Fakultät für Informatik
Technische Universität München

Deep learning based medical image segmentation and classification for artificial intelligence healthcare

Yu Zhao

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Nils Thuerey

Prüfende der Dissertation:
1. Prof. Dr. Bjoern H. Menze
2. Prof. Dr. Georg Langs

Die Dissertation wurde am 06.10.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 17.03.2021 angenommen.

Abstract

As one of the most advanced iterations of machine learning methods, deep learning has recently shown record-breaking performance in previously difficult tasks such as image analysis, natural language processing, speech recognition, and information retrieval. Unlike traditional machine learning methods that rely on hand-crafted features, deep learning has the advantage of being able to learn salient feature representations automatically and effectively. Considering the contradiction between the dramatic increase of healthcare data and the limitation of medical experts who can address those data and make clinical decisions, deep learning with the data-driven nature has gained increasing attention in the healthcare domain. Deep learning is well suited for medical data due to its ability to identify informative patterns in sparse, noisy data and the advantage of requiring little efforts of data pre-processing and feature engineering. Current researchers have demonstrated the potential of deep learning to achieve competitive and even superior performance compared to human experts on multiple tasks in medical image analysis. However, the success of deep learning in healthcare is still in the early stages, and innovation works such as the network architecture, learning framework, and training procedure are needed to tackle different specific clinical tasks.

This thesis focuses on developing novel deep learning based methods to address medical image segmentation and classification issues such as small organ segmentation, prostate cancer lesion characterization, parkinsonian syndrome diagnosis, lymph node metastasis prediction, and microsatellite instability prediction. The collected medical image data in these clinical tasks are of different modalities including computed tomography (CT), positron emission tomography (PET), and whole slide imaging (WSI) and these tasks have different clinical backgrounds. Therefore, different deep learning frameworks and strategies are proposed and employed. Specifically, the detailed contributions can be summarized as follows. (1) We propose an automatic approach for small organ segmentation with limited training data using two cascaded steps — localization and segmentation. The localization stage involves the extraction of the region of interest after the registration of images to a common template

and during the segmentation stage, a novel *knowledge-aided* convolutional neural network is proposed to improve segmentation accuracy. (2) Besides, we develop an end-to-end deep neural network to characterize the prostate cancer lesions on ^{68}Ga-PSMA-11 PET/CT imaging automatically. (3) In the lymph node metastasis prediction, we propose a multiple instance learning method based on deep graph convolutional network and weakly supervised feature selection for histopathological image classification. (4) Furthermore, in the parkinsonian syndrome diagnosis task, a 3D deep convolutional neural network is utilized on ^{18}F-fluorodeoxyglucose (FDG) PET images for the automated differential diagnosis of idiopathic Parkinson's disease from multiple system atrophy and progressive supranuclear palsy. And we depicted in saliency maps the decision mechanism of the deep learning method to assist the physiological interpretation of deep learning performance. (5) Finally, we developed a deep-learning-based multiple instance learning method to predict microsatellite instability from histopathology images.

Zusammenfassung

Auf dem Gebiet der künstlichen Intelligenz und unter den Methoden des maschinellen Lernens hat Deep Learning innerhalb kürzester Zeit Rekordleistungen verzeichnet bei der Lösung von zuvor schwierigen Aufgaben im Bereich der Bildanalyse, maschineller Verarbeitung natürlicher Sprachen oder Spracherkennung und Informationsabruf. Im Gegensatz zu herkömmlichen Methoden des maschinellen Lernens, die sehr von manuellen Eingaben abhängig sind, hat Deep Learning den Vorteil, dass die entscheidenden Merkmale der Lösungsalgorithmen automatisch und effektiv gelernt werden können. Angesichts der Herausforderung durch eine dramatischen Zunahme von Datenmengen aus dem Gesundheitswesen und der zugleich eingeschränkten Kapazität medizinischer Experten, die diese Daten interpretieren und klinische Entscheidungen treffen können, hat das Deep Learning im Gesundheitswesen zunehmend an Bedeutung gewonnen. Deep Learning eignet sich gut für medizinische Datenverarbeitung, da es informative Muster in zerstreuten, verrauschten Daten identifizieren kann und bietet den Vorteil wenig Aufwand für die Vorverarbeitung sowie das Feature Engineering zu benötigen. Aktuelle Forschungen haben das Potenzial von Deep Learning demonstriert, bei verschiedenen Aufgaben in der medizinischen Bildanalyse wettbewerbsfähig gegenüber menschlichen Experten zu sein und sogar dem Experten gegenüber überlegene Leistungen zu erzielen. Der Erfolg des Deep Learnings im Gesundheitswesen befindet sich jedoch noch im Frühstadium, weitere innovative Arbeiten im Bereich der Netzwerkarchitektur, der Programmbibliotheken und das Optimieren von Trainingsverfahren sind noch erforderlich, um spezifische klinische Aufgaben bewältigen zu können.

Diese Arbeit konzentriert sich auf die Entwicklung neuer Deep-Learning-basierter Methoden zur Bearbeitung medizinischer Bildsegmentierungs- und Klassifizierungsprobleme wie die Segmentierung kleiner Organe, Charakterisierung von Prostatakrebs-Läsionen, Diagnose des Parkinson-Syndroms und Vorhersage von Lymphknotenmetastasen sowie von Mikrosatelliteninstabilität. Die in diesen klinischen Aufgaben gesammelten medizinischen Bilddaten weisen unterschiedliche Modalitäten auf, einschließlich Computertomographie (CT), Positronenemissionstomographie (PET) und Whole Slide Imaging (WSI). Diese Aufgaben haben unterschiedliche klinische Hin-

tergründe. Daher werden verschiedene Deep-Learning Modelle und Strategien vorgeschlagen und umgesetzt. Insbesondere können die Beiträge wie folgt zusammengefasst werden: (1) Wir stellen einen automatischen Ansatz für die Segmentierung kleiner Organe mit begrenzten Trainingsdaten vor, den wir in zwei hintereinandergeschaltenen Schritten unternehmen - Lokalisierung und Segmentierung. Der Lokalisierungsschritt umfasst die Extraktion des Interessenbereichs nach erfolgter Registrierung von Bildern auf eine gemeinsame Vorlage. Für den Segmentierungsschritt wird ein neuartiges wissensunterstütztes faltendes neuronales Netzwerk vorgestellt, um die Genauigkeit der Segmentierung zu verbessern. (2) Außerdem haben wir ein End-to-End tiefes neuronales Netzwerk entwickelt, um Prostatakrebs-Läsionen bei der ^{68}Ga-PSMA-11 PET/CT Bildgebung automatisch zu charakterisieren. (3) Bei der Vorhersage von Lymphknotenmetastasen stellen wir eine Mehrfachinstanz-Lernmethode vor, die auf einem Deep-Graph-Faltungsnetzwerk und einer schwach überwachten Merkmalsauswahl für die histopathologische Bildklassifizierung basiert. (4) Darüber hinaus wurde bei der Diagnose des Parkinson-Syndroms ein tiefes 3D-Faltungsneuronennetzwerk auf ^{18}F-Fluorodeoxyglucose (FDG)-PET-Bildern verwendet, um diese automatisch von den Differentialdiagnosen der idiopathischen Parkinson-Krankheit aus multipler Systematrophie sowie der progressiven supranukleären Lähmung zu unterscheiden. Wir stellen in Saliency-Maps den Entscheidungsmechanismus der Deep-Learning-Methode dar, um die physiologische Interpretationen der Deep-Learning-Leistungen zu unterstützen. (5) Schließlich entwickelten wir ein Deep-Learning-basiertes Modell mit einem Multiplen-Instanz-Lernansatz, um Mikrosatelliteninstabilität basierend auf histopathologischen Bildern vorherzusagen.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Bjoern H. Menze, who introduced me to the field of medical image analysis and offered me the great support for my Ph.D. study during the past four years. I am grateful for his advice and guidance. It is my honor to work with him and thank him for sharing personal experiences in research and providing a flexible research environment, I enjoyed a happy study life at the Technical University of Munich.

Additionally, I would like to thank Dr. Kuangyu Shi, Prof. Axel Rominger, and Dr. Ali Afshar-Oromieh who supervised me during my exchange at the University of Bern. I would also like to express my appreciation to Dr. Jianhua Yao and Dr. Fan Yang for their guidance and support during my internship at Tencent China. I am also grateful to Prof. Wendong Xu for his great help in our collaborative research project and the letter of recommendation for me.

Besides, I would like to acknowledge the co-authors of my publications. I highly appreciate their input, advice, comments, and feedback. Special thanks go to Hongwei Li, Xiaobin Hu, Anjany Sekuboyina whose suggestions and ideas inspired me a lot throughout my doctoral studies, to Giles Tetteh for his patience, generous help, and guidance in the early stage of my research, and to Dr. Mingming Wu, Dr. Dhritiman Das, Diana Waldmannstetter, and Dr. Pedro A. Gómez for their help during my writing this dissertation.

Furthermore, I would like to thank my colleagues and friends at the image-based biomedical modeling group. To name a few, Esther Alberts, Markus Rempfler, Lina Xu, Patrick Christ, Marie Bieth, Marie Piraud, Xin Liu, Jana Lipkova, Cagdas Ulas, Yusuf Yilmaz, Judith Zimmermann, Patrick Bilic, Amir Hossein Bayat, Carolin Pirkl, Fernando Navarro, Suprosanna Shit, Oliver Schoppe, and Ivan Ezhov. Thank them for the contribution to the enjoyable memories of Munich.

Then, a particular acknowledgment to the Chinese Scholarship Council, which provided funds for my four-year study in Germany.

Last but not least, I would like to deeply thank my family, in particular, my parents, Mr. Long Zhao and Mrs. Qingli Yu, for their unconditional love and support

through all my life.

Contents

List of Figures

Acronyms

AD . . . . . Alzheimer's disease

CNN . . . . . convolutional neural network

CPUs . . . . . central processing units

CT . . . . . computed tomography

DNN . . . . . deep neural network

EHR . . . . . electronic health record

EM . . . . . expectation-maximization

FCN . . . . . fully convolutional network

fMRI . . . . . functional magnetic resonance imaging

GPUs . . . . . graphics processing units

IPD . . . . . idiopathic Parkinson's disease

MAS . . . . . multi-atlas segmentation

MIL . . . . . multiple instance learning

MRI . . . . . magnetic resonance imaging

MSA . . . . . multiple system atrophy

PET . . . . . positron emission tomography

PSP . . . . . progressive supranuclear palsy

ROI . . . . . region of interest

SMI . . . . . standard multiple instance

US . . . . . ultrasound

VPF . . . . . vantage point forests

WSI . . . . . whole slide image

1

Introduction

Nowadays, healthcare providers are generating and capturing an explosion of healthcare-related data containing valuable signals and information, with the digitization of medical records, increasing affordability of molecular testing and widespread use of medical imaging technologies such as ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and histology slides. Considering the large amount of time and economic costs required to train medical experts, as well as the time constraints caused by the fatigue of these experienced experts, it becomes a huge challenge to analyze all collected medical data by healthcare professionals in the traditional way. Therefore, there is currently a more urgent demand for artificial intelligence technologies to assist human doctors in automatic diagnosis, treatment planning and prognostic predictions, alleviate their daily workload, and provide patients with efficient and personalized care. Machine learning, as a dominant problem-solving technique of artificial intelligence therefore has gained widespread attention due to its advantage to integrate, analyze and make predictions based on large and heterogeneous data sets. With recent advances in machine learning and the support of the increasing multi-modality patient data, healthcare service can be gradually transformed into personalized healthcare, which means the diagnosis, management decision, and therapy are specially personalized based on collected information of a patient [1].

Deep learning [2], as the most recent form of machine learning, can exploit informative feature representations in a self-taught way by automatically learning from data, which is different from the procedure in transitional machine learning approaches by designing hand-crafted features according to domain-specific knowledge [3]. Due to the advantages of incorporating the feature engineering into a learning process and requiring little efforts for data pre-processing, deep learning is rapidly becoming popular. In addition, with the continuous development of computing power (high-tech central processing units (CPUs) and graphics processing units (GPUs)) and the availability of a huge amount of data, it has achieved record-breaking

performance on multiple tasks in medical image analysis [4, 5, 6, 7, 8]. Moreover, the deep-learning systems can achieve real-time results showing the potential to make healthcare more efficient [3]. Notably, the authors in reference [9] reported a dermatologist-level approach based on the deep neural network for the classification of skin cancer. Reference [10] proposed an end-to-end deep learning method which can achieve cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms. More recent studies have demonstrated the successes of the deep learning in breast cancer prediction [11] and the prediction of the lung cancer risk [12], where the deep neural networks were on-par or outperformed human experts. In the pathologic analysis field, which is regarded as the golden standard in cancer diagnosis, deep learning also showed its potential in tasks such as the prediction of microsatellite instability directly from histology in gastrointestinal cancer [10] and the diagnosis of the prostate cancer, basal cell carcinoma, and breast cancer metastases to axillary lymph nodes [13].

Medical image data are employed throughout the entire process of healthcare, such as diagnosis, treatment planning, intraoperative navigation and postoperative monitoring. According to the estimation by IBM researchers, medical images currently account for at least 90% of all medical data, making it the largest data source in the healthcare [14]. Although deep learning methods are developing rapidly, deep learning has not yet fulfilled its potential in medical image analysis. Moreover, the deep learning models are challenged for lacking sufficient interpretability which is a major bottleneck to the widespread acceptance of these models in clinical practice. Therefore, in this thesis, we focus on leveraging the deep learning methods for medical image analysis, in particular, medical image segmentation and medical classification (computer-aided diagnosis based on medical images) and investigating the interpretability of deep neural networks.

1.1 Medical Image Segmentation

1.1.1 Motivation and Challenges

Image segmentation, as the foundation of quantitative image analysis, aims at partitioning an image into multiple semantic regions, enabling localization and quantification. The result of image segmentation can be a set of sub-regions or contours that collectively cover the entire image. The automatic image segmentation has become one of the main goals of artificial intelligence in medical imaging since manual segmentation is tedious, time-consuming and requires expert knowledge [5]. This technology can benefit an amount of healthcare application such as computer-aided disease diagnosis, treatment planning, lesion quantification, surgery monitoring and navigation, and disease progression [15, 16, 17, 18, 19, 20, 21]. Although, various aspects of segmentation approaches have been reported and reviewed, the problem remains challenging and there is no general successful solution, due to following issues:

- Inhomogeneity: In many cases, anatomical structures are inhomogeneous with respect to spatial repetitiveness of individual pixel/voxel intensities, texture, or grouped co-occurrences of pixels/voxels, which may confuse the segmentation approaches.
- Low contrast: Low contrast medical images bring challenges for algorithms to determine the boundaries of the interested anatomical structures and unrelated background.
- Noise: Noise disturbs uniformity in the intensity range of medical images and increases uncertainty and hence makes the segmentation difficult.
- Shape variability: The significant shape variability of a kind of objects-of-interest in images usually affects the performance of segmentation methods, especially those leveraging shape priors.
- Sample imbalance: The sample imbalance problems in medical image segmentation are normally twofold: (1) the number of voxels (pixels) of one class dominants over other classes; (2) the number of samples with regular distribution (e.g., normal voxels) dominants over that with irregular distribution (e.g., rarely appeared voxels).
- Lack of annotated data: learning-based segmentation methods, especially the supervised-learning methods, are data-driven, it needs big data to train them and prevent them from over-fitting. However, in practice, there is always a lack

of sufficient high-quality annotated data since manually annotation requires special domain knowledge and is time-consuming.

1.1.2 Previous Work

Previous attempts for medical image segmentation can be categorized into the following classes: (1) rule-based segmentation, (2) statistical-inference-based segmentation, (3) deformable-model-based segmentation, (4) atlas-based segmentation, (5) machine-learning-based segmentation with hand-crafted features, (6) deep-learning-based segmentation with automatically learned features.

Rule-based segmentation methods rely on a set of heuristic rules to divide an image into sub-regions. Feature thresholding methods [22] are the most simple and straightforward rule-based segmentation methods. Region-growing and region split-and-merge are two typical rule-based methods that iteratively grow, merge and/or split the current regions in accord with a set of predefined rules after the initial setting of growing seed or partitions [15, 23, 24]. **statistical-inference-based segmentation** formats the problem with parametric or nonparametric probability models of appearance and shape of target objects together with the corresponding optimization such as Bayesian or maximum likelihood inference [15, 25]. Popular nonparametric probability is built using the k-nearest neighbor and Parzen-window estimators [26]. Popular parametric models employ analytical representations that allow for computationally optimal numerical parameter searching such as the maximum likelihood estimates with the Gaussian model and expectation-maximization (EM) techniques with the Gaussian mixture model [15, 27]. **Deformable-model-based segmentation** methods incorporate both the shapes and appearances of the target objects as discriminative features. The involved deformable model is a curve in a 2-D image or a surface in a 3-D image to outline a target object. It propagates an evolution towards the object boundary under the constraint of keeping the evolved curve smooth and unified. Typical deformable models can be divided into two classes: parametric [28, 29] and geometric [30, 31]. The main limitation of these above-mentioned rule-based or unsupervised model-based segmentation method is that they cannot adapt well on sophisticated tasks and many of these methods are inaccurate, time-consuming and sensitive to the initialization.

Atlas-based segmentation approaches utilize atlases (image-label pairs) to predict the segmentation of the test image. Given a test image, atlas images are separately registered to the given unlabeled image. Then the corresponding atlas labels are transformed with the resulting registration transforms and fused to provide an estimated label [32]. The main issue of atlas-based segmentation algorithms is the

high computation cost and sensitivity to registration accuracy.

Machine-learning-based segmentation with hand-crafted features. Segmentation can be formulated as an optimization issue to find the best shape model fitting the target image evaluated with defined similarity measures or as a pixel/voxel by pixel/voxel classification problem. Conventional learning-based segmentation methods leverage hand-crafted features including intensity, texture, and context features together with machine learning technologies such as support vector machine [33], random forest [34], vantage point forest [35], dictionary learning [36] for medical image segmentation. However, all these methods rely on well-designed features, which require careful feature engineering and expert domain knowledge.

Deep-learning-based segmentation with automatically learnt features. The recent developments in deep learning dramatically improve the performance of the medical image segmentation. Segmentation approach based on the convolutional neural network (CNN) become popular. CNNs have the advantage of extracting informative feature representations automatically and effectively from image instead of relying on hand-crafted features [37]. fully convolutional network (FCN) [38] and its variants [39, 40, 41] are a widely used architecture in medical image segmentation. Another popular kind of architectures are encoder-decoder based that consists of a down-sampling path, an up-sampling path and skip-connections. U-Net [42] for 2D image, 3D U-Net/V-Net [43, 44] and their variants [45, 46, 47, 48] belong to this category.

1.2 Medical Image Classification

Classification is also known as computer-aided diagnosis. Receiving the right diagnosis is the first step leading to appropriate care [49]. Even when there is adequate access to therapies, time to examine patients, and clinical professionals, the diagnostic error occurs and is not limited to rare conditions. For instance, cardiac chest pain, tuberculosis, dysentery, and complications of childbirth are commonly not detected [1]. With the increasing collected data during routine care, machine-learning-based methods have been used to computer-aided diagnosis [50, 51, 52]. However, conventional machine-learning-based methods have limitations due to their requirement of feature engineering. In practice, less skilled clinicians may not elicit the information necessary for a model to assist them meaningfully.

Deep neural networks, especially the convolutional neural networks, can learn informative features automatically from medical image data, which boosts the development of computed aided diagnosis recently. Researchers from Google Deepmind reported the success of leveraging deep learning to make a diagnosis recommendation

for patients with sight-threatening retinal diseases based on three-dimensional optical coherence tomography scans. The proposed deep neural network achieved experts-exceeding performance after training on a dataset with 14,884 scans [53]. Cheng et at. proposed a deep-learning-based denoising auto-encoder for the differentiation of breast ultrasound lesions and lung CT nodules, which resulted in state-of-the-art performance and showed noise tolerance advantage [54]. Deep learning has also been introduced into the diagnosis of the Alzheimer's disease (AD) based on magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI) scans. For pathology images, deep learning has also been leveraged in the prediction of microsatellite instability [10] and the diagnosis of the prostate cancer [13].

Although deep learning has achieved preliminary success in many medical image classification tasks, There are still challenges that need to be solver to fulfill the potential of deep learning in computer-aided diagnosis. One of these challenges is that deep-learning-based classification approaches require a large amount of data to minimize overfitting and improve the performances, however, it is normally difficult to achieve these big medical image datasets in practice, especially when addressing low-incidence serious diseases. Besides, the medical images are always with high dimensions (3D for whole-body CT/PET/MRI scans) or large size (whole slide images), which poses challenges for computing resources, memory, and the optimization of the network. Additionally, increasing concern regarding the interpretability of deep neural networks has been raised currently and the black-box nature of deep neural networks hinders the acceptance of them in the clinical community. Thus, more interpretable, computing efficient, and training-data-saving deep-learning-based classification strategies are needed to deal with the above-mentioned challenges.

1.3 Summary of Contributions

This thesis is set in the context of medical image segmentation and classification based on the deep neural networks. We have focused on two challenging medical image segmentation tasks, i.e, the segmentation of small organs based on CT scans, the characterization of prostate cancer lesions based on the ^{68}Ga-Prostate specific-membrane antigen(PSMA)-11 PET/CT, and three challenging medical image classification tasks, i.e, lymph node metastasis prediction based on the pathological slides from patients with colorectal cancer, the differential diagnosis of parkinsonian syndrome based on the ^{18}F-fluorodeoxyglucose PET scans, and the prediction of microsatellite instability from histopathology images. Different deep learning frameworks and strategies were proposed and employed to solve the above tasks.

In the following, we give a brief introduction to the setting of each publication-based chapter and summarize its content and contributions.

Chapter 3: Knowledge-aided Convolutional Neural Network for Small Organ Segmentation

Accurate and automatic organ segmentation plays an important role in diverse medical applications, including computer-aided diagnosis, computer-aided interventions and radiotherapy planning [55, 56]. Multiple approaches have been proposed for organ segmentation and obtained high accuracy when segmenting large organs such as the liver, lungs and kidneys. For these organs, state-of-the-art methods achieve good performance with Dice similarity coefficients (DSCs) of $> 90\%$ [18, 57, 58, 59]. However, for small organs such as the gallbladder, pancreas, and adrenal glands, accurate segmentation remains challenging due to their limited fraction in the image, high anatomical variability and inhomogeneity.

For image segmentation, deep classification networks with sliding patches during segmentation have been used initially [60], which leads to redundant computations and long inference times. Fully convolutional network (FCN) was later developed to realize semantic segmentation using fully convolution layers, deconvolution layers and skip architecture [38]. FCN-like networks were first applied to medical image segmentation in [42], the proposed U-Net obtained competitive performance on neuronal structure and cell segmentation. Subsequently, more and more FCN based methods have been proposed and gained significant success in different medical image segmentation problems [43, 58, 61]. Although deep-learning-based methods are able to reach impressive segmentation performance when trained on sufficient data, medical data is often scarce, as it is difficult to obtain, and needs to be labeled by experts. Moreover, in CNNs, there is a trade-off between a large receptive field and accurate voxel-wise segmentation. For example, the use of more pooling layers increases the receptive field (and consequently the model capacity), but also leads to loss of small-scale details in the image [62, 63]. It should be noted that traditional methods do not suffer from this trade-off since they work with the voxel-information kept intact.

Besides the deep-learning-based segmentation method, two kinds of traditional methods have been widely used in medical image segmentation: multi-atlas segmentation (MAS) methods and forest-based methods. MAS utilizes atlases (image-label pairs) to predict the segmentation of the target image. Given a test image, all atlas images are separately registered to the unlabeled target image. Then the corresponding atlas labels are transformed with the resulting registration transforms and fused to provide an estimated label. MAS has proven to be a successful tool due to its robust

performance in different anatomical structures and its applicability to relatively small training datasets [32, 64, 65, 66, 67]. However, they are usually time-consuming since each atlas image has to be non-linearly registered to the test image. Alternatively, forest-based methods such as atlas forests [68], random forests [34] and vantage point forests (VPF) [35] usually employ contextual features with tree-based classifier to do voxel-wise segmentation. For instance, VPF utilizes local binary patterns (LBP) and the binary robust independent elementary feature (BRIEF) [69, 35] as features and reaches state-of-the-art accuracy on large abdominal organs segmentation. In [70], the authors improved the vantage point forests by including the regional context to segment the small organs. This method outperforms traditional VPF.

We contribute an end-to-end *knowledge-aided* convolutional neural network (KaCNN) combining the effort of both deep learning and traditional methods to enhance the segmentation performance of small organs on limited training data. On the one hand, the traditional part can be seen as offering complementary knowledge to the deep neural network, for example, the contextual information which cannot be easily obtained within the limited field of view of the CNN without using contracting pooling layers. On the other hand, the deep neural network can be seen as refining the result of the traditional part. As segmentation of small organs on limited training data is a challenging task, we propose to segment these organs with cascaded localization and segmentation steps. In the localization stage, images are first registered to a common space and a bounding box identifies the region of interest (ROI) for the more refined segmentation step. In the second stage, KaCNN is used to predict a segmentation for the ROI. Finally, the obtained segmentation is transformed back to its original space as the final segmentation result.

Chapter 4: Deep Neural Network for Automatic Characterization of Lesions on ^{68}Ga-PSMA PET/CT Images

Prostate cancer (PC) is one of the most common cancers worldwide [71] and has become the third most frequent cause of cancer-related mortality among men in developed countries [72]. By introducing serum PSA-levels as a screening tool, PC is usually diagnosed in the early stage, with a 5-year survival rate of almost 100% for local disease. If not detected and treated at stage 1, when the cancer is confined to the prostate gland, malignant tumour cells may spread to other regions by invading the hematic and lymphatic systems, whereupon the 5-year survival rate declines to 29% for metastatic PC [73, 74]. Chemotherapy imparts some increase in survival in patients with treat advanced metastatic PC, but current treatments are not curative [75].

Prostate specific-membrane antigen (PSMA) is a type II transmembrane glycoprotein, which is constitutively expressed by normal prostate cells and significantly upregulated in prostate cancer cells [72]. Internalization of PSMA can concentrate bound ligands within the cancer cell, thus presenting a mechanism for targeted radiotherapy. Indeed, PSMA has emerged as a major target for the theranostic approach [76]. In various studies, diagnostic positron emission tomography (PET) imaging with the PSMA ligand ^{68}Ga-PSMA-11 is followed by treatment with ^{131}I, ^{177}Lu, ^{213}Bi and ^{225}Ac labelled PSMA-ligands for therapy in PC [77, 78, 79, 80, 81]. The efficacy of ^{177}Lu-PSMA-617 has been recently validated in a phase II clinical trial [82].

Despite the encouraging early results for PSMA-targeted radioligand therapy (RLT), treatment planning of this novel therapy is very challenging compared to planning for conventional external beam radiotherapy, due to abundance and systemic spread of the lesions. Indeed, RLT proved to be suboptimal for 30% of a group treated PC patients [83]. Therefore, there is a need for improved treatment planning to optimize the RLT outcome. A critical step for treatment planning is to assess with some accuracy the tumour burden, which necessarily entails detection and segmentation of the lesions to diagnostic PET. Usually, patients who undergo PSMA-targeted RLT have a high number of metastases. Therefore, time-consuming manual segmentation methods are impractical in routine practice. A first approach towards a semiautomatic segmentation method was developed to characterize the tumour burden of bone metastases, namely the bone-PET-index (BPI), on ^{68}Ga-PSMA-11 PET/CT images. This procedure informed the planning of ^{223}Ra-dichloride therapy by segmentation of osseous lesions using an SUV-based threshold on the PET image, with masking of the skeleton based on the CT images [84]. However, this method does not generalize to other types of lesions such as lymph node metastases, where prior anatomical information is more difficult to obtain. Indeed, it is extremely challenging to segment a high number of PSMA-positive lesions of heterogeneous size and tracer uptake, with distribution in a variety of anatomical contexts with different background activity. Until now, there are no successful computer-aided methods to evaluate tumour load for the treatment planning of PSMA-targeted RLT.

In this chapter, we developed 3D deep supervised residual U-Net (DS-Res-U-Net) to automatically characterize local recurrence, bone lesions, and lymph node metastasis synchronously. For proof-of-concept, we focused on the detection of lesions in the pelvic area. We tested the proposed deep learning method on a dataset of PSMA PET-CT scans collected from three different centers.

Chapter 5: Predicting Lymph Node Metastasis Based on Multiple Instance Learning with Deep Graph Convolution

Colorectal cancer (CRC) remains the third most common malignancy and is a leading cause of cancer-related mortality in the world, whereas the overall outcomes have been improved due to the development of new cancer treatment and management [85]. Lymph node metastasis (LNM) from colorectal cancer is a major factor in patient management and prognosis [86, 87, 88]. Patients diagnosed with LNM should undergo lymph node dissection surrounding the colon region [89]. Since most patients have polyp biopsy during the colonoscopy exam, the prediction of LNM from polyp biopsy has great clinical value and can potentially detect the metastasis early and prevent it from further spreading. However, the prediction of LNM status remains challenging and there is a lack of knowledge that indicates useful features for LNM prediction. Therefore, an approach that can automatically learn the informative features from the pathological image is required for the prediction of LNM.

Pathology plays a role of the mainstay in modern medicine, especially cancer treatment, where pathology image analysis is the golden standard for diagnosis. With the development of electronic health record (EHR) technologies, the glass slides can be digitized into whole slide images (WSIs) using digital slide scanners, which paves a way for introducing computer-aided procedures like deep learning-based AI system in the pathology analysis field. However, an obstacle that cannot be ignored is that the size of a whole slide image (WSI) is usually very large (around 100000×50000 pixels in our case). Given the current computational resource, it is infeasible to load the WSI into the deep neural networks. The multiple instance learning (MIL) offers a suitable and effective way to solve this problem. To meet the formalization of the MIL, A WSI is usually divided into a set of image patches (for instance 512×512 pixels) and then the WSI is regarded as a bag containing multiple patches and each patch is regarded an instance in MIL. Multi-instance learning is a typical weakly-supervised learning [90], which has been widely employed in different tasks, including object detection [91, 92, 93], semantic segmentation [94, 95], scene classification [96, 97], medical diagnosis [13, 98], etc. In the MIL task, the training dataset is composed of bags, where each bag contains a set of instances. The goal of MIL is to learn a model for predicting the bag label. Different from conventional fully-supervised machine learning problems, where each instance has a confident label, only the bag-level label is available in MIL. Furthermore, instances in a bag are not necessarily relevant, sometimes even providing confusing information. For example, some instances do not contain discriminative information related to its bag class, or they are more related to other classes of bags [99].

Based on which level the discriminative information is at (instance-level or bag-level) and how the relevant information is extracted (implicitly or explicitly), MIL algorithms can be categorized into three groups, i.e., instance-space paradigm, bag-space paradigm, and embedded-space paradigm [99]. The instance-space paradigm tends to focus on local information, which learns instance classifier at the first stage and then achieves the bag-level classifier by simply aggregating instance-level results. These instance-space methods are mostly based on the standard multiple instance (SMI) assumption [100], i.e., a bag is positive only if it contains at least one positive instance and otherwise is negative [101, 102, 103]. However, this key-instance based SMI assumption is inappropriate in our application where the classification should be based on the global bag information instead of an individual instance. The bag-space paradigm and embedded-space paradigm, on the other hand, extract discriminative information from the whole bag. The difference between these two paradigms lies in the way to exploit the bag-level information. The bag-space paradigm implicitly utilizes bag-to-bag distance/similarity, while the embedded-space paradigm explicitly embeds the information of a bag into a feature space.

