IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 21, 2019, accepted October 7, 2019, date of publication October 15, 2019, date of current version October 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947470

A Multidimensional Choledoch Database and Benchmarks for Cholangiocarcinoma Diagnosis

QING ZHANG[1], QINGLI LI[1,2], (Senior Member, IEEE), GUANZHEN YU[3], LI SUN[1], MEI ZHOU[1], AND JUNHAO CHU[1]

[1]Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China
[2]Engineering Center of SHMEC for Space Information and GNSS, Shanghai 200241, China
[3]Longhua Hospital, Shanghai 200030, China

Corresponding author: Qingli Li (qlli@cs.ecnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61975056, in part by the Shanghai Natural Science Foundation under Grant 19ZR1416000, and in part by the Science and Technology Commission of Shanghai Municipality under Grant 14DZ2260800 and Grant 18511102500.

ABSTRACT Histopathological examination is very important for diseases diagnosis and treatment. With the development of artificial intelligence, more and more pathological databases have been reported for histopathological diagnosis because database is quite crucial for the validation and testing of feature extraction, statistical analysis and deep learning algorithms. However, most of these databases are either gray images or RGB color images of tissue sections contain limited information of samples which limited the performance of most current deep learning algorithms. There are few publicly available pathological databases that include more than two modalities for the same subject. This paper introduces a database for both microscopy hyperspectral and color images of cholangiocarcinoma, including 880 scenes from 174 individuals, among which 689 scenes are samples with part of cancer areas, 49 scenes full of cancer areas, and 142 scenes without cancer areas. In addition, all cancer areas have been precisely labeled by experienced pathologists. The contributions of this work: a) A comprehensive and up-to-date review on pathological imaging systems and databases; b) Detailed description of the proposed the multidimensional Choledoch Database and login method; c) The multidimensional Choledoch Database has been published and can be downloaded after registration and made an entry on the website.

INDEX TERMS Database, artificial intelligence, hyperspectral imaging, pathology.

I. INTRODUCTION

Histopathological examination usiually been regarded as the 'gold standard' of tumor diagnosis and therapy. Traditional histopathological examination is performed by pathologist with light microscopy, which is time-consuming and laborious. With the development of image processing and artificial intelligence technology, deep learning has made great progress in pathological analysis in recent years. For example, some studies on pathological and normal voice identification [1], gastric cancer diagnosis [2], pathological retina images segmentation [3], gait analysis [4], and pathological cells recognition [5] have got high recognition accuracy. However, there still exist certain problems in the pathological diagnosis of the current artificial intelligence technology and the data resources are the most important one [6]. For most of these studies, a standard pathological database is very important to extract features, do data analysis and get the satisfied diagnosis results. In addition, the availability is also significant to guarantee the AI research results [7], [8]. So there are several researchers publishing large numbers of pathological database to support the development of algorithms such as brain database [9] and lung database [10]. In recent years, there are several latest pathological databases released for the validation and testing of feature extraction, deep learning algorithms and statistical analysis, e.g. the VOice ICar fEDerico II (VOICED) [8] and the National Cancer Data Base (NCDB) [11] and also a hyperspectral database for brain cancer [12]. Among all these databases, some classic pathological databases such as Stanford Tissue Microarray Databases (TMAD), Medical image databases of nuclear segmentation, Colorectal Histology MNIST, The University College London (UCL) Low-density Lipoprotein Receptor Gene (LDLR) Variant Database, BreakHis Database,

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Peng.

and The Cancer Genome Atlas (TCGA) have been set up for pathological analysis. However, most of these datasets are all traditional color pathological images which can just provide limited information and then limit the improvement of accuracy of pathological image processing algorithms from the perspective of information source. In contrast, microscopy hyperspectral images contain both spatial and spectral information and have been utilized to identify brain cancer [13], oral cancer [14], prostate cancer [15], tongue tumor [16], et al. Although these studies have got certain experimental results, most of them are based on traditional image processing methods as there are few microscopy hyperspectral pathological datasets. In view of these problems, this paper introduced a dataset contains both the microscopy hyperspectral images and the RGB color images of the same field of view of cholangiocarcinoma tissues which is unusual in the academic sector. On top of this, all the images in the dataset are labeled manually by experienced pathologists to generate annotations files so as to provide a new kind of data source with both spatial and spectral information for the pathological diagnosis of cholangiocarcinoma by artificial intelligence.

This paper presents a multidimensional choledoch database which contains both microscopy hyperspectral and RGB color images with fine labeling. These images can be classified into three types: L (samples part of cancer areas, with annotation files), N (samples full of cancer areas) and P (samples without cancer areas). In addition, diverse non-blank regions are collected on each tissue specimen, as well as one blank region collected for microscopy hyperspectral data calibration.

II. RELATED WORK

A. EXISTING PATHOLOGICAL IMAGING SYSTEMS

High resolution whole-slide-imaging (WSI) system was first introduced in 1999 [17] with the development of computer technology, robotic microscopes, digital cameras and the software converting images of histological sections on glass slides prepared from frozen, HE (hematoxylin and eosin) stained, formalin fixed and paraffin embedded tissue into digital files [18]. Then it has been widely used in several biomedical applications, especially in pathology [19]–[21]. WSI systems are composed of illumination systems, microscopic optical components, and a focusing system placing in front of a camera precisely [22]. There are various kinds of focusing systems (virtual slide) the particular scanners can be used, such as tiling, line scanning, dual sensor scanning, dynamic focusing, and array scanning [23]. Most of current WSI imaging systems use tile-based scanning or line-based scanning methods to get the whole slide image of tissue sections [24].

