A SHAP Quotient Game for explaining Raman Spectroscopy classification models

Marco Piazza[1][0000−0002−8337−0415], Mauro Passacantando[1][0000−0003−2098−8362], Marzia Bedoni[2][0000−0003−2618−3661], and Enza Messina[1][0000−0002−4062−0824]

[1] University of Milano-Bicocca, Milan 20125, Italy
{marco.piazza,mauro.passacantando,enza.messina}@unimib.it
[2] IRCCS Fondazione Don Carlo Gnocchi ONLUS, Milan, 20148, Italy
mbedoni@dongnocchi.it

Abstract. Recent advances in healthcare technologies have enabled the collection of large amounts of biological samples across various techniques and applications. In particular, over the past few years, *Raman Spectroscopy* (RS) has been successfully applied to the early-stage diagnosis of diseases. However, the inherent complexity and variability of spectra make manual analysis challenging, even for domain experts. For this reason, there has been exponential growth in the number of studies applying *Machine Learning* (ML) for automatic spectra classification. Despite their success, ML models are often considered *black-box* systems, meaning it is difficult to interpret the patterns they learn from the training data. To address this limitation, approaches in the field of *eXplainable Artificial Intelligence* (XAI) have been proposed. Among these, *SHAP*, a method based on cooperative game theory, is one of the most widely used. While SHAP provides strong theoretical guarantees and good empirical results, its computational complexity poses significant challenges for problems characterized by a large number of features. In this paper, we propose a novel XAI method inspired by SHAP and based on the concept of *Quotient Game* (QG) from cooperative game theory with coalition structure. Our approach is specifically designed to effectively handle spectral data charachterized by high-dimensionality. We validate the effectiveness of our method using two real-world datasets of saliva Raman spectra: one for *Covid-19* diagnosis and another for the identification of *Parkinson's* and *Alzheimer's* diseases. The explanations generated by our Quotient Game approach are compared with state-of-the-art methods through computational experiments.

Keywords: eXplainable AI · Raman Spectroscopy · Game Theory

1 Introduction

Raman Spectroscopy is a technique based on the inelastic scattering of monochromatic light to observe low-frequency modes in a molecular system [4]. The resulting scattering pattern serves as a 'fingerprint', revealing information about

the sample's chemical composition, including the presence, concentrations, and interactions of its molecules. This information can be exploited for different purposes in different fields, including food industry for fraud detection [16] and forensics for crime scene analysis [12]. One of the most promising applications of Raman Spectroscopy emerged in the last few years in the healthcare domain for disease diagnosis. Indeed, the spectral information acquired from biological samples can be exploited for early diagnosis and for monitoring emerging disease by detecting certain biomarkers associated with the suspected condition. These applications typically involve a variety of target samples, such as blood-based fluids (serum, plasma) or human tissues [8,11]. In recent research [5,6], the analysis of saliva samples has demonstrated potential for identifying the presence of relevant biomarkers and their concentration, making saliva one of the most promising targets for analysis, especially considering its straightforward accessibility. However, the acquired spectrum is difficult to interpret, not only because of the noise introduced during acquisition but also because of the complex interactions of different molecules that are present within the biological sample. Recent works successfully applied Machine Learning and Deep Learning (DL) for building accurate classification models of Raman samples, enabling the development of a fast, cheap, non-invasive, diagnostic tools. The adoption of a diagnostic system based on ML/DL may be limited due to their lack of Explainability, as these models functions as 'black boxes'. While this may not always be a critical issue, in the healthcare domain, some degree of understanding is essential. Indeed, both patients and clinicians are more likely to trust an automatic diagnosis when it is accompanied by a justification. Moreover, under the GDPR and the AI Act [3], this is already a legal requirement in the European Union.

The field of eXplainable Artificial Intelligence emerged during the last decade with the objective of providing an explanation for the predictions of ML and DL models. Despite numerous works being proposed over the years to handle images and tabular data, adapting these methods to non-traditional data, such as spectra, remains particularly challenging. A recent survey [10] analyzed XAI applications in the RS field. The most frequently employed method is SHAP [14], primarily due to two reasons: its robust theoretical foundation and the availability of a user-friendly Python library. However, SHAP has a very high computational complexity that make unfeasible the exact computation in the context of RS where typical problems have thousand of features. To address this limitation, the majority of SHAP-based approaches applied to RS [3,9,23,25] implement a feature selection or reduction strategy before classification and explanation, thereby limiting the effectiveness of the results. Others Explainability methods include approaches specifically designed for Neural Networks (NN), with the two most prominent being grad-CAM [15,19,21] and Integrated Gradients [24]. These approaches require the use of a specific model, and sometimes, a specific architectural configuration of this model and this limits their applicability.

