









Deep learning-based hyperspectral microscopic imaging for cholangiocarcinoma detection and classification

SIKHAKOLLI SRAVAN KUMAR,^{1,†}  OMM PRAKASH SAHOO,^{2,†} 
GAGAN MUNDADA,^{2,†}  SURESH AALA,¹  DORABABU SUDARSA,³
OM JEE PANDEY,² SUNIL CHINNADURAI,^{1,6} OSAMU MATOBA,⁴ 
INBARASAN MUNIRAJ,^{5,7}  AND ANUJ DESHPANDE^{1,8}

¹Department of Electronics and Communication Engineering, SRM University AP, Andhra Pradesh-522019, India

²Department of Electronics Engineering, Indian Institute of Technology-BHU, Varanasi, Uttar Pradesh 221005, India

³Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh 522302, India

⁴Center of Optical Scattering Image Science, Department of Systems Science, Graduate School of System Informatics, Kobe University, Rokkodai 1-1, Nada, Kobe 657-8501, Japan

⁵LiFE Lab, Department of ECE, Alliance School of Applied Engineering, Alliance University, Bengaluru, Karnataka 562106, India

⁶sunil.c@srmmap.edu.in

⁷inbarasan.m@gmail.com

⁸deshpande.anuj24@gmail.com

[†] Authors contributed equally to this work.

Abstract: Cholangiocarcinoma is one of the rarest yet most aggressive cancers that has a low 5-year survival rate (2% - 24%) and thus often requires an accurate and timely diagnosis. Hyperspectral Imaging (HSI) is a recently developed, promising spectroscopic-based non-invasive bioimaging technique that records a spatial image (x, y) together with wide spectral (λ) information. In this work, for the first time we propose to use a three-dimensional (3D)U-Net architecture for Hyperspectral microscopic imaging-based cholangiocarcinoma detection and classification. In addition to this architecture, we opted for a few preprocessing steps to achieve higher classification accuracy (CA) with minimal computational cost. Our results are compared with several standard unsupervised and supervised learning approaches to prove the efficacy of the proposed network and the preprocessing steps. For instance, we compared our results with state-of-the-art architectures, such as the Important-Aware Network (IANet), the Context Pyramid Fusion Network (CPFNet), and the semantic pixel-wise segmentation network (SegNet). We showed that our proposed architecture achieves an increased CA of 1.29% with the standard preprocessing step i.e., flat-field correction, and of 4.29% with our opted preprocessing steps.

© 2024 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Cancer is one of the most critical and devastating diseases that affect millions of lives worldwide. According to the World Health Organization, cancer is the second leading cause of death, responsible for an estimated 9.6 million deaths in the year 2018 [1]. The apprehension for cancer continues to grow as the new cases are estimated to increase by about 70% in the next two decades. These alarming statistics highlight the urgent need for comprehensive strategies to prevent, diagnose, and treat cancer effectively. Efforts in cancer research, such as early detection and improved personalized treatment options, are essential to combat this deadly disease and improve the quality of life for individuals [2]. Cholangiocarcinoma, also known as bile duct

cancer, is a rare but fatal cancer that affects the duct vessels that carry bile, i.e., a fluid that is made and released by the liver, which helps in digestion. According to the American Cancer Society, the relative 5-year survival rate for all stages of cholangiocarcinoma combined is around 24% [3]. Alarmingly, new cases of bile duct cancer also appear to be increasing with high rates of liver fluke infection and chronic liver diseases (in the form of cirrhosis and hepatitis). Traditionally, techniques such as cancer antigen 19-9 blood test [4], ultrasound [5], computed tomography (CT) [6], magnetic resonance imaging (MRI) [7], and endoscopic retrograde cholangiopancreatography (ERCP) [8] are used to identify tumours or blockages that occur at the bile ducts. In addition to these techniques, recently, hyperspectral imaging (HSI) has emerged as a promising alternative tool for accurately identifying (non-invasively) the locations of tumours within the body [9]. It was shown that HSI captures both spatial (x, y) and spectral (λ) information simultaneously. Thereafter, several studies were performed for oncological assessments [10]. Recently, HSI was also combined with optical microscopes to get the additional salient features of the image samples [11–16].

In conjunction with classical diagnostic techniques, machine learning (ML) [17] and deep learning (DL) techniques [18] have been widely applied for automatic bile duct classification problems [19]. It is known that these techniques are broadly classified into supervised and unsupervised learning algorithms. Generally, the supervised learning (SL) technique is widely preferred over the unsupervised learning (USL) technique due to its capability to achieve higher classification accuracy, which is critical for image-guided surgeries. In light of this, several investigations have been carried out for the classification of cholangiocarcinoma. For instance, Quing Li *et al.* [20] proposed the use of a deep neural network (DNN) for the classification of cholangiocarcinoma tissues imaged under a microscope. In their study, the authors used DNN to extract the key spectral features from the HSI dataset and employed a three-layer network for classification. They reported 90.5% accuracy in distinguishing cholangiocarcinoma from healthy tissues. In [21], Deng *et al.* used a residual neural network (i.e., RESNET50) to classify cholangiocarcinoma tissues using a patch-based approach. The authors reported an accuracy of 92.6%, in classifying normal and malignant tissues. Further, J Wang *et al.* [22] proposed a spatial-spectral-based hyperspectral generative adversarial network (SSHGAN) to automatically generate digital annotations of the microscopic hyperspectral images, i.e., HE-CAM5.2 stained tissues. In addition to these, a study conducted by C. Bench *et al.* [23] explored the possibility of combining Raman with infrared-based HSI for segmenting the biopsied specimens of real colon cancer and synthetic fat muscle tissues. For this work, the authors proposed to use end-to-end deep convolutional autoencoders (CAEs).

