Nanoscale Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Nanoscale Adv.*, 2024, **6**, 5171

Hyperspectral enhanced imaging analysis of nanoparticles using machine learning methods

Kaeul Lim and Arezoo Ardekani *

Nanoparticle (NP)-based technologies have gained significant attention in targeted drug delivery, encompassing chemotherapies, photodynamic therapy, and immunotherapy. Hyperspectral imaging (HSI) emerges as a label-free, minimally invasive, and high-throughput technique for quantitative NP analysis. Despite its growing importance, the application of HSI to nanoparticle analysis, especially for label-free characterization and classification, remains limited. Here, we propose a novel method integrating hyperspectral imaging with a spectral noise reduction method and machine learning (ML) for robust nanoparticle classification. There are many challenges to extracting information from noisy and overlapping particles in HSI data. To surmount these challenges, we propose a spectral angle matching (SAM) algorithm to effectively denoise hyperspectral datasets. Complementing this, we employ a support vector machine (SVM) algorithm for classification, leveraging preprocessed HSI data to extract unique spectral signatures. Our hyperspectral imaging classification of multiple nanoparticle types reveals distinct spectral characteristics inherent to each class. The classification accuracy reaches 99.9% for single nanoparticle types, highlighting the efficiency of our method. In the case of classifying multiple particle types, the overall accuracy also reaches 99.9%. Visualization of the NP classification map further demonstrates the efficacy of our model. The application of the SAM-SVM algorithm in hyperspectral analysis outperforms traditional SVM methods in classifying multiple samples, highlighting the potential of our nanoparticle analysis. Our findings not only address the challenges posed by noisy and overlapping particles but also demonstrate the potential of hyperspectral imaging in advancing real-time and label-free detection systems for diverse biomedical applications.

Received 8th March 2024
Accepted 15th August 2024

DOI: 10.1039/d4na00205a

rsc.li/nanoscale-advances

Introduction

Nanoparticles (NPs), structures with at least one spatial dimension less than 100 nanometers, have ubiquitous applications in advanced materials engineering and biomedical engineering. Nanoparticle characterization typically focuses on size, morphology, and surface charge using microscopic techniques. The demand for a novel and precise characterization method is crucial, especially for distinguishing properties among nanoparticles with similar size and shape. Spectral imaging technologies have become increasingly attractive due to their noninvasive nature. In particular, Raman spectroscopy (RS) enables to probe very small quantities of materials with high spatial resolution at the micro- and nanoscales from the spectral features.[1–4] Although applications of RS in macrolevel diagnosis have been attempted, it is very time-consuming to scan a large area of interest, as scan windows are small and must be repeated. Thus, in general, using RS as a mapping tool is limited[5,6] in analyzing nanoparticle systems due to slow acquisition times. In contrast, hyperspectral imaging (HSI) techniques have emerged as smart analytical tools by integrating both spectroscopic and imaging techniques into one system to cope with the increasing demand to attain both spectral and spatial information.[7–9] This holds many potential advantages for region identification on a large scale, given that it scans large areas quickly.

HSI can capture a large spectral range, from the ultraviolet to the infrared, providing abundant information for each image pixel.[10,11] Hyperspectral camera sensors measure the light reflected, absorbed, and scattered by materials illuminated by a light source. HSI is a powerful tool expanding into the realm of biological and medical sciences, which holds a proven record of success in astronomy, geosciences, agriculture, and environmental monitoring, among other applications. Since HSI was originally developed for remote sensing and space applications,[12,13] machine learning-based classification for HSI is mostly focused on remote sensing data;[14–18] only a few studies investigate pharmaceutical applications at the nanoscale.[19–22] Combining the spatial-scanning hyperspectral imaging methodology with dark field microscopy is considered highly advantageous for optical studies of nanoscale materials. As an emerging imaging modality for medical applications, HSI offers great potential for classifying nanoparticles without the need for labeling.[7,23–25]

School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, USA. E-mail: ardekani@purdue.edu

HSI yields a hyperdata cube with continuous spectral and spatial information in one measurement, allowing non-contact sensing. When the HSI system scans a sample, it provides higher spectral resolution and more continuity between spectral bands than traditional multispectral fluorescence microscopes.[8] The spectral signature is the consequence of molecular absorption and particle scattering, allowing to distinguish between materials with different characteristics. In problems with an unknown target spectrum, the continuous spectral data is enough to reveal each sample's unique spectral signature. This allows distinguishing spatially and spectrally overlapping components in nanoparticle samples.[26,27] Moreover, this method is sensitive to subtle spectral changes, ensuring discrimination of chemical or biological entities.

HSIs often suffer from various types of noise, such as random noise, stripe noise, and dead pixels.[13,28,29] These issues can potentially hinder classification model development and yield misleading results. Therefore, it is critical to address these challenges and ensure the availability of high-quality data for the subsequent analysis.[30,31] Various preprocessing methods are commonly applied to hyperspectral data to improve the performance of the classification model. One of the most common preprocessing techniques in signal and image processing is the application of spectral range normalization, such as Min–max scaling and Standard Normal Variate (SNV) normalization. So far, there is no preprocessing method applied to the nanoscale. This paper proposes a modified spectral angle mapper (SAM) based image preprocessing method to denoise hyperspectral images. The SAM utilizes spectral angular information of hyperspectral image data and calculates the spectral similarity between the image spectrum and the reference spectra signature. The SAM is suitable for building a fast, efficient, and universal framework to adapt to different HSI data. The reference spectrum can either be attained from a manufacturer or extracted directly from the HSI data.[32,33]

We show the effect of the HSI denoising technique as a preprocessing step for HSI classification. The image preprocessing step before machine learning-based classification can greatly enhance classification performance by extracting only representative features. Various machine learning-based methods have been developed for hyperspectral image classification based on statistical parameters such as artificial neural networks (ANNs),[34] minimum distance classifiers,[35] *k*-nearest neighbors (*k*-NN),[36] Gaussian maximum likelihood estimators,[37] convolutional neural networks (CNNs),[38] and support vector machines (SVMs).[39] Each of these methods brings unique strengths in dealing with hyperspectral data.