Tile-based scanning method obtains a large number of square image frames that are assembled into a mosaic pattern relying on a robotics-controlled motorized slide stage. However, there is usually 2%-5% overlap in one image due to the precise movement of the slide stage. Then each tile will be captured by a charge coupled device (CCD) and stored into the memory. Finally, these tiles will be automatically associated with each other to ensure proper alignment so that the patterns can be stitched correctly to get a whole slide digital image of the histological section [19], [24].

Line-based scanning method utilizes a servomotor-based slide stage which moves in a linear and jitter-free manner along a single axis. Subsequently, a set of images in the form of long and uninterrupted lines will be produced after multiple successive passes at different locations on the slide. Different from the tile-based scanning, this method can greatly simplifies the alignment process as the number of lines and the degrees of freedom associated with each line are greatly reduced [19].

Investigators can acquire images of different resolution depending on the microscopic objective for scanning (eg, $\times$ 4, $\times$ 20, $\times$ 40, $\times$ 100), the numerical aperture of the objective and the quality of the CCD [25]. As a result, the diagnostic accuracy of the images is equal to the one taken from glass slides [26]–[28] and these images can reach a resolution of less than 0.5 μm [23]. In addition, the software of WSI is gradually becoming more and more convenient now. It allows investigators to freely navigate the images of histological section they want and also preview the images that are magnified in real time. They can capture either the regions they want automatically after they have circled on the virtual slides, [22] or the whole slide, and even several slides [19], [29]. However, WSI systems require high-quality specimen preparation and the specimen has to be cleaned carefully so that high-quality images can be obtained [30]. Additionally, almost all the WSI scanners capture color images only [29] and then investigators detect, segment and analyze the lesion areas based on these images acquired.

B. EXISTING DATABASES SUMMARY AND COMPARISON

Several researchers have been making efforts to develop pathological databases to keep in pace with pathological image analysis algorithms. Table 1 provides a brief overview of the most popular databases supplied for researches on pathological detection, classification and segmentation algorithms and also used as training data for deep learning algorithms. As shown in Table 1, most of current pathological databases are color images captured by WSI systems.

Stanford Tissue Microarray Database (TMAD) [31] are widely used in several pathological image processing algorithms and are constantly updated [32]–[35]. Now TMAD has contained about 600,000 pathological images across sarcoma, breast, lymphoma, and so on. Among TMAD, there are some special databases such as MMMP (Multi-Dimensional Microscopic Molecular Profiling) database [36]. MMMP was published in 2015 which can measure several individual molecular properties in the same histologic section at subcellular resolution. In this database, the tissue microarray includes 102 human tissues with a panel of 15 informative antibodies, as well as 5 histochemical stains plus DAPI of several organs such as colon, lung and breast.

TABLE 1. A comprarison of popular pathological databases.

Database Name & References	Hyperspectral	Color	Imaging Precision	#Sample	Organ	Highlights
MMMP Database[36]		✔	Subcellular resolution	102	Colon, lung and breast	Stitching images on intact tissue sections of several organs
Medical image databases of nuclear segmentation[33]			Nuclear resolution	30	breast, liver, kidney, prostate, bladder, colon and stomach	30 cropped images of over 21,000 nuclei
Colorectal Histology MNIST[32]		✔	Individual cell	625	colon	5000 eight types human colorectal cancer
UCL LDLR Variant Database[38]		✔	gene	1288	receptor	LDLR variants and their potential pathogenicity
BreakHis Database[39]		✔	Individual cell	82	breast	7909 breast cancer images for benign and malignant tumors
TCGA[40]		✔	Individual cell	Over 11000	lung, ovarian and so on	more than 11000 cases across 33 tumor types
Multidimensional Choledoch Database	✔	✔	Individual cell	174	bile duct	880 scenes of multidimensional images

Medical image database of nuclear which were published in 2018 and available in the website (http://nucleisegmentationbenchmark.weebly.com/website) comprise of labeled hematoxylin-eosin staining (H&E) images culled from digitalized tissue samples from 30 full-section imaging system, which contain samples of benign and diseased breast, liver, kidney, prostate, bladder, colon, and stomach. This database contains high quality features for nuclear morphometry and can be used for computational pathology, such as density, nucleus-mass ratio, size, shape, and pleomorphism. The features extracted from nuclear segmentation images of digital microstructure can be used to access tumor grade and predict therapeutic efficacy. What's more, 30 cropped images of more than 21,000 nuclei in this dataset were all labeled and validated by pathologist so that the nuclear can be segmented accurately [33].

Colorectal Histology MNIST (Mixed National Institute of Standards and Technology) published in 2016 contains 5000 histological images for eight different types of tissues of human colorectal cancer (http://creativecommons.org/licenses/by/4.0/) [32]. The eight types of tissues consist of tumor epithelium, sample stroma (homogeneous composition, includes tumor stroma, extra-tumoral stroma and smooth muscle), complex stroma (including single tumor cells and/or few immune cells), debris (containing necrosis, hemorrhage and mucus), immune cells (containing immune-cell conglomerates and sub-mucosal lymphoid follicles), normal mucosal glands, adipose tissue, and background (no tissue). In addition, contiguous tissue areas were manually annotated and tessellated [32]. The database has been classified into 8 types with which researchers can make comparison among different classification algorithms using this database.

