Explainable Deep Multi-View Methodology for HEDP Foam Quality Assessment

Nadav Schneider
Ben-Gurion University, IAEC
Israel
nadavsch@post.bgu.ac.il

Muriel Tzdaka, Galit Strum, Guy Lazovski, Galit Bar, Gilad Oren, Raz Gvishi
SNRC
Israel
rgvishi@soreq.gov.il

Gal Oren
Technion, NRCN
Israel
galoren@cs.technion.ac.il

Abstract

Physical experiments often involve multiple imaging representations, such as X-ray scans, microscopic or spectroscopic images, and diffraction patterns. Deep learning models have been widely used for supervised analysis in these experiments. Combining these different image representations is frequently required to analyze and make a decision properly. Consequently, multi-view data has emerged – datasets where each sample is described by multiple feature representations or views from different angles, sources, or modalities. These problems are addressed with the concept of multi-view learning. Understanding the decision-making process of deep learning models is essential for reliable and credible analysis. Hence, many explainability methods have been devised recently. Nonetheless, there is a lack of proper explainability in multi-view models, which are challenging to explain due to their architectures.

In this paper, we suggest four different multi-view architectures for the vision domain, each suited to another problem, with different relations between its views, and present a methodology for explaining these models. To demonstrate the effectiveness of our methodology, we focus on the domain of High Energy Density Physics (HEDP) experiments, where multiple imaging representations are used to assess the quality of foam samples. We expand the existing dataset and apply our methodology to classify the foam samples using the suggested multi-view architectures. Through experimental results, we show an improvement by accurate architecture choice on both accuracy (78% to 84%) and AUC (83% to 93%) while presenting a trade-off between performance and explainability. Specifically, we demonstrate that our approach enables the explanation of individual one-view models through model-agnostic techniques, providing insights into the decision-making process of each view. This comprehensive understanding enhances the interpretability of the overall multi-view model. The sources of this work are available at: https://github.com/Scientific-Computing-Lab-NRCN/Multi-View-Explainability.git.

Workshop on AI for Accelerated Materials Discovery – AI4Mat, 2024

1 Introduction

1.1 Multi-View Learning in Physics

Physical experiments often involve multiple imaging representations, such as photographs, X-ray scans Gales & Bentley (2004), microscopic or spectroscopic images of various types Ge et al. (2020), and diffraction patterns Ferri et al. (2005). Deep learning models have been widely used for supervised analysis in these experiments De Bézenac et al. (2019); Das & Tesfamariam (2022); Sadowski & Baldi (2018); Rusanovsky et al. (2022). Combining these different image representations is essential for accurate classification and comprehensive analysis of complex phenomena Sleeman IV et al. (2022); Rusanovsky et al. (2023). From high-energy physics experiments requiring data from different detectors Schneider et al. (2022), to astrophysics utilizing telescopes operating at various wavelengths Reza (2021), multi-view data plays a critical role in classifying celestial objects. In material science, combining data from X-ray diffraction, electron microscopy, and spectroscopy helps characterize materials and study crystal structures Ge et al. (2020); Sumon et al. (2022); Sumon (2022). Additionally, multi-view data is crucial in medical physics for precise disease diagnosis using imaging modalities like X-ray, MRI, CT, and ultrasound Li et al. (2021); Hosseinzadeh (2022). Fluid dynamics, biophysics, quantum optics, remote sensing, and other domains also benefit from multi-view data, enabling a deeper understanding of intricate physical processes and phenomena across different scales and disciplines. Consequently, multi-view data has emerged — datasets where each sample is described by multiple feature representations or *views* from different angles, sources, or modalities. These problems are addressed with the concept of *multi-view learning* Li et al. (2016); Xu et al. (2013); Sun (2013); Zhao et al. (2017); Yan et al. (2021b).

Multi-view learning uses multiple feature sets or representations of data to improve the performance of machine learning algorithms. In the context of deep learning, there are several types of multi-view approaches, including multi-view Convolutional Neural Networks (CNNs), multi-view Auto-encoders, multi-view Generative Adversarial Networks (GANs), multi-view Graph Neural Networks (GNNs), multi-view Deep Belief Nets (DBNs), and multi-view Recurrent Neural Networks (RNNs), while each approach has its own unique advantage Yan et al. (2021a).

Multi-view CNNs aim to learn high-level feature representations by integrating information from different views of the target data Seeland et al. (2021). To correctly classify these problems, the model has to consider each view and combine them into a joint decision. There are two main fusion strategies in multi-view CNNs: (1) the one-view-one-network strategy, where each view is processed by a separate CNN, and (2) the multi-view-one-network strategy, where multiple views are fed into the same CNN Kan et al. (2016); Seeland et al. (2021). We are going to distinguish between different real-life tasks and propose different architectures based on these two strategies.

