Manifold Learning for Hyperspectral Images

Fethi Harkat*†, Tiphaine Deuberet*, Guillaume Gey*, Valérie Perrier†, Kévin Polisano†

* Detection Technology Plc, A Grid, Otakaari 5A, 02150 Espoo, Finland
† Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

***Abstract*—Traditional feature extraction and projection techniques, such as Principal Component Analysis, struggle to adequately represent X-Ray Transmission (XRT) Multi-Energy (ME) images, limiting the performance of neural networks in decision-making processes. To address this issue, we propose a method that approximates the dataset topology by constructing adjacency graphs using the Uniform Manifold Approximation and Projection. This approach captures nonlinear correlations within the data, significantly improving the performance of machine learning algorithms, particularly in processing Hyperspectral Images (HSI) from X-ray transmission spectroscopy. This technique not only preserves the global structure of the data but also enhances feature separability, leading to more accurate and robust classification results.**

***Index Terms*—Hyperspectral imaging, X-ray detection, Topology, Manifold learning, Dimensionality reduction, Deep learning, Feature extraction, Computer Vision.**

I. Introduction

Recent advances in Hyperspectral Images (HSI) analysis have primarily focused on reflection spectroscopy in the visible or near-infrared light domains. However, these methods often underperform in X-ray transmission (XRT) spectroscopy [1] due to the distinct noise characteristics in this energy domain, the low-energy resolution of X-ray detectors, and poor counting statistics. Traditional methods, such as the Principal Component Analysis (PCA) [2], and Non-negative Matrix Factorization (NMF) [3] have shown limitations in characterizing these datasets, which in turn affects the efficiency of neural networks. Over the past decade, manifold learning approaches have gained increasing interest for HSI analysis [4]. These approaches typically aim to project data onto a low-dimensional space to reveal intrinsic structures, reduce noise, and improve classification performance. Studies have demonstrated the effectiveness of manifold-based techniques, such as t-distributed Stochastic Neighbor Embedding (t-SNE) [5] combined with Convolutional Neural Networks (CNN) [6], in enhancing non-linear dimensionality reduction [7] [8] and classification performance [9], [10]. In this work, we propose to apply the Uniform Manifold Approximation and Projection (UMAP) [11] [12] technique to X-ray HSI due to its ability to more accurately approximate the dataset's topology, thus better capturing nonlinear correlations and improving the performance of subsequent machine learning algorithms.

Our detector consists of an array of semi-conductor active areas, which we will refer to as "detector pixels" to avoid confusion with image processing terminology. Each detector typically contains 512 detector pixels. The data acquisition process involves using an X-ray source to illuminate an object placed on a moving belt, while a detector captures the X-rays transmitted through the object during its motion (Fig. 1). During the process, both the source and the detector remain static. The resulting images have a shape of $[H, W, C]$ where H represents the time axis, W the detector pixel axis, and C the energy band axis.

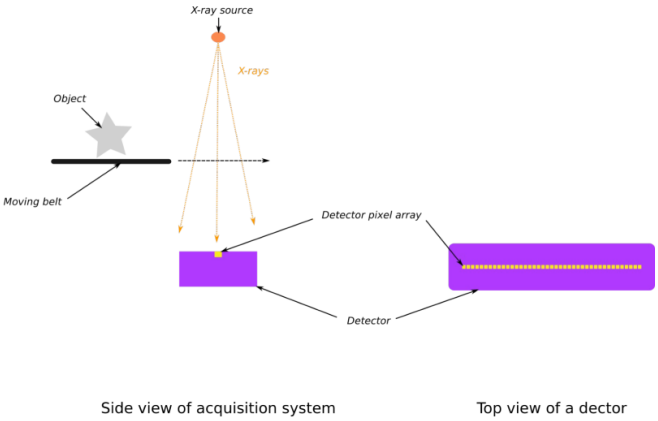

Fig. 1: Scheme of the experimental setting (left) view from aside. A top view of the detector is also given (right) to make the data collection process clearer to the reader.