The University College London (UCL) Low-density Lipoprotein Receptor Gene (LDLR) Variant Database was first set up in 1996 and gradually updated with the addition of different variants (https://grenada.lumc.nl/LOVD2/UCL Heart/home.php?select_db = LDLR) [37]. Now the database contains over 1288 various variants from several familial hypercholesterolemia (FH) patients [38]. Among these variants, 55% belong to exonic substitutions, 22% belong to exonic small rearrangements, 11% belong to large rearrangements, 10% belong to intronic variants, and 2% belong to promoter variants [38]. This database can be used for pathogenicity prediction.

BreakHis Database was released in 2014 and is composed of 7909 images which can be sorted into benign and malignant tumors. (available on the website http://web.inf.ufpr.br/vri/breast-cancer-database) [39]. The benign breast tumors are divided into four histological distinct types: adenosis (A), tubular adenoma (TA), fibroadenoma (F), and phyllodes tumor (PT). Malignant tumors can also be classified into four histological distinct types: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC) [39]. Therefore, it can be used to study classification algorithms on breast tumors.

The Cancer Genome Atlas (TCGA) started in 2006 which contained three cancer types: lung, ovarian, and glioblastoma [40]. During the past ten years, TCGA has studied more than 11,000 cases which include 33 tumor types, such as colon and endometrial cancer (https://cancergenome.nih.gov/publications). Therefore, it is more convenient for researchers to train their algorithms using data from TCGA for cancer diagnosis and catalog specific genomic and molecular changes [40].

III. THE MULTIDIMENSIONAL CHOLEDOCH DATABASE

A. HARDWARE AND SOFTWARE CONFIGURATION OF ACQUISITION SYSTEM

To set up the multidimensional choledoch database, a microscopy hyperspectral imaging system is developed and used to capture both the hyperspectral and RGB color

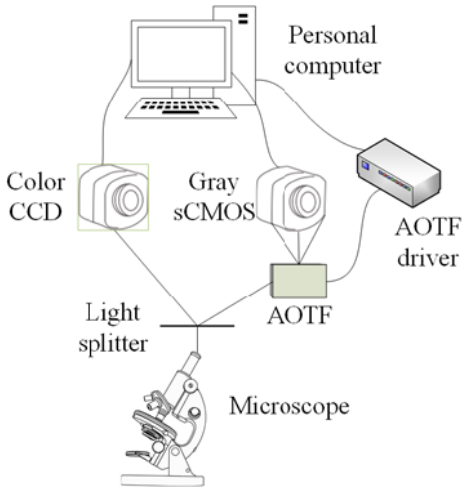

FIGURE 1. Schematic diagram of the microscopy hyperspectral imaging system.

images of choledoch tissues. As shown in Fig. 1, the imaging system consists of a microscope (Nikon 80i, Nikon Corp.), an acousto-optic tunable filter (AOTF) adapter (VA310-.37-.80-L, Brimrose Corp.), an SPF Model AOTF controller (VFI130-140SPFB2C2exSTS, Brimrose Corp.), a gray scientific complementary metal oxide semiconductor (sCMOS, Dhyana 400D, Tucsen Corp.), a color charge coupled device detector (color CCD, DigiRetina 16, Tucsen Corp.), and a personal computer [41]. The light transmitted from the choledoch tissue slice is collected by the microscope with the objective lens of 20×, then filtered by the AOTF and imaged on the sCMOS. Different single band images are captured by sCMOS with the wavelength switching from 550 nm to 1000 nm at the narrow bandwidth via the AOTF. These single band images consist of two-dimensional spatial information and one-dimensional spectral information as illustrated in Fig. 2. These single band images can be visualized as a three-dimensional cube because of its intrinsic structure, where the cube face is a function of the spatial coordinates and the depth is a function of the wavelength. In order to compare with those traditional RGB color image based pathological diagnosis methods, the RGB color images with the same field of view of hyperspectral images are captured by a color CCD and stored into the database. The multidimensional imaging software running on the Microsoft .NET Framework 4.0 is programmed in C# with the character of friendly user interface to control all hardware work automatically for image capture.

B. THE MULTIDIMENSIONAL CHOLEDOCH DATABASE

The multidimensional choledoch database contains both microscopy hyperspectral images and RGB color images of choledoch tissues captured by the microscopy hyperspectral imaging system. The choledoch tissues stained with HE (hematoxylin and eosin) are provided by Changhai hospital, Shanghai, China with the approval of the ethics committee and the slide thickness is 10 microns. This database contains 880 scenes of multidimensional images

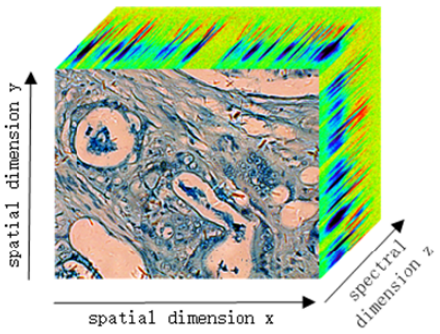

FIGURE 2. Microscopy hyperspectral data cube.

captured from choledoch tissues of 174 patients. Among these multidimensional images, 689 scenes are images contain part of cancer areas, 49 scenes are full of cancer areas, and 142 scenes are images without cancer areas. As an example, one scene of image 031368c-20x-roi2 are shown in Fig. 3. It contains a microscopy hyperspectral data cube and a RGB color image of the same field of view. Fig. 4 shows different spectra extracted from the same microscopy hyperspectral data cube. From the figures it can be seen that there are differences among different parts of Choledoch section in different single band images and spectra which can be used for diseases identification and the development of three-dimensional algorithms because researchers can extract the spectral information.