Despite the considerable success of deep learning-based multi-view approaches, they still encounter significant challenges. One such challenge is the requirement for abundant training data for each view to effectively capture diverse representations and correlations, which can be both time-consuming and costly Zhao et al. (2017). Additionally, a major limitation lies in the lack of explanations for the decisions made by the complex architectures of deep learning models Yan et al. (2021b); Joshi et al. (2021); Rahate et al. (2022); Saeed & Omlin (2023), presenting a considerable obstacle in critical applications where interpretability is crucial*.

1.2 Multi-view Model Explainability

Explainability in the context of deep learning is the research field of explaining deep learning models' decisions. An ability which is often crucial in high credibility obligated systems, like autonomous driving Zablocki et al. (2022), medical imaging Shachor et al. (2020); Bekker et al. (2016) and physics-oriented problems Roscher et al. (2020) in particular. Explaining deep learning models'

*We note that interpretable models can be easily understood by humans without any additional aids or techniques, as their decision-making process is transparent and straightforward. On the other hand, explainable models do provide explanations for their predictions, but a deeper understanding may require supplementary tools or techniques, as their internal mechanics may not be immediately apparent to humans. In the context of this work, the two terms are applicable and interchangeable in different contexts, as we are trying to interpret the multi-view model decisions in the experiments while we also present an explainability methodology to allow such interpretability.

decisions is a growing research field, and these models are no longer considered as a complete "black box" Samek et al. (2017); Hödl et al. (2023).

Nonetheless, the explainability of multi-view models remained mostly unaddressed. This is due to the challenges caused inherently by these architectures and will be discussed in section 3. While many explanation methods are proposed Joshi et al. (2021), and some even applied for multi-modals Park et al. (2018), they are limited and are not suited for any multi-view model, and specifically, to multi-view CNNs in the vision domain, especially when particular classification should be made. In experimental physics-oriented problems, multiple image representations are extracted through an experiment, and a certain classification should be made (subsection 1.1). Explainability is also critical to ensure reliable and credible analysis.

Thus, current explanations in computer vision multi-view classification models mostly rely on the two common types of explainers that are able to achieve local (such as Local interpretable Model-Agnostic Explanations (LIME) Ribeiro et al. (2016)) or global (such as SHapley Additive exPlanations (SHAP) Lundberg & Lee (2017)) interpretability for each specific view *separately* (while disregarding the actual "big picture" the multi-view model was designed to provide initially). Note that both LIME and SHAP algorithms are model-agnostic algorithms, meaning they can be applied to any machine learning model, regardless of its complexity or architecture. This is because they only consider the input examples and their corresponding predictions.

An explanation may appear in various options in the different domains. In computer vision tasks, the input image pixels are marked with relevant areas for the model. These marked areas can be represented as contributing areas for a specific class where each class has its own color and even as a heat map where a gradient of colors represents the model's focus during its prediction. These types of explanations are common both in classification and segmentation tasks. This allows users to visually understand which regions of the image are influencing the model's prediction for a particular class. These color-coded areas provide either local or global interpretability (again, per view), offering insights into the decision-making process for individual instances Schorr et al. (2021). However, when dealing with multi-view models, this approach becomes problematic. In multi-view models, the input comprises multiple feature representations or "views" from different sources or modalities. The challenge lies in combining these views into a single decision, which often involves a view-pooling step where features are blended, leading to the loss of individual view-specific information. As a result, the model's attention or focus on specific areas in each view is not easily discernible, and explanations may not accurately reflect the true contribution of each view to the final prediction. Thus, achieving explainability in multi-view models remains a challenging task due to their inherent complexities Roscher et al. (2020).

1.3 Contribution

In this paper, we suggest four different multi-view CNN architectures for the vision domain, each suited to another problem, with different relations between its views, and present a relatively simple methodology for explaining these models and interpreting the decision-making process for experimental physics. Then we apply the methodology fully to a real-world physics experiment problem of High Energy Density Physics foam quality classification and generate several explanations (Figure 1). The main contributions of this paper are as follows:

1. Proposal of four novel multi-view architectures for the vision domain, each suited to specific problem types and relations between views, leading to improved performance and interpretability in physical experiments and their decision-making processes.

2. Comprehensive trade-off analysis between performance and explainability, offering insights into the strengths and limitations of each multi-view architecture and guiding future model design and optimization.

3. Demonstrating the proposed architectures and methodology in a real-world physics experiment involving High Energy Density Physics (HEDP) foam quality classification, showcasing significant accuracy and AUC improvements compared to previous work.

4. Potential applicability of the proposed methodology and architectures to other scientific domains beyond HEDP experiments, expanding the scope of multi-view model applications in various fields.

