Review

Applications of Entropy in Data Analysis and Machine Learning: A Review

Salomé A. Sepúlveda-Fontaine and José M. Amigó [*]

Review

Applications of Entropy in Data Analysis and Machine Learning: A Review

Salomé A. Sepúlveda-Fontaine [1] **and José M. Amigó** [2,*]

[1]   Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, 03202 Elche (Spain)

[2]   Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, 03202 Elche (Spain)

*   Correspondence: jm.amigo@umh.es

Abstract: Since its origin in the thermodynamics of the 19th century, the concept of entropy has also permeated other fields of physics and mathematics, such as Classical and Quantum Statistical Mechanics, Information Theory, Probability Theory, Ergodic Theory and the Theory of Dynamical Systems. Specifically, we are referring to the classical entropies: the Boltzmann-Gibbs, von Neumann, Shannon, Kolmogorov-Sinai and topological entropies. In addition to their common name, which is historically justified (as we briefly describe in this review), other commonality of the classical entropies is the important role that they have played and are still playing in the theory and applications of their respective fields and beyond. Therefore, it is not surprising that, in the course of time, many other instances of the overarching concept of entropy have been proposed, most of them tailored to specific purposes. Following the current usage, we will refer to all of them, whether classical or new, simply as entropies. Precisely, the subject of this review is their applications in data analysis and machine learning. The reason for these particular applications is that entropies are very well suited to characterize probability mass distributions, typically generated by finite-state processes or symbolized signals. Therefore, we will focus on entropies defined as positive functionals on probability mass distributions and provide an axiomatic characterization that goes back to Shannon and Khinchin. Given the plethora of entropies in the literature, we have selected a representative group, including the classical ones. The applications summarized in this review finely illustrate the power and versatility of entropy in data analysis and machine learning.

Keywords: entropy; entropic measures; data analysis; machine learning; deep learning.

---

1. Introduction

1.1. Aims and Scope

Entropy is a concept that appears in different areas of physics and mathematics with different meanings. Thus, entropy is a measure of: (i) disorder in Statistical Mechanics, (ii) uncertainty in Information and Probability Theories, (iii) (pseudo-)randomness in the Theory of Measure-preserving Dynamical Systems, and (iv) complexity in Topological Dynamics. This versatility explains why entropy has found extensive applications across various scientific disciplines since its inception in the 19th century.

Precisely, this paper aims to provide an up-to-date overview of the applications of entropy in data analysis and machine learning, where entropy stands here not only for the traditional instances but also for more recent proposals inspired by them. In data analysis, entropy is a powerful tool for detection of dynamical changes, segmentation, clustering, discrimination, etc. In machine learning, it is used for classification, feature extraction, optimization of algorithms, anomaly detection, and more. The ability of entropy to provide insights into data structure and algorithm performance has led to a widespread search for further applications and new proposals tailored to specific needs, both in data analysis and machine learning.

This being the case, the present review will be useful for researchers in the above two fields, interested in the theoretical basics and/or the current applications of entropy. Along with established applications, the authors also have taken into account innovative proposals to reflect the intense research activity on entropy that is currently underway.

At this point, the reader may be wondering what entropy is. A search for the word "entropy" on the Internet returns a large number of results, some of them also called entropy metrics, entropy-like measures or entropy-based indices in the literature. So, what is entropy actually?

1.2. Classical Entropies and Generalized Entropies

Historically, the word "entropy" was introduced by the German physicist Clausius in Thermo-dynamics in 1865 to designate the amount of internal energy in a system that cannot be transformed into work. In particular, entropy determines the equilibrium of a thermodynamical system, namely, the state of maximum entropy consistent with the macroscopic constraints. In the second half of the 19th century, entropy was given a microscopic interpretation in the foundational works of Boltzmann and Gibbs on Statistical Mechanics. In 1927, von Neumann generalized the Boltzmann-Gibbs entropy to the then-emerging theory of Quantum Mechanics [1]. In 1948, the word entropy appeared in a completely different context: Information Theory. If entropy is a measure of disorder in Statistical Mechanics; in the seminal paper of Shannon [2], the creator of Information Theory, entropy stands for the average uncertainty about the outcome of a random variable (or the information conveyed by knowing it). Albeit in different realms, the coincidence in names is explained because Shannon's formula (see equation (1) below) is formally the same as Gibbs' for the entropy of a system in thermal equilibrium with a heat bath at constant temperature [3].

This abridged history of entropy continues with Kolmogorov, who crafted Shannon's entropy into a useful invariant in Ergodic Theory [4], and Sinai, who adapted Kolmogorov's proposal to the theory of measure-preserving dynamical systems [5]. In turn, Adler, Konheim and McAndrew [6] generalized the Kolmogorov-Sinai (KS) entropy from measure-preserving dynamics to topological dynamics under the name of topological entropy. According to the Variational Principle, topological entropy is a tight upper bound of the KS entropies defined over certain probability measures [7].

To get down to the mathematical formulas, let $\mathcal{P}$ be the set of probability mass distributions $\{p_1, ..., p_W\}$ for all $W \geq 2$. Then, the Shannon entropy of the probability distribution $\mathrm{p}_W = \{p_1, ..., p_W\}$ is defined as

$$S(\mathrm{p}_W) = S(p_1, ..., p_W) = -\sum_{i=1}^{W} p_i \log p_i \tag{1}$$

where the choice of the logarithm base fixes the unit of the entropy. The usual choices being 2 (bit), e (nat) or 10 (dit). If $p_i = 0$, then $0 \log 0 := \lim_{x \to 0+} x \log x = 0$. Mathematically, equation (1) is the expected value of the *information function* $I(X) = -\log p(X)$, where X is a random variable with probability distribution p_W. Since entropy is the cornerstone of Information Theory, Shannon also justified definition (1) by proving in his seminal paper [2] that it is unique (except for a positive factor) under a few, general assumptions. In their modern (equivalent) formulation, these assumptions are called the *Shannon-Khinchin axioms* [8], that we state below.

A positive functional H on $\mathcal{P}$, i.e., a map $H : \mathcal{P} \to \mathbf{R}^+$ ($\mathbf{R}^+$ being the non-negative real numbers), is an entropy if it satisfies the following properties:

SK1 *Continuity.* $H(p_1, ..., p_W)$ depends continuously on all variables for each W.
SK2 *Maximality.* For all W,

$$H(p_1, ..., p_W) \leq H(\tfrac{1}{W}, ..., \tfrac{1}{W}).$$

SK3 *Expansibility.* For all W and $1 \leq i \leq W$,

$$H(0, p_1, ..., p_W) = H(p_1, ..., p_i, 0, p_{i+1}, ..., p_W) = H(p_1, ..., p_i, p_{i+1}, ..., p_W).$$

SK4 *Strong additivity* (or *separability*). For all W, U,

$$H(p_{11}, \ldots, p_{1U}, p_{21}, \ldots p_{2U}, \ldots, p_{W1}, \ldots, p_{WU}) \tag{2}$$

$$= H(p_{1*}, p_{2*}, \ldots, p_{W*}) + \sum_{i=1}^{W} p_{i*} H\left(\frac{p_{i1}}{p_{i*}}, \frac{p_{i2}}{p_{i*}}, \ldots, \frac{p_{iU}}{p_{i*}}\right),$$

where $p_{i*} = \sum_{j=1}^{U} p_{ij}$.

Axiom SK4 can be formulated in a more compact way as

$$H(X, Y) = H(X) + H(Y|X),$$

where X and Y are random variables with probability distributions $\{p_{i*} : 1 \leq i \leq W\}$ and $\{p_{*j} = \sum_{i=1}^{W} p_{ij} : 1 \leq j \leq U\}$ respectively, $H(X, Y) = H(p_{11}, \ldots, p_{1U}, \ldots, p_{W1}, \ldots, p_{WU})$ and $H(Y|X)$ is the entropy of Y conditional on X, i.e., the expected value of the conditional distributions $p(y|x)$, averaged over the conditioning variable X [9]. In particular, if X and Y are independent (i.e., $p_{ij} = p_{i*} p_{*j}$), then $H(Y|X) = H(Y)$ and

$$H(X, Y) = H(X) + H(Y). \tag{3}$$

If H satisfies equation (3) for independent random variables X and Y, then it is called *additive*.

It was proved in [2] and [8] that a positive functional H on $\mathcal{P}$ that fulfills Axioms SK1-SK4 is necessarily of the form

$$H(p_1, \ldots, p_W) = -k \sum_{i=1}^{W} p_i \log p_i =: S_{BGS}(p_1, \ldots, p_W) \tag{4}$$

for every $W \geq 2$, where k is a positive constant. For historical reasons, S_{BGS} is usually called the *Boltzmann-Gibbs-Shannon entropy*. In Physics, k is the Boltzmann constant $k_B = 1.3806504(24) \times 10^{-23}$ J/K and log is the natural logarithm. In Information Theory, $k = 1$ and log is the base 2 logarithm when dealing with digital communications. The particular case

$$S_{BGS}(1/W, \ldots, 1/W) = k \log W, \tag{5}$$

obtained for uniform distributions is sometimes referred to as the *Boltzmann entropy*, although the expression (5) is actually due to Planck [10]. According to Axiom SK2, the Boltzmann entropy is the maximum of S_{BGS}.

The same conclusion about the uniqueness of S_{BGS} can be derived using other equivalent properties [11]. Since we are not interested in physical applications here, we set $k = 1$ and generally refer to S_{BGS} as Shannon's entropy.

In 1961 Rényi proposed a generalization of Shannon's entropy by using a different, more general definition of expectation value [12,13]: For any real $\alpha > 0$, $\alpha \neq 1$, *Rényi entropy* R_α is defined as

$$R_\alpha(p_1, \ldots, p_W) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{W} p_i^\alpha\right). \tag{6}$$

So, Rényi entropy is actually a family of entropies; in particular, $R_1 := \lim_{\alpha \to 1} R_\alpha = S_{BGS}$. Other limiting cases are $R_0 := \lim_{\alpha \to 0} R_\alpha = \log W$, called *Hartley* or *max-entropy*, which coincides with the Boltzmann entropy (5) except for the value of the constant k, and $R_\infty := \lim_{\alpha \to \infty} R_\alpha = \min_{1 \leq i \leq W}(-\ln p_i)$, called the *min-entropy*. These names are due to the decreasing monotonicity of Rényi's entropy with respect to the parameter: $R_\alpha > R_\beta$ for $\alpha < \beta$.

It is easy to show that Rényi's entropy satisfies Axioms SK1-SK3 but not SK4. Instead of strong additivity, R_α satisfies *additivity*:

$$R_\alpha(\mathrm{p}_U \times \mathrm{q}_W) = R_\alpha(\mathrm{p}_U) + R_\alpha(\mathrm{q}_W),$$

see equation (3).

A final milestone in this short history of entropy is the introduction of *non-additive entropies* by Havrda and Charvát in Information Theory [14] and Tsallis in Statistical Mechanics [15], which are equivalent and usually called the Tsallis entropy:

$$T_q(p_1, ..., p_W) = \frac{1}{1-q}\left(\sum_{i=1}^{W} p_i^q - 1\right) \tag{7}$$

for any real $q > 0, q \neq 1$. Again, Tsallis entropy is a family of entropies that satisfies Axioms SK1-SK3 but not SK4. Instead, T_q is "q-additive", meaning that

$$T_q(\mathrm{p}_U \times \mathrm{q}_W) = T_q(\mathrm{p}_U) + T_q(\mathrm{q}_W) + (1-q)T_q(\mathrm{p}_U)T_q(\mathrm{q}_W).$$

As Rényi's entropy, the Tsallis entropy is a generalization of Shannon's entropy in the sense that $T_1 := \lim_{q \to 1} T_q = S_{BGS}$. Formally, T_q can be obtained from S_{BGS} by replacing the logarithm in equation (4) by the "q-logarithm" [13].

The appearance of generalizations of the Shannon entropy prompted the weaker concept of *generalized entropy*: a positive functional on probability distributions that satisfies Axioms SK1-SK3. Therefore, the BGS entropy, together with the Rényi and Tsallis entropies are examples of generalized entropies. Shannon's uniqueness theorem can then be rephrased by saying that, the only generalized entropy that is strongly additive is the BGS entropy. Axioms SK1-SK3 are arguably the minimal requirements for a positive functional on probability mass distributions to be called an entropy. Most "entropies" proposed since the formulation of Rényi and Tsallis' entropies are precisely generalized entropies in the axiomatic sense.

To wrap up this short account of the classical and generalized entropies, let us mention that Shannon's, Rényi's and Tsallis' entropies (and other entropies for that matter) have counterparts for continuous-valued random variables and processes (i.e., defined on probability densities). These "differential" versions are formally obtained by replacing probability mass functions by probability densities and summations by integrations in equations (1), (6) and (7), respectively. Although also useful in applications, differential entropies may lack important properties of their discrete counterparts. For example, *differential* (Shannon's) *entropy* lacks positivity [9].

1.3. Methodology and Organization of this Review

As said above, the primary objective of this work is to review the applications of entropy in the fields of data analysis and machine learning. In view of the many versions of entropy currently in use we had to make a selection of them based on their relevance in general, and the interest of their applications, in particular. Apart from the group of classical entropies mentioned in Section 1.2), the remaining entropies selected for this review can be classified into the following three groups.

G1 Entropies based on Shannon's entropy. These are entropies that use Shannon's function (1) on probability distributions obtained from the data by a number of techniques. This group comprises: *Dispersion entropy* and *fluctuation-based dispersion entropy*, *energy entropy* and *empirical mode decomposition energy entropy*, *Fourier entropy* and *fractional Fourier entropy*, *graph entropy*, *permutation entropy*, *spectral entropy*, and *wavelet entropy*.

G2 Entropies based on other information-theoretical concepts, such as the correlation integral, divergences, unconditional or conditional mutual information, etc. This group comprises: *Approximate*

entropy, *cross entropy* and *categorical cross entropy*, *excess entropy*, *kernel entropy*, *relative entropy* or *Kullback-Leibler divergence*, and *transfer entropy*.

G3 Entropies tailored to specific needs or inspired by other entropies. This group comprises: *Bubble entropy*, *entanglement entropy*, *fuzzy entropy*, *intrinsic mode entropy*, *Kaniadakis entropy*, *Rao's quadratic entropy*, *rank-based entropy*, *sample entropy*, and *tone entropy*.

In Section 2, each of the 33 selected entropies is assigned a subsection in alphabetical order. For brevity, mathematical definitions are reminded only when convenient; otherwise, we give a qualitative account and refer the reader to the original publications or standard bibliography for the formulas. The corresponding applications to data analysis and machine learning are explained with brief but sufficient description and provided with specific references. Practical issues such as the choice of parameters or computational implementations are beyond the scope of this review. For a general bibliography on entropy, we refer the reader to the excellent review "The Entropy Universe" by M. Ribeiro et al. [16], as well as the reviews [13,17,18].

2. Applications in Data Analysis and Machine Learning

In this section, the selected entropies are sorted alphabetically. To streamline the exposition, multiscale and weighted versions [19] are included in the same section as the original entropy. The corresponding applications to data analysis and machine learning are tagged with some keywords in alphabetical order so that reverse search (i.e., searching for entropies for a given application) can also be easily performed.

2.1. Approximate Entropy

Approximate entropy was proposed by Pincus in 1991 [20] to analyze medical data. Loosely speaking, approximate entropy is a heuristic implementation of the correlation integral with time series data [18]. The approximate entropy depends on a parameter $r > 0$, sometimes called tolerance, which is a cut-off that defines the concept of proximity between points via the Heaviside step function $\Theta(r - \|x_i - x_j\|)$ (where $\Theta(z) = 1$ if $z \geq 0$, and 0 otherwise). It quantifies the change in the relative frequencies of length-k time-delay vectors with increasing k. A modified version of approximate entropy was proposed in 2000 under the name sample entropy (Section 2.25). See [21] for a tutorial.

Applications

- **Alzheimer's disease.** Approximate entropy has been used in the non-linear analysis of EEGs and MEGs from patients with Alzheimer's disease [22], [23].
- **Anesthetic drug effects.** Another field of applications is the quantification of anesthetic drug effects on the brain activity as measured by EEGs, including comparative testing of different anesthetics [24].
- **Emotion recognition.** Along with other entropies, approximate entropy has been used for EEG-based human emotion recognition [25].
- **Epileptic seizure detection.** Approximate entropy has also been used as biomarker in algorithms for epileptic EEG analysis, in particular, for epileptic seizure detection [26], [27], [28].
- **Physiological time series.** See [29] for an overview of applications of approximate entropy to the analysis of physiological time series.
- **Sleep research.** The applications of approximate entropy include sleep research, in particular, the separation of sleep stages based on EEG data [30].

2.2. Bubble Entropy

Bubble entropy is a metric that evaluates changes in the order of data segments in time series when a new element is added. It was proposed by Manis et al. in 2017 [31] as "an entropy almost free of parameters", inspired by permutation entropy (Section 2.20) and rank-based entropy (Section 2.22). Bubble entropy relies on the Bubble Sort algorithm [31], which compares and swaps adjacent items until ordered.

Applications

- **Biomedical applications.** Due to its minimal dependency on parameters, bubble entropy is particularly useful in biomedical applications (e.g., analysis of heart rate variability) to distinguish healthy from pathological conditions [31,32].
- **Fault bearing detection.** Bubble entropy is used to reinforce the accuracy of fault bearing diagnosis through the Gorilla Troops Optimization (GTO) algorithm for classification [33]. A similar application can be found for the so-called Improved Hierarchical Refined Composite Multiscale Multichannel Bubble Entropy [34].
- **Feature extraction.** Bubble entropy is compared with dispersion entropy (Section 2.6) in the extraction of single and double features in [35].