Particularly, SVM has been widely applied to identify features with the multiclass problem and is an effective method of statistical learning theory. The challenge of limited training samples relative to the abundance of spectral bands, coupled with the high correlation among these bands, often compromises classification accuracy. The inherent noise among spectral bands also makes hyperspectral image classification challenging. The attraction of this method lies in its ability to locate the optimal hyperplane between the class of interest and the rest of the classes. It achieves separation in a new high-dimensional feature space by considering only the training samples that lie on the edge of the class distributions, referred to as support vectors. This technique can overcome the difficulties present in classifying the limited hyperspectral data. Using kernel functions further enhances the classifier's flexibility, making it robust against outliers.

Conversely, CNNs, while favored for their automatic feature extraction from images, tend to require a large number of labeled datasets for effective generalization. Deep learning models are computationally intensive and often require significant computational resources, especially during training. In this paper, we mainly propose the effect of the preprocessing step alongside machine learning-based classification. The classification was executed using the SVM technique driven by its distinct advantages over other methods. SVM possesses effective generalization ability, making it adept at recognizing patterns and making accurate predictions on previously unseen data.[40–42]

Addressing the challenge of overfitting in small datasets, the risk of overtraining or overfitting models is a common concern. SVM effectively mitigates this risk, demonstrating resilience against overtraining even when faced with limited data points. This characteristic ensures that our classification model remains reliable, avoiding excessive adaptation to the peculiarities within the training set. This instills confidence in the accuracy of our nanoparticle classification results. SVM can minimize the risks of overtraining with small datasets and efficiently handles the computational demands of hyperspectral image analysis. This strategic choice enhances the reliability and effectiveness of our machine learning-based approach to nanoparticle classification.

In this paper, we present a novel approach aimed at reducing image noise within HSI data, focusing on its applicability in nanoparticle classification. We introduce the Spectral Angle Matching-Support Vector Machine (SAM-SVM) method, which combines spectral similarity analysis to enhance the accuracy of HSI image classification. Through quantitative comparisons with traditional SVM techniques without the preprocessing step, our study aims to improve the classification performance while ensuring time efficiency, particularly in the context of complex nanoparticle analysis. The proposed SAM-SVM method has the potential to revolutionize hyperspectral imaging for nanoscale applications, offering a rapid, label-free classification approach crucial for advancements in drug delivery and biomedical applications.

The remainder of this paper is organized as follows: the experimental setup, preprocessing procedures, machine learning model, and comprehensive analysis of findings. By addressing the unique challenges posed by nanoscale materials, our work contributes to the advancement of hyperspectral imaging for precise nanoparticle characterization, filling a crucial gap in the current state of research.

Methods

Image acquisition and preprocessing

HSI system obtains scattered light spectra from a sample through line-by-line spatial scanning, where each pixel's

information represents the spectrum at that location. The core components of a typical HSI system are shown in Fig. 1. The HSI image cube has two spatial (x,y) and one spectral (λ) dimension, which can be viewed as a stack of spectral sub-images as illustrated in Fig. 1. The HSI system consists of an illumination unit, a light-dispersive device (spectrophotometer), a camera, and a computer equipped with image acquisition software. The sensor is positioned above the target, the same as the light source. A pair of polarizers are positioned in front of the sensor and the light source to remove saturation and isolate information from the sample.

The proposed nanoparticle spectral analysis consists of HSI image acquisition, preprocessing, extraction of spectral profiles, building an ML-based classification model, and then nanoparticle classification. The simultaneous identification of multiple nanoparticles within a sample is conducted. The overall analysis process is shown in Fig. 2.

Dark-field images were recorded by using an enhanced dark-field illumination system (CytoViva, Auburn, AL) attached to the Nikon ECLIPSE Ni-E microscope. The system consisted of a CytoViva 150 dark-field condenser in place of the microscope's original condenser attached *via* a fiber optic light guide to the lamp source. A 60× oil immersion color-corrected objective (Nikon UPlanAPO fluorite, N.A. 1.35–0.55) was integral to the system. A 150 W DC-regulated halogen fiber optics light source (Dolan Jenner DC-950, Massachusetts, USA) was used, which covers a wavelength range from approximately 360 nm to 2400 nm. The hyperspectral image of the sample was acquired with a resolution of 2 nm in the wavelength window of visible near-infrared (VNIR, 400 nm–1000 nm) using a 60× objective lens in each 5.2 nm-sized image pixel.

The acquired HSI image undergoes correlation in both the spatial and spectral domains. The hyperspectral data is first preprocessed to adapt them for subsequent feature extraction. Spectral information representing the physicochemical properties of the sample is extracted directly from the segmented objects in the image, serving as the main region of interest. In most circumstances, the extracted spectral data contain noise and variability, and this variability is one of the most challenging problems in HSI data analysis. If the extracted data exhibits a low signal-to-noise ratio, preprocessing steps become imperative. The denoising process employs a combination of background subtraction and inpainting methods for both spatial and spectral enhancement. Fig. 2 shows the nanoparticle classification workflow. The image preprocessing module is illustrated in Fig. 2b–d. An example of the preprocessing result is shown in Fig. 3.

Spectral angle mapper (SAM) is a similarity measure used for HSI data, which groups samples according to a library of reference spectra.[43,44] Reference spectra can be selected from a library of reference spectra or estimated based on the particle segmentation data. Since HSI is beneficial as a label-free method, it is crucial to be able to extract reliable reference spectra from unknown samples. The SAM-based algorithm also proposes a reliable method for estimating the reference spectra and validating the method.