This work focuses on three main objectives:

- Constructing an expressive representation of spectral data,
- Reducing the dimensionality of this representation,
- Evaluating it using deep learning.

In this paper, we investigate the application of UMAP to XRT HSI data. The proposed methodology details the approaches and tools used in our analysis. In Section II, we present the UMAP method and its application in our approach. Section III provides an overview of our experiments, and finally, in Section IV, we present our results and insights.

II. Methodology

A. UMAP:Uniform Manifold Approximation and Projection

UMAP [11] is a powerful technique for dimensionality reduction and visualization, widely used in data analysis and visualization [13], [14]. It serves as an alternative to t-SNE, with the added advantage of preserving both local and global structures in high-dimensional data. UMAP has gained

Email of the corresponding author: fethi.harkat@univ-grenoble-alpes.fr

popularity due to its ability to retain meaningful relationships between data points while remaining computationally efficient.

The method is based on manifold learning, which assumes that data lies on a lower-dimensional manifold embedded on a high-dimensional space. UMAP seeks to find a lower-dimensional representation that preserves the intrinsic structure of data. This is achieved by constructing a fuzzy topological representation of neighborhood relationships. Using fuzzy set theory [15], UMAP estimates the probability that data points are neighbors. The algorithm then minimizes the cross-entropy between pairwise probabilities in the high-dimensional space and their counterparts in the low-dimensional embedding, ensuring that the global structure is effectively preserved.

In our experiments, we used the Parametric UMAP [16] which employs a Neural Network (NN) to learn a mapping from the original data to its embedding while optimizing the same objective as traditional UMAP. This approach enables fast embedding of new data, as the trained network directly maps inputs without requiring re-optimization, making it particularly suitable for large or dynamic datasets.

B. Our approach

Our approach (Fig. 2) consists of the following steps:

1) Fit the UMAP model with images of size $[H, W, C]$ where H, W and C are respectively the width, height and number of bands of raw images, with the target dimension D, where $D \leq C$.
2) Project all images using the UMAP model onto the learnt lower-dimensional space of size D, and reshape them to their new shape $[H, W, D]$.
3) Feed the projected images into a specific CNN model to perform various machine learning tasks such as segmentation or regression.

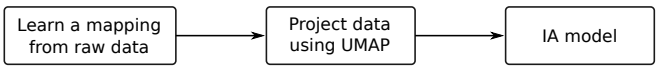

Fig. 2: Pipeline of our proposed method.

III. Numerical experiments

To evaluate the effectiveness of our proposed methodology, we assess its performance across three different datasets and tasks. Through these experiments, we aim to demonstrate the adaptability and efficiency of our approach for both segmentation and regression challenges in hyperspectral imaging (HSI). Below, we describe each dataset's characteristics and the respective machine learning task.

A. Segmentation of Cigarettes

The Cigarettes dataset was created internally and consists of HSI images of random luggage, which may contain either a pack or a carton of cigarettes (positive samples) or none (negative samples). The objective of this experiment is to segment cigarettes in the images when they are present. Binary masks of cigarettes inside the luggage have been carefully annotated for each relevant image.

Fig. 3: Comparison between raw spectral bands (top row) and UMAP-projected bands (bottom row).

Acquiring such images requires extensive material setup and complex procedures, making the process both time-consuming and costly. Since our focus is to demonstrate the effectiveness of our preprocessing approach rather than optimizing the CNN's performance, we opted for a simplified physical approximation to expand our training set. The Beer-Lambert law [17] describes photons attenuation in a medium:

$$I = I_0 e^{-\sum \mu_i L_i},$$

where I is the output intensity, I_0 is the input intensity, μ_i and L_i are respectively the attenuation coefficient and the medium thickness of $i-th$ material. A direct consequence of this formulation is that the product of two normalized images of different materials is equivalent to the normalized image of both materials superimposed, provided all images were acquired under the same conditions. We assume that the detector response is appropriately accounted for by the so-called white-normalization, commonly used in X-ray imaging. This normalization, also known as "full-flux normalization" consists of dividing acquired images by an image taken without an object. Using this approximation, we generated synthetic training images by separately capturing luggage and cigarettes, normalizing them, and summing the results.