2.3. Categorical cross entropy

Categorical cross entropy (CCE) is a variant of the cross entropy (Section 2.4) that uses categorical labels instead of real-valued entries. For clarity, we consider CCE in this section and the conventional cross entropy in the next.

Applications

- **Deep Learning.** CCE is used in deep neural networks when dealing with noisy labels. An improved categorical cross entropy (ICCE) is also used in this case [36].
- **Multi-class classification.** CCE is used in the analysis of multi-channel time-series multi-class classification. It is the standard loss function for tasks such as image classification, text classification, and speech recognition [37].
- **Reinforcement learning.** CCE is used as an improvement of value function training (using classification instead of regression) mainly in games [38].
- **Semi-supervised learning.** CCE is used in pseudo-labelling to optimize convolutional neural networks parameters [39].

2.4. Cross Entropy

The cross entropy between the probability distributions $p = (p_1, ..., p_W)$ and $q = (q_1, ..., q_W)$ is defined as

$$C(p, q) = -\sum_{i=1}^{W} p_i \log q_i. \tag{8}$$

It is related to the Shannon entropy $S(p)$ (equation 1) and the Kullback-Leibler divergence $D(p\|q)$ (Section 2.23, equation (15)) through the equation $C(p, q) = S(p) + D(p\|q)$. As $D(p\|q)$, the cross entropy is used to quantify the difference between two probability distributions: the true distribution q (actual labels) and the predicted distribution p (model predictions). A typical example is determining whether an email is spam or not.

Applications

- **Deep learning.** Cross entropy is a standard loss function for training deep neural networks, particularly those involving softmax activation functions. It is very useful for applications such as object detection, language translation, and sentiment analysis [40]. In this regard, empirical evidence with limited and noisy data suggests that to measure the top-κ error (a common measure of performance in machine learning performed with deep neural networks trained with the cross entropy loss), the loss function must be smooth, meaning that it should incorporate a smoothing parameter ϵ to handle small probability events [41].
- **Feature selection.** Cross entropy is used to select significant features of binary values from highly imbalanced large datasets via a framework called FMC Selector [42].
- **Image analysis.** Wavelet analysis together with cross entropy are used in image segmentation, object recognition, texture analysis (e.g., fabric defect detection) and pattern classification [43].

- **Learning-to-rank methods.** In [44] the author proposes a learning-to-rank loss function that is based on cross entropy. Learning-to-rank methods form a class of ranking algorithms that are widely applied in information retrieval.
- **Multiclass classification.** Cross entropy is used to enhance the efficiency of solving support vector machines for multi-class classification problems [45].
- **Semi-supervised clustering.** Cross entropy is employed along with the information bottleneck method in semi-supervised clustering. It is robust to noisy labels and automatically determines the optimal number of clusters under mild conditions [46].

2.5. Differential Entropy

As said in Section 1.2, differential entropy is the continuous counterpart of Shannon entropy: If X is a continuous random variable with density $\rho(x)$ and support set S, then the differential entropy of X is [9]

$$h(X) = h(\rho) = -\int_S \rho(x) \log \rho(x) dx. \tag{9}$$

Therefore, differential entropy is used in the analysis of continuous random variables and processes, e.g., analog signals.

Applications

- **Anomaly detection.** Differential entropy can measure changes in the probability density function of an analog signal which reveals an anomaly in the source, whether it is a mechanical system or a patient [47].
- **Emotion recognition.** Differential entropy has been used in [25] to extract features in EEG-based human emotion recognition.
- **Feature selection.** A feature selection algorithm based on differential entropy to evaluate feature subsets has been proposed in [48]. This algorithm effectively represents uncertainty in the boundary region of a fuzzy rough model and demonstrates improved performance in selecting optimal feature subsets, thereby enhancing classification accuracy; see [49] for an implementation.
- **Generative models.** Variational Autoencoders and other generative models leverage differential entropy to model the latent space of continuous data distributions. These models can learn better representations of the input data, thus improving the performance [50].
- **Mutual information.** Differential entropy is instrumental to compute the mutual information of continuous-valued random variables and processes, e.g., autoregressive processes. Seen in speech processing (linear prediction), seismic signal processing and biological signal processing [51].
- **Probabilistic models.** Differential entropy is utilized in probabilistic models such as Gaussian Mixture Models to describe the uncertainty and distribution of continuous variables. This approach is applicable to image processing and network inference as well [52].

2.6. Dispersion Entropy

Dispersion entropy characterizes time series data based on the Shannon Entropy. Introduced by Rostaghi and Azami in 2016 [53] to address limitations in other entropy measures, dispersion entropy includes a transformation that translates data into discrete symbols [16].

Applications

- **Feature extraction.** Multiscale fuzzy dispersion entropy is applied in fault diagnosis of rotating machinery to capture the dynamical variability of time series across various scales of complexity [54].
- **Image classification.** A multiscale version of dispersion entropy called $MDispEn_{2D}$ has been used with biomedical data to measure the impact of key parameters that may greatly influence the entropy values obtained in image classification [55].

- **Signal classification.** Other generalization of dispersion entropy, namely, fractional fuzzy dispersion entropy, has been proposed as a fuzzy membership function for signal classification tasks [56].
- **Signal denoising.** Dispersion entropy is also used in signal denoising via (i) adaptive techniques and group-sparse total variation [57], or (ii) empirical mode decomposition with adaptive noise [58].
- **Time series analysis.** Multiscale graph-based dispersion entropy is a generalization of dispersion entropy used to analyze multivariate time series data in graph and complex network frameworks, e.g., weather and two-phase flow data; it combines temporal dynamics with topological relationships [59].

2.7. Energy Entropy and Empirical Mode Decomposition Energy Entropy

Energy entropy is the Boltzmann-Gibbs-Shannon entropy of a normalized distribution of energy levels or modes. So it is mainly used in data-driven analysis of physical systems, whether in natural sciences or technology.

In nonlinear time series analysis, one uses empirical mode decomposition (EMD) to decompose a time series into a set of intrinsic mode functions, each representing a simple oscillatory mode inherent to the data. This decomposition is adaptive and data-driven, making it suitable for analysing complex signals without requiring predefined basis functions [60]. As its name suggests, EMD energy entropy combines energy entropy with EMD to provide information about the energy distribution across different intrinsic mode functions derived from a signal.

Applications

- **Chatter detection.** This application involves detecting vibrations and noise in machining operations that can indicate chattering. In this regard, energy entropy can detect chatter in robotic milling [61].
- **Fault prediction in vibration signal analysis.** EMD energy entropy has been employed to predict early fault of bearings in rotating machinery [62,63].
- **Feature extraction.** Energy entropy is calculated via the empirical decomposition of the signal into intrinsic mode functions and serves as a feature for machine learning models used in chatter detection. The chatter feature extraction method is based on the largest energy entropy [64].
- **Time-series forecasting.** EMD energy entropy was used in [65] to predict short-term electricity consumption by taking into account the data variability, i.e., that power consumption data is non-stationary, nonlinear, and influenced by the season, holidays, and other factors. In [66], this entropy was the tool to distinguish two kinds of financial markets.

2.8. Entanglement Entropy

Entanglement entropy originated in Quantum Mechanics as a measure of the degree of quantum entanglement between two subsystems of a quantum system. To be more precise, entanglement entropy is the von Neumann entropy (Section 2.32) of the reduced density matrix for any of the subsystems [67]; see [68] for the estination of entanglement entropy through supervised learning. In addition to its important role in quantum mechanics and quantum information theory, entanglement entropy is increasingly finding more applications in machine learning as well. This is why we have included entanglement entropy in this review.

Applications

- **Feature extraction.** In quantum machine learning, entanglement entropy is used for feature extraction by representing data in a form that highlights quantum correlations and thus, leveraging the quantum properties of the data [69].
- **Quantum models.** Entanglement entropy is used in quantum models to quantify unknown entanglement by using neural networks to predict entanglement measures of unknown quantum states based on experimentally measurable data: moments or correlation data produced by local measurements [70].

2.9. Excess Entropy

The *mutual information* between the random variables X and Y measures the average reduction in uncertainly about one of variables that results from learning the value of the other. It is defined as [9]

$$I(X;Y) = I(Y;X) = S(X) + S(Y) - S(X,Y) \geq 0, \tag{10}$$

where $S(...)$ is Shannon's entropy (1).

Excess entropy (also called dual total correlation) is a non-negative generalization of mutual information to more than two random variables defined as [71]

$$D(X_1, ..., X_N) = S(X_1, ..., X_N) - \sum_{i=1}^{N} S(X_i | X_1, ..., X_{i-1}, X_{i+1}, ..., X_N) \tag{11}$$

where $S(X_i|...)$ is the Shannon conditional entropy of X_i given the other variables. Of course, $D(X_1, X_2) = I(X_1; X_2)$. An alternative definition called *total excess entropy* was introduced by Crutchfield and Packard in [72].

Applications

- **Image segmentation.** Excess entropy is used to measure the structural information of a 2D or 3D image and then determine the optimal threshold in a segmentation algorithm proposed in [73]. The working hypothesis of this thresholding-based segmentation algorithm is that the optimal threshold corresponds to the maximum excess entropy (i.e., to a segmentation with maximum structure).
- **Machine learning.** In [74], the authors present a method called machine-learning iterative calculation of entropy, for calculating the entropy of physical systems by iteratively dividing the system into smaller subsystems and estimating the mutual information between each pair of halves.
- **Neural estimation in adversarial generative models.** Mutual Information Neural Estimator is a scalable estimator used in high dimensional continuous data analysis that optimizes mutual information. The authors apply this estimator to Generative Adversarial Networks (GANs) [75].
- **Time series analysis.** In this application, total excess entropy is used for classifying stationary time series into long-term and short-term memory. A stationary sequence with finite block entropy is long-term memory if its excess entropy is infinite [76].

2.10. Fluctuation-Based Dispersion Entropy

Fluctuation-based dispersion entropy (FDispEn) was introduced by Azami and Escudero in 2018 [77] to quantify the complexity and dynamical variability of time-series by incorporating information about fluctuations into the original dispersion entropy framework. This enhancement allows for a more detailed analysis of the signal's inherent dynamics.

Applications

- **Fault diagnosis.** The so-called refined composite moving average FDispEn is used in machinery fault diagnosis by analysing vibration signals [78].

  Refined composite multiscale FDispEn and supervised manifold mapping are used in fault diagnosis for feature extraction in planetary gearboxes [79].

  Multivariate hierarchical multiscale FDispEn along with multi-cluster feature selection and Gray-Wolf Optimization-based Kernel Extreme Learning Machine helps diagnose faults in rotating machinery. It also captures the high-dimensional fault features hidden in multichannel vibration signals [80].
- **Feature extraction.** FDispEn gives rise to hierarchical refined multi-scale fluctuation-based dispersion entropy, used to extract underwater target features in marine environments and weak target echo signals, thereby improving the detection performance of active sonars [81].

- **Robustness in spectrum sensing.** Reference [82] proposes a machine learning implementation of spectrum sensing using an improved version of the FDispEnt as a feature vector. This improved version shows enhanced robustness to noise.
- **Signal classification.** FDispEn helps distinguish various physiological states of biomedical time series and it is commonly used in biomedicine. It is also used to estimate the dynamical variability of the fluctuations of signals applied to neurological diseases [83]. Fluctuation-based reverse dispersion entropy is applied to signal classification combined with k-nearest neighbor [84].
- **Time series analysis.** FDispEn is used to quantify the uncertainty of time series to account for knowledge on parameters sensitivity and studying the effects of linear and nonlinear mapping on the defined entropy in [77]. FDispEn is defined as a measure for dealing with fluctuations in time series. Then, the performance is compared to complexity measures such as permutation entropy (Section 2.20), sample entropy (Section 2.25), and Lempel-Ziv complexity [9], [85].

2.11. Fourier Entropy

The Fourier entropy $h(f)$ of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is the Shannon entropy of its power spectrum $\{\hat{f}(S)^2 : S \subset [n]\}$, where $S \subset [n]$ stands for the 2^n subsets of $\{1, 2, ..., n\}$ (including the empty set) and $\hat{f}$ is the Fourier transform of f. By Parseval's Theorem, $\sum_{S \subset [n]} \hat{f}(S)^2 = 1$, so the power spectrum of f is a probability distribution.

Applications

- **Decision trees:** The Fourier Entropy-Influence Conjecture, made by Friedgut and Kalai [86], says that the Fourier entropy of any Boolean function f is upper bounded, up to a constant factor, by the total influence (or average sensitivity) of f. This conjecture, applied to decision trees, gives interesting results that boil down to $h(f) = O(\log(L(f)))$, meaning that $|h(f)| < C \log(L(f))$, where $L(f)$ denotes the minimum number of leaves in a decision tree that computes f and $C > 0$ is independent of f and $L(f)$ [87]. Another similar application to decision trees can be found in [88].
- **Learning theory.** The Fourier Entropy-Influence Conjecture is closely related to the problem of learning functions in the membership model. It is said that if a function has small Fourier entropy it means that its Fourier transform is concentrated on a few characters, i.e. the function can be approximated by a sparse polynomial, a class that is very important in the context of learning theory [89]. Learning theory provides the mathematical foundation for understanding how algorithms learn from data, guiding the development of machine learning models.

2.12. Fractional Fourier Entropy

Fractional Fourier entropy contains two ingredients: the fractional Fourier transforms of the data [90] and the Shannon entropy of probability distributions related to the resulting frequency spectrum, e.g., the power spectrum.

Applications

- **Anomaly detection in remote sensing.** Fractional Fourier entropy is used in hyperspectral remote sensing to distinguish signals from background and noise [91], [92].
- **Artificial intelligence.** Two-dimensional fractional Fourier entropy helps to diagnose COVID-19 by extracting features from chest CT images [93].
- **Biomedical image classification.** Fractional Fourier entropy has proven helpful in detecting pathological brain conditions. By using it as a new feature in Magnetic Resonance Imaging, the classification of images is improved in time and cost [94].
- **Deep learning.** Fractional Fourier entropy is used in the detection of gingivitis via feed-forward neural networks. It reduces the complexity of image extraction before classification and can obtain better image eigenvalues [95].

- **Emotion recognition.** Fractional Fourier entropy, along with two binary support vector machines, helps improve the accuracy of emotion recognition from physiological signals in electrocardiogram and galvanic skin responses [96].
- **Multilabel classification.** Fractional Fourier entropy has been used in a tea-category identification system, which can automatically determine tea category from images captured by a 3 charge-coupled device digital camera [97].

2.13. Fuzzy Entropy

Fuzzy entropy, introduced by Ishikawa and Mieno in 1979 [98], quantifies the uncertainty or fuzziness within a system, thus extending entropy concepts to fuzzy sets to better represent uncertainty. In 2014, Zheng proposed multiscale fuzzy entropy [99]. Fuzzy entropy was introduced into fuzzy dynamical systems in [100].

Applications

- **Clustering and time-series analysis.** Fuzzy entropy is used in problems of robustness against outliers in clustering techniques in [101].
- **Data analysis.** Fuzzy entropy is proposed in [102] to assess the strength of fuzzy rules with respect to a dataset, based on the greatest energy and smallest entropy of a fuzzy relation.
- **Fault detection.** Fuzzy entropy (along with dispersion entropy, Section 2.6) was the best performer in a comparative study of entropy-based methods for detecting motor faults [103]. Multiscale fuzzy entropy is used to measure complexity in time series in rolling bearing fault diagnosis [104].
- **Feature selection and mathematical modelling.** Fuzzy entropy is used in feature selection to evaluate the relevance and contribution of each feature in Picture Fuzzy Sets [105].
- **Image classification.** Fuzzy entropy, in the form of multivariate multiscale fuzzy entropy, is proposed and tested in [106] for the study of texture in color images and their classification.
- **Image segmentation.** Fuzzy entropy is the objective function of a colour image segmentation technique based on an improved cuckoo search algorithm [107].

2.14. Graph Entropy

Graph entropy was introduced by Korner in 1971 [108] to quantify the complexity or information content of a graph. It is usually defined as the Shannon entropy (although any other entropy would do) of a probability distribution over the graph's vertex set. In addition to applications in data analysis and machine learning, graph entropy is also applied in combinatorics; see [109] for a survey of graph entropy.

A particular case of graph entropy is the horizontal visibility (HV) graph entropy of a time series, which is the graph entropy of the so-called HV graph of the time series [110]. In particular, this method is useful for distinguishing between different types of dynamical behaviours in nonlinear time series, such as chaotic versus regular dynamics [111].

Applications

- **Dimension reduction and feature selection.** Graph entropy gives rise to the Conditional Graph Entropy that helps in the alternating minimization problem [112].
- **Graph structure.** Graph entropy is used to measure the information content of graphs, as well as to evaluate the complexity of the hierarchical structure of a graph [113].
- **Graph-based time series analysis.** Graph entropy can be used in time series analysis in conjunction with any method that transforms time series into graphs. An example is the HV graph entropy presented above; see [114] and references therein.
- **Node embedding dimension selection.** Graph entropy is applied in Graph Neural Networks through the Minimum Graph Entropy algorithm. It calculates the ideal node embedding dimension of any graph [115].