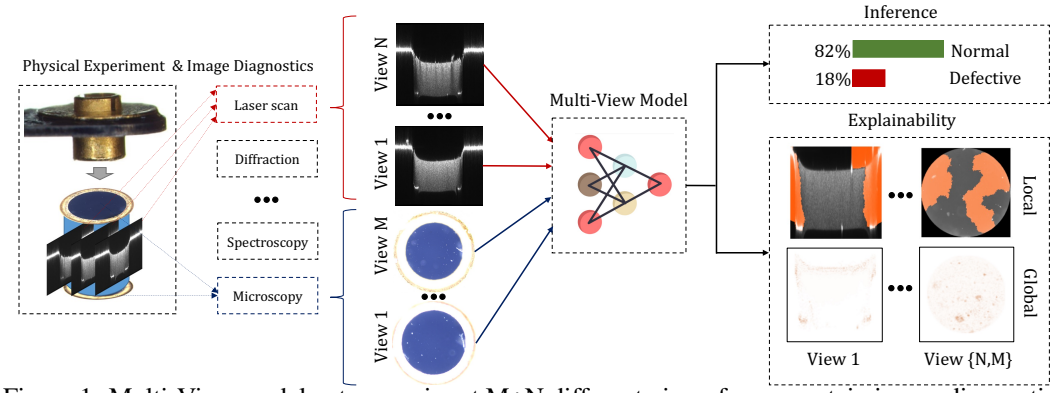
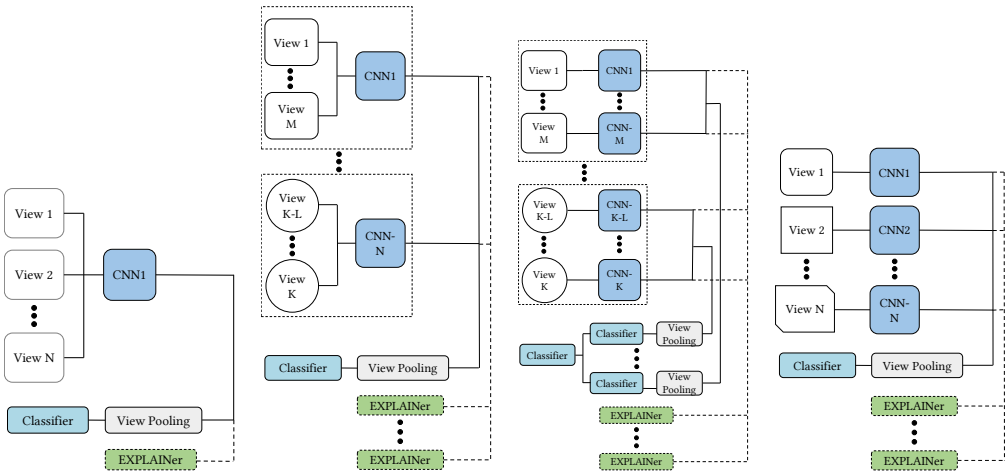


Figure 1: Multi-View model gets as an input M+N different views from a certain image diagnostic. In this example, foam quality is assessed for HEDP experiments. There are two visually different sub-group views. Then, the model outputs a classification and the explanation behind it, explainability is both local and global.



(a) Completely Similar Views (b) Similar Sub-Groups (c) Partly Similar Groups (d) Completely Different Views

Figure 2: The four different suggested multi-view architectures which are suited to different problems: (a) assumes views are similar, (b) assumes visually similar sub-groups, (c) assumes similar sub-groups but also different in a way, and (d) assumes views are completely different. EXPLAINer content is demonstrated in Figure 3. Note that CNN can be replaced with any other feature extractor.

2 Multi-View Architectures

We begin by presenting and defining four different relevant multi-view architectures, each assuming various relations between the views and, as such, fits another type of multi-view explainability (Figure 2). We observe that multi-view architecture suitability to the relevant needed explainability type directly correlates to the overall performance (see subsection 4.2), as the model understands the influence of the diverse features differently in each architecture (see section 3):

1. **Completely Similar Views (CSV)** — This approach assumes the different views are visually similar. Therefore, one feature extractor applies to all of the views under the assumption that it could efficiently extract relevant features from all of them. A pooling view through all of the outputs occurs, and the resulting vector is connected to a classifier. This architecture is mostly useful when there are multi-views of one object in one form (Figure 2a).

2. **Similar Sub-Groups (SSG)** — This approach assumes clustering views to two or more sub-groups based on similar visual attributes is possible. Each group has its own feature extractor, and therefore,

each feature extractor is optimized for its specific sub-group. Each sub-group can be visually different and still preserves an optimal extraction for each (Figure 2b).

3. **Partly Similar Groups (PSG)** — This approach assumes clustering views to two or more sub-groups based on similar visual attributes is possible. Nonetheless, views in sub-groups still have their unique features. Hence, each view has its own feature extractor. Then, each sub-group is passed through a view-pooling layer and a classifier, leading to a complete separation from other sub-groups during the classification process. Finally, each sub-group outputs a softmax decision to an additional classifier which then outputs the final decision (Figure 2c).