[3] `https://eur-lex.europa.eu/eli/reg/2024/1689/oj`

As mentioned before, applying SHAP-based methods to high-dimensional data is challenging due to the computational complexity. In spectral data, the challenge is further amplified by the fact that feature positions correspond to specific Raman-shifts, meaning that attributions must account for the continuous and structured nature of the spectrum, rather than just treating features as independent values. In this work, we propose a novel game formulation of SHAP based on cooperative games with coalition structure [2], which is capable of accounting for the Raman-based structure of spectral data. The objective of our method is two-fold: firstly, it allows the efficient application of SHAP-based methods directly to raw spectra without the need for feature selection or reduction that would compromise their structure; secondly, it provides results at the level of Raman-shift intervals, minimizing the impact of noise inherent in spectral data. Indeed, it is common for molecules and basic components of the spectrum to be associated with a Raman-shift interval, rather than a single feature, as for example, peaks in a fixed position. The proposed method has been validated on two real-world datasets: one for the identification of Covid-19 [5] and the other one for the diagnosis of Parkinson and Alzheimer [6]. Our method has been compared with several state-of-the-art approaches, namely SHAP [14], LIME [20] and grad-CAM [21].

Since one of the main challenges in XAI is the lack of a standardized evaluation procedure, we propose an evaluation framework that integrates both quantitative methodologies using state-of-the-art metrics, alongside qualitative assessments that compare XAI results with domain knowledge. The computational experiments suggest that the novel proposed method is more effective than the baseline in identifying sensitive regions and assigning them a reliable importance score.

2 SHAP and Quotient Game

The proposed approach was inspired by SHAP [14], which is based on game theory. Specifically, it treats the features of a machine learning problem as players in a transferable utility (TU) cooperative game and solves the game using the *Shapley Value* [22]. In a TU cooperative game, players work together to maximize a collective reward, and a solution corresponds to an array that assigns each player their share of the reward. The Shapley Value provides a 'fair' solution by assigning each player his weighted average marginal contribution among all possible coalitions.

A key challenge in applying the Shapley Value is its computational complexity, as it requires enumerating all possible coalitions, resulting in 2^n combinations, where n is the number of players (features). The current literature addresses this problem by sampling a subset of coalitions. While this approach reduces complexity, it comes at the cost of slightly reduced robustness and introduces another challenge: sampling is effective for problems with few features, but as the number of features increases, each player appears in too few relevant coalitions to ensure a reliable Shapley Value estimation. To address this limita-

tion, we apply a different game formulation based on cooperative game theory with a coalition structure [2].

2.1 SHAP

The idea behind SHAP is to treat the features of an ML problem as the players in a cooperative game with transferable utility. SHAP is a *local* method, which aims to explain a single instance in the dataset at a time. Given a dataset D and one instance $X \in D$, SHAP seeks to explain the instance X. Each instance X is composed of n features denoted as $x_1, \ldots, x_n$, which are treated as players of a cooperative game, and we denote $N = \{1, \ldots, n\}$. The objective of SHAP is to assign a value ϕ_i to each feature x_i, which represents the importance of that feature in contributing to the final classification. The model we want to explain is represented as f and its output as $f(X)$. The value $v(S)$ of each coalition $S \subseteq N$ of features is defined as $f_S(X) - \overline{f(X)}$, where $f_S(X)$ represents the model output evaluated considering only the features of coalition S and $\overline{f(X)}$ is the baseline output of a random classifier. Considering that the ML model is fixed and already trained, the removal of a feature is not feasible; for this reason, different methods to simulate the absence of a feature were discussed in the literature [7]. The importance of each feature x_i is then determined by calculating its Shapley Value given by

$$\phi_i(v) = \sum_{S \subseteq N \setminus i} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [v(S \cup i) - v(S)]. \tag{1}$$

2.2 Cooperative Game with Partitions: Quotient Game

The concept of Quotient Game applies to TU cooperative games adhering to some coalition structure. A game with a coalition structure is a game in which cooperation between some players is assumed, as a result, these players cannot be separated between different coalitions. In game theory literature there are multiple ways to deal with this category of game, the most naive approach is by not considering the partition assumption and solving the game with traditional solution concepts; other possibilities modify the original solution concepts' formulations, by encountering for the coalition structure, such as the Owen Value [18]. In this work, we focus on the concept of Quotient Game, in which the solution is computed at a higher level by considering a novel game in which the partitions act as the players of the game.