- **Time series analysis.** HV graph entropy along with sample entropy has been used in [116] to identify abnormalities in the EEGs of alcoholic subjects. HV transfer entropy was proposed in [117] to estimate the direction of the information flow between pairs of coupled time series.

2.15. Havrda–Charvát Entropy

Havrda–Charvát (HC) entropy, also known as the Havrda–Charvát α-entropy, was introduced by Havrda and Charvát [14] in 1967 in Information Theory. It is formally identical to the Tsallis entropy (Section 2.31, introduced by Tsallis in 1988 in Statistical Mechanics [15]. Similar to Rényi entropy (Section 2.24), HC entropy is a family of entropies parameterized by $\alpha > 0$ and generalizes the Shannon entropy in the sense that the HC entropy coincides with the Shannon entropy in the limit $\alpha \to 1$.

Although nowadays the most popular name for this entropy is Tsallis entropy, we present in this section applications published in articles that refer to this entropy as the Havrda–Charvát entropy.

Applications

- **Computer vision.** An HC entropy-based technique for group-wise registration of point sets with unknown correspondence is used in graphics, medical imaging and pattern recognition. By defining the HC entropy for cumulative distribution functions (CDFs), the corresponding CDF-HC divergence quantifies the dissimilarity between CDFs estimated from each point-set in the given population of point sets [118].
- **Financial time series.** Weighted HC entropy outperforms regular HC entropy when used as a complexity measure in financial time series. The weights turn out to be useful for showing amplitude differences between series with the same order mode (i.e. similarities in patterns or specific states) and robust to noise [119].
- **Image segmentation and classification.** HC entropy is applied as loss function in image segmentation and classification tasks using convolutional neural networks in [120].
- **Loss functions in deep learning.** HC entropy can be used to design loss functions in deep learning models. These loss functions are particularly useful in scenarios with small datasets, common in medical applications [121].

2.16. Intrinsic Mode Entropy

Intrinsic mode entropy (IME) computes the sample entropy (Section 2.25) over different scales of intrinsic mode functions extracted by the empirical mode decomposition method (Section 2.7). It was introduced by Amoud et al. in 2007 [122].

Applications

- **Language gesture recognition.** IME is used in [123] to analyse data from a 3-dimensional accelerometer and a five-channel surface electromyogram of the user's dominant forearm for automated recognition of Greek sign language gestures.
- **Neural data analysis.** An IME version with improved discriminatory capacity in the analysis of neural data is proposed in [124].
- **Time series analysis.** IME is used in nonlinear time series analysis to efficiently characterize the underlying dynamics [122]. As any multiscale entropy, IME is particularly useful for the analysis of physiological time series [125]. See also [126] for an application to the analysis of postural steadiness.

2.17. Kaniadakis Entropy

Kaniadakis entropy, also known as κ-entropy due to its dependence on a parameter $0 < \kappa < 1$, was introduced by the physicist G. Kaniadakis in 2002 [127] to address the limitations of classical entropy in systems exhibiting relativistic effects. It is defined as

$$S_\kappa(p_1, ..., p_W) = \sum_{i=1}^{W} \frac{p_i^{1-\kappa} - p_i^{1+\kappa}}{2\kappa}. \tag{12}$$

Kaniadakis entropy is a relativistic generalization of the BGS entropy (4) in the sense that the latter is recovered in the $\kappa \to 0$ limit [128]. Like other physical entropies such as Tsallis' and von Neumann's, Kaniadakis entropy has also found interesting applications in applied mathematics.

Applications

- **Image segmentation.** Kaniadakis entropy is used in image thresholding to segment images with long-tailed distribution histograms; the parameter κ is selected via a swarm optimization search algorithm [129].
- **Images threshold selection.** Kaniadakis entropy can be used to construct an objective function for image thresholding. By using the energy curve and the Black Widow optimization algorithm with Gaussian mutation, this approach can be performed on both grayscale and colour images of different modalities and dimensions [130].
- **Seismic imaging.** Application of the Maximum Entropy Principle with S_κ leads to the Kaniadakis distribution, a deformation of the Gaussian distribution that has application, e.g., in seismic imaging [131].

2.18. Kernel Entropy

Kernel (or kernel-based) entropy is an evolution of the approximate entropy (Section 2.1) that consists of replacing the Heaviside step function in the definition of proximity by other functions (or "kernels") to give more weight to the nearest neighbors. Thus, in the Gaussian kernel entropy, the Gaussian kernel

$$\ker(i, j; r) = \exp\left(-\frac{\|x_i - x_j\|^2}{10r^2}\right) \tag{13}$$

is used. Here x_i, x_j are entries of a time series and r is the parameter of the approximate entropy. Other popular kernels include the spherical, Laplacian and Cauchy functions [132].

Of course, the same refinement can be done in the computation of the sample entropy (Section 2.25), an improvement of the approximate entropy. To distinguish between the two resulting kernel entropies, one speaks of kernel-based approximate or sample entropy.

Applications

- **Complexity of time-series.** The authors of [133] present experimental evidence that Gaussian kernel entropy outperforms approximate entropy when it comes to analyze the complexity of time series.
- **Fetal heart rate discrimination.** In [134] the authors compare the performance of several kernel entropies on fetal heart rates discrimination, with the result that the circular and Cauchy kernels outperform other, more popular kernels, such as the Gaussian or the spherical ones.
- **Parkinson's disease.** Gaussian kernel entropy, along with other nonlinear features, is used in [135] in the task of automatic classification of speech signals from subjects with Parkinson's disease and a control set.
- **Pathological speech signal analysis.** Reference [132] is a study of several approaches in the field of pathological speech signal analysis. Among the new pathological voice measures, the authors include different kernel-based approximate and sample entropies.

2.19. Kolmogorov-Sinai Entropy

As mentioned in Section 1.2, Kolmogorov-Sinai entropy is a classical entropy introduced by Kolmogorov in ergodic theory in 1958 [136] and extended by Sinai to the theory of measure-preserving dynamical systems in 1959 [5]. Although it is defined using refinements of dynamical partitions of the state space, it is usually estimated via Pesin's formula [137], which involves the strictly positive Lyapunov exponents of the system; see, e.g., [138] and [139] for other numerical methods. Kolmogorov-Sinai entropy is a fundamental invariant in the theory of metric dynamical systems.

Applications

- **Time series analysis.** The perhaps main practical application of the Kolmogorov-Sinai entropy is the analysis of nonlinear, real-valued time series, where it is used to characterize the underlying dynamical system, in particular, its chaotic behavior. Recent practical examples include short-term heart rate variability [140], physical models of the vocal membranes [139], autonomous driving [141], and EEG-based human emotion recognition [25,142].

2.20. Permutation Entropy

The conventional (or Shannon) permutation entropy of a time series was introduced by Bandt and Pompe in 2002 [143]. It is the Shannon entropy (Section 2.26) of the probability distribution obtained from the ordinal patterns of length $L \geq 2$ in the time series, i.e., the rankings (or permutations) of the series entries in sliding windows of size L. Therefore, permutation entropy depends on the parameter L. In the particular case that the time series is an orbit generated by a selfmap f of a one-dimensional interval, then the permutation entropy rate converges to the Kolmogorov entropy of f (Section 2.19) when $L \to \infty$ [144]. This means that the permutation entropy calculated with a sufficiently large L (and divided by $L-1$) is a good estimator of the Kolmogorov entropy of one-dimensional dynamics. Furthermore, permutation entropy is easy to program, relatively robust to noise, and can be computed practically in real time since knowledge of the data range is not needed [145]. See, e.g., [146–149] for references on theoretical and practical aspects of permutation entropy.

If, instead of the Shannon entropy of the ordinal patterns distribution, we use other entropy, e.g., Rényi entropy (Section 2.24 or Tsallis entropy (2.31, then we obtain the corresponding "permutational version": permutation Rényi entropy, permutation Tsallis entropy, and more [150],[151]. There are also "weighted versions" that take into account not only the rank order of the entries in a window but also their amplitudes; see, e.g., [152]. In turn, "multiscale versions" (including multiscale permutation Rényi and Tsallis entropy) account for multiple time scales in time series by using different time delays [153–155].

Applications

- **Analysis of EEGs.** One of the first applications of permutation entropy was the analysis of EEGs of subjects with epilepsy because normal and abnormal signals (during epileptic seizures) have different complexities [156]. Furthermore, since permutation entropy can be computed in virtually real time, it has been used to predict seizures in epilepsy patients by tracking dynamical changes in EEGs [147]. Further examples can be found in the reviews [18,147]. Results can be improved using permutation Rényi and Tsallis entropy due to their additional, fine-tunable parameter [150,157].
- **Determinism detection.** Time series generated by one-dimensional maps have necessarily forbidden ordinal patterns of all sufficiently large lengths L [146]. Theoretical results under some provisos and numerical results in other cases show that the same happens with higher dimensional maps [158,159]. Therefore, the scaling of permutation entropy with L can distinguish noisy deterministic signals from random signals [146,160].
- **Emotion recognition.** Permutation entropy is used to help in tasks of feature extraction in EEGs [25].
- **Nonlinear time series analysis.** Permutation entropy has been extensively used in the analysis of continuous-valued time series for its straightforward discretization of the data and ease of calculation. Numerous applications can be found, e.g., in [147–149] and the references therein.
- **Obstructive sleep apnea.** A combination of permutation entropy-based indices and other entropic metrics was used in [161] to distinguish subjects with obstructive sleep apnea from a control group. The data consisted of heart rate and beat-to-beat blood pressure recordings.
- **Prediction.** Permutation entropy has been used along with variational modal decomposition to predict wind power [162].

- **Speech signals.** In their seminal paper [143], Bandt and Pompe used precisely permutation entropy to analyze speech signals and showed that it is robust with respect to the window length, sampling frequency and observational noise.
- **The causality-complexity plane.** Permutation entropy together with the so-called statistical complexity builds the causality-complexity plane, that has proven to be a powerful tool to discriminate and classify time series [163]. By using variants of the permutation entropy and the statistical complexity, the corresponding variants of the causality-complexity plane are obtained, possibly with enhanced discriminatory abilities for the data at hand [149].
- **Unstructured data.** Nearest-neighbor permutation entropy is an innovative extension of permutation entropy tailored for unstructured data, irrespective of their spatial or temporal configuration and dimensionality, including, e.g., liquid crystal textures [164].

2.21. Rao's Quadratic Entropy

Rao's quadratic entropy (RQE, not to be confused with Rényi's collision entropy R_2, Section 2.24, also called quadratic entropy) was proposed in 1982 [165] as a measure of diversity in biological populations. Given W species, RQE is defined as

$$\text{RQE}(p_1, ..., p_W) = \sum_{i,j=1}^{W} \delta_{i,j} p_i p_j \tag{14}$$

where $\delta_{i,j}$ is the difference between the i-th and the j-th specie and $\{p_1, ..., p_W\}$ is the probability distribution of the W species in the multinomial model.

Applications

- **Environmental monitoring.** RQE helps calculate the environmental heterogeneity index and assist prioritization schemes [166].
- **Genetic diversity metrics.** RQE is used to measure diversity for a whole collection of alleles to accommodate different genetic distance coding schemes and computational tractability in case of large datasets [167].
- **Unsupervised classification.** RQE is used as a framework in the support vector data description algorithm for risk management, enhancing knowledge in terms of interpretation, optimization, among others [168].

2.22. Rank-Based Entropy

Rank-based entropy (RbE), introduced by Citi et al. in 2014 [169] within the framework of multiscale entropy, measures the disorder or uncertainty in a dataset by analysing the ranks of the data rather than their absolute values, it is useful for ordinal data so that a proper rank can be created. This measure captures the complexity and variability in the data by examining how the ranks of the data points are distributed.

Applications

- **Anomaly detection.** RbE is applied in mixed data analysis to check the influence of categorical features, using Jaccard index for anomaly ranking and classification [170].
- **Feature selection.** RbE is used in the Entropy-and-Rank-based-Correlation framework to select features, e.g., in the detection of fruit diseases [171].
- **Mutual information.** RbE is used to rank mutual information in decision trees for monotonic classification [172].
- **Node importance.** RbE is employed in the analysis of graphs to rank nodes taking into account the local and global structure of the information [173].

- **QSAR models.** RbE is employed in Quantitative Structure-Activity Relationship models (QSAR) to analyse their stability via "rank order entropy", suggesting that certain models typically used should be discarded [174].
- **Time series analysis.** RbE is used in terms of correlation entropy to test serial independence in [175]. A multiscale version was used in [169] to study data of heart rate variability.
- **Time series classification.** RbE helps classify order of earliness in time series to generate probability distributions in different stages [176].

2.23. Relative Entropy

Relative entropy, also known as Kullback-Leibler (KL) divergence [9], is an information-theoretical measure that quantifies the difference or "distance" between two probability distributions. Specifically, if p $= (p_1, ..., p_W)$ and q $= (q_1, ..., q_W)$ are two probability distributions, then the relative entropy or KL divergence from p to q, $D(\mathrm{p}\|\mathrm{q})$, is defined as

$$D(\mathrm{p}\|\mathrm{q}) = \sum_{i=1}^{W} p_i \log \frac{p_i}{q_i}. \tag{15}$$

It follows that $D(\mathrm{p}\|\mathrm{q}) \geq 0$ (Gibb's inequality) and $D(\mathrm{p}\|\mathrm{q}) = 0$ if and only if p $=$ q. Note that $D(p\|q)$ is not a distance in the strict sense because $D(\mathrm{p}\|\mathrm{q}) \neq D(\mathrm{q}\|\mathrm{p})$ in general, although it can be easily symmetrized by taking any mean (arithmetic, geometric, harmonic,...) of $D(p\|q)$ and $D(q\|p)$. Therefore, if p is a true probability distribution approximated with q, then $D(\mathrm{p}\|\mathrm{q})$ is a measure of the approximation error. As another useful example, if p is a bivariate joint distribution and q is the product distribution of the two marginals, then $D(p\|q)$ is the mutual information between the random variables defined by the marginal distributions, equation (10).

See [11] for a generalizacion of divergence (or relative entropy), where the log p_i/q_i in equation (15) is replaced by $f(p_i/q_i)$ where f is a convex function on $(0, \infty)$ with $f(1) = 0$.

Applications

- **Anomaly detection.** KL divergence has been used for plane control in Software-Defined Networking as a method to detect Denial of Service attacks in [177].
- **Bayesian networks.** The efficient computation of the KL divergence of two probability distributions, each one coming from a different Bayesian network (with possibly different structures), has been considered in [178].
- **Feature selection.** The authors of [179] show that the KL divergence is useful in information-theoretic feature selection due to the fact that maximising conditional likelihood corresponds to minimising KL-divergence between the true and predicted class posterior probabilities.
- **Multiscale errors.** KL divergence is a useful metric for multiscale errors [48]. A recent application to the study of the behavior of various nonpolar liquids via the Relative Resolution algorithm can be found in [180].
- **Parameter minimization in ML.** Parameters that minimize the KL divergence minimize also the cross entropy and the negative log likelihood. So, the KL divergence is useful in optimization problems where the loss function is a cross-entropy [181].

2.24. Rényi Entropy

Rényi entropy R_α, where $\alpha > 0$ and $\alpha \neq 1$, was introduced by Alfréd Rényi in 1961 [12] as a generalization of Shannon entropy in the sense that R_1 is set equal to the latter by continuity; see Section 1.2 and the review [13] for detail. The parameter α allows for different emphasis on the probabilities of events, making it a versatile measure in information theory and its applications. Thus, for $\alpha < 1$ the central part of the distribution is flattened, i.e., high probability events are suppressed, and low-probability events are enhanced. The opposite happens when $\alpha < 1$. As a function of the parameter α, the Rényi entropy is non-increasing. Particular cases include the Hartley entropy or max-entropy $R_0 = \lim_{\alpha \to 0} R_\alpha$, the collision or quadratic entropy R_2, and the min-entropy $R_\infty = \lim_{\alpha \to \infty} R_\alpha$.

Applications

- **Anomaly Detection.** Lower values of α highlight rare events, making this measure useful for identifying anomalies [182,183]. In particular, Rényi entropy is used in network intrusion detection for detecting botnet-like malware based on anomalous patterns [184].
- **Automated identification.** Average Renyi entropy, along with other entropic measures, have been used as inputs for SVM algorithms to classify focal or non-focal EEGs of subjects affected by partial epilepsy [185].
- **Clustering.** Rényi entropy can provide robust similarity measures that are less sensitive to outliers [182].
- **Extreme entropy machines.** Rényi's quadratic entropy R_2 is used in the construction of extreme entropy machines to improve classification problems [186].
- **Feature selection and character recognition.** Adjustment of the parameter α can help to emphasize different parts of the underlying probability distribution and hence the selection of the most informative features. Rényi entropy is used for feature selection in [182,187]. Max-entropy is used in [188] for convolutional feature extraction and improvement of image perception.
- **Medical time series analysis** Applications of the Rényi entropy in time series analysis reach from epilepsy detection in EEG (see, e.g., [26]) over artifact rejection in multichannel scalp EEG (see [189] and references therein) to early diagnosis of Alzheimer's disease in MEG data (see, e.g., [190].

2.25. Sample Entropy

Sample Entropy was introduced by Richman and Moorman in 2000 [191] as an improvement over approximate entropy (Section 2.1), namely, its calculation is easier and independent of the time series length. Sample entropy is the negative natural logarithm of the conditional probability that close sequences of m points remain close when one more point is added, within a tolerance $r > 0$.