4. **Completely Different Views (CDV)** — This approach assumes each view has its own unique visual appearance, and thus, one feature extractor could not learn to extract relevant features from all views. Therefore, a feature extractor is applied for each view, then view pooling and a classifier (Figure 2d).

3 Multi-View Explainability

3.1 Problem Definition

Given the multi-view architectures, defined in section 2, we move to the needed explainability. Multi-view models' explainability is not trivial due to feature maps loss of their corresponding view after the feature map combination (view-pooling) process. A mixture of these feature maps is necessary to exploit all the data effectively and to reach maximal performance (subsection 1.1). Consequently, the multi-view models' explanation is inherently limited. Providing effective and comprehensive explainability for general and the suggested multi-view architectures (which are relevant in the context of physics experiments, subsection 1.1) is the problem we want to address.

3.2 Suggested Solution

We propose a methodology for explaining multi-view models given the different multi-view architectures (Figure 2). To overcome the view-pooling, in which afterward, features are blended and contribution from each view is lost, we propose the following steps (also presented formally in Algorithm 1): (1) Train the feature extractor inside the multi-view mechanism, (2) Freeze its weights and attach the trained feature extractor for each view, (3) Attach a classifier layer at the end of each view, (4) Train the resulting one-view models, and (5) Activate an explanation algorithm for each one-view model.

Steps 2-4 are represented by EXPLAINer in Figure 2 and illustrated in-depth in Figure 3. To explain each view for each multi-view architecture properly, slight modifications were performed:

1. **CSV** — One feature extractor learns to extract relevant features in a relatively accurate way. Hence, the frozen feature extractor is attached to all views, and then there is one EXPLAINer box to all of the views (Figure 2a).

2. **SSG** — Each sub-group has one feature extractor and, therefore, will be concatenated to one EXPLAINer (Figure 2b).

3. **PSG** — Views in sub-groups are visually similar but also different in a way; therefore, each view in a sub-group has its own feature extractor, and then each feature extractor is concatenated to an EXPLAINer (Figure 2c).

4. **CDV** — Each view has its own EXPLAINer box since each view has its own unique feature extractor (Figure 2d).

The proposed methodology enables relevant feature extraction from each view, sub-group, or group, depending on the applied multi-view architecture. The resulting explainability is the model's attention, meaning each marked area contributes to one of the classes, but we can not decide whom to which. By that, and contrary to general purpose multi-view classification, which does not use the devised architectures, it is possible to give an accurate explainability (either local or global) per EXPLAINer and interpretability of the entire multi-view classification decision-making visually. EXPLAINer output is a classification but is only used through model agnostic techniques to mark the contributing areas in the images. This is by extracting the features treasured inside the feature extractors (CNNs in our case).

Algorithm 1 Multi-View Explainability

```
procedure MVEXPLAIN(CNNs, classifiers)
    for CNN in CNNs do
        CNN.freeze()
        input ← Empty list
        for view in CNNs[i].input do
            cnn_output ← CNNs[i].forward(view)
            input.append(cnn_output)
        classifiers[i].fit(input)
    for classifier in classifiers do
        Apply LIME to classifier
        Apply SHAP to classifier
```

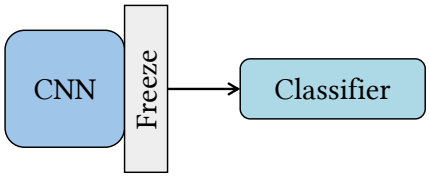


Figure 3: This scheme defined as EXPLAINer in Figure 2 and Figure 4. EXPLAINer action is freezing the concatenated feature extractor weights and training a classifier (Algorithm 1).

We note that there is a trade-off between the explainability capacity of the suggested architectures and their performance. This is due to the fine-tuned one-view classifier, which trains based on a frozen feature extractor on one of the views. This training inherently limits the model's performance. A classifier can not properly classify when it has not been trained on the other views as well. Nonetheless, the explainability is fine-tuned to the specific view, and connections are built between the frozen feature extractor and the corresponding relevant features in the view, which is then emphasized. In the following use case of HEDP experiments (section 4, we will elaborate on this trade-off in practice.