In this chapter, we outline the potential of deep learning in discovering the characteristic pattern for the differential diagnosis of the LNM. An AI system based on the deep neural network and embedded-space multiple instance learning was built for automatic predicting the patient with LNM positive or negative.

Chapter 6: A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes

The idiopathic Parkinson's disease (IPD) is one of the most common age-related neurodegenerative disorders affecting more than 6 million people worldwide [104]. The symptoms and signs of idiopathic Parkinson's disease are usually similar to those of atypical parkinsonian syndromes such as multiple system atrophy (MSA) and progressive supranuclear palsy (PSP), especially in the early stages of disease [105]. Pathological studies report that approximately 20-30% of patients misdiagnosed as having IPD turn out to have either MSA or PSP [105, 106]. The precise diagnosis is necessary in clinical practice since the misdiagnosis can lead to significant consequences for clinical patient care and research trials [107, 108]. The accurate differentiation of parkinsonian disorders is important for the determination of therapy strategies and the management of the disease.

Positron emission tomography (PET) using ^{18}F-fluorodeoxyglucose (FDG) can detect a wide spectrum of neurobiological abnormalities and is reported to have

advantages in differential diagnosis of parkinsonism before structural damage to brain tissue [109]. Principal component analysis (PCA) was applied to extract PD-related pattern (PDRP), MSA-related pattern (MSARP), and PSP-related pattern (PSPRP). These patterns, which were used as features for a machine learning method of logistic regression, have been found as effective surrogates to discriminate between classical PD, atypical parkinsonian syndromes and healthy control subjects [110]. However, the PCA decomposition takes the 3D image volume of a subject as a squeezed 1D vector and the high-level spatial interrelation is not considered any more during the pattern extraction.

Compared to conventional handcrafted feature extraction, the emerging of deep learning moves an advanced step forward to discover new characteristic features in data automatically and effectively [2, 9]. Previous attempts to introduce deep learning in PD diagnosis were based on either morphological imaging (MRI or ultrasound) or very specific dopamine transporter (DaT) SPECT. The high selectivity of DaT SPECT in striatum restricts the space for AI in the development of imaging-based biomarkers and has been only applied for the differentiation between PD and normal subjects [111, 112]. None of the previous studies were designed to extract useful biomarkers based on the 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) to support the differential diagnosis of parkinsonism.

In this chapter, we extensively explore the potential of deep learning on ^{18}F-FDG PET imaging, which is sensitive to a wide spectrum of neurobiological abnormalities. A 3D deep residual convolutional neural network was built for automatic classification of IPD, MSA, and PSP. Additionally, we utilize the saliency map to afford a pathophysiological insight into the decision mechanism of the deep learning method.

Chapter 7: Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer

Microsatellite instability (MSI) is a hypermutator phenotype that occurs in tumors with DNA mismatch repair deficiency (dMMR) [113], which is reported as a hallmark of hereditary Lynch syndrome (LS)-associated cancers [114] and observed in about 15% of colorectal cancer (CRC) [115]. Microsatellite instability (MSI) has been recently approved by the U.S. Food and Drug Administration (FDA) as a favorable predictor of anti-PD-1 immunotherapy in pan-cancer and it is also a prognostic factor for colorectal cancer (CRC). However, current MSI identification methods are not available for all patients, because it requires additional genetic or immunohistochemical tests which are costly and time-consuming.

In clinical practice, pathology slides are produced for almost every patient diagnosed with cancer, which can be digitized into whole slide images (WSIs) [11]. WSI not only reveals the tissue spatial arrangement of tumor cells at low magnification, but also the cell structure at high magnification [12]. Furthermore, histopathology images also show the immunologic microenvironment of tumors [13]. The cell level phenotypes presented in WSI are affected by genotypes such as MSI at the molecular scale. Therefore, learning informative features from pathology slides may provide an opportunity for the detection of MSI.

In this study, we developed a multiple-instance-learning (MIL)-based deep learning model to predict MS status from histopathology images, and utilize transfer learning for model fine-tuning across different populations to improve its generalizability. Furthermore, we designed an interpretable analysis pipeline to link the image phenotypes to genotypes.

1.4 Organization

This is a publication-based thesis with the following structure: Chapter 1 introduces to the topic of deep learning, medical image segmentation and classification along with current challenges, and summarizes our contributions. Chapter 2 gives a brief summary of relevant terminology and key concepts from the deep neural network and convolutional neural network, which are used throughout this manuscript.

Chapter 3 to 7 are composed of five publications [116, 117, 118, 119, 120] in their original form. They have been published as peer-reviewed journals and conference proceedings, and are therefore self-contained. Each of these chapters starts with a brief summary, containing the full citation of the original publication, a short synopsis introducing the content of the corresponding publication, and the author's contributions.

Chapter 8 offers discussion and conclusions over the presented material and suggest directions for future work. Finally, a complete list of publications that have been written during the time period of this doctoral thesis can be found in Appendix A.

2

Background

The main themes of this thesis consist of the deep neural network and the convolutional neural network. This chapter aims at giving a brief summary of the key concepts and notation used throughout this thesis, but it is not intended to be a representative overview of the most important concepts of each field. For a more complete and in-depth discussion on deep neural network (DNN) and convolutional neural network (CNN), please refer to [2, 37, 121].

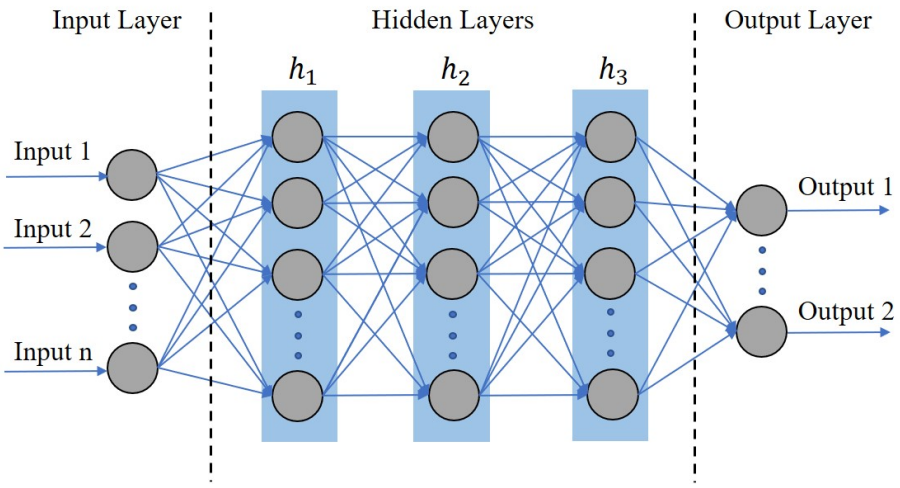


Figure 2.1: A typical architecture of neural network. It consists of input layer, hidden layers, and output layer. These layers are composed of a number of connected computational units called neurons

2.1 Neural Network

Neural networks are composed of a number of connected computational units, which are called neurons and organized in layers. As shown in Fig. 2.1, a typical neural network consists of an input layer where data enters the network, hidden layers transforming the data as it flows through, and an output layer producing results. The network is trained to produce useful predictions through recognizing informative patterns in training data set, supervised by comparing the predicted results to the actual labels under the format of an objective function [5]. Each neuron in the network can be defined as follows:

$$\hat{f}(\mathbf{x}) = h(\mathbf{w}^{\mathbf{T}}\mathbf{x} + b) \tag{2.1}$$

where $\mathbf{x}$ represents the input vector, $\mathbf{w} = (w_1, \cdots, w_n)$ is the weight vector and b is the bias. $h(\cdot)$ is the non-linear activation function. The commonly used activation functions are:

(1) Sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

(2) Tanh

$$\sigma(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.3}$$

(3) Rectified Linear Unit (ReLU)

$$\sigma(x) = \max(0, x) \tag{2.4}$$

(4) Leaky ReLU

$$\sigma(x) = \max(0.1x, x) \tag{2.5}$$

(5) Maxout

$$\sigma(x) = \max(w_1^{\mathbf{T}}x + b_1, w_2^{\mathbf{T}}x + b_2) \tag{2.6}$$

(6) Exponentional Linear Unit (ELU)

$$\sigma(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^{x} - 1), & x < 0 \end{cases} \tag{2.7}$$

Combining all the neurons in one layer, a single layer network can approximate any continuous function $\hat{f}(x)$ on a compact subset of $\mathbb{R}^n$, which can be formatted as a linear combination of N individual neurons as follows:

$$\hat{f}(\mathbf{x}) = \sum_{i=0}^{N-1} v_i h(\mathbf{w}_i^{\mathbf{T}}\mathbf{x} + b_i) \tag{2.8}$$

where v_i is the combination weights between neurons. We can summarize all trainable parameters of the network as:

$$\theta = (v_0, b_0, \mathbf{w_0}, \cdots, v_N, b_N, \mathbf{w_N})^{\mathbf{T}}. \tag{2.9}$$

In order to increase the model capacity and representation ability, we can introduce more hidden layers into the network between the input layer and output layer as shown in Fig. 2.1. These layers are connected between each other to be a deep neural network. One theory of neural networks is that both shallow and deep networks are able to approximate arbitrarily well any continuous function on a compact domain. However, the deep networks can approximate the class of compositional functions with the same accuracy as shallow networks but with an exponentially lower number of training parameters as well as VC-dimension and deep architecture also has benefits for feature representation such as the recombination of the weights along different paths and re-using latent features [4, 122]. A neural network with a number of layers can be defined as:

$$\begin{aligned} \hat{f}(\mathbf{x}; \mathbf{\Theta}) &= (f_m \circ \cdots \circ f_1)(\mathbf{x}) \\ &= h^m \left(h^{m-1} \left(\cdots \left(h^2 \left(h^1(\mathbf{w}_1^{\mathbf{T}}\mathbf{x} + b_1\right) + b_2\right) + b_{m-1}\right) + b_m\right) \end{aligned} \tag{2.10}$$

where $\Theta = \{\mathbf{w}_1, \cdots, \mathbf{w}_m, b_1, \cdots, b_m\}$ is the parameter set.

For neural networks, the parameters Θ are learned in the training phase from the data. The parameter learning phase can be formulated as an optimization problem of minimizing the error between the prediction and the ground truth called loss function. The optimization problem is nonlinear and nonconvex, and there is no analytic solution for the parameter set Θ. Therefore, the gradient descent algorithm is utilized to learn the parameters iteratively. In neural networks, the back-propagation strategy [123] can efficiently evaluate the gradient and then the parameters Θ can be updated as:

$$\mathbf{\Theta}^{(\tau+\mathbf{1})} = \Theta^{(\tau)} - \eta \nabla \mathbb{E}(\Theta^{(\tau)}). \tag{2.11}$$

where η is the learning rate, τ represents the iteration index, and $\mathbb{E}$ is the loss function.

2.2 Convolutional Neural Networks

The convolutional neural network is a specially designed deep neural network for address image data, which is recently spotlighted in computer vision tasks including medical image segmentation and classification. It can leverage spatial and structural information in 2D or 3D images of its input and it benefits from mechanisms of the

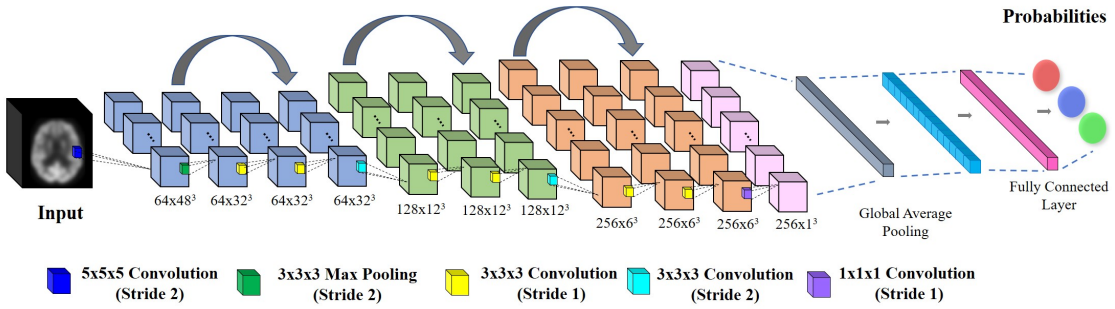


Figure 2.2: A sample segmentation convolutional netwrok architecture. It consists of convolutional layers, pooling layers, fully connected layers and skip-connections.

local receptive field, weight sharing and sub-sampling [3]. Typical image classification network consists of convolutional layers, pooling layers, fully connected layers and skip-connection strategy. Besides these layers, typical image segmentation network usually also includes up-sampling layers to up-sample feature map for generating final segmentation maps. For image classification, the milestone architectures include AlexNet [124], VGGNet [125], Inception [126], ResNet [127], and DenseNet [128]. And, for image segmentation, the popular network architectures are the fully convolutional network (FCN) [38], U-Net [42], 3D U-Net/V-Net [43, 44] and their variants [45, 46, 47, 48]. Currently, the above architectures are widely-used as the cornerstone to design specific solutions. Fig. 2.2 illustrates the structure of a sample image classification convolutional neural network and Fig. 2.3 demonstrates a sample image segmentation convolutional neural network.

2.2.1 Convolutional layer

Convolutional layers play a central role in building the convolutional neural networks. The convolutional layer detects local features at different positions from the previous layer and maps the learned information into a new feature map. During the convolution in the networks, the input from the previous layer is split into perceptrons, creating local receptive fields and finally compressing the perceptrons in feature maps, which indicate the locations and strength of a detected feature in an input.

In l^{th} layer, assuming there is a group of N^l filters, each filter detects a particular feature at every position on the input. The output of layer l denoted as $Y^{(l)}$ will consists of N^l feature maps, where the i^{th} feature map $Y_i^{(l)}$ can be computed as:

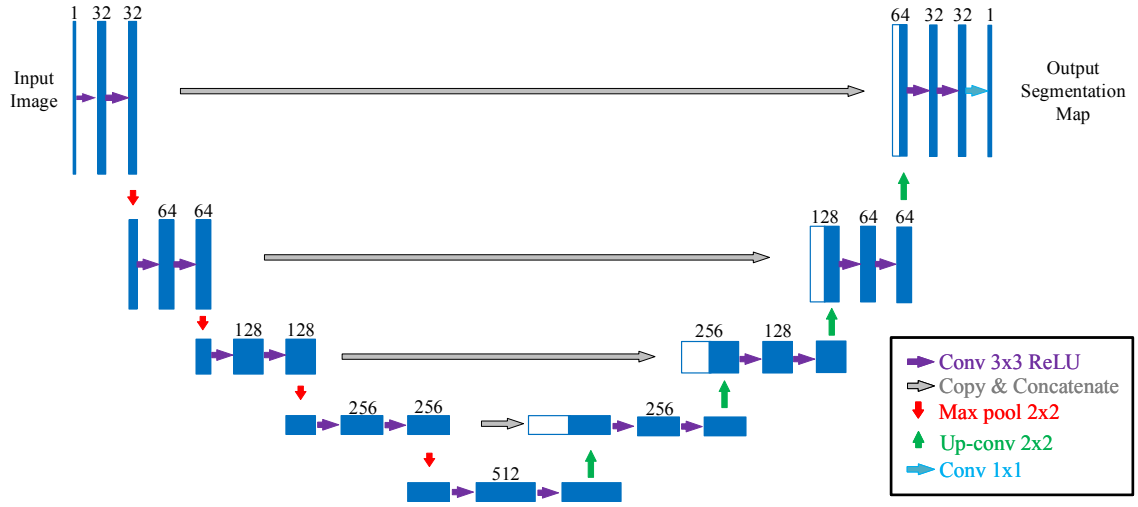


Figure 2.3: A sample classification convolutional netwrok architecture: U-Net. Different operations are denoted by different arrows. The multi-channel feature maps are shown in blue and the copied feature maps are shown in write. The digit above the feature maps denotes the number of channels.

$$Y_i^{(l)} = h\left(\sum_{j=1}^{N^{(l-1)}} K_{i,j}^{(l)} * Y_j^{(l-1)} + B_i^{(l)}\right) \tag{2.12}$$

Where $K_{i,j}^{(l)}$ is the applied convolutional kernel, $B_i^{(l)}$ is a bias matrix, h is the activation function illustrated in section 2.1. Currently, there are multiple attempts to improve the convolutional layer and detailed informance can be found in [129] for deformable convolution, [130] for depth-wise separable convolution, and [131] for the dilated convolution.

2.2.2 Pooling layer

The pooling layer in the convolutional network aims at downsampling the feature maps to reduce the amount of parameters and computation cost in the network, which is also beneficial for controlling overfitting. Pooling operations take a small region with a defined size as the input and output a single number for representing this region. The representative value of the receptive field is usually computed by employing the max function called max-pooling or the average function called average

pooling. In the convolutional neural network, another approach for achieving the downsampling effect of pooling is using convolutions with stride [3, 5].

2.2.3 Up-sampling layer

Up-sampling layers are widely used in image segmentation networks to up-sample the input feature map to a higher resolution. The first approach is **re-sampling and interpolation** that re-scale an input feature map to the desired size with the interpolation method such as bilinear interpolation. Another approach is **unpooling** [132], regarded as the reverse of pooling. In the unpooling layer, an approximate inverse of the previous pooling layer is obtained by recording the position of each maximum activation value within each pooling region and this position information was then used to place the reconstructions of the previous layer into right locations. The third approach is **transpose convolution** [133]. The transpose convolution is regarded as the reverse of the convolution operation but it is not actually a proper mathematical deconvolution. In the transpose convolution, the kernel is placed over the input and values of the input are multiplied successively by the kernel weights for producing the up-sampled result.

2.2.4 Skip connection

The skip-connections used in the proposed network introduce connections that skip one or more layers. They are helpful for simplifying the optimization of a network. Deeper models tend to hit obstacles during the training process. The gradient signal vanishes with increasing network depth. But the skip-connections propagate the gradient throughout the model, which can alleviate the vanishing gradient problem [127, 128]. In segmentation networks with encoder-decoder architecture, The skip-connections are usually utilized between each decoder-encoder pair brings the features with higher spatial resolution from shallow layers of the encoder part directly to the layers of the decoder part for detection and segmentation.

2.2.5 Loss function

The loss function supervises the training of the networks. In the image classification task, widely-used loss functions include the categorical cross-entropy loss and Focal loss [134]. Focal loss is a cross-entropy loss that weighs the contribution of each sample to the loss based on the classification error to tackle the class imbalance issue.

The categorical cross-entropy loss is defined as:

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\delta(y_i = c)\log(P(y_i = c)) \tag{2.13}$$

where N denotes the data number and C represents the categories number. $\delta(y_i = c)$ is the indicator function and $P(y_i = c)$ is the predicted probability by the model. The Focal loss is defined as:

$$L_{FL} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\delta(y_i = c)\left(1 - P(y_i = c)\right)^{\gamma}\log(P(y_i = c)) \tag{2.14}$$

where γ is the focusing parameter to control the rate of down-weighting the easy examples. When $\gamma = 0$, L_{FL} is equivalent to L_{CE}.

Typical loss functions for image segmentation consist of above-mentioned Cross-Entropy loss and Focal loss, as well as, Dice loss, Tversky loss, shape aware loss, and boundary loss. Assume $p(x_i)$ is the prediction probability of a voxel x_i and $g(x_i)$ is the corresponding ground truth at the same voxel. The Dice loss is given by

$$L_{Dice}(\mathbf{X}) = -\frac{2\sum_{x_i\in\mathbf{X}} p(x_i)g(x_i) + \varepsilon}{\sum_{x_i\in\mathbf{X}} p(x_i) + \sum_{x_i\in\mathbf{X}} g(x_i) + \varepsilon} \tag{2.15}$$

where $\mathbf{X}$ is the training images, ε is a small term to prevent the loss function from the issue of dividing by 0. The formulations of the cross-entropy loss and Focal loss are given in (2.13) and (2.14) and detailed information for other segmentation loss functions can be found in [135] and [136].

3

Knowledge-aided Convolutional Neural Network for Small Organ Segmentation

This chapter has been published as **peer-reviewed journal paper:**.

Y. Zhao, H. Li, S. Wan, A. Sekuboyina, X. Hu, G. Tetteh, M. Piraud, and B. Menze. "Knowledge-aided convolutional neural network for small organ segmentation." In: *IEEE journal of biomedical and health informatics* 23.4 (2019), pp. 1363–1373. DOI: 10.1109/JBHI.2019.2891526

Synopsis: This work deals with the problem of small organ segmentation with limited training data. We introduce an automatic approach including two cascaded steps — localization and segmentation. The localization stage involves the extraction of the region of interest after the registration of images to a common template and during the segmentation stage, a novel *knowledge-aided* convolutional neural network is proposed to improve segmentation accuracy.

Contributions of thesis author: algorithm design and implementation, computational experiments and composition of manuscript.

Accepted manuscript for the IEEE (IEEE journal of biomedical and health informatics, 2019)
Published version: DOI: 10.1109/JBHI.2019.2891526

Knowledge-aided Convolutional Neural Network for Small Organ Segmentation

Yu Zhao, Hongwei Li, Shaohua Wan*, Anjany Sekuboyina, Xiaobin Hu, Giles Tetteh, Marie Piraud and Bjoern Menze

***Abstract*—Accurate and automatic organ segmentation is critical for computer-aided analysis towards clinical decision support and treatment planning. State-of-the-art approaches have achieved remarkable segmentation accuracy on large organs such as the liver and kidneys. However, most of these methods do not perform well on small organs such as the pancreas, gallbladder and adrenal glands, especially when lacking sufficient training data. This paper presents an automatic approach for small organ segmentation with limited training data using two cascaded steps — localization and segmentation. The localization stage involves the extraction of the region of interest after the registration of images to a common template and during the segmentation stage, a voxel-wise label map of the extracted region of interest is obtained and then transformed back to the original space. In the localization step, we propose to utilize a graph-based groupwise image registration method to build the template for registration so as to minimize the potential bias and avoid getting a fuzzy template. More importantly, a novel *knowledge-aided* convolutional neural network is proposed to improve segmentation accuracy in the second stage. This proposed network is flexible and can combine the effort of both deep learning and traditional methods, consequently achieving better segmentation relative to either of individual methods. The ISBI 2015 VISCERAL challenge dataset is used to evaluate the presented approach. Experimental results demonstrate that the proposed method outperforms cutting-edge deep learning approaches, traditional forest-based approaches and multi-atlas approaches in the segmentation of small organs.**

***Index Terms*—medical image segmentation, convolutional neural networks, knowledge-aided, deep learning**

I. Introduction

Cancer in advanced stage usually metastasizes from where it formed to other parts of the body including related organs. For instance, advanced lymphoma could metastasize to the bone marrow, spleen, or extralymphatic organs [1]. Accurate segmentation of abdominal organs enables the comprehensive measurement of shapes and volumes of target organs, which are the indicators of disorders supporting the clinical decision [2], [3]. Moreover, segmentation of treatment volumes or high-risk organs also plays a central role in radiation treatment planning [2], [4]. In traditional clinical practice, radiologists often manually delineate and segment organs slice-by-slice, which is tedious, time-consuming and prone to intra- and inter-observer variability. Therefore automatic multi-organ segmentation in abdominal scans is an important task. Multiple approaches have been proposed for organ segmentation and obtained remarkable performance with Dice scores of over 90% when segmenting large organs, including the liver, lungs, and kidneys [5]–[7]. However, for small organs such as the pancreas, gallbladder, and adrenal glands, accurate segmentation remains challenging due to their limited fraction of the entire image (class-imbalance issue in segmentation), in-homogeneity, and variation in their size, shape, and appearance among different subjects. Besides, their localization is not as stable as large organs, so prior spatial knowledge on the absolute location within the anatomical reference is not as helpful as for large organs. This paper hence focuses on fully automatic small organ segmentation when limited data is available.

Current work in multi-organ segmentation can be divided into two categories: registration-based and classification-based method. Registration-based methods include statistical shape models (SSM) [8], probabilistic atlases (PA) [9], and multi-altas segmentation techniques (MAS) [6]. Classification-based methods include two groups, i.e., traditional methods which require hand-crafted feature such as forest-based methods [10], [11] and deep convolutional neural networks [12], [13].

SSM approaches work by constructing statistical shape or appearance models, while PA methods incorporate global spatial information and inter-organ spatial relationships during segmentation. However, existing SSM and PA cannot handle large inter-subject variability, and MAS outperforms these two single model/atlas approaches in most cases [14]. MAS techniques utilize atlases (image-label pairs) to predict the segmentation of the test image. Given a test image, atlas images are separately registered to the given unlabeled image. Then the corresponding atlas labels are transformed with the resulting registration transforms and fused to provide an estimated label [15]. MAS has been proven to be a successful tool due to its robust performance in different anatomical structures and its applicability to relatively small training datasets [15]–[18]. However, MAS approaches are time-consuming because it is necessary to have all the atlases available and register the target image with each atlas separately during segmentation time.

Alternatively, classification-based methods such as atlas forests [19], random forests [10] and vantage point forests (VPF) [11] are not plagued by the same problem as registration-based approaches. These methods usually employ local appearance features with the classifier to perform voxel-

Yu Zhao is with the Department of Computer Science, Technische Universität München, Munich, Germany and the School of Electronic and Information Engineering, Beihang University, Beijing, China e-mail: (yu.zhao@tum.de).

Hongwei Li, Anjany Sekuboyina, Xiaobin Hu, Giles Tetteh, Marie Piraud and Bjoern Menze are with the Department of Computer Science, Technische Universität München, Munich, Germany e-mail: (bjoern.menze@tum.de).

Shaohua Wan is with the School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, China e-mail: (shaohua.wan@ieee.org).

The corresponding author is marked with *

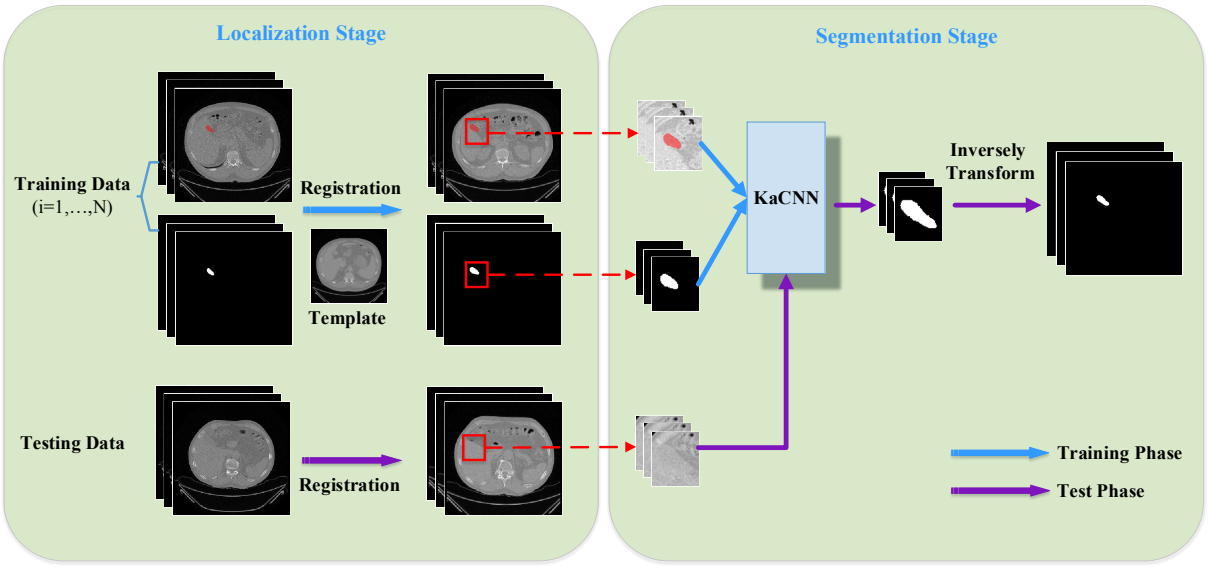


Fig. 1. The overall framework of the proposed approach. It consists of the localization step and the segmentation step. The localization stage involves the extraction of the region of interest after the registration of images to the common template and during the segmentation stage, a voxel-wise label map of the extracted region of interest is obtained by using the proposed KaCNN and then transformed back to the original space.

wise segmentation. For instance, VPF utilizes local binary patterns (LBP) and the binary robust independent elementary feature (BRIEF) [11], [20] as features and reaches high accuracy in the segmentation of large abdominal organs. In [21], the authors improved the vantage point forests by including the regional context to segment the small organs. This method outperforms the traditional VPF. Even though traditional classification-based methods reduce the computational time required in image registration, they face challenges in constructing variability- and deformation-invariant features for characterizing anatomy and their segmentation quality is often inferior to the MAS method.

Recently, convolutional neural networks (CNNs) have achieved significant success in diverse computer vision applications, including object detection, image classification, and segmentation [12], [22]. Unlike conventional methods which usually rely on hand-crafted features, CNNs have the advantage of learning salient feature representations automatically and effectively [23]. For image segmentation, deep classification networks with sliding patches have been used initially [24], which leads to redundant computations and long inference times. Fully convolutional network (FCN) [12] was later developed to perform semantic segmentation using convolution and deconvolution architecture. U-Net [13] equipped with skip-connections was proposed for medical image segmentation. Very recently, various FCN-based methods have been proposed and achieved significant progress in various medical image segmentation problems [25]–[28]. Although deep-learning based methods can reach impressive segmentation performance when trained on sufficient data, medical data is often scarce, as it is difficult to obtain, and needs to be labeled by experts. Moreover, in CNNs, there is a trade-off between a large receptive field and accurate voxel-wise segmentation. For example, the use of more pooling layers increases the receptive field (and consequently the model capacity), but also leads to the loss of spatial information in the image [29].