Applications

- **Automated identification.** Average sample entropy and other entropy measures are used as input for an SVM algorithm to classify focal and non-focal EEG signals of subjects with epilepsy [185].
- **Fault diagnosis.** Sample entropy has been used for multi-fault diagnosis in lithium batteries [192].
- **Image classification.** Sample entropy, in the form of multivariate multiscale sample entropy, is used for classifying RGB colour images to compare textures, based on a threshold to measure similarity [106].
- **Image texture analysis.** Two-dimensional sample entropy has shown to be a useful texture feature quantifier for the analysis of biomedical images [193].
- **Mutual information.** Modified sample entropy has been used in skin blood flow signals to analyse mutual information and, hence, study the association of microvascular dysfunction in different age groups [194].
- **Neurodegenerative disease classification.** Sample entropy is used to classify neurodegenerative diseases. Gait signals, support vector machines and nearest neighbours are employed to process the features extracted using sample entropy [195].
- **Short signal analysis.** Quadratic sample entropy is an evolution of sample entropy that has been proposed for the analysis of short length biomedical signals [196], [197].
- **Time-series analysis.** Sample entropy, often in the form of multiscale sample entropy, is a popular tool in time series analysis, in particular with biomedical data [198]. For example, it is used for the fast diagnosis and monitoring of Parkinson's disease [199] and human emotion recognition [25] using EEGs. A modified version of multiscale sample entropy has recently been used for diagnosing epilepsy [200]. See [29] for an overview of applications of sample entropy to the analysis of physiological time series.
- **Weather forecasting.** Sample entropy is applied in weather forecasting by using transductive feature selection methods based on clustering-based sample entropy [201].

2.26. Shannon Entropy

Shannon entropy was introduced by Claude Shannon in his foundational 1948 paper "A Mathematical Theory of Communication" [2] as the cornerstone of digital and analog Information Theory. Indeed, Shannon entropy informs the core theorems of Information Theory [9]. In the case of discrete probability distributions, the Shannon entropy measures the uncertainty about the outcome of a random variable with the given distribution or, alternatively, the expected information conveyed by that outcome, maximum uncertainty (or minimum information) being achieved by uniform distributions. It also quantifies the rate of information growth produced by a data source modelled as a stationary random process. In the case of continuous probability distributions, the Shannon entropy is called differential entropy and its applications to data analysis and machine learning were the subject of Section 2.5.

In this section we only consider discrete probability distributions, i.e., finite state random variables and processes (possibly after a discretization or symbolization of the data). A typical example are real-valued time series, where the Shannon entropy is used to measure their complexity and, hence, distinguish between different dynamics. The Shannon entropy of certain probability mass distributions (e.g., ordinal patterns of a time series, power spectrum of a signal, eigenvalues of a matrix) may have particular names (permutation, spectral, von Neumann entropies); in this case, the applications of such entropies are presented in the corresponding sections.

 Applications

- **Accurate prediction.** Shannon entropy is employed in machine learning models to improve the accuracy of predictions of molecular properties in the screening and development of drug molecules and other functional materials [202].
- **Anomaly detection.** Shannon entropy is employed in sensors (Internet of Things) to identify anomalies using the CorrAUC algorithm [203].
- **Artificial intelligence.** Shannon entropy contributes to the creation of the Kolmogorov Learning Cycle, which acts as a framework to optimize "Entropy Economy", helped by the intersection of Algorithmic Information Theory (AIT) and Machine Learning (ML). This framework enhances the performance of the Kolmogorov Structure Function, leading to the development of "Additive AI". By integrating principles from both AIT and ML, this approach aims to improve algorithmic efficiency and effectiveness, driving innovation in AI by balancing information theory with practical machine learning applications [204].
- **Automated identification.** Average Shannon entropy and other entropy measures are used as inputs of an SVM algorithm to classify focal or non-focal EEG signals of subjects with epilepsy [185].
- **Fault bearing diagnosis.** Multi-scale stationary wavelet packet analysis and the Fourier amplitude spectrum are combined to obtain a new discriminative Shannon entropy feature that is called stationary wavelet packet Fourier entropy in [205]. Features extracted by this method are then used to diagnose bearing failure.
- **Feature selection.** Shannon's mutual information entropy is used to design an entire information-theoretic framework that improves the selection of features in [179]. Shannon entropy is employed in biological science to improve classification of data in clustering of genes using microarray data [206].
- **Hard clustering.** Shannon entropy is used as a criterion to measure the confidence in unsupervised clustering tasks [207].
- **Natural language processing.** Shannon entropy quantifies the predictability (or redundancy) of a text. Therefore, it is instrumental in language modelling, text compression, information retrieval, among others [2,9]. For example, it is used in [208] for keyword extraction, i.e., to rank the relevance of words.
- **Policy learning.** Shannon entropy acts as a regularization inside of an iterative policy optimization method for certain quadratic linear control scenarios [209].

- **Signal analysis.** Shannon entropy is used as the cost functional of compression algorithms in sound and image processing [210].
- **Statistical inference.** According to the Maximum Entropy Principle of Jaynes [211], "in making inferences on the basis of partial information we must use the probability distribution which has maximum entropy subject to whatever is known". This principle has been traditionally applied with the Shannon entropy and several moment constraints of a probability distribution to infer the actual distribution [9,18].

2.27. Spectral Entropy

Spectral entropy, proposed by Kapur and Kesavan in 1992 [212], is an entropy based on the Shannon entropy. Here, the probability distribution is the (continuous or discrete) power spectrum in a representative frequency band, obtained from a signal or time series via the Fourier transform, and conveniently normalized. Hence, the spectral entropy characterizes a signal by the distribution of power among its frequency components.

Applications

- **Audio analysis.** Spectral entropy has been applied for robust audio content classification in noisy signals [213]. Specifically, spectral entropy is used to segment input signals into noisy audio and noise. Also, spectral entropy (in the form of Multiband Spectral Entropy Signature) has been shown to outperform other approaches in the task of sound recognition [214].
- **Damage event detection.** Spectral entropy detects damage in vibration recordings from a wind turbine gearbox [215].
- **Data time compression.** Spectral entropy has been successfully applied to identify important segments in speech, enabling time-compression of speech for skimming [216].
- **Deep learning synchronization.** Spectral entropy evaluates synchronization in neuronal networks, providing analysis of possibly noisy recordings collected with microelectrode arrays [217].
- **Feature extraction.** Spectral entropy is used to extract features from EEG signals in [25] (emotion recognition) and [218] (assessment of the depth of anaesthesia).
- **Hyperspectral anomaly detection.** Hyperspectral Conditional Entropy enters into the Entropy Rate Superpixel Algorithm, which is used in hyper spectral-spatial data to recognize unusual patterns [219].
- **Signal detection.** Spectral entropy has been used to detect cetacean vocalization in marine audio data [220]. The time frequency decomposition was done with the short time Fourier transform and the continuous wavelet transform.

2.28. Tone Entropy

Tone entropy was proposed by Oida et al. in 1997 [221] to study heart period fluctuations in electrocardiograms. Specifically, tone entropy is the Shannon entropy of a probability distribution derived from the percentage index of heart period variation. Therefore, its applications are mainly in cardiology.

Applications

- **Biomedical analysis.** Tone entropy has been employed to study the autonomic nervous system in age groups at high-risk of cardiovascular diseases [222]. In [223], tone entropy was used to study the influence of gestational ages on the development of the foetal autonomic nervous system by analyzing the foetal heart rate variability.
- **Time series.** Tone entropy has been used in time series analysis to differentiate between physiologic and synthetic interbeat time series [224].

2.29. Topological and Topology-Based Entropies

Topological entropy was introduced in [6] to measure the complexity of continuous dynamics on topological spaces. On metric spaces, topological entropy measures the exponential growth rate of the number of distinguishable orbits with finite precision. See, e.g., [225,226] for exact formulas and fast algorithms to compute the topological entropy of piecewise monotone maps and multimodal maps.

In time series analysis and digital communication technology, one is mostly interested in spaces with a finite number of states. Such time series can be the result symbolizing a continuous-valued time series. In this case, the states are usually called letters (or symbols), the state space is called alphabet and the blocks of letters are called words (which correspond to the "admissible" or "allowed" strings of letters). If $A(n)$ is the number of words of length n, then the topological entropy of the time series is defined as

$$h_{top} = \lim_{n \to \infty} \log(A(n)/n), \tag{16}$$

where the base of the logarithm is usually 2 or e.

Along with the above "conventional" topological entropies in dynamical systems and time series analysis, there are a number of adhoc entropies in time series analysis, sometimes also called topological entropies. This name is due to the fact that those entropies are based on topological properties extracted from the data, for example, via graphs or persistent homology. For clarity, here we refer to them as topology-based entropies. Examples include graph entropy and horizontal visibility graph entropy (Section 2.14). See [227] for an account of topological methods in data analysis.

Applications

- **Cardiac dynamics.** Given a (finite) time series, the *out-link entropy* is derived from the adjacency matrix of its ordinal network. This entropy has been used to classify cardiac dynamics in [228].
- **Convolutional neural networks.** The authors of [229] propose a method for quantitatively clarifying the status of single unit in convolutional neural networks using algebraic topological tools. Unit status is indicated via the calculation of a topology-based entropy, called *feature entropy*.
- **Damage detection.** Persistent entropy can also address the damage detection problem in civil engineering structures. In particular, to solve the supervised classification damage detection problem [230].
- **Detection of determinism.** Permutation topological entropy (i.e., the topological entropy of the distribution of ordinal patterns of length $L \geq 2$ obtained from a time series) can be used to detect determinism in continuous-valued time series. Actually, it suffices to check the growth of ordinal patterns with increasing L's, since this growth is exponential for deterministic signals (h_{top} converges to a finite number) and factorial for random ones (h_{top} diverges) [146,158].
- **Financial time series.** Topological entropy has been applied to horizontal visibility graphs of financial time series in [231].
- **Similarity of piecewise linear functions.** Piecewise linear functions are a useful mathematical tool in different areas of applied mathematics, including signal processing and machine learning methods. In this regard, *persistent entropy* (a topological entropy based on persistent homology) can be used to measure their similarity [232].

2.30. Transfer Entropy

A relevant question in time series analysis of coupled random or deterministic processes is the causality relation, i.e., which process is driving and which is responding. Transfer entropy, introduced by Schreiber in 2000 [233], measures the information exchanged between two processes in both directions separately. It can be considered as an information-theoretical (or nonlinear) implementation of Granger causality [234].

Given two stationary random processes $\mathbf{X} = (X_t)_{t \geq 0}$ and $\mathbf{Y} = (Y_t)_{t \geq 0}$, the transfer entropy from $\mathbf{Y}$ to $\mathbf{X}$, $T_{\mathbf{Y} \to \mathbf{X}}$, is the reduction of uncertainty in future values of $\mathbf{X}$, given past values of $\mathbf{X}$, due to

the additional knowledge of past values of **Y**. For simplicity, we consider here the simplest (lowest dimensional) case:

$$T_{\mathbf{Y} \to \mathbf{X}} = S(X_{t+1}|X_t) - S(X_{t+1}|X_t, Y_t) \tag{17}$$

where $S(X_{t+1}|...)$ is the conditional Shannon entropy of the variable X_{t+1} on the other variable(s) [9]. If the process **Y** is not causal to **X** (i.e., they are independent), then $S(X_{t+1}|X_t, Y_t) = S(X_{t+1}|X_t)$ and $T_{\mathbf{Y} \to \mathbf{X}} = 0$; otherwise, $S(X_{t+1}|X_t, Y_t) < S(X_{t+1}|X_t)$ and $T_{\mathbf{Y} \to \mathbf{X}} > 0$. Observe that $T_{\mathbf{Y} \to \mathbf{X}}$ is not an entropy proper but a conditional mutual information, namely: $T_{\mathbf{Y} \to \mathbf{X}} = I(X_{t+1}; Y_t \mid X_t)$ [9,18].

Applications

- **Accelerated training in Convolutional Neural Networks (CNN).** The authors of [235] propose a training mechanism for CNN architectures that integrates transfer entropy feedback connections. In this way, the training process is accelerated as fewer epochs are needed. Furthermore, it generates stability, hence, it can be considered a smoothing factor.
- **Improving accuracy in Graph Convolutional Neural Networks (GCN).** The accuracy of a GCN can be improved by using node relational characteristics (such as heterophily), degree information, and feature-based transfer entropy calculations. However, depending on the number of grah nodes, the computation of the transfer entropy can significantly increase the computational load [236].
- **Improving neural network performance.** A small, few-layer artificial neural network that employs feedback can reach top level performance on standard benchmark tasks, otherwise only obtained by large feed-forward structures. To show this, the authors of [237] use feed-forward transfer entropy between neurons to structure feedback connectivity.
- **Multivariate time series forecasting.** Transfer entropy is used to establish causal relationships in multivariate time series converted into graph neural networks, each node corresponding to a variable and edges representing the casual relationships between the variables. Such neural networks are then used for prediction [238].
- **Time series analysis.** The main application of transfer entropy since its formulation has been the analysis of multivariate time series (whether biomedical, physical, economical, financial, ...) for revealing causal relationships via information directionality. See [239] and the references therein for the conceptual underpinnings and practical applications.

2.31. Tsallis Entropy

Tsallis entropy T_q, defined in Section 1.2, equation (7), was introduced by Tsallis in 1988 in Statistical Mechanics [15]. The parameter q takes the real values $q > 0$, $q \neq 1$; T_q converges to the Shannon entropy when $q \to 1$ [13]. Tsallis entropy is identical in form to the Havrda–Charvát entropy (Section 2.15).

Tsallis entropy is particularly useful for describing systems with non-extensive properties, such as long-range interactions, non-Markovian processes, and fractal structures. In machine learning, Tsallis entropy is used to improve algorithms in areas such as clustering, image segmentation, and anomaly detection by changing and fine tuning the parameter q. See [240] for the general properties of the Tsallis entropy.

Applications

- **Anomaly detection.** Tsallis entropy is used in network intrusion detection by detecting botnet-like malware based on anomalous patterns in the network [184].
- **Clustering.** A Tsallis entropy based categorical data clustering algorithm is proposed in [241]. It is shown there, that when the attributes have a power law behavior the proposed algorithm outperforms existing Shannon entropy-based clustering algorithms.
- **Feature selection.** Tsallis-entropy-based feature selection is used in [242] to identify significant features, which boosts the classification performance in machine learning. The authors propose an algorithm to optimize both the classifier (a Support Vector Machine) and Tsallis entropy parameters, so improving the classification accuracy.

- **Image segmentation.** Tsallis Entropy is maximized to use it for segmenting images by maximizing the entropy within different regions of the image [243].
- **Pre-seismic signals.** Tsallis entropy has been used in [244] to analyze pre-seismic electromagnetic signals.

2.32. Von Neumann Entropy

Von Neumann entropy [1] is the equivalent of Shannon entropy in Quantum Statistical Mechanics and Quantum Information Theory. It is defined via the density matrix of an ensemble of quantum states, which (i) is Hermitian, (ii) has unit trace, and (iii) is positive semidefinite. Therefore, the eigenvalues of the density matrix build a probability distribution and, precisely, the von Neumann entropy of the ensemble is the Shannon entropy of the probability distribution defined by the eigenvalues of the corresponding density matrix.

The same approach can be used with any matrix with the properties (i)-(iii), for example, the Pearson correlation matrix of a Markov chain (divided by the dimension of the matrix) or the Laplacian matrix of a graph. This explains the use of von Neumann entropy in classical data analysis as well.

Applications

- **Feature selection and dimensionality reduction.** Von Neumann entropy is employed in the case of kernelized relevance vector machines to asses dimensionality reduction for better model performances [245].
- **Graph-based learning.** In [246] the authors propose a method to identify vital nodes in hypergraphs that is based on von Neumann entropy. More precisely, this method is based on the high-order line graph structure of hypergraphs and measures changes in network complexity using von Neumann entropy.
- **Graph similarity and anomaly detection.** The von Neumann graph entropy (VNGE) is used to measure the information divergence and distance between graphs in a sequence. This is used for various learning tasks involving network-based data. The Fast Incremental von Neumann Graph Entropy algorithm reduces the computation time of the VNGE, making it feasible for real-time applications and large datasets [247].
- **Network analysis.** Von Neumann entropy is used in [248] to build visualization histograms from the edges of networks and then component analysis is performed on a sample for different networks.
- **Pattern recognition in neurological time series.** Von Neumann entropy has been used (together with other entropies) in [249] for automated pattern recognition in neurological conditions, a crucial task in patient monitoring and medical diagnosis.

2.33. Wavelet Entropy

Wavelet entropy was introduced by Rosso et al. in 2001 [250]. It combines the wavelet transform with the concept of entropy to analyse the complexity and information content of multi-frequency signals. More precisely, the wavelet transform decomposes a signal into components at various scales, capturing both time and frequency information, while wavelet entropy quantifies the degree of disorder or unpredictability in those components, thus providing insights into the signal structure and complexity. Wavelet entropy is defined by the Shannon formula, but here the probability distribution corresponds to different resolution levels.

Applications

- **Emotion recognition.** Wavelet entropy can detect little variations in signals and it has been used in [25] to develop an automatic EEG classifier.
- **Fault detection.** Wavelet entropy is applied to monitor the condition of machinery and detect faults by analysing vibration signals [251].
- **Feature extraction.** Wavelet entropy is used to extract features from biomedical signals such as EEG and ECG to identify different physiological states or detect abnormalities [250].

