.10 times in the number of model parameters. The proposed pipeline offers a versatile, high-precision, privacy-preserved, and efficient solution for analyzing time-series data for pattern recognition in neurological conditions, which can be applied to various scenarios in healthcare.

CRediT authorship contribution statement

Yushan Huang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yuchen Zhao:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Alexander Capstick:** Writing – review & editing, Validation, Methodology, Data curation. **Francesca Palermo:** Writing – review & editing, Validation, Methodology, Data curation. **Hamed Haddadi:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Payam Barnaghi:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Payam Barnaghi reports financial support was provided by EPSRC PROTECT Project. Hamed Haddadi reports financial support was provided by EPSRC OpenPlus Fellowship. Yushan Huang reports financial support was provided by China Scholarship Council. Payam Barnaghi reports financial support was provided by MRC and Alzheimer's Society. Payam Barnaghi reports was provided by the Great Ormond Street Hospital Children's Charity Award.

Acknowledgments

This project is supported by the EPSRC PROTECT Project, United Kingdom (grant number: EP/W031892/1), EPSRC OpenPlus Fellowship (EP/W005271/1), and the UK DRI Care Research and Technology Centre funded by MRC and Alzheimer's Society (grant number: UKDRI-7002). The raw data from the Minder dataset was accessed using the DCARTE library developed by Dr. Eyal Soreq at the UK Dementia Research Institute's Care Research and Technology Centre. Yushan Huang is funded by the China Scholarship Council. Payam Barnaghi's research is also partially supported by an award from the Great Ormond Street Hospital (GOSH), United Kingdom (grant number 21PP30).

Appendix

A.1. Entropy and entropy variants

A.1.1. Approximate entropy

For non-Markovian chain systems, Approximate Entropy (ApEn) can be used to quantify the complexity of the system. Given a time series dataset $\{u(i) : 1 \le i \le N\}$ with N samples, form the sequence in order to generate an m-dimension vector:

$$u'(i) = [u(i), u(i+1), \dots, u(i+m-1)] \quad i = 1, N-m+1 \tag{A.1}$$

Define the distance between the vectors $u'(i)$ and $u'(j)$ to be the maximum of the differences between the corresponding elements of the two vectors:

$$d[u'(i), u'(j)] = \max_{k=0,m-1}[|u'(i+k) - u'(j+k)|] \tag{A.2}$$

Given a threshold p, count the number of $d[u'(i), u'(j)] <= p$, denoted as $A_N^m(p)$, and calculate the ratio of $A_N^m(p)$ to $N-m+1$, denoted as $B_N^m(p)$:

$$B_l^m(p) = \frac{A_l^m(p)}{N-m+1} \tag{A.3}$$

Calculate the average value of $B_N^m(p)$:

$$B^m(p) = \frac{1}{N-m+1}\sum_{N=1}^{N-m+1} B_N^m(p) \tag{A.4}$$

Increase the dimension from m to $m+1$, and repeat the above steps. For sequences of finite length, an estimate of the sample entropy can be obtained as [58]:

$$\mathrm{ApEn}(m, r, N) = B^m(p) - B^{m+1}(p) \tag{A.5}$$

A.1.2. Increment entropy

The Incremental Entropy (IncrEn) algorithm is a method for calculating the entropy of a sequence of data points incrementally rather than computing the entropy of the entire sequence all at once. Given a time series dataset $\{u(i) : 1 \le i \le N\}$ with N samples. Construct an increment time series $\{v(i), 1 \le i \le N-1\}$ by $v(i) = x(i+1) - x(i)$ from $u(i)$. Hence, for a positive integer m, $N-m$ vectors of dimension m are derived from an incremental time series. These vectors, denoted as $V(k) = [v(k), v(k+1), \dots, v(k+m-1)], 1 \le k \le N-m$, represent contiguous segments of the time series. Each element in a vector $V(k)$ is mapped onto a word of two letters. The sign of each component is represented by $v'_{k+j} = \mathrm{sgn}(v(k+j)), j = 1 \cdots, m-1$, and the magnitude of each component in relation to the other components within the vector is represented by $q_{k+j}, j = 1, \dots, m-1$ for a quantifying resolution r. As a result, $N-m$ words, $w_k, 1 \le k \le N-m$, are generated. Each word, consisting of $2 \times m$ letters, can have $(2r+1)^m$ variations, depending on the values of m and r. The frequency of occurrence of each unique word w_n is defined as:

