Statistical and Symbolic Neuroaesthetics Rules Extraction From EEG Signals

M. Coccagna[1], F. Manzella[2], S. Mazzacane[1], G. Pagliarini[2], and G. Sciavicco[2]

[1] Dept. of CIAS Interdepartmental Research Center, University of Ferrara, Italy
[2] Dept. of Mathematics and Computer Science, University of Ferrara, Italy

Abstract. Neuroaesthetics, as defined by Zeki in 1999, is the scientific approach to the study of aesthetic perceptions of art, music, or any other experience that can give rise to aesthetic judgments. One way to understand the processes of neuroaesthetics is studying the electroencephalogram (EEG) signals that are recorded from subjects while they are exposed to some expression of art, and study how the differences among such signals correlate to the differences in their subjective judgments; typically, such studies are conducted on limited data with a purely statistical signal analysis. In this paper we consider a larger data set which was previously used in an experiment on beauty perception; we apply a novel machine learning-based data analysis methodology that allows us to extract symbolic *like/dislike* rules on the voltage at the most relevant frequencies from the most relevant electrodes. Our approach is not only novel in this particular area, but it is also reproducible and allows us to treat large quantities of data.

Keywords: Interpretable Machine Learning · Neural Activity Classification · Modal Decision Trees · Neural Correlates

1 Introduction

Neuroaesthetics was defined by Zeki [21] as the scientific approach to the study of the perception of beauty in a broad sense, and later formalized as a field of study by Nalbantian [14]. The approaches to neuroaesthetics vary very much in the recent scientific literature, but they can be essentially divided into *top-down* processes, in which the essence of beauty undergoes an axiomatic treatment in which the subjective feeling is broken down in its constituting elements, and *bottom-up* ones, in which some kind of objective data is analyzed and related to the subjective expression of beauty. While the former, e.g. as in [20], is certainly fascinating from an epistemological point of view (i.e., it tries to answer the question of whether *beauty can be defined*), the latter has the advantage of being based on objective data, being systematic, and taking advantage from modern analysis techniques. For these reasons, bottom-up approaches are more common.

Bottom-up strategies aimed to understand how the brain interprets the sense of beauty can be classified into (functional) MRI (*magnetic resonance*) image

processing and interpretation, and EEG (*electroencephalogram*) signal processing. In the first case, the specific aim is to understand the which are the involved areas of the brain, how they are activated, and by what are they influenced. In [4], Di Dio, Macaluso, and Rizzolatti designed and carried on an experiment using functional MRI; by analyzing the data produced by 14 subjects, they concluded that the sense of beauty is mediated by two non-mutually exclusive processes: one based on a joint activation of sets of cortical neurons (objective beauty), the other based on the activation of the amygdala, driven by one's own emotional experiences (subjective beauty). Functional MRI was also used by Chatterjee et al. [1], in which the authors used a AI-based system to create artificial face stimuli in order to measure the response to beauty in the brains of 13 subjects. Ishizu and Zeki, in [8], considered functional MRI data from 21 subjects, and established that the experience of beauty is transversal to being stimulated by, for example, vision or hearing, at least in terms of functional brain area. Huang et al., on the other hand, asked the question of whether during the vision of art the experience of beauty is influenced by the apriori knowledge of an artwork being an original masterpiece or rather a copy [7], and their result seem to indicate that such a knowledge has zero or very low influence. Jacobs, Renken, and Cornelissen asked the question of whether the subjective sense of beauty is influenced by the internal state of the subject [9], and the concluded that, instead, there seems to be areas specifically devoted to aesthetic assessment, irrespective of the stimulus type. Kühn and Gallinat researched and investigated concurrence across 40 studies reporting brain regions which seem correlated with self-reported judgements of subjective pleasantness [10]. A few, more recent contributions focus specifically on the use of statistical/machine learning related techniques used to assess the problem of establishing the brain mechanisms that regulate the sense of beauty. Among these, we mention Reng and Geng's work [15], based on a learning process founded on labels' distributions among attractiveness rating, and Seresinhe, Prais, and Moat's contribution [18], based on using a deep learning strategy to understand and predict the sense of scenic of a subject looking at an outdoor image. When it comes to EEG signal processing, on the other hand, most work is quite recent. In [6] Hadjidimitriou and Hadjileontiadis used different feature extraction approaches and classifiers and focused on the discrimination between subjects' EEG responses to self-assessed liked or disliked music. In [2], Chew, Teo, and Mountstephens used neural networks to classify EEG signals recorded from subjects while viewing 3D shapes and expressing their level of appreciation. Finally, in [5], Guo et. al. designed and executed an experiment to measure and classify EEG signals recorded while the subjects were exposed to 3D prototype led lamps, with the aim of predicting the subjects' sense of like/dislike. Li and Zhang's review [22] on computational neuroaesthetics lists and discusses most of this work.