Note that most of these studies were performed either on grayscale or conventional colour (RGB) images of tissue sections. Intuitively, the spectral information contained in these image types is limited, and as a consequence, the performance of the machine and deep learning algorithms is severely restricted. As mentioned earlier, HSI overcomes this limitation by combining optical imaging with spectroscopy thereby capturing both spatial and spectral information. Thus, the hyperspectral microscopic choledochal dataset [20] was captured to provide better spatial as well as spectral information about the cholangiocarcinoma. Thereafter, in the literature, a few supervised learning-based studies have been proposed for this dataset. For instance, authors in [24] proposed an important-aware network (IANet), which used a pre-trained model (ResNet32) and a hierarchical fusion pyramid to extract multi-layer features from the microscopic hyperspectral pathology images. In the IANet model, an important-aware fusion (IAF) module was developed and implemented in conjunction with the skip connection to emphasize important spatial and semantic features. Thus, IANet was shown to have achieved a maximum classification accuracy of 88.20% [24] when compared with other state-of-the-art methods such as CPFNet [25] and SegNet [26].

Furthermore, it is known that DL networks excel in performance when combined with some preprocessing steps [27]. For instance, a study on deep learning-based adaptive filtering for noise reduction in X-ray computed tomography [28] demonstrated the significance of preprocessing steps. Similarly, a study in [29] showcased the effectiveness of deep learning in combination with filtering approaches for ground-point filtering in aerial laser scanning point clouds. Furthermore, preprocessing techniques such as illumination equalization and top-hat transformation were combined with deep learning approaches to efficiently detect the microaneurysms from colour fundus images [30]. Thus, in addition to the standard preprocessing carried out on the HSI dataset, e.g., flatfield correction, we propose to use a few additional steps, such as dimensionality reduction (DR) [31], filtering (or denoising) [32], and data normalisation [33] to achieve better classification performance. To validate our claim, we propose to use the 3D U-Net architecture [34] and note that our model architecture was enabled with skip connections. Additionally, a few bespoke preprocessing algorithms were added for accurate HSI-based microscopic cholangiocarcinoma classification. We have tested and compared the performance of our proposed approach against the state-of-the-art (SOTA) SL algorithms such as IANet, CPFNet, and SegNet and with a few standard USL algorithms, such as K-means clustering, non-negative matrix factorization (NMF) [35], Otsu thresholding [36], Spectral clustering [37], and Agglomerative clustering [38].

This paper is structured as follows: in Section 2, we describe the chosen HSI dataset and opted preprocessing steps. A brief discussion of the chosen unsupervised and supervised learning approaches is given in Section 3. Section 4 presents our results. The discussion and conclusion are articulated in Sections 5 and 6, respectively.

2. Imaging and data preprocessing

In this section, we discussed the details of the chosen dataset and opted preprocessing steps. As mentioned earlier, we used the multidimensional Choledoch database [20] in this work. It is obtained from the authors of [39]. To note, the authors in [20] developed a unique imaging system by combining a color charged-coupled device (CCD)-based commercial microscope with an acousto-optic tunable filter (AOTF)-coupled grayscale scientific complementary metal-oxide-semiconductor (sCMOS) sensor to image choledochal tissues. The AOTF controller was used to switch the wavelength from 550 to 1000 nm at a narrow bandwidth, e.g., 60 bands. Thus, both the two-dimensional spatial information (i.e., intensity image) and the one-dimensional spectral information, representing a three-dimensional (3D) cube, can be captured simultaneously. More details about the about imaging system can be found at [20]. Note that the dataset includes 880 scenes captured from 174 individuals. Out of these scenes, 689 are samples with part of cancer areas (L), 49 scenes full of cancer areas (N), and 142 scenes without cancer areas (P). Further, experienced pathologists have precisely labelled all cancer areas in the chosen dataset. In our simulations, we used images from the 'L folder', where each image consists of two regions i.e., one cancerous and the other noncancerous, which are ideal for performing binary classification. Figure 1 depicts some of the images taken from the database. To note, the size of a single-band image from an HSI data cube is 1024×1280 and each image is captured at different viewing angles.

It is worth mentioning the fact that flatfield correction (FFC) is one of the fundamental processes applied to the HSI dataset [40] to correct for variable illumination across an image. The FFC process is mentioned in Algorithm 1. It is known that variable illumination arises due to several factors, such as variations in the light source, optical path, sensor sensitivity, vignetting, etc. To overcome these limitations, FFC is often applied, which involves the division of the acquired image by a reference flatfield image (i.e., a blank image without any sample) to compensate for uneven illumination. This process helps scale the data to a limited range and remove systematic noise, improving image quality, which assists us in accurate analysis [40]. The effect of FFC is demonstrated in Fig. 2. In addition to this standard preprocessing step, in

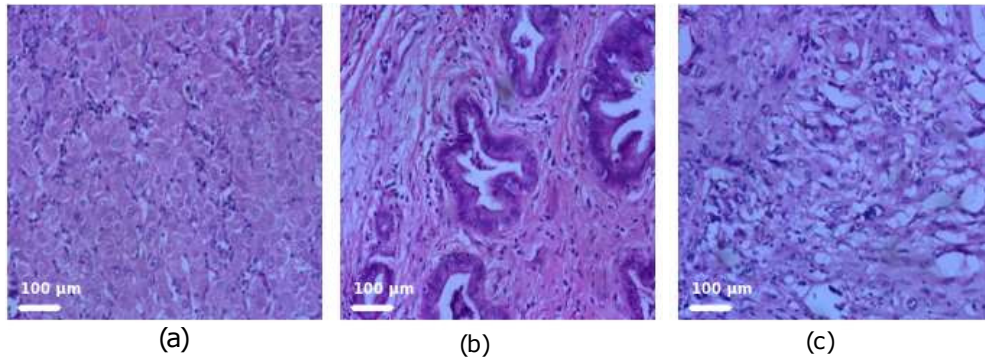


Fig. 1. RGB images taken from the Choledoch database [20]: (a) Normal; (b) Partially cancer (to be tested with USL and SL algorithms); and (c) Full Cancer image. Note that the scale bar is uniform for every microscopic image that has been used in this article.

this work, for the first time, we propose to use a set of bespoke algorithms to achieve highest classification accuracy when compared with SOTA architectures.