$$p(w_n) = \frac{Q(w_n)}{N-m} \tag{A.6}$$

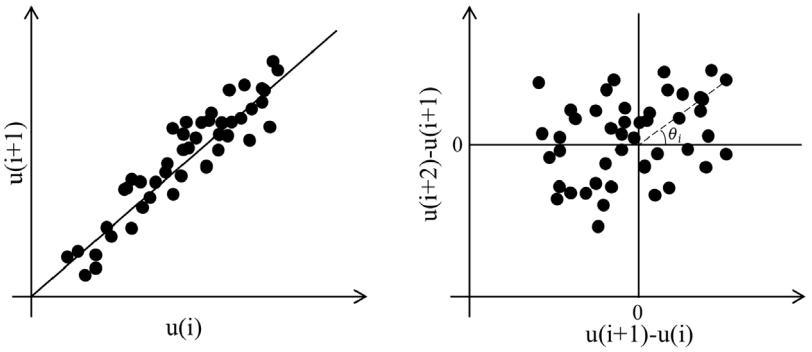
where $Q(w_n)$ signifies the count of the unique word w_n within the $\{w_k\}$. The Increment Entropy (IncrEn) of order m (where m is equal to or greater than 2) and resolution R is defined as:

$$IncrEn(m) = -\sum_{n=1}^{(2R+1)^m} p(w_n)\log p(w_n) \tag{A.7}$$

A.1.3. Dispersion entropy

Dispersion entropy (DE) can be used to describe the complexity of time series data. For time series with low regularity, DE can reflect the degree of disorder of the series [59]. Given a time series dataset $\{u(i) : 1 \le i \le N\}$ with N samples. Map $u(i)$ to $y(i)$ between 0 and 1 by normal cumulative distribution function (NCDF):

$$y_j = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{u_j} e^{-((t-\mu)^2/2\sigma^2)dt} \tag{A.8}$$



(a) Lorenz plot (b) SODP

Fig. A.16. The phase space representation of an HRV signal.

In which the parameter μ is the expectation of $u(i)$, while the parameter σ is its standard deviation. Map y to the range of $[1, 2, \ldots, c]$, and obtain a new sequence $z_j^{(c)}$:

$$z_j^{(c)} = \text{int}\left(cy_j + 0.5\right) \tag{A.9}$$

In which, c is the number of categories, and int is the rounding function. Then construct the embedding vector $z_i^{(m,c)}$ by:

$$z_i^{(m,c)} = \left(z_i^{(c)}, z_{i+d}^{(c)}, \ldots, z_{i+(m-1)d}^{(c)}\right), \quad i = 1, 2, \ldots, N-(m-1)d \tag{A.10}$$

In which, m is the embedding dimension, c is the number of classes, and d is the time delay. Then each $z_j^{(m,c)}$ is mapped to dispersion pattern $\pi_{v_0 v_1 \cdots v_{m-1}}(v = 1, 2, \ldots, c)$, in which $z_i^{(c)} = v_0$, $z_{i+d}^{(c)} = v_1$, ..., and $z_{i+(m-1)d}^{(c)} = v_{m-1}$. The number of possible dispersion of each $z_j^{(m,c)}$ is c^m.