IV. HOW TO USE THE DATABASE

A. IMAGE NAMING CONVENTION

In the multidimensional choledoch database, the image filename illustrates the image information. The naming convention is interpreted in Fig. 4. The format is comprised of 3 kinds of samples: L-samples with part of cancer areas, N-samples with full of cancer areas, and P-normal samples without cancer areas. Each scene contains both microscopy hyperspectral image and RGB color image. Among these data, L images have been labeled and annotation files have been generated. The filename consists of three parts which are separated by dash as illustrated by the example of '030406-20x-roi1'. The serial number '030406' in the first part represents the number of tissue specimen. The second part '20x' represents the magnification of the objective lens used. The third part 'roi1' represents different field of view of the same pathological section.

B. IMAGE FORMATTING

There are four kinds of files in the multidimensional choledoch database and different files have different sizes as shown in TABLE 2. Among these files, '.raw' files occupy most of the memory because each raw file contains 60 bands filled with spectral information. The image size of one scene of color image is 2304 × 1728 while the image size of single band image of the microscopy hyperspectral date cube is 1280 × 1024. Files in '.hdr' format contain description information of '.raw' files. Among several parameters in the

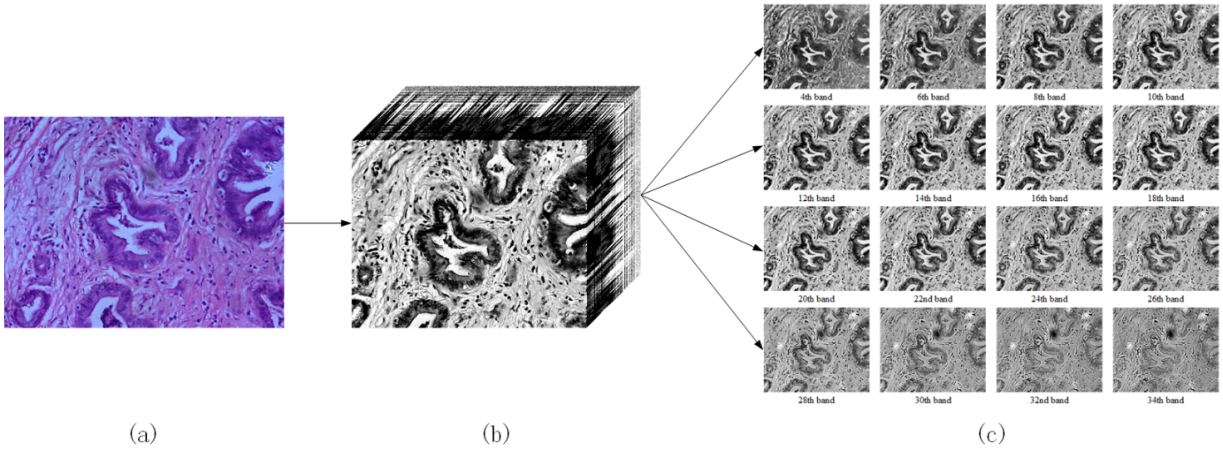

FIGURE 3. A scene of multidimensional image of the dataset. (a) RGB image, (b) microscopy hyperspectral data cube, (c) 16 single band images extracted from hyperspectral date cube.

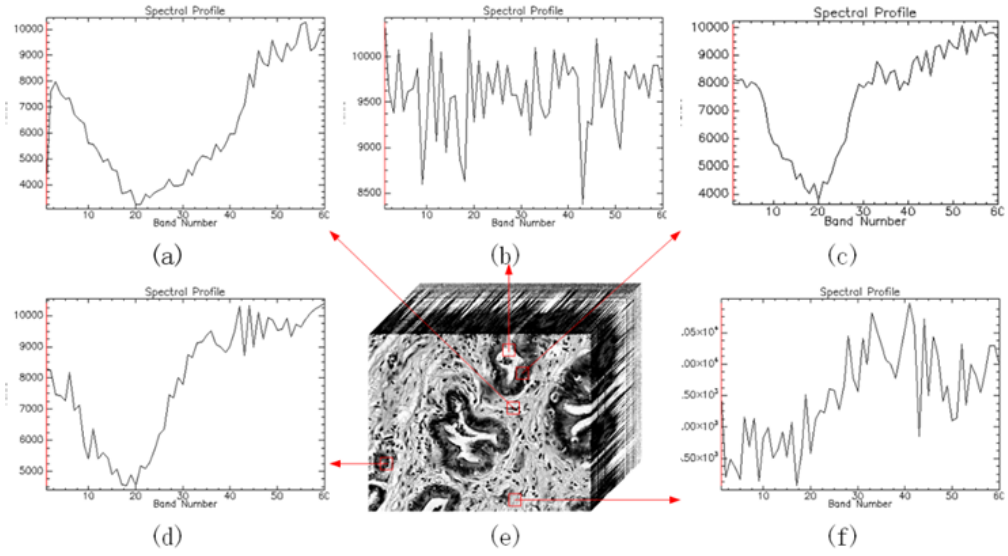

FIGURE 4. Spectra extracted from the microscopy hyperspectral data cube.

TABLE 2. Various attributes of the files in the multidimentional choledoch database.

Categories	File format	Average file size
Hyperspectral image	.hdr	1KB
	.raw	153.6MB
RGB image	.jpg	1.15MB
Annotations file	.xml	5.14KB

'.hdr' file, 'band = 60' indicates that the hyperspectral daye cube contains 60 bands. 'data type = 2' means that each pixel in each single band image has two bytes. What's more, 'interleave = bsq' is on behave of that the images are stored in BSQ (band sequential format) format.