Formally, let (N, v) be a TU cooperative game where the set of players N is partitioned into m subsets $N_1, \ldots, N_m$, i.e., $N = \bigcup_{i=1}^{m} N_i$ and $N_i \cap N_j = \emptyset$ for any $i, j = 1, \ldots, m$ with $i \neq j$. The quotient game is the game (M, v^p), where the new set of players $M = \{1, \ldots, m\}$ is composed of the partitions and the characteristic function is defined as $v^p(S) = v(\bigcup_{i \in S} N_i)$ for any $S \subseteq M$.

Relying on this game formulation, it is possible to compute the solution with the solution concepts available for traditional games; in this work, we focus on the Shapley Value as defined in (1).

The coalition structure may be created in several ways. Examples from the literature include the use of hierarchical clustering, mutual information, or the tree-based partition strategy. Considering the challenges associated with the automatic identification of molecular fingerprints within the spectrum, we decided to implement a straightforward strategy: the number of partitions has been fixed to eight and the input spectrum has been divided in equal-dimension intervals. This proved to be an optimal balance between the complexity of the problem and the clarity of the results.

3 Datasets and experimental Settings

3.1 Datasets

Both datasets consist of Raman spectra extracted from saliva samples. Each spectrum contains 991 measurement points corresponding to intensity values (Y-axis in Figures 2 and 3) obtained through the Raman procedure. These intensity values are measured at specific Raman shift positions (X-axis in Figures 2 and 3) in units of cm cm^{-1}. For the datasets included in this study, the Raman shift positions range approximately from 400 cm^{-1} to 1600 cm^{-1}.

The *Covid-19* dataset involved a total of 101 subjects, categorized as follows: 30 patients affected by COVID-19 (COV+), 37 subjects negative for SARS-CoV-2 with a confirmed previous episode of COVID-19 (COV−), and 34 age- and sex-matched healthy controls (CTRL). More details regarding the acquisition protocol and patient selection are available in [5]. The *Parkinson and Alzheimer* dataset (PD-AD) includes 24 patients with Parkinson's disease (PD), 10 patients with Alzheimer's disease (AD), and 33 control subjects (CTRL). More details regarding the acquisition protocol and patient selection are available in [6]. To reduce noise from specific acquisitions, 25 spectra were obtained per patient, resulting in 2408 spectra for the Covid-19 dataset and 1751 spectra for the PD-AD task.

3.2 Experimental Settings

The experimental setting is defined as follows. Each dataset was divided into training set and testing set through a *Leave One Patient Out-Cross Validation* (LOPO-CV) approach. Once the classifier has been trained the explanations have been generated on the unique patient in the testing set; the procedure was then repeated for all patients. Since the focus of this work is on model explainability we validated our method on the best performing model, based on Convolutional Neural Network (CNN). CNNs not only offer superior classification performance than traditional ML but also preserve the original spectral shape, which is essential for a full interpretation. For reproducibility purposes, we made our code and models publicly available at `https://github.com/MIND-Lab/Aime25-QuotientGame`.

4 Results

Since the XAI field still lacks a standardized and recognized evaluation procedure, we proposed a framework to evaluate and compare explanations of different approaches. To assess the correctness of explanations we employ *Faithfulness* [1] and *Remove-One-And-Retrain* (ROAR) [13]. Then, we compare the results to ensure they align with known patterns and domain knowledge.

Faithfulness is a lightweight metric for evaluating the correctness of explanations by measuring how well feature importance scores align with their impact on classification probability. Specifically, it quantifies the correlation between a feature's importance score and the drop in classification probability when that feature is removed. The metric, baed on Pearson correlation, ranges from -1 to $+1$ and is computed for each spectrum of the dataset. The average results for this metric are presented in Table 1.

The average Faithfulness value of the QG approach is higher than the other baseline methods. The distribution of the results across the all dataset is depicted in Figure 1. For the PD-AD dataset, QG shows a skewed distribution toward high values, supporting the quality of the results. In the Covid-19 dataset, although the QG distribution is smoother, it remains largely in the positive range. In contrast, Grad-CAM exhibits a bimodal distribution with peaks in both the positive and negative regions. This pattern explains its low average value and high standard deviation. The presence of two peaks suggests the existence of two subgroups within the dataset, potentially corresponding to correctly and incorrectly classified samples, which requires further investigation.