At least two considerations can be drawn from reviewing the literature. First, most of the existing work on computational neuroaesthetics focuses on artificial, 2D images or 3D shapes, designed with experimental purposes, and not submerged into a realistic context. Second, EEG signals have been analyzed with

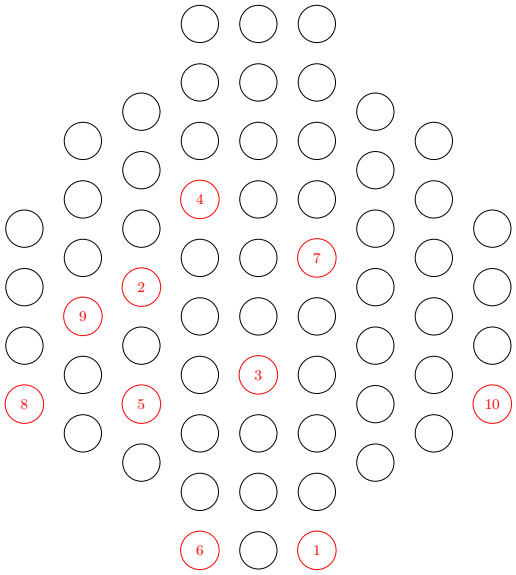


Fig. 1. Selected electrodes disposition

functional machine learning techniques that do not offer interpretability of the results, but only their statistical value, and do not allow, in general, the extraction of rules. The data used in this paper have been collected during a real-world experiment, analysing the explicit and implicit reactions of participants, using different kinds of sensors, during their visit to the exhibition *Painting affections: sacred painting in Ferrara between the '500 and the '700*, set up at the Estense Castle in Ferrara from the 26th of January to the 26th of December 2019 [12, 16] (ISRCTN: 70216542). The EEG data are part of a larger bio-signals data set recorded to evaluate the participants' reactions during the observation of paintings, including ECG (*electrocardiogram*), EDA (*electrodermal activity*), two different tools for EEG recording, the *gaze pattern* during the observations, the participants' main features (age, gender, education, familiarity with art, etc.) and their explicit judgments about paintings. Here, we focus on a total of 248 trials (16 subjects, exposed up to 18 paintings each), and we analyze their EEG recording by means of a new machine learning technique, called *temporal random forests* [13, 17], with the purpose of, first, establishing which electrodes the EEG signal carries the most information about the subjects' personal beauty experience, and, second, if symbolic rules can be extracted from the classification models.

2 Data Origin, Description, and Preparation

EEG data have been recorded through two main tools, to analyse their affordability and easiness of use: a dry electrode EEG cap *Waveguard™touch* by *ANT Neuro*

in the 64-channel variant, as well as a *Myndplay* headsets, managed through *OpenVibe* software. In this paper, we focus on the recording from the cap only, and our data set consists of 16 people (10 males, 6 females, average age 32.19). For each subject a single long recording was made from the beginning to the end of the exhibition, therefore a slicing operation was necessary based on the timing the subject was standing in front of each painting to collect only the EEG data relative to a specific painting (about 60 seconds each). Being recorded at a sampling rate of 512 Hz, each resulting data slice consisted of a time series of 30720 points. Each subject was asked to express a integer in the range [0, 50] as an expression of his/her *liking score* of the painting. Since not all pairs subject/painting were recorded, the resulting data set is composed by 248 instances, each instance being described by 64 time series (one for each electrode) of 30720 points and labeled by the related liking score.