Each scene of microscopy hyperspectral image contains 60 band. Fig. 5 shows the 30th band images extracted from the hyperspectral data cube on which the pathologists labeled manually to generate annotations files, that is, '.xml' files. '.xml' files include coordinates of all the points that can build up areas filled with tumor. As there are several tumor areas on one scene usually, the coordinates of different tumor areas are separated by the information of objects in the '.xml' file. These annotations files can also be transferred from microscopy hyperspectral images to RGB color images with coordinate transformation method as shown in Fig. 6 and Fig. 7. These labeled RGB color images can be used to train the deep learning methods and compared with those hyperspectral based methods. Therefore, the multidimensional choledoch database is important for researchers to evaluate both the RGB color image based and hyperspectral based deep learning pathology diagnosis algorithms.

C. AN EXAMPLE TO USE THE DATABASE

In order to demonstrate the availability of the database, we provide an example to explain how to use the released

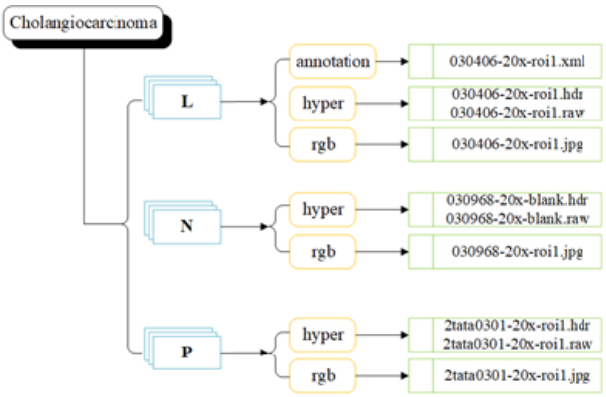

FIGURE 5. Database naming convention.

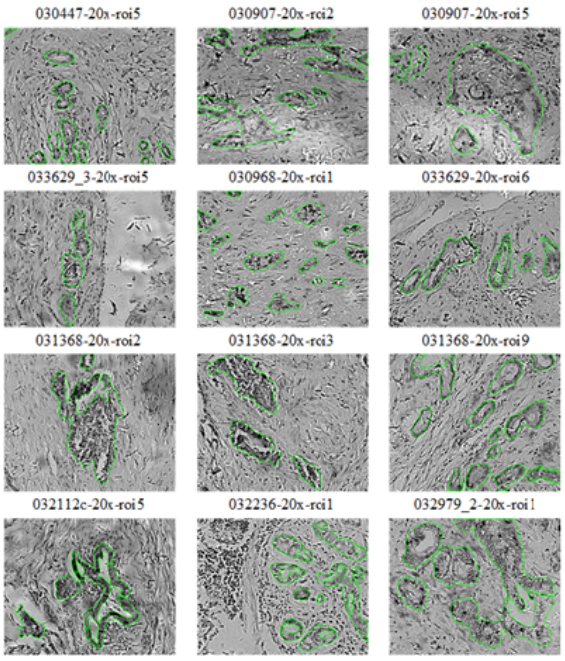

FIGURE 6. Single band images with annotations extracted from microscopy hyperspectral images of the multidimensional Choledoch database.

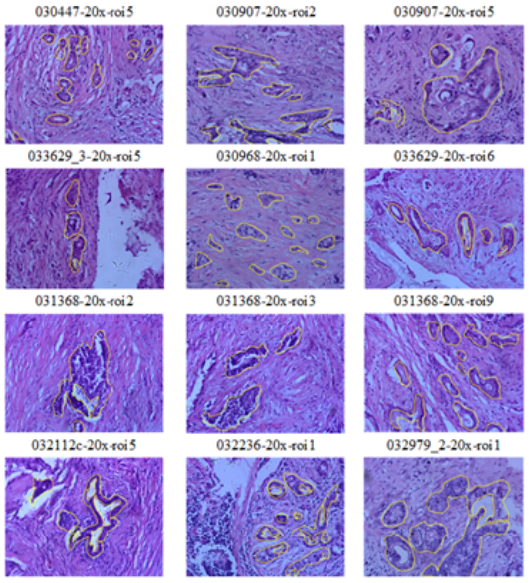

FIGURE 7. RGB color images with transferred annotations in the same field of view.

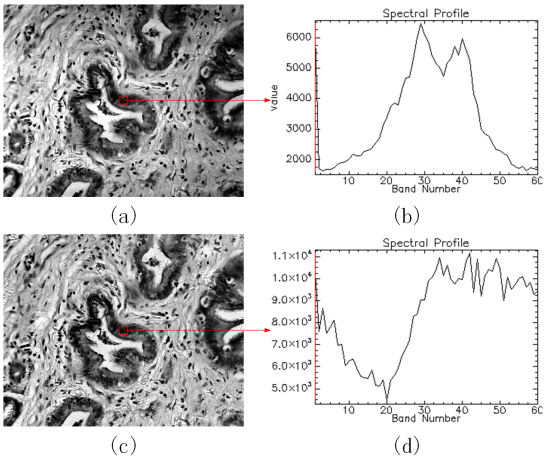

FIGURE 8. Preprocessing of the microscopy hyperspectral image. The 20th single band image before (a) and after preprocessing (c); spectrum extracted from the data cube before (b) and after preprocessing (d).

database. To train a deep learning method using the database, a data preprocessing procedure is needed to remove the effects of noises, artifacts, and inhomogeneous spectral response of the microscopy hyperspectral images [42]. These effects in the original data can be corrected by some flat-field correction methods [43]. The database contains the calibration data cube for each scene of microscopy hyperspctral images which captured under the same illumination with the samples. Fig. 8 shows the 20th single band image, the spectrum before and after preprocessing. From the figure, it can be seen that the influence of the inhomogeneous spectral response of the system can be removed and the spectrum of samples can be retrieved. Then, these corrected microscopy hyperspectral images can be used for further processing such as segmentation, classification, and training some algorithms.