The validity of such approach is debatable given the assumptions underlying the Beer-Lambert law. Because of this specific context, we used this technique to build 320 train images, and 30 test images. We therefore have 2 test sets. The first one is composed of 21 real images acquired with manually placed cigarette cartons inside the luggage. This set is used to measure the validity of this data generation technique against real data. The second one has been generated using images of cartons and luggage that were not used in the training set. This set to evaluate the performance of the pipeline in the context of the crude approximation we used. The results for both test sets will be presented in Section IV.

To reduce the computational cost, the input images were cropped resized from $[16, 1000, 512]$ to $[16, 152, 128]$ using 2D average pooling. The trained Parametric UMAP then projected the downsized images onto a subspace of dimension $D = 5$

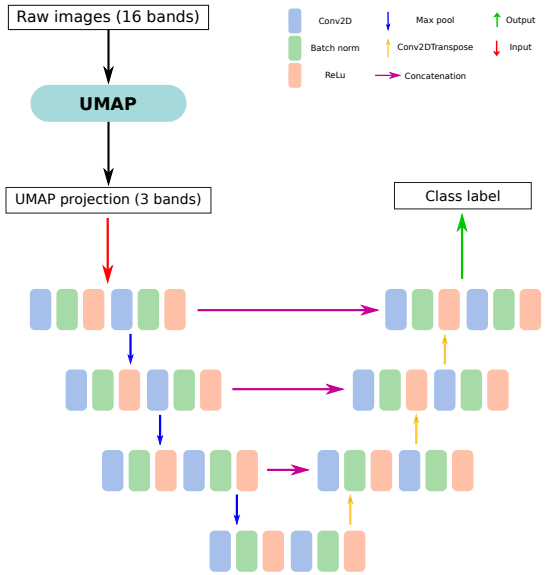

Fig. 4: The U-Net architecture consists of four downsampling blocks, each with a Double Convolution followed by Max Pooling (stride = 2), progressively increasing the number of filters from 8 to 64. The upsampling blocks include an Upsampling (scale = 2), concatenation with the corresponding output, and a Double Convolution to refine features. The final convolution reduces the output to a single prediction channel.

before feeding them into the CNN. A sample of these bands is shown in Figure 3.

For the segmentation task, a U-Net model [18] is used with Binary Cross Entropy as the loss function. The symmetric architecture (see Fig. 4) of U-Net includes an encoder, capturing context by compressing the input into feature maps, and a decoder, reconstructing the spatial dimensions to produce the segmented output. UMAP hyperparameters where left to their default values.

B. *Chemical composition of Stones*

The Stones dataset consists of individual stone images from an industrial mine, each composed of two main chemical components. Each image includes a binary mask, mass (grams), and concentration percentages of components A and B, with a third component C accounting for residual elements (see Fig. 5). The goal is to predict mass and concentrations of A and B from the HSI and binary mask. C can be derived from A and B and always has the lowest concentration, therefore it is not a relevant prediction target. As B concentration dominates, standard error averaging is unsuitable for performance evaluation, necessitating a more balanced metric. To ensure a better estimation of performances over all the targets, we therefore define the *Special Harmonic* (SH) score:

$$\text{SH-score} = M \, \frac{\prod_{i=1}^{M} S_i}{\sum_{i=1}^{M} S_i} \tag{1}$$

where M is the number of regression targets, and S_i the associated individual score. The SH-score as well as the S_i are