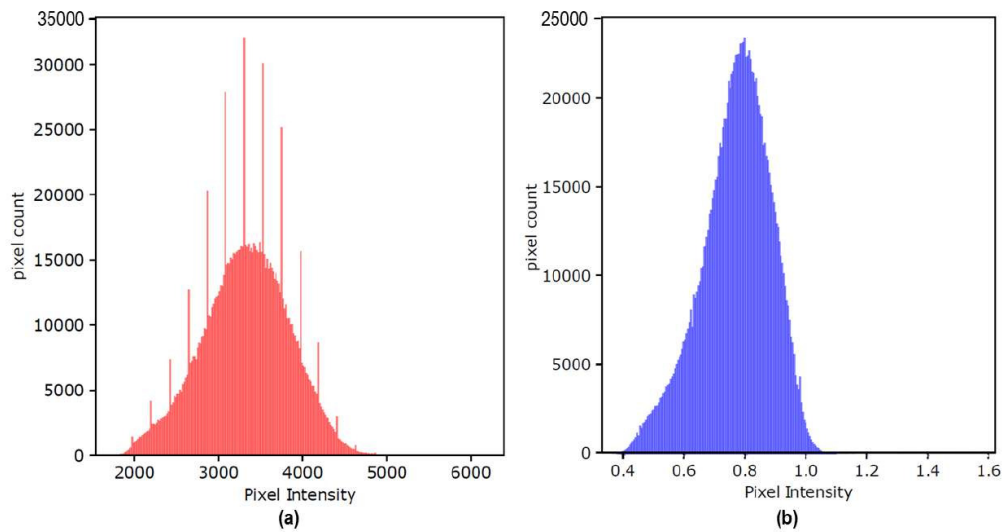


Fig. 2. Impact of flatfield correction (FFC): (a) Raw histogram, and (b) Flatfield corrected histogram

Algorithm 1. Flat Field Correction

```

1: Input: image, flat_field_ref           ▷ input image & flat field reference image
2: Ensure: The shape of image and flat_field_ref are same.
3: for each layer in image (i.e., spectral information) do
4:   if corresponding layer of flat_field_ref is not zero then
5:     layer of corrected_image  $\leftarrow \frac{\text{layer of } image}{\text{corresponding layer of } flat\_field\_ref}$ 
6:   end if
7: end for
8: Output: corrected_image

```

Firstly, to reiterate, HSI records the 3D information, i.e., two spatial (x, y) and a wide spectral bandwidth (λ) data, e.g., 60 bands in the currently used dataset. It is worth mentioning the fact that in most cases, significant information can be obtained from fewer bands, such as 20–30 bands. Intuitively, in such cases, dimensionality reduction (DR) approaches can be applied to remove redundant information from the original dataset. In the literature, several DR techniques have been proposed, for instance, principal component analysis (PCA) and its variations such as SPCA, KPCA and IPCA, linear discriminant analysis (LDA), and singular value decomposition (SVD), to name a few [31]. Analogously, in our chosen dataset, no significant information was perceived in the last few layers (bands). Thus, to remove redundant spectral information (with minimal information loss) after experimenting with multiple aforementioned DR techniques, we have opted for PCA, which gave us the best results in reducing the original spectral bands to 40% while preserving 98% of the key features. As a second step after DR, we applied classical median filtering [41] approach to reduce salt-and-pepper noise that was observed in the dataset. Finally, we normalized our dataset using a standard scaler function, thereby preventing any individual feature from dominating the learning process [42]. Figure 3 depicts our proposed workflow. Note that these preprocessing steps were applied in conjunction with the supervised learning algorithms.

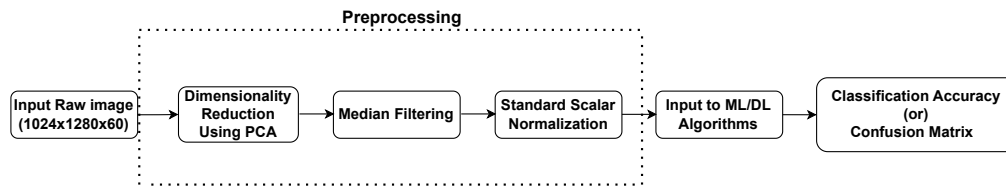


Fig. 3. Proposed workflow for the opted SL techniques.

3. Methodology

In this section, we first briefly discuss the chosen unsupervised and supervised learning approaches. We then present the schematic of the proposed 3D U-Net architecture.

3.1. Unsupervised learning approach

It is known that unsupervised learning (USL) algorithms classify images without any ground truth (or labels), as this approach relies on the identification of patterns and forming clusters within the data itself [43]. In our analysis, we used various well-known clustering techniques, such as K-means clustering, non-negative matrix factorization (NMF) [35], Otsu thresholding [44], spectral clustering [37], and agglomerative clustering [38]. As these techniques are well discussed in the literature, details of those are not described here, but readers are recommended to check Ref. [45] and references therein for more details.

3.2. Supervised learning approach

Unlike USL, supervised learning (SL) demands labelled data (ground truth) for training, thus being shown to achieve better classification accuracy for several applications. Figure 3 in depicts our proposed workflow. In this work, we used support vector machine (SVM) [46], DNN [47], and customized 3D U-Net [48] for SL-based analysis. To note, SVM and DNN are well-analyzed for the classification problems in the literature; therefore, a description of these techniques is not presented here. However, readers are recommended to read Ref. [49].

3.2.1. Model architecture for SL

After applying the opted preprocessing steps to the HSI data cube, pixel-wise features were extracted from each spectral band to train SVM and DNN. These features were then fed into the model as an input and binary classification was carried out accordingly. The predicted value is compared with the corresponding pixel in the annotated image for validation. An SVM model with the kernel as ‘RBF’, gamma = 1, and a regularization parameter C = 100, after grid search was considered for training. The optimal hyperparameters for SVM are tabulated in Table 1.

Table 1. Optimal Hyperparameters for SVM.