Unlike Faithfulness, the ROAR methodology evaluates the variation in model classification performance when the ML model is retrained from scratch after removing one or more features. In this study, the F1-score was monitored, and the analysis focused on the two intervals with the highest average importance. This metric is computed once for a single dataset, however to get more robust results we replicated the computation trough a 10-fold cross validation procedure. The results for this metric are presented in Table 2.

The ROAR metric evaluated on PD-AD task exhibits lower values compared to the Covid-19 task, suggesting a more accurate identification of critical regions.

	Covid-19	PD-AD
SHAP	0.00 (± 0.30)	0.01 (±0.35)
LIME	*0.12 (± 0.42)*	0.02 (±0.39)
Grad-CAM	0.07 (± 0.52)	*0.05 (± 0.39)*
QG	**0.31 (± 0.37)**	**0.45 (±0.33)**

Table 1: Average and Standard Deviation of faithfulness (with p-value < 0.05). Best and second best results are highlighted.

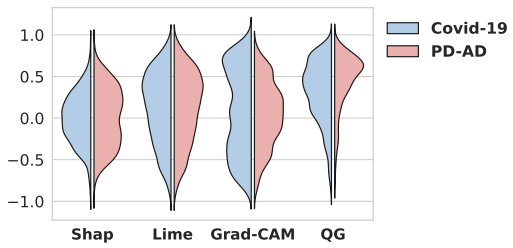
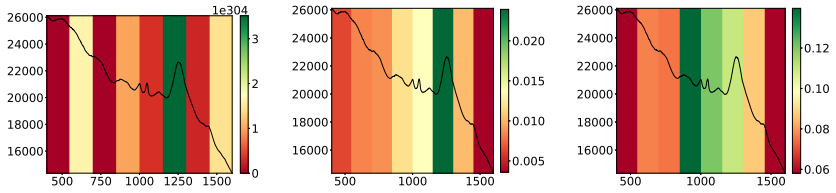


Fig. 1: Distribution of faithfulness metric.



(a) SHAP (b) Grad-CAM (c) Quotient Game

Fig. 2: Comparison between average heatmaps identified by methods SHAP, Grad-CAM and Quotient Game for Covid-19 task.

We also observe that Grad-CAM outperforms the proposed Quotient Game on the Covid-19 dataset only in the best-performing interval, whereas the Quotient Game outperforms Grad-CAM in all other cases.

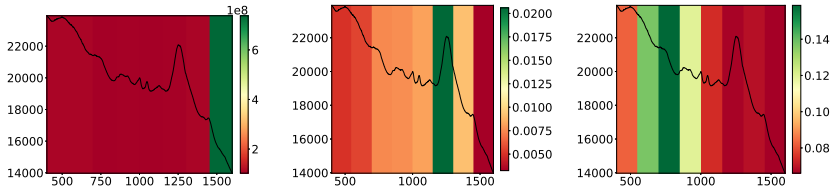
Since a standard evaluation procedure is not yet well-established, we compared our findings with known patterns and domain knowledge from the scientific literature. We found that according prior studies [5,6], domain experts expect to find the most critical peaks at approximately *1048* cm^{-1} and *1126* cm^{-1} for Covid-19, and at *750* cm^{-1} for Parkinson and Alzheimer.

In the Covid-19 dataset, Grad-CAM and SHAP identified the most important area in the interval between ∼1015 and ∼1138 cm^{-1}, which closely aligns with the critical region identified in the literature. In contrast, QG highlighted an area below 1000 cm^{-1} as the most important but identified the second and third most relevant intervals between 892 cm^{-1} and 1138 cm^{-1}, which include the sensitive intervals. The average spectrum, along with the heatmap representing the importance score for the Covid-19 task, is shown in Figure 2.

The graphical results(reported in Figure 3) for PD-AD classification shows that QG identified the two most relevant intervals between 646 cm^{-1} and 892 cm^{-1}, which perfectly match the critical peak at 750 cm^{-1}. While other methods failed to identify this relevant region. On one side, SHAP identifies as relevant the last portion of the spectrum, which is known to be dominated by noise.