Calculate the relative frequency for each potential dispersion pattern:

$$P\left(\pi_{v_0 v_1, \ldots, v_{m-1}}\right) = \frac{\text{num}\left(\pi_{v_0 v_1, \ldots, v_{m-1}}\right)}{N-(m-1)d} \tag{A.11}$$

Finally, based on Shannon's entropy, DE can be obtained by [60]:

$$DE(u, m, c, d) = -\sum_{\pi=1}^{c^m} p\left(\pi_{v_0, \ldots, v_{m-1}}\right) \ln\left(p\left(\pi_{v_0, \ldots, v_{m-1}}\right)\right) \tag{A.12}$$

A.1.4. Phase entropy

Phase entropy (PhEn) is developed to detect the complexity of physiological signals. For example, given a time series dataset $\{u(i) : 1 \le i \le N\}$ with N samples, we can represent the data by the Lorenz plot, as Fig. A.16(a) shows. In the Poincaré plot, if we replace the sequence u_i by $u_{i+1} - u_i$, then we can get SODP plot, as Fig. A.16(b) shows. Specifically, from a given time series u_i, we can obtain Y_i and X_i by [61]:

$$\begin{aligned} Y_i &= u_{i+2} - u_{i+1} \\ X_i &= u_{i+1} - u_i \end{aligned} \tag{A.13}$$

Then compute the slope angle of each scatter point as shown in Fig. A.16(b).

$$\theta_i = \tan^{-1}\frac{Y_i}{X_i} \tag{A.14}$$

Then the probability distribution p_i can be calculated by:

$$p_i = \frac{S_{\theta_i}}{\sum_{i=1}^{k} S_{\theta_i}} \tag{A.15}$$

Finally, based on Shannon's entropy, the PhEn can be calculated as [61]:

$$\text{PhEn} = \frac{-1}{\log N}\sum_{i=1}^{k} p(i)\log p(i) \tag{A.16}$$

A.1.5. Slope entropy

Slope entropy (SlopEn) is a method for measuring the complexity of time series data, which is primarily based on transferring the original time series data to a series of single-threshold and symbolic patterns [62,63]. SlopEn is initially applied in the fields of medicine and biological signal processing, particularly in the analysis of electrocardiograms (ECG) and electroencephalograms (EEG). Its core concept is based on estimating the uncertainty or complexity of data by examining changes in the slopes of time-series data. This means that slope entropy not only considers the absolute values of data points but also delves into the relative changes between data points.

Compared to other traditional linear metrics, SlopEn demonstrates significant advantages in several aspects. First, it provides a tool to capture the inherent dynamics and nonlinear characteristics of time series, making it an effective measure of complexity. Second, it focuses on changes in slopes, enabling an intuitive understanding of the dynamics of data through it, which is crucial for interpreting and comprehending patterns and trends in time series. Additionally, SlopeEn exhibits good robustness, maintaining the stability of its measurement even in the presence of noise or interference from external factors. When considering time-series signals and their nonlinear and complex nature, these signals often contain rich information and intrinsic dynamics that can be captured through changes in slopes. Slope entropy allows for a deeper understanding of this information, offering a comprehensive and in-depth insight into the complexity and intrinsic dynamics of the data.

Given a time series dataset $\{u(i) : 1 \le i \le N\}$ with N samples. Decompose u into j subsequences according to the embedded dimension m:

$$u_i^m = \left\{u_i, u_{i+1}, \ldots, u_{i+m-1}\right\} \tag{A.17}$$

In which, $i = \{1, 2, \ldots, j\}$, $j = N - m + 1$. Define two soft threshold parameters δ and γ to calculate the symbolic patterns of u_i^m, where $0 < \delta < \gamma$.