Hyper Parameters	Optimized Values
C	100
gamma	1
Kernel	RBF

Similarly, a sequential DNN model with four dense layers was considered, where the node sizes were selected as 48, 96, 48, and 29 with two dropout layers. To note, 64% of the entire dataset was considered for training, 16% was given for validation, and testing was carried out using the remaining 20% of the dataset. The optimal hyperparameters for DNN and 3D U-Net are given in Table 2, respectively. Further, the training epoch was stopped when the loss values with the training data were performed under the validation data.

Our customized 3D U-Net architecture is shown in Fig. 4. As can be seen, the architecture consists of two main parts: encoder and decoder. In the encoding phase, each block comprises two 3D convolutional layers that leverage the volumetric information in the HSI cube, separated by a dropout layer, which helps prevent overfitting and introduces non-linearity to the network. Following this layer, a max-pooling layer is used to extract the key information from the data in each block. Conversely, in the decoding phase, each block starts with a transpose layer for upsampling, followed by concatenation with skip connections [50] (from the encoder) to retain spatial details. Two 3D convolutional layers, also with a dropout layer in between, complete each decoding block. Further, ReLU and Sigmoid activation functions are used as activation functions, while the ADAM optimizer is employed for efficient weight updates during training [32]. The binary cross-entropy loss function is utilized to measure the disparity between predicted segmentation masks and ground truth data, ensuring the model learns effectively for tasks such as segmentation.

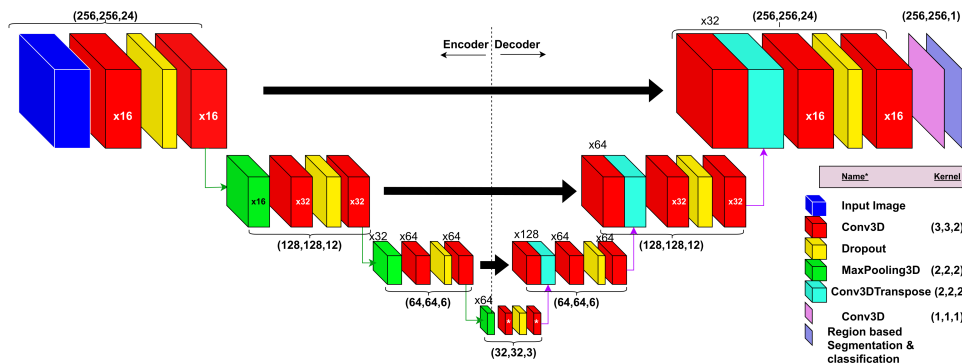


Fig. 4. The proposed 3D U-Net architecture.

Table 2. Optimal Hyperparameters for DL models devised.

Hyper Parameters	3D U-Net	DNN
Shape	Encoder: Conv3D(16, 32, 64, 256) Decoder: Conv3D(256, 64, 32, 16)	Conv2D(48, 96, 48, 29, 1)
Activation functions	Relu, Sigmoid	Relu, Sigmoid
Learning rates	0.005	0.001
Optimizers	Adam	Adam
No. of epochs	5	2
Batch size	2	500

4. Results

The evaluation results of various clustering and classification techniques used in our analysis are discussed in this section. The chosen frameworks were implemented using Python 3.9, with interactive development facilitated by Jupyter Notebooks [51]. To optimize the computational efficiency, TensorFlow 2.4 [52], designed for GPU acceleration, was utilized. An NVIDIA V100 GPU device was used to train the 3D U-Net model on a supercomputer. To note, it took approximately 3.5 hours to train the entire database before it started overfitting. The SVM and DNN models were trained on a CPU with 32GB of RAM hardware, with a time duration of around 20 minutes each. The visualization of the simulation results was carried out using the ‘matplotlib’ library [53].

4.1. Evaluation indices

It is known that the silhouette score (SS) measures the similarity of an object from its cluster and/or from the other clusters [54]. Generally, SS values are given between -1 and 1, where -1 refers to the poorly classified and 1 denotes the perfect classification. Mathematically, SS is given as follows:

$$SS(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

where $SS(i)$ is the silhouette score for the i^{th} data point. $a(i)$ represents the average distance between the i^{th} data point and all other data points within the same cluster, and $b(i)$ denotes the average distance between the i^{th} data point and all other data points in the nearest cluster, i.e., the cluster with the smallest average distance. Table 3 shows a quantitative comparison, in terms of classification accuracy (CA), for the chosen clustering techniques using SS [55]. In addition to SS and CA, we opted for a few standard metrics [56] such as structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), root mean squared error (RMSE), and correlation to compare the results of USL algorithms. The quantified values of the chosen USL algorithms are given in Table 4 and the corresponding results for USL-based approaches are shown in Fig. 5. Note that, for K-means clustering, we tried with different k-values ranging from 1 to 5. Out of them, $k = 2$ gave us the best result with the adopted elbow method [57]. As seen from Table 3, the spectral clustering method has a good SS, while agglomerative clustering achieves better classification accuracy. Also from Table 4 it is observed that agglomeration clustering outperforms the other listed unsupervised methods [56].

Similarly, for each proposed SL model, we used the confusion matrix [58]. It is known that the confusion matrix provides a breakdown of the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for each class in the dataset. By analyzing the confusion matrix, one can gain insights into the strengths and weaknesses of each model. In addition to the confusion matrix, we also calculated the other crucial parameters, such as sensitivity (SN), specificity (SP), classification accuracy (CA), precision (PN), recall, intersection over union