As segmentation of small organs on limited training data is a challenging task, we propose a new procedure with cascaded localization and segmentation steps and a novel *knowledge-aided* convolutional neural network to enhance the segmentation accuracy. This work exploits the advantages of all the registration-based, traditional classification-based, and deep convolutional neural network while having relatively low computational cost and alleviating the requirement of large training data to train the deep convolutional neural network. In the two-stage procedure, the localization step can decrease the influence of the class-imbalance issue and facilitate the subsequent segmentation step. The detailed contributions of this work are as follows:

- We present a *knowledge-aided* convolutional neural network (KaCNN) that combines both deep-learning and traditional methods to enhance the segmentation performance of small organs on limited training data. From one point of view, the traditional part can be seen as offering complementary knowledge to the deep neural network, (for example, the contextual information which cannot be easily extracted within the limited field of view of the CNN without using contracting pooling layers) and from another point of view, the deep neural network part can be regarded as refining the result of the traditional part.
- We propose to segment these organs with cascaded localization and segmentation steps. In the localization step, we register images to the common template, and a bounding box of the region of interest is obtained for the more refined segmentation step. In the segmentation step, KaCNN is employed to predict a segmentation for the ROI. Finally, the final segmentation of the test image is generated by transforming and fusing the obtained segmentation of ROI back into the original space. The

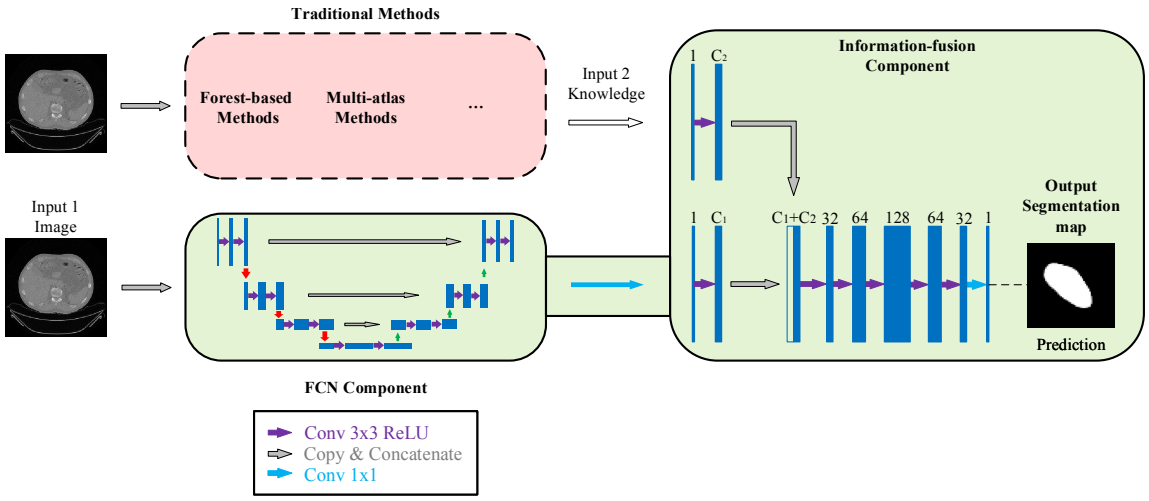


Fig. 2. Proposed *knowledge-aided* convolutional neural network (KaCNN). It consists of a FCN component (left, green background) and an information-fusion component (right, green background). In the information-fusion component, different operations are denoted by different arrows. The multi-channel feature maps are shown in blue and the copied feature maps are shown in write. The digit above the feature maps denotes the number of channels. The architecture of the FCN component is illustrated in Fig. 3.

proposed cascaded framework takes advantage of multi-atlas segmentation but avoids its high computational cost regarding registration during inference. In conventional multi-atlas segmentation method, numerous registration processes are needed for aligning the atlas images to each test image [30]. However, in this proposed framework, we only need to align the test image to the common template and transform the obtained segmentation map back for each given test image during the inference phase. To minimize the potential bias and avoid getting a fuzzy template, we propose to utilize a graph-based groupwise image registration method [31] to build the common template for registration in the localization step.

- We validate the proposed method on the ISBI 2015 VISCERAL challenge dataset [32]. The results indicate that a superior segmentation performance can be achieved than state-of-the-art deep learning approaches, traditional forest-based approaches, and multi-atlas approaches in the segmentation of small organs.

This paper is structured as follows. In section II, we demonstrate details of the localization step as well as the proposed *knowledge-aided* convolutional neural network. Section III presents the dataset, pre-processing strategies, experimental details and evaluation methods. Subsequently, the obtained results are discussed in Section IV. Section V then gives the discussions including effectiveness and future directions of this works. Finally, Section VI concludes the entire work.

II. Method

A. Overview

We present the localization and the segmentation stages and describe their combination resulting in the KaCNN. The overall framework is illustrated in Fig. 1.

During the training phase, all training images and their corresponding labels are first registered to a generated common template (II-B1). Once registered, a bounding box covering the target organ is chosen based on the information from training labels (II-B2). The cropped training images and labels are then used to train the proposed *knowledge-aided* convolutional neural network. During inference, given a test image, we register it to the common template, segment it with the learnt network and then inversely transform the obtained result back to the original space. In this proposed approach, all the training part is addressed offline. In order to further reduce inference time, we opt for the affine registration in this approach, which is computationally efficient and works well for aligning target organs to similar regions.

The architecture of the proposed knowledge-aided neural network is given in Fig. 2, which consists of an FCN component and an information-fusion component. The FCN component outputs a feature map and then the information-fusion component combines it with the input complementary knowledge together to predict the final segmentation result. The FCN component is flexible and can be one of the state-of-the-art segmentation convolutional networks. Moreover, the KaCNN model is designed to combine knowledge provided by different methods, such as the segmentation probability map obtained by forest-based methods, multi-atlas segmentation methods, etc. In this work, we use a variant of the U-Net [13] as the FCN component and use the segmentation probability map offered by the vantage point forests method using LBP and BRIEF features [11] as the complementary knowledge.

The details of each component of the proposed method are described as follows: the template construction and bounding box selection methods in the localization step are presented in II-B1 and II-B2. The FCN component and the information-fusion component of the proposed KaCNN are illustrated in

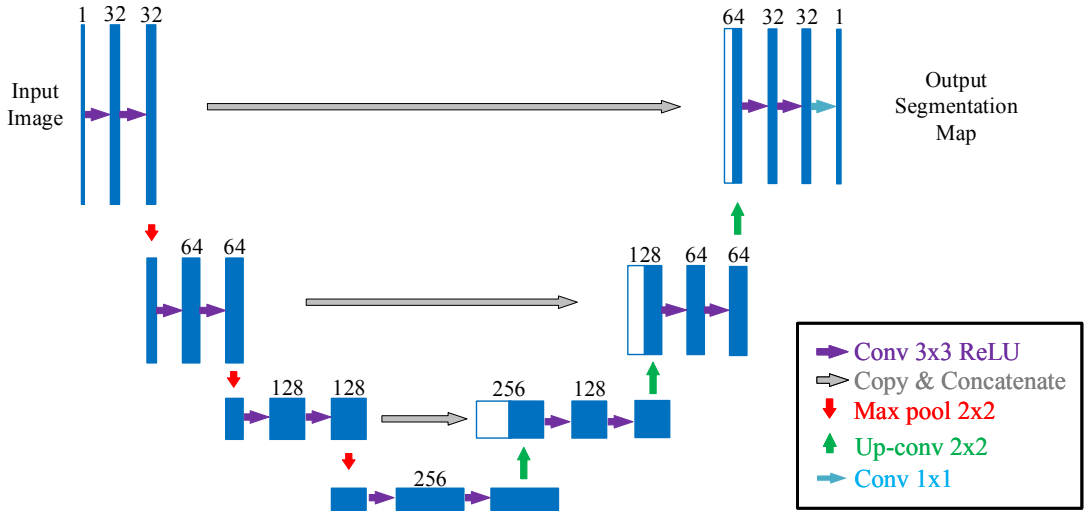


Fig. 3. The FCN component of the Knowledge-aided Convolutional Neural Network (modified U-Net). Different operations are denoted by different arrows. The multi-channel feature maps are shown in blue and the copied feature maps are shown in write. The digit above the feature maps denotes the number of channels.

II-C1 and II-C2 respectively. Finally, the loss function of KaCNN is described in II-C3.

B. Template Construction and Localization

1) Template Construction: It is crucial to find an appropriate template for the following image registration steps. It is common to use the average of a group of images or aligned images as the unbiased common template. However, the average of multiple images (even after registration) is usually fuzzy, since the inter-subject variations are hard to be totally removed via registration. Taking a fuzzy group mean as a template to guide the following registration will undermine the accuracy. In this work, we initially select a subset images which are around the latent group center and apply a graph-based groupwise image registration method [31] to iteratively align these selected images to become close to each other. Finally, we choose the group center image of the aligned images, instead of the average of these images, as the final template, which is unbiased and sharp.

For 3D images, let $\mathbf{U} = \{\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_N\}$ represents the training set containing N images. A subset of M images $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_M\}$ is firstly selected from $\mathbf{U}$ by following step (1) to step (4) and then the selected images are used to construct the template by following step (5) and step (6). The image-selection strategy used here is mainly for reducing computing cost. The details of each step are: (1) Calculate the average image as follows:

$$\mathbf{I}_\mu = \frac{1}{N} \sum\nolimits_{\mathbf{I}_i \in \mathbf{U}} \mathbf{I}_i. \tag{1}$$

(2) Choose the closest image to $\mathbf{I}_\mu$ from $\mathbf{U}$ as $\mathbf{F}_1$ in the subset. The Euclidean distance [33] is used to evaluate the distance between image $\mathbf{I}_i$ and $\mathbf{I}_j$, which is defined as:

$$d(\mathbf{I}_i, \mathbf{I}_j) = \left(\| vec(\mathbf{I}_i) - vec(\mathbf{I}_j) \|_2 \right)^2 \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm, $vec(\mathbf{I}_i)$ and $vec(\mathbf{I}_j)$ are vectorized $\mathbf{I}_i$ and $\mathbf{I}_j$ respectively. (3) Select the consecutive image $\mathbf{F}_{m+1}$ $(m+1 \leq M)$ from the remaining training images in $\mathbf{U}$ based on the distances to the existing images $\mathbf{F}_1, \ldots \mathbf{F}_m$ in $\mathbf{F}$. $\mathbf{F}_{m+1}$ is the result of the following problem:

$$\begin{aligned} &\underset{\mathbf{I}_i}{\operatorname{argmin}} \frac{1}{m} \sum_{\mathbf{I}_j \in \mathbf{F}} d(\mathbf{I}_i, \mathbf{I}_j), \\ &s.t.\ \mathbf{I}_i \in \mathbf{U},\ \mathbf{I}_i \notin \{\mathbf{F}_1, \ldots, \mathbf{F}_m\}. \end{aligned} \tag{3}$$

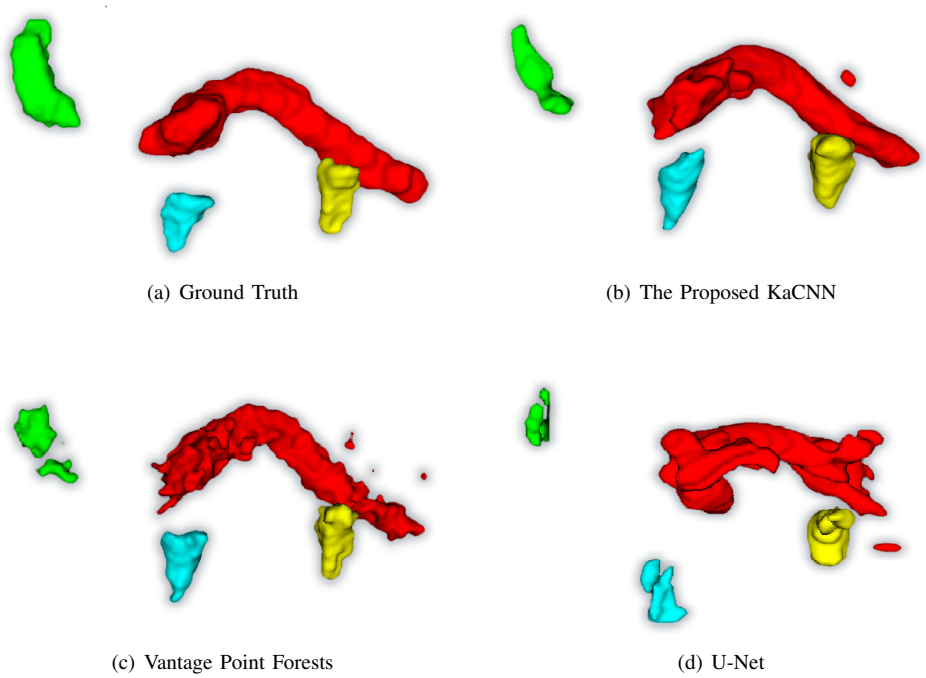
(4) Repeat step (3) until M images $\{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_M\}$ are selected. (5) Apply the graph-based groupwise image registration method [31] to iteratively align $\{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_M\}$ to be closer to each other. (6) Finally, the group center image of the aligned images $\{\mathbf{F'}_1, \mathbf{F'}_2, \ldots, \mathbf{F'}_M\}$ is selected as the template $\mathbf{T}$.

2) Localization: The localization step aims at obtaining roughly the organ region in images, which helps to focus the field of view on the region of interest, avoid over-segmenting similar neighbouring structures, and reduce the computational complexity. Let $\mathbf{A}_1 = (\mathbf{I}_1, \mathbf{L}_1), \ldots, \mathbf{A}_N = (\mathbf{I}_N, \mathbf{L}_N)$ denote N atlases (images and the corresponding labels). $\mathbf{A}_1, \ldots, \mathbf{A}_N$ are first registered to the common template $\mathbf{T}$, the aligned images and labels are denoted as $\mathbf{I'}_1, \ldots, \mathbf{I'}_N$ and $\mathbf{L'}_1, \ldots, \mathbf{L'}_N$. We can obtain the probabilistic atlases by counting the frequency with which each label k occurs at every location $\mathbf{x}_i$ across the N aligned labels as follows:

$$p_k(\mathbf{x}_i) = \frac{\sum_{n=1}^{N} \delta(\mathbf{L'}_n(\mathbf{x}_i) = k)}{N} \tag{4}$$

where $\delta(\cdot)$ is the indicator function, $\mathbf{x}_i$ represents the location of voxel i. Subsequently, we make a bounding box which includes all voxels where $p_k(\mathbf{x}_i) > \Theta$ (Θ is a small positive threshold) as the location of organ k. In order to make the location robust, a security margin is added to the obtained bounding box.

C. Knowledge-aided Convolutional Neural Network



(a) Ground Truth (b) The Proposed KaCNN

(c) Vantage Point Forests (d) U-Net

Fig. 4. 3D organ segmentations obtained by the proposed KaCNN, vantage point forests and U-Net. The pancreas is shown in red, the gallbladder in green, the left adrenal gland in yellow and the right adrenal gland in cyan.

1) Fully-Convolutional-Network Component : Recently, U-Net [13] has become a popular and successful convolutional neural network architecture for medical image segmentation. We utilize a modified U-Net architecture as the FCN component of the proposed KaCNN model. In the U-Net architecture, the number of the down-sampling stages is an important factor influencing the segmentation performance. On the one hand, if the network has too many down-sampling layers, the information of the small object may disappear after one of the deeper down-sampling layers. On the other hand, an appropriate number of down-sampling stages correspond to a large valid receptive field, which helps the network capture more contextual information [29]. After comparing U-Net architectures with varying down-sampling stages (2, 3, and 4), we find that the architecture with three down-sampling layers works better than other architectures giving consideration to the performance of to both the small and large organs in our task. In addition, we also carefully determine the number of channels of each layer. The detailed architecture is shown in Fig. 3, it is composed of a down-sampling path which consists of three repeated encoder stacks and an up-sampling path which consists of three repeated decoder stacks. In each encoder stack, there are two 3×3 convolutions and a 2×2 max pooling operation with stride 2 for down-sampling, where each convolution is followed by a rectified linear unit (ReLU) as the activation function. While each decoder stack consists of a transposed convolution with kernel size 2×2 and a stride of 2, a concatenation operation and two 3×3 convolutions with ReLU. The transposed convolution is utilized for up-sampling and the concatenation operation is used for fusing the feature maps from the encoder stack into the corresponding decoder stack. At the last layer, a 1×1 convolution with softmax activation is employed to map obtained features to the segmentation probability map.

2) Information-fusion Component: After obtaining the feature map from the FCN component and complementary knowledge from traditional methods, they are combined through the information-fusion component for predicting the final segmentation probability map. The architecture of this component is given in Fig. 2, which begins with two convolutions with kernel size 3×3, each followed by a ReLU. One of these convolutions with C_1 filters is applied on feature maps obtained from the FCN component, while another one with C_2 filters is applied on the complementary knowledge. A concatenation with the obtained feature maps is then applied. After that, we utilize multiple 3×3 convolutions with ReLU. The last layer includes a convolution with kernel size 1×1 and the softmax activation. Filter numbers C_1 and C_2 are used to adjust the contribution of information from the FCN component and its counterpart since they determine the ratio of each kind of feature maps fed into the following block. For instance, when the amount of the training data is relatively sufficient, we can choose a larger C_1 to make the deep-learning based FCN component contribute more and vice versa. It should be noted that the adjustment by using C_1 and C_2 gives an initial setting, more refined adjustment work can be addressed adaptively based on the automatically learned weights of the network in the information-fusion component.

3) Loss Function: Small organs usually occupy only a small part of the image. This class-imbalance issue will result in sub-optimal performance. To solve this problem, we employ the Dice loss. Dice as a loss function was initially proposed in

[34] to tackle the class-imbalance issue. Then, the Dice loss and its variations have been widely used in different medical image segmentation tasks [35]–[37] and have been proved to be well adaptable to the high imbalance problem. Assume $p(x_i)$ is the prediction probability of a voxel x_i and $g(x_i)$ is the corresponding ground truth at the same voxel. The Dice loss is given by

$$DL(\mathscr{X}) = -\frac{2\sum\limits_{x_i \in \mathscr{X}} p(x_i)g(x_i) + \varepsilon}{\sum\limits_{x_i \in \mathscr{X}} p(x_i) + \sum\limits_{x_i \in \mathscr{X}} g(x_i) + \varepsilon} \tag{5}$$

where $\mathscr{X}$ is the training images, ε is a small term to prevent the loss function from the issue of dividing by 0.

III. Experiments

A. Dataset

The VISCERAL challenge dataset [32] is a public dataset for benchmarking state-of-the-art multi-organ segmentation methods. It contains twenty non-contrast-enhanced whole-body CT (CTwb) and twenty contrast-enhanced CT (CTce) volumes. The CTwb images with the field of view from the head to the knee are obtained from patients diagnosed with bone marrow neoplasms. The CTce scans, which are acquired from patients with malignant lymphoma, are obtained after improving tissue contrast with an injection of an iodine-containing contrast agent. Their field of view ranges from the corpus mandibulate to the pelvis. These images are annotated by physicians for up to twenty anatomical structures, and this paper focuses on the relatively small ones: the pancreas, gallbladder, left and right adrenal glands, which are more difficult to segment. Part of these small organs are not annotated because they are not visible in the modality. The number of available small organ annotations in different modalities is shown in Table I. The quality of obtained annotations is checked by three radiologists and two medical doctors [32].

B. Pre-processing

1) Data sampling: To guarantee that all the volumes have a uniform voxel size, the images are resampled to 1 mm isotropic resolution in all three dimensions. When the ground-truth of a particular organ is not provided in a volume, it is removed when evaluating the performance on this organ.

2) Gaussian Smoothing: For the VPF method, in order to avoid the influence of noisy artifacts, Gaussian smoothing with a Gaussian kernel with $\sigma_p = 3$ is employed to smooth the images when building the BRIEF and LBP features.

3) Gaussian Normalization: To normalize the intensity distributions for reducing variation across subjects, the Gaussian normalization is employed to normalize data fed to U-Net and KaCNN model. For the scan of each patient, the mean value and standard deviation were calculated based on intensities of all voxels. Then each image volume was normalized to zero mean and unit standard deviation.

C. Experimental Setup and Parameter Selection

The evaluation is performed with leave-one-out cross-validation. The dataset is randomly split into a training set, a validation set, and a test set. Each of the images is used once for testing and the remaining images are separated into training and validation sets in the ratio of $4:1$. Every cross-validation includes the training and evaluation of the models from scratch.

In the localization step, the threshold Θ is set at 0.05 and the security margins are 5% length of each side of the bounding box. We use Elastix [38] for registration with parameterization described in [39].

In the segmentation step, we compare the proposed KaCNN to the modified U-Net shown in Fig. 3, forest-based VPF [11], and two MAS methods with different label fusion techniques: joint label fusion [15] and majority voting [30]. It is necessary to note that we have tested the performance of the original U-Net architecture [13] and different modified U-Net architectures when they work alone on the dataset and we have chosen the best one of them (described in II-C1) as the FCN component of the KaCNN, which is also compared with the KaCNN here. The related parameters of the above methods are given in Table II. In the remainder of this paper, the MAS method with joint label fusion is denoted by MAS-JLF and the MAS method with majority voting is denoted by MAS-MV. The KaCNN model is implemented with the Keras library [40]. The weights are randomly initialized and trained using Adam optimizer. Considering the limited amount of available data, the KaCNN is implemented in a 2D slice-wise fashion, which also reduces the computational complexity. All axial slices of each volume (Input 1 in Fig. 2) together with the corresponding slices of the segmentation probability map obtained by VPF method (Input 2 in Fig. 2) are fed one-by-one into the KaCNN network. We use the same number of feature maps from the FCN and the VPF methods, i.e. $C_1 = C_2 = 32$, at the beginning of the information-fusion component and allow the KaCNN network adaptively adjust their contribution by automatically learning weights of the network.

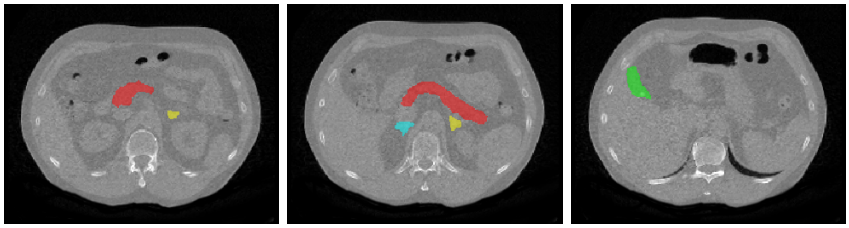
D. Evaluation

We evaluate the segmentation results with the Dice Similarity Coefficient (DSC):

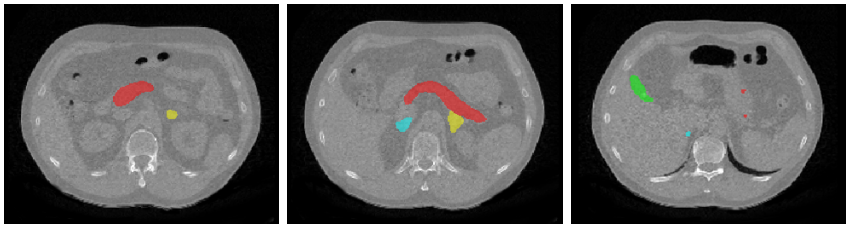
$$DSC = \frac{2\,|\mathbf{R} \cap \mathbf{G}|}{(|\mathbf{R}| + |\mathbf{G}|)} \tag{6}$$

where $\mathbf{R}$ and $\mathbf{G}$ represent the predicted and ground-truth segmentation respectively. In clinical practice, sensitivity is more important than specificity for tasks such as radiotherapy planning and false positive removal in positron-emission tomography (PET) images [21]. Therefore, the True Positive Rate (TPR) is another important evaluation guideline, which is given by:

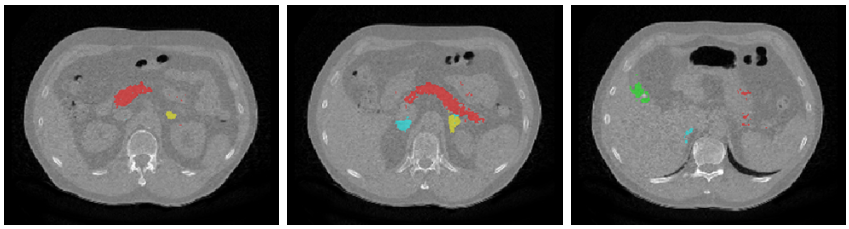
$$TPR = \frac{|\mathbf{R} \cap \mathbf{G}|}{|\mathbf{G}|}. \tag{7}$$



(a) Ground Truth

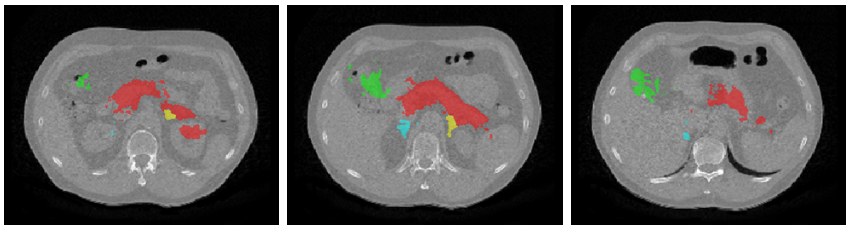


(b) The Proposed KaCNN



(c) Vantage Point Forests

(d) U-Net



(e) Multi-atlas Joint Label Fusion

Fig. 5. Illustration of the segmentation obtained for four subjects. The pancreas is shown in red, the gallbladder in green, the left adrenal gland in yellow and the right adrenal gland in cyan.

TABLE I
THE NUMBER OF AVAILABLE ANNOTATIONS OF DIFFERENT MODALITIES AND ORGANS (L-ADGLAND: LEFT ADRENAL GLAND, R-ADGLAND: RIGHT ADRENAL GLAND)

Organ / Image Modality	Pancreas	Gallbladder	L-AdGland	R-AdGland
CTwb	20	18	16	15
CTce	18	19	18	18

IV. RESULTS

The performance of the proposed KaCNN is compared to other methods in Table III to VI. The segmentation results of these methods on the CTwb dataset are presented in Table III and V and their segmentation results on CTce

TABLE II
PARAMETERS OF DIFFERENT METHODS

Parameter	Value
knowledge-aided convolutional neural network	
Learning rate	$1e^{-5}$ (Adam)
Batch size	50
N_{epoch}	100
U-Net	
Learning rate	$1e^{-5}$ (Adam)
Batch size	50
N_{epoch}	100
Vantage Point Forests	
Number of trees N_{tree}	15
Leaf size $\mathcal{L}_{min}$	15
Number of k-nearest neighbours	40
Binary feature length	640 bits
Training grid stride	4
Testing grid stride	2
Joint Label Fusion	
Appearance patch radius	$2 \times 2 \times 2$
Local search radius	$3 \times 3 \times 3$
Regularization term α	0.1
Exponent β	2

dataset are illustrated in Table IV and VI. In general, we can observe that the proposed KaCNN dramatically improves the segmentation performance on all small organs tested here (pancreas, gallbladder, and the left and right adrenal glands), compared to widely used deep-learning network U-Net, forest-based vantage point forests, multi-atlas methods MAS-JLF and MAS-MV. When comparing the performance of KaCNN with that of the U-Net and vantage point forests, it can be seen that after successfully combining the information of vantage point forests and U-Net, the proposed KaCNN outperforms each of these stand-alone methods. Fig. 4 illustrates this observation, where 3D segmentation results of these three methods are presented. The segmentation of vantage point forests is more accurate than that of U-Net but its surface is not smooth. After combing the effort of both methods, the proposed KaCNN obtained a smooth segmentation and better performance than both of them. In Fig. 5, three axial slices are selected to further illustrate the segmentations obtained by the proposed KaCNN and other three methods.

A. KaCNN versus Convolutional Neural Network (U-Net)

The proposed KaCNN and U-Net alone are compared in Table III and Table IV. From these tables, we can observe that the KaCNN obtains 0.189 higher Dice Similarity Coefficient and 0.181 higher True Positive Rate on average compared to the U-Net on CTwb data. Similarly, on the CTce data, the KaCNN obtains on average 0.194 higher Dice Similarity Coefficient and 0.204 higher True Positive Rate compared to the U-Net. Fig. 5(d) shows that the U-Net fails to avoid under-segmentation of the pancreas (see in the second slice) and over-segmentation of similar neighbouring structures (see in the third slice) on this sample. While after bringing more contextual information to the U-Net architecture, the KaCNN significantly alleviates the under- and over-segmentation problems (see in Fig. 5(b)).

B. KaCNN versus Forest-based Vantage Point Forests

Compared to vantage point forests, it can be seen that the KaCNN refines the result of the vantage point forests, which achieves 0.061 higher Dice Similarity Coefficient and 0.147 higher True Positive Rate on average when evaluated on CTwb data. Similarly, on the CTce data, the KaCNN achieves 0.067 higher Dice Similarity Coefficient and 0.144 higher True Positive Rate on average. Fig. 5(b) and Fig. 5(c) give a visible illustration of this refinement. With the help of the KaCNN network, the raw result of the vantage point forests obtains a smooth border and less over- and under-segmentation. Another advantage of KaCNN compared to vantage point forests is that it alleviates the burden of tuning the threshold for the probability map to produce the final label [41]. Indeed, the performance changes slightly when the threshold changes within a wide range around 0.5, which is used in this work, while the vantage point forests method needs to carefully tune the threshold parameter on the validation data for different organs. This benefit is attributed to the softmax function in the last layer of the KaCNN architecture (Fig. 2). It forces outputs to be either close to 1 or close to 0, so as to simplify the classification [41].

C. KaCNN versus Multi-atlas Segmentation Methods (MAS-JLF and MAS-MV)

We can see that MAS-MV underperforms other methods in Table III and Table IV. The MAS-JLF algorithm yields more accurate segmentation than the MAS-MV algorithm for the smaller organs. However, there is a significant gap between its performance and that of KaCNN, which can be seen directly when comparing Fig. 5(b) and Fig. 5(e). The proposed KaCNN outperformed MAS-JLF for all small organs on both CTce and CTwb data.

D. KaCNN versus Other Previously Proposed Work

In Table V and Table VI, we compare KaCNN with other previously proposed methods which have been applied to the same dataset. The short description of these methods could be found in Table V and Table VI, which belong to two categories — multi-atlas based methods including [43], [39] and [45] and forest-based methods including [44] and [21]. From this two tables, we see that KaCNN outperforms these approaches on the segmentation of small organs, which we attribute to two factors: (1) The localization step helps to focus the field of view on the region of interest and avoid over-segmenting similar neighbouring structures. (2) The proposed method takes advantage of multi-atlas segmentation in localization stage (registration of both training images and test image to the common template makes them similar to each other) and then combines the effort of forest-based vantage point forests and deep-learning based U-Net using KaCNN in segmentation stage.

TABLE III
THE AVERAGE DSC AND TPR OBTAINED BY DIFFERENT METHODS EVALUATED ON THE **CTWB** DATASET

Method \ Organ	Pancreas		Gallbladder		Left Adrenal Gland		Right Adrenal Gland	
	DSC	TPR	DSC	TPR	DSC	TPR	DSC	TPR
KaCNN	**0.583**	**0.591**	**0.473**	**0.530**	**0.472**	**0.528**	**0.390**	**0.498**
U-Net	0.351	0.426	0.404	0.441	0.307	0.465	0.101	0.090
VPF	0.520	0.475	0.377	0.361	0.418	0.369	0.358	0.355
MAS-JLF	0.416	0.494	0.148	0.334	0.355	0.428	0.306	0.383
MAS-MV	0.332	0.481	0.073	0.233	0.214	0.229	0.193	0.269

TABLE IV
THE AVERAGE DSC AND TPR OBTAINED BY DIFFERENT METHODS EVALUATED ON THE **CTCE** DATASET

Method \ Organ	Pancreas		Gallbladder		Left Adrenal Gland		Right Adrenal Gland	
	DSC	TPR	DSC	TPR	DSC	TPR	DSC	TPR
KaCNN	**0.588**	**0.690**	**0.624**	**0.695**	**0.403**	**0.505**	**0.434**	**0.548**
U-Net	0.336	0.373	0.503	0.546	0.275	0.468	0.159	0.234
VPF	0.521	0.571	0.561	0.575	0.369	0.383	0.331	0.334
MAS-JLF	0.407	0.539	0.139	0.372	0.365	0.498	0.307	0.462
MAS-MV	0.311	0.362	0.087	0.212	0.201	0.477	0.133	0.376

TABLE V
THE AVERAGE DSCS OF THE PROPOSED KACNN AND OTHER PREVIOUSLY PROPOSED METHODS EVALUATED ON THE **CTWB** DATA (RESULTS OF [39], [43] HAVE BEEN GENERATED ON THE PRIVATE VISCERAL TEST SET THAT IS NOT PUBLICLY AVAILABLE.)