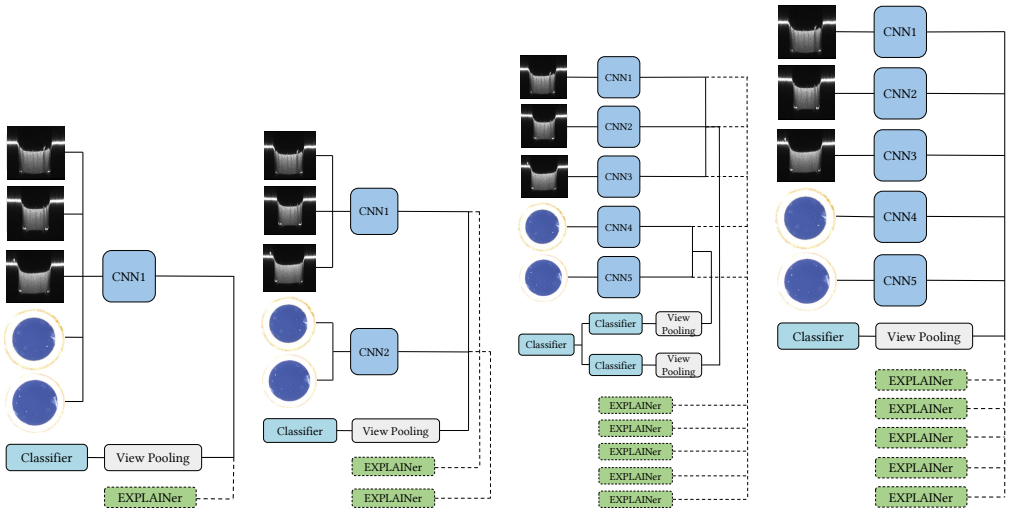
4 High Energy Density Physics Use Case

To demonstrate our methodology, we aimed to enhance both performance and explainability compared to previous research in this field by Schneider et al. (2022). In that work, a real-world problem has been addressed in the domain of High Energy Density Physics (HEDP) experiments Mussack et al. (2011); Baer (1992); Johns et al. (2021); Wu et al. (2021); Harel et al. (2020). HEDP involves a dynamic wave-front propagating inside a low-density foam, which is commonly aerogel. Aerogels are a large family of materials, generally defined as extremely low-density solids (more than 90% porosity, less than 200mg/cm^3 density) Figovsky et al. (2017); Schaefer et al. (1986).

A significant limitation of these aerogels is their tendency to suffer from cracks Sedova et al. (2015); Haj-Ali et al. (2016); Sedova et al. (2020). Therefore, it is required to characterize each sample for its quality as a foam. A single aerogel specimen includes five images (Figure 1). An expert searches for characteristics such as scratches, dirt, and dark stains in the top and bottom view images acquired by an optical microscope, implying a deep hole inside the foam. As a complementary, the profile images – acquired by laser scanning confocal microscopy (LSCM) – help confirm or refute the initial assumptions. This complicated, harsh, and exhausting work of manually classifying the foams' quality through the image set is mandatory to decide whether foams can be used in experiments (such as HEDP). Expert labels the quality levels to either Normal (Valid to use in experiments), Defective (Not valid), or Normal-Defective (contains defective characteristics, yet they are not sufficient to determine a defective classification).

The authors in Schneider et al. (2022) present a multi-view deep-learning solution for the specified problem. However, explainability for the multi-view model is absent, and its architecture is not suited to the given task. The authors use Completely Similar Views (CSV) architecture and several other one-view model configurations are suggested. Input alters accordingly: top and bottom views only, profiles only, and full group (all-views). The given explanation is partial and limited to one-view models only. An issue was presented with the use of normal-defective examples. On the one hand, due to the limited number of examples, ignoring them is not ideal. On the other hand, these examples might confuse the models. Finally, models were trained both with and without normal-defective examples.

In our work, and for clarity, we expand the previous dataset (subsection 4.1) and exclude normal-defective examples, focusing on the all-views configuration. We also demonstrate how different multi-view architecture affects learning, and how better results and interpretability are achieved when



(a) Completely Similar Views (b) Similar Sub-Groups (c) Partly Similar Groups (d) Completely Different Views

Figure 4: Demonstrating the Multi-View architectures correspondingly to Figure 2 on a real-life physical problem. In this case, there are 5 different views and 2 visually different sub-groups in the foam quality assessment for the HEDP experiments domain. (a) refers to all different views as visually similar, which is wrong. (b) does the correct splitting between the sub-groups, (c) can be correct to some degree and (d) has redundant feature extractors assuming each view is completely different than the other. EXPLAINer is connected to each feature extractor to explain its decision.

using the proper architecture. The suggested architectures applied for the given use case are presented in Figure 4.

4.1 Dataset

As previously mentioned, additional data has been added, turning the data set to a total of 123 labeled examples, each containing 5 views (615 images in total). This is in contrast to the previous work, with only 95 labeled examples. The data set was originally created for an expert's usage and not for a learning model. Repeatable methods for adequate image acquisition do not exist, this leads to a large variance between examples, and therefore, to adjust the data set into a learning model, a massive pre-processing is required. We are using the same pre-processing made in Schneider et al. (2022), and the only change is data set expansion.

4.2 Experimental Results

Since more data has been made available, normal-defective examples, which might confuse the model, have been removed. For a valid comparison, the previous model — CSV multi-view without normal-defective examples model has been retrained on the expanded data set. The data set split is a 70:30 train-to-test ratio. Overall, without the normal-defective examples, there are 59 training examples, 29 normal and 33 defectives, and 19 test examples — 10 normal and 9 defectives.