To evaluate the database, the Neural Net (NN) [44], [45] and Support Vector Machine (SVM) [46]–[48] algorithms are used to segment tumor areas from the preprocessed images [49]. The parameters of each algorithm used in this experiment are illustrated in TABLE 3. The comparison between the annotation image and segmentation results of NN and SVM are shown in Fig. 9. Several accuracy parameters are calculated and listed in Table 4. From the tables and figures it can be seen that the multidimensional choledoch database can be used to evaluate algorithms and develop new identification methods.

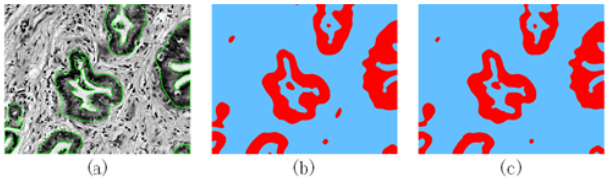

FIGURE 9. Comparison between the annotation data and the results of segmentation algorithms. (a) single band image with annotation; (b) NN; (c) SVM.

TABLE 3. Parameters for NN and SVM.

Algorithm	Neural Net (NN)		Support Vector Machine (SVM)	
Parameters	Activation	Logistic	Kernel Type	Polynomial
	Training Threshold contribution	0.9000	Degree of Kernel Polynomial	2
	Training Rate	0.2000	Bias in Kernel Function	1.000
	Training Momentum	0.9000	Gamma in Kernel Function	0.017
	Training RMS Exit Criteria	0.1000	Penalty Parameter	100.00
	Number of Hidden Layers	1	Pyramid Levels	0
	Number of Training Iterations	1000	Classification Probability Threshold	0.00

TABLE 4. Accuracy parameters of NN and SVM.

Parameters	Kappa	accuracy	sensitivity	specificity	precision
NN	0.8270	0.9375	0.9173	0.9433	0.8234
SVM	0.8364	0.9427	0.8834	0.9598	0.8637

D. HOW TO ACCESS

Availability of the database is of great importance for research in AI field. The multidimensional choledoch database presented in this paper can be accessed through the website http://bio-hsi.ecnu.edu.cn/. To access to the database, one should register and get the authorization from the administrator. Then both the microscopy hyperspectral images and RGB color images of choledoch with annotations can be downloaded by login the website.

V. CONCLUSION

Histopathological analysis is usually regarded as the 'gold standard' of tumor diagnosis and clinical treatment. In recent years, artificial intelligence (AI) has been used to perform pathologic diagnosis and made great progress. However, the data sources most of these methods used are color images captured by traditional light microscopy, which limits the performance of these methods as this kind of images contain limited pathological information. As a result, two-dimension algorithms for image processing have been developed into maturity and researcher are committed to studying three-dimension algorithms to obtain more accurate results with more information. Nevertheless, few three-dimensional databases are published online for researches. In this paper, a multidimensional choledoch database which contains both microscopy hyperspectral images and RGB color images at the same field of view is proposed for deep learning studies. All images in this database have been evaluates and labeled by experienced pathologists which is suitable for training neural networks. This database is very useful for researchers to investigate new multidimensional deep learning algorithms for pathologic diagnosis as it contains morphology, spectrum, and biochemical changes information of samples. And that's the point of this paper, so we operated two simple algorithms on our database just to prove the availability of the choledoch database rather than investigate the algorithm profoundly. Yet we will do some research on image processing algorithms, especially three-dimensional AI algorithms, based on the database we announce in this paper to detect, segment, or recognize tumor areas more precisely next.

Up to now, the presented multidimensional choledoch database is the first public choledoch pathology database contains both microscopy hyperspectral and RGB color images with annotations of choledoch sections. The public availability of this database will enhance the development and evaluation of new deep learning algorithms of pathologic diagnosis of cholangiocarcinoma.

ACKNOWLEDGMENT

The authors have no relevant financial interests or conflicts of interest to disclose.

REFERENCES

[1] B. Sabir, F. Rouda, Y. Khazri, B. Touri, and M. Moussetad, "Improved algorithm for pathological and normal voices identification," *Int. J. Elect. Comput. Eng.*, vol. 7, no. 1, pp. 238–243, 2017.
[2] Q.-K. Liang, Y. Nan, S. Xiang, L. Mei, W. Sun, and G.-Z. Yu, "Feasibility of deep learning image-based segmentation algorithm in pathological section of gastric cancer," *Acad. J. Second Military Med. Univ.*, vol. 39, no. 8, pp. 903–908, 2018.
[3] B. S. Y. Lam and H. Yan, "A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields," *IEEE Trans. Med. Imag.*, vol. 27, no. 2, pp. 237–246, Feb. 2008.
[4] D. A. Bruening and S. T. Ridge, "Automated event detection algorithms in pathological gait," *Gait Posture*, vol. 39, no. 1, pp. 472–477, 2014.
[5] B. Liu, C. Yin, Z. Liu, Z. Zhang, J. Gao, M. Zhu, J. Gu, and K. Xu, "Microscopic image analysis and recognition on pathological cells," in *Proc. Can. Conf. Elect. Comput. Eng.*, Apr. 2007, pp. 1022–1025.
[6] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.
[7] N. Duchateau, M. Sermesant, H. Delingette, and N. Ayache, "Model-based generation of large databases of cardiac images: Synthesis of pathological cine MR sequences from real healthy cases," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 755–766, Mar. 2018.
[8] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "A new database of healthy and pathological voices," *Comput. Elect. Eng.*, vol. 68, pp. 310–321, May 2018.
[9] S. Ortega, G. M. Callicó, M. L. Plaza, R. Camacho, H. Fabelo, and R. Sarmiento, "Hyperspectral database of pathological in-vitro human brain samples to detect carcinogenic tissues," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 369–372.