Method	Description	Organ			
		Pancreas	Gallbladder	L-AdGland	R-AdGland
KaCNN	Deep neural network combining the U-Net and VPF	**0.583**	**0.473**	**0.472**	**0.390**
Bieth et al. [21]	Vantage point forests using regional context and shape voting	0.481	0.288	0.220	0.294
Peter et al. [44]	Scale-adaptive Random Forest	0.246	0.012	0.080	0.018
Gass et al. [43]	Multiatlas registration via Markov Random Field	0.438	0.102	0.165	0.138
Jiménez et al. [39]	Multiatlas registration, anatomical spatial correlations	0.408	0.276	0.373	0.355

TABLE VI
THE AVERAGE DSCS OF THE PROPOSED KACNN AND OTHER PREVIOUSLY PROPOSED METHODS EVALUATED ON THE **CTCE** DATA (RESULTS OF [39], [43], [45] HAVE BEEN GENERATED ON THE PRIVATE VISCERAL TEST SET THAT IS NOT PUBLICLY AVAILABLE.)

Method	Description	Organ			
		Pancreas	Gallbladder	L-AdGland	R-AdGland
KaCNN	Deep neural network combining the U-Net and VPF	**0.588**	**0.624**	**0.403**	**0.434**
Gass et al. [43]	Multiatlas registration via Markov Random Field	0.465	0.334	0.204	0.164
Jiménez et al. [39]	Multiatlas registration, anatomical spatial correlations	-	0.566	-	-
Kéchichian et al. [45]	Atlas registration, clustering, graph cut w/spatial relations	0.155	0.281	0.000	0.007

V. DISCUSSION

A. Computational Complexity

In clinical applications, the expected automatic multi-organ segmentation approach should not only have the ability to achieve high accuracy but also retain low computation time. We evaluated the computational cost of the proposed method and other existing methods. The results are demonstrated in Table VII. All of the experiments were conducted on a GNU/Linux server running Ubuntu 16.04, with Intel Core i7-6700 CPU and 32GB RAM. The networks were trained on a single NVIDIA Titan-Xp GPU with 12GB RAM. For the fair comparison, we evaluate the computational cost of each method in our proposed two-step framework.

Compared to the traditional MAS-JLF and MAS-MV methods, the proposed localization strategy dramatically reduces the computational cost of the registration. In conventional multi-atlas segmentation method, numerous registration processes are needed for aligning the atlas images to each test image, while we only need to align the test image to a common template during the inference phase. In the segmentation phase, the proposed KaCNN takes 16 seconds more than the U-Net on average, which is the time for vantage point forest to obtain the input knowledge of the KaCNN.

Totally, the segmentation of 4 organs of the whole abdominal volume takes around 83 seconds, which is efficient to be used in clinical applications such as diagnostic tasks. In future work, the localization time could be reduced by modifying specific organ localization methods or by employing deep-learning based registration.

B. Effectiveness and future work

Given that the small organs usually represent a limited fraction of the image, we apply a cascaded localization step and segmentation step for automatic segmentation. This cascaded

framework takes advantage of aligning training images and test images to be similar to each other by registration to a template constructed by a graph-based groupwise image registration. We see a scope of improvement in the localization part of our approach, which is currently affine registration-based. This is validated when the segmentation component of the KaCNN is tested on manually-selected bounding box ROIs, we see a considerable improvement in performance, e.g., average Dice Similarity Coefficients on the CTwb data are 0.656 (Pancreas), 0.731 (Gallbladder), 0.605 (left adrenal gland) and 0.602 (right adrenal gland), and that on CTce data are 0.620 (Pancreas), 0.734 (Gallbladder), 0.601 (left adrenal gland) and 0.599 (right adrenal gland). Therefore, in spite of advantages entailing our approach (simultaneous localization of multiple ROIs, ability to work with limited data) we could resort to an improvement by modifying some organ localization methods [41], [46] or more complex registration techniques such as hierarchical registration [47], tree-based registration [48] and deep-learning based registration. [49], [50].

Once localized, a novel knowledge-aided convolutional neural network architecture is proposed, which combines the effort of deep-learning based methods and traditional methods for small organ segmentation. This proposed architecture outperforms each of individual components by a significant margin (> 6 DSC points on the VISCERAL data), thereby illustrating the significance of the fusion. As a next step, since the organs of interest are small and thanks to our ROI localization, 3D segmentation is a feasible extension. Another direction of interest lies in merging more than two paths in the information-fusion component, which is also a potential way to enhance the performance of the KaCNN. Another interesting direction worth exploring is to employ a Region Proposal Network for localization and combine it with our proposed KaCNN into a single end-to-end model. Besides, in the pre-processing step, the Gaussian filter is not edge-preserving. Therefore, we will consider other edge-preserving methods such as the bilateral filtering and Wavelet transform based filtering for image denoising [51].

VI. Conclusion

In this paper, we proposed a *knowledge-aided* convolutional neural network (KaCNN) specially designed for small organ segmentation when limited training data is available. A novel knowledge-aided convolutional network architecture is proposed, which combines the assets of deep-learning based methods and traditional methods for small organ segmentation and outperforms either of those method taken individually. We apply cascaded localization and segmentation steps for automatic organ segmentation. This cascaded framework takes advantage of aligning training and test images to be similar to each other by registrations and the localization step plays a role of focusing the field of view on the region of interest, avoid over-segmentation of similar neighbouring structures and facilitate the following segmentation. Experimental results on the ISBI 2015 VISCERAL challenge dataset demonstrated that the proposed method outperforms deep-learning based U-Net and traditional methods in the segmentation of small organs on the same dataset.

Acknowledgment

We would like to gratefully acknowledge NVIDIA Corporation for the donation of a Titan XP GPU used for this research.

References

[1] P. S. Steeg and D. Theodorescu, "Metastasis: a therapeutic target for cancer," *Nature Reviews Clinical Oncology*, vol. 5, no. 4, p. 206, 2008.

[2] B. van Ginneken, C. M. Schaefer-Prokop, and M. Prokop, "Computer-aided diagnosis: how to move from the laboratory to the clinic," *Radiology*, vol. 261, no. 3, pp. 719–732, 2011.

[3] K. Men, J. Dai, and Y. Li, "Automatic segmentation of the clinical target volume and organs at risk in the planning ct for rectal cancer using deep dilated convolutional neural networks," *Medical physics*, vol. 44, no. 12, pp. 6377–6389, 2017.

[4] G. Sharp, K. D. Fritscher, V. Pekar, M. Peroni, N. Shusharina, H. Veeraraghavan, and J. Yang, "Vision 20/20: perspectives on automated image segmentation for radiotherapy," *Medical physics*, vol. 41, no. 5, 2014.

[5] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. DAnastasi *et al.*, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 415–423.

[6] Z. Wang, K. K. Bhatia, B. Glocker, A. Marvao, T. Dawes, K. Misawa, K. Mori, and D. Rueckert, "Geodesic patch-based segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 666–673.

[7] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE transactions on medical imaging*, vol. 32, no. 9, pp. 1723–1730, 2013.

[8] J. J. Cerrolaza, M. Reyes, R. M. Summers, M. Á. González-Ballester, and M. G. Linguraru, "Automatic multi-resolution shape modeling of multi-organ structures," *Medical image analysis*, vol. 25, no. 1, pp. 11–21, 2015.

[9] C. Chu, M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, Y. Hayashi, Y. Nimura, D. Rueckert, and K. Mori, "Multi-organ segmentation based on spatially-divided probabilistic atlas from 3d abdominal ct images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 165–172.

[10] E. Geremia, B. H. Menze, O. Clatz, E. Konukoglu, A. Criminisi, and N. Ayache, "Spatial decision forests for ms lesion segmentation in multi-channel mr images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2010, pp. 111–118.

[11] M. P. Heinrich and M. Blendowski, "Multi-organ segmentation using vantage point forests and binary context features," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 598–606.

[12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[14] V. Zografos, A. Valentinitsch, M. Rempfler, F. Tombari, and B. Menze, "Hierarchical multi-organ segmentation without registration in 3d abdominal ct images," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2015, pp. 37–46.

[15] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 3, pp. 611–623, 2013.

[16] Y. Song, G. Wu, K. Bahrami, Q. Sun, and D. Shen, "Progressive multi-atlas label fusion by dictionary evolution," *Medical image analysis*, vol. 36, pp. 162–171, 2017.

[17] G. Sanroma, O. M. Benkarim, G. Piella, O. Camara, G. Wu, D. Shen, J. D. Gispert, J. L. Molinuevo, M. A. G. Ballester, A. D. N. Initiative *et al.*, "Learning non-linear patch embeddings with neural networks for label fusion," *Medical image analysis*, vol. 44, pp. 143–155, 2018.

[18] G. Wu, M. Kim, G. Sanroma, Q. Wang, B. C. Munsell, D. Shen, A. D. N. Initiative *et al.*, "Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition," *NeuroImage*, vol. 106, pp. 34–46, 2015.

TABLE VII
THE COMPUTATIONAL COST OF PROPOSED APPROACH AND THE COMPARED METHODS

Approach \ Average Time	Localization (sec)	Segmentation (sec)	Total Time (sec)
KaCNN	66.00	16.75	82.75
U-Net	66.00	0.70	66.70
VPF	66.00	16.00	82.00
MAS-JLF	66.00	232.00	298.00
MAS-MV	66.00	0.15	66.15
traditional MAS-JLF	-	-	3660.00
traditional MAS-MV	-	-	3360.00

[19] D. Zikic, B. Glocker, and A. Criminisi, "Atlas encoding by randomized forests for efficient label propagation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 66–73.

[20] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2012.

[21] M. Bieth, E. Alberts, M. Schwaiger, and B. Menze, "From large to small organ segmentation in ct using regional context," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2017, pp. 1–9.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[24] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[25] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[26] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3d deeply supervised network for automated segmentation of volumetric medical images," *Medical image analysis*, vol. 41, pp. 40–54, 2017.

[27] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[28] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[29] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár, "Learning to refine object segments," in *European Conference on Computer Vision*. Springer, 2016, pp. 75–91.

[30] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[31] S. Ying, G. Wu, Q. Wang, and D. Shen, "Groupwise registration via graph shrinkage on the image manifold," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 2323–2330.

[32] O. Jimenez-del Toro, H. Müller, M. Krenn, K. Gruenberg, A. A. Taha, M. Winterstein, I. Eggel, A. Foncubierta-Rodríguez, O. Goksel, A. Jakab *et al.*, "Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: Visceral anatomy benchmarks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2459–2475, 2016.

[33] L. Wang, Y. Zhang, and J. Feng, "On the euclidean distance of images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1334–1339, 2005.

[34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 565–571.

[35] H. Li, G. Jiang, R. Wang, J. Zhang, Z. Wang, W.-S. Zheng, and B. Menze, "Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images," *arXiv preprint arXiv:1802.05203*, 2018.

[36] M. Drozdzal, G. Chartrand, E. Vorontsov, M. Shakeri, L. Di Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury, "Learning normalized inputs for iterative estimation in medical image segmentation," *Medical image analysis*, vol. 44, pp. 1–13, 2018.

[37] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 240–248.

[38] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[39] O. A. J. del Toro, Y. D. Cid, A. Depeursinge, and H. Müller, "Hierarchic anatomical structure segmentation guided by spatial correlations (anatseg-gspac): Visceral anatomy3." in *VISCERAL Challenge@ ISBI*, 2015, pp. 22–26.

[40] F. Chollet *et al.*, "Keras," https://github.com/keras-team/keras, 2015.

[41] B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever, and I. Išgum, "Convnet-based localization of anatomical structures in 3-d medical images," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1470–1481, 2017.

[42] J. S. Bridle, "Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters," in *Advances in neural information processing systems*, 1990, pp. 211–217.

[43] T. Gass, G. Szekely, and O. Goksel, "Multi-atlas segmentation and landmark localization in images with large field of view," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2014, pp. 171–180.

[44] L. Peter, O. Pauly, P. Chatelain, D. Mateus, and N. Navab, "Scale-adaptive forest training via an efficient feature sampling scheme," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 637–644.

[45] R. Kéchichian, S. Valette, M. Sdika, and M. Desvignes, "Automatic 3d multiorgan segmentation via clustering and graph cut using spatial relations and hierarchically-registered atlases," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2014, pp. 201–209.

[46] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Medical image analysis*, vol. 17, no. 8, pp. 1293–1303, 2013.

[47] O. A. J. del Toro and H. Müller, "Hierarchic multi–atlas based segmentation for anatomical structures: Evaluation in the visceral anatomy benchmarks," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2014, pp. 189–200.

[48] H. Jia, P.-T. Yap, and D. Shen, "Iterative multi-atlas-based multi-image segmentation with tree-based registration," *NeuroImage*, vol. 59, no. 1, pp. 422–430, 2012.

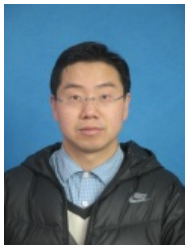
[49] B. Gutiérrez-Becker, D. Mateus, L. Peter, and N. Navab, "Learning optimization updates for multimodal registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 19–27.

[50] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen, "Deformable image registration based on similarity-steered cnn regression," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 300–308.

[51] P. Jain and V. Tyagi, "A survey of edge-preserving image denoising methods," *Information Systems Frontiers*, vol. 18, no. 1, pp. 159–170, 2016.

Yu Zhao is a third-year Ph.D. student of Computer Science at Technische Universität München (TUM). He is working in Image-Based Biomedical Modeling (IBBM) group with Prof. Bjoern H. Menze. Prior to coming to TUM, he received a Msc degree in signal and information processing and a BSc. degree in physics both from Beihang University (BUAA). His research focuses on the medical computer vision and application of machine learning, in particular medical image segmentation and high-level vision tasks.

Hongwei Li is a first-year Ph.D student of Computer Science at Technische Universität München (TUM) under the supervision of Prof. Bjoern Menze. He was a visiting research student in University of Dundee for six months as a part of his master program in Sun Yat-sen University. His research interests include machine learning and biomedical image processing.



Shaohua Wan received his Ph.D. degree from School of Computer, Wuhan University. From 2015, he worked as a postdoc at State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology. From 2016 to 2017, he worked as a visiting scholar at Decomplementarytment of Electrical and Computer Engineering at the Technical University of Munich. At present, he is an associate professor and master advisor at school of Information and Safety Engineering, Zhongnan University of Economics and Law. His main research interests include massive data computing for sensor networks and Internet of Things and cognitive robotics.

Anjany Sekuboyina is a second-year Ph.D. student of Computer Science at Technische Universität München (TUM) working in Image-Based Biomedical Modeling (IBBM) group with Prof. Bjoern H. Menze and Dr. Jan S. Kirschke. He has a Masters degree from the Indian Institute of Science, Bangalore with a specialization in signal processing. He works on the cross domains of medical image analysis, computer vision, and machine learning with special focus on multi-modal spine images.

Xiaobin Hu received the M.S degree in physics from the Hunan University in September 2017. He is a first-year Ph.D. student in chair of Image-Based Biomedical Modelling, Fakultät für Informatik in Technische Universität München (TUM) for biomedical information analysis. His research focuses on the semantic segmentation of Brain image, and probability theory of deep learning and machine learning. His main interests are computer vision, machine learning, and uncertain optimization.

Giles Tetteh Giles Tetteh has a MSc. degree in mathematical sciences from the African Institute for Mathematical Sciences (AIMS) - Ghana. He is currently a third-year doctoral candidate in computer science with the Image-Based Biomedical Modeling (IBBM) group of Prof. Bjoern H. Menze at the Technische Universität München (TUM). His general research area is application of deep learning methods to medical image analysis with special focus on brain vessel analysis.

Marie Piraud received a Ph.D. degree in Physics in 2012 from Université Paris-Sud, Orsay, France, and consequently joined the Ludwig-Maximilian University of Munich, Germany. In 2016, she joined the Image-Based Biomedical Modeling group of the Technical University of Munich, as a senior researcher. Her current research interests include multi-scale modelisation and computer vision with deep learning techniques, both applied to the medical realm.

Bjoern H. Menze is an Assistant Professor and the Head of the Image-based Biomedical Modeling and Computational Physiology Group at the Department of Computer Science, Technische Universität München, Munich, Germany, and a Visiting Research Scientist of the Asclepios team at the Inria Sophia-Antipolis, Nice, France. He is also a Research Affiliate of the Medical Vision group at CSAIL, Massachusetts Institute of Technology (MIT), Cambridge, MA. He received the M.Sc. degree in physics from Uppsala University, Uppsala, Sweden, in 2002, the M.Sc. degree from Heidelberg University, Heidelberg, Germany, in 2004, and the Ph.D. degree in computer science from the Interdisciplinary Center for Scientific Computing, Heidelberg, in 2007. He subsequently moved to Boston, MA, USA, where he was a Post-Doctoral Researcher at Harvard University, Cambridge, Harvard Medical School, Boston, and MIT. He was then as a Researcher at Inria Sophia-Antipolis, Nice, France, and a Senior Researcher and Lecturer at ETH Zurich, Zurich, Switzerland. He is developing methods in medical image computing at the interface of computational pathophysiology, medical computer vision, and machine learning. In this, he focuses on applications in multimodal clinical neuroimaging and the patient-adaptive modeling of tumor growth. He organized workshops at MICCAI, NIPS, and CVPR in the fields of medical computer vision and neuroimage processing, served as a Guest Editor of Medical Image Analysis, and as a Program Committee Member of the International Conference on Medical Image Computing and Computer Assisted Intervention.

4

Deep Neural Network for Automatic Characterization of Lesions on ^{68}Ga-PSMA PET/CT Images

This chapter has been published as **peer-reviewed conference paper:**.

Y. Zhao, A. Gafita, G. Tetteh, F. Haupt, A. Afshar-Oromieh, B. Menze, M. Eiber, A. Rominger, and K. Shi. "Deep Neural Network for Automatic Characterization of Lesions on ^{68}Ga-PSMA PET/CT Images." In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 951–954. DOI: 10.1109/EMBC.2019.8857955

Synopsis: This work develops an end-to-end deep neural network to characterize the prostate cancer lesions on ^{68}Ga-PSMA-11 PET/CT imaging automatically.

Contributions of thesis author: algorithm design and implementation, computational experiments and composition of manuscript.

Accepted manuscript for the IEEE (41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society , 2019)
Published version: DOI: 10.1109/EMBC.2019.8857955

Deep Neural Network for Automatic Characterization of Lesions on ^{68}Ga-PSMA PET/CT Images

Yu Zhao[1], Andrei Gafita[2], Giles Tetteh[1], Fabian Haupt[3], Ali Afshar-Oromieh[3,5], Bjoern Menze[1], Matthias Eiber[2], Axel Rominger[3,4], and Kuangyu Shi[1,3]

***Abstract*— The emerging PSMA-targeted radionuclide therapy provides an effective method for the treatment of advanced metastatic prostate cancer. To optimize the therapeutic effect and maximize the theranostic benefit, there is a need to identify and quantify target lesions prior to treatment. However, this is extremely challenging considering that a high number of lesions of heterogeneous size and uptake may distribute in a variety of anatomical context with different backgrounds. This study proposes an end-to-end deep neural network to characterize the prostate cancer lesions on PSMA imaging automatically. A ^{68}Ga-PSMA-11 PET/CT image dataset including 71 patients with metastatic prostate cancer was collected from three medical centres for training and evaluating the proposed network. For proof-of-concept, we focus on the detection of bone and lymph node lesions in the pelvic area suggestive for metastases of prostate cancer. The preliminary test on pelvic area confirms the potential of deep learning methods. Increasing the amount of training data may further enhance the performance of the proposed deep learning method.**

I. INTRODUCTION

Prostate cancer (PC) is the third most frequent cause of cancer-related mortality among men in developed countries [1]. Despite effective primary treatment, malignant tumour cells may spread to other regions by invading the hematic and lymphatic systems, leading to the advanced stage of the disease with a 5-year survival rate of 29% for metastatic PC [2]. Several methods including chemotherapeutic agents have been employed to treat advanced metastatic PC, which showed benefits in terms of survival rate [3]. Prostate specific-membrane antigen (PSMA) is a type II transmembrane glycoprotein which is expressed by normal prostate cells and significantly upregulated in prostate cancer cells [1]. PSMA undergoes constitutive internalization, therefore can serve as a target not only for imaging but also in the therapy framework. Therefore, PSMA has become a target of interest in the theragnostic approach [4].

Despite the encouraging results of PSMA-targeted radioligand therapy (RLT), treatment planning of this novel therapy is much more challenging compared to conventional external beam radiotherapy, due to systemic spread of the lesions. It is reported that 30% of the treated patients turned out to be suboptimal [5]. Therefore, there is a need to improve the treatment planning to optimize the RLT outcome. A critical step for treatment planning is to assess tumour burden and namely, to characterize the lesions. Usually, patients who undergo PSMA-targeted RLT have a high number of metastases. Therefore, a manual characterize method is most likely impossible since it is time-consuming. A first approach towards a semiautomatic method was developed to characterize the tumour load of bone metastases, namely bone-PET-index (BPI), on ^{68}Ga-PSMA-11 PET/CT images for ^{223}Ra-dichloride therapy by segmenting using an SUV-based threshold on PET and restricting the result to the skeleton segmented from the CT images [6]. However, this method cannot be easily extended to other types of lesions such as lymph node metastases, where prior anatomical information is more difficult than for bone skeleton. It is extremely challenging to characterize a high number of lesions of heterogeneous size and uptake distributing in a variety of anatomical context with different backgrounds. Until now, there are no successful computer-aided methods to evaluate tumour load for the treatment planning of PSMA-targeted RLT.

Deep learning has achieved significant success in diverse computer vision applications, including object detection, image classification, and segmentation. Unlike conventional methods which usually rely on expert-designed features, convolutional neural networks (CNNs) have the advantage of being able to automatically and effectively learn salient feature representations [7]. For image segmentation, deep classification networks with sliding patches have been used initially [8], which leads to redundant computations and long inference times. Fully convolutional network (FCN) [9] was later developed to realize semantic segmentation using fully convolution layers, deconvolution layers and skip architecture. FCN-like networks were then applied to medical image segmentation, the proposed U-Net [10] obtained competitive performance on neuronal structure and cell segmentation. Subsequently, more and more FCN-based methods have been proposed and gained significant success in different medical image segmentation problems [11]–[13]. Two cascaded FCN, i.e. W-Net, has been applied to automatically detect and segment bone lesions on ^{68}Ga-Pentixafor PET/CT images [14]. However, this method is still restricted to the characterization of less challenging bone lesions.

In this study, we propose a 3D deep supervised residual U-Net (DS-Res-U-Net) to automatically characterize local recurrence, bone lesions, and lymph node metastasis syn-

[1]Department of Computer Science, Technische Universität München, Munich, Germany yu.zhao@tum.de
[2]Department of Nuclear Medicine, Technische Universität München, Munich, Germany
[3]Department of Nuclear Medicine, University of Bern, Bern, Switzerland
[4]Department of Nuclear Medicine, Ludwig Maximilian University of Munich, Munich, Germany
[5]Department of Nuclear Medicine, Heidelberg University Hospital, Heidelberg, Germany

chronously. For proof-of-concept, this work focuses on the detection of lesions in the pelvic area. By taking full advantage of the 3D spatial information, our method can learn representative features with higher discrimination capability than those learned from 2D CNNs. The proposed deep learning method was then evaluated on a dataset collected from three different centres.

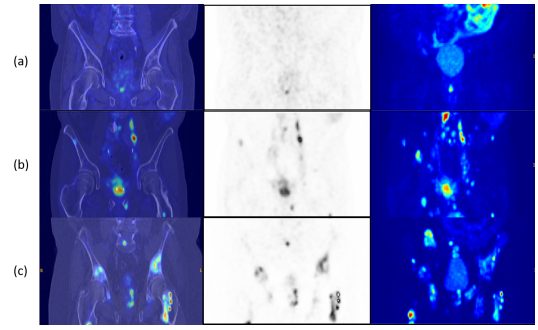


Fig. 1. Typical examples illustrating the recurrent PC patients with lesions in the prostate, bone and lymph node. The first column shows fused PET/CT slices, the middle column shows PET slices only in reversed colormap, the last column shows the maximum intensity projection (MIP) of PET scan.

II. MATERIAL AND METHOD

A. Patients, Imaging and Pre-processing

In this study, a dataset of 71 patients with metastatic prostate cancer was included, which was collected from Technical University of Munich (25 Patients), University of Munich (16 Patients) and the University of Bern (30 Patients). All the patients underwent ^{68}Ga-PSMA-11 PET/CT imaging from the head to the thigh approximately 60 min after intravenous injection of ^{68}Ga-PSMA-11. A low-dose CT was obtained for attenuation correction. PET emission data were acquired using a 3D model, followed with decay and scatter correction, and was iteratively reconstructed with attenuation correction. The PET images were normalized by injection activity and body weight to standard uptake values (SUVs). The PET image was co-registered to CT and all the imaging data of the three centres were resliced to the same pixel size $1 \times 1 \times 1$ mm^3. A Gaussian normalization upon Hounsfield unit of CT and tracer uptake distribution of PET was applied to both PET and CT images. All the lesions were annotated by a nuclear medicine physician and his assistants in fused PET/CT scans.

Fig. 1 illustrates typical examples of the PET/CT dataset from patients with different types of PC lesions including local, bone metastasis and lymph node metastasis. In the dataset, 21 patients have primary PC and local recurrent tumour (Fig. 1 a); Most patients underwent prostatectomy, they may have cured local tumours after surgery, while lymph node or bone metastases recur (Fig. 1 b); Fifteen patients had all three types of lesions mentioned above (Fig. 1 c). Therefore, to evaluate the overall state of the disease, all these kinds of lesions need to be detected and analysed, which is a challenging task for the following reasons: 1) The original low contrast of the prostate gland to the background impairs the detection of the local tumour. 2) The unexpected occurrence of lymph nodes or bone lesions across the body with large heterogeneity in shape, size, and intensity make the detection more difficult and 3) Compared to the background, the lesions usually occupy a small fraction of the image, which leads to high imbalance issue for the segmentation.

B. Deep Neural Network

We propose a 3D deep supervised residual U-Net to characterize PC lesions. As shown in Fig. 2, the network consists of an encoder-decoder architecture adopted as the mainstream network, residual connections, and deep supervision layers. Similar to the U-Net [10], the mainstream network comprises a down-sampling path including three repeated encoder stacks and an up-sampling path including three repeated decoder stacks. In each encoder stack, there are a residual module and a $3\times3\times3$ convolutional layers with stride 2 for down-sampling the feature maps. Each residual module includes three $3 \times 3 \times 3$ convolutional layers and two dropout layers in between. Each decoder stack starts with a transposed convolution with kernel size $2 \times 2 \times 2$ and a stride of 2 for up-sampling the feature map resolution and halving the number of feature maps simultaneously. The remaining decoder stack consists of a $3 \times 3 \times 3$ convolution followed by a $1 \times 1 \times 1$ convolution which halves the number of feature maps. With the concatenation operations, feature maps from the encoder stack are fused into the corresponding decoder stack. At the last layer, a $1 \times 1 \times 1$ convolution with softmax activation is employed to map obtained features to the segmentation probability map.

Inspired by [15], we exploit deep supervision in the up-sampling path for improving the prediction accuracy, averting the problem of vanishing gradients and accelerating the optimization convergence rate of the whole network. Throughout the network, we employ the leaky ReLU as the activation function following the convolution layers and utilize instance normalization [16] instead of the traditional batch normalization since it is more stable than batch normalization if working on small batch size.

C. Loss Function

One challenge in medical image segmentation is the class imbalance in the data. In the prostate cancer dataset, the cancer lesions usually occupy only a small part of the image. This class-imbalance issue will result in sub-optimal performance. The Dice loss has been widely used in different medical image segmentation tasks and has been proved to be well adapted to the high imbalance problem [17]. Therefore, in this work, we choose to use a multi-class Dice loss function [18] to tackle this issue. Assuming p is the output of the network and g is the one hot encoding of the ground truth segmentation map, the multi-class Dice loss is defined as:

$$L_{dc} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i p_i^k g_i^k}{\sum_i p_i^k \sum_i g_i^k} \quad (1)$$

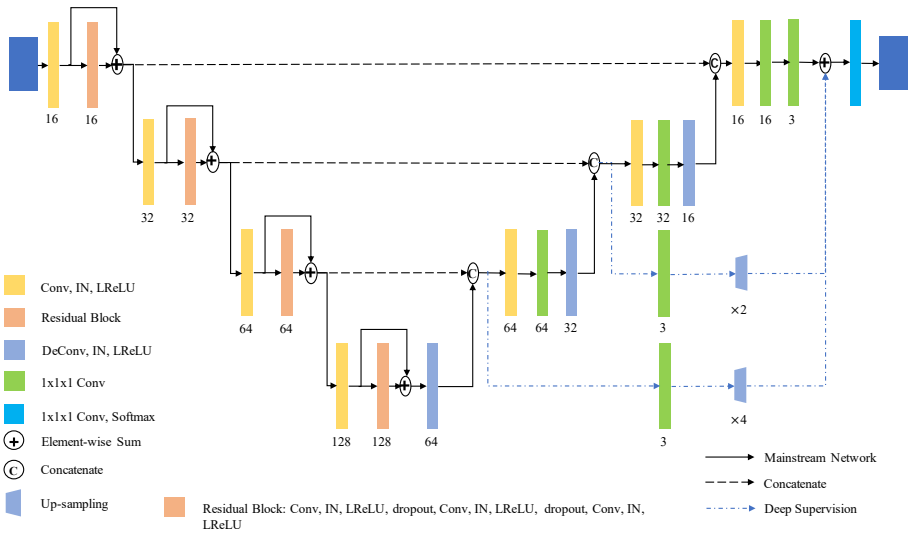


Fig. 2. The network structure of deep supervised residual U-Net. The context pathway (left) aggregates high-level information that is subsequently localized precisely in the localization pathway (right). We inject gradient signals deep into the network through deep supervision. The digits below the operations denote the number of channels. (Conv: convolutional layer, DeConv: transposed convolution, IN: instance normalization, LReLU: leaky ReLU)

where p and g have shape $I \times K$ with $i \in I$ being the number of pixels in the training patch and $k \in K$ being the classes.

D. Lesion Detection Based On the Segmentation Result

The proposed lesion detection approach is based on the segmentation result obtained by the proposed network. Different from the conventional detection task in computer vision community, which has a regular bounding box or sphere as the ground truth, we use the manually annotated areas as the ground truth, which is irregular but more accurate. In this work, we assume that the lesion is a topologically connected area. If two areas are disconnected, they will be considered as two lesions. During lesion detection, all connected areas of different kind of lesions (bone lesion, lymph node lesion, and local lesion) are recognized after receiving the segmentation result. Then we filtered out small connected areas, whose volume is less than a threshold T_V, to avoid oversaturation by noises. Therefore, connected lesion areas in the obtained segmentation map are determined as positive or negative as follows:

$$d_{l,i} = \begin{cases} 1 & if \quad V_{l,i} > T_V \\ 0 & if \quad V_{l,i} < T_V \end{cases} \tag{2}$$

where l represents the lesion type, $V_{l,i}$ denotes $i\ th$ connected area, T_V is the threshold.

Besides, the lesion detection accuracy was evaluated based on the overlap between a predicted lesion and ground truth. A lesion was considered as correctly detected when the overlap ratio is higher than a threshold T_E.

III. Results and Discussion

The evaluation was performed with six-fold cross validation. For each training/testing split, the training folds were further split into a training set and validation set in a ratio of 4:1 (The validation set was used for parameter tuning purpose). We train the network with randomly sampled $64 \times 64 \times 64$ voxel patches. The PET scan and CT scan are assembled as two channels of the input images for the network. ADAM optimizer was used during training with an initial learning rate $lr_{init} = 10^{-2}$. To regularize the network, we utilized the early stopping strategy with the patience of 20, which is a method employed to detect the convergence of training thereby avoiding overfitting. We filtered out connected masses, whose volume is less than $T_V = 25mm^3$, to avoid oversaturation by noises.

Detection results of the proposed network for three different kinds of prostate cancer lesions are illustrated in Table 1. We have tested the influence of the threshold T_E on the detection performance. We found that the influence on the bone lesion and the lymph node lesion is unobvious and we can obtain the best detection performance when $T_E = 0.1$. As can be seen from Table 1, the proposed method achieves F1 score as high as 98%, 79% and 73% on the bone lesion, lymph node lesion and the local lesion respectively, which demonstrates the effectiveness of the proposed method on detection of the prostate cancer lesions.

In Fig. 3, we further illustrate the segmentation results obtained by the proposed network. Slices in different views are selected to demonstrate the results of bone lesions, lymph-node metastases, and local recurrence, respectively, where the true positive, false positive and false negative are ma rked with different colours. Typically, false negatives occur when the lesion is too small while the contrast is not high enough to identify its presence. False positives are highly intensity driven, which considers the nonspecific high tracer uptake as a lesion by mistake.

IV. Conclusions

This study exploits the deep learning method for the detection of prostate cancer lesions in ^{68}Ga-PSMA-11 PET/CT scans. We proposed an end-to-end 3D deep supervised

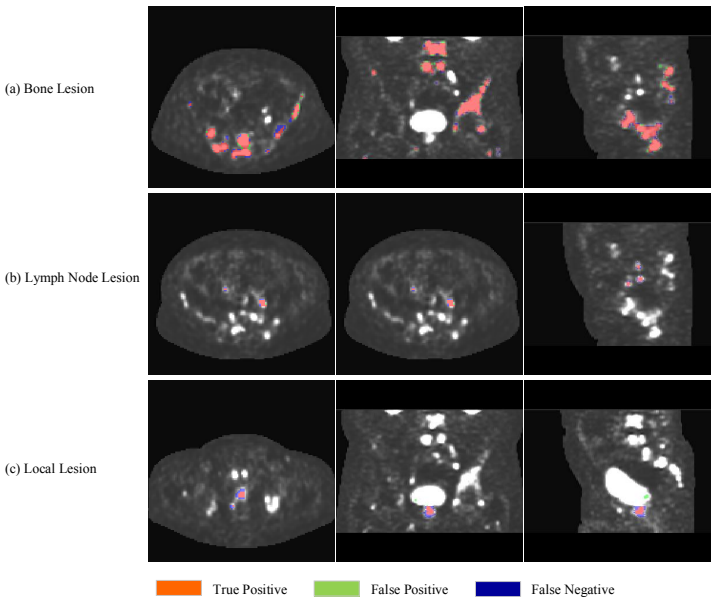


Fig. 3. Exemplary segmentation results of the proposed network. The left, middle and right column demonstrate the axial, sagittal, and coronal views separately.

TABLE I

DETECTION ACCURACY OF THE PROPOSED NETWORK FOR THREE DIFFERENT KINDS OF PROSTATE CANCER LESIONS

	Bone Lesion	Lymph Node Lesion	Local Lesion
Precision	0.97	0.75	0.88
Recall	0.98	0.84	0.63
F1 Score	0.98	0.79	0.73
Total Lesions (71 patients)	955	232	45

residual convolutional neural network, which encodes richer spatial information and extracts more discriminative representations via the hierarchical architecture trained with 3D samples. ^{68}Ga-PSMA-11 PET/CT scans from three different centres were used to train and evaluate the proposed networks. The preliminary test on pelvic area confirmed the potential of deep learning methods. Currently, more data are in collection and annotation. Increasing the amount of training data may further enhance the performance of the developed deep learning methods.

REFERENCES

[1] T. Maurer *et al.*, "Current use of PSMAPET in prostate cancer management," *Nature Reviews Urology*, vol. 13, no. 4, p. 226, 2016.

[2] K. D. Bernacki *et al.*, "The utility of psma and psa immunohistochemistry in the cytologic diagnosis of metastatic prostate carcinoma," *Diagnostic cytopathology*, vol. 42, no. 7, pp. 570–575, 2014.

[3] G. Attard *et al.*, "Prostate cancer," *The Lancet*, vol. 387, no. 10013, pp. 70 – 82, 2016.

[4] M. Weineisen *et al.*, "68ga-and 177lu-labeled psma i&t: optimization of a psma-targeted theranostic concept and first proof-of-concept human studies," *Journal of Nuclear Medicine*, vol. 56, no. 8, pp. 1169–1176, 2015.

[5] M. Eiber *et al.*, "Prostate-specific membrane antigen ligands for imaging and therapy," *J Nucl Med*, vol. 58, no. Supplement 2, pp. 67S–76S, 2017.

[6] M. Bieth *et al.*, "Exploring new multimodal quantitative imaging indices for the assessment of osseous tumor burden in prostate cancer using 68ga-psma pet/ct," *Journal of Nuclear Medicine*, vol. 58, no. 10, pp. 1632–1637, 2017.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[8] D. Ciresan *et al.*, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[9] J. Long *et al.*, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[10] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[11] D. Shen *et al.*, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[12] M. Hatt *et al.*, "The first miccai challenge on pet tumor segmentation," *Medical image analysis*, vol. 44, pp. 177–195, 2018.

[13] J. E. Corral *et al.*, "Su1337-deep learning to diagnose intraductal papillary mucinous neoplasms (ipmn) with mri," *Gastroenterology*, vol. 154, no. 6, pp. S–524, 2018.

[14] L. Xu *et al.*, "Automated whole-body bone lesion detection for multiple myeloma on 68ga-pentixafor pet/ct imaging using deep learning methods," *Contrast media & molecular imaging*, vol. 2018, 2018.

[15] Q. Dou *et al.*, "3d deeply supervised network for automated segmentation of volumetric medical images," *Medical image analysis*, vol. 41, pp. 40–54, 2017.

[16] V. L. D. U. A. Vedaldi, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[17] F. Milletari *et al.*, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 565–571.

[18] F. Isensee *et al.*, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 287–297.

5

Predicting Lymph Node Metastasis Based on Multiple Instance Learning with Deep Graph Convolution

This chapter has been published as **peer-reviewed conference paper:**.

Y. Zhao, F. Yang, Y. Fang, H. Liu, N. Zhou, J. Zhang, J. Sun, S. Yang, B. Menze, X. Fan, et al. "Predicting Lymph Node Metastasis Using Histopathological Images Based on Multiple Instance Learning With Deep Graph Convolution." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 4837–4846

Synopsis: This work deals with the problem of lymph node metastasis prediction. we propose a multiple instance learning method based on deep graph convolutional network and weakly supervised feature selection (FS-GCN-MIL) for histopathological image classification. The proposed method consists of three components, including instance-level feature extraction, instance-level feature selection, and bag-level classification.

Contributions of thesis author: algorithm design and implementation, computational experiments and composition of manuscript.

Accepted manuscript for the IEEE (Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition , 2020)
Published version: DOI: 10.1109/CVPR42600.2020.00489

Predicting Lymph Node Metastasis Using Histopathological Images Based on Multiple Instance Learning with Deep Graph Convolution

Yu Zhao[1,2*], Fan Yang[1*], Yuqi Fang[3], Hailing Liu[4], Niyun Zhou[1], Jun Zhang[1], Jiarui Sun[1], Sen Yang[1], Bjoern Menze[2], Xinjuan Fan[4] and Jianhua Yao[1]
[1]Tencent AI Lab, [2]Technical University of Munich, [3]The Chinese University of Hong Kong
[4]Sixth Affiliated Hospital of Sun Yat-sen University

Abstract

Multiple instance learning (MIL) is a typical weakly-supervised learning method where the label is associated with a bag of instances instead of a single instance. Despite extensive research over past years, effectively deploying MIL remains an open and challenging problem, especially when the commonly assumed standard multiple instance (SMI) assumption is not satisfied. In this paper, we propose a multiple instance learning method based on deep graph convolutional network and feature selection (FS-GCN-MIL) for histopathological image classification. The proposed method consists of three components, including instance-level feature extraction, instance-level feature selection, and bag-level classification. We develop a self-supervised learning mechanism to train the feature extractor based on a combination model of variational autoencoder and generative adversarial network (VAE-GAN). Additionally, we propose a novel instance-level feature selection method to select the discriminative instance features. Furthermore, we employ a graph convolutional network (GCN) for learning the bag-level representation and then performing the classification. We apply the proposed method in the prediction of lymph node metastasis using histopathological images of colorectal cancer. Experimental results demonstrate that the proposed method achieves superior performance compared to the state-of-the-art methods.

1. Introduction

Recently, weakly-supervised learning (WSL) has gained greater attention in the machine learning field since it significantly reduces the workload of human annotation. Multi-instance learning (MIL) is a typical weakly-supervised learning [48], which has been widely employed in different tasks, including object detection [37, 38, 18], semantic segmentation [43, 33], scene classification [42, 19], medical diagnosis[5, 31], etc. In the MIL task, the training dataset is composed of bags, where each bag contains a set of instances. The goal of MIL is to learn a model for predicting the bag label. Different from conventional fully-supervised machine learning problems, where each instance has a confident label, only the bag-level label is available in MIL. Furthermore, instances in a bag are not necessarily relevant, sometimes even providing confusing information. For example, some instances do not contain discriminative information related to the bag class, or they are more related to other classes of bags.[2]

Based on which level the discriminative information is at (instance-level or bag-level) and how the relevant information is extracted (implicitly or explicitly), MIL algorithms can be categorized into three groups, i.e., instance-space paradigm, bag-space paradigm, and embedded-space paradigm [41, 2]. The instance-space paradigm tends to focus on local information, which learns instance classifier at the first stage and then achieves the bag-level classifier by simply aggregating instance-level results. These instance-space methods are mostly based on the standard multiple instance (SMI) assumption [27], i.e., a bag is positive only if it contains at least one positive instance and otherwise is negative [46, 3, 30]. However, this key-instance based SMI assumption is inappropriate in applications where the classification is based on the global bag information instead of an individual instance. The bag-space paradigm and embedded-space paradigm, on the other hand, extract discriminative information from the whole bag. The difference between these two paradigms lies in the way to exploit the bag-level information. The bag-space paradigm implicitly utilizes bag-to-bag distance/similarity, while the embedded-space paradigm explicitly embeds the information of a bag into a feature space [41].

In this paper, we propose a novel embedded-space multiple instance learning method with feature selection and graph convolutional network for image classification. The method has three major components: instance-level feature

*Yu Zhao and Fan Yang contributed equally and Jianhua Yao is the corresponding author (jianhuayao@tencent.com).

extraction, instance-level feature selection, and bag-level classification. Our method is developed for the prediction of lymph node metastasis using histopathological images of colorectal cancer. Lymph node metastasis (LNM) from colorectal cancer is a major factor in patient management and prognosis [13, 9, 39]. Patients diagnosed with LNM should undergo lymph node dissection surrounding the colon region [7]. This research has great clinical value since LNM pre-surgical detection indicates the necessity of lymph node dissection to prevent further spreading. This is a challenging task and we tackle it in the following aspects. (1) The size of a whole slide image (WSI) is usually very large (around 100000×50000 pixels in our case). Given the current computational resource, it is infeasible to load the WSI into the deep neural networks. Therefore we divide the WSI into a set of image patches (512×512 pixels) and treat the WSI as a bag of patches. In this way, the prediction problem is formalized as an embedded-space multiple instance learning task. (2) To the best of our knowledge, there is no prior work or knowledge that indicates useful features for metastasis prediction. Similar as other image classification works, we employ deep neural networks to automatically extract latent features from the image [26]. Moreover, in our case where the instance labels are not available, conventional methods usually utilize a pre-trained model (e.g., trained on ImageNet) as the feature extractor. However, the domain gap between natural scene images and histopathological images may compromise the performance of the pre-trained model on histopathological images [34]. To solve this problem, we develop a combination model of variational autoencoder and generative adversarial network (VAE-GAN) to train the feature extractor in a self-supervised way. (3) Generating effective representation for all instances in a bag is not a trivial problem. Various methods such as max pooling, average pooling, log-sum-exp pooling [41], dynamic pooling [44], and adaptive pooing [29] have been proposed. However, these operators are either non-trainable or too simple. In this paper, we apply the graph convolutional network (GCN) for generating the bag representation, which is fully trainable. (4) Features extracted from the instances are redundant. We propose a novel feature selection approach to remove the indiscriminative instance-level features.

To summarize, the contributions of this paper include:

1) We propose an embedded-space deep multiple instance learning method with GCN for the prediction of lymph node metastasis in colorectal cancer on histopathological images. To the best of our knowledge, this is the first method tackling this challenging clinical problem.

2) We design a VAE-GAN model to generate instance and use the trained encoder component of the VAE-GAN as the feature extractor. With this setting, we can train the feature extractor in a self-supervised way, without knowing the instance label.

3) We develop a novel feature selection approach working on instances to select discriminative features for final bag representation. The proposed method utilizes a histogram to build a bag-level representation of this feature and then use the maximum mean discrepancy (MMD) [14, 40] of the obtained bag-level representation between positive and negative bags to evaluate the feature significance.

4) We apply a GCN for generating the bag representation and bag-level classification, which is fully trainable.

2. Related Work

2.1. Deep Multiple Instance Learning

Combined with deep features, multiple instance learning has shown great representation power in recent studies [42, 36, 41, 6, 16, 17]. Wu *et al*. [42] utilized a MIL neural network to simultaneously learn the object proposals and text annotation. Sun el al. [36] proposed a weakly supervised MIL network for object recognition on natural images, which solved the inaccurate instance label problem in data augmentation. Hou *et al*. [17] showed that a decision classifier based on MIL can boost the performance in classifying glioma and non-small-cell lung carcinoma by aggregating instance-level predictions. In this paper, we follow this line of research and employ MIL to solve a challenging clinical problem of predicting lymph node metastasis from histopathological images of the primary tumor region.

2.2. Bag Representation

In a MIL task, generating bag representation is a crucial step. Pooling methods such as max pooling, average pooling, and log-sum-exp pooling [41] are typically adopted in this step. However, these pooling methods are not trainable which may limit their applicability. A novel dynamic pooling method iteratively renewing bag information from instance was proposed in [44]. Kraus *et al*. [24] proposed a noisy-and pooling layer against outliers, which demonstrated promising results in microscopy images. Zhou *et al*. [47] proposed an adaptive pooling method that can be dynamically adjusted to various classes in video caption tagging. These methods are partly trainable with restricted flexibility. Ilse *et al*. [19] proposed a fully trainable method utilizing the attention mechanism to allocate weights to instances. However, this method considers the bag representation as a weighted sum of instance features, which is just a linear combination. Different from the approaches mentioned above, our work builds the bag representation with

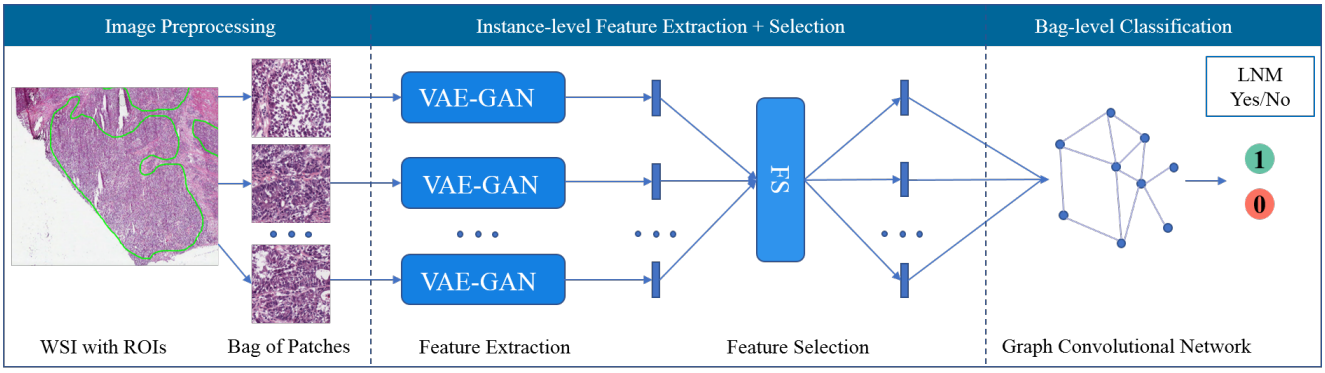


Figure 1. The overall framework of the proposed approach. It consists of image preprocessing, instance-level feature extraction, instance-level feature selection, and bag-level classification. VAE-GAN works as an instance-level feature extractor. The feature selection procedure selects discriminative instance-level features. The GCN is used to synthesize selected instance-level features, generate bag representation and perform the final classification.

a GCN, which is fully trainable and integrates the patch information into a complex and high-level representation.

2.3. Pathological Image Analysis

In clinical practice, pathology image analysis is the gold standard for cancer diagnosis. Nowadays, the development of deep neural networks has made many breakthroughs in automatic pathology image analysis and assisted diagnosis [50, 5, 21]. As mentioned above, MIL is naturally suitable for pathological image analysis due to the vast image. Authors in [21] recently reported an instance-space MIL method applied in the prediction of microsatellite instability (MSI). In the first step of this method, they assigned the bag label to its patches and then trained a ResNet with the patch-label pairs. In the second stage, the trained ResNet was used to generate the patch-level prediction and then all these patch-level prediction results are aggregated with the majority voting strategy. Another instance-space approach is the Whole Slide Histopathological Images Survival Analysis framework (WSISA) [50] for survival prediction. This method first unsupervisely clustered patches into different clusters and then selected useful clusters by evaluating patch-level classification performance. After that, it aggregated patch-level features from selected clusters to make the patient-level prediction. The success of the above two methods implies that these tasks meet the SMI assumption. However, these instance-space MIL methods have their limitations. They are suitable for the tasks where discriminative information is considered to lie at the instance level and there exist key instances whose labels are strongly related to bag-level labels. The above conditions do not hold in our problem.

Recently, Campanella *et al.* [5] proposed an embedded-space MIL method with applying a recurrent neural network (RNN) to integrate patch information extracted from the WSI. They treated each WSI as a bag and considered all the patches of the WSI as sequential inputs to the RNN. This model was trained on three huge datasets and successfully classified the sub-types of three cancers. Meanwhile, attention-based deep multiple instance learning is another currently proposed embedded-space MIL method [19]. It achieved the state-of-the-art performance on classifying epithelial cells in colon cancer by training on large-scale data. These two methods utilize different approaches to integrate instance information for bag representation, which are both end-to-end trainable. However, the end-to-end method requires the network to extract the instance features and generate bag representation simultaneously with only bag-level classification error as supervision, which makes the network hard to train, especially when lacking sufficient training data. Therefore, we propose a feature selection component in our method to remove the redundant and unhelpful features to alleviate the workload of the network for generating the bag representation and performing the bag-level classification. Considering the specialty of our problem, we also equip our MIL method with GCN to take advantage of the structure information among instances.

3. Methods

3.1. Overview

The overall framework of the proposed method is illustrated in Fig. 1. The whole pipeline consists of four steps: image preprocessing, instance-level feature extraction with VAE-GAN (3.2), instance-level feature selection (3.3), and bag-level classification with graph convolutional network (3.4). In the image preprocessing step, the tumor areas manually annotated by pathologists are selected as the regions of interest (ROIs) referring to the manually annotated labels obtained from clinical experts in our team. Then, the ROIs are divided into non-overlapping patches of size 512×512. The details of the other three components are illustrated in

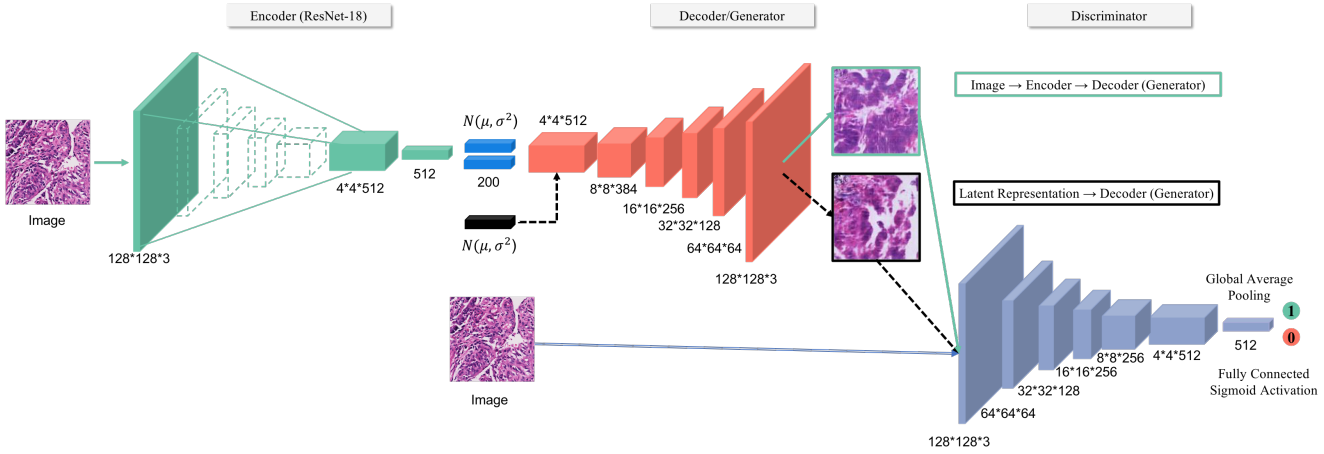


Figure 2. The architecture of the VAE-GAN. The ResNet-18 is used as the encoder of the VAE. The decoder of VAE and generator of GAN share the same component in VAE-GAN.

the remainder of this section.

3.2. VAE-GAN

A variational autoencoder (VAE) [23] is comprised of an encoder which encodes input data x to a latent representation h, and a decoder which decodes the latent representation h back to the original data-space. In order to regularize the encoder of the VAE, a prior over the latent distribution $p(h)$ is usually imposed. In this work, we use the Gaussian distribution, i.e. $N \sim (0, I)$ to regularize the encoder. The VAE loss [23, 25] is formulated as:

$$\begin{aligned}\mathscr{L}_{VAE} &= \mathscr{L}^{pixel}_{LLike} + \mathscr{L}_{KL} \\ &= -E_{q(h|x)}[log(p(x|h))] + D_{KL}(q(h|x)||p(h)),\end{aligned} \tag{1}$$

where D_{KL} denotes the Kullback-Leibler divergence.

A generative adversarial network (GAN) consists of a generator network G which aims to map the latent representation h to data space, and a discriminator network D which aims to distinguish the generated fake data from the real data. The loss of GAN is defined as:

$$\mathscr{L}_{GAN} = log(D(x)) + log(1 - D(G(h))) \tag{2}$$

A VAE-GAN is a combination of a VAE and a GAN, where the decoder of VAE and generator of GAN share the same component [25]. In a VAE-GAN architecture, the original VAE reconstruction error $\mathscr{L}^{pixel}_{LLike}$ is replaced by the reconstruction error expressed in the GAN discriminator. To be specific, let $Dis_l(x)$ denote the hidden representation of the l^{th} layer of the discriminator. A Gaussian observation model for $Dis_l(x)$ with mean $Dis_l(\tilde{x})$ and identity covariance is introduced:

$$p(Dis_l(x)||h) = N(Dis_l(x)|Dis_l(\tilde{x}), I) \tag{3}$$

where $\tilde{x} \sim Decoder(h)$ is the sample from the decoder of x, then the reconstruction error in the GAN discriminator can be denoted as follows:

$$\mathscr{L}^{Dis_l}_{LLike} = -E_{q(h|x)}[logp(Dis_l(x)|h)] \tag{4}$$

Using $\mathscr{L}^{Dis_l}_{LLike}$ replacing $\mathscr{L}^{pixel}_{LLike}$, we can obtain the loss function of entire VAE-GAN [25]:

$$\mathscr{L} = \lambda_{Dis} * \mathscr{L}^{Dis_l}_{LLike} + \lambda_{KL} * \mathscr{L}_{KL} + \lambda_{GAN} * \mathscr{L}_{GAN} \tag{5}$$

where λ_{Dis}, λ_{KL} and λ_{GAN} are the hyperparamters of the VAE-GAN loss.

Different from the conventional goal of GANs, the main function of the VAE-GAN in our work is for training or fine-tuning the encoder component which will be used as an instance-level feature extractor. The detailed architecture of our VAE-GAN can be found in Fig. 2. The widely utilized ResNet-18 [15] acts as the encoder. The decoder of VAE and generator of GAN share the same component, which incorporates five up-sampling stacks. Each up-sampling stack contains a transposed convolution followed by batch normalization and the rectified linear unit (ReLU) as the activation function. The discriminator is made up of five down-sampling stacks. In each encoder stack, there is one convolution layer followed by the batch normalization and LeakyReLU activation function.

3.3. Feature Selection

The feature selection procedure chooses the most discriminative instance-level features for generating the bag representation. This step is especially important in medical image analysis tasks due to lack of training data. Removing redundant or irrelevant features can also alleviate the workload and simplify the following learning task. Unlike most feature selection problems where there are feature-label pairs, in our task, the instance-level feature has no

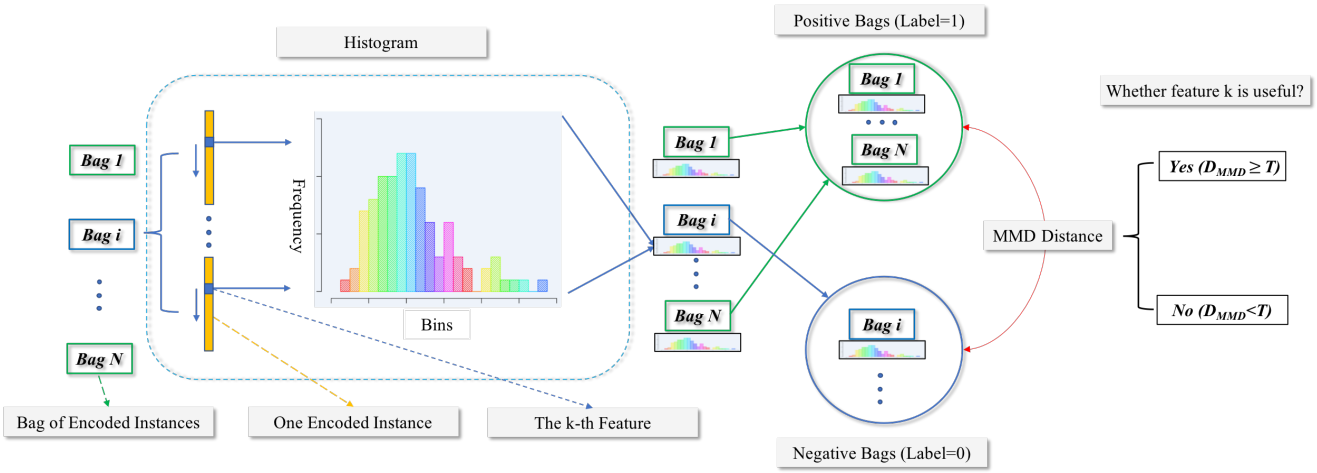


Figure 3. The pipeline of the feature selection component. Each bag has a various number of instances (features vectors). When evaluating the discriminating value of a feature for bag-level classification, the histogram acts as a bridge connecting the instance-level feature and the bag-level label. For instance, when evaluating the k-th feature, feature k is chosen as the representation of the instance. A histogram of this feature is calculated in each bag and then these histograms are utilized as the representations of a bag. After that, The MMD distance are calculated between positive bags and negative bags using these bag representations. The feature is regarded as discriminative if the MMD distance is large.

associated label and is just assigned a bag-level label. We need to build a bridge between the extracted instance feature and the bag label. As demonstrated in Fig. 3, we utilize the histogram [2] as the bridge and the maximum mean discrepancy [4] as the criterion to evaluate the feature importance.

Assume we have N bag-label pairs denoted as $\{X_1, X_2, ..., X_N\}$ and $\{Y_1, Y_2, ..., Y_N\}$, where the i^{th} bag contains K_i instances represented as $\{x_1^i, x_2^i, ..., x_{K_i}^i\}$. In our case, $x_j^i \in \mathbb{R}^D$ is the j^{th} extracted instance features of i^{th} bag and $Y_i \in \{0, 1\}$ denotes whether there exists LNM in the bag. To make it easier to express, we use $F = [f_1, f_2, ..., f_D]$ to represent the extracted instance features, i.e, x_j^i is a sampler of F. Our goal is to evaluate the importance of each extracted instance feature $f_k = F[k]$, which is achieved through the following two steps: (1) Generate a histogram of every feature in each bag with N_b bins of equal widths (which reflects the distribution of the feature in a bag). (2) Use the histogram as the bag representation and calculate the difference of obtained histograms between positive and negative labels to assess the discriminating value of a feature for classification.

3.3.1 Histogram Generation

For feature f_k, we calculate the maximum value and minimum value of this feature among all instances in all bags.

$$f_k^{max} = max\{x_j^i[k]\}, (i = 1, ..., N, j = 1, ..., K_i) \quad (6)$$

$$f_k^{min} = min\{x_j^i[k]\}, (i = 1, ..., N, j = 1, ..., K_i) \quad (7)$$

Then, we divide the range $[f_k^{min}, f_k^{max}]$ into N_b bins of equal widths and map each bag X_i into a histogram $H_k^i = (h_1^{i,k}, ..., h_{N_b}^{i,k})$, where h_o^i indicates the percentage of instances in X_i with feature f_k located in the o^{th} bin.

$$h_o^{i,k} = \frac{1}{K_i} \sum_{x_j^i \in \mathbf{X_i}} f_o(x_j^i[k]), \quad (8)$$

where $o = 1, ..., N_b$ and $j = 1, ..., K_i$. $f_o(x_j^i[k]) = 1$ if $x_j^i[k]$ is located in the o^{th} bin, otherwise $f_o(x_j^i[k]) = 0$.

3.3.2 Feature Evaluation

After obtaining the histograms of feature f_k of all bags $\{H_k^1, ..., H_k^N\}$, we evaluate the importance of the feature by the MMD distance, which is defined as:

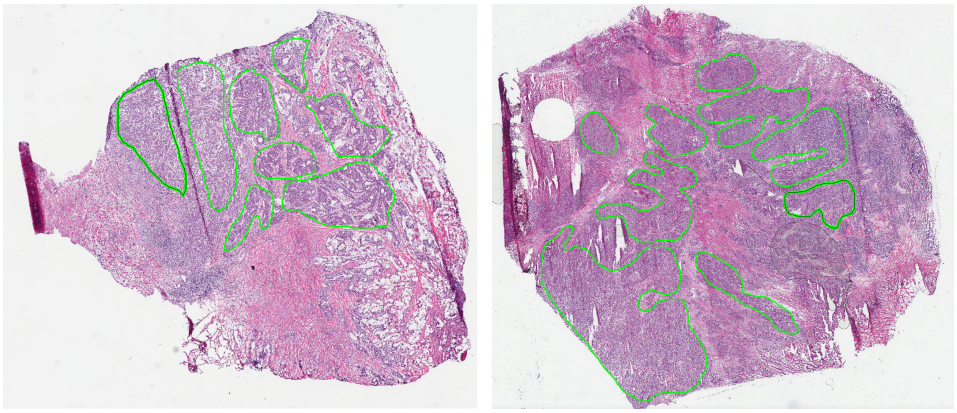
$$D(f_k) = \| \frac{1}{|G_P|} \sum_{X_i \in G_P} \phi(H_k^i) - \frac{1}{|G_N|} \sum_{X_j \in G_N} \phi(H_k^j) \|, \quad (9)$$

where G_P and G_N are the groups of all positive bags and negative bags respectively and ϕ is a mapping function. Bigger MMD distance means it is easier to discriminate the positive group from the negative group.

3.4. GCN-based Multiple Instance Learning

3.4.1 Graph Construction

We formulate our proposed network, i.e. GCN-based MIL as follows. Similar to [49], we utilize a heuristic approach to construct a graph from a bag of instance features $[x_1^i, x_2^i, ..., x_K^i]$ (K can variate for different bags.). First the adjacency



(a) Positive sample (b) Negative sample

Figure 4. Examples illustrating the whole slide image (WSI) from the cancer genome atlas (TCGA) dataset. The regions within the green contours are the colorectal cancer regions which are annotated by clinical experts. Sub-figure (a) shows a positive sample, i.e. WSI from patient with Lymph node metastasis (LNM). Sub-figure (b) demonstrates a negative sample.

matrix A can be obtained:

$$A_{mn} = \begin{cases} 1 & if \quad dist(x_p^i, x_q^i) < \gamma \\ 0 & otherwise \end{cases} \tag{10}$$

where $dist(x_p^i, x_q^i)$ is the distance between p^{th} and q^{th} instance feature in bag i. Here we use Euclidean distance to calculate $dist$. γ determines whether there is an edge connecting two instances. γ = 0 represents there is no edge connecting x_p^i and x_q^i while $\gamma = +\infty$ represents the input is a fully connected graph. At the same time, the instance features [x_1^i, x_2^i, ..., x_K^i] of bag i are considered to be the nodes of a graph. Then we obtain the graph of bag i as:

$$G_i = G(A_i, E_i) \tag{11}$$

where $A_i \in \{0,1\}^{K \times K}$ represents the adjacency matrix, $E_i \in \mathbb{R}^{K \times D}$ means node feature matrix constructed from a bag of X_i (D is the feature dimension).

3.4.2 Spectral Graph Convolution

Given a graph $G = (V, E)$, its normalized graph Laplacian $L = I - D^{-1/2}AD^{-/2}$. D is the degree matrix of G and A is the adjacency matrix mentioned above. Following the work in [11], we formulated kernel as a M^{th} order polynomial of diagonal Λ, and $diag(\Lambda)$ is the spectrum of graph Laplacian L:

$$g\theta(\Lambda^M) = \sum_{m=0}^{M-1} \theta_m \Lambda^m \tag{12}$$

Spectral convolution on graph G with vertex features $X \in \mathbb{R}^{N \times F}$ as layer input can be further obtained [10]:

$$Y = ReLU(g_\theta(L^M)X) \tag{13}$$

where ReLU is the commonly used activation function, $Y \in \mathbb{R}^{N \times F}$ is a graph of the identical number of vertices with convolved features. Chebyshev expansion is used to approximate $g_\theta(L)$ in order to accelerate filtering [11].

3.4.3 Network Architecture

Our GCN-based MIL network employs three stacked graph convolution layers (3.4.2), each followed by a ReLU activation and a self-attention graph pooling layer [28], to generate the bag representation (node embedding of the graph). After that, two fully connected layers with ReLU and a sigmoid activation function are utilized to achieve the bag-level classification. The categorical cross-entropy loss is used to optimize the network, which is defined as:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} \delta(y_i = c) \log(P(y_i = c)) \tag{14}$$

where N denotes the data number and C represents the categories number. $\delta(y_i = c)$ is the indicator function and $P(y_i = c)$ is the predicted probability by the model.

4. Results and Discussion

4.1. Dataset

In this study, the Colon Adenocarcinoma (COAD) cohort of the Cancer Genome Atlas (TCGA) dataset [20] is used to evaluate our proposed method. This publicly released dataset contains 425 patients with colorectal cancer. For each patient, a H&E-stained histology WSI is acquired from the tumor region. Based on the clinical tumor node metastasis (TNM) staging information, these patients can be categorized into two groups: one without lymph node metastasis (patients in N_0 stage) and one with lymph node

metastasis (patients in stage from N_1 to N_4). There are 174 positive samples (patients with LNM) and 251 negative samples (patients without LNM) in the dataset. Fig. 4 shows one positive sample and one negative sample. Even experienced doctors and pathologists in our team cannot distinguish the patient with or without LNM if only relying on the histopathological image.

4.2. Evaluation

The area under the receiver operating characteristic curve (ROCAUC) together with the accuracy, precision, recall and F1-score are used to evaluate the performance of our proposed method and the state-of-the-art approaches. Specifically, these metrics are defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

$$\text{F1-score} = \frac{2 * (Recall * Precision)}{Recall + Precision} \tag{18}$$

Where TP, FP, TN, and FN represent the True Positive, False Positive, True Negative and False Negative respectively. Among these, ROCAUC is more comprehensive when comparing the performance of different methods.

Table 1. Hyperparameters of the proposed methods.

Hyperparameters	Value
VAE-GAN	
Loss weight λ_{Dis}	1
Loss weight λ_{KL}	1
Loss weight λ_{GAN}	1
Feature Selection	
Histogram bin number N_b	50
Feature selection rate	50%
GCN	
Distance threshold γ	$0.5 * \max_{x_p, x_q \in X} \{dist(x_p, x_q)\}$

4.3. Experimental Setup

In our experiment, the entire dataset (425 samples) is randomly divided into a training set (354 samples) and a test set (71 samples) in a ratio of 5:1. We perform five-fold cross-validation on the training dataset for parameter tuning purposes. We implement the VAE-GAN and GCN with PyTorch [32] and PyTorch Geometric library [12]. We randomly initialize the VAE-GAN under the default setting of PyTorch and resize the input images to 128×128 pixels. The Adam optimizer [22] is used to train both the VAE-GAN and GCN. To tackle the class imbalance problem during the bag-level classification stage, we employ the "WeightedRandomSampler" strategy [32] to prepare each training batch. Other related hyperparameters of each stage of the proposed method are given in Table 1.

4.4. Ablation Study

To evaluate the effectiveness of different components in the proposed method, we conduct ablation studies. We experiment on the following configurations: (A) Our proposed method: VAE-GAN + FS + GCN. (B) VAE-GAN + GCN: our framework without feature selection. (C) Pre-trained ResNet + FS + GCN: our framework but using pre-trained ResNet-18 (on ImageNet) as the instance-level feature extractor. (D) Pre-trained ResNet + GCN: using the pre-trained ResNet-18 (on ImageNet) as instance-level feature extractor and then using the GCN for bag-level classification directly without feature selection. (E) Pre-trained ResNet + GCN (end-to-end): This configuration is similar to (D), while the difference lies in that (E) is an end-to-end network, i.e, the back-propagated loss from the GCN can guide the training of the instance-level feature extraction network (ResNet-18).

The results of all these configurations are illustrated in Table 2. Comparing (A) to (B) and (C) to (D), it is noted that the use of the proposed feature selection procedure improves ROCAUC by 3.3% and 3.0% respectively. Similarly, when comparing (A) to (C) and (B) to (D), the VAE-GAN results in 7.3% and 7.1% performance gain. The comparison between (A) and (E) shows the two-stage method performs better than the end-to-end approach. Although the end-to-end configuration (E) can extract instance-level features and generate bag representation together, tackling both tasks simultaneously imposes heavy workload to optimize the whole network with only bag-level classification supervision, especially when lacking sufficient training data. Therefore, the two-stage method which separately handles instance-level feature extraction and bag-level representation is more appropriate for our task.

4.5. Comparison with State-of-the-Art Methods

Table 3 demonstrates the comparison between our proposed method and other state-of-the-art methods including (1) T-stage + LR, (2) Histomics + Histogram [8], (3) ResNet + voting [21], (4) WSISA [45], (5) ResNet + RNN [5], (5) Attention based MIL [19]. We reimplement all the previous methods based on the literatures and open source codes. From the table, we can observe that our approach outperforms these methods.

T-stage + LR and Histomics + Histogram [8] are two machine learning algorithms based on hand crafted features. T-

Table 2. The results of the ablation study.

Method \ Metric	Accuracy	Precision	Recall	F1-score	ROCAUC
(A) VAE-GAN + FS + GCN (OUR)	**0.6761**	**0.575**	0.7931	**0.6667**	**0.7102**
(B) VAE-GAN + GCN	0.5775	0.4902	**0.8621**	0.6250	0.6773
(C) Pre-trained ResNet + FS + GCN	0.4225	0.4032	**0.8621**	0.5495	0.6371
(D) Pre-trained ResNet + GCN	0.5634	0.4792	0.7931	0.5974	0.6067
(E) Pre-trained ResNet + GCN (End-to-End)	0.4648	0.4182	0.7931	0.5476	0.6010

Table 3. Comparisons between our proposed method and the state-of-the-art approaches.

Method \ Evaluation	Accuracy	Precision	Recall	F1-score	ROCAUC
Our	**0.6761**	0.575	0.7931	**0.6667**	**0.7102**
T-stage + LR	0.6357	**0.7143**	0.1887	0.2985	0.6471
Histomics + Histogram [8]	0.6124	0.5484	0.3208	0.4048	0.6157
ResNet + Voting [21]	0.5891	0.5	0.3208	0.3908	0.5824
WSISA [45]	0.5969	0.5152	0.3208	0.3953	0.5792
ResNet + RNN [5]	0.4109	0.4109	**1**	0.5824	0.5
Attention based MIL [19]	0.5891	0.5	0.3208	0.3908	0.5457

stage is a factor describing the invasion depth of the tumor into the intestinal wall [1]. Recent researches report that the depth of tumor invasion is related to the lymph node metastasis [35]. The T-stage + LR method utilizes the T-stage information as the feature and adopts the logistic regression to predict lymph node metastasis of patients in colorectal cancer. Histomics + Histogram method [8] extracts cell morphologic features including nucleus shape, intensity, texture, and the spatial relationship between nuclei as features. It utilizes a histogram to analyze the cell distribution in the WSI and uses the lasso regression [8] to predict the prognosis of patients.

The T-stage + LR method is only based on the T-stage feature, which lacks the information of local cancer cell texture and global histology of the tumor. The Histomics + Histogram method utilizes specifically-designed features, which limits its extension ability. For instance, the cell morphologic features maybe not suitable for the LNM prediction on colorectal cancer because cancer cells are similar and do not change during the propagation.

As mentioned in section 2.3, the ResNet + voting method [21] and WSISA are typical instance-space MIL methods. These methods work well if the discriminative information is considered to lie at the instance level and there exist key instances which are strongly related to the bag-level labels. However, the conditions do not hold in our task and therefore these two methods perform poorly.

The performances of the ResNet + RNN and Attention-based MIL methods are inferior in the LNM prediction task compared to our proposed method. This might be due to three reasons: (1) Extracting instance-level features and generating bag representation together impose heavy workload on the end-to-end network. (2) The network has learnt many unhelpful features which should be removed before generating the bag representation. (3) The RNN and attention mechanism are not as good as the GCN in the bag-level classification stage.

5. Conclusion

In this paper, we investigate a challenging clinical task of automatic prediction of lymph node metastasis using histopathological images of colorectal cancer. To achieve this, we develop a deep GCN-based MIL method combined with a feature selection strategy. Experimental results demonstrate that our method benefits from our proposed components, including (1) VAE-GAN for instance-level feature extraction, (2) instance-level feature selection and (3) GCN-based MIL for bag representation and bag-level classification. Our approach shows superior performance compared to the state-of-the-art methods. In the future, it would be meaningful to develop a unified GCN model for performing a joint instance selection and instance-level feature selection with the weak label in a bag level.

References

[1] Mahul B Amin, Frederick L Greene, Stephen B Edge, Carolyn C Compton, Jeffrey E Gershenwald, Robert K Brookland, Laura Meyer, Donna M Gress, David R Byrd, and

David P Winchester. The eighth edition ajcc cancer staging manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: a cancer journal for clinicians*, 67(2):93–99, 2017. 8

[2] Jaume Amores. Multiple instance classification: Review, taxonomy and comparative study. *Artificial intelligence*, 201:81–105, 2013. 1, 5

[3] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann. Support vector machines for multiple-instance learning. In *Advances in neural information processing systems*, pages 577–584, 2003. 1

[4] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006. 5

[5] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019. 1, 3, 7, 8

[6] Marc-André Carbonneau, Eric Granger, and Ghyslain Gagnon. Bag-level aggregation for multiple-instance active learning in instance classification problems. *IEEE transactions on neural networks and learning systems*, 30(5):1441–1451, 2018. 2

[7] George J Chang, Miguel A Rodriguez-Bigas, John M Skibber, and Virginia A Moyer. Lymph node evaluation and survival after curative resection of colon cancer: systematic review. *Journal of the National Cancer Institute*, 99(6):433–441, 2007. 2

[8] Jun Cheng, Jie Zhang, Yatong Han, Xusheng Wang, Xiufen Ye, Yuebo Meng, Anil Parwani, Zhi Han, Qianjin Feng, and Kun Huang. Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. *Cancer research*, 77(21):e91–e100, 2017. 7, 8

[9] Kenneth SH Chok and Wai Lun Law. Prognostic factors affecting survival and recurrence of patients with pt1 and pt2 colorectal cancer. *World journal of surgery*, 31(7):1485–1490, 2007. 2

[10] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997. 6

[11] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852, 2016. 6

[12] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. 7

[13] Leonard L Gunderson, Daniel J Sargent, Joel E Tepper, Norman Wolmark, Michael J O'Connell, Mirsada Begovic, Cristine Allmer, Linda Colangelo, Steven R Smalley, Daniel G Haller, et al. Impact of t and n stage and treatment on survival and relapse in adjuvant rectal cancer: a pooled analysis. *Journal of Clinical Oncology*, 22(10):1785–1796, 2004. 2

[14] Philip Haeusser, Thomas Frerix, Alexander Mordvintsev, and Daniel Cremers. Associative domain adaptation. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 2

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4

[16] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. Efficient multiple instance convolutional neural networks for gigapixel resolution image classification. *arXiv preprint arXiv:1504.07947*, page 7, 2015. 2

[17] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2424–2433, 2016. 2

[18] Fang Huang, Jinqing Qi, Huchuan Lu, Lihe Zhang, and Xiang Ruan. Salient object detection via multiple instance learning. *IEEE Transactions on Image Processing*, 26(4):1911–1922, 2017. 1

[19] Maximilian Ilse, Jakub M Tomczak, and Max Welling. Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712*, 2018. 1, 2, 3, 7, 8

[20] Cyriac Kandoth, Michael D McLellan, Fabio Vandin, Kai Ye, Beifang Niu, Charles Lu, Mingchao Xie, Qunyuan Zhang, Joshua F McMichael, Matthew A Wyczalkowski, et al. Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471):333, 2013. 6

[21] Jakob Nikolas Kather, Alexander T Pearson, Niels Halama, Dirk Jäger, Jeremias Krause, Sven H Loosen, Alexander Marx, Peter Boor, Frank Tacke, Ulf Peter Neumann, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature medicine*, page 1, 2019. 3, 7, 8

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[23] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 4

[24] Oren Z Kraus, Jimmy Lei Ba, and Brendan J Frey. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics*, 32(12):i52–i59, 2016. 2

[25] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*, 2015. 4

[26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 2

[27] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808, 2007. 1

[28] Junhyun Lee, Inyeop Lee, and Jaewoo Kang. Self-attention graph pooling. In *International Conference on Machine Learning*, pages 3734–3743, 2019. 6

[29] D. Liu, Y. Zhou, X. Sun, Z. Zha, and W. Zeng. Adaptive pooling in multi-instance learning for web video annotation. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 318–327, Oct 2017. 2

[30] Guoqing Liu, Jianxin Wu, and Zhi-Hua Zhou. Key instance detection in multi-instance learning. In *Asian Conference on Machine Learning*, pages 253–268, 2012. 1

[31] Siyamalan Manivannan, Caroline Cobb, Stephen Burgess, and Emanuele Trucco. Subcategory classifiers for multiple-instance learning and its application to retinal nerve fiber layer visibility classification. *IEEE transactions on medical imaging*, 36(5):1140–1150, 2017. 1

[32] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017. 7

[33] Deepak Pathak, Philipp Krahenbuhl, and Trevor Darrell. Constrained convolutional neural networks for weakly supervised segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1796–1804, 2015. 1

[34] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning with applications to medical imaging. *Advances in neural information processing systems*, 2019. 2

[35] TJ Saclarides, Achyut K Bhattacharyya, C Britton-Kuzel, D Szeluga, and SG Economou. Predicting lymph node metastases in rectal cancer. *Diseases of the colon & rectum*, 37(1):52–57, 1994. 8

[36] Miao Sun, Tony X Han, Ming-Chang Liu, and Ahmad Khodayari-Rostamabad. Multiple instance learning convolutional neural networks for object recognition. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 3270–3275. IEEE, 2016. 2

[37] Peng Tang, Xinggang Wang, Angtian Wang, Yongluan Yan, Wenyu Liu, Junzhou Huang, and Alan Yuille. Weakly supervised region proposal network and object detection. In *Proceedings of the European conference on computer vision (ECCV)*, pages 352–368, 2018. 1

[38] Fang Wan, Chang Liu, Wei Ke, Xiangyang Ji, Jianbin Jiao, and Qixiang Ye. C-mil: Continuation multiple instance learning for weakly supervised object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2199–2208, 2019. 1

[39] Hao Wang, Xian-Zhao Wei, Chuan-Gang Fu, Rong-Hua Zhao, and Fu-Ao Cao. Patterns of lymph node metastasis are different in colon and rectal carcinomas. *World Journal of Gastroenterology: WJG*, 16(42):5375, 2010. 2

[40] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018. 2

[41] Xinggang Wang, Yongluan Yan, Peng Tang, Xiang Bai, and Wenyu Liu. Revisiting multiple instance neural networks. *Pattern Recognition*, 74:15–24, 2018. 1, 2

[42] Jiajun Wu, Yinan Yu, Chang Huang, and Kai Yu. Deep multiple instance learning for image classification and auto-annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3460–3469, 2015. 1, 2

[43] Gang Xu, Zhigang Song, Zhuo Sun, Calvin Ku, Zhe Yang, Cancheng Liu, Shuhao Wang, Jianpeng Ma, and Wei Xu. Camel: A weakly supervised learning framework for histopathology image segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10682–10691, 2019. 1

[44] Yongluan Yan, Xinggang Wang, Xiaojie Guo, Jiemin Fang, Wenyu Liu, and Junzhou Huang. Deep multi-instance learning with dynamic pooling. In *Asian Conference on Machine Learning*, pages 662–677, 2018. 2

[45] Jiawen Yao, Xinliang Zhu, and Junzhou Huang. Deep multi-instance learning for survival prediction from whole slide images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 496–504. Springer, 2019. 7, 8

[46] Qi Zhang and Sally A Goldman. Em-dd: An improved multiple-instance learning technique. In *Advances in neural information processing systems*, pages 1073–1080, 2002. 1

[47] Yizhou Zhou, Xiaoyan Sun, Dong Liu, Zhengjun Zha, and Wenjun Zeng. Adaptive pooling in multi-instance learning for web video annotation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 318–327, 2017. 2

[48] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2017. 1

[49] Zhi-Hua Zhou, Yu-Yin Sun, and Yu-Feng Li. Multi-instance learning by treating instances as non-iid samples. In *Proceedings of the 26th annual international conference on machine learning*, pages 1249–1256. ACM, 2009. 5

[50] Xinliang Zhu, Jiawen Yao, Feiyun Zhu, and Junzhou Huang. Wsisa: Making survival prediction from whole slide histopathological images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7234–7242, 2017. 3

6

A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes

This chapter has been published as **peer-reviewed conference paper:**.

Y. Zhao, P. Wu, J. Wang, H. Li, N. Navab, I. Yakushev, W. Weber, M. Schwaiger, S.-C. Huang, P. Cumming, et al. "A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes on ^{18}F-FDG PET Images." In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 3531–3534. DOI: 10.1109/EMBC.2019.8856747

Synopsis: This work deals with the problem of parkinsonian syndrome diagnosis. we developed a 3D deep convolutional neural network utilizing ^{18}F-fluorodeoxyglucose (FDG) PET images for the automated differential diagnosis of idiopathic Parkinson's disease (IPD) from multiple system atrophy (MSA) and progressive supranuclear palsy (PSP). And we depicted in saliency maps the decision mechanism of the deep learning method to assist the physiological interpretation of deep learning performance.

Contributions of thesis author: algorithm design and implementation, computational experiments and composition of manuscript.

Accepted manuscript for the IEEE (41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society , 2019)
Published version: DOI: 10.1109/EMBC.2019. 8856747

A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes on ^{18}F-FDG PET Images

Yu Zhao[1], Ping Wu[2], Jian Wang[3], Hongwei Li[1] Nassir Navab[1], Igor Yakushev[4], Wolfgang Weber[4], Markus Schwaiger[4], Sung-Cheng Huang[5], Paul Cumming[6,7], Axel Rominger[6], Chuantao Zuo[2], Kuangyu Shi[1,6]

***Abstract*— Idiopathic Parkinsons disease and atypical parkinsonian syndromes have similar symptoms at early disease stages, which makes the early differential diagnosis difficult. Positron emission tomography with ^{18}F-FDG shows the ability to assess early neuronal dysfunction of neurodegenerative diseases and is well established for clinical use. In the past decades, machine learning methods have been widely used for the differential diagnosis of parkinsonism based on metabolic patterns. Unlike these conventional machine learning methods relying on hand-crafted features, the deep convolutional neural networks, which have achieved significant success in medical applications recently, have the advantage of learning salient feature representations automatically and effectively. This advantage may offer more appropriate invisible features extracted from data for the enhancement of the diagnosis accuracy. Therefore, this paper develops a 3D deep convolutional neural network on ^{18}F-FDG PET images for the automated early diagnosis. Furthermore, we depicted in saliency maps the decision mechanism of the deep learning method to assist the physiological interpretation of deep learning performance. The proposed method was evaluated on a dataset with 920 patients. In addition to improving the accuracy in the differential diagnosis of parkinsonism compared to state-of-the-art approaches, the deep learning methods also discovered saliency features in a number of critical regions (e.g., midbrain), which are widely accepted as characteristic pathological regions for movement disorders but were ignored in the conventional analysis of FDG PET images.**

I. INTRODUCTION

Parkinson's disease (PD) is one of the most common age-related neurodegenerative disorders, and approximately 7 to 10 million people worldwide are suffering from this disease [1]. On the other hand, very similar clinical signs can appear in patients with atypical parkinsonian syndromes, such as multiple system atrophy (MSA) and progressive supranuclear palsy (PSP) [2]. It has been reported that approximately 20-30% of patients believed to have PD turn out to have either MSA or PSP following pathological examinations [2], [3]. This misdiagnosis can lead to significant consequences for clinical patient care and research trials [4], [5]. The development of biomarkers for accurate differentiation of parkinsonian disorders is important for the determination of therapy strategies and the management of the disease.

[1]Department of Computer Science, Technische Universität München, Munich, Germany yu.zhao@tum.de
[2]PET Center, Huashan Hospital, Fudan University, Shanghai, China
[3]Dept. Neurology, Huashan Hospital, Fudan University, Shanghai, China
[4]Dept. Nuclear Medicine, Technical University of Munich, Germany
[5]Dept. Molecular and Medical Pharmacology, UCLA, Los Angeles, USA
[6]Dept. Nuclear Medicine, University of Bern, Bern, Switzerland
[7]School of Psychology and Counselling and IHBI, Queensland University of Technology, Brisbane, Australia

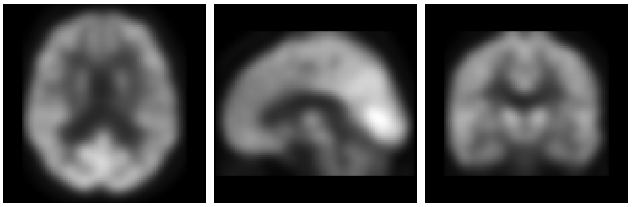
Positron emission tomography (PET) detects abnormal functional alterations of PD using specific in vivo biomarkers and has been reported of advantage in differential diagnosis far before structural damages to the brain tissue are present. ^{18}F-FDG PET visualizes the brains glucose metabolism to assess neuronal dysfunction. During the last decades, metabolic pattern analysis have been developed on ^{18}F-FDG PET for the early and accurate differential diagnosis of parkinsonism [6]. Principal component analysis (PCA) was applied to extract PD-related pattern (PDRP), MSA-related pattern (MSARP), and PSP-related pattern (PSPRP). These patterns, which were used as features for a machine learning method of logistic regression, have been found as effective surrogates to discriminate between classical PD, atypical parkinsonian syndromes and healthy control subjects [7].

Machine learning is objective and robust in digging out information, which may be filtered cognitively or disregarded in data [8]. The extraction of discriminative features is considered as one most critical factor determining the performance of machine learning [9]. The above mentioned PCA-based pattern analysis has been confirmed as effective features during the machine learning for the differentiation of parkinsonism subtypes [7]. However, the PCA decomposition takes the 3D image volume of a subject as a squeezed 1D vector and the high-level spatial interrelation is not considered during the pattern extraction. The potential of machine learning and pattern recognition is therefore not fully explored.

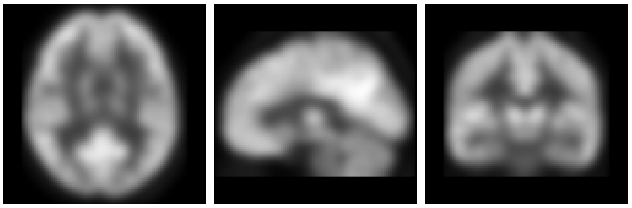
In addition to conventional handcrafted feature extraction, the emerging of deep learning moves an advanced step forward to discover new characteristic features in data automatically and effectively [10], [11]. It replicates and extends human powers of perception for information from data (e.g., images) and may surpass human performance in some situations [11], [12], [13]. Thus, it has brought record-breaking performance in many applications such as image recognition, natural language understanding, robotics, and self-driving cars [10]. It has been also applied in computerized diagnosis on medical imaging, such as differential diagnosis [14], abnormality detection [15] and prognosis [16]. Among the deep learning methods, convolutional neural network (CNN) increasingly draws attention in pattern recognition[17], [18]. It can identify automatically the discriminative features, and the learning procedure is less dependent on prior knowledge. Thus, it is powerful in discovering new invisible characteristic patterns for differentiation. However, it has not been

applied in the differential diagnosis of parkinsonism.

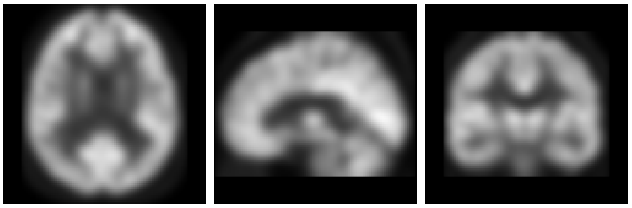
This study explores the potential of deep learning in discovering the characteristic pattern for the differential diagnosis of the parkinsonian syndrome. A 3D deep residual convolutional neural network was built for automatic classification of multiple system atrophy (MSA), progressive supranuclear palsy (PSP) and idiopathic Parkinson's disease (IPD). More importantly, we integrate saliency maps to explore the decision mechanism of the deep learning method.



(a) Multiple System Atrophy (MSA)



(b) Progressive Supranuclear Palsy (PSP)



(c) Idiopathic Parkinson's Disease (IPD)

Fig. 1. Typical examples illustrating ^{18}F-FDG PET images from patients with the multiple system atrophy (MSA), progressive supranuclear palsy (PSP) and idiopathic Parkinson's disease (IPD). The left, middle and right column demonstrate the axial, sagittal, and coronal views separately. The first row (Fig. 1 a) presents the patient with MSA. The middle row (Fig. 1 b) is a patient with PSP. The last row (Fig. 1 c) shows a patient with IPD.

II. MATERIAL AND METHOD

A. Patients, Imaging and Pre-processing

920 patients with evident parkinsonian features underwent ^{18}F-FDG PET imaging with tentative diagnosis and follow-up by movement disorders specialists, yielding the diagnosis of idiopathic Parkinson's disease (IPD, n=502) and the APS multiple system atrophy (MSA, n=239) and progressive supranuclear palsy (PSP, n=179). All patients were referred for ^{18}F fluorodeoxyglucose (FDG) positron emission tomography (PET) imaging at Shanghai Huashan Hospital. A 10-minute 3-dimensional brain emission scan was acquired at 45-minute post injection of approximately 185 MBq ^{18}F FDG (Siemens Biograph 64 HD PET/CT, Siemens, Germany). Attenuation correction was performed using low-dose CT before the emission scan. Following corrections for scatter, dead time, and random coincidences, PET images were reconstructed by using 3-dimensional filtered back projection with Gaussian Filter (FWHM 3.5mm). Fig. 1 illustrates examples of ^{18}F-FDG PET images from patients with the multiple system atrophy (MSA), progressive supranuclear palsy (PSP) and idiopathic Parkinson's disease (IPD). The clinical diagnosis was determined based on the existing clinical research criteria [19], [20].

B. Deep Neural Network

We propose a 3D residual convolutional neural network for automated differential diagnosis of the parkinsonian syndrome. As shown in Fig. 2, the network comprises a down-sampling path including three repeated encoder stacks (in the purple box). The down-sampling path aggregates increasingly abstract information as features for accurate differentiation of parkinsonian disorders. In each encoder stack, there are a residual module and a $3 \times 3 \times 3$ convolutional layers with stride 2 for down-sampling the feature maps. Each residual module includes three $3 \times 3 \times 3$ convolutional layers and two dropout layers in between. At the end of the proposed network, a global average pooling is performed, and then a fully connected layer with softmax activation is employed to map obtained features to the classification probability.

The residual connections used in the proposed network introduce connections that skip one or more layers. They are helpful for simplifying the optimization of a network. Deeper models tend to hit obstacles during the training process. The gradient signal vanishes with increasing network depth. But the residual connections propagate the gradient throughout the model, which can alleviate the vanishing gradient problem [17]. Throughout the network, we employ leaky ReLU as the activation function following the convolution layers and utilize instance normalization [21] instead of the traditional batch normalization since it is more stable than batch normalization if working on small batch size.

In this work, we choose to use the categorical cross-entropy loss, which is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\delta(y_i = c)\log(P(y_i = c)) \qquad (1)$$

where N denotes the data number and C represents the categories number. The term $\delta(y_i = c)$ is the indicator function of the i th observation belonging to the c th category. The $P(y_i = c)$ is the predicted probability by the model.

C. Saliency Maps

The saliency maps which visualize the dominant locations are employed as a computer-aided analysis tool in our study. We generated the saliency maps of testing images by using Guided Back-propagation method in [22]. It visualized the part of an input image that mostly activates a given neuron and used a simple backward pass of the activation of a single neuron after a forward pass through the network. To this end, it computed the gradient of the activation w.r.t. the image. These gradients were then used to highlight input regions that cause the most changes to the output.

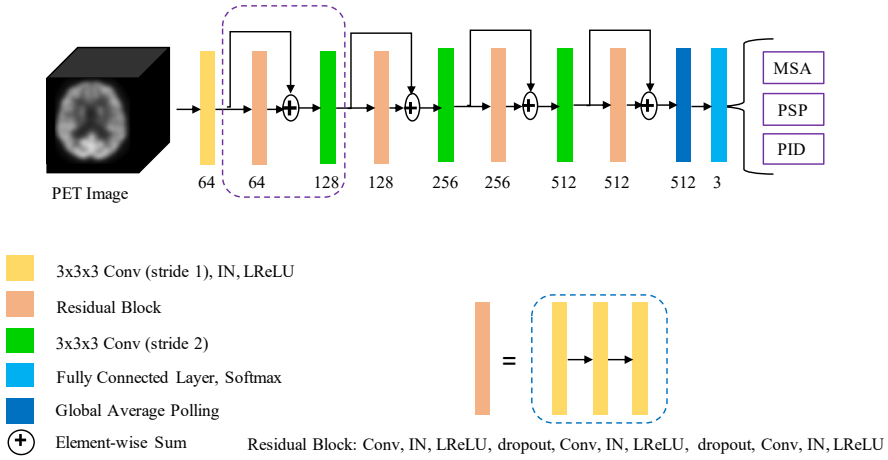


Fig. 2. The structure of the proposed network. Different operations are denoted by different colors. The digits below the operations denote the number of channels. The purple box represents a encoder stack. (Conv: Convolutional layer, IN: instance normalization, LReLU: leaky ReLU)

III. Results and Discussion

A. Experimental Setup and Performance Evaluation

The whole dataset with 920 patients can be divided into two groups. One group consists the 378 patients with tentative diagnoses (IPD: 199, MSA: 84, PSP: 95), and another group is composed with the remaining 542 patients with definite diagnoses (IPD: 303, MSA: 155, PSP: 84). We firstly pre-train the proposed network on the first group with 378 patients. Then we further train and evaluate the network based on the second group including 542 patients with definite diagnoses.

The evaluation was performed with six-fold cross-validation. We implemented the network with the Keras library [23]. ADAM optimizer was used during training with an initial learning rate $lr_{init} = 10^{-4}$. The learning rate was reduced by a factor of 2 once learning stagnates. We choose mini-batches with the size of 30. To regularize the network, we utilized the early stopping strategy with the patience of 20, which is a method employed to detect the convergence of training thereby avoiding overfitting.

We use four standard metrics to evaluate the performance of the diagnosis, i.e., sensitivity, specificity, negative predictive value (NPV) and negative predictive value (NPV), which are defined as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$NPV = \frac{TN}{TN + FN} \tag{5}$$

Where TP, FP, TN, and FN represent the True Positive, False Positive, True Negative and False Negative respectively.

B. Diagnosis Accuracy and Saliency Maps

Diagnosis results of the proposed network for three different kinds of parkinsonian disorders are illustrated in Table 1. As can be seen from this table, the proposed framework achieved 97.7% sensitivity, 94.1% specificity, 95.5% PPV and 97.0% NPV for the classification of IPD, versus 96.8%, 99.5%, 98.7%, and 98.7% for the classification of MSA, and 83.3%, 98.3%, 90.0%, and 97.8% for the classification of PSP respectively. These preliminary experimental results demonstrate the effectiveness of the proposed method on the diagnosis of the parkinsonian syndrome. When comparing the proposed method to state-of-art methods in [7], [24], we can also find that the proposed method achieves competitive and often superior accuracy than these methods. For the proposed method, the results were obtained on a bigger dataset than that in [24] and the results in [7] are obtained from a different dataset which is not openly available. Nonetheless, the current test gives a good orientation to their respective performance. The prospective randomized comparison would be preferred for further assessment of the potential of the deep learning method.

In Fig. 3, we further illustrate the calculated mean saliency maps of a group of subjects (N=15) who are with different parkinsonian disorders. Sample slices in axial view are selected to demonstrate regions driving the differential classification of MSA, PSP, and IPD for the deep convolutional neural network. As can be seen in this figure, these regions include the right prefrontal cortex (Fig. 3 (A)), the bilateral thalamus and putamen (Fig. 3 (B)), and the midbrain and left lingual gyrus (Fig. 3(C)). It is interesting to note that the deep learning methods discovered saliency features in the midbrain. This region is widely accepted as characteristic pathological region for movement disorders but was ignored in the conventional analysis of FDG PET images. In other words, the proposed network may be more sensitive to capture the different pathological abnormalities during the differential diagnosis of parkinsonian disorders.

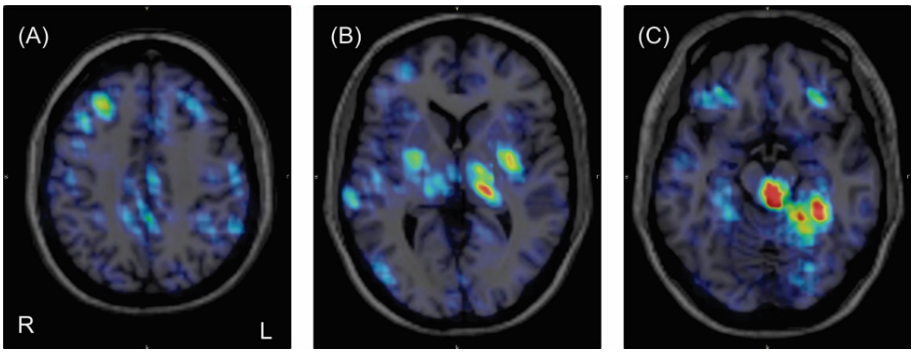


Fig. 3. Visualization of mean saliency map of 15 patients showing regions driving the differential diagnosis of IPD, MSA, and PSA for deep learning. As can be seen in this figure, these regions include the right prefrontal cortex (Fig. 3 (A)), the bilateral thalamus and putamen (Fig. 3 (B)), and the midbrain and left lingual gyrus (Fig. 3 (C)).

TABLE I

DIAGNOSIS ACCURACY OF THE PROPOSED NETWORK FOR THREE DIFFERENT KINDS OF PARKINSONIAN DISORDERS

	Idiopathic Parkinsons Disease (IPD)	Multiple System Atrophy (MSA)	Progressive Supranuclear Palsy (PSP)
Sensitivity	97.7%	96.8%	83.3%
Specificity	94.1%	99.5%	98.3%
PPV	95.5%	98.7%	90.0%
NPV	97.0%	98.7%	97.8%

IV. CONCLUSIONS

In this paper, we developed a 3D deep residual convolutional neural network for automated differential diagnosis of idiopathic Parkinson's disease and atypical parkinsonism. Experimental results based on a dataset with 920 patients have demonstrated that the proposed method can achieve excellent diagnostic accuracy. The greatest salience was in expected regions of the basal ganglia, but initial findings also implicate high order visual cortex and prefrontal cortex. The method is currently under detailed assessment in a separate group of several hundred parkinsonian patients, and with emphasis on the interpretation of the saliency maps in diagnosis.

REFERENCES

[1] M. Broadstock *et al.*, "Latest treatment options for alzheimers disease, parkinsons disease dementia and dementia with lewy bodies," *Expert opinion on pharmacotherapy*, vol. 15, no. 13, pp. 1797–1810, 2014.

[2] A. J. Hughes *et al.*, "The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service," *Brain*, vol. 125, no. 4, pp. 861–870, 2002.

[3] A. J. Hughes, Y. Ben-Shlomo, *et al.*, "What features improve the accuracy of clinical diagnosis in parkinson's disease a clinicopathologic study," *Neurology*, vol. 42, no. 6, pp. 1142–1142, 1992.

[4] K. CHAO, "Subthalamic nucleus deep brain stimulation in a patient with levodopa-responsive multiple system atrophy," *J Neurosurg*, vol. 100, pp. 553–556, 2004.

[5] V. Lambrecq *et al.*, "Deep-brain stimulation of the internal pallidum in multiple system atrophy," *Revue neurologique*, vol. 164, no. 4, pp. 398–402, 2008.

[6] M. Niethammer and D. Eidelberg, "Metabolic brain networks in translational neurology: concepts and applications," *Annals of neurology*, vol. 72, no. 5, pp. 635–647, 2012.

[7] C. C. Tang *et al.*, "Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis," *The Lancet Neurology*, vol. 9, no. 2, pp. 149–158, 2010.

[8] S. Wang and R. M. Summers, "Machine learning and radiology," *Medical image analysis*, vol. 16, no. 5, pp. 933–951, 2012.

[9] J.-Z. Cheng *et al.*, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, p. 24454, 2016.

[10] Y. LeCun *et al.*, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[11] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[12] D. Silver *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, p. 484, 2016.

[13] M. Moravčík *et al.*, "Deepstack: Expert-level artificial intelligence in heads-up no-limit poker," *Science*, vol. 356, no. 6337, pp. 508–513, 2017.

[14] H.-I. Suk *et al.*, "Latent feature representation with stacked auto-encoder for ad/mci diagnosis," *Brain Structure and Function*, vol. 220, no. 2, pp. 841–859, 2015.

[15] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.

[16] H. C. Hazlett *et al.*, "Early brain development in infants at high risk for autism spectrum disorder," *Nature*, vol. 542, no. 7641, p. 348, 2017.

[17] K. He *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] J. Long *et al.*, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[19] S. Gilman *et al.*, "Second consensus statement on the diagnosis of multiple system atrophy," *Neurology*, vol. 71, no. 9, pp. 670–676, 2008.

[20] P. G. Spetsieris *et al.*, "Highly automated computer-aided diagnosis of neurological disorders using functional brain imaging," in *Medical Imaging 2006: Image Processing*, vol. 6144. International Society for Optics and Photonics, 2006, p. 61445M.

[21] V. L. D. U. A. Vedaldi, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[22] J. T. Springenberg *et al.*, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

[23] F. Chollet *et al.*, "Keras," https://github.com/keras-team/keras, 2015.

[24] P. Wu *et al.*, "Deep learning on 18f-fdg pet imaging for differential diagnosis of parkinsonian syndromes," *Journal of Nuclear Medicine*, vol. 59, no. supplement 1, pp. 624–624, 2018.

7

Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer

This chapter has been published as **peer-reviewed journal paper:**.

R. Cao*, F. Yang*, S.-C. Ma*, L. Liu*, Y. Zhao*, Y. Li*, D.-H. Wu, T. Wang, W.-J. Lu, W.-J. Cai, H.-B. Zhu, X.-J. Guo, Y.-W. Lu, J.-J. Kuang, W.-J. Huan, W.-M. Tang, K. Huang, J. Huang, J. Yao, and Z.-Y. Dong. "Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer." In: *Theranostics* 10 (2020), pp. 11080–11091. DOI: 10.7150/thno.49864

Synopsis: This work developed a multiple-instance-learning (MIL)-based deep learning model to predict microsatellite instability in coloretal cancer from histopathology images.

Contributions of thesis author: algorithm design and implementation, computational experiments and composition of manuscript.

IVYSPRING INTERNATIONAL PUBLISHER

Theranostics

2020; 10(24): 11080-11091. doi: 10.7150/thno.49864

Research Paper

Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer

Rui Cao[1#], Fan Yang[2#], Si-Cong Ma[3#], Li Liu[1#], Yu Zhao[2,10#], Yan Li[4#], De-Hua Wu[3], Tongxin Wang[5], Wei-Jia Lu[2], Wei-Jing Cai[6], Hong-Bo Zhu[1], Xue-Jun Guo[3], Yu-Wen Lu[3], Jun-Jie Kuang[3], Wen-Jing Huan[7], Wei-Min Tang[7], Kun Huang[8,9], Junzhou Huang[2], Jianhua Yao[2]✉ and Zhong-Yi Dong[3]✉

1. Information Management and Big Data Center, Nanfang Hospital, Southern Medical University, Guangzhou, China.
2. AI Lab, Tencent, Shenzhen, China.
3. Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, China.
4. Department of Pathology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China.
5. Indiana University Bloomington, Bloomington, USA.
6. Tongshu Biotechnology Co., Ltd. Shanghai, China.
7. Tencent Healthcare, Shenzhen, China.
8. Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA.
9. Regenstrief Institute, Indianapolis, IN, USA.
10. Department of Computer Science, Technical University of Munich, Munich, Germany.

#These authors contributed equally to this study.

✉ Corresponding authors: Zhong-Yi Dong (E-mail: dongzy1317@foxmail.com) Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou, 510515, China; and Jianhua Yao (E-mail: jianhua_yao@yahoo.com) AI Lab, Tencent, Building 12A, Shengtaiyuan, Nanshan District, Shenzhen, 518057, China.

Received: 2020.06.24; Accepted: 2020.08.25; Published: 2020.09.02

Abstract

Microsatellite instability (MSI) has been approved as a pan-cancer biomarker for immune checkpoint blockade (ICB) therapy. However, current MSI identification methods are not available for all patients. We proposed an ensemble multiple instance deep learning model to predict microsatellite status based on histopathology images, and interpreted the pathomics-based model with multi-omics correlation.

Methods: Two cohorts of patients were collected, including 429 from The Cancer Genome Atlas (TCGA-COAD) and 785 from an Asian colorectal cancer (CRC) cohort (Asian-CRC). We established the pathomics model, named Ensembled Patch Likelihood Aggregation (EPLA), based on two consecutive stages: patch-level prediction and WSI-level prediction. The initial model was developed and validated in TCGA-COAD, and then generalized in Asian-CRC through transfer learning. The pathological signatures extracted from the model were analyzed with genomic and transcriptomic profiles for model interpretation.

Results: The EPLA model achieved an area-under-the-curve (AUC) of 0.8848 (95% CI: 0.8185-0.9512) in the TCGA-COAD test set and an AUC of 0.8504 (95% CI: 0.7591-0.9323) in the external validation set Asian-CRC after transfer learning. Notably, EPLA captured the relationship between pathological phenotype of poor differentiation and MSI ($P < 0.001$). Furthermore, the five pathological imaging signatures identified from the EPLA model were associated with mutation burden and DNA damage repair related genotype in the genomic profiles, and antitumor immunity activated pathway in the transcriptomic profiles.

Conclusions: Our pathomics-based deep learning model can effectively predict MSI from histopathology images and is transferable to a new patient cohort. The interpretability of our model by association with pathological, genomic and transcriptomic phenotypes lays the foundation for prospective clinical trials of the application of this artificial intelligence (AI) platform in ICB therapy.

Key words: microsatellite instability, colorectal cancer, pathomics, multi-omics, ensembled patch likelihood aggregation (EPLA)

Introduction

Microsatellite instability (MSI) is a hypermutator phenotype that occurs in tumors with DNA mismatch repair deficiency (dMMR) [1], which is reported as a hallmark of hereditary Lynch syndrome (LS)-associated cancers [2] and observed in about 15% of colorectal cancer (CRC) [3]. MSI has been identified as a favorable prognostic factor but a negative predictor for adjuvant chemotherapy in stage II CRC [4]. More importantly, recent studies have demonstrated MSI or dMMR is correlated to an increased neoantigen burden that sensitizes the tumor to immune checkpoint blockade (ICB) treatment [5]. Further investigations have suggested that the benefit of ICB treatment for patients with MSI is not limited to specific tumor types but to all solid tumors [6], which established the crucial role of MSI in predicting the efficacy of immunotherapy for advanced solid tumors, especially CRC.

MSI or dMMR testing has traditionally been performed in patients with CRC and endometrial cancer to screen for LS–associated cancer predisposition [7]. Recently, with the U.S. Food and Drug Administration (FDA) designation of MSI/dMMR as a favorable predictor of anti-programmed death-1 (PD-1) therapy [8], the clinical demand for MSI/dMMR testing has increased dramatically. However, in clinical practice, not every patient is tested for MSI, especially in those cancers with lower occurrences of MSI or in patients in developing countries, because it requires additional genetic or immunohistochemical tests which are costly and time-consuming. Additionally, various existing MSI testing methods show different sensitivities and specificities, leading to the disunity of results [9, 10]. Therefore, there are both opportunities and challenges that lie ahead in developing an MSI testing method that is available for all cancer patients.

The emergence of computational pathology have provided an opportunity for the detection of MSI because pathology slides are produced for almost every patient diagnosed with cancer; these slides can be digitized into whole slide images (WSIs) [11]. WSI not only reveals the tissue spatial arrangement of tumor cells at low magnification, but also the cell structure at high magnification [12]. Furthermore, histopathology images also show the immunologic microenvironment of tumors [13]. The cell level phenotypes presented in WSI are affected by genotypes such as MSI at the molecular scale. With the continuous penetration of artificial intelligence (AI) into the field of medical imaging, researchers have sought solutions based on deep learning, a research area in AI, in a wide range of medical problems, such as prediction of gene mutations [14] and tumor-infiltrating lymphocytes [12], and cancer screening [15, 16]. Whereas traditional machine learning depends largely on human-selected features [17], deep learning can learn features from the data, which makes it possible for researchers to discover untapped information [18, 19]. Previous studies have suggested that deep learning can discover regions that contribute to microsatellite (MS) status with special pathomorphological characteristics [20], but the applicability of the model in the Asian population remains in question because of the great variation in demographics and data preparation. The inability to interpret the extracted signatures and the predictions made by the model is considered to be one of the major issues that limit the acceptance of AI models in medicine [21].

In this study, we developed a multiple-instance-learning (MIL)-based deep learning model to predict MS status from histopathology images. The model, for which we proposed as Ensemble Patch Likelihood Aggregation (EPLA), combined both deep learning and traditional machine learning techniques. It was trained using the TCGA-COAD data set, and then transfer learning was implemented to fine tune the model using an Asian-CRC cohort curated locally, which enhanced the generalizability of this model. More importantly, we also demonstrated the interpretability of the model by identifying the crucial pathological signatures generated by the MIL model and linking them with MSI genomic and transcriptomic profiles.

Materials and Methods

Patient cohorts and dataset partition

In this study, whole slide images (WSIs) of two large cohorts were collected, and an MS label was assigned to each WSI based on the patient's microsatellite measurement. The first cohort (TCGA-COAD), retrieved from The Cancer Genome Atlas, comprised 429 frozen tissue slides diagnosed as colon adenocarcinoma (COAD) with stage I to IV. MSI score of each sample within the cohort was measured using the MSIsensor algorithm based on tumor-normal paired genome sequencing data [22]; tumors with MSIsensor scores of ≥ 10 were defined as MSI, whereas those with MSIsensor scores of < 10 were defined as microsatellite stability (MSS) [23]. In this cohort, 358 cases were labeled as MSS and 71 cases were labeled as MSI. The second cohort (Asian-CRC), collected from Tongshu Biotechnology Co., Ltd, consisted of 785 formalin-fixed paraffin-embedded (FFPE) sections diagnosed with CRC of all stages, which were provided from three medical centers in

China. Patients in the Asian-CRC group were analyzed by an MSI detection kit (Shanghai Tongshu Biotechnology Co., Ltd.) that detects five microsatellite loci (BAT-25, BAT-26, D5S346, D2S123 and D17S250) based on multiplex PCR-capillary electrophoresis [24]; tumors with instability in ≥ 2 out of five microsatellite loci were classified into the MSI-high (MSI-H) group, and the rest were assigned into the MSI-low (MSI-L)/MSS group, following the recommendations and guidelines on MSI testing for CRC [24, 25]. Thus, 164 cases were identified as MSI-H, and 621 cases were identified as MSI-L/MSS. The details of the two cohorts are summarized in Table S1. This study was approved by the Institutional Ethical Review Boards of Nanfang Hospital (NFEC-2020-055), and patient consents were obtained.

The TCGA-COAD cohort was split into separate training and test sets at a 7:3 ratio using stratified sampling, in order to maintain the same ratio of positive to negative samples in the training set and test set. The training set was used for hyper-parameter tuning based on cross-validation, whereas the test set was used for the evaluation of generalization performance, and the independent Asian-CRC cohort for external validation.

ROI delineation, tiling, and data preprocessing

All WSIs were digitalized at 20× objective lens with a predefined pixel resolution (~0.5μm/pixel). In order to reduce the influence of unrelated areas and alleviate the workload of the classification method, regions of carcinoma (ROIs) on WSIs were manually annotated by expert pathologists, according to the following rules: (1) the tumor cells should occupy more than 80% of a ROI, i.e., the interstitial component is less than 20%; and (2), obvious interfering factors, including creases, bleeding, necrosis and blurred areas, should be excluded. The annotation was performed using Aperio ImageScope (Aperio Technologies, Inc.).

Given the extremely large image size (typically 100,000 × 50,000 pixels) of a WSI, the WSIs were subsequently tiled into 512×512 patches. Only patches having a greater than 80% overlap with the carcinoma ROI were used for the following analysis. The number of patches per WSI in TCGA-COAD ranges from 22 to 2357 (average 224), whereas the Asian-CRC ranges from 5 to 3718 (average 338) (Table S1).

Data augmentation and normalization were applied for training patches, whereas only normalization was employed for test patches. Data augmentations used in our work included random horizontal flipping and random affine transformation of the patches (keeping the center invariant). Finally, the augmented patches were center cropped to 224 pixels × 224 pixels similar to Campanella's study [26], following a z-score normalization on RGB channels.

Multiple Instance Learning (MIL)-based deep learning pipeline

Our MIL-based deep learning pipeline presented two predictions: patch-level and WSI-level. Due to the large image size and heterogeneity in tumors, the WSI was first divided into small patches, and then the patch likelihoods were aggregated in an ensemble classifier to obtain the WSI-level prediction. Therefore, our method was termed Ensemble Patch Likelihood Aggregation (EPLA).

During the patch-level prediction, a residual convolutional neural network (ResNet-18) was trained to compute the patch likelihood in a MIL paradigm where the patches were assigned with the WSI's label. Binary cross-entropy (BCE) loss was utilized to optimize the network using a mini-batch gradient descent method.

We developed two independent MIL methods to aggregate the patch likelihoods: Patch Likelihood Histogram (PALHI) pipeline and Bag of Words (BoW) pipeline, which were inspired by the histogram-based method and the vocabulary-based method, respectively. In PALHI, a histogram of the occurrence of the patch likelihood was applied to represent the WSI, whereas in BoW, each patch was mapped to a TF-IDF floating-point variable, and a TF-IDF feature vector was computed to represent the WSI. Traditional machine learning classifiers were then further trained using these feature vectors to predict the MS status for each WSI. Here, Extreme Gradient Boosting (xgboost), a kind of gradient boosted decision tree, was employed in the PALHI pipeline. Naïve Bayes (NB) was used in the BoW pipeline. During the training of the WSI-level classifier, the hyperparameters were determined based on the cross-validation on the training set, using WSI-level ROCAUC as the performance metric. During WSI-level prediction, the results of PALHI and BoW classifiers were then ensembled to obtain the final prediction [27].

The initial parameters of the model were trained in the training set of TCGA-COAD, and the transfer learning technique was implemented using the Asian-CRC data to generalize the model across cohorts with a high degree of heterogeneity. The transfer learning was conducted by reusing the model weights in the patch-level discriminators and then fine-tuning the weights using a small amount of labeled Asian-CRC data. In addition, we gradually added more Asian-CRC data for model fine-tuning to explore the impact on model performance. All codes

were implemented in Python 3.6.5 and run on a workstation with Nvidia GPUs (P40). As for the minimal requirement, a desktop with CPUs and the above dependencies can run our algorithm for inference, which is widely available and easy-to-use for physicians and biologists. The average time for the completion of a single patient test is 0.5118s on a P40 workstation and 20.9291s on a regular CPU machine (i5-9500, 3.00GHz, 16GB).

Multi-omics correlation analysis of pathological signatures

Identification of pathological signatures of importance

The occurrence histograms in the PALHI and the TF-IDF feature vector in BoW were the pathological signatures generated by our model. The importance of each signature was measured by its contribution weight to the final WSI-level prediction for discovering top pathological signatures. The top pathological signatures were evaluated by Wilcoxon Rank Sum tests for significance and then sent for genomic and transcriptomic correlation analysis.

Genomic correlation analysis

The DNA mutation profile of TCGA-COAD was retrieved from cBioPortal [28]. The synonymous mutations were excluded from the following correlation analysis. For a particular gene set, as long as there was a non-synonymous mutation in any of its gene members, it would be defined as deficient.

The relationship between MSI and some mutation indexes has been reported in previous literature, including INDEL and tumor mutation burden (TMB) [29]. INDEL mutations refer to a variant type caused by sequence insertion (INS) or deletion (DEL) and can be calculated as the frequency of DEL and INS mutations. As the mutation data was profiled by the whole exome sequencing, TMB is defined and calculated as the total number of somatic nonsynonymous mutations divided by size of the exonic region of the entire genome [30]. To explore the relationship between the pathological signatures and these known genomic biomarkers, they were first normalized to a range of 0 to 1 and then visualized in a heat map using the R package *pheatmap*, during which unsupervised clustering was applied using Ward's minimum variance method.

Transcriptomic correlation analysis

The mRNA expression profile of TCGA-COAD, retrieved from cBioPortal, was normalized using the RSEM method [31]. Gene co-expression network analysis (WGCNA) is a bioinformatics method based on expression data and is typically used to identify gene modules with highly synergistic changes [32]. We first constructed a gene co-expression network for the mRNA expression profile using the R package *WGCNA*, during which the soft threshold for the network was set to the recommended value selected by the function *pickSoftThreshold* (Figure S1). Setting the minimum module size to 100 and other parameters to default, we identified 24 transcriptomic modules (Figure S2). The biological functions of the modules were annotated by the Gene Ontology (GO) over-representation test using the R package *clusterProfiler* [33], during which the Benjamini-Hochberg method was used to adjust *P* value for controlling false discover rate. Only those GO terms with adjusted *P* values lower than 0.05 were considered significantly enriched in a particular module. After that, we calculated Spearman's rank correlation coefficients for each pair of modules and pathological signatures to recognize the modules of interest.

An immune cytolytic activity (CYT) score, defined as the geometric mean of transcript levels of *GZMA* and *PRF1* [34], as well as a $CD8^+$ T-effector gene set (*CD8A*, *IFNG*, *GZMA*, *PRF1*, *CXCL9*, *CXCL10*, *TBX21*, *GZMB*) [35] was quantified from the RNA-seq data, and subsequently associated with pathological signatures to characterize the correlation with anti-tumor immunity.

Statistical analysis

The ROC curves were drawn using *pROC* and *ggplot2* in R (version 3.6.1). The area under the ROC curve and confidence intervals were calculated in *pROC*. The significance of AUC differences was tested using the Wald test statistic [36]. The optimal cutoff points of the ROC curves were estimated using the Youden Index [37]. The Wilcoxon Rank Sum test was used to compare two paired groups and visualized as a boxplot using R package *ggpubr*. Spearman's rank correlation coefficients were used for correlation analysis.

Results

Development and performance evaluation of EPLA model

The pathomics-based model named EPLA was developed in the training set of the TCGA-COAD cohort (7:3 for training and test), which consisted of two consecutive stages: patch-level prediction and WSI-level prediction (Figure 1). Briefly, a WSI was annotated to delineate the region of carcinoma (ROI). The ROI was tiled into patches, which were subsequently fed to a residual convolutional neural network (ResNet-18) to obtain the patch-level MSI prediction. Then, we trained two independent MIL

pipelines to integrate multiple patch-level predictions into an MSI score at the WSI level: the PAtch Likelihood HIstogram (PALHI) pipeline and the Bag of Words (BoW) pipeline. To obtain the optimal convex combination of the two MIL methods, we employed ensemble learning to eventually obtain the predicted MS status of the patient (Figure 1).

The performance of the EPLA model was measured in the TCGA-COAD test set. Two representative heat maps providing the patch level prediction, for an MSI case and an MSS case respectively, are shown in Figure 2A. The EPLA model achieved an AUC of 0.8848 (95% CI: 0.8185-0.9512) at the WSI level (Figure 2B) and outperformed the state-of-the-art Deep-Learning based Majority Voting method (denoted as DL-based MV) in Kather's study [20], which trained a ResNet for patch-level predictions and then took the majority of these predictions as the final MS status of the patient (Figure 2C). To directly compare our method with the DL-based MV in the same test set, we implemented DL-based MV method in the TCGA-COAD cohort, and achieved an AUC of 0.8457 (95% CI: 0.7591-0.9323) consistent with the result in Kather's study (Figure 2C).

We further compared the specificity and sensitivity of the two components of the EPLA (i.e., PALHI and BoW) to that of DL-based MV (Table S2). We found that BoW achieved higher specificity (89.5% vs 75.2%) and PALHI was superior in terms of sensitivity (86.4% vs 81.8%). The ensembled EPLA classifier combined the advantage of its two components and thus obtained both superior specificity and sensitivity compared to the DL-based MV (Table S2). Representative heat maps of the discrepant cases are shown in Figure S3. These cases were correctly predicted by EPLA but mistakenly classified by DL-based MV.

Additionally, an exploratory analysis was undertaken to identify the pathological phenotype recognized by EPLA. Of note, EPLA captured the relationship between the degree of differentiation (poor, middle or high differentiation) and MS status. Tumors with higher MSIsensor score or were predicted as MSI by EPLA model showed high proportion of poor differentiation, while lower MSIsensor score or predicted MSS tumors were demonstrated increasing proportion of high and middle differentiation ($P < 0.001$), which supports the inner relationship between EPLA model and pathological morphology (Figure 2D).

External validation of EPLA in an Asian-CRC cohort

We further measured the generalizability of our model in an Asian-CRC cohort. It was noteworthy that there existed great differences between the Asian-CRC cohort and the TCGA-COAD cohort, not only in patient race but also in the slide preparation techniques (Table S1). As a consequence, the EPLA model trained on TCGA-COAD only achieved an AUC of 0.6497 (95% CI: 0.6061-0.6933) on the external validation data set Asian-CRC (Figure 3A). Considering the wide variations in medical practice, we therefore applied transfer learning to generalize the EPLA model by fine-tuning our model using only 10% of cases from Asian-CRC, and thus achieved an AUC of 0.8504 (95% CI: 0.8158-0.885) in the remaining

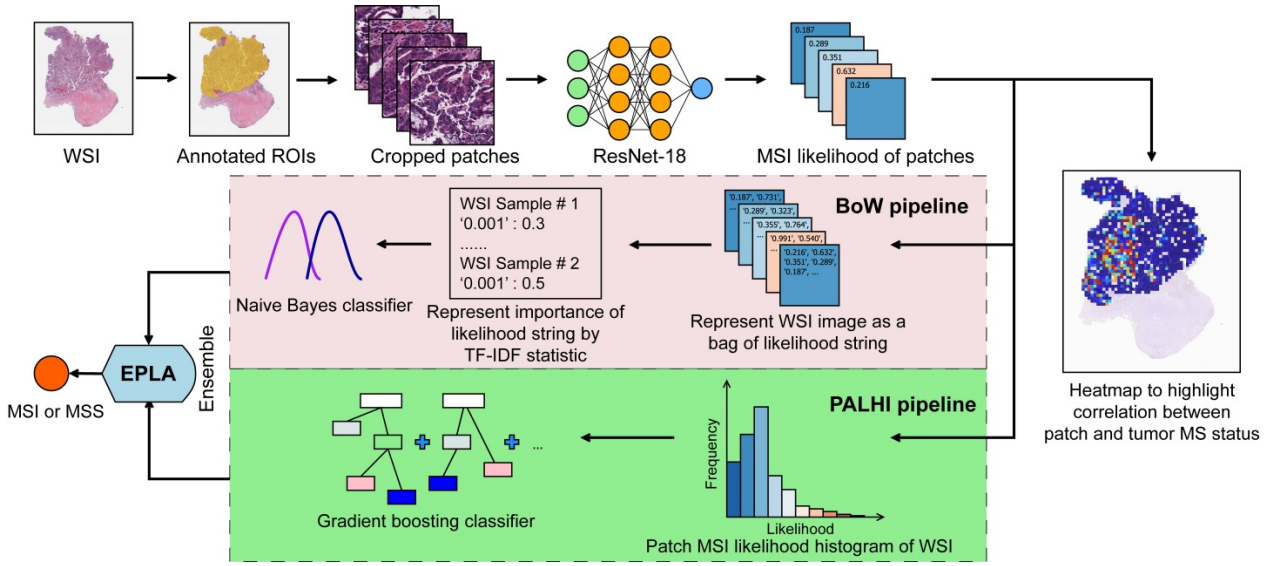
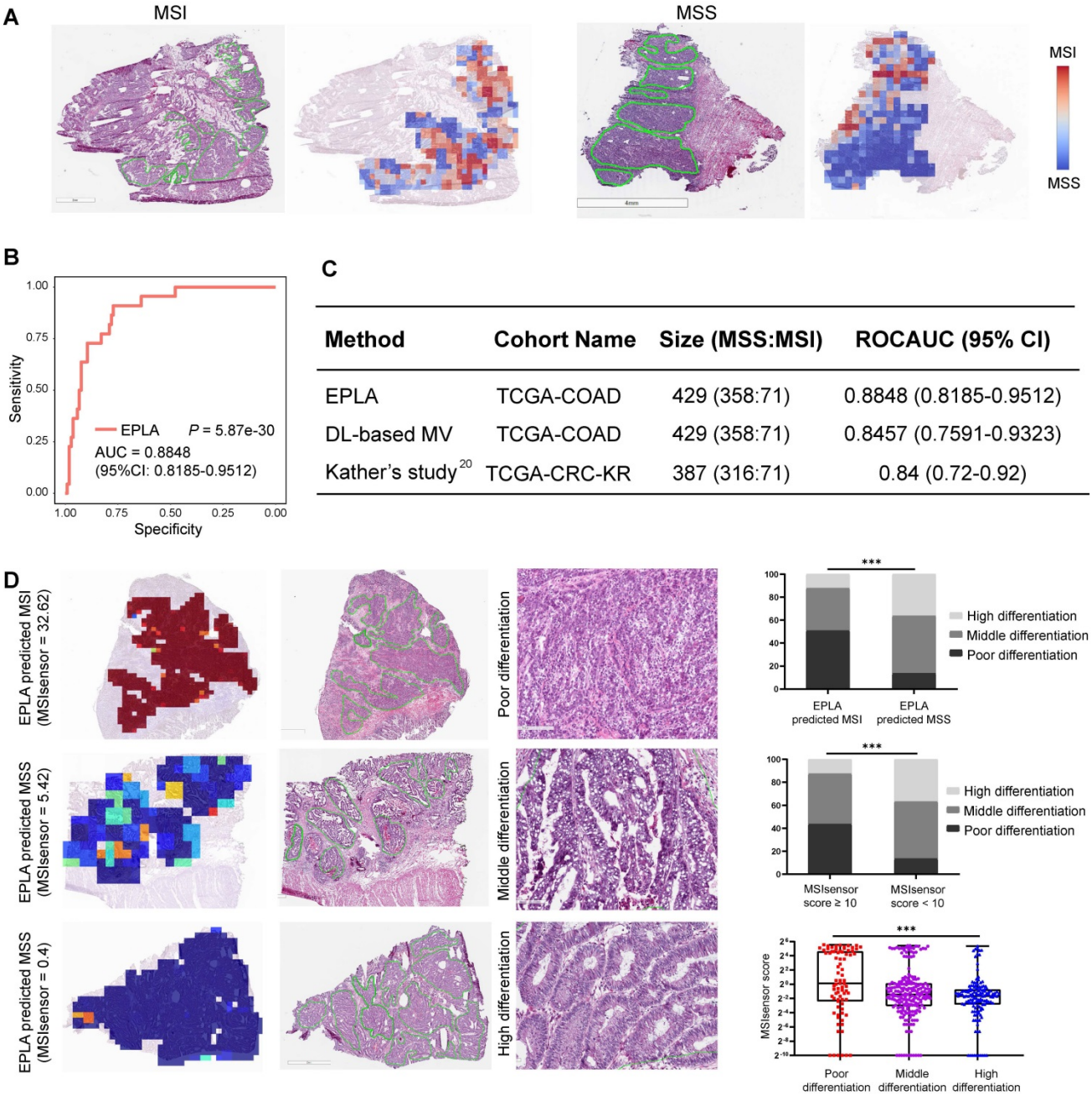


Figure 1. Overview of the Ensemble Patch Likelihood Aggregation (EPLA) model. A whole slide image (WSI) of each patient was obtained and annotated to highlight the regions of carcinoma (ROIs). Then, patches were tiled from ROIs, and the MSI likelihood of each patch was predicted by ResNet-18, during which a heat map was shown to visualize the patch-level prediction. Then, PALHI and BoW pipelines integrated the multiple patch-level MSI likelihoods into a WSI-level MSI prediction, respectively. Finally, ensemble learning combined the results of the two pipelines and made the final prediction of the MS status.

data set (Figure 3A-B). Moreover, we analyzed the performance of the EPLA model for MS status prediction across tumor stages; the EPLA model achieved high prediction performance in both non-metastatic and metastatic CRC cases, with an AUC of 0.8768 (95% CI: 0.8427-0.9110) in the stage I-III subgroup and an AUC of 0.8242 (95% CI: 0.7460-0.9023) in the stage IV subgroup, indicating the robustness of the model in predicting MS status of CRC (Figure S4).

We subsequently evaluated the amount of data needed for transfer learning by increasing the proportion of cases from Asian-CRC for model fine tuning. The performance of the fine-tuned model steadily improved, resulting in 0.8627 (95% CI: 0.8208-0.9045), 0.8967 (95% CI: 0.8596-0.9338), 0.9028 (95% CI: 0.8534-0.9522) and 0.9264 (95% CI: 0.8806-0.9722) AUCs in the ratios of 30%, 40%, 60% and 70%, respectively, implying that transfer learning was an effective measure to overcome the heterogeneity between different cohorts (Figure 3C).



Method	Cohort Name	Size (MSS:MSI)	ROCAUC (95% CI)
EPLA	TCGA-COAD	429 (358:71)	0.8848 (0.8185-0.9512)
DL-based MV	TCGA-COAD	429 (358:71)	0.8457 (0.7591-0.9323)
Kather's study[20]	TCGA-CRC-KR	387 (316:71)	0.84 (0.72-0.92)

Figure 2. Validation of the EPLA and comparison with DL-based MV in the TCGA cohort. (**A**) Representative heat maps of MSI and MSS cases at the patch-level prediction stage. Color bars show the MSI likelihood of each patch. (**B**) Receiver operating characteristic (ROC) curve of EPLA. The *P* value was calculated by the Wald test. (**C**) Summary of EPLA and DL-based MV. DL-based MV was re-implemented from a voting-based model in Ref.20. The last line of the table summarizes the performance of the original DL-based MV model. (**D**) Correlation of the degree of differentiation with EPLA-predicted MS status and MSIsensor score. DL-based MV, deep-learning based majority voting; EPLA, Ensemble Patch Likelihood Aggregation. Significance values: *** $P < 0.001$.

A

Method	Cohort Name	Size (MSS:MSI)	Transfer Learning?	ROCAUC (95% CI)
EPLA	Asian-CRC	785 (621:164)	Yes	**0.8504 (0.8158-0.8850)**
EPLA	Asian-CRC	785 (621:164)	No	0.6497 (0.6061-0.6933)

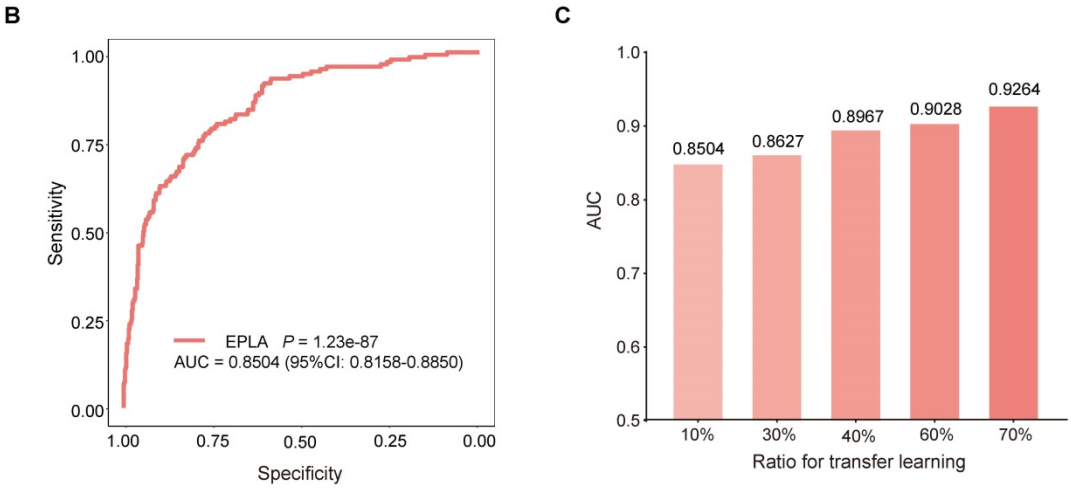


Figure 3. Generalization performance of the EPLA in an Asian cohort. (**A**) Summary of the performance of EPLA in Asian-CRC with or without transfer learning. When using transfer learning, 10% of cases from Asian-CRC were used for model fine-tuning. (**B**) The Receiver operating characteristic (ROC) curve of EPLA in the Asian-CRC after transfer learning. (**C**) ROCAUCs of the model in Asian-CRC with increasing proportions of cases for transfer learning. EPLA, Ensemble Patch Likelihood Aggregation; CRC, colorectal cancer.

Identification of top pathological signatures from the EPLA model

To gain insight into the MSI prediction mechanism of the model, we explored the contribution of the pathological signatures extracted from the EPLA model to the prediction of MSI in TCGA-COAD. The ranking of significance of the top ten pathological signatures is shown in Figure 4A. Given that the top five pathological signatures (FEA#197, FEA#198, FEA#001, FEA#188 and FEA#200) were significantly more important than the others, they were selected for subsequent analysis. Among them, FEA#001 had a significantly higher value ($P < 0.0001$) for patients in the MSS group, while the other four (FEA#188/197/198/200) had significantly higher values ($P < 0.0001$) in the MSI group (Figure 4B). Then we employed molecular-level association analysis to link the pathological signatures and the genetic alterations, which enhanced the clinical interpretation and application value of our AI method.

Association of the EPLA related pathological signatures and genomic landscape

Cluster analysis in Figure 4C shows that patients with a high value of FEA#001 were mainly MSS with normal function in DNA repair-related pathways consisting of mismatch repair (MMR), DNA damage response and repair (DDR), and homologous recombination deficiency (HRD). On the contrary, patients with high levels of FEA#188/197/198/200 were mainly due to MSI with deficient DNA repair related pathways, namely deficient-MMR (dMMR), deficient-DDR (dDDR) and deficient-HRD (dHRD). In addition, mutations of several representative genes in these pathways, including *POLE*, *BRCA1*, and *BRCA2*, also demonstrated a consistent finding. Moreover, since recent evidence suggested MSI was significantly related to TMB, especially INDEL mutation load [29], we assessed the relation between the pathological signatures and these known biomarkers and found that high TMB and INDEL mutation load were often accompanied by low FEA#001 and high FEA#188/197/198/200 (Figure 4C).

Association of the EPLA related pathological signatures and transcriptomic pathway

We applied weighted gene co-expression network analysis (WGCNA) and identified 24 modules (Figure 5A). Gene ontology (GO) enrichment analyses were performed to annotate the modules (Table S3), among which 18 modules with biological function are retained for further analyses (Figure 5A).

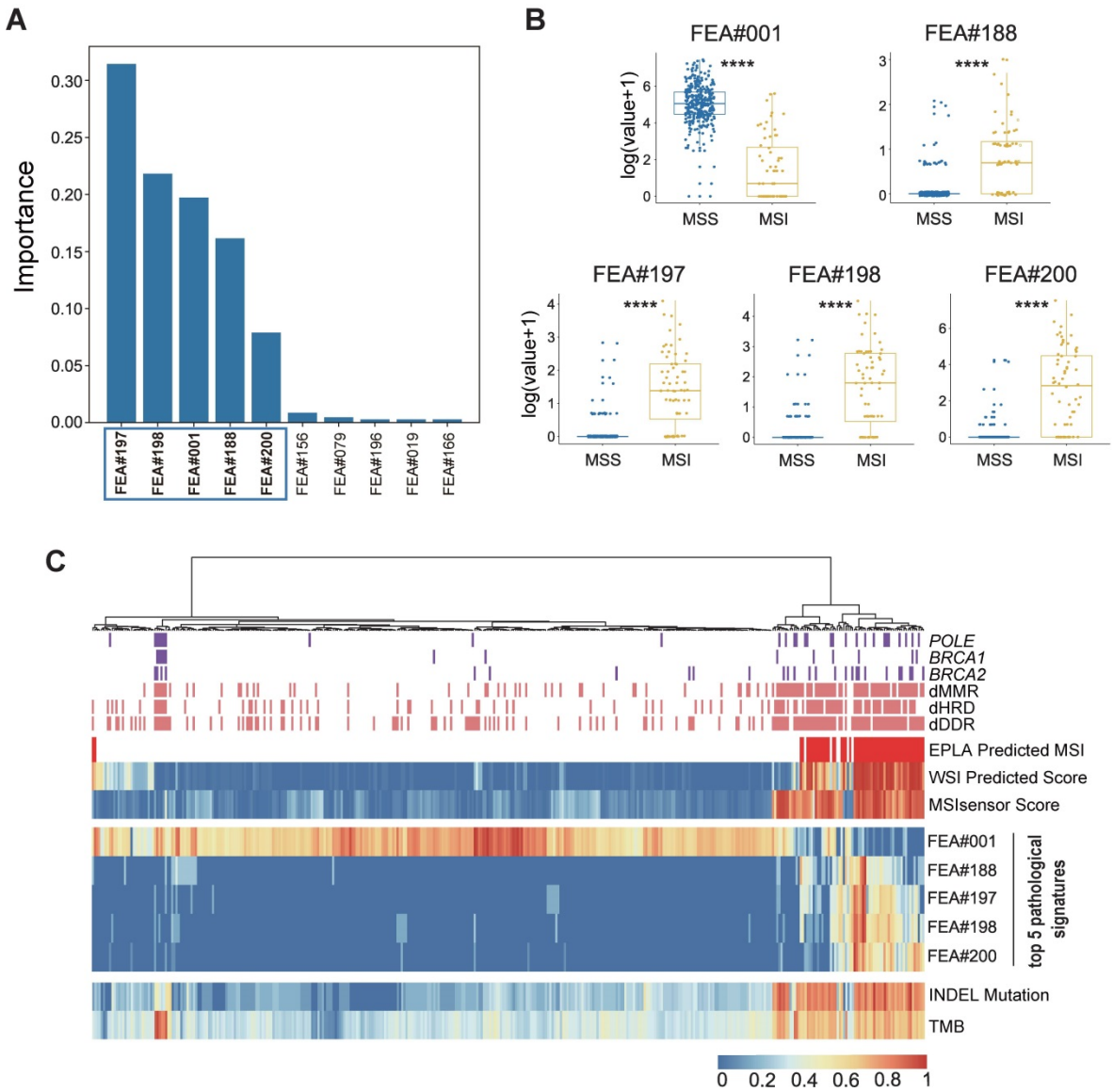


Figure 4. Identification and genomic correlation analysis of top pathological signatures. (**A**) Importance ranking of the top ten pathological signatures extracted from EPLA. (**B**) Boxplots of the five pathological signatures between MSI and MSS groups. Significance values: **** $P < 0.0001$. (**C**) Heat map with unsupervised clustering showing the correlation between genomic landscape and top pathological signatures in each patient. Each column corresponds to a patient in the TCGA-COAD cohort. All continuous variables are normalized to a range of 0 to 1. EPLA, Ensemble Patch Likelihood Aggregation; FEA, feature; INDEL: insertion-deletion, TMB: tumor mutation burden, MMR: mismatch repair, DDR: DNA damage response and repair, and HRD: homologous recombination deficiency.

Spearman's rank correlation between the 18 annotated WGCNA modules and the top five pathological signatures showed that 7 out of 18 modules are of significance, including ME12, ME8, ME21, ME14, ME13, ME18, and ME16, which were positively correlated to FEA#188/197/198/200, but negatively correlated to FEA#001 (Figure 5B). By referring to the significantly enriched GO terms of the correlated modules, we found that those molecules enriched in ME13 and ME8 were mainly related to the biological processes of immune activation, such as T cell activation and regulation of leukocyte activation (Figure 5C). As for ME12, some biological processes related to the signaling of inflammatory cytokines were significantly enriched, where the most notable was the interferon-gamma (IFN-γ) mediated pathway, namely the core IFN-γ-JAK-STAT1 signaling, which might contribute to the combination function of increased antigen processing and presentation (Figure 5C). Representative GO terms enriched in other correlated modules are shown in Figure S5.

Further investigation into the transcriptomic association of pathological signatures was conducted from the perspective of anti-tumor immunity. A strong correlation of pathological signatures with cytolytic activity (CYT) was demonstrated, which was in line with the result observed between CYT and MS status (Figure 5D). Moreover, a high degree of relevance also existed between the pathological

signatures and CD8+ T-effector genes, consistent with the finding regarding MS status (Figure 5E). Collectively, these results indicate that the pathological signatures of the model could, to some extent, reflect the anti-tumor activity of MSI, which potentiates the efficacy of immune checkpoint inhibitors [29].

Discussion

MSI testing can provide important information for clinical decision-making in a variety of cancers. However, the requirement of additional genetic or immunohistochemical tests limits its access to the general population. In this study, we developed a pathomics-based deep learning model which we term Ensemble Patch Likelihood Aggregation (EPLA) to predict MS status of CRC directly from histopathology images that are ubiquitously available in clinical practice, making it possible for every patient with a pathological diagnosis to receive an MSI evaluation. Furthermore, we proposed the use of transfer learning for model fine-tuning in a different population, improving its generalizability. We also explored the model interpretability from the perspective of genome and transcriptome association, giving a molecular biological explanation of our model.

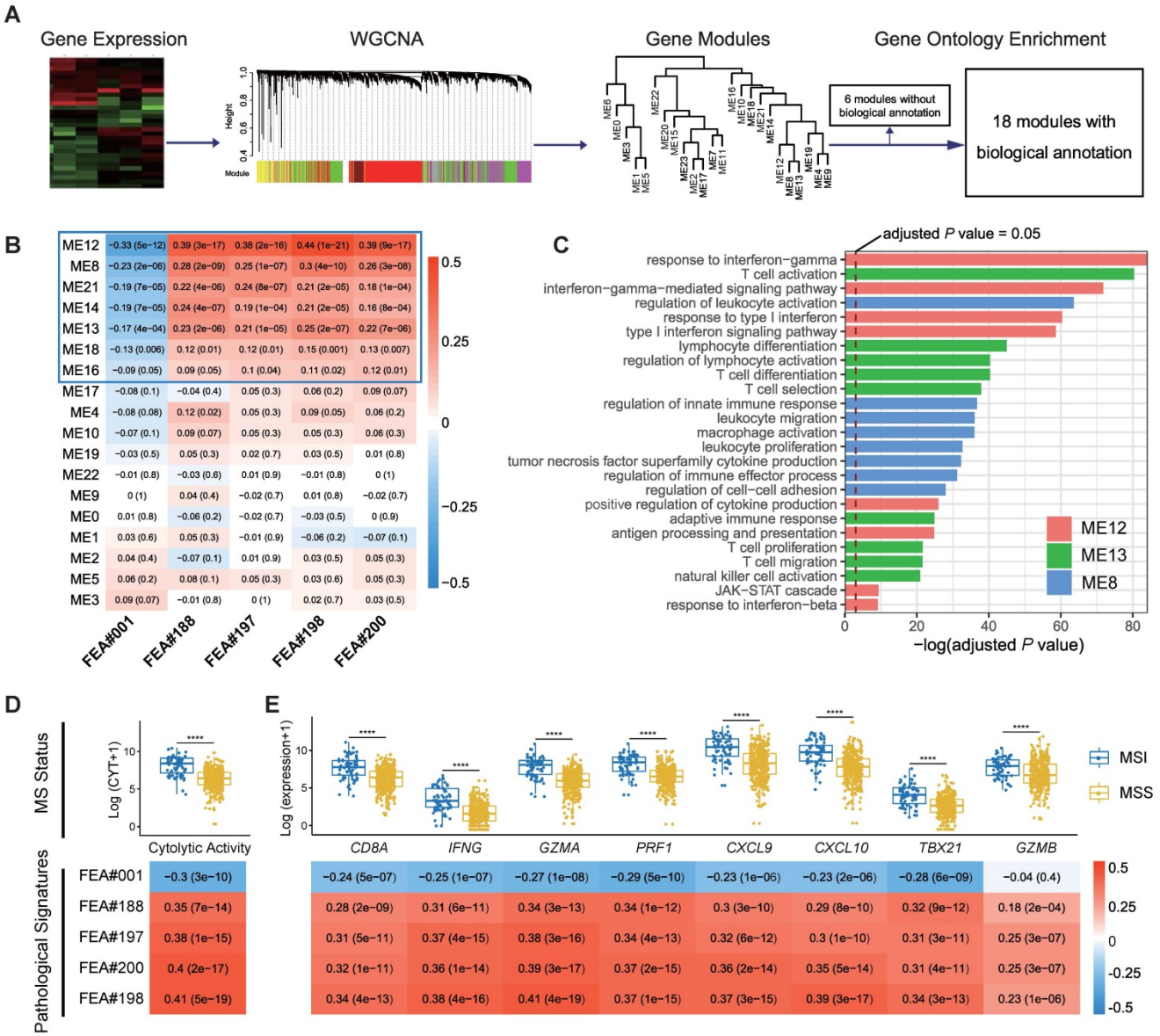


Figure 5. Correlation of top pathological signatures with WGCNA-identified modules and anti-tumor immunity. (**A**) Weighted gene co-expression network analysis (WGCNA) based on gene expression data identified gene modules with highly synergistic changes. The biological functions of these modules were annotated using Gene Ontology (GO) analyses. (**B**) Heat map of correlation coefficients (corresponding *P* values in brackets) for each pair of annotated modules and top pathological signatures. (**C**) Significantly-enriched GO terms of ME8, ME12 and ME13. The dotted line indicates the level with an adjusted *P* value of 0.05. Correlation of cytolytic activity (CYT) (**D**) and CD8+ T-effector genes (**E**) with MS status and top pathological signatures. The heat maps show Spearman's rank correlation coefficients, where a transition from red to blue represents positive to negative correlations. Significance values in boxplots: **** $P < 0.0001$.

In the development of a state-of-the-art method, Kather *et al.* proposed a deep-learning based majority voting method to predict MSI from histology under the assumption that all patches contribute equally to the prediction of MS status. Such assumptions might not be valid and could limit the prediction accuracy. In practice, although hundreds of patches are tiled from each WSI, most of them do not contribute much to the final prediction. In contrast, only a few key patches make the majority contribution. Our model based on multiple instance deep learning has the ability to automatically adjust the contribution of each patch to the overall WSI-level prediction in a learnable way by giving key patches higher weights, resulting in higher performances over the DL-based MV method in terms of AUC, sensitivity, and specificity. The superiority of multiple instance deep learning over DL-based MV method was also confirmed in a cohort of stomach adenocarcinoma collected from TCGA, implying the feasibility of EPLA in predicting microsatellite status across tumor types (Figure S6). Moreover, the influence of the magnification on the performance of the EPLA model was analyzed in the TCGA-COAD cohort. Notably, there was a performance degradation of our model using 5× magnification or 10× magnification, indicating that WSIs at 20× magnification better preserved the information of the microenvironment in tumors (Table S4). Therefore, we recommend this model being applied on WSIs at 20× magnification, which is also the commonly used magnification at clinical practice.

In clinical practice, different data sets could be vastly different due to the disparities between patient populations and data acquisition processes, resulting in a large performance gap for AI algorithms [38]. For example, in Kather's study, an MSI classifier trained on TCGA, which was mainly made up of Western populations, and only achieved an AUC less than 0.70 in the KCCH cohort, a Japanese cohort [20]. Furthermore, the histology slides in TCGA-COAD were flash-frozen slides that utilize water crystallization during the freezing process, often resulting in an altered appearance of the tissue structure as compared to the FFPE slides used in Asian-CRC which provided more tissue structure clarity. This data difference could not be effectively eliminated by only color normalization (data not shown), indicating that more advanced techniques, such as transfer learning, are necessary. As expected, EPLA showed performance degradation in Asian-CRC by simply applying the model trained on TCGA-COAD, but the results improved significantly after transfer learning. It is of clinical significance that using only a minority of the new domain data for model fine-tuning can already improve the AUC to a satisfactory level and further improvement can be expected if even more data are included, which proves that our model can be easily generalized to the complicated clinical environment, regardless of race, preparation techniques, and data acquisition techniques.

Deep learning models are often criticized for their poor interpretability, especially in mission-critical applications, such as healthcare [38]. Only those models with certain interpretability can be understood, verified, and trusted by clinicians in clinical practice [21]. To solve this problem, the pathological signatures, defining stable or unstable of a cancer specimen, were built during the training of the MIL model. We visualized the patches corresponding to these signatures and connected them into contours, which in turn guided us to discover the morphological features that are critical for MS status. In this way, the correlation between morphological features and the predicted MS status was investigated, through which we found that EPLA captured the information of poor differentiation in MSI tumors, in accordance with the previous finding [39]. More importantly, we proposed a comprehensive molecular-level analysis including genomic and transcriptomic association analysis with pathological signatures found by AI for clinical interpretation, which could also be easily applied on other gene mutation prediction tasks. In terms of our task, the genomic association between DNA repair pathways and MSI cancers, which is exquisitely sensitive to ICB, has been verified previously [40]. Moreover, MSI and its resultant TMB have been reported to underlie the response to PD-1 blockade immunotherapy [41, 42], and the INDEL mutation load is particularly associated with the extent of the response [29]. Inspired by these discoveries, we confirmed the strong correlation between the pathological signatures identified by the model and these genomic biomarkers of MSI (Figure 4C and Figure S7). Despite advances in understanding of MSI at the genomic level, the process and mechanisms at the transcriptomic level remain relatively understudied. Researches on the anti-tumor effects of MSI have suggested an increased activation and infiltration of immune cells, together with an enhanced cytolytic activity as well as an up-regulation of $CD8^+$ T-effector genes [29, 43]. Remarkably, by analyzing the WGCNA-identified modules, the pathological signatures were demonstrated with high relevance to the expression level of IFN-γ-JAK-STAT1 signaling pathway, whose pivotal role in immune activation and response to immunotherapy has been supported by extensive evidence [44]. Although IFN-γ

at the same time induces feedback of up-regulation of PD-L1 on both tumor and immune cells, anti-PD-L1 therapy pertinently blocks the suppressive mechanisms, and thus inclines the balance of immune microenvironment to the inflamed phenotype [45]. Furthermore, we provided evidence of a tight connection of the pathological signatures with anti-tumor activity from the perspective of transcriptomic profiles, consistent with the relationship between MS status and immunity.

The nature of AI models has limitations in our model. The performance of deep learning models largely depends on the size and quality of the training set. We still need to expand the training data to improve the accuracy and generalizability of the model. Although the model has been verified in TCGA and an Asian cohort respectively, a large prospective clinical trial is necessary before we can deploy it as a routine MSI testing method in clinical practice.

Conclusions

In this study, we developed a pathomics-based model for MSI prediction directly from pathological images without the need for genetic or immuno-histochemical tests. Using these images allows the evaluation of MS status in many more patients than was previously possible. Through the model, we identified five pathohistological imaging signatures to predict the MS status. The reliability of the model was verified in two independent cohorts and the interpretability of the model was illustrated by exploring the correlation between the pathological signatures and multi-omics characterizations. Ongoing work is attempting to further validate our model in large-cohort, prospective clinical trials.

Abbreviations

MSI: microsatellite instability; MSS: microsatellite stability; MS: microsatellite; CRC: colorectal cancer; MIL: multiple instance learning; EPLA: ensembled patch likelihood aggregation; MMR: mismatch repair; WSI: whole slide image; AI: artificial intelligence; ROI: regions of carcinoma; BCE: binary cross-entropy; PALHI: patch likelihood histogram; BoW: bag of words; NB: naïve bayes; TMB: tumor mutation burden; GO: gene ontology; CYT: cytolytic activity; DL-based MV: deep-learning based majority voting; ROC: receiver operating characteristic; AUC: area under the curve; DDR: DNA damage response and repair; HRD: homologous recombination deficiency; FEA: feature; WGCNA: weighted gene co-expression network analysis.

Supplementary Material

Supplementary figures and tables.
http://www.thno.org/v10p11080s1.pdf
Supplementary table S3.
http://www.thno.org/v10p11080s2.xlsx

Acknowledgments

This study was supported by the National Natural Science Foundation for Young Scientists of China (No. 81802863), the Natural Science Foundation of Guangdong Province (No. 2018030310285), the Outstanding Youths Development Scheme of Nanfang Hospital, Southern Medical University (No. 2017J003), the Key Area Research and Development Program of Guangdong Province, China (No. 2018B010111001) and Science and Technology Program of Shenzhen, China (No. ZDSYS201802021814180). We would like to thank the Shanghai Tongshu Biotechnology Co., Ltd. for MSI detection and