3. Discussion

In Section 1 of this review, we provided a brief historical account of the concept of entropy. Due to its seemingly unconnected appearances in Thermodynamics, Statistical Mechanics, Information Theory, Dynamical Systems, etc., this was necessary to clarify its role in data analysis and machine learning, the subject of the present review. To this end, we gave an axiomatic characterization of entropy and generalized entropies, and acknowledged the inclusion in our review of several entropy-based probability functionals that also go by the name of entropy in the literature and are useful for the mentioned applications. In the case of univariate arguments, entropy quantifies the uncertainty, complexity, and information content of the data. In the case of bivariate and multivariate arguments, entropy quantifies similarity, distance, and information flow.

In Section 2 we collected a representative sample of 33 entropies (including the classical entropies visited in Section 1.2) to show the versatility and potential of the general concept of entropy in practical issues. Indeed, applications such as biomedical signal analysis, fault diagnosis, feature extraction, anomaly detection, optimization cost, and more highlight the diversity of the applications of entropy in data analysis and machine learning.

Section 2 also showed that, more than 150 years after its formulation, entropy and its applications remain the subject of intense research. In fact, new concepts of entropy, evolutions and generalizations are constantly being proposed in the literature to address new challenges. As a result, entropy is being applied to current topics of applied mathematics, in particular, data analysis and machine learning.

4. Conclusions

We conclude this review with some final remarks.

- The choice of a particular entropy-based method depends in general on the application. Some methods may be more popular than others because they are designed for the purpose or dataset at hand, or simply because they have some computational advantage. In this regard, Section 2 presented possible candidates for different applications but no performance comparison between them was discussed. In fact, such a comparison would require a case-by-case test as, for example, in Reference [103], where the authors study motor fault detection with the approximate, dispersion, energy, fuzzy, permutation, sample and Shannon entropies (see Section 2.13 for the best performer). Along with the selection of the "right" entropy, a common concern among practitioners is the choice of parameters and hyperparameters. A combination of methods and parameter settings may be also a good approach in practice [161] [249].
- We have not included applications of entropy to cryptography in this review because they belong to the general field of Data Science (through data security) rather than to Data Analysis. Entropy (mainly Shannon's and Rényi's entropies) have been applied to measure the "randomness" of encrypted messages in the so-called chaotic cryptography, which applies ideas from Chaos Theory and chaos synchronization to the masking of analog signals [252]. To deal with digital signals, new tools such as discrete entropy [253] and discrete Lyapunov exponents [254] have also been developed for application to chaotic cryptography, inspired by their conventional counterparts.
- We have not delved into the numerical methods to compute entropies from the data. Being functionals of probability distributions, most methods for computing entropy are based on the estimation of data probabilities. In the case of discrete-valued data, the probabilities are usually estimated via relative frequencies (the maximum likelihood estimator) and possibly extrapolation methods in case of undersampling [146]. In the case of continuous-valued data, the probability densities are usually estimated via kernel density estimation (also called Parzen-Rosenblatt windowing) [255]. Furthermore, there are methods that do not rely on probability estimation like, e.g., Lempel-Ziv complexity, which resorts to pattern matching [85]. Also, in some particular cases, the entropy can be estimated via spectral information. For example, the quadratic Rényi entropy R_2 can be directly estimated via inner product matrices and principal component analysis [256]. See [257] for a general review on the estimation of entropy.

References

1. von Neumann, J. *Mathematische Grundlagen der Quantenmechanik*; Springer Verlag: Berlin, Heidelberg, New York, 1971; p. 150. doi:https://doi.org/10.1007/978-3-642-96048-2.

2. Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal* **1948**, *27*, 379–423. doi:https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

3. Gibbs, J.W. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*; Charles Scribner's Sons: Farmington Hills, MI, USA, 1902.

4. Kolmogorov, A.N. Entropy per unit time as a metric invariant of automorphisms. *Doklady of the Russian Academy of Sciences* **1959**, *124*, 754–755.

5. Sinai, Y.G. On the Notion of Entropy of a Dynamical System. *Doklady of the Russian Academy of Sciences* **1959**, *124*, 768–771.

6. Adler, R.L.; Konheim, A.G.; McAndrew, M.H. Topological entropy. *Transactions of the American Mathematical Society* **1965**, *114*, 309–319.

7. Walters, P. *An Introduction to Ergodic Theory*; Springer Verlag: New York, 2000.

8. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Dover: New York, 1957.

9. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2005; pp. 243–244. doi:https://doi.org/10.1002/047174882X.ch2.

10. Planck, M. Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum. *Verhandlungen der Deutschen Physikalischen Gesellschaft* **1900**, *2*, 237–245.

11. Csiszár, I. Axiomatic Characterization of Information Measures. *Entropy* **2008**, *10*, 261–273.

12. Rényi, A. On measures of entropy and information. Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, 1961, Vol. 1, pp. 547–561.

13. Amigó, J.M.; Balogh, S.G.; Hernández, S. A Brief Review of Generalized Entropies. *Entropy* **2018**, *20*. doi:https://doi.org/10.3390/e20110813.

14. Havrda, J.; Charvát, F. Quantification method of classification processes. *Kybernetika* **1967**, *3*, 30–35.

15. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **1988**, *52*, 479–487. doi:http://dx.doi.org/10.1007/BF01016429.

16. Ribeiro, M.; Henriques, T.; Castro, L.; Souto, A.; Antunes, L.; Costa-Santos, C.; Teixeira, A. The Entropy Universe. *Entropy* **2021**, *23*. doi:https://doi.org/10.3390/e23020222.

17. Katok, A. Fifty years of entropy in dynamics: 1958–2007. *Journal of Modern Dynamics* **2007**, *1*, 545–596. doi:http://dx.doi.org/10.3934/jmd.2007.1.545.

18. Amigó, J.M.; Keller, K.; Unakafova, V. On entropy, entropy-like quantities, and applications. Frontiers in Entropy Across the Disciplines; Freeden, W.; Nashed, M., Eds. World Scientific, 2022, pp. 197–231. doi:https://doi.org/10.1142/12920.

19. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Physical Review E* **2005**, *71*, 021906. doi:https://doi.org/10.1103/physreve.71.021906.

20. Pincus, J.M. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* **1991**, *88*, 2297–2301. doi:https://doi.org/10.1073/pnas.88.6.2297.

21. Delgado-Bonal, A.; Marshak, A. Approximate Entropy and Sample Entropy: A Comprehensive Tutorial. *Entropy* **2019**, *21*, 541. doi:https://doi.org/10.3390/e21060541.

22. Hornero, R.; Abásolo, D.; Escudero, J.; Gómez, C. Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with Alzheimer's disease. *Philosophical Transactions of the Royal Society of London, Series A* **2009**, *367*, 317–336.

23.   Morabito, F.C.; Labate, D.; La Foresta, F.; Bramanti, A.; Morabito, G.; Palamara, I. Multivariate Multi-Scale Permutation Entropy for Complexity Analysis of Alzheimer's Disease EEG. *Entropy* **2012**, *14*, 1186–1202.

24.   Liang, Z.; Wang, Y.; Sun, X.; Li, D.; Voss, L.J.; Sleigh, J.W.; Hagihira, S.; Li, X. EEG Entropy Measures in Anesthesia. *Frontiers in Computational Neuroscience* **2015**, *9*, 16.

25.   Patel, P.R.; Annavarapu, R.N. EEG-based human emotion recognition using entropy as a feature extraction measure. *Brain Informatics* **2021**, *8*, 20. doi:https://doi.org/10.1186/s40708-021-00141-5.

26.   Kannathal, N.; Choo, M.L.; Acharya, U.R.; Sadasivan, P.K. Entropies for Detection of Epilepsy in EEG. *Computer Methods and Programs in Biomedicine* **2005**, *80*, 187–194.

27.   Srinivasan, V.; Eswaran, C.; Sriraam, N. Approximate Entropy-Based Epileptic EEG Detection Using Artificial Neural Networks. *IEEE Transactions on Information Technology in Biomedicine* **2007**, *11*, 288–295.

28.   Jouny, C.C.; Bergey, G.K. Characterization of Early Partial Seizure Onset: Frequency, Complexity and Entropy. *Clinical Neurophysiology* **2012**, *123*, 658–669.

29.   Richman, J.; Moorman, J. Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy. *American Journal of Physiology - Heart and Circulatory Physiology* **2000**, *278*, H2039–H2049. doi:https://doi.org/10.1152/ajpheart.2000.278.6.H2039.

30.   Acharya, U.R.; Faust, O.; Kannathal, N.; Chua, T.; Laxminarayan, S. Non-linear Analysis of EEG Signals at Various Sleep Stages. *Computer Methods and Programs in Biomedicine* **2005**, *80*, 37–45.

31.   Manis, G.; Aktaruzzaman, M.; Sassi, R. Bubble Entropy: An Entropy Almost Free of Parameters. *IEEE Transactions on Biomedical Engineering* **2017**, *64*, 2711–2718. doi:https://doi.org/10.1109/TBME.2017.2664105.

32.   Manis, G.; Bodini, M.; Rivolta, M.W.; Sassi, R. A Two-Steps-Ahead Estimator for Bubble Entropy. *Entropy* **2021**, *23*. doi:https://doi.org/10.3390/e23060761.

33.   Gong, J.; Yang, X.; Wang, H.; Shen, J.; Liu, W.; Zhou, F. Coordinated method fusing improved bubble entropy and artificial Gorilla Troops Optimizer optimized KELM for rolling bearing fault diagnosis. *Applied Acoustics* **2022**, *195*, 108844. doi:https://doi.org/10.1016/j.apacoust.2022.108844.

34.   Gong, J.; Yang, X.; Qian, K.; Chen, Z.; Han, T. Application of improved bubble entropy and machine learning in the adaptive diagnosis of rotating machinery faults. *Alexandria Engineering Journal* **2023**, *80*, 22–40. doi:https://doi.org/10.1016/j.aej.2023.08.006.

35.   Jiang, X.; Yi, Y.; Wu, J. Analysis of the synergistic complementarity between bubble entropy and dispersion entropy in the application of feature extraction. *Frontiers in Physics* **2023**, *11*. doi:https://doi.org/10.3389/fphy.2023.1163767.

36.   Li, P.; He, X.; Song, D.; Ding, Z.; Qiao, M.; Cheng, X.; Li, R. Improved Categorical Cross-Entropy Loss for Training Deep Neural Networks with Noisy Labels. Pattern Recognition and Computer Vision; Ma, H.; Wang, L.; Zhang, C.; Wu, F.; Tan, T.; Wang, Y.; Lai, J.; Zhao, Y., Eds. Springer International Publishing, 2021, pp. 78–89. doi:https://doi.org/10.1007/978-3-030-88013-2_7.

37.   Spindelböck, T.; Ranftl, S.; von der Linden, W. Cross-Entropy Learning for Aortic Pathology Classification of Artificial Multi-Sensor Impedance Cardiography Signals. *Entropy* **2021**, *23*, 1661. doi:https://doi.org/10.3390/e23121661.

38.   Farebrother, J.; Orbay, J.; Vuong, Q.; Taïga, A.A.; Chebotar, Y.; Xiao, T.; Irpan, A.; Levine, S.; Castro, P.S.; Faust, A.; Kumar, A.; Agarwal, R. Stop Regressing: Training Value Functions via Classification for Scalable Deep RL. *arXiv:2403.03950* **2024**, [2403.03950]. doi:https://doi.org/10.48550/arXiv.2403.03950.

39.   Arazo, E.; Ortego, D.; Albert, P.; O'Connor, N.E.; McGuinness, K. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. *arXiv:1908.02983* **2020**. doi:https://doi.org/10.48550/arXiv.1908.02983.

40.   Mao, A.; Mohri, M.; Zhong, Y. Cross-entropy loss functions: theoretical analysis and applications. *arXiv:2304.07288*. doi:https://doi.org/10.48550/arXiv.2304.07288.

41.   Berrada, L.; Zisserman, A.; Kumar, M.P. Smooth Loss Functions for Deep Top-k Classification. *arXiv:1802.07595* **2018**, [1802.07595]. doi:https://doi.org/10.48550/arXiv.1802.07595.

42.   Wang, Z.; Zhu, Q. *A Cross-Entropy Based Feature Selection Method for Binary Valued Data Classification*; Springer International Publishing, 2022; pp. 1406–1416. doi:https://doi.org/10.1007/978-3-030-96308-8_130.

43.   Kim, S.C.; Kang, T.J. Texture classification and segmentation using wavelet packet frame and Gaussian mixture model. *Pattern Recognition* **2007**, *40*, 1207–1221. doi:https://doi.org/10.1016/j.patcog.2006.09.012.

44.   Bruch, S. An Alternative Cross Entropy Loss for Learning-to-Rank. *arXiv:1911.09798* **2021**. doi:https://doi.org/10.48550/arXiv.1911.09798.

45.   Santosa, B. Multiclass Classification with Cross Entropy-Support Vector Machines. *Procedia Computer Science* **2015**, *72*, 345–352. doi:https://doi.org/10.1016/j.procs.2015.12.149.

46.   Śmieja, M.; Geiger, B.C. Semi-supervised cross-entropy clustering with information bottleneck constraint. *Information Sciences* **2017**, *421*, 254–271. doi:https://doi.org/10.48550/arXiv.1705.01601.

47.   Orchard, M.E.; Olivares, B.; Cerda, M.; Silva, J.F. Anomaly Detection based on Information-Theoretic Measures and Particle Filtering Algorithms. *Annual Conference of the Prognostics and Health Management (PHM) Society* **2012**, *4*. doi:https://doi.org/10.36001/phmconf.2012.v4i1.2113.

48.   Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer Science + Business Media, 2006.

49.   Qu, Y.; Li, R.; Deng, A.; Shang, C.; Shen, Q. Non-unique Decision Differential Entropy-Based Feature Selection. *Neurocomputing* **2020**, *393*, 187–193. doi:https://doi.org/10.1016/j.neucom.2018.10.112.

50.   Grassucci, E.; Comminiello, D.; Uncini, A. An Information-Theoretic Perspective on Proper Quaternion Variational Autoencoders. *Entropy* **2021**, *23*, 856. doi:https://doi.org/10.3390/e23070856.

51.   Gibson, J. Entropy Power, Autoregressive Models, and Mutual Information. *Entropy* **2018**, *20*, 750. doi:https://doi.org/10.3390/e20100750.

52.   Robin, S.; Scrucca, L. Mixture-based estimation of entropy. *Computational Statistics & Data Analysis* **2023**, *177*, 107582. doi:https://doi.org/10.1016/j.csda.2022.107582.

53.   Rostaghi, M.; Azami, H. Dispersion Entropy: A Measure for Time-Series Analysis. *IEEE Signal Processing Letters* **2016**, *23*, 610–614. doi:https://doi.org/10.1109/LSP.2016.2542881.

54.   Rostaghi, M.; Khatibi, M.M.; Ashory, M.R.; Azami, H. Refined Composite Multiscale Fuzzy Dispersion Entropy and Its Applications to Bearing Fault Diagnosis. *Entropy* **2023**, *25*, 1494. doi:https://doi.org/10.3390/e25111494.

55.   Furlong, R.; Hilal, M.; O'Brien, V.; Humeau-Heurtier, A. Parameter Analysis of Multiscale Two-Dimensional Fuzzy and Dispersion Entropy Measures Using Machine Learning Classification. *Entropy* **2021**, *23*, 1303. doi:https://doi.org/10.3390/e23101303.

56.   Hu, B.; Wang, Y.; Mu, J. A new fractional fuzzy dispersion entropy and its application in muscle fatigue detection. *Mathematical Biosciences and Engineering* **2024**, *21*, 144–169. doi:https://doi.org/10.3934/mbe.2024007.

57.   Dhandapani, R.; Mitiche, I.; McMeekin, S.; Mallela, V.S.; Morison, G. Enhanced Partial Discharge Signal Denoising Using Dispersion Entropy Optimized Variational Mode Decomposition. *Entropy* **2021**, *23*, 1567. doi:https://doi.org/10.3390/e23121567.

58.   Li, G.; Yang, Z.; Yang, H. A Denoising Method of Ship Radiated Noise Signal Based on Modified CEEM-DAN, Dispersion Entropy, and Interval Thresholding. *Electronics* **2019**, *8*, 597. doi:https://doi.org/10.3390/electronics8060597.

59.   Fabila-Carrasco, J.S.; Tan, C.; Escudero, J. Graph-Based Multivariate Multiscale Dispersion Entropy: Efficient Implementation and Applications to Real-World Network Data. *arXiv:2405.00518* **2024**, [2405.00518]. doi:https://doi.org/10.48550/arXiv.2405.00518.

60.   Ge, H.; Chen, G.; Yu, H.; Chen, H.; An, F. Theoretical Analysis of Empirical Mode Decomposition. *Symmetry* **2018**, *10*, 623. doi:https://doi.org/10.3390/sym10110623.

61.   Liu, C.; Zhu, L.; Ni, C. Chatter detection in milling process based on VMD and energy entropy. *Mechanical Systems and Signal Processing* **2018**, *105*, 169–182. doi:https://doi.org/10.1016/j.ymssp.2017.11.046.