The different multi-view models were trained and executed on a 32GB Tesla V100 GPU, using Pytorch Paszke et al. (2019). The training was carried out with a batch size of 4, 150 epochs, a cross-entropy loss function, a learning rate of 0.00005, and an Adam optimizer. Accuracy and AUC were measured with the predefined settings for the different architectures. Loss and accuracy trends of each multi-view model and their performance are presented in Table 1.

4.2.1 Evaluation

In Schneider et al. (2022), all-views model without normal-defective examples and yielded on the test set accuracy of 78% and AUC of 75%. After re-training with the same settings and new additional data, the accuracy remained at 78% while its AUC raised to 83%. We show that the SSG architecture

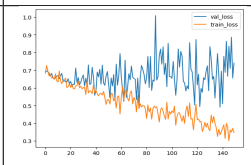
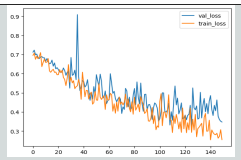
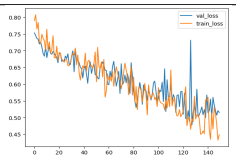
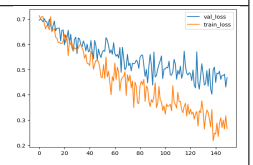
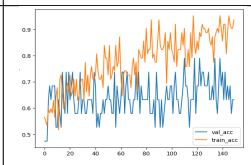
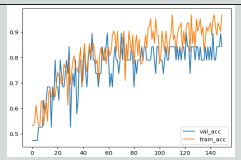
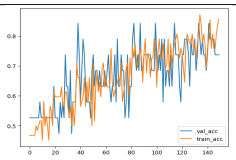
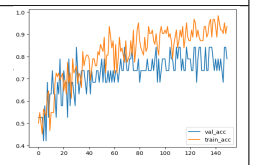
	CSV	SSG	PSG	CDV
Loss Trend				
Acc Trend				
Acc	78	**84**	78	84
AUC	83	**93**	81	84

Table 1: Learning curves (accuracy and loss as functions of epochs) of all models. The best model's cells are shadowed. These empirical results demonstrate that SSG is the best architecture for the given use case. While CSV and CDV develop overfit in epoch 60, SSG and PSG continue to converge. SSG architecture confirmed to result in the best performance with an accuracy of 84% and AUC of 93% compared to 78% and 81% by PSG.

is best suited for the given problem and therefore achieves the best accuracy – 84%, and AUC – 93%. This is compared to the previous work that used a non-optimal architecture (CSV) for the given problem. Loss and accuracy trends of the new architecture converge better and indicate better learning (Table 1). Note that AUC indicates the model's generalization capability and constitutes a crucial metric due to the tiny number of examples in the test set.

4.2.2 Explainability

We train the one-view classifiers according to Algorithm 1 and generate an explanation for each view based on the most relevant features. These models can be used to explain the prediction both locally (Figure 5) and globally (Figure 6), in a way that collectively the experiment multi-view deep-learning classification model turns interpretable for further decision-making. As mentioned, the explanation is limited to merely attention areas since the one-view models can not correctly classify based on one-view only. There is a trade-off between performance and explainability; SSG leads to the highest results but there is one CNN for multiple views, and therefore, explainability per view is damaged. CDV on the contrary has CNN for each view, therefore, explainability will be better, however, the performance is not optimal.

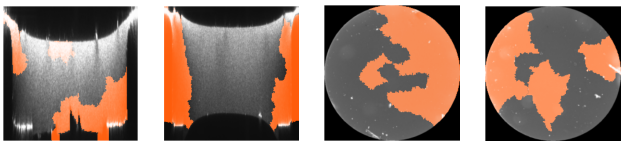


Figure 5: Multi-View local explainability with LIME. Areas of interest are marked with orange in each view. Cracks in the profiles can be watched, indicating a contribution towards defective foam decision. On the other hand, empty areas are marked in the top-bottom views, contributing towards a normal decision.

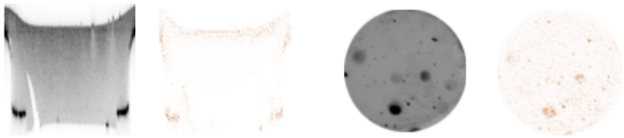


Figure 6: Multi-View global explainability of a profile and top-bottom views with SHAP. Pixels that contribute to the prediction are marked with orange.

5 Conclusions and Future Work

In this paper, we proposed a novel approach for explainability in multi-view deep learning models applied in the domain of High Energy Density Physics (HEDP) experiments with foam samples. We introduced four multi-view architectures: Completely Similar Views (CSV), Similar Sub-Groups (SSG), Partly Similar Groups (PSG), and Completely Different Views (CDV), tailored for different physics-oriented problems. Our methodology involved training one-view classifiers with frozen feature extractors for each view, enabling effective feature extraction and interpretation. The SSG architecture showed the best performance, achieving an accuracy of 84% and an AUC of 93% in classifying foam samples.