The SAM algorithm determines the spectral similarity between the reference spectra and test spectra by calculating the angle between the two spectra, treating them as vectors in a space with dimensions equal to the number of bands.[45–47] The spectral angle α is calculated by using:

$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{n_b} t_i r_i}{\sqrt{\sum_{i=1}^{n_b} t_i^2}\sqrt{\sum_{i=1}^{n_b} r_i^2}}\right) \quad (1)$$

where n_b represents the number of bands in the image, t represents the pixel spectrum, and r represents the reference spectrum. According to the availability of prior information, such as the public spectral library, we derive a known reference spectral profile from the public spectral library or derive

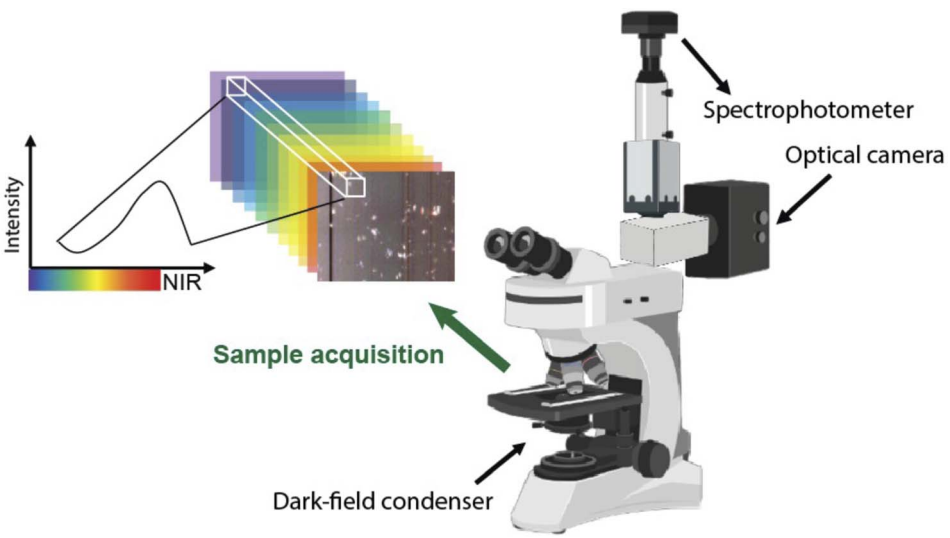

Fig. 1 The components of an HSI system (right) and HSI data cube (left).

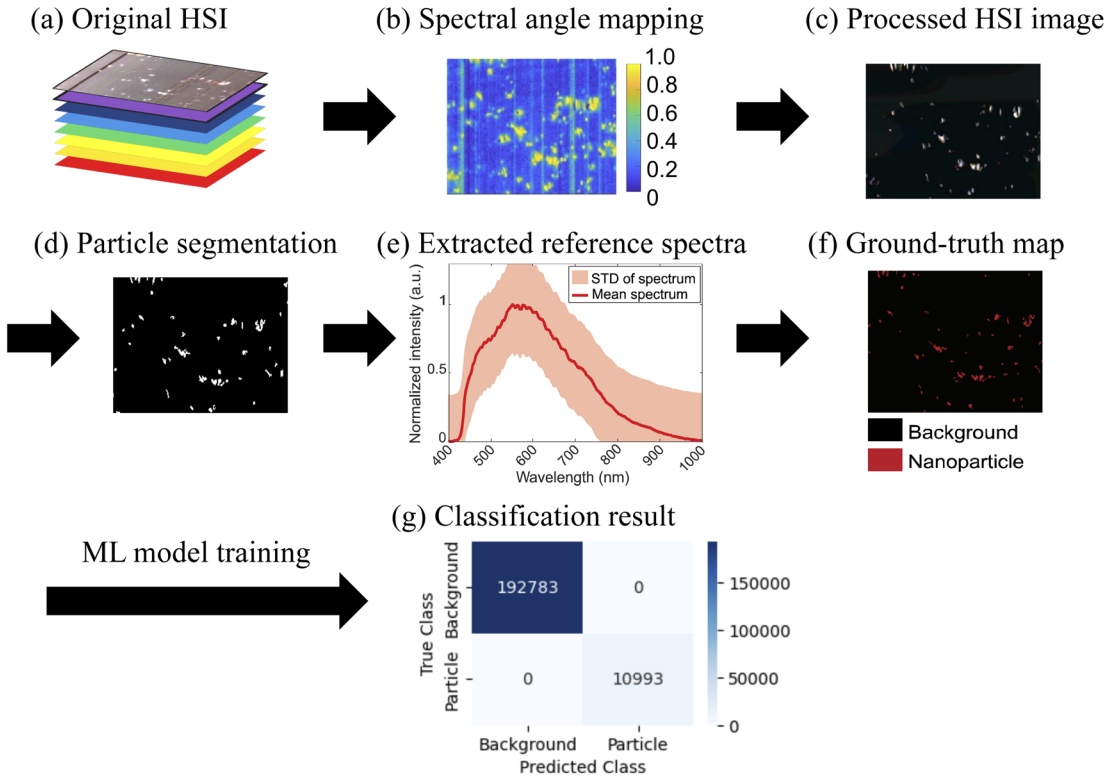

Fig. 2 Flowchart of our method (a) original hyperspectral image datacube (b) spectral angle mapping (SAM) score map (c) SAM-based pre-processed hyperspectral image (d) particle segmentation result (e) segmented particles' reference spectra (f) ground truth image for machine learning model training (g) example result of the SAM-SVM classification.

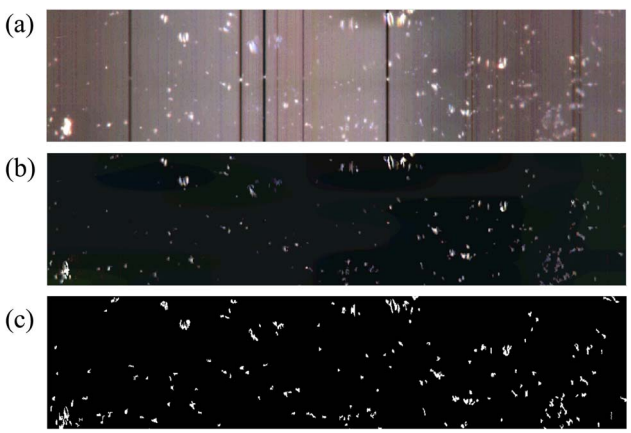

Fig. 3 Hyperspectral image processing and analysis.

a spectral signature profile from our HSI data. This adaptability is particularly valuable when dealing with unlabeled NP samples. This technique effectively addresses the challenge of blind source or unlabeled data, as it operates under the assumption of little to no prior knowledge about the components that comprise the HSI signal.