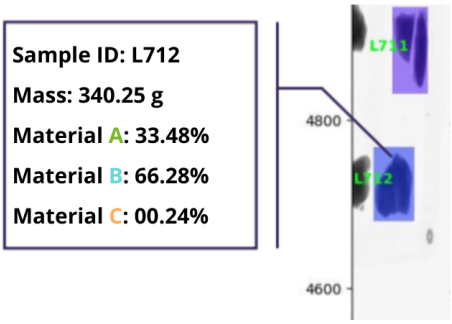

Fig. 5: Labels of Stones: each stone has an ID along with information about its total mass and material concentrations. The material C is residual elements and always has a very low concentration.

bounded to $[0, 1]$, where 1 is a perfect result. S_i are defined as follows:

$$S_i = \exp\left(-\frac{\sum_{j=1}^{N} |p_{ij} - t_{ij}|}{\sum_{j=1}^{N} t_{ij}}\right), \tag{2}$$

where p_{ij} and t_{ij} respectively stands for the prediction and the target value associated to i-*th* regression target for j-*th* sample, N being the total number of samples. The rationale behind this formulation of individual score S_i is to collapse quickly in case of poor performances, thus giving a solid estimation of performance robustness through the SH-score. We decided to name this score as such because of its close resemblance with the *Harmonic Mean* stripped from its cross-products in the denominator. To avoid any unnecessary complexity in this work, we decided to use a simple sum of weighted squared errors as the minimization objective for the stochastic gradient descent:

$$Loss = \sum_{i=1}^{M} w_i \left(\frac{p_i - t_i}{t_i}\right)^2 \tag{3}$$

where p_i, t_i and w_i respectively stands for the prediction, target and weight values associated to i-*th* regression target. The w_i have been empirically set to $\{0.5, 5, 5\}$, corresponding respectively to the mass, the material A and B targets.

The Stones dataset consists of 1,453 training images and 514 test images. To reduce computational load, 4% of the dataset is used for Parametric UMAP training, selecting a balanced subset based on mass and concentrations. The UMAP hyperparameters are set to a minimum distance of 0.5 and 50 neighbors. The minimum distance controls how closely points can be placed, while the number of neighbors determines how UMAP balances local and global structures by defining the neighborhood size considered when learning the data's underlying manifold.

We tested UMAP projections with 10, 20, and 30 bands. The UMAP-projected data was then fed into a CNN encoder followed by a decision-making layer adapted to the targets of the regression task. The CNN encoder is made of four stacks of customized Inception [19] blocks (Figure 6). For the UMAP

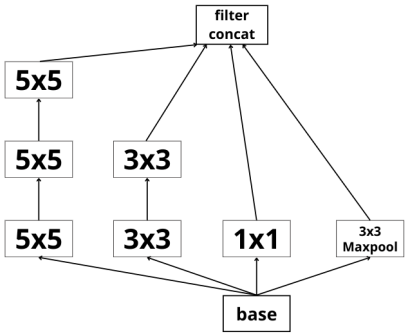

Fig. 6: Customized filters used for the regression task.

method evaluation, all the 514 test images were projected onto the respective subspace obtained from each fitted UMAP models. The results were compared to those obtained feeding the CNN with the 64 raw data bands, a 20-bands PCA, and a 20-bands NMF.

C. Indian Pines dataset

The Indian Pines dataset [20] is a hyperspectral image dataset captured by the AVIRIS sensor over the Indian Pines test site in Indiana, USA, in 1992. It consists of a single 145x145 pixel image, with each pixel containing 220 spectral bands ranging from 0.4 to 2.5 micrometers. The dataset includes 16 land-cover classes such as corn, soybeans, and forests, with ground truth labels for a subset of the image. It is commonly used for testing algorithms in hyperspectral image analysis, including classification and feature extraction, due to its high dimensionality and complexity.