62.   Gao, Z.; Liu, Y.; Wang, Q.; Wang, J.; Luo, Y. Ensemble empirical mode decomposition energy moment entropy and enhanced long short-term memory for early fault prediction of bearing. *Measurement* **2022**, *188*, 110417. doi:https://doi.org/10.1016/j.measurement.2021.110417.

63.   Yu, Y.; Dejie, Y.; Junsheng, C. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *Journal of Sound and Vibration* **2006**, *294*, 269–277. doi:https://doi.org/10.1016/j.jsv.2005.11.002.

64.   Yang, Z.; Luo, S.; Zhong, P.; Chen, R.; Pan, C.; Li, K. An EMD and IMF Energy Entropy-Based Optimized Feature Extraction and Classification Scheme for Single Trial EEG Signal. *Journal of Mechanics in Medicine and Biology* **2023**, *23*, 2340063. doi:https://doi.org/10.1142/S0219519423400638.

65.   Zhu, G.; Peng, S.; Lao, Y.; Su, Q.; Sun, Q. Short-Term Electricity Consumption Forecasting Based on the EMD-Fbprophet-LSTM Method. *Mathematical Problems in Engineering* **2021**, *2021*, 6613604. doi:https://doi.org/10.1155/2021/6613604.

66.   Gao, J.; Shang, P. Analysis of complex time series based on EMD energy entropy plane. *Nonlinear Dynamics* **2019**, *96*, 465–482. doi:https://doi.org/10.1007/s11071-019-04800-5.

67. Headrick, M. Lectures on entanglement entropy in field theory and holography. *arXiv:1907.08126* **2019**, [1907.08126]. doi:https://doi.org/10.48550/arXiv.1907.08126.

68. Rieger, M.; Reh, M.; Gärtner, M. Sample-efficient estimation of entanglement entropy through supervised learning. *Physical Review A* **2024**, *109*, 012403. doi:https://doi.org/10.1103/PhysRevA.109.012403.

69. Liu, Y.; Li, W.J.; Zhang, X.; Lewenstein, M.; Su, G.; Ran, S.J. Entanglement-Based Feature Extraction by Tensor Network Machine Learning. *Frontiers in Applied Mathematics and Statistics* **2021**, *7*, 716044.

70. Lin, X.; Chen, Z.; Wei, Z. Quantifying Unknown Quantum Entanglement via a Hybrid Quantum-Classical Machine Learning Framework. *Physical Review A* **2023**, *107*. doi:https://doi.org/10.48550/arXiv.2204.11500.

71. Abdallah, S.A.; Plumbley, M.D. A measure of statistical complexity based on predictive information with application to finite spin systems. *Physics Letters A* **2012**, *376*, 275–281. doi:https://doi.org/10.1016/j.physleta.2011.10.066.

72. Crutchfield, J.P.; Packard, N.H. Symbolic dynamics of noisy chaos. *Physica D: Nonlinear Phenomena* **1983**, *7*, 201–223. doi:https://doi.org/10.1016/0167-2789(83)90127-6.

73. Bardera, A.; Boada, I.; Feixas, M., S.M. Image Segmentation Using Excess Entropy. *Journal of Signal Processing Systems* **2009**, *54*, 205–214. doi:https://doi.org/10.1007/s11265-008-0194-6.

74. Nir, A.; Sela, E.; Beck, R.; Bar-Sinai, Y. Machine-Learning Iterative Calculation of Entropy for Physical Systems. *Proceedings of the National Academy of Sciences* **2020**, *117*, 30234–30240. doi:https://doi.org/10.1073/pnas.2017042117.

75. Belghazi, M.I.; Baratin, A.; Rajeswar, S.; Ozair, S.; Bengio, Y.; Courville, A.; Hjelm, R.D. MINE: Mutual Information Neural Estimation. *arXiv:1801.04062* **2021**.

76. Xiang, X.; Zhou, J. An Excess Entropy Approach to Classify Long-Term and Short-Term Memory Stationary Time Series. *Mathematics* **2023**, *11*, 2448. doi:https://doi.org/10.3390/math11112448.

77. Azami, H.; Escudero, J. Amplitude- and Fluctuation-Based Dispersion Entropy. *Entropy* **2018**, *20*. doi:https://doi.org/10.3390/e20030210.

78. Chen, Y.; Chen, J.; Qiang, Y.; Yuan, Z.; Yang, J. Refined composite moving average fluctuation dispersion entropy and its application on rolling bearing fault diagnosis. *Review of Scientific Instruments* **2023**, *94*, 105110. doi:https://doi.org/10.1063/5.0165430.

79. Su, H.; Wang, Z.; Cai, Y.; Ding, J.; Wang, X.; Yao, L. Refined Composite Multiscale Fluctuation Dispersion Entropy and Supervised Manifold Mapping for Planetary Gearbox Fault Diagnosis. *Machines* **2023**, *11*, 47. doi:https://doi.org/10.3390/machines11010047.

80. Zhou, F.; Han, J.; Yang, X. Multivariate hierarchical multiscale fluctuation dispersion entropy: Applications to fault diagnosis of rotating machinery. *Applied Acoustics* **2021**, *182*, 108271. doi:https://doi.org/10.1016/j.apacoust.2021.108271.

81. Li, Z.; Lan, T.; Li, Z.; Gao, P. Exploring Relationships between Boltzmann Entropy of Images and Building Classification Accuracy in Land Cover Mapping. *Entropy* **2023**, *25*, 1182. doi:https://doi.org/10.3390/e25081182.

82. Baldini, G.; Chareau, J.M.; Bonavitacola, F. Spectrum Sensing Implemented with Improved Fluctuation-Based Dispersion Entropy and Machine Learning. *Entropy* **2021**, *23*, 1611. doi:https://doi.org/10.3390/e23121611.

83. Azami, H.; Arnold, S.E.; Sanei, S.; Chang, Z.; Sapiro, G.; Escudero, J.; Gupta, A.S. Multiscale Fluctuation-based Dispersion Entropy and its Applications to Neurological Diseases. *ArXiv:1902.10825* **2019**, [1902.10825]. doi:https://doi.org/10.48550/arXiv.1902.10825.

84. Jiao, S.; Geng, B.; Li, Y.; Zhang, Q.; Wang, Q. Fluctuation-based reverse dispersion entropy and its applications to signal classification. *Applied Acoustics* **2021**, *175*, 107857. doi:https://doi.org/10.1016/j.apacoust.2020.107857.

85. Amigó, J.M.; Szczepanski, J.; Wajnryb, E.; Sanchez-Vives, M.V. Estimating the entropy of spike trains via Lempel-Ziv complexity. *Neural Computation* **2004**, *16*, 717–736. doi:https://doi.org/10.1162/089976604322860677.

86. Friedgut, E.; Kalai, G. Every Monotone Graph Property Has A Sharp Threshold. *Proceedings of the American Mathematical Society* **1999**, *124*, 2993–3002. doi:http://dx.doi.org/10.1090/S0002-9939-96-03732-X.

87. Chakraborty, S.; Kulkarni, R.; Lokam, S.; Saurabh, N. Upper bounds on Fourier entropy. *Theoretical Computer Science* **2016**, *654*, 92–112. doi:https://doi.org/10.1016/j.tcs.2016.05.006.

88.   O'Donnell, R.; Wright, J.; Zhou, Y. The Fourier Entropy–Influence Conjecture for Certain Classes of Boolean Functions. Automata, Languages and Programming; Aceto, L.; Henzinger, M.; Sgall, J., Eds. Springer Berlin Heidelberg, 2011, pp. 330–341. doi:https://doi.org/10.1007/978-3-642-22006-7_28.

89.   Kelman, E.; Kindler, G.; Lifshitz, N.; Minzer, D.; Safra, M. Towards a Proof of the Fourier–Entropy Conjecture? 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), 2020, pp. 247–258. doi:https://doi.org/10.1109/FOCS46700.2020.00032.

90.   Almeida, L.B. The fractional Fourier transform and time-frequency representations. *IEEE Transactions on Signal Processing* **1994**, *42*, 3084–3091. doi:https://doi.org/10.1109/78.330368.

91.   Tao, R.; Zhao, X.; Li, W.; Li, H.C.; Du, Q. Hyperspectral Anomaly Detection by Fractional Fourier Entropy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2019**, *12*, 4920–4929. doi:https://doi.org/10.1109/JSTARS.2019.2940278.

92.   Zhang, L.; Ma, J.; Cheng, B.; Lin, F. Fractional Fourier Transform-Based Tensor RX for Hyperspectral Anomaly Detection. *Remote Sensing* **2022**, *14*, 797. doi:https://doi.org/10.3390/rs14030797.

93.   Wang, S.H.; Zhang, X.; Zhang, Y.D. DSSAE: Deep Stacked Sparse Autoencoder Analytical Model for COVID-19 Diagnosis by Fractional Fourier Entropy. *ACM Transactions on Management Information System (TMIS)* **2021**, *13*. doi:https://doi.org/10.1145/3451357.

94.   Wang, S.; Zhang, Y.; Yang, X.; Sun, P.; Dong, Z.; Liu, A.; Yuan, T.F. Pathological Brain Detection by a Novel Image Feature—Fractional Fourier Entropy. *Entropy* **2015**, *17*, 8278–8296. doi:https://doi.org/10.3390/e17127877.

95.   Yan, Y. Gingivitis detection by fractional Fourier entropy with optimization of hidden neurons. *International Journal of Cognitive Computing in Engineering* **2020**, *1*, 36–44. doi:https://doi.org/10.1016/j.ijcce.2020.09.003.

96.   Panahi, F.; Rashidi, S.; Sheikhani, A. Application of fractional Fourier transform in feature extraction from Electrocardiogram and Galvanic Skin RESPONSE for emotion recognition. *Biomedical Signal Processing and Control* **2021**, *69*, 102863. doi:https://doi.org/10.1016/j.bspc.2021.102863.

97.   Zhang, Y.; Yang, X.; Cattani, C.; Rao, R.V.; Wang, S.; Phillips, P. Tea Category Identification Using a Novel Fractional Fourier Entropy and Jaya Algorithm. *Entropy* **2016**, *18*, 77. doi:https://doi.org/10.3390/e18030077.

98.   Ishikawa, A.; Mieno, H. The fuzzy entropy concept and its application. *Fuzzy Sets and Systems* **1979**, *2*, 113–123. doi:https://doi.org/10.1016/0165-0114(79)90020-4.

99.   Zheng, J.; Cheng, J.; Yang, Y.; Luo, S. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mechanism and Machine Theory* **2014**, *78*, 187–200. doi:https://doi.org/10.1016/j.mechmachtheory.2014.03.014.

100.   Markechová, D.; Riečan, B. Entropy of Fuzzy Partitions and Entropy of Fuzzy Dynamical Systems. *Entropy* **2016**, *18*, 19. doi:https://doi.org/10.3390/e18010019.

101.   D'Urso, P.; De Giovanni, L.; Vitale, V. Robust DTW-based entropy fuzzy clustering of time series. *Annals of Operations Research* **2023**. doi:https://doi.org/10.1007/s10479-023-05720-9.

102.   Di Martino, F.; Sessa, S. Energy and Entropy Measures of Fuzzy Relations for Data Analysis. *Entropy* **2018**, *20*, 424. doi:https://doi.org/10.3390/e20060424.

103.   Aguayo-Tapia, S.; Avalos-Almazan, G.; Rangel-Magdaleno, J.J. Entropy-Based Methods for Motor Fault Detection: A Review. *Entropy* **2024**, *26*, 1–21. doi:https://doi.org/10.3390/e26040299.

104.   Jinde, Z.; Chen, M.J.; Junsheng, C.; Yang, Y. Multiscale fuzzy entropy and its application in rolling bearing fault diagnosis. *Zhendong Gongcheng Xuebao/Journal of Vibration Engineering* **2014**, *27*, 145–151.

105.   Kumar, R.; Bisht, D.C.S. Picture fuzzy entropy: A novel measure for managing uncertainty in multi-criteria decision-making. *Decision Analytics Journal* **2023**, *9*, 100351. doi:https://doi.org/10.1016/j.dajour.2023.100351.

106.   Lhermitte, E.; Hilal, M.; Furlong, R.; O'Brien, V.; Humeau-Heurtier, A. Deep Learning and Entropy-Based Texture Features for Color Image Classification. *Entropy* **2022**, *24*, 1577. doi:https://doi.org/10.3390/e24111577.

107.   Tan, Z.; Li, K.; Wang, Y. An improved cuckoo search algorithm for multilevel color image thresholding based on modified fuzzy entropy. *Journal of Ambient Intelligence and Humanized Computing* **2021**. doi:https://doi.org/10.1007/s12652-021-03001-6.

108.   Korner, J. Coding of an information source having ambiguous alphabet and the entropy of graphs. *Transactions of the 6th Prague conference on Information Theory* **1971**, pp. 411–425.

109.   Simonyi, G. Graph Entropy: A Survey. *Combinatorial Optimization* **1995**, *20*, 399–444.

110. Luque, B.; Lacasa, L.; Ballesteros, F.; Luque, J. Horizontal visibility graphs: exact results for random time series. *Physical Review E* **2009**, *80*, 046103. doi:https://link.aps.org/doi/10.1103/PhysRevE.80.046103.

111. Lacasa, L.; Just, W. Visibility graphs and symbolic dynamics. *Physica D: Nonlinear Phenomena* **2018**, *374-375*, 35–44. doi:https://doi.org/10.1016/j.physd.2018.04.001.

112. Harangi, V.; Niu, X.; Bai, B. Conditional graph entropy as an alternating minimization problem. *arXiv:2209.00283* **2023**, [2209.00283]. doi:https://doi.org/10.48550/arXiv.2209.00283.

113. Wu, J.; Chen, X.; Xu, K.; Li, S. Structural Entropy Guided Graph Hierarchical Pooling. *arXiv:2206.13510* **2022**, [2206.13510]. doi:https://doi.org/10.48550/arXiv.2206.13510.

114. Juhnke-Kubitzke, M.; Köhne, D.; Schmidt, J. Counting Horizontal Visibility Graphs. *arXiv:2111.02723* **2021**, [2111.02723]. doi:https://doi.org/10.48550/arXiv.2111.02723.

115. Luo, G.; Li, J.; Su, J.; Peng, H.; Yang, C.; Sun, L.; Yu, P.S.; He, L. Graph Entropy Guided Node Embedding Dimension Selection for Graph Neural Networks. *arXiv:2105.03178* **2021**, [2105.03178]. doi:https://doi.org/10.48550/arXiv.2105.03178.

116. Zhu, G.; Li, Y.; Wen, P. Analysis of alcoholic EEG signals based on horizontal visibility graph entropy. *Brain Informatics* **2014**, *1*, 19–25. doi:https://doi.org/10.1007/s40708-014-0003-x.

117. Yu, M.; Hillebrand, A.; Gouw, A.; Stam, C.J. Horizontal visibility graph transfer entropy (HVG-TE): A novel metric to characterize directed connectivity in large-scale brain networks. *NeuroImage* **2017**, *156*, 249–264. doi:https://doi.org/10.1016/j.neuroimage.2017.05.047.

118. Chen, T.; Vemuri, B.; Rangarajan, A.; Eisenschenk, S.J. Group-Wise Point-Set Registration Using a Novel CDF-Based Havrda-Charvát Divergence. *International Journal of Computer Vision* **2010**, *86*, 111–124. doi:https://doi.org/10.1007/s11263-009-0261-x.

119. Shi, Y.; Wu, Y.; Shang, P. Research on weighted Havrda–Charvat's entropy in financial time series. *Physica A: Statistical Mechanics and its Applications* **2021**, *572*, 125914. doi:https://doi.org/10.1016/j.physa.2021.125914.

120. Brochet, T.; Lapuyade-Lahorgue, J.; Bougleux, S.; Salaun, M.; Ruan, S. Deep learning using Havrda-Charvat entropy for classification of pulmonary endomicroscopy. *arXiv:2104.05450* **2021**, [2104.05450]. doi:https://doi.org/10.48550/arXiv.2104.05450.

121. Brochet, T.; Lapuyade-Lahorgue, J.; Huat, A.; Thureau, S.; Pasquier, D.; Gardin, I.; Modzelewski, R.; Gibon, D.; Thariat, J.; Grégoire, V.; Vera, P.; Ruan, S. A Quantitative Comparison between Shannon and Tsallis–Havrda–Charvat Entropies Applied to Cancer Outcome Prediction. *Entropy* **2022**, *24*, 436. doi:https://doi.org/10.3390/e24040436.

122. Amoud, H.; Snoussi, H.; Hewson, D.; Doussot, M.; Duchene, J. Intrinsic Mode Entropy for Nonlinear Discriminant Analysis. *IEEE Signal Processing Letters* **2007**, *14*, 297–300. doi:https://doi.org/10.1109/LSP.2006.888089.

123. E., K.V.; Hadjileontiadis, L.J. Intrinsic mode entropy: An enhanced classification means for automated Greek Sign Language gesture recognition. 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, pp. 5057–5060. doi:https://doi.org/10.1109/IEMBS.2008.4650350.

124. Hu, M.; Liang, H. Intrinsic mode entropy based on multivariate empirical mode decomposition and its application to neural data analysis. *Cognitive Neurodynamics* **2011**, *5*, 277–284. doi:https://doi.org/10.1007/s11571-011-9159-8.

125. Hu, M.; Liang, H., Multiscale Entropy: Recent Advances. In *Complexity and Nonlinearity in Cardiovascular Signals*; Barbieri, R.; Scilingo, E.; Valenza, G., Eds.; Springer International Publishing, 2017; pp. 115–138. doi:https://doi.org/10.1007/978-3-319-58709-7_4.