The spatial distribution of illumination intensity becomes inhomogeneous and nontrivial for a large captured area. Normalization of spectra can lead to minimizing bias from nonuniform spatial illumination, different particle types, or different light intensities. Measurement normalization greatly affects performance, especially when machine learning is used. Moreover, it is necessary in order to convert HSI measurements into reflectance ratios. In addition, in the case of NP spectra, noise affects the location of a peak instead of the height of the peak. Therefore, that peak shift might disappear with spectral filtering.[44]

The normalized reflectance spectrum is utilized for the preprocessing step. Normalizing the spectral values of every pixel P_λ at a certain wavelength (λ) to the sum of the spectral values of all pixels at all wavelengths (n) using the following equation results in a spectral value X_λ, which is independent of the illumination spectral power distribution, illumination direction, and object geometry.[48] This way, bias is removed from reflectance measurements.

$$X_\lambda = \frac{P_\lambda}{\sum_{\lambda=1}^{n} P_\lambda} \quad (2)$$

To extract spectral data, particle segmentation was carried out, the main purpose of which was to separate only nanoparticles from the background. Nanoparticles are identified and differentiated based on the spectral angle value in each pixel. The entire scanned region (1024 × 200 pixels) was designated as the region of interest (ROI) for each particle sample. The sample concentration is 0.099 mg mL^{-1}, with a sample volume of 15 μL.

The result of particle segmentation is shown in Fig. 2d. The average spectra of all segmented particle pixels within each ROI were then calculated and presented in Fig. 2e. We acquired more than three sample datasets to extract representative mean spectra of each nanoparticle shown in Table 1.

This preprocessed method is validated through a spectra comparison between the estimated average spectra in Fig. 4 and the fluorescent particle manufacture catalog's spectra. The estimated mean spectra in Fig. 4 are in agreement with the catalog spectra, specifically its spectra peak location. Yellow-green fluorescent particles show a reflectance peak at 515 nm, and europium chelate fluorescent particle shows a reflectance peak at 605 nm. In this manner, we can also obtain the representative spectral signature of an unknown sample. The average reflectance spectra of three nanoparticles were used to generate a ground-truth map for the SVM model in the ML-based classification step.

Machine learning-based classification

The classification was performed after the above-mentioned preprocessing stages. We addressed the hyperspectral image classification problem using the support vector machine technique. Among the large variety of classifiers available, SVM was selected in this study because of its (i) great generalization ability, (ii) low overtraining risk due to small datasets, and (iii) low computational load.[49–51]

SVM works by mapping data of low-dimension space into a higher-dimension space in which a separating hyperplane is constructed to realize linear classification. It separates the data into different categories by finding the best hyperplane and maximizing the distance between points. In practice, most classification problems cannot be solved by using a simple hyperplane as the decision boundary. In such a case, a more complex and elaborate decision boundary is required. By introducing a kernel function, the computational complexity will be effectively reduced. The most typical transformation function is the radial basis function (RBF) kernel.

The RBF kernel is one of the most powerful, useful, and popular kernels in the SVM family of classifiers. Unlike linear or polynomial kernels, RBF is more complex and efficient at the same time that it can combine multiple polynomial kernels of different degrees to project the nonlinearly separable data into higher dimensional space so that it can be separable using a hyperplane. The RBF kernel works by mapping the data into a high-dimensional space by finding the dot products and squares of all features in the dataset and then performing the classification using the basic idea of linear SVM. For projecting the data into a higher dimensional space, the RBF kernel uses the so-called radial basis function, which can be written as:

$$K(X_1,X_2) = \exp(-\gamma\|X_1 - X_2\|^2) \quad (3)$$

where $\|X_1 - X_2\|^2$ is the squared Euclidean distance and γ is a user-defined parameter. RBF is especially effective for addressing classification tasks due to its fast convergence speed and its capability to approximate any continuous functions with arbitrary precision. In this study, the SVM model was established to discriminate different nanoparticle types. Model performance was evaluated in terms of classification accuracy, including calibration accuracy and prediction accuracy.

As noted above, both the spectral and the spatial features influence a pixel's class label prediction. On the other hand, as the geographically close pixels tend to belong to the same class, predicting the class label of a pixel should take into account the class labels of the surrounding pixels. Hence, a good hyperspectral image classification method should consider both the spectral and spatial features together. Fig. 5, 6 and 7 show preprocessed images and ground truth images for the training of the SVM model. We split the data into train and test (80 : 20) to ensure the classification algorithm is able to generalize to unseen data well. For multi-class classification, the regularization parameters C and γ are decided. We also choose pseudo-random number generation for shuffling the data for probability estimates.[52,53] The major task of the confusion matrix, also known as the error matrix, is to compare whether the classification result matches the ground truth or not.

$$X_{i,j} = \begin{bmatrix} x11 & x12 & \dots & x1c \\ x21 & x22 & \dots & x2c \\ \vdots & \vdots & \ddots & \vdots \\ xc1 & xc2 & \dots & xcc \end{bmatrix}$$

Here, c denotes the number of categories, X_{ij} $(i, j = 1, 2, \dots, n)$ represents the number of samples, and X_{ij} represents the samples that were successfully predicted.

Each pixel in hyperspectral images contains a spectrum covering the whole spectral range of the hyperspectral imaging system. In order to better evaluate the performance of this method in the experiment, Overall accuracy (OA) is used to evaluate the classification performance of the model. Overall accuracy refers to the percentage of correctly predicted sample pixels compared to the total number of pixels. The number of correctly classified pixels is distributed along the diagonal of the confusion matrix, and the total number of pixels is equal to the total number of pixels of the ROI.

$$\mathrm{OA} = \frac{\sum_{j=1}^{n} h_{jj}}{N} \times 100\% \quad (4)$$

where h_{jj} is the number of correctly classified pixels distributed along the diagonal of the confusion matrix, N is the total number of samples, and n is the number of categories.