The EEG data was processed by applying the Short Time Fourier Transform (STFT), using different parametrizations for the two phases: *best electrode selection* and *knowledge extraction*. For the former, the range of frequencies $0 - 256$ Hz was explored, corresponding to the entire available spectrum for a sampling rate of 512 Hz (i.e., the Nyquist frequency is 256 Hz [19]). This spectrum was divided into 60 equally wide bands of frequencies $(F_1, \ldots, F_{60})$, with a resulting width of 4.267 Hz. The STFT time window size was set to 50 ms with a step time of 25 ms, and the resulting time series were 2665 points long. For the knowledge extraction phase a narrower frequency range was fixed ($0 - 51.2$ Hz), composed by five relevant wave patterns as suggested by the literature: δ, ranging in $0 - 4$ Hz (usually associated with slow-wave sleep), θ, ranging in $4 - 7$ Hz (usually associated with phase 1 and 2 of non-REM sleep and with REM sleep), α, ranging in $7 - 13$ Hz (usually associated with waking state with closed eyes and instants prior to fall asleep), β, ranging in $14 - 30$ Hz (usually associated with intense mental activity), and γ, ranging in $30 - 49$ Hz (usually associated with states of particular stress). After trying multiple configurations for the width of the STFT frequency window, the one providing the best results was ultimately found to be 3.01 Hz, which divides the whole frequency range into 17 bands $(\overline{F}_1, \ldots, \overline{F}_{17})$. Similarly, multiple settings for the time window parameter were explored, and we finally selected a size of 300 ms and a step time of 100 ms. The resulting time series were 638 points long. In the end, we found that the first quarter of each series, corresponding to the first 1 seconds, actually contains all relevant information, and therefore we limited the data set to the the first 159 points, further reduced to 15 points by means of a moving average filter.

The scores among the data set were non-normally distributed ($p = 10^{-56}$, Kolmogorov-Smirnov test). For the purpose of learning, we treated the resulting problem as a classification problem, by discretizing the liking scores into three categories (*dislike*, *neutral*, *like*), with thresholds, respectively, of 16 and 33, both included, and, finally, by excluding the middle category (neutral). The resulting data set was composed by 81 *positive* instances (*like*) and 86 negative instances (*dislike*).

measure	symbol
mean	M
max	MAX
min	MIN
mode of z-scored distribution (5-bin histogram)	$Z5$
mode of z-scored distribution (10-bin histogram)	$Z10$
longest period of consecutive values above the mean	C
time intervals between successive extreme events above the mean	A
time intervals between successive extreme events below the mean	B
first $1/e$ crossing of autocorrelation function	FC
first minimum of autocorrelation function	FM
tot. power in lowest 1/5 of frequencies in the Fourier power spectrum	TP
centroid of the Fourier power spectrum	CE
mean error from rolling 3-sample mean forecasting	ME
time-reversibility statistic $\langle(x_{t+1-x_t})^3\rangle t$	TR
automutual information ($m = 2$, $\tau = 5$)	AI
first minimum of the automutual information function	$FMAI$
proportion of successive differences exceeding 0.04σ (Mietus 2002)	PD
longest period of successive incremental decreases	LP
Entropy of two successive letters in equiprobable 3-letter symbolization	EN
change in correlation length after iterative differencing	CC
exponential fit to successive distances in 2-d embedding space	EF
ratio of slower timescale fluctuations that scale with DFA (50% sampling)	$FDFA$
ratio of slower timescale fluctuations that scale with linearly rescaled range fits	FLF
trace of covariance of transition matrix between symbols in 3-letter alphabet	TC
periodicity measure of (Wang et al. 2007)	PM

Table 1. Twenty-five statistical measures for time series (including 22 measures from [11]).