[10] G. Han, X. Liu, F. Han, I. N. T. Santika, Y. Zhao, X. Zhao, and C. Zhou, "The LISS—A public database of common imaging signs of lung diseases for computer-aided detection and diagnosis research and medical education," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 648–656, Feb. 2015.

[11] M. H. Kamel, J. Tao, J. Su, M. I. Khalil, N. K. Bissada, B. Schurhamer, P. E. Spiess, and R. Davis, "Survival outcomes of organ sparing surgery, partial penectomy, and total penectomy in pathological T1/T2 penile cancer: Report from the national cancer data base," *Urologic Oncol., Seminars Original Invest.*, vol. 36, no. 2, pp. 82.e7–82.e15, 2018.

[12] H. Fabelo *et al.*, "*In-vivo* hyperspectral human brain image database for brain cancer detection," *IEEE Access*, vol. 7, pp. 39098–39116, 2019.

[13] H. Fabelo *et al.*, "Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations," *PLoS ONE*, vol. 13, no. 3, 2018, Art. no. e0193721.

[14] J.-R. Duann, C.-I. Jan, M. Ou-Yang, C.-Y. Lin, J.-F. Mo, Y.-J. Lin, M.-H. Tsai, and J.-C. Chiou, "Separating spectral mixtures in hyperspectral image data using independent component analysis: Validation with oral cancer tissue sections," *Proc. SPIE*, vol. 18, no. 12, 2013, Art. no. 126005.

[15] H. Akbari, L. Halig, D. M. Schuster, B. Fei, A. Osunkoya, V. Master, P. Nieh, and G. Chen, "Hyperspectral imaging and quantitative analysis for prostate cancer detection," *Proc. SPIE*, vol. 17, no. 7, 2012, Art. no. 076005.

[16] Z. Liu, H. Wang, and Q. Li, "Tongue tumor detection in medical hyperspectral images," *Sensors*, vol. 12, no. 1, pp. 162–174, 2012.

[17] J. Ho, A. V. Parwani, D. M. Jukic, Y. Yagi, L. Anthony, and J. R. Gilbertson, "Use of whole slide imaging in surgical pathology quality assurance: Design and pilot validation studies," *Hum. Pathol.*, vol. 37, no. 3, pp. 322–331, 2006.

[18] B. F. Boyce, "An update on the validation of whole slide imaging systems following FDA approval of a system for a routine pathology diagnostic service in the United States," *Biotechn. Histochem.*, vol. 92, no. 6, pp. 381–389, 2017.

[19] N. Farahani, A. Parwani, and L. Pantanowitz, "Whole slide imaging in pathology: Advantages, limitations, and emerging perspectives," *Pathol. Lab. Med. Int.*, vol. 7, pp. 23–33, Jun. 2015.

[20] L. Pantanowitz, J. H. Sinard, W. H. Henricks, L. A. Fatheree, A. B. Carter, L. Contis, B. A. Beckwith, A. J. Evans, A. Lal, and A. V. Parwani, "Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the college of American pathologists pathology and laboratory quality center," *Arch. Pathol. Lab. Med.*, vol. 137, no. 12, pp. 1710–1722, 2013.

[21] G. Bueno, O. Déniz, M. D. M. Fernández-Carrobles, N. Vállez, and J. Salido, "An automated system for whole microscopic image acquisition and analysis," *Microsc. Res. Techn.*, vol. 77, no. 9, pp. 697–713, 2014.

[22] F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: Whole-slide imaging and beyond," *Annu. Rev. Pathol., Mech. Disease*, vol. 8, pp. 331–359, Jan. 2013.

[23] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, K. A. Erps, A. K. Bhattacharyya, Y. Yagi, and J. R. Gilbertson, "Overview of telepathology, virtual microscopy, and whole slide imaging: Prospects for the future," *Hum. Pathol.*, vol. 40, no. 8, pp. 1057–1069, Aug. 2009.

[24] J. D. Webster and R. W. Dunstan, "Whole-slide imaging and automated image analysis: Considerations and opportunities in the practice of pathology," *Vet. Pathol.*, vol. 51, no. 1, pp. 211–223, 2014.

[25] T. C. Cornish, R. E. Swapp, and K. J. Kaplan, "Whole-slide imaging: Routine pathologic diagnosis," *Adv. Anatomic Pathol.*, vol. 19, no. 3, pp. 152–159, 2012.

[26] S. Al-Janabi, A. Huisman, M. Nap, R. Clarijs, and P. J. Van Diest, "Whole slide images as a platform for initial diagnostics in histopathology in a medium-sized routine laboratory," *J. Clin. Pathol.*, vol. 65, no. 12, pp. 1107–1111, 2012.

[27] L. Pantanowitz, P. N. Valenstein, A. J. Evans, K. J. Kaplan, J. D. Pfeifer, D. C. Wilbur, L. C. Collins, and T. J. Colgan, "Review of the current state of whole slide imaging in pathology," *J. Pathol. Inform.*, vol. 2, p. 36, Aug. 2011.