Table 1 The accuracy of SVM classification for the single nanoparticle type against the background

Dataset	Traditional SVM accuracy (%)	SAM-SVM accuracy (%)
Yellow green fluorescent particles	80.0	99.9
Europium chelate fluorescent particles	80.0	99.9
Flash red fluorescent particles	78.5	99.9

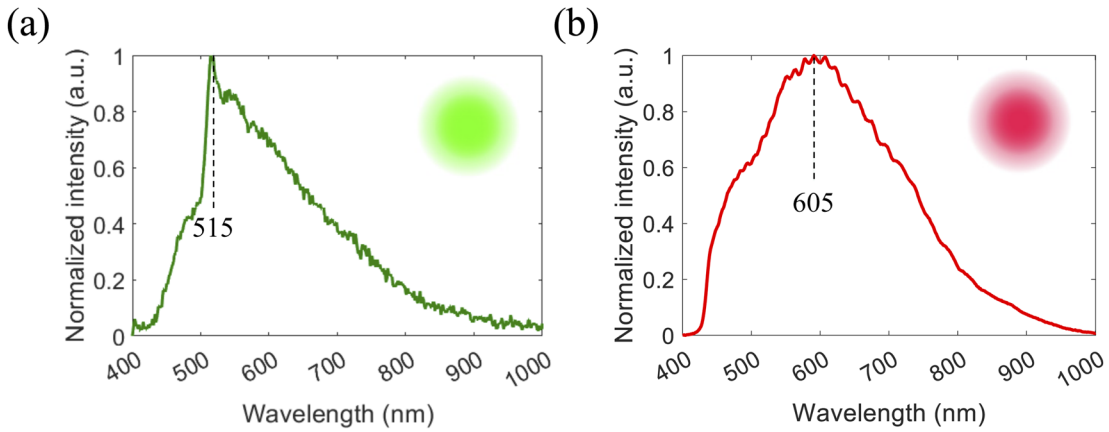

Fig. 4 Spectral validation (a) yellow-green fluorescent particles (b) europium chelate fluorescent particles.

The advantage of acquiring spectral and spatial information simultaneously provides the feasibility of predicting chemical, physical, and category information of each pixel within the samples, based on the established calibration models. The average spectra from all pixels within the sample were used for visualization in this study. Generating classification maps made it possible to visualize category information for the samples, which was beneficial for a convenient and intuitive distinction of different nanoparticle types.

Results and discussion

The SVM-based classifier was employed for each of the three related classification problems: (1) a single nanoparticle type against the background (2) two different nanoparticles, and (3) three different nanoparticles. In order to verify the advantages of our method in nanoparticle classification using hyperspectral images, our method is compared with the traditional SVM method. While our method shares the same model structure as the SVM in this paper, the distinction lies in the preprocessed HSI data and its corresponding ground truth data. The comparison of classification accuracy with the traditional standalone SVM method is shown in Table 1. We acquired three datasets for each nanoparticle type as shown in Tables 1 and 2 to train the machine learning model. The samples were imaged at 60× using reflected darkfield microscopy. The hyperspectral image consists of 476 spectral channels across a broad range of wavelengths (400–1000 nm) with 1000 × 200 pixel images at each wavelength. We also tested the accessibility of particle size by our method to maintain optimal classification accuracy. The nanoparticle samples and single particle classification results corresponding to their size are summarized in Table 2. Fig. 5 shows a classification example for a single particle type, specifically utilizing europium chelate fluorescent particles. Our method successfully identifies nanoparticles as small as 44 nm in size, exemplifying the effect of our method and its further availability.

Table 2 The summary of single nanoparticle classification samples

Particle size	500 nm	300 nm	100 nm	44 nm
Overall accuracy (%)	99.9	99.9	99.9	99.9

The main objective is to identify and classify different NPs based on the signature of spectrum profiles. In the training phase of this work, 80% of pixels were randomly selected and labeled by ground-truth data to determine the weights and biases. The other 20% of pixels were then used in the testing phase to evaluate the classification performance.

Table 1 shows the classification accuracy of two different methods for the single particle classification case. The overall accuracy of our SVM model after applying SAM is improved by over 20% compared with the traditional SVM model, which is visually shown in Fig. 5. Fig. 5 shows the input data including (a) the original HSI image before image preprocessing, (b) the ground truth image to train the machine learning methods, (c) a classification example of the traditional SVM method, (d) a classification example of our SAM-SVM method, (e) the confusion matrix of the traditional SVM result, and (f) the confusion matrix for our method. The results highlight the significance of the preprocessing step to achieve a better classification performance. The overall accuracy is 99%, which is the ratio of pixels correctly classified and the total number of testing pixels on the three datasets, corresponding to the single particle classification case. The confusion matrix analysis shows that the SAM-SVM model is performing well. The classification results in Fig. 5(d) clearly show that the present method works well in the presence of noisy points or missing data points. The classification accuracy is much higher than the traditional SVM method, as shown in Table 1.

Fig. 6 illustrates the classification results for two different nanoparticles analyzed by using our SAM-SVM method. Two classes of interest are considered, namely: NP1 (europium chalate fluorescent nanoparticle) and NP2 (yellow green fluorescent nanoparticle). The ground truth image was generated by applying the SAM method based on the reference spectral profile presented in Fig. 4. Accurate classification becomes particularly demanding in the analysis of nanoparticle mixtures