126. Amoud, H.; Snoussi, H.; Hewson, D.; Duchêne, J. Intrinsic Mode Entropy for postural steadiness analysis. 4th European Conference of the International Federation for Medical and Biological Engineering. Springer, Berlin, Heidelberg, 2009, Vol. 22, *IFMBE Proceedings*, pp. 212–215. doi:https://doi.org/10.1007/978-3-540-89208-3_53.

127. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Physical Review E* **2002**, *66*, 056125. doi:https://doi.org/10.1103/PhysRevE.66.056125.

128. Kaniadakis, G. Relativistic Roots of κ-Entropy. *Entropy* **2024**, *26*, 406. doi:https://doi.org/10.3390/e26050406.

129. Lei, B.; Fan, J.I. Adaptive Kaniadakis entropy thresholding segmentation algorithm based on particle swarm optimization. *Soft Computing* **2020**, *24*, 7305–7318. doi:https://doi.org/10.1007/s00500-019-04351-2.

130. Jena, B.; Naik, M.K.; Panda, R. A novel Kaniadakis entropy-based multilevel thresholding using energy curve and Black Widow optimization algorithm with Gaussian mutation. 2023 International Conference in

Advances in Power, Signal, and Information Technology (APSIT), 2023, pp. 86–91. doi:https://doi.org/10.1109/APSIT58554.2023.10201718.

131. da Silva, S.L.E.F.; de Araújo, J.M.; de la Barra, E.; Corso, G. A Graph-Space Optimal Transport Approach Based on Kaniadakis κ-Gaussian Distribution for Inverse Problems Related to Wave Propagation. *Entropy* **2023**, *25*, 990. doi:https://doi.org/10.3390/e25070990.

132. Mekyska, J.; Janousova, E.; Gomez-Vilda, P.; Smekal, Z.; Rektorova, I.; Eliasova, I.; Kostalova, M.; Mrackova, M.; Alonso-Hernandez, J.B.; Faundez-Zanuy, M.; López-de Ipiña, K. Robust and complex approach of pathological speech signal analysis. *Neurocomputing* **2015**, *167*, 94–111. doi:http://dx.doi.org/10.1016/j.neucom.2015.02.085.

133. Xu, L.S.; Wang, K.Q.; Wang, L. Gaussian kernel approximate entropy algorithm for analyzing irregularity of time-series. 2005 International Conference on Machine Learning and Cybernetics, 2005, Vol. 9, pp. 5605–5608. doi:https://doi.org/10.1109/ICMLC.2005.1527935.

134. Zaylaa, A.; Saleh, S.; Karameh, F.; Nahas, Z. andBouakaz, A. Cascade of nonlinear entropy and statistics to discriminate fetal heart rates. Proceedings of the 2016 3rd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA). IEEE Xplore, 2016, pp. 152–157. doi:https://doi.org/10.1109/ACTEA.2016.7560130.

135. Orozco-Arroyave, J.; Arias-Londono, J.; Vargas-Bonilla, J.; Nöth, E. Analysis of speech from people with Parkinson's disease through nonlinear dynamics. Advances in Nonlinear Speech Processing. NOLISP 2013. Lecture Notes in Computer Science, vol 7911; Drugman, T.; Dutoit, T., Eds. Springer, 2013, pp. 112–119. doi:https://doi.org/10.1007/978-3-642-38847-7_15.

136. Kolmogorov, A.N. A New Metric Invariant of Transitive Dynamical Systems and Automorphisms of Lebesgue Spaces. *Proceedings of the Steklov Institute of Mathematics* **1986**, *169*, 97–102.

137. Pesin, Y. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys* **1977**, *32*, 55. doi:https://doi.org/10.1070/rm1977v032n04abeh001639.

138. Stolz, I.; Keller, K. A General Symbolic Approach to Kolmogorov-Sinai Entropy. *Entropy* **2017**, *19*, 675. doi:https://doi.org/10.3390/e19120675.

139. Shiozawa, K.; Tokuda, I. Estimating Kolmogorov-Sinai entropy from time series of high-dimensional complex systems. *Physics Letters A* **2024**, *510*, 129531. doi:https://doi.org/10.1016/j.physleta.2024.129531.

140. Karmakar, C.; Udhayakumar, R.; Palaniswami, M. Entropy Profiling: A Reduced-Parametric Measure of Kolmogorov-Sinai Entropy from Short-Term HRV Signal. *Entropy* **2020**, *22*, 1396. doi:https://doi.org/10.3390/e22121396.

141. Kiss, G.; Bakucz, P. Using Kolmogorov Entropy to Verify the Description Completeness of Traffic Dynamics of Highly Autonomous Driving. *Applied Sciences* **2024**, *14*, 2261. doi:https://doi.org/10.3390/app14062261.

142. Aftanas, L.I.; Lotova, N.V.; Koshkarov, V.I.; Pokrovskaja, V.L.; Popov, S.A.; Makhnev, V.P. Non-linear analysis of emotion EEG: calculation of Kolmogorov entropy and the principal Lyapunov exponent. *Neuroscience Letters* **1997**, *226*, 13–16. doi:https://doi.org/10.1016/S0304-3940(97)00232-2.

143. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical Review Letters* **2002**, *88*, 174102. doi:https://link.aps.org/doi/10.1103/PhysRevLett.88.174102.

144. Bandt, C.; Keller, G.; Pompe, B. Entropy of interval maps via permutations. *Nonlinearity* **2002**, *15*, 1595. doi:http://dx.doi.org/10.1088/0951-7715/15/5/312.

145. Riedl, M.; Müller, A.; Wessel, N. Practical considerations of permutation entropy. *The European Physical Journal Special Topics* **2013**, *222*, 249–262. doi:https://doi.org/10.1140/epjst/e2013-01862-7.

146. Amigó, J.M. *Permutation Complexity in Dynamical Systems*; Springer Verlag, 2010. doi:https://doi.org/10.1007/978-3-642-04084-9.

147. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy* **2012**, *14*, 1553–1577. doi:https://doi.org/10.3390/e14081553.

148. Amigó, J.M.; Keller, K.; Kurths, J. Recent Progress in Symbolic Dynamics and Permutation Complexity —Ten Years of Permutation Entropy. *The European Physical Journal Special Topics* **2013**, *222*, 241–247. doi:http://dx.doi.org/10.1140/epjst/e2013-01839-6.

149. Amigó, J.M.; Rosso, O.A. Ordinal methods: Concepts, applications, new developments, and challenges—In memory of Karsten Keller (1961–2022). *Chaos* **2023**, *33*, 080401. doi:https://doi.org/10.1063/5.0167263.

150. Mammone, N.; Duun-Henriksen, J.; Kjaer, T.W.; Morabito, F.C. Differentiating Interictal and Ictal States in Childhood Absence Epilepsy through Permutation Rényi Entropy. *Entropy* **2015**, *17*, 4627–4643. doi:https://doi.org/10.3390/e17074627.

151. Zunino, L.; Pérez, D.G.; Kowalski, A.; Martin, M.T.; Garavaglia, M.; Plastino, A.; Rosso, O.A. Fractional Brownian motion, fractional Gaussian noise, and Tsallis permutation entropy. *Physica A: Statistical Mechanics and its Applications* **2008**, *387*, 6057–6068. doi:https://doi.org/10.1016/j.physa.2008.07.004.

152. Stosic, D.; Stosic, D.; Stosic, T.; Stosic, B. Generalized weighted permutation entropy. *Chaos* **2022**, *32*. doi:http://dx.doi.org/10.1063/5.0107427.

153. Yin, Y.; Sun, K.; He, S. Multiscale permutation Rényi entropy and its application for EEG signals. *PLOS ONE* **2018**, *13*, 1–15. doi:https://doi.org/10.1371/journal.pone.0202558.

154. Li, C.; Shang, P. Multiscale Tsallis permutation entropy analysis for complex physiological time series. *Physica A* **2019**, *523*, 10–20. doi:https://doi.org/10.1016/j.physa.2019.01.031.

155. Azami, H.; Escudero, J. Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings. *Biomedical Signal Processing and Control* **2016**, *23*, 28–41. doi:https://doi.org/10.1016/j.bspc.2015.08.004.

156. Keller, K.; Lauffer, H. Symbolic Analysis of High-Dimensional Time Series. *International Journal of Bifurcation and Chaos* **2003**, *13*, 2657–2668. doi:https://doi.org/10.1142/S0218127403008168.

157. Keller, K.; Mangold, T.; Stolz, L.; Werner, J. Permutation Entropy: New Ideas and Challenges. *Entropy* **2017**, *19*, 134. doi:https://doi.org/10.3390/e19030134.

158. Amigó, J.M.; Kennel, M.B. Forbidden ordinal patterns in higher dimensional dynamics. *Physica D* **2008**, *237*, 2893–2899. doi:doi.org/10.1016/j.physd.2008.05.003.

159. Amigó, J.M.; Keller, K. Permutation entropy: One concept, two approaches. *The European Physical Journal Special Topics* **2013**, *222*, 263–273. doi:https://doi.org/10.1140/epjst/e2013-01840-1.

160. Carpi, L.C.; Saco, P.M.; Rosso, O.A. Missing ordinal patterns in correlated noises. *Physica A* **2010**, *389*, 2020–2029. doi:https://doi.org/10.1016/j.physa.2010.01.030.

161. Pilarczyk, P.; Graff, G.; Amigó, J.M.; Tessmer.; Narkiewicz, K.; Graff, B. Differentiating patients with obstructive sleep apnea from healthy controls based on heart rate-blood pressure coupling quantified by entropy-based indices. *Chaos* **2023**, *33*, 103140. doi:https://doi.org/10.1063/5.0158923.

162. Qu, Z.; Hou, X.; Hu, W.; Yang, R.; Ju, C. Wind power forecasting based on improved variational mode decomposition and permutation entropy. *Clean Energy* **2023**, *7*, 1032–1045. doi:https://doi.org/10.1093/ce/zkad043.

163. Rosso, O.A.; Larrondo, H.A.; Martin, M.T.; Plastino, A.; Fuentes, M.A. Distinguishing Noise from Chaos. *Physical Review Letters* **2007**, *99*, 154102. doi:https://doi.org/10.1103/PhysRevLett.99.154102.

164. Voltarelli, L.G.J.M.; Pessa, A.A.B.; Zunino, L.; Zola, R.S.; Lenzi, E.K.; Perc, M.; Ribeiro, H.V. Characterizing unstructured data with the nearest neighbor permutation entropy. *Chaos* **2024**, *34*, 053130. doi:https://doi.org/10.1063/5.0209206.

165. Rao, C.R. Diversity and Dissimilarity Coefficients: A Unified Approach. *Theoretical Population Biology* **1982**, *21*, 24–43. doi:https://doi.org/10.1016/0040-5809(82)90004-1.

166. Doxa, A.; Prastacos, P. Using Rao's quadratic entropy to define environmental heterogeneity priority areas in the European Mediterranean biome. *Biological Conservation* **2020**, *241*, 108366. doi:https://doi.org/10.1016/j.biocon.2019.108366.

167. Smouse, P.E.; Banks, S.C.; Peakall, R. Converting quadratic entropy to diversity: Both animals and alleles are diverse, but some are more diverse than others. *PLOS ONE* **2017**, *12*, 1–19. doi:https://doi.org/10.1371/journal.pone.0185499.

168. Dionne, G.; Koumou, G. Machine Learning and Risk Management: SVDD Meets RQE. Technical Report 18-6, HEC Montreal, Canada Research Chair in Risk Management, 2018.

169. Citi, L.; G., G.; Mainardi, L. Rank-based Multi-Scale Entropy analysis of heart rate variability. Proceedings of the Computing in Cardiology 2014, 2014, pp. 597–600.

170. Garchery, M.; Granitzer, M. On the influence of categorical features in ranking anomalies using mixed data. *Procedia Computer Science* **2018**, *126*, 77–86. doi:https://doi.org/10.1016/j.procs.2018.07.211.

171. Khan, M.A.; Akram, T.; Sharif, M.; Alhaisoni, M.; Saba, T.; Nawaz, N. A probabilistic segmentation and entropy-rank correlation-based feature selection approach for the recognition of fruit diseases. *EURASIP Journal on Image and Video Processing* **2021**, *2021*, 14. doi:https://doi.org/10.1186/s13640-021-00558-2.

172. Hu, Q.; Che, X.; Zhang, L.; Zhang, D.; Guo, M.; Yu, D. Rank Entropy-Based Decision Trees for Monotonic Classification. *IEEE Transactions on Knowledge and Data Engineering* **2012**, *24*, 2052–2064. doi:https://doi.org/10.1109/TKDE.2011.149.

173. Liu, S.; Gao, H. The Structure Entropy-Based Node Importance Ranking Method for Graph Data. *Entropy* **2023**, *25*, 941. doi:https://doi.org/10.3390/e25060941.

174. McLellan, M.R.; Ryan, M.D.; Breneman, C.M. Rank order entropy: why one metric is not enough. *Journal of Chemical Information and Modeling* **2011**, *51*, 2302–2319. doi:https://doi.org/10.1021/ci200170k.

175. Diks, C.; Panchenko, V. Rank-based Entropy Tests for Serial Independence. *Studies in Nonlinear Dynamics & Econometrics* **2008**, *12*. doi:https://doi.org/10.2202/1558-3708.1476.

176. C., S.; Li, H.; M., S.; S., H. A Ranking-Based Cross-Entropy Loss for Early Classification of Time Series. *IEEE Transactions on Neural Networks and Learning Systems* **2024**, *35*, 11194–11203. doi:https://doi.org/10.1109/tnnls.2023.3250203.

177. Niknami, N.; Wu, J. Entropy-KL-ML: Enhancing the Entropy-KL-Based Anomaly Detection on Software-Defined Networks. *IEEE Transactions on Network Science and Engineering* **2022**, *9*, 4458–4467. doi:https://doi.org/10.1109/TNSE.2022.3202147.

178. Moral, S.; Cano, A.; Gómez-Olmedo, M. Computation of Kullback–Leibler Divergence in Bayesian Networks. *Entropy* **2021**, *23*, 1122. doi:https://doi.org/10.3390/e23091122.

179. Brown, G.; Pocock, A.; Zhao, M.J.; Luján, M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *The Journal of Machine Learning Research* **2012**, *13*, 27–66.

180. Chaimovich, M.; Chaimovich, A. Relative Resolution: An Analysis with the Kullback–Leibler Entropy. *Journal of Chemical Theory and Computation* **2024**, *20*, 2074–2087. doi:https://doi.org/10.1021/acs.jctc.3c01052.

181. Draelos, R. Connections: Log-Likelihood, Cross-Entropy, KL-Divergence, Logistic Regression, and Neural Networks. *GlassBox Medicine* **2019**. Published online: 7 December 2019. Retrieved on 20 April 2024.

182. De La Pava Panche, I.; Alvarez-Meza, A.M.; Orozco-Gutierrez, A. A Data-Driven Measure of Effective Connectivity Based on Renyi's α-Entropy. *Frontiers in Neuroscience* **2019**, *13*, 1277. doi:https://doi.org/10.3389/fnins.2019.01277.

183. Rioul, O. The Interplay between Error, Total Variation, Alpha-Entropy and Guessing: Fano and Pinsker Direct and Reverse Inequalities. *Entropy* **2023**, *25*, 978. doi:https://doi.org/10.3390/e25070978.

184. Berezinski, P.; Jasiul, B.; Szpyrka, M. An Entropy-Based Network Anomaly Detection Method. *Entropy* **2015**, *17*, 2367–2408. doi:https://doi.org/10.3390/e17042367.

185. Sharma, R.; Pachori, R.B.; Acharya, U.R. Application of Entropy Measures on Intrinsic Mode Functions for the Automated Identification of Focal Electroencephalogram Signals. *Entropy* **2015**, *17*, 669–691. doi:https://doi.org/10.3390/e17020669.

186. Czarnecki, W.M.; Tabor, J. Extreme entropy machines: Robust information theoretic classification. *Pattern Analysis and Applications* **2017**, *20*, 383–400. doi:https://doi.org/10.1007/s10044-015-0497-8.

187. Sluga, D.; Lotrič, U. Quadratic Mutual Information Feature Selection. *Entropy* **2017**, *19*, 157. doi:https://doi.org/10.3390/e19040157.

188. Gowdra, N.; Sinha, R.; MacDonell, S. Examining convolutional feature extraction using Maximum Entropy (ME) and Signal-to-Noise Ratio (SNR) for image classification. IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, 2020, pp. 471–476. doi:https://doi.org/10.1109/IECON43393.2020.9254346.

189. Mammone, N.; La Foresta, F.; Morabito, F.C. Automatic Artifact Rejection from Multichannel Scalp EEG by Wavelet ICA. *IEEE Sensors Journal* **2012**, *12*, 533–542. doi:https://doi.org/10.1109/JSEN.2011.2115236.

190. Poza, J.; Escudero, J.; Hornero, R.; Fernandez, A.; Sanchez, C.I. Regional Analysis of Spontaneous MEG Rhythms in Patients with Alzheimer's Disease Using Spectral Entropies. *Annals of Biomedical Engineering* **2008**, *36*, 141–152. doi:https://doi.org/10.1007/s10439-007-9402-y.

191. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropY. *Journal of Physiology – Heart and Circulatory Physiology* **2000**, *278*, 2039–H2049. doi:https://doi.org/10.1152/ajpheart.2000.278.6.h2039.

192. Shang, Y.; Lu, G.; Kang, Y.; Zhou, Z.; Duan, B.; Zhang, C. A multi-fault diagnosis method based on modified Sample Entropy for lithium-ion battery strings. *Journal of Power Sources* **2020**, *446*, 227275. doi:https://doi.org/10.1016/j.jpowsour.2019.227275.

193. Silva, L.E.V.; Senra Filho, A.C.S.; Fazan, V.P.S.; Felipe, J.C.; Murta Junior, L.O. Two-dimensional sample entropy: assessing image texture through irregularity. *Biomedical Physics & Engineering Express* **2016**, *2*, 045002. doi:https://doi.org/10.1088/2057-1976/2/4/045002.

194. Liao, F.; Jan, Y.K. Using Modified Sample Entropy to Characterize Aging-Associated Microvascular Dysfunction. *Frontiers in Physiology* **2016**, *7*. doi:https://doi.org/10.3389/fphys.2016.00126.

195. Nam N., Q.D.; Liu, A.B.; Lin, C.W. Development of a Neurodegenerative Disease Gait Classification Algorithm Using Multiscale Sample Entropy and Machine Learning Classifiers. *Entropy* **2020**, *22*, 1340. doi:https://doi.org/10.3390/e22121340.

196. Lake, D.E. Renyi Entropy Measures of Heart Rate Gaussianity. *IEEE Transactions on Biomedical Engineering* **2005**, *53*, 21–27. doi:https://doi.org/10.1109/TBME.2005.859782.

197. Lake, D.E. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *American journal of Physiology - Heart and Circulatory Physiology* **2011**, *300*, H319–H325. doi:https://doi.org/10.1152/ajpheart.00561.2010.

198. Humeau-Heurtier, A. Evaluation of Systems' Irregularity and Complexity: Sample Entropy, Its Derivatives, and Their Applications across Scales and Disciplines. *Entropy* **2018**, *20*, 794. doi:https://doi.org/10.3390/e20100794.

199. Belyaev, M.; Murugappan, M.; Velichko, A.; Korzun, D. Entropy-Based Machine Learning Model for Fast Diagnosis and Monitoring of Parkinson's Disease. *Sensors* **2023**, *23*, 8609. doi:https://doi.org/10.3390/s23208609.

200. Lin, G.; Lin, A. Modified multiscale sample entropy and cross-sample entropy based on horizontal visibility graph. *Chaos, Solitons & Fractals* **2022**, *165*, 112802. doi:https://doi.org/10.1016/j.chaos.2022.112802.

201. Karevan, Z.; Suykens, J.A.K. Transductive Feature Selection Using Clustering-Based Sample Entropy for Temperature Prediction in Weather Forecasting. *Entropy* **2018**, *20*, 264. doi:https://doi.org/10.3390/e20040264.

202. Guha, R.; Velegol, D. Harnessing Shannon entropy-based descriptors in machine learning models to enhance the prediction accuracy of molecular properties. *Journal of Cheminformatics* **2023**, *15*, 54. doi:https://doi.org/10.1186/s13321-023-00712-0.

203. DeMedeiros, K.; Hendawi, A.; Alvarez, M. A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks. *Sensors* **2023**, *23*, 1352. doi:https://doi.org/10.3390/s23031352.

204. Evans, S.C.; Shah, T.; Huang, H.; Ekanayake, S.P. The Entropy Economy and the Kolmogorov Learning Cycle: Leveraging the intersection of Machine Learning and Algorithmic Information Theory to jointly optimize energy and learning. *Physica D* **2024**, *461*, 134051. doi:https://doi.org/10.1016/j.physd.2024.134051.

205. Rodriguez, N.; Barba, L.; Alvarez, P.; Cabrera-Guerrero, G. Stationary Wavelet-Fourier Entropy and Kernel Extreme Learning for Bearing Multi-Fault Diagnosis. *Entropy* **2019**, *21*, 540. doi:https://doi.org/10.3390/e21060540.

206. Soltanian, A.R.; Rabiei, N.; Bahreini, F. Feature Selection in Microarray Data Using Entropy Information. In *Computational Biology*; Husi, H., Ed.; Exon Publications, 2019; chapter 10. doi:https://doi.org/10.15586/computationalbiology.2019.ch10.

207. Hoayek, A.; Rullière, D. Assessing Clustering Methods Using Shannon's Entropy. *hal-03812055v1* **2023**.

208. Yang, Z.; Lei, J.; Fan, K.; Lai, Y. Keyword extraction by entropy difference between the intrinsic and extrinsic mode. *Physica A* **2013**, *392*, 4523–4531. doi:https://doi.org/10.1016/j.physa.2013.05.052.

209. Guo, X.; Li, X.; Xu, R. Fast Policy Learning for Linear Quadratic Control with Entropy Regularization. *arXiv:2311.14168* **2023**, [2311.14168]. doi:https://doi.org/10.48550/arXiv.2311.14168.

210. Coifman, R.; Wickerhauser, M. Entropy-Based Algorithms for Best Basis Selection. *IEEE Transactions on Information Theory* **1992**, *38*, 713–718. doi:https://doi.org/10.1109/18.119732.

211. Jaynes, E.T. Information Theory and Statistical Mechanics. *Physical Review* **1957**, *106*, 620–630. doi:https://doi.org/10.1103/PhysRev.106.620.

212. Kapur, J.N.; Kesavan, H.K. Entropy Optimization Principles and Their Applications. In *Entropy and Energy Dissipation in Water Resources*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 3–20. doi:https://doi.org/10.1007/978-94-011-2430-0_1.

213. Wang, K.C. Robust Audio Content Classification Using Hybrid-Based SMD and Entropy-Based VAD. *Entropy* **2020**, *22*, 183. doi:https://doi.org/10.3390/e22020183.

214. Manzo-Martínez, A.; Gaxiola, F.; Ramírez-Alonso, G.; Martínez-Reyes, F. A Comparative Study in Machine Learning and Audio Features for Kitchen Sounds Recognition. *Computación y Sistemas* **2022**, *26*, 4244. doi:https://doi.org/10.13053/cys-26-2-4244.

215. Civera, M.; Surace, C. An Application of Instantaneous Spectral Entropy for the Condition Monitoring of Wind Turbines. *Applied Sciences* **2022**, *12*, 1059. doi:https://doi.org/10.3390/app12031059.

216. Ajmal, M.; Kushki, A.; Plataniotis, K.N. Time-Compression of Speech in Information Talks Using Spectral Entropy. Eighth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '07), 2007, pp. 80–80. doi:https://doi.org/10.1109/WIAMIS.2007.80.

217. Kapucu, F.E.; Välkki, I.; Mikkonen, J.E.; Leone, C.; Lenk, K.; Tanskanen, J.M.A.; Hyttinen, J.A.K. Spectral Entropy Based Neuronal Network Synchronization Analysis Based on Microelectrode Array Measurements. *Frontiers in Computational Neuroscience* **2016**, *10*, 112. doi:https://doi.org/10.3389/fncom.2016.00112.

218. Ra, J.S.; Li, T.; Li, Y. A novel spectral entropy-based index for assessing the depth of anaesthesia. *Brain Informatics* **2021**, *8*, 1–12. doi:https://doi.org/10.1186/s40708-021-00130-8.

219. Liu, S.; Li, Z.; Wang, G.; Qiu, X.; Liu, T.; Cao, J.; Zhang, D. Spectral–Spatial Feature Fusion for Hyperspectral Anomaly Detection. *Sensors* **2024**, *24*, 1652. doi:https://doi.org/10.3390/s24051652.

220. Rademan, M.W.; Versfeld, D.J.J.; du Preez, J.A. Soft-Output Signal Detection for Cetacean Vocalizations Using Spectral Entropy, K-Means Clustering and the Continuous Wavelet Transform. *Ecological Informatics* **2023**, *74*, 101990. doi:https://doi.org/10.1016/j.ecoinf.2023.101990.

221. Oida, E.; Moritani, T.; Yamori, Y. Tone-Entropy Analysis on Cardiac Recovery After Dynamic Exercise. *Journal of Applied Physiology* **1997**, *82*, 1794–1801. doi:https://journals.physiology.org\linebreak /doi/full/10.1152/jappl.1997.82.6.1794##.

222. Khandoker, A.H.; Al Zaabi, Y.; Jelinek, H.F. What Can Tone and Entropy Tell Us About Risk of Cardiovascular Diseases? 2019 Computing in Cardiology (CinC), 2019, pp. 1–4. doi:https://doi.org/10.22489/CinC.2019.120.

223. Khandoker, A.; Karmakar, C.; Kimura, Y.; Endo, M.; Oshio, S.; Palaniswami, M. Tone Entropy Analysis of Foetal Heart Rate Variability. *Entropy* **2015**, *17*, 1042–1053. doi:https://doi.org/10.3390/e17031042.

224. Karmakar, C.K.; Khandoker, A.H.; Palaniswami, M. Multi-scale Tone Entropy in differentiating physiologic and synthetic RR time series. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 6135–6138. doi:https://doi.org/10.1109/EMBC.2013.6610953.

225. Amigó, J.M.; Giménez, Á. A Simplified Algorithm for the Topological Entropy of Multimodal Maps. *Entropy* **2014**, *16*, 627–644. doi:https://doi.org/10.3390/e16020627.

226. Amigó, J.M.; Giménez, A. Formulas for the topological entropy of multimodal maps based on min-max symbols. *Discrete and Continuous Dynamical Systems B* **2015**, *20*, 3415–3434. doi:https://doi.org/10.3934/dcdsb.2015.20.3415.

227. Lum, P.; Singh, G.; Lehman, A.; Ishkanov, T.; Vejdemo-Johansson, M.; Alagappan, M.; Carlsson, J.; Carlsson, G. Extracting insights from the shape of complex data using topology. *Scientific Reports* **2013**, *3*, 1236. doi:https://doi.org/10.1038/srep01236.

228. McCullough, M.; Small, M.; Iu, H.H.C.; Stemler, T. Multiscale Ordinal Network Analysis of Human Cardiac Dynamics. *Philosophical Transactions of the Royal Society A* **2017**, *375*, 20160292. doi:http://dx.doi.org/10.1098/rsta.2016.0292.

229. Zhao, Y.; Zhang, H. Quantitative Performance Assessment of CNN Units via Topological Entropy Calculation. *arXiv:2103.09716* **2022**, [2103.09716]. doi:https://doi.org/10.48550/arXiv.2103.09716.

230. Jiménez-Alonso, J.F.; López-Martínez, J.; Blanco-Claraco, J.L.; González-Díaz, R.; Sáez, A. A topological entropy-based approach for damage detection of civil engineering structures. 5th International Conference on Mechanical Models in Structural Engineering (CMMoST 2019), 2019, pp. 55–62.

231. Rong, L.; Shang, P. Topological entropy and geometric entropy and their application to the horizontal visibility graph for financial time series. *Nonlinear Dynamics* **2018**, *92*, 41–58. doi:https://doi.org/10.1007/s11071-018-4120-6.

232. Rucco, M.; Gonzalez-Diaz, R.; Jimenez, M.; Atienza, N.; Cristalli, C.; Concettoni, E.; Ferrante, A.; Merelli, E. A new topological entropy-based approach for measuring similarities among piecewise linear functions. *Signal Processing* **2017**, *134*, 130–138.

233. Schreiber, T. Measuring Information Transfer. *Physical Review Letters* **2000**, *85*, 461–464. doi:https://link.aps.org/doi/10.1103/PhysRevLett.85.461.

234. Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* **1969**, *37*, 424–438.

235. Moldovan, A.; Caţaron, A.; Răzvan, A. Learning in Convolutional Neural Networks Accelerated by Transfer Entropy. *Entropy* **2021**, *23*, 1281. doi:https://doi.org/10.3390/e23091218.

236. Moldovan, A.; Caţaron, A.; Andonie, R. Transfer Entropy in Graph Convolutional Neural Networks. *arXiv:2406.06632* **2024**. doi:https://doi.org/10.48550/arXiv.2406.06632.

237. Herzog, S.; Tetzlaff, C.; Wörgötter, F. Transfer entropy-based feedback improves performance in artificial neural networks. *arXiv:1706.04265* **2017**, [arXiv:cs.LG/1706.04265]. doi:https://doi.org/10.48550/arXiv.1706.04265.

238. Duan, Z.; Xu, H.; Huang, Y.; Feng, J.; Wang, Y. Multivariate Time Series Forecasting with Transfer Entropy Graph. *Tsinghua Science and Technology* **2023**, *28*, 141–149. doi:10.26599/TST.2021.9010081.

239. Amblard, P.O.; Michel, O.J.J. The relation between Granger causality and directed information theory: A Review. *Entropy* **2013**, *15*, 113–143. doi:https://doi.org/10.3390/e15010113.

240. Alomani, G.; Kayid, M. Further Properties of Tsallis Entropy and Its Application. *Entropy* **2023**, *25*, 199. doi:https://doi.org/10.3390/e25020199.

241. Sharma, S.; Bassi, I. Efficacy of Tsallis Entropy in Clustering Categorical Data. 2019 IEEE Bombay Section Signature Conference (IBSSC), 2019, pp. 1–5. doi:https://doi.org/10.1109/IBSSC47189.2019.8973057.

242. Wu, D.; Jia, H.; Abualigah, L.; Xing, Z.; Zheng, R.; Wang, H.; Altalhi, M. Enhance Teaching-Learning-Based Optimization for Tsallis-Entropy-Based Feature Selection Classification Approach. *Processes* **2022**, *10*, 360. doi:https://doi.org/10.3390/pr10020360.

243. Naidu, M.S.R.; Kumar, P.R. Tsallis Entropy Based Image Thresholding for Image Segmentation. In *Computational Intelligence in Data Mining*; Springer, Singapore, 2017; Vol. 556, pp. 371–379. doi:https://doi.org/10.1007/978-981-10-3874-7_34.

244. Kalimeri, M.; Papadimitriou, C.; Balasis, G.; Eftaxias, K. Dynamical complexity detection in pre-seismic emissions using nonadditive Tsallis entropy. *Physica A* **2008**, *387*, 1161–1172. doi:https://doi.org/10.1016/j.physa.2007.10.053.

245. Belanche-Muñoz, L.A.; Wiejacha, M. Analysis of Kernel Matrices via the von Neumann Entropy and Its Relation to RVM Performances. *Entropy* **2023**, *25*, 154. doi:https://doi.org/10.3390/e25010154.

246. Hu, F.; Tian, K.; Zhang, Z.K. Identifying Vital Nodes in Hypergraphs Based on Von Neumann Entropy. *Entropy* **2023**, *25*, 1263. doi:https://doi.org/10.3390/e25091263.

247. Chen, P.Y.; Wu, L.; Liu, S.; Rajapakse, I. Fast Incremental von Neumann Graph Entropy Computation: Theory, Algorithm, and Applications. *arXiv:1805.11769* **2019**, [1805.11769]. doi:https://doi.org/10.48550/arXiv.1805.11769.

248. Ye, C.; Wilson, R.C.; Hancock, E.R. Network analysis using entropy component analysis. *Journal of Complex Networks* **2017**, *6*, 831. doi:https://doi.org/10.1093/comnet/cnx051.

249. Huang, Y.; Zhao, Y.; Capstick, A.; Palermo, F.; Haddadi, H.; Barnaghi, P. Analyzing entropy features in time-series data for pattern recognition in neurological conditions. *Artificial Intelligence in Medicine* **2024**, *150*, 102821. doi:https://doi.org/10.1016/j.artmed.2024.102821.

250. Rosso, O.A.; Blanco, S.; Yordanova, J.; Kolev, V.; Figliola, A.; Schürmann, M.; Başar, E. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods* **2001**, *105*, 65–75. doi:https://doi.org/10.1016/S0165-0270(00)00356-3.

251. Hu, P.; Zhao, C.; Huang, J.; Song, T. Intelligent and Small Samples Gear Fault Detection Based on Wavelet Analysis and Improved CNN. *Processes* **2023**, *11*, 2969. doi:https://doi.org/10.3390/pr11102969.

252. Cuomo, K.M.; Oppenheim, A.V.; Strogatz, S.H. Synchronization of Lorenz-Based Chaotic Circuits with Applications to Communications. *IEEE Transactions on Circuits and Systems II* **1993**, *40*, 626–633. doi:https://doi.org/10.1109/82.246163.

253. Amigó, J.M.; Kocarev, L.; Tomovski, I. Discrete entropy. *Physica D* **2007**, *228*, 77–85. doi:https://doi.org/10.1016/j.physd.2007.03.001.

254. Amigó, J.M.; Kocarev, L.; Szczepanski, J. Discrete Lyapunov exponent and resistance to differential cryptanalysis. *IEEE Transactions on Circuits and Systems II* **2007**, *54*, 882–886. doi:https://doi.org/10.1109/TCSII.2007.901576.

255. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, second edition ed.; Springer: New York, 2009. doi:https://doi.org/10.1007/b94608.

256. Jenssen, R. Kernel Entropy Component Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2010**, *32*, 847–860. doi:https://doi.org/10.1109/TPAMI.2009.100.

257. Paninski, L. Estimation of Entropy and Mutual Information. *Neural Computation* **2003**, *15*, 1191–1253. doi:https://doi.org/10.1162/089976603321780272.