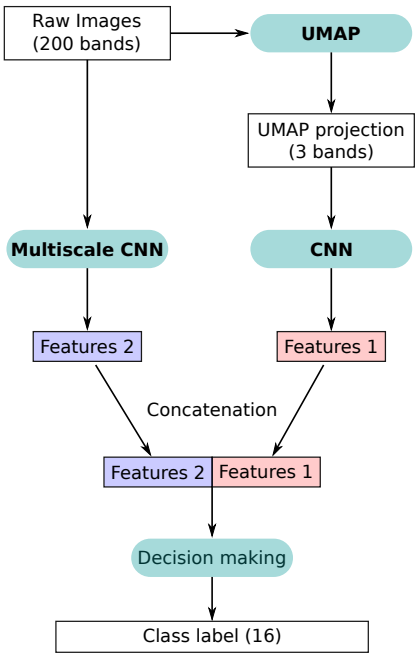

Fig. 7: Multiscale CNN model with two parallel branches and their respective inputs, the UMAP projection and the raw image.

For this experiment, we used a multiscale CNN model [21] with two branches and two inputs (see Fig. 7). Multiscale feature fusion helped leveraging diverse spatial structures and rich texture features through extensive neighborhood associations. As before, data were projected onto a low dimension representation using UMAP, then fed into the CNN architecture. Results were compared to those obtained using t-SNE and PCA as data reduction techniques. In all cases, the target dimension was set to $D = 3$. UMAP hyperparameters where left to their default values for this experiment as well.

IV. Results

A. Cigarettes dataset

As mentioned in Section III-A, we tested our approach using both a real acquisition dataset and an artifical one. We used the well-known Intersection Over Union (IoU) and Dice Score as segmentation metrics [22]. Results are summarized in Table I.

Data	Input Features	Bands	IoU	Dice Score
Synthetic Dataset	PCA	5	0.41	0.43
	NMF	5	0.22	0.33
	Our method	**5**	**0.83**	**0.90**
	Raw Data	16	0.76	0.86
Real Dataset	PCA	5	0.18	0.30
	NMF	5	0.20	0.31
	Our method	**5**	**0.32**	**0.49**
	Raw Data	16	0.23	0.37

TABLE I: Scores of the Cigarettes the segmentation task.

UMAP significantly outperforms PCA and NMF for this HSI dataset in both metrics, despite the low number of bands used. We also observe a significant performance gap between the synthetic and real datasets, which results from the very crude approximation we used to compensate for our lack of data. Even though this outcome was expected, we observe that the ranking of results remains consistent.

B. Stones dataset

Feature	**Bands**	**SH-Score**	$\mathbf{S}_A$	$\mathbf{S}_B$
Our method	10	0.65	0.57	0.97
Our method	**20**	**0.68**	**0.61**	**0.97**
Our method	30	0.63	0.55	0.97
PCA	20	0.61	0.54	0.97
NMF	20	0.60	0.51	0.97
RAW	64	0.61	0.52	0.97

TABLE II: Scores for the regression task on the Stones dataset.

The results of the experiment described in Section III-B is summarized in Table II. First, as we said earlier, the concentration of material B is dominating in our stones. This is why its associated score S_B (eq. (2)) is close to the maximum possible value, and identical in all our experiments. The score S_A is therefore more relevant to determine the success or failure of our approach. Second, those results illustrate the relevance of our definition of the SH-score (eq. (1)) as it gives values closer to the poorest score than a traditional mean,

thus giving better insights on the global performance of the pipeline. Last, we see that once again the UMAP approach gives the best results by a significant margin for a given number of bands.

C. Indian pines dataset

The results obtained after mixing the raw images and the UMAP features are shown Fig. 8, illustrating the accuracy of different strategies across epochs. The UMAP method outperforms other approaches, including the t-SNE which is also a manifold learning-based method. In addition to its accuracy advantage, UMAP also offers computational benefits compared to t-SNE, which has been too costly in term of running time and ressources to be applied to the other two datasets.