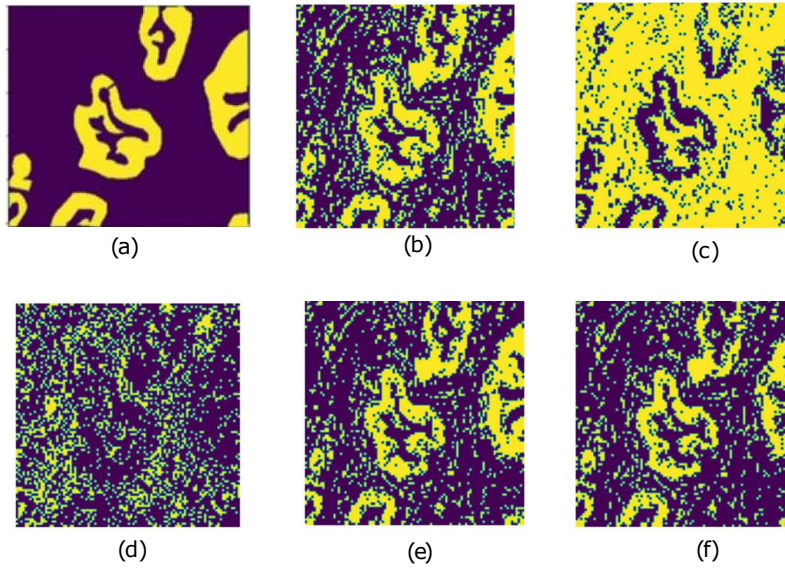


Fig. 5. Results obtained with different USL algorithms: (a) Ground truth, (b) K-means, (c) NMF, (d) Otsu, (e) Spectral Clustering, (f) Agglomerative Clustering.

Table 3. Comparison of USL classifiers in terms of SS and CA .

Method	SS	CA
K-means Clustering	0.3188	57.02
NMF	0.2184	33.13
OTSU Threshold	0.1113	10.89
Spectral Clustering	0.3280	60.29
Agglomerative Clustering	0.3181	61.89

Table 4. Image similarity indices for USL algorithms.

Method	SSIM	PSNR	RMSE	Correlation
K-means Clustering	0.8538	19.2983	0.1084	0.9569
NMF	0.2131	7.4207	0.4256	0.3458
OTSU Threshold	0.3769	11.3831	0.2697	0.7143
Spectral Clustering	0.9120	21.6709	0.0825	0.9745
Agglomerative Clustering	0.9985	25.7930	0.0072	0.9989

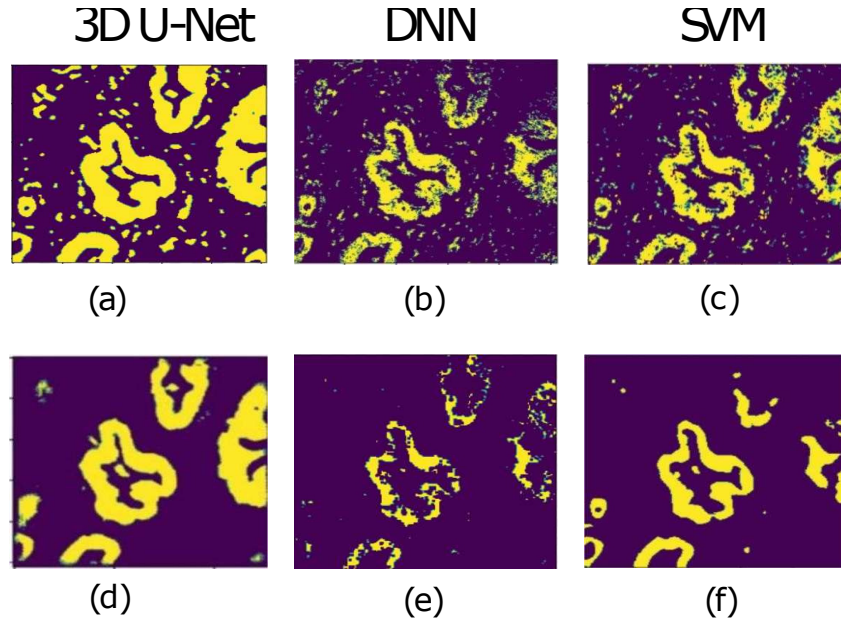


Fig. 6. Results observed with different SL algorithms: (a-c) without opted preprocessing steps; and (d-f) with opted preprocessing steps for 3D U-Net, DNN and SVM, respectively. Note that the ground truth image used for SL analysis is Fig. 5(a).

(IoU), and dice [59]. The mathematical representations of the aforementioned parameters are given as follows:

$$SN = \frac{TP}{TP + FN} \quad (2)$$

$$SP = \frac{TN}{TN + FP} \quad (3)$$

$$CA = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$PN = \frac{TP}{TP + FP} \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (6)$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (7)$$

The quantified values of the chosen SL algorithms are given in Table 5 and the corresponding results for the SL-based approach are shown in Fig. 6. Furthermore, as can be seen from Table 5, the proposed 3D U-Net achieves the highest classification accuracy of 89.49% (i.e., 1.29% more than the IANet) applied just after the flat-field correction, i.e., without additional opted preprocessing steps. However, by adding the opted preprocessing steps, classification accuracy is increased to 92.49% for 3D U-Net which is the best-proposed value for this dataset so far, which also proves the significance of our opted preprocessing steps. Similarly, the other parameters, such as SN, SP, PN, IoU and Dice, also show the efficiency of the proposed network (see Table 5). It is worth mentioning the fact that SVM and DNN (with opted preprocessing) show commendable performance in estimating the shape of objects and clustering them efficiently.

Nevertheless, we estimated a higher number of FNs in these approaches. Consequently, despite showing good values for SN, they show underperformance in terms of SP, CA, IoU and Dice metrics. Holistically, our customized 3D U-Net outperformed the SOTA models that have been presented in this work. It is evident from Fig. 6 that the opted preprocessing (i.e., PCA, filtering, and normalization) has a significant impact on the chosen dataset. For instance, Fig. 6(a-c) shows the classified results without the opted preprocessing where several outliers are present, which may lead to poor classification accuracy and other parameters, while Fig. 6(d-f) shows the best classification results that resemble the ground truth, Fig. 5(a). Further, as can be seen from Table 3 and Table 5, the supervised learning techniques outperform the unsupervised approach in terms of classification accuracy on the chosen dataset, even without applying the opted preprocessing steps.

Table 5. Comparison of various DL and ML algorithms. The symbol (*) denotes the results obtained without the opted preprocessing steps.