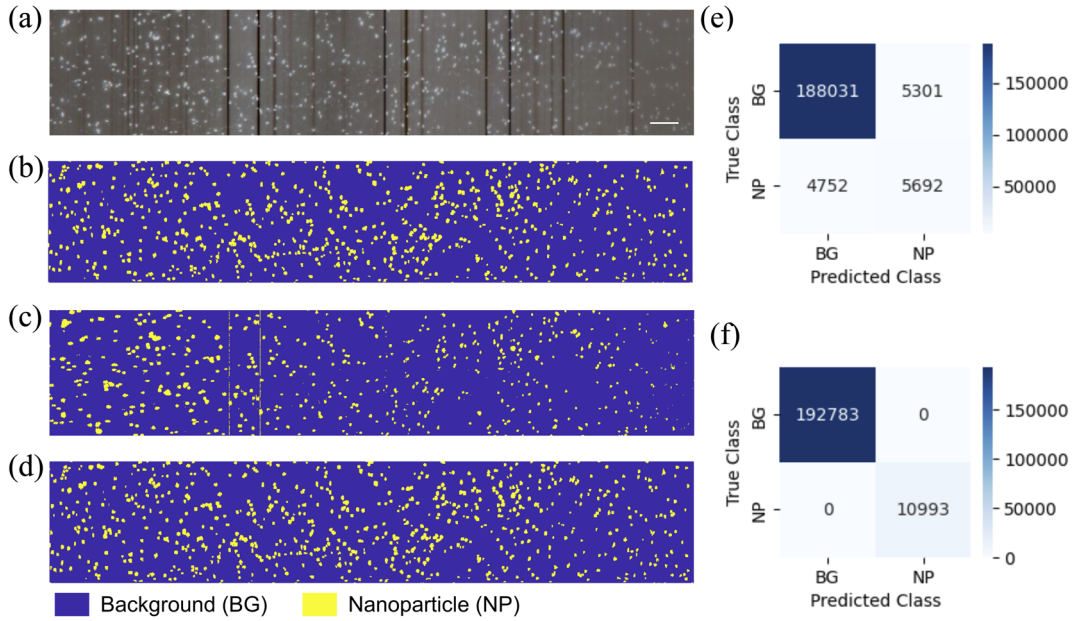

Fig. 5 Classification maps for europium chelate fluorescent particles (a) raw image (b) ground truth image (c) classification result of the traditional SVM (d) classification result of our method (e) the confusion matrix of the traditional SVM result (f) the confusion matrix for our result.

due to overlapping particles, necessitating the segmentation of individual nanoparticles. To address this issue and minimize the misclassification of overlapping particles, we employed the SVM classification method using spectral features. In Fig. 6, we can see that the particle classification prediction accuracy exceeds 90%. Only a small number of pixels were found to be misclassified, attesting to the robustness and effectiveness of our method.

Fig. 7 illustrates the classification results for three different nanoparticles. Three classes of interest are considered, namely: NP1 (europium chelate fluorescent nanoparticle), NP2 (yellow green fluorescent nanoparticle), and NP3 (plum purple fluorescent nanoparticle). We can find that the particle classification accuracy exceeds 90%.

While SVM methods offer a solid foundation for classification, they typically struggle with the high degree of spectral similarity found among nanoparticles. In specific, it was limited to classifying overlapped nanoparticles with multiple types. Similarly, approaches employing CNNs, despite their prowess in feature extraction, require substantial labeled datasets and are computationally intensive. In summary, our SAM-SVM model exhibits outstanding nanoparticle classification performance, driven by its robust preprocessing steps and optimized classification parameters. The SAM-SVM model demonstrates

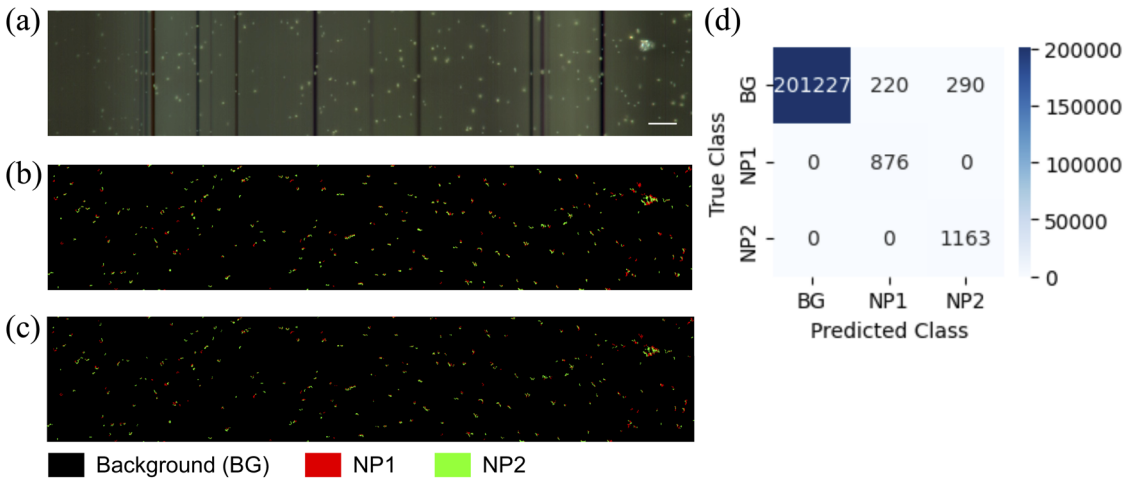

Fig. 6 Classification maps for the two fluorescent particle mixture case (a) raw image (b) ground truth image (c) classification result (d) the confusion matrix for our result.

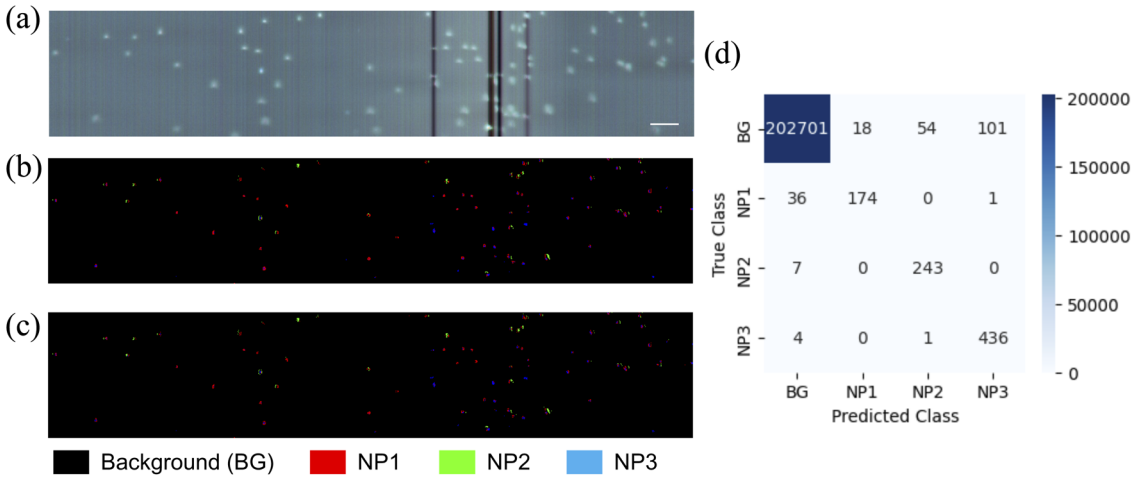

Fig. 7 Classification maps for the three fluorescent particle mixture case (a) raw image (b) ground truth image (c) classification result (d) the confusion matrix for our result.

exceptional performance, particularly in the challenging task of classifying nanoparticles with overlapping features.