Define $d = u_{i+1} - u_i$, and compare d with the two soft threshold parameters δ and γ, then five patterns can be obtained:

$$\begin{cases} \text{pattern} = 2, & \gamma < d, \\ \text{pattern} = 1, & \delta < d \le \gamma, \\ \text{pattern} = 0, & |d| \le \delta, \\ \text{pattern} = -1, & -\gamma \le d < -\delta, \\ \text{pattern} = -2, & d < -\gamma. \end{cases} \tag{A.18}$$

Based on the five patterns, we can get 5^{m-1} sequence combinations. The relative frequency p_n of the combination can be calculated by the number of occurrences f_n of each combination:

$$p_n = \frac{f_n}{j}, n = 1, 2, \ldots, 5^{m-1} \tag{A.19}$$

Finally, SlopEn can be calculated based on the Shannon's entropy:

$$\mathrm{SE}(m,\gamma,\delta) = -\sum_{n=1}^{5^{m-1}} p_n \ln p_n \tag{A.20}$$

References

[1] Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang G-Z. Big data for health. IEEE J Biomed Health Inf 2015;19(4):1193–208.

[2] Motwani A, Shukla PK, Pawar M. Ubiquitous and smart healthcare monitoring frameworks based on machine learning: A comprehensive review. Artif Intell Med 2022;102431.

[3] Zhang C, Song D, Chen Y, Feng X, Lumezanu C, Cheng W, et al. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In: Proceedings of the AAAI conference on artificial intelligence. 2019, p. 1409–16.

[4] Piccialli F, Giampaolo F, Prezioso E, Camacho D, Acampora G. Artificial intelligence and healthcare: Forecasting of medical bookings through multi-source time-series fusion. Inf Fusion 2021;74:1–16.

[5] Rafiei MH, Gauthier LV, Adeli H, Takabi D. Self-supervised learning for electroencephalography. IEEE Trans Neural Netw Learn Syst 2022.

[6] Shah HA, Saeed F, Yun S, Park J-H, Paul A, Kang J-M. A robust approach for brain tumor detection in magnetic resonance images using finetuned efficientnet. IEEE Access 2022;10:65426–38.

[7] Huang Y, Zhao Y, Haddadi H, Barnaghi P. Using entropy measures for monitoring the evolution of activity patterns. 2022, arXiv preprint arXiv:2210.01736.

[8] Palermo F, Chen Y, Capstick A, Fletcher-Loyd N, Walsh C, Kouchaki S, et al. Tihm: An open dataset for remote healthcare monitoring in dementia. Sci Data 2023;10(1):606.

[9] Parkinson ME, Doherty R, Curtis F, Soreq E, Lai HH, Serban A-I, et al. Using home monitoring technology to study the effects of traumatic brain injury in older multimorbid adults. Ann Clinical Transl Neurol 2023;10(9):1688–94.

[10] Murugappan M, Murugappan S. Human emotion recognition through short time electroencephalogram (EEG) signals using fast Fourier transform (FFT). In: 2013 IEEE 9th international colloquium on signal processing and its applications. IEEE; 2013, p. 289–94.

[11] Lawhern V, Hairston WD, McDowell K, Westerfield M, Robbins K. Detection and classification of subject-generated artifacts in eeg signals using autoregressive models. J Neurosci Methods 2012;208(2):181–9.

[12] Xu S, Wang Z, Sun J, Zhang Z, Wu Z, Yang T, et al. Using a deep recurrent neural network with EEG signal to detect parkinson's disease. Ann Transl Med 2020;8(14).

[13] Michielli N, Acharya UR, Molinari F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. Comput Biol Med 2019;106:71–81.

[14] Ghosh SM, Bandyopadhyay S, Mitra D. Nonlinear classification of emotion from EEG signal based on maximized mutual information. Expert Syst Appl 2021;185:115605.

[15] Krishnan PT, Raj ANJ, Balasubramanian P, Chen Y. Schizophrenia detection using MultivariateEmpirical mode decomposition and entropy measures from multichannel EEG signal. Biocybern Biomed Eng 2020;40(3):1124–39.

[16] McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. PMLR; 2017, p. 1273–82.

[17] Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Phys Rev E 2001.

[18] Bousseljot R-D. Nutzung der EKG-signaldatenbank CARDIODAT der PTB über das internet. Biomed Tech/Biomed Eng (Biomed Eng) 1995;40:317–8.

[19] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–20.