For future work, we believe that extending the methodology and architectures to other scientific domains, such as medical imaging and materials science, will validate the approach's efficiency in various contexts. Exploring hybrid architectures that combine multiple multi-view approaches offers the potential for more robust and effective models, capitalizing on the strengths of different techniques. Lastly, incorporating human-in-the-loop approaches leverages domain expertise, enhancing overall model performance and trust. Embracing these directions will promote the field of explainable multi-view models and foster their broader adoption in various scientific disciplines.

Acknowledgments

This research was supported by the Pazy Foundation and the Lynn and William Frankel Center for Computer Science. Computational support was provided by the NegevHPC project Park (2019). The authors would like to thank Hannah Zilverberg, Matan Rusanovsky, Israel Hen, and Gabi Dadush for their technical support.

References

MR Baer. A numerical study of shock wave reflections on low density foam. *Shock Waves*, 2(2): 121–124, 1992.

Alan Joseph Bekker, Moran Shalhon, Hayit Greenspan, and Jacob Goldberger. Multi-view probabilistic classification of breast microcalcifications. *IEEE Transactions on Medical Imaging*, 35(2): 645–653, 2016. doi: 10.1109/TMI.2015.2488019.

Sourav Das and Solomon Tesfamariam. State-of-the-art review of design of experiments for physics-informed deep learning. *arXiv preprint arXiv:2202.06416*, 2022.

Emmanuel De Bézenac, Arthur Pajot, and Patrick Gallinari. Deep learning for physical processes: Incorporating prior scientific knowledge. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124009, 2019.

Fabio Ferri, Davide Magatti, Alessandra Gatti, Morten Bache, Enrico Brambilla, and Luigi A Lugiato. High-resolution ghost image and ghost diffraction experiments with thermal light. *Physical review letters*, 94(18):183602, 2005.

Oleg Figovsky et al. *Green nanotechnology*. Jenny Stanford Publishing, 2017.

SG Gales and CD Bentley. Image plates as x-ray detectors in plasma physics experiments. *Review of scientific instruments*, 75(10):4001–4003, 2004.

Mengshu Ge, Fei Su, Zhicheng Zhao, and Dong Su. Deep learning analysis on microscopic imaging in materials science. *Materials Today Nano*, 11:100087, 2020.

Rami Haj-Ali et al. Mechanical characterization of aerogel materials with digital image correlation. *Microporous and Mesoporous Materials*, 226:44–52, 2016.

Re'em Harel, Matan Rusanovsky, Yehonatan Fridman, Assaf Shimony, and Gal Oren. Complete deep computer-vision methodology for investigating hydrodynamic instabilities. In *High Performance Computing: ISC High Performance 2020 International Workshops, Frankfurt, Germany, June 21–25, 2020, Revised Selected Papers 35*, pp. 61–80. Springer, 2020.

Hamidreza Hosseinzadeh. Deep multi-view feature learning for detecting covid-19 based on chest x-ray images. *Biomedical Signal Processing and Control*, 75:103595, 2022.

Stefan Hödl, William Robinson, Yoram Bachrach, Wilhelm Huck, and Tal Kachman. Explainability techniques for chemical language models, 2023.

Heather Marie Johns et al. A temperature profile diagnostic for radiation waves on omega-60. *High Energy Density Physics*, 39:100939, 2021.

Gargi Joshi, Rahee Walambe, and Ketan Kotecha. A review on explainability in multimodal deep neural nets. *IEEE Access*, 9:59800–59821, 2021.

Meina Kan et al. Multi-view deep network for cross-view classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4847–4855, 2016.

Ye Li, Junyu Chen, Justin L Brown, S Ted Treves, Xinhua Cao, Frederic H Fahey, George Sgouros, Wesley E Bolch, and Eric C Frey. Deepamo: a multi-slice, multi-view anthropomorphic model observer for visual detection tasks performed on volume images. *Journal of Medical Imaging*, 8 (4):041204–041204, 2021.

Yingming Li, Ming Yang, and Zhongfei Zhang. Multi-view representation learning: A survey from shallow methods to deep methods. *CoRR*, abs/1610.01206, 2016. URL `http://arxiv.org/abs/1610.01206`.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

K Mussack et al. Simulating hohlraum dynamics and radiation flow for pleiades experiments on nif. In *APS Division of Plasma Physics Meeting Abstracts*, volume 53, pp. GP9–124, 2011.

Dong Huk Park, Lisa Anne Hendricks, Zeynep Akata, Anna Rohrbach, Bernt Schiele, Trevor Darrell, and Marcus Rohrbach. Multimodal explanations: Justifying decisions and pointing to the evidence. *CoRR*, abs/1802.08129, 2018. URL `http://arxiv.org/abs/1802.08129`.