[28] T. W. Bauer, L. Schoenfield, R. J. Slaw, L. Yerian, Z. Sun, and W. H. Henricks, "Validation of whole slide imaging for primary diagnosis in surgical pathology," *Arch. Pathol. Lab. Med.*, vol. 137, no. 4, pp. 518–524, 2013.

[29] A. J. Evans, M. E. Salama, W. H. Henricks, and L. Pantanowitz, "Implementation of whole slide imaging for clinical purposes: Issues to consider from the perspective of early adopters," *Arch. Pathol. Lab. Med.*, vol. 141, no. 7, pp. 944–959, 2017.

[30] R. W. Dunstan, K. A. Wharton, Jr., C. Quigley, and A. Lowe, "The use of immunohistochemistry for biomarker assessment—Can it compete with other technologies?" *Toxicologic Pathol.*, vol. 39, no. 6, pp. 988–1002, 2011.

[31] R. J. Marinelli, K. Montgomery, C. L. Liu, N. H. Shah, W. Prapong, M. Nitzberg, Z. K. Zachariah, G. J. Sherlock, Y. Natkunam, R. B. West, and M. van de Rijn, "The Stanford tissue microarray database," *Nucleic Acids Res.*, vol. 36, pp. D871–D877, Nov. 2008.

[32] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 27988.

[33] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[34] R. A. Metcalf, S. Zhao, M. W. Anderson, Z. S. Lu, I. Galperin, R. J. Marinelli, A. M. Cherry, I. S. Lossos, and Y. Natkunam, "Characterization of D-cyclin proteins in hematolymphoid neoplasms: Lack of specificity of cyclin-D2 and D3 expression in lymphoma subtypes," *Mod. Pathol.*, vol. 23, pp. 420–433, Jan. 2010.

[35] M. Anthimopoulos, S. Christodoulidis, L. Ebner, T. Geiser, A. Christe, and S. Mougiakakou, "Semantic segmentation of pathological lung tissue with dilated fully convolutional networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 714–722, Mar. 2019.

[36] V. G. Onete, M. G. Besselink, C. M. Salsbach, C. H. Van Eijck, O. R. Busch, D. J. Gouma, I. H. de Hingh, E. Sieders, C. H. Dejong, J. G. Offerhaus, and I. Q. Molenaar, "Impact of centralization of pancreatoduodenectomy on reported radical resections rates in a nationwide pathology database," *HPB*, vol. 17, no. 8, pp. 736–742, 2015.

[37] K. E. Heath, M. Gahan, R. A. Whittall, and S. E. Humphries, "Low-density lipoprotein receptor gene (LDLR) world-wide Website in familial hypercholesterolaemia: Update, new features and mutation analysis," *Atherosclerosis*, vol. 154, no. 1, pp. 243–246, 2001.

[38] E. Usifo, S. E. A. Leigh, R. A. Whittall, N. Lench, A. Taylor, C. Yeats, C. A. Orengo, A. C. R. Martin, J. Celli, and S. E. Humphries, "Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: Update and pathological assessment," *Ann. Hum. Genet.*, vol. 76, no. 5, pp. 387–401, 2012.

[39] F. A. Spanhol, L. S. Oliveira, L. Heutte, and C. Petitjean, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[40] C. Hutter and J. C. Zenklusen, "The cancer genome atlas: Creating lasting value beyond its data," *Cell*, vol. 173, no. 2, pp. 283–285, 2018.

[41] Q. Li, D. Xu, X. He, Y. Wang, Z. Chen, H. Liu, Q. Xu, and F. Guo, "AOTF based molecular hyperspectral imaging system and its applications on nerve morphometry," *Appl. Opt.*, vol. 52, no. 17, pp. 3891–3901, 2013.

[42] Y.-Q. Zheng, "Precise spectral calibration for hyperspectral imager," *Guangxue Jingmi Gongcheng/Opt. Precis. Eng.*, vol. 18, no. 11, pp. 2348–2354, 2010.

[43] Q. Li, X. He, Y. Wang, H. Liu, D. Xu, and F. Guo, "Review of spectral imaging technology in biomedical engineering: Achievements and challenges," *Proc. SPIE*, vol. 18, no. 10, Oct. 2013, Art. no. 100901.

[44] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 990–1002, 2015.

[45] D. Zhang, Y. Song, D. Liu, H. Jia, S. Liu, Y. Xia, H. Huang, and W. Cai, "Panoptic segmentation with an end-to-end cell R-CNN for pathology image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 11071, 2018, pp. 237–244.

[46] T. Wu, M. H. Bae, M. Zhang, R. Pan, and A. Badea, "A prior feature SVM-MRF based method for mouse brain segmentation," *Neuroimage*, vol. 59, no. 3, pp. 2298–2306, Feb. 2012.

[47] P. Banu, N. Sakthieswaran, P. Suresh, and K. Kanagalakshmi, "SVM based lung tumor segmentation using Otsu's thresholding," *Int. J. Pharmacy Technol.*, vol. 8, no. 3, pp. 18611–18620, 2016.

[48] T. Ramakrishnan and B. Sankaragomathi, "A professional estimate on the computed tomography brain tumor images using SVM-SMO for classification and MRG-GWO for segmentation," *Pattern Recognit. Lett.*, vol. 94, pp. 163–171, Jul. 2017.

[49] M. K. Bhowmik, U. R. Gogoi, G. Majumdar, D. Bhattacharjee, D. Datta, and A. K. Ghosh, "Designing of ground-truth-annotated DBT-TU-JU breast thermogram database toward early abnormality prediction," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1238–1249, Jul. 2018.