[20] Huang Y. EntropyPipeline. 2023, https://github.com/yushan-huang/EntropyPipeline.

[21] Niu M, Zhao Y, Haddadi H. Effective abnormal activity detection on multivariate time series healthcare data. In: Proceedings of the 29th annual international conference on mobile computing and networking. 2023, p. 1–3.

[22] Hussain W, Sadiq MT, Siuly S, Rehman AU. Epileptic seizure detection using 1 D-convolutional long short-term memory neural networks. Appl Acoust 2021;177:107941.

[23] Park J, Lee K, Park N, You SC, Ko J. Self-attention LSTM-FCN model for arrhythmia classification and uncertainty assessment. Artif Intell Med 2023;102570.

[24] Shankar V, Yousefi E, Manashty A, Blair D, Teegapuram D. Clinical-gan: Trajectory forecasting of clinical events using transformer and generative adversarial networks. Artif Intell Med 2023;138:102507.

[25] Ding S, Zhu H, Jia W, Su C. A survey on feature extraction for pattern recognition. Artif Intell Rev 2012;37(3):169–80.

[26] Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27(3):379–423.

[27] Powell G, Percival I. A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems. J Phys A 1979;12(11).

[28] Richman JS, Lake DE, Moorman JR. Sample entropy. In: Methods in enzymology, vol. 384, Elsevier; 2004, p. 172–84.

[29] Nurwulan NR, Jiang BC. Multiscale entropy for physical activity recognition. In: Proceedings of the 2020 2nd Asia Pacific information technology conference. 2020, p. 73–7.

[30] Bao L, Intille SS. Activity recognition from user-annotated acceleration data. In: International conference on pervasive comp.. Springer; 2004, p. 1–17.

[31] Howedi A, Lotfi A, Pourabdollah A. Exploring entropy measurements to identify multi-occupancy in activities of daily living. Entropy 2019;21(4):416.

[32] Lampard D. A stochastic process whose successive intervals between events form a first order Markov chain—I. J Appl Probab 1968;5(3):648–68.

[33] Enshaeifar S, Zoha A, Markides A, Skillman S, Acton ST, Elsaleh T, et al. Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques. PLoS One 2018;13(5).

[34] Dewar R. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. J Phys A 2003;36(3):631.

[35] Butner J, Amazeen PG, Mulvey GM. Multilevel modeling of two cyclical processes: extending differential structural equation modeling to nonlinear coupled systems.. Psychol Methods 2005;10(2):159.

[36] Vallacher RR, Nowak A. The emergence of dynamical social psychology. Psychol Inq 1997;8(2):73–99.

[37] Kim D-K, Bae Y, Lee S, Jeong H. Learning entropy production via neural networks. Phys Rev Lett 2020;125(14).

[38] Bengtsson I, Życzkowski K. Geometry of quantum states: an introduction to quantum entanglement. Cambridge University Press; 2017.

[39] Felippe H, Viol A, de Araujo D, da Luz M, Palhano-Fontes F, Onias H, et al. The von Neumann entropy for the pearson correlation matrix: A test of the entropic brain hypothesis. 2021, arXiv preprint arXiv:2106.05379.

[40] MacDuffee CC. The theory of matrices, vol. 5, Springer Science & Business Media; 2012.

[41] Mei K, Tan M, Yang Z, Shi S. Modeling of feature selection based on random forest algorithm and pearson correlation coefficient. J Phys Conf Series 2022;2219(1):012046.

[42] Liu Y, Mu Y, Chen K, Li Y, Guo J. Daily activity feature selection in smart homes based on pearson correlation coefficient. Neural Process Lett 2020;51:1771–87.

[43] Kachuee M, Fazeli S, Sarrafzadeh M. Ecg heartbeat classification: A deep transferable representation. In: IEEE international conference on healthcare informatics. IEEE; 2018, p. 443–4.