<i>Method</i>	<i>SN</i>	<i>SP</i>	<i>CA</i>	<i>PN</i>	<i>IOU</i>	<i>Dice</i>
3D U-Net	86.09	94.50	92.49	83.11	73.27	84.57
3D U-Net *	74.19	94.31	89.49	80.38	62.82	77.16
DNN	81.56	94.74	91.59	82.99	69.83	82.27
DNN *	52.45	95.13	84.93	77.23	45.43	62.47
SVM	83.58	88.88	87.62	70.28	61.76	76.36
SVM *	64.23	95.15	87.75	80.63	55.65	71.5
IANet[24]	80.43	91.61	88.20	—	66.91	79.82
CPFNet[25]	80.17	91.01	87.48	—	65.46	78.63
SegNet[26]	79.28	90.25	86.77	—	64.09	77.66

5. Discussion and future directions

From our analysis, the 3D U-Net showed the best performance, which can be justified as follows: The 3D U-Net architecture leverages volumetric data and captures spatial context and spectral context also across three dimensions, enabling an exhaustive understanding of the neighbourhood pixels. In addition to this, we have added long skip connections [60] which helps us mitigate the spatial feature loss, which is likely to happen at the decoder end of the 3D U-Net. Unlike traditional 2D approaches (those are listed in Table 5) which may overlook critical spatial information, 3D U-Net considers volumetric data, leading to more accurate object segmentation. By operating directly on volumetric HSI data, the model minimizes information loss and maintains the integrity of spatial relationships, resulting in more accurate object detection. Further, the inherent multi-scale feature fusion capabilities of 3D U-Net facilitate the integration of information from different spatial resolutions. By combining features at multiple scales, the model can effectively capture fine-grained details and global context, enhancing the accuracy of cholangiocarcinoma detection and classification. It is also inherently robust to spatial transformations and object size and shape variations. This robustness is advantageous in real-world scenarios where objects may appear at different scales or orientations. The ability to handle such variations contributes to improved detection accuracy across diverse environments and datasets. In addition to this, we also opted for a few preprocessing steps to achieve higher classification accuracy on the microscopic hyperspectral dataset. After an exhaustive analysis, we finalized the preprocessing approaches that are best suited for the chosen dataset. As mentioned earlier, the efficacy of the proposed architecture was tested against some of the SOTA networks, such as SegNet, CPFNet, and IANet. To reiterate, we followed pixel-wise training for SVM and DNN and 3D volumetric

data training for 3D U-Net, considering the spatial and spectral attributes of the pixel together with semantic segmentation.

It is worth mentioning the fact that there are a few limitations to our suggested network. One of the main limitations is that the efficacy of the model may be hampered by the extremely small amount of available data. In addition to this, filtering the chosen spectral bands is a laborious task, which could add subjectivity and impact the model's performance. Notwithstanding these drawbacks, the suggested method excels because it can reliably and efficiently produce SOTA results on the chosen dataset. Furthermore, it is known that performance degradation occurs when training is carried out with excessively imbalanced datasets, which was mitigated by adding skip connections.

Our future work includes investigating vision transformers, attention mechanisms for U-Nets, and dilated convolutions to simplify the network. For instance, vision transformers have emerged as a viable substitute for convolutional neural networks (CNNs) for various computer vision applications. Likewise, by including attention methods, U-Nets can better concentrate on pertinent aspects when segmenting data. This method may enhance U-Net's performance in applications involving hyperspectral imaging. Furthermore, dilated convolutions have been investigated to efficiently extract features and broaden the receptive field in deep learning without appreciably raising the number of parameters or processing expenses. Dilated convolutions allow the acceptance domain to be expanded without incurring excessive computational costs or parameter additions, resulting in a lighter network. It is known that they can capture long-range relationships, and vision transformers have become a possible substitute for CNNs in a variety of computer vision applications. By including attention mechanisms in U-Nets, performance in HSI classification tasks may be improved, as it allows for improved concentration on pertinent features during data segmentation. Dilated convolutions also provide a cost-effective way to increase the receptive field and extract features in deep learning without appreciably raising computing overhead or parameter values. When combined, these approaches may improve our proposed method.

6. Conclusion

In this study, we tested and compared the feasibility of several unsupervised and supervised learning approaches for cholangiocarcinoma classification using the hyperspectral microscopic choledoch dataset. From our analysis, it was estimated that the unsupervised algorithms were unreliable as they were producing more outliers and failed to achieve good classification accuracy on the chosen dataset. The best-performing unsupervised algorithm was 3D U-Net with 92.49% accuracy. Unlike unsupervised approaches, supervised learning approaches such as 3D U-Net architecture have been shown to achieve higher classification accuracy. To further validate our claim, we compared our results with state-of-the-art (SOTA) networks such as SegNet, CPFNet, and IANet and showed that the proposed 3D U-Net achieves at least 4.29% more classification accuracy than the SOTA models. It is worth mentioning the fact that this superiority is achieved by incorporating the opted preprocessing steps into the conventional 3D U-Net model. We, therefore, believe that our approach will ultimately help surgeons in accurately identify resection margins with precision in image-guided surgery.

Funding. SRM University; Andhra Pradesh.

Acknowledgments. The authors thank Prof. Qingli Li and their research group at East China Normal University, Shanghai, China, for providing access to their dataset. Sravan Kumar Sikhakolli and Suresh Aala are recipients of a University Research Fellowship (URF) from SRM University, AP, India. Omjee Pandey acknowledges the National Supercomputing Mission (NSM) for providing computing resources for PARAM Shivay at the Indian Institute of Technology (BHU), Varanasi, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and the Department of Science and Technology, Government of India.

Authors contributions: S.S.K., I.M., Conceptualization and methodology; S.S.K., O.P.S., G.M., S.A., and D.S., software implementation; I.M., O.M., A.D., validation, formal analysis, and project administration; S.C., A.D., and O.J.P.,

resource allocations; All authors contributed equally to the original draft preparation and have read and agreed to the published version of the manuscript.

Disclosures. The authors declare no conflict of interest.