3 Statistical Analysis Phase

We modeled this problem as a multivariate temporal series classification problem. Naïve symbolic treatment of time series consists of simple feature extraction for further analysis; typical examples of interesting features are minimum, maximum, or average of a series. As suggested [3, 11], elementary features can be combined with more elaborate ones, in order to derive a systematic, statistical treatment of pure temporal data. Prior to the knowledge extraction phase, we aimed to highlight which electrodes and which measures are more prone to have a role in the problem. To this end, we performed a statistical analysis based on the assumption that, when coupled with a given measure, a frequency is informative if it displays a high variance across the data set. We proceeded as follows:

- We first considered the electrodes listed as $E_1, \ldots, E_{64}$ in Fig. 2, and we computed the variance of the value of each one of the measures in Tab. 1 for each one of the electrodes, on each one of the frequencies $F_1, \ldots, F_{60}$; then we aggregated the result by electrode by averaging the variances, and sorted the electrodes themselves. As a result, we obtained a list of electrodes that displayed, on average, the highest variance across the whole spectrum

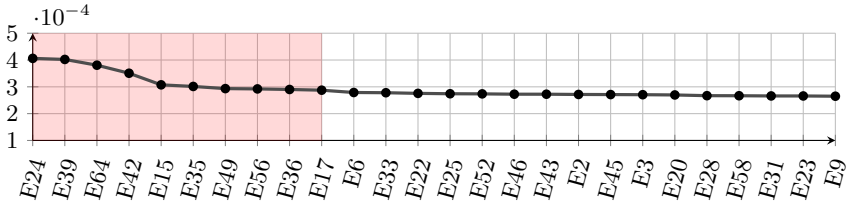


Fig. 2. Normalized mean variance of each electrode shown in descending order.

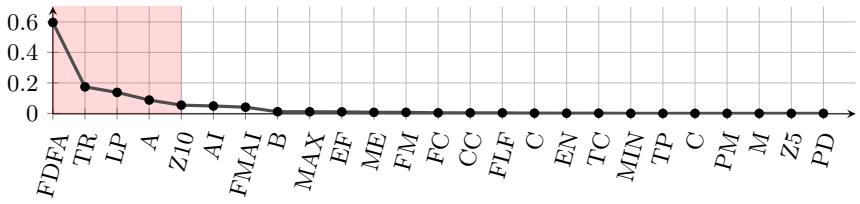


Fig. 3. Normalized mean variance of each measure shown in descending order.

of frequencies. We proceeded by selecting the first 10 electrodes, which are the ones highlighted in Fig. 2.

- Then, we averaged the variance of each of the measures in Tab. 1 across all frequencies $F_1, \ldots, F_{60}$ and the 10 selected electrodes, to identify the 5 most informative ones, as in Fig. 3.

Each pair electrode/measure gives can be compared among the two classes (like,dislike) across each of the 60 frequencies band. An graphical example of such a comparison can be seen in Fig. 4, namely for F_3, and for the five most informative electrodes. As it can be noticed, the differences are subtle, which is an indication of the difficulty of the problem if approached with purely, univariate statistical methods.

4 Knowledge Extraction

Time series classification can be approached in several ways. Methods for classifying time series can be roughly divided into *symbolic* and *functional*. Symbolic methods aim to extract a logical characterization of the classes in terms of the behaviour of the series, while functional ones approach the classification problem by extracting a mathematical function of the series. Time series classification methods can also be divided into *native* or *feature-based*. Native methods consider time series as they are, without performing any modification of the signals. Feature-based method, on the other hand, focus on extracting statistically interesting measures of the signals and use those for the classification phase. Feature-based methods are far more common, and they are both symbolic and functional; their major drawback is the lack of interpretability of the results in

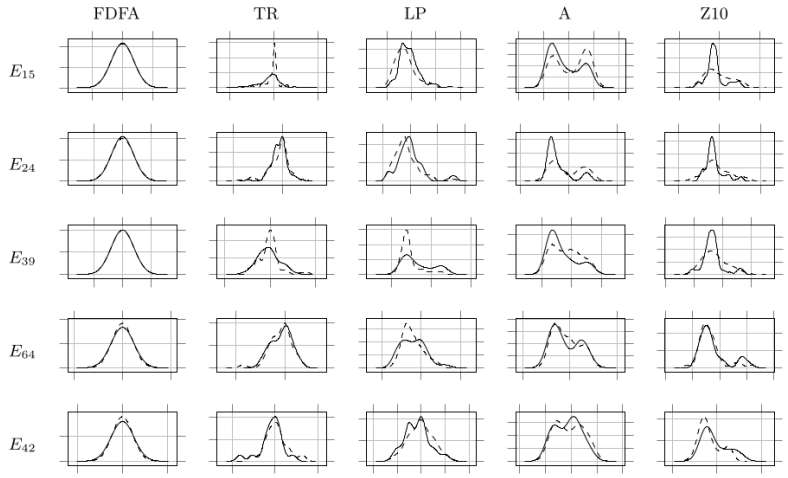


Fig. 4. Distribution of each of the selected variables by measure.

the functional case, and the low predictive capabilities in the symbolic one. Native methods are more scarce, and the most common ones among them, that is, distance-based methods, do not offer, in general, a real grasp of the underlying problem, despite their general good behaviour in terms of performances.

In [13, 17], a new class of symbolic, native time series classification methods was proposed. Despite their short history, *temporal decision trees* showed a good compromise between interpretability and performances. The key points that define temporal decision trees are:

- They follow the general pattern and schema of conventional decision trees. Decisions are taken on a data set in order to maximize the amount of *information gain* in a greedy fashion, starting from the original training data set and obtaining, at each step, smaller, and more informative subsets. When the data set associated with a node is too small, or too pure in terms of class, it is converted into a leaf, and labeled with the majority class (generating, as in the classical case, a certain amount of misclassifications). Classical techniques, up to and including *pre- and post-pruning*, can be applied, at least in a limited form, to propositional and temporal decision trees alike.
- Unlike conventional decision trees, decisions are relativized to intervals on time series. So, while conventional decision trees treat times series by extracting features from them, and then taking decisions on such features, temporal decision trees take decisions directly on time series, in a native way. Consider, for example, the mean; while a conventional decision tree may separate the data set using the fact that the mean of a specific variable on the whole time period exceeds some value (e.g. *if the mean value of a*

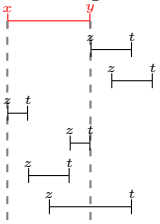
symbol	Allen's relation	graphical representation
$\langle A \rangle$	$[x,y]R_A[z,t] \Leftrightarrow y = z$	
$\langle L \rangle$	$[x,y]R_L[z,t] \Leftrightarrow y < z$	
$\langle B \rangle$	$[x,y]R_B[z,t] \Leftrightarrow x = z, t < y$	
$\langle E \rangle$	$[x,y]R_E[z,t] \Leftrightarrow y = t, x < z$	
$\langle D \rangle$	$[x,y]R_D[z,t] \Leftrightarrow x < z, t < y$	
$\langle O \rangle$	$[x,y]R_O[z,t] \Leftrightarrow x < z < y < t$	

Fig. 5. Allen's interval relations and their notation in temporal decision trees.

variable is more than that value, then...), a temporal decision tree may do so using the existence of an interval in which the mean of a specific variable exceeds some value (e.g. *if the mean value of a variable is more than that value between the instants x and y, then...*).

- Like conventional decision trees, a temporal decision tree has a clear logical interpretation, but makes use of a more complex logic than propositional logic, which allows one to express properties over intervals, and their relations. There are thirteen relations between two intervals, known as *Allen's* relations (see Fig. 5, in which we show only the six *direct* relations of the type $\langle X \rangle$; their inverses, denoted with $\langle \overline{X} \rangle$, can be obtained by switching the roles of each interval, and the thirteenth, *equals*, is denoted $\langle = \rangle$), and a temporal decision tree is able to learn interval patterns which we can formalize using suitable symbols to denote Allen's relations (see Fig. 5, first column).

In [13, 17] it was shown that temporal decision trees perform better than their propositional counterparts, and, while retaining a very high level of interpretability, are able to extract classification models that are comparable with those extracted by non-interpretable approaches.

We used temporal decision trees, in their *random forest* generalization, to build models based with the selected electrodes and the selected measures, using the frequency bands $\bar{F}_1, \ldots, \bar{F}_{17}$ (from δ to γ). We used forests with 100 trees, run in 10-fold cross validation mode, with four different configurations in terms of selected frequencies. The results of our experiment can be seen in Tab. 6. As it can be noticed, the best results are obtained with β frequencies ($\bar{F}_6$ to $\bar{F}_{10}$), followed by γ ($\bar{F}_{11}$ to $\bar{F}_{17}$), and then by the combination with all frequencies. The standard deviations across the 10 repetitions are not too high, indicating that the experiment is quite solid. Some of the seeds (e.g., seed 2) show very high values (e.g., accuracy 77.5%). More important than the statistical value of our models are the rules that can be extracted from them. As we have recalled, these have the form of formulas in the logic of Allen's relations. Among the many possible rules that a model such as our encompasses, we have selected two examples for dislike (d) and like (l):

	accuracy					κ index			
seed	*all*	β	γ	$\beta+\gamma$	seed	*all*	β	γ	$\beta+\gamma$
1	68.75	70.62	68.75	69.38	1	37.50	41.25	37.50	38.75
2	73.12	77.50	70.00	70.62	2	46.25	55.00	40.00	41.25
3	66.88	68.75	61.88	66.88	3	33.75	37.50	23.75	33.75
4	57.50	60.00	58.13	58.13	4	15.00	20.00	16.25	16.25
5	65.00	64.38	66.88	62.50	5	30.00	28.75	33.75	25.00
6	66.25	68.12	68.12	67.50	6	32.50	36.25	36.25	35.00
7	61.88	63.75	60.00	66.25	7	23.75	27.50	20.00	32.50
8	70.00	66.88	68.12	68.75	8	40.00	33.75	36.25	37.50
9	72.50	71.88	68.12	72.50	9	45.00	43.75	36.25	45.00
10	64.38	68.12	70.00	70.62	10	28.75	36.25	40.00	41.25
avg	66.63	**68.00**	67.31	66.00	avg	33.25	36.00	32.00	34.63
std	4.79	4.82	4.33	4.27	std	9.58	9.64	8.66	8.54

Fig. 6. Results.

$$\varphi_d = [L](A(A_{20}) \geq -0.7 \rightarrow ([B](TR(A_{51} > -3.04 \rightarrow \langle \overline{A} \rangle (Z10(A_{80} < 0.66)))))$$
$$\varphi_l = [L]A(A_{122} > -0.71) \wedge [L]TR(A_6 < 3.65) \wedge [L](TR(A_{84} < 4.30))$$

where, in the first rule, which has an estimated validation confidence of 0.75, A_{20} (resp., A_{51}, A_{80}) is the pair $E_{64}/\bar{F}_7$ (resp., $E_{35}/\bar{F}_{11}$, $E_{36}/\bar{F}_{13}$), and in the second rule, which has an estimated confidence of 0.8, A_{122} (resp., A_6, A_{84}) is the pair $E_{56}/\bar{F}_3$ (resp., $E_2/\bar{F}_6$, $E_{15}/\bar{F}_{16}$).

5 Conclusions

In this paper we have applied a new methodology to extract knowledge from EEG signals, in order to study a neuroaesthetic problem. Our data consisted of the EEG recordings of subjects exposed to artistic paintings, and we analyzed them to extract rules that allows one to establish if the subject liked, or disliked, the painting. In a way, this work contributes to the more general neuroaesthetic problem of synthesizing a *theory of beauty*. While our results must be considered only preliminary, they constitute a first step towards using a new generation of knowledge extraction methods, by means of which one trades, in a limited way, some degree of performance of a prediction model (i.e., reliability of the prediction) to obtain, again in a limited way, explicit knowledge.

References

1. Chatterjee, A., Thomas, A., Smith, S., Aguirre, G.: The neural response to facial attractiveness. Neuropsychology **23**(2), 135–143 (2009)
2. Chew, L., Teo, J., Mountstephens, J.: Aesthetic preference recognition of 3d shapes using EEG. Cognitive Neurodynamics **10**(2), 165–173 (2016)

3. Christ, M., Braun, N., Neuffer, J., Kempa-Liehr, A.: Time series feature extraction on basis of scalable hypothesis tests. Neurocomputing **307**, 72–77 (2018)
4. Dio, C.D., Macaluso, E., Rizzolatti, G.: The golden beauty: Brain response to classical and renaissance sculptures. Plos One (11), 1–9 (2007)
5. Guo, F., Li, M., Hu, M., Li, F., Lin, B.: Distinguishing and quantifying the visual aesthetics of a product: An integrated approach of eye-tracking and EEG. International Journal of Industrial Ergonomics **71**, 47–56 (2019)
6. Hadjidimitriou, S., Hadjileontiadis, L.: Toward an EEG-based recognition of music liking using time-frequency analysis. IEEE Transactions on Biomedical Engineering **59**(12), 3498–3510 (2012)
7. Huang, M., Bridge, H., Kemp, M., Parker, A.: Human cortical activity evoked by the assignment of authenticity when viewing works of art. Frontiers in Human Neuroscience **5**, 1–20 (2011)
8. Ishizu, T., Zeki, S.: Toward a brain-based theory of beauty. Plos One **6**(7), 1–10 (2011)
9. Jacobs, R., Renken, R., Cornelissen, F.: Neural correlates of visual aesthetics – beauty as the coalescence of stimulus and internal state. Plos One **7**(2), 1–8 (2012)
10. Kühn, S., Gallinat, J.: The neural correlates of subjective pleasantness. Neuroimage **61**, 289 – 294 (2012)
11. Lubba, C., Sethi, S., Knaute, P., Schultz, S., Fulcher, B., Jones, N.: Catch22: Canonical time-series characteristics - selected through highly comparative time-series analysis. Data Min. Knowl. Discov. **33**(6), 1821–1852 (2019)
12. M. Coccagna, e.a.: Environment and people perceptions: The experience of nevart, neuroestethics of the art vision. In: Proc. of Global Challenges in Assistive Technology: Research, Policy & Practice. pp. 204 – 205 (2019)
13. Manzella, F., Pagliarini, G., Sciavicco, G., Stan, I.: Interval temporal random forests with an application to COVID-19 diagnosis. In: Proc. of the 28th International Symposium on Temporal Representation and Reasoning. LIPIcs, vol. 206, pp. 7:1–7:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2021)
14. Nalbantian, S.: Neuroaesthetics: Neuroscientific theory and illustration from the arts. Interdisciplinary Science Reviews **33**(4), 357–368 (2008)
15. Ren, Y., Geng, X.: Sense beauty by label distribution learning. In: Proc. of the 26th International Joint Conference on Artificial Intelligence. pp. 2648–2654 (2017)
16. S. Mazzacane, e.a.: Neuroaesthetics of art vision: An experimental approach to the sense of beauty. Journal of Clinical Trials **10**, 1–8 (2020)
17. Sciavicco, G., Stan, I.: Knowledge extraction with interval temporal logic decision trees. In: Proc. of the 27th International Symposium on Temporal Representation and Reasoning. LIPIcs, vol. 178, pp. 9:1–9:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2020)
18. Seresinhe, C., Preis, T., Moat, H.: Using deep learning to quantify the beauty of outdoor places. Royal Society Open Science **4**(7), 1–14 (2017)
19. Shannon, C.: Communication in the presence of noise. Proceedings of the IRE **37**(1), 10–21 (1949)
20. Sidhu, D., McDougall, K., Jalava, S., Bodner, G.: Prediction of beauty and liking ratings for abstract and representational paintings using subjective and objective measures. Plos One **13**(7), 1–15 (2018)
21. Zeki, S.: Inner Vision: An Exploration of Art and the Brain. Oxford University Press (1999)
22. Zhang, J., Li, R.: Review of computational neuroaesthetics: bridging the gap between neuroaesthetics and computer science. Brain Informatics **7**(16), 1–17 (2020)