[44] Sun S, Chen H, Luo G, Yan C, Dong Q, Shao X, et al. Clustering-fusion feature selection method in identifying major depressive disorder based on resting state EEG signals. IEEE J Biomed Health Inf 2023.

[45] Abdellatef E, Emara HM, Shoaib MR, Ibrahim FE, Elwekeil M, El-Shafai W, et al. Automated diagnosis of EEG abnormalities with different classification techniques. Med Biol Eng Comput 2023;1–23.

[46] Chawla P, Rana SB, Kaur H, Singh K, Yuvaraj R, Murugappan M. A decision support system for automated diagnosis of Parkinson's disease from EEG using FAWT and entropy features. Biomed Signal Process Control 2023;79:104116.

[47] Chimamiwa G, Alirezaie M, Banaee H, Köckemann U, Loutfi A. Towards habit recognition in smart homes for people with dementia. In: Ambient intelligence: 15th European conference, amI 2019, rome, Italy, November 13–15, 2019, proceedings 15. Springer; 2019, p. 363–9.

[48] Khalilpour S, Ranjbar A, Menhaj MB, Sandooghdar A. Application of 1-D CNN to predict epileptic seizures using eeg records. In: 2020 6th international conference on web research. IEEE; 2020, p. 314–8.

[49] Farsi L, Siuly S, Kabir E, Wang H. Classification of alcoholic EEG signals using a deep learning method. IEEE Sens J 2020;21(3):3552–60.

[50] Volicer L, Harper DG, Manning BC, Goldstein R, Satlin A. Sundowning and circadian rhythms in alzheimer's disease. Am J Psychiatry 2001;158(5):704–11.

[51] Jui SJJ, Deo RC, Barua PD, Devi A, Soar J, Acharya UR. Application of entropy for automated detection of neurological disorders with electroencephalogram signals: A review of the last decade (2012–2022). IEEE Access 2023.

[52] Li W, Zhao Y, Wang Q, Zhou J. Twenty years of entropy research: A bibliometric overview. Entropy 2019;21(7):694.

[53] Ke C, Jin Y, Evans H, Lober B, Qian X, Liu J, et al. Prognostics of surgical site infections using dynamic health data. J Biomed Informat 2017;65:22–33.

[54] Tan Q, Ye M, Ma AJ, Yip TC-F, Wong GL-H, Yuen PC. Importance-aware personalized learning for early risk prediction using static and dynamic health data. J Am Med Inf Assoc 2021;28(4):713–26.

[55] Ranjbari S, Arslanturk S. Integration of incomplete multi-omics data using knowledge distillation and supervised variational autoencoders for disease progression prediction. J Biomed Inform 2023;147:104512.

[56] Liu H, Sun J, Liu L, Zhang H. Feature selection with dynamic mutual information. Pattern Recognit 2009;42(7):1330–9.

[57] Luo C, Li T, Chen H, Lv J, Yi Z. Fusing entropy measures for dynamic feature selection in incomplete approximation spaces. Knowl-Based Syst 2022;252:109329.
[58] Chen X, Solomon IC, Chon KH. Comparison of the use of approximate entropy and sample entropy: applications to neural respiratory signal. In: 2005 IEEE engineering in medicine and biology 27th annual conference. IEEE; 2006, p. 4212–5.
[59] Rostaghi M, Azami H. Dispersion entropy: A measure for time-series analysis. IEEE Signal Process Lett 2016;23(5):610–4.
[60] Chakraborty M, Mitra D, et al. Automated detection of epileptic seizures using multiscale and refined composite multiscale dispersion entropy. Chaos Solut Fractals 2021;146.
[61] Rohila A, Sharma A. Phase entropy: A new complexity measure for heart rate variability. Physiol Measur 2019;40(10).
[62] Li Y, Mu L, Gao P. Particle swarm optimization fractional slope entropy: a new time series complexity indicator for bearing fault diagnosis. Fractal Fractional 2022;6(7).
[63] Li Y, Tang B, Yi Y. A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy. Appl Acoust 2022;196:108899.