	Covid Dataset		PD-AD Dataset	
	Best Interval	Second Best Interval	Best Interval	Second Best Interval
SHAP	0.84 (±0.13)	0.82 (±0.13)	0.67 (±0.21)	*0,67 (±0.20)*
LIME	*0.80 (±0.14)*	0.83 (±0.12)	*0.64 (±0.24)*	0.67 (±0.24)
Grad-CAM	**0.77 (±0.14)**	*0.80 (±0.14)*	0.70 (±0.22)	0.67 (±0.21)
Quotient Game	0.85 (±0.11)	**0.79 (±0.14)**	**0.62 (± 0.24)**	**0.65 (± 0.22)**

Table 2: Average and Standard Deviation value for ROAR, resulted from the 10-fold cross validation procedure. The first two best intervals for average importance score have been considered. Note that for this metric lower values are better. Best and second best results are highlighted.



(a) SHAP (b) Grad-CAM (c) Quotient Game

Fig. 3: Comparison between average heatmaps identified by methods SHAP, Grad-CAM and Quotient Game for Parkinson and Alzheimer task.

On the other, Grad-CAM identifies as relevant an interval near 1000 cm^{-1} that does not align with expert expectations. For both tasks, we reported the global results, as no significant differences were observed across classes. This is likely due to discriminative features being concentrated in specific spectral regions, with similar key peaks across patients from different classes.

The qualitative analysis supports the findings of the quantitative analysis, particularly confirming the results obtained using the ROAR metric. In fact, when considering this metric, the best results are achieved using the interval identified by Grad-CAM for the Covid-19 dataset, and the one identified by Quotient Game for the PD-AD dataset. These results also suggest that the classification model correctly learned to distinguish between different classes of patients by leveraging the information contained in appropriate spectral intervals.

5 Conclusions

In this work, we presented a novel approach for applying eXplainable Artificial Intelligence to spectral data characterized by a high number of features. Our method leverages cooperative game theory and the concept of the Quotient Game to compute importance scores that extend beyond individual feature levels.

To validate this approach, we considered two case studies in the domain of Raman Spectroscopy for automatic diagnosis. The first case study focused on Covid-19 detection, while the second aimed at identifying Parkinson's and Alzheimer's diseases. Both studies utilized spectra extracted from saliva samples, which are noisier compared to more invasive sample types. To evaluate the Quotient Game approach and compare it with the state-of-the-art, we proposed a combined qualitative and quantitative analysis framework. Computational experiments demonstrated that our approach is robust in identifying sensitive Raman-shift intervals associated with specific molecules. Additionally, qualitative analysis revealed interesting connections between domain knowledge and XAI results.

Future work includes applying refined optimization-based approaches to identify intervals in Raman spectra and exploring solution concepts from game theory

literature for games with coalition structure, such as the Owen Value [18] and the Banzhaf-Owen Index [17]. Additionally, to validate the proposed method in high-dimensional contexts, it is essential to extend its application to 2D data.

Acknowledgements

This work has been supported by Fondazione Regionale per la Ricerca Biomedica, project CORSAI ID 383 - JTC 2021 ERA PerMed, GA 779282 - CUP: H45E22000030006. MP was partially supported by the MUR under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy.