Rotem Industrial Park. NegevHPC Project. `https://www.negevhpc.com`, 2019. [Online].

Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Anil Rahate, Rahee Walambe, Sheela Ramanna, and Ketan Kotecha. Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions. *Information Fusion*, 81:203–239, 2022.

Moonzarin Reza. Galaxy morphology classification using automated machine learning. *Astronomy and Computing*, 37:100492, 2021.

Marco Tulio Ribeiro et al. "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, San Francisco, CA, USA,*, pp. 1135–1144, 2016.

Ribana Roscher, Bastian Bohn, Marco F Duarte, and Jochen Garcke. Explainable machine learning for scientific insights and discoveries. *Ieee Access*, 8:42200–42216, 2020.

Matan Rusanovsky, Ofer Beeri, and Gal Oren. An end-to-end computer vision methodology for quantitative metallography. *Scientific Reports*, 12(1):4776, 2022.

Matan Rusanovsky, Ofer Be'eri, Shai Avidan, and Gal Oren. Universal semantic-less texture boundary detection for microscopy (and metallography). In *NeurIPS 2023 Workshop on Machine Learning and the Physical Sciences*, 2023.

Peter Sadowski and Pierre Baldi. Deep learning in the natural sciences: applications to physics. In *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State: International Conference Commemorating the 40th Anniversary of Emmanuil Braverman's Decease, Boston, MA, USA, April 28-30, 2017, Invited Talks*, pp. 269–297. Springer, 2018.

Waddah Saeed and Christian Omlin. Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, 263:110273, 2023.

Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU Journal: ICT Discoveries - Special Issue 1 - The Impact of Artificial Intelligence (AI) on Communication Networks and Services*, 1:1–10, 10 2017.

Dale W Schaefer et al. Structure of random porous materials: silica aerogel. *Physical review letters*, 56(20):2199, 1986.

Nadav Schneider, Matan Rusanovsky, Raz Gvishi, and Gal Oren. Determining hedp foams' quality with multi-view deep learning classification. In *2022 IEEE/ACM International Workshop on Artificial Intelligence and Machine Learning for Scientific Applications (AI4S)*, pp. 19–25. IEEE, 2022.

Christian Schorr, Payman Goodarzi, Fei Chen, and Tim Dahmen. Neuroscope: An explainable ai toolbox for semantic segmentation and image classification of convolutional neural nets. *Applied Sciences*, 11(5):2199, 2021.

A Sedova et al. Reinforcing silica aerogels with tungsten disulfide nanotubes. *The Journal of Supercritical Fluids*, 106:9–15, 2015.

Anastasiya Sedova et al. Silica aerogels as hosting matrices for ws2 nanotubes and their optical characterization. *Journal of materials science*, 55(18):7612–7623, 2020.

Marco Seeland et al. Multi-view classification with convolutional neural networks. *Plos one*, 16(1), 2021.

Yaniv Shachor, Hayit Greenspan, and Jacob Goldberger. A mixture of views network with applications to multi-view medical imaging. *Neurocomputing*, 374:1–9, 2020. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2019.09.027. URL `https://www.sciencedirect.com/science/article/pii/S0925231219312895`.

William C Sleeman IV, Rishabh Kapoor, and Preetam Ghosh. Multimodal classification: Current landscape, taxonomy and future directions. *ACM Computing Surveys*, 55(7):1–31, 2022.

Borhan Uddin Sumon, Damien Muselet, Sixiang Xu, and Alain Trémeau. Multi-view learning for material classification. *Journal of Imaging*, 8(7):186, 2022.

Md Borhan Uddin Sumon. Deep learning methods for classification of photometric images of materials. Master's thesis, NTNU, 2022.

Shiliang Sun. A survey of multi-view machine learning. *Neural computing and applications*, 23(7): 2031–2038, 2013.

Yaoxing Wu et al. Dynamics of laser produced plasma from foam targets for future nanolithography devices and x-ray sources. *Scientific Reports*, 11(1):1–14, 2021.

Chang Xu et al. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.

Xiaoqiang Yan, Shizhe Hu, Yiqiao Mao, Yangdong Ye, and Hui Yu. Deep multi-view learning methods: A review. *Neurocomputing*, 448:106–129, 2021a. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2021.03.090. URL `https://www.sciencedirect.com/science/article/pii/S0925231221004768`.

Xiaoqiang Yan, Shizhe Hu, Yiqiao Mao, Yangdong Ye, and Hui Yu. Deep multi-view learning methods: A review. *Neurocomputing*, 448:106–129, 2021b.

Éloi Zablocki, Hédi Ben-Younes, Patrick Pérez, and Matthieu Cord. Explainability of deep vision-based autonomous driving systems: Review and challenges. *International Journal of Computer Vision*, 130(10):2425–2452, 2022.

Jing Zhao et al. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.