Conclusion

This paper proposes a novel approach, leveraging HSI combined with ML algorithms for the precise classification of multiple nanoparticles that exhibit similar shapes and sizes. Our study explores the untapped potential of hyperspectral imaging and machine learning-based classification specifically tailored for nanoparticles. Both the spectral and spatial features of HSI are exploited to enhance the overall accuracy of image classification with the SVM. The SAM is applied to preprocess the HSI data and generate ground truth images. Our novel HSI classification method, SAM-SVM, combines spectral angle mapping and support vector machine. Compared to existing methods, our approach yields a significant improvement in overall accuracy and classification results. This model achieves the highest overall accuracy of 99.9% in both single particle classification and multiple particle classification, affirming the efficacy of our algorithm in the challenging nanoscale domain.

The principal advantages of this technique are its non-contact, non-invasive, and label-free nature. This not only establishes its practicality but also positions it as a transformative tool for the characterization of nanoscale materials. In future studies, this technique can offer a reliable and efficient method for distinguishing nanoparticles in diverse practical applications. In the pharmaceutical industry, our method can significantly enhance the development and quality control of drug delivery systems by enabling the characterization of nanoparticle carriers. Similarly, in materials science, the ability to accurately classify nanoparticles opens up new avenues for creating advanced materials with tailored properties. We aim to broaden the scope of our method to include the classification of various biological particles. This expansion has the potential to catalyze substantial progress in biomedical research, particularly in the characterization and development of nanoparticle-based therapeutics and diagnostic tools. By harnessing the capabilities of HSI and ML, our work lays a foundation for cutting-edge research at the intersection of nanotechnology and biomedical sciences, promising breakthroughs in diverse applications.

Data availability

Data for the paper titled "Hyperspectral enhanced imaging analysis of nanoparticles using machine learning methods," including hyperspectral images and ground truth maps for nanoparticle classification, are available in the HSI-Nanoparticle-Datasets repository at [https://github.com/AreZooArdekani/HSI-Nanoparticle-Datasets].

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the NSF Center for Bioanalytic Metrology (supported by the National Science Foundation under Grant IIP-1916691) for funding this project and to industry members for valuable discussions.

References

1 C. S. Kumar, *Raman Spectroscopy for Nanomaterials Characterization*, Springer Science & Business Media, 2012.
2 G. Gouadec and P. Colomban, Raman spectroscopy of nanomaterials: How spectra relate to disorder, particle size and mechanical properties, *Prog. Cryst. Growth Char. Mater.*, 2007, **53**, 1–56.
3 M. Boodaghidizaji, et al., Characterizing viral samples using machine learning for Raman and absorption spectroscopy, *MicrobiologyOpen*, 2022, **11**, e1336.