Data availability. Data and model training parameters presented in the article will be made available upon reasonable request to corresponding authors.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, *et al.*, “Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *Ca-Cancer J. Clin.* **68**, 394–424 (2018).
2. Z. Xu, W. Zhang, C. Quesada, *et al.*, “Longitudinal monitoring of angiogenesis in a murine window chamber model in vivo,” *Tissue Eng., Part C* **30**(3), 93–101 (2024).
3. A. C. Society, “Bile duct cancer survival rates,” <https://www.cancer.org/cancer/types/bile-duct-cancer/detection-diagnosis-staging/survival-by-stage.html> (2022). Online; accessed 12-January-2024.
4. X.-L. Qin, Z.-R. Wang, J.-S. Shi, *et al.*, “Utility of serum ca19-9 in diagnosis of cholangiocarcinoma: in comparison with cea,” *World J. Gastroenterol.* **10**(3), 427 (2004).
5. J. E. Eaton, C. L. Welle, Z. Bakhshi, *et al.*, “Early cholangiocarcinoma detection with magnetic resonance imaging versus ultrasound in primary sclerosing cholangitis,” *Hepatology* **73**(5), 1868–1881 (2021).
6. H. Zhang, J. Zhu, F. Ke, *et al.*, “Radiological imaging for assessing the respectability of hilar cholangiocarcinoma: a systematic review and meta-analysis,” *BioMed Res. Int.* **2015**, 1 (2015).
7. M. Saleh, M. Virarkar, V. Bura, *et al.*, “Intrahepatic cholangiocarcinoma: pathogenesis, current staging, and radiological findings,” *Abdom Radiol.* **45**(11), 3662–3680 (2020).
8. D. Ledro-Cano, “Suspected choledocholithiasis: endoscopic ultrasound or magnetic resonance cholangiopancreatography? a systematic review,” *Eur. J. Gastroenterol. Hepatol.* **19**(11), 1007–1011 (2007).
9. N. Bedard, M. Pierce, A. El-Naggar, *et al.*, “Opinion article title: Emerging roles for multimodal optical imaging in early cancer detection: A global challenge running title: Multimodal optical imaging for global cancer detection,” *Ann Gillenwater* **1500**, 4 (2010).
10. M. E. Martin, M. B. Wabuye, K. Chen, *et al.*, “Development of an advanced hyperspectral imaging (hsi) system with applications for cancer detection,” *Ann. Biomed. Eng.* **34**(6), 1061–1068 (2006).
11. X. Zhou, L. Ma, W. Brown, *et al.*, “Automatic detection of head and neck squamous cell carcinoma on pathologic slides using polarized hyperspectral imaging and machine learning,” in *Medical Imaging 2021: Digital Pathology*, vol. 11603 (SPIE, 2021), pp. 165–173.
12. Z. Xu, Y. Jiang, J. Ji, *et al.*, “Classification, identification, and growth stage estimation of microalgae based on transmission hyperspectral microscopic imaging and machine learning,” *Opt. Express* **28**(21), 30686–30700 (2020).
13. S. Li, C. Jiao, Z. Xu, *et al.*, “Determination of geographic origins and types of lindera aggregata samples using a portable short-wave infrared hyperspectral imager,” *Spectrochim. Acta, Part A* **279**, 121370 (2022).
14. J. Luo, H. Zhang, E. Forsberg, *et al.*, “Confocal hyperspectral microscopic imager for the detection and classification of individual microalgae,” *Opt. Express* **29**(23), 37281–37301 (2021).
15. H. Zhang, J. Luo, S. Hou, *et al.*, “Incoherent broadband cavity-enhanced absorption spectroscopy for sensitive measurement of nutrients and microalgae,” *Appl. Opt.* **61**(12), 3400–3408 (2022).
16. Z. Xu, C. S. Locke, R. Morris, *et al.*, “Development of a semi-anthropomorphic photoacoustic calcaneus phantom based on nano computed tomography and stereolithography 3d printing,” *J. Orthop. Res.* **42**(3), 647–660 (2024).
17. Z. Xu, Y. Jiang, and S. He, “Multi-mode microscopic hyperspectral imager for the sensing of biological samples,” *Appl. Sci.* **10**(14), 4876 (2020).
18. C. Jiao, Z. Xu, Q. Bian, *et al.*, “Machine learning classification of origins and varieties of tetrastigma hemsleyanum using a dual-mode microscopic hyperspectral imager,” *Spectrochim. Acta, Part A* **261**, 120054 (2021).
19. H. Namburu, V. N. Munipalli, M. Vanga, *et al.*, “Cholangiocarcinoma classification using medisawhsi: A breakthrough in medical imaging,” in *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, (2024), pp. 1–6.
20. Q. Zhang, Q. Li, G. Yu, *et al.*, “A multidimensional choledoch database and benchmarks for cholangiocarcinoma diagnosis,” *IEEE Access* **7**, 149414–149421 (2019).
21. Y. Deng, J. Yin, Y. Wang, *et al.*, “Resnet-50 based method for cholangiocarcinoma identification from microscopic hyperspectral pathology images,” in *Journal of Physics: Conference Series*, vol. 1880 (IOP Publishing, 2021), p. 012019.
22. J. Wang, X. Mao, Y. Wang, *et al.*, “Automatic generation of pathological benchmark dataset from hyperspectral images of double stained tissues,” *Opt. Laser Technol.* **163**, 109331 (2023).
23. C. Bench, J. Nallala, C.-C. Wang, *et al.*, “Unsupervised segmentation of biomedical hyperspectral image data: tackling high dimensionality with convolutional autoencoders,” *Biomed. Opt. Express* **13**(12), 6373–6388 (2022).
24. W. Zeng, W. Li, and R. Tao, “Janet: Important-aware network for microscopic hyperspectral pathology image segmentation,” in *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, (IEEE, 2022), pp. 1–6.
25. S. Feng, H. Zhao, F. Shi, *et al.*, “Cpfnet: Context pyramid fusion network for medical image segmentation,” *IEEE Trans. Med. Imaging* **39**(10), 3008–3018 (2020).

26. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
27. R. Helin, U. G. Indahl, O. Tomic, *et al.*, "On the possible benefits of deep learning for spectral preprocessing," *J. Chemom.* **36**(2), e3374 (2022).
28. T.-C. Lee, J. Zhou, and Z. Yu, "Deep learning based adaptive filtering for projection data noise reduction in x-ray computed tomography," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072 (SPIE, 2019), pp. 229–233.
29. A. Nurunnabi, F. Teferle, J. Li, *et al.*, "An efficient deep learning approach for ground point filtering in aerial laser scanning point clouds," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLIII-B1-2021**, 31–38 (2021).
30. M. Tavakoli, S. Jazani, and M. Nazar, "Automated detection of microaneurysms in color fundus images using deep learning with different preprocessing approaches," in *Medical imaging 2020: imaging informatics for healthcare, research, and applications*, vol. 11318 (SPIE, 2020), pp. 110–120.
31. J. Khodr and R. Younes, "Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas," in *2011 4th international congress on image and signal processing*, vol. 4 (IEEE, 2011), pp. 1875–1883.
32. V. C. Dodda, L. Kuruguntla, K. Elumalai, *et al.*, "A denoising framework for 3d and 2d imaging techniques based on photon detection statistics," *Sci. Rep.* **13**(1), 1365 (2023).
33. I. A. Cruz-Guerrero, R. Leon, L. Granados-Castro, *et al.*, "Reflectance calibration with normalization correction in hyperspectral imaging," in *2022 25th Euromicro Conference on Digital System Design (DSD)*, (IEEE, 2022), pp. 855–862.
34. S. Qamar, H. Jin, R. Zheng, *et al.*, "A variant form of 3d-unet for infant brain segmentation," *Future Generation Computer Systems* **108**, 613–623 (2020).
35. N. Kumar, P. Uppala, K. Duddu, *et al.*, "Hyperspectral tissue image segmentation using semi-supervised nmf and hierarchical clustering," *IEEE Trans. Med. Imaging* **38**(5), 1304–1313 (2018).
36. S. Sehgal, S. Kumar, and M. H. Bindu, "Remotely sensed image thresholding using otsu & differential evolution approach," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, (IEEE, 2017), pp. 138–142.
37. P. Zheng, H. Zhou, J. Liu, *et al.*, "Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture," *Appl. Energy* **349**, 121607 (2023).
38. Z. Liu, "Unsupervised clustering for intrinsic mode functions selection in hyperspectral image classification," *Multimed. Tools Appl.* **83**(13), 37387–37407 (2024).
39. L. Sun, M. Zhou, Q. Li, *et al.*, "Diagnosis of cholangiocarcinoma from microscopic hyperspectral pathological dataset by deep convolution neural networks," *Methods* **202**, 22–30 (2022).
40. Y. Wang, Y. Gu, and X. Li, "A novel low rank smooth flat-field correction algorithm for hyperspectral microscopy imaging," *IEEE Trans. Med. Imaging* **41**(12), 3862–3872 (2022).
41. H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sensing* **44**(2), 397–408 (2006).
42. A. Alizadeh Naeini, M. Babadi, and S. Homayouni, "Assessment of normalization techniques on the accuracy of hyperspectral data clustering," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLII-4/W4**, 27–30 (2017).
43. B. Melit Devassy, S. George, and P. Nussbaum, "Unsupervised clustering of hyperspectral paper data using t-sne," *J. Imaging* **6**(5), 29 (2020).
44. X. Tan, M. Cui, Y. Xu, *et al.*, "Hyperspectral image classification using otsu thresholding and morphological operations," *Remote Sens.* **11**, 466 (2019).
45. Z. Ghahramani, "Unsupervised learning," in *Summer school on machine learning*, (Springer, 2003), pp. 72–112.
46. M. A. Hearst, S. T. Dumais, E. Osuna, *et al.*, "Support vector machines," *IEEE Intell. Syst. Their Appl.* **13**(4), 18–28 (1998).
47. V. Sze, Y.-H. Chen, T.-J. Yang, *et al.*, "Efficient processing of deep neural networks: A tutorial and survey," *Proc. IEEE* **105**(12), 2295–2329 (2017).
48. Ö. Çiçek, A. Abdulkadir, and S. S. Lienkamp, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19*, (Springer, 2016), pp. 424–432.
49. R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, (2006), pp. 161–168.
50. L. Qian, C. Wen, Y. Li, *et al.*, "Multi-scale context unet-like network with redesigned skip connections for medical image segmentation," *Computer Methods and Programs in Biomedicine* **243**, 107885 (2024).
51. G. Van Rossum, "Python programming language," in *USENIX annual technical conference*, vol. 41 (Santa Clara, CA, 2007), pp. 1–36.
52. M. Abadi, A. Agarwal, P. Barham, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv*, arXiv:1603.04467 (2016).
53. J. D. Hunter, "Matplotlib: A 2d graphics environment," *Comput. Sci. Eng.* **9**(3), 90–95 (2007).
54. B. Everitt, *An Introduction to Cluster Analysis* (John Wiley & Sons, Chichester, West Sussex, UK, 2011).
55. P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics* **20**, 53–65 (1987).

56. M. Kokare, B. Chatterji, and P. Biswas, "Comparison of similarity metrics for texture image retrieval," in *TENCON 2003. Conference on convergent technologies for Asia-Pacific region*, vol. 2 (IEEE, 2003), pp. 571–575.
57. C. Shi, B. Wei, S. Wei, *et al.*, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *J. Wireless Com. Network* **2021**(1), 31 (2021).
58. S. Visa, B. Ramsay, A. L. Ralescu, *et al.*, "Confusion matrix-based feature selection," *Maics* **710**, 120–127 (2011).
59. R. Trevethan, "Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice," *Front. Public Health* **5**, 307 (2017).
60. G. Xu, X. Wang, X. Wu, *et al.*, "Development of skip connection in deep neural networks for computer vision and medical image analysis: A survey," *arXiv*, arXiv:2405.01725 (2024).