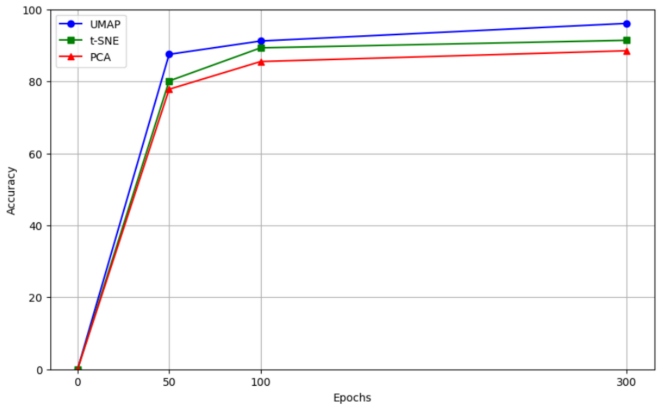

Fig. 8: Accuracy evolution during CNN training for Indian Pines dataset.

V. Conclusion

We saw that UMAP-based approaches overperform traditional linear approaches in all our experiments, like PCA and NMF. This is a result to be expected for X-ray transmission spectroscopy, where the energy bands are known to exhibit substantial non-linear behaviour, much stronger than in optical domain hyperspectral imaging. By approximating dataset topology through adjacent graph construction, our approach captures at least partially non-linear correlations in the data, enhancing the performance of machine learning models. It should be noted that the method perform very well as it is, hyperparameters tuning being necessary for only one experiment. By showing that non-linear correlations between energy bands have to be accounted for in data processing for this kind of images, we hope this contribution will stimulate further researches in toplogical data analysis for improved decision-making and more accurate data representation in complex applications.

References

[1] X. Ou, X. Chen, X. Xu, L. Xie, X. Chen, Z. Hong, H. Bai, X. Liu, Q. Chen, L. Li *et al.*, "Recent development in x-ray imaging technology: Future and challenges," *Research*, 2021.

[2] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 11, pp. 559–572, 1901.

[3] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, vol. 13, 2000.

[4] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 55–66, 2014.

[5] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[6] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[7] C. O. S. Sorzano, J. Vargas, and A. P. Montano, "A survey of dimensionality reduction techniques," *arXiv preprint arXiv:1403.2877*, 2014.

[8] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2663–2693, 2022.

[9] C. Bachmann, T. Ainsworth, and R. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.

[10] L. Gao, D. Gu, L. Zhuang, J. Ren, D. Yang, and B. Zhang, "Combining t-distributed stochastic neighbor embedding with convolutional neural networks for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1368–1372, 2020.

[11] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," Sep. 2020, arXiv:1802.03426 [cs, stat]. [Online]. Available: http://arxiv.org/abs/1802.03426

[12] J. Healy and L. McInnes, "Uniform manifold approximation and projection," *Nature Reviews Methods Primers*, vol. 4, no. 1, p. 82, 2024.

[13] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, "Dimensionality reduction for visualizing single-cell data using UMAP," *Nature biotechnology*, vol. 37, no. 1, pp. 38–44, 2019.

[14] A. Diaz-Papkovich, L. Anderson-Trocmé, and S. Gravel, "A review of UMAP in population genetics," *Journal of Human Genetics*, vol. 66, no. 1, pp. 85–91, 2021.

[15] H.-J. Zimmermann, *Fuzzy set theory—and its applications*. Springer Science & Business Media, 2011.

[16] T. Sainburg, L. McInnes, and T. Q. Gentner, "Parametric UMAP embeddings for representation and semisupervised learning," *Neural Computation*, vol. 33, no. 11, pp. 2881–2907, 2021.

[17] F. Miller, A. Vandome, and J. McBrewster, *Beer-Lambert Law*. VDM Publishing, 2009. [Online]. Available: https://books.google.fr/books?id=XO-DQgAACAAJ

[18] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of u-net," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[20] P. University, "Indian pines test site, nw tippecanoe county, indiana," 1992, hyperspectral image data collected by AVIRIS sensor, available through Purdue University's MultiSpec software.

[21] S. Li, Y. Liu, X. Sui, C. Chen, G. Tjio, D. S. W. Ting, and R. S. M. Goh, "Multi-instance multi-scale CNN for medical image classification," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22*. Springer, 2019, pp. 531–539.

[22] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022.