4 M. K. Maruthamuthu, A. H. Raffiee, D. M. De Oliveira, A. M. Ardekani and M. S. Verma, Raman spectra-based deep learning: A tool to identify microbial contamination, *MicrobiologyOpen*, 2020, **9**, e1122.
5 N. Kuhar, S. Sil, T. Verma and S. Umapathy, Challenges in application of Raman spectroscopy to biology and materials, *RSC Adv.*, 2018, **8**, 25888–25908.
6 R. Smith, K. L. Wright and L. Ashton, Raman spectroscopy: an evolving technique for live cell studies, *Analyst*, 2016, **141**, 3590–3600.
7 G. Lu and B. Fei, Medical hyperspectral imaging: a review, *J. Biomed. Opt.*, 2014, **19**, 010901.
8 L. Gao and R. T. Smith, Optical hyperspectral imaging in microscopy and spectroscopy–a review of data acquisition, *J. Biophotonics*, 2015, **8**, 441–456.
9 M. B. Sinclair, D. M. Haaland, J. A. Timlin and H. D. Jones, Hyperspectral confocal microscope, *Appl. Opt.*, 2006, **45**, 6283–6291.
10 G. A. Roth, S. Tahiliani, N. M. Neu-Baker and S. A. Brenner, Hyperspectral microscopy as an analytical tool for nanomaterials, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 2015, **7**, 565–579.
11 J. Wang, X. Wang, K. Zhang, K. Madani and C. Sabourin, Morphological band selection for hyperspectral imagery, *Geosci. Rem. Sens. Lett. IEEE*, 2018, **15**, 1259–1263.
12 A. F. Goetz, Three decades of hyperspectral remote sensing of the earth: A personal view, *Rem. Sens. Environ.*, 2009, **113**, S5–S16.
13 A. Plaza, et al., Recent advances in techniques for hyperspectral image processing, *Rem. Sens. Environ.*, 2009, **113**, S110–S122.
14 L. Bruzzone, M. Chi and M. Marconcini, Semisupervised support vector machines for classification of hyperspectral remote sensing images, *Hyperspectral Data Exploitation: Theory and Applications*, 2007, pp. 275–311.
15 W. Wang, S. Dou, Z. Jiang and L. Sun, A fast dense spectral–spatial convolution network framework for hyperspectral images classification, *Rem. Sens.*, 2018, **10**, 1068.
16 J. Yang, Y. Zhao, J. C.-W. Chan and C. Yi, Hyperspectral image classification using two-channel deep convolutional neural network, In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 2016, pp. 5079–5082.
17 K. Pooja, R. R. Nidamanuri and D. Mishra, Multi-scale dilated residual convolutional neural network for hyperspectral image classification, In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE, 2019, pp. 1–5.
18 X. Yang, et al., Synergistic 2d/3d convolutional neural network for hyperspectral image classification, *Rem. Sens.*, 2020, **12**, 2033.
19 E. D. SoRelle, et al., A hyperspectral method to assay the microphysiological fates of nanomaterials in histological samples, *Elife*, 2016, **5**, e16352.
20 P. Zamora-Perez, et al., Hyperspectral-enhanced dark field analysis of individual and collective photo-responsive gold–copper sulfide nanoparticles, *Nanoscale*, 2021, **13**, 13256–13272.
21 A. Yakovliev, et al., Hyperspectral multiplexed biological imaging of nanoprobes emitting in the short-wave infrared region, *Nanoscale Res. Lett.*, 2019, **14**, 1–11.
22 N. Fairbairn, A. Christofidou, A. G. Kanaras, T. A. Newman and O. L. Muskens, Hyperspectral darkfield microscopy of single hollow gold nanoparticles for biomedical applications, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4163–4168.
23 C. Juntunen, I. M. Woller, A. R. Abramczyk and Y. Sung, Deep-learning-assisted Fourier transform imaging spectroscopy for hyperspectral fluorescence imaging, *Sci. Rep.*, 2022, **12**, 1–11.
24 D. Gosavi, B. Cheatham and J. Sztuba-Solinska, Label-free detection of human coronaviruses in infected cells using enhanced darkfield hyperspectral microscopy (edhm), *J. Imaging*, 2022, **8**, 24.
25 D. T. Dicker, et al., Differentiation of normal skin and melanoma using high resolution hyperspectral imaging, *Cancer Biol. Ther.*, 2006, **5**, 1033–1038.
26 F. R. Bertani, et al., Label-free and non-invasive discrimination of hacat and melanoma cells in a co-culture model by hyperspectral confocal reflectance microscopy, *J. Biophot.*, 2016, **9**, 619–625.
27 R. Lansford, G. Bearman and S. E. Fraser, Resolution of multiple green fluorescent protein color variants and dyes using two-photon microscopy and imaging spectroscopy, *J. Biomed. Opt.*, 2001, **6**, 311–318.
28 H. Su, Q. Du, G. Chen and P. Du, Optimized hyperspectral band selection using particle swarm optimization, *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.*, 2014, **7**, 2659–2670.
29 X. Lu, Y. Wang and Y. Yuan, Graph-regularized low-rank representation for destriping of hyperspectral images, *IEEE Trans. Geosci. Rem. Sens.*, 2013, **51**, 4009–4018.
30 B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi and J. Chanussot, Noise reduction in hyperspectral imagery: Overview and application, *Rem. Sens.*, 2018, **10**, 482.
31 D. Cozzolino, P. Williams and L. Hoffman, An overview of pre-processing methods available for hyperspectral imaging applications, *Microchem. J.*, 2023, 109129.
32 J. Schwarz and K. Staenz, Adaptive threshold for spectral matching of hyperspectral data, *Can. J. Rem. Sens.*, 2001, **27**, 216–224.
33 A. P. Crosta, C. Sabine and J. V. Taranik, Hydrothermal alteration mapping at bodie, california, using AVIRIS hyperspectral data, *Rem. Sens. Environ.*, 1998, **65**, 309–319.
34 Y. Zhong and L. Zhang, An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery, *IEEE Trans. Geosci. Rem. Sens.*, 2011, **50**, 894–909.
35 W. Li, S. Prasad, E. W. Tramel, J. E. Fowler and Q. Du Decision fusion for hyperspectral image classification based on minimum-distance classifiers in thewavelet domain. In *2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP)*, IEEE, 2014, pp. 162–165.

36 B. Tu, et al., Knn-based representation of superpixels for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.*, 2018, **11**, 4032–4047.
37 D. Saqui et al., Methodology for band selection of hyperspectral images using genetic algorithms and Gaussian maximum likelihood classifier, In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, IEEE, 2016, pp. 733–738.
38 L. Fang, G. Liu, S. Li, P. Ghamisi and J. A. Benediktsson, Hyperspectral image classification with squeeze multibias network, *IEEE Trans. Geosci. Rem. Sens.*, 2018, **57**, 1291–1301.
39 Y. Chen, N. M. Nasrabadi and T. D. Tran, Hyperspectral image classification via kernel sparse representation, *IEEE Trans. Geosci. Rem. Sens.*, 2012, **51**, 217–231.
40 C. Campbell and Y. Ying, *Learning with Support Vector Machines*, Springer Nature, 2022.
41 I. Steinwart and A. Christmann, *Support Vector Machines*, Springer Science & Business Media, 2008.
42 C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., *A Practical Guide to Support Vector Classification*, 2003.
43 H. Fabelo et al., Dermatologic hyperspectral imaging system for skin cancer diagnosis assistance, In *2019 XXXIV Conference on Design of Circuits and Integrated Systems (DCIS)*, IEEE, 2019, pp. 1–6.
44 L. L. Randeberg, E. L. P. Larsen and L. O. Svaasand, Characterization of vascular structures and skin bruises using hyperspectral imaging, image analysis and diffusion theory, *J. Biophotonics*, 2010, **3**, 53–65.
45 Q. Du, Modified fisher's linear discriminant analysis for hyperspectral imagery, *Geosci. Rem. Sens. Lett. IEEE*, 2007, **4**, 503–507.
46 H. Sima, P. Liu, L. Liu, A. Mi and J. Wang, Sparse representation classification based on flexible patches sampling of superpixels for hyperspectral images, *Math. Probl Eng.*, 2018, 8264961.
47 J. Jiang, L. Huang, H. Li and L. Xiao, Hyperspectral image supervised classification via multi-view nuclear norm based 2d pca feature extraction and kernel elm, In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 2016, pp. 1496–1499.
48 G. Polder, G. Van Der Heijden, H. Van der Voet and I. Young, Measuring surface distribution of carotenes and chlorophyll in ripening tomatoes using imaging spectrometry, *Postharvest Biol. Technol.*, 2004, **34**, 117–129.
49 B. Schölkopf, A. J. Smola, F. Bach and et al., *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond*, MIT press, 2002.
50 C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4, Springer, 2006.
51 M. E. Paoletti et al., Multicore implementation of the multi-scale adaptive deep pyramid matching model for remotely sensed image classification, In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 2017, pp. 2247–2250.
52 J. Platt et al., Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 1999, 10, pp. 61–74.
53 C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.*, 2011, **2**, 1–27.