References

1. Alvarez-Melis, D., Jaakkola, T.S.: Towards robust interpretability with self-explaining neural networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. p. 7786–7795 (2018)
2. Aumann, R.J., Dreze, J.H.: Cooperative games with coalition structures. Int. J. Game Theory **3**(4), 217–237 (1974)
3. Bellantuono, L., Tommasi, R., Pantaleo, E., Verri, M., Amoroso, N., Crucitti, P., Di Gioacchino, M., Longo, F., Monaco, A., Naciu, A.M., Palermo, A., Taffon, C., Tangaro, S., Crescenzi, A., Sodo, A., Bellotti, R.: An eXplainable Artificial Intelligence analysis of Raman spectra for thyroid cancer diagnosis. Scientific Reports **13**, 16590 (2023)
4. Butler, H.J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., Esmonde-White, K., Fullwood, N.J., Gardner, B., Martin-Hirsch, P.L., Walsh, M.J., McAinsh, M.R., Stone, N., Martin, F.L.: Using Raman spectroscopy to characterize biological materials. Nature Protocols **11**(4), 664–687 (2016)
5. Carlomagno, C., Bertazioli, D., Gualerzi, A., Picciolini, S., Banfi, P.I., Lax, A., Messina, E., Navarro, J., Bianchi, L., Caronni, A., Marenco, F., Monteleone, S., Arienti, C., Bedoni, M.: COVID-19 salivary Raman fingerprint: innovative approach for the detection of current and past SARS-CoV-2 infections. Scientific Reports **11**, 4943 (2021)
6. Carlomagno, C., Bertazioli, D., Gualerzi, A., Picciolini, S., Andrico, M., Rodà, F., Meloni, M., Banfi, P.I., Verde, F., Ticozzi, N., Silani, V., Messina, E., Bedoni, M.: Identification of the Raman Salivary Fingerprint of Parkinson's Disease Through the Spectroscopic–Computational Combinatory Approach. Frontiers in Neuroscience **15**, 704963 (2021)
7. Chen, H., Covert, I.C., Lundberg, S.M., Lee, S.I.: Algorithms to estimate Shapley value feature attributions. Nat. Mach. Intell. **5**(6), 590–601 (2023)
8. Chen, S., Zhu, S., Cui, X., Xu, W., Kong, C., Zhang, Z., Qian, W.: Identifying non-muscle-invasive and muscle-invasive bladder cancer based on blood serum surface-enhanced Raman spectroscopy. Biomedical Optics Express **10**(7), 3533 (2019)
9. Chi, J., Bu, X., Zhang, X., Wang, L., Zhang, N.: Insights into cottonseed cultivar identification using Raman spectroscopy and explainable machine learning. Agriculture **13**(4), 768 (2023)
10. Contreras, J., Bocklitz, T.: Explainable artificial intelligence for spectroscopy data: a review. Pflügers Archiv-European Journal of Physiology **477**, 603–615 (2025)

11. Deepaisarn, S., Vong, C., Perera, M.: Exploring Machine Learning Pipelines for Raman Spectral Classification of COVID-19 Samples. In: 2022 14th International Conference on Knowledge and Smart Technology (KST). vol. 9, pp. 51–56 (2022)
12. Dies, H., Raveendran, J., Escobedo, C., Docoslis, A.: Rapid identification and quantification of illicit drugs on nanodendritic surface-enhanced Raman scattering substrates. Sensors and Actuators B: Chemical **257**, 382–388 (2018)
13. Hooker, S., Erhan, D., Kindermans, P.J., Kim, B.: A benchmark for interpretability methods in deep neural networks. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. pp. 9737–9748 (2019)
14. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 4768–4777 (2017)
15. Mansouri, T.S., Wang, H., Mariotti, D., Maguire, P.: Distinguishing methane from other hydrocarbons using machine learning and atmospheric pressure plasma optical emission spectroscopy. J. Phys. D Appl. Phys. **57**(34), 345202 (2024)
16. Mohamadi Monavar, H., Afseth, N.K., Lozano, J., Alimardani, R., Omid, M., Wold, J.P.: Determining quality of caviar from Caspian Sea based on Raman spectroscopy and using artificial neural networks. Talanta **111**, 98–104 (2013)
17. Owen, G.: Modification of the Banzhaf-Coleman index for games with a priori unions. In: Power, Voting, and Voting Power, pp. 232–238. Physica-Verlag HD (1981)
18. Owen, G.: Values of Games with a Priori Unions. In: Lecture Notes in Economics and Mathematical Systems, pp. 76–88. Springer Berlin Heidelberg (1977)
19. Passos, D., Mishra, P.: An automated deep learning pipeline based on advanced optimisations for leveraging spectral classification modelling. Chemometr. Intell. Lab. Syst. **215**, 104354 (2021)
20. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 1135–1144 (2016)
21. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision. pp. 618–626 (2017)
22. Shapley, L.S.: A Value for n-Person Games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games (AM-28), Volume II, pp. 307–318. Princeton University Press, Princeton (1953)
23. Shibu, C.J., Sreedharan, S., Arun, K.M., Kesavadas, C., Sitaram, R.: Explainable artificial intelligence model to predict brain states from fNIRS signals. Front. Hum. Neurosci. **16**, 1029784 (2022)
24. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. p. 3319–3328 (2017)
25. Tideman, L.E.M., Migas, L.G., Djambazova, K.V., Patterson, N.H., Caprioli, R.M., Spraggins, J.M., Van de Plas, R.: Automated biomarker candidate discovery in imaging mass spectrometry data through spatially localized Shapley additive explanations. Anal. Chim. Acta **1177**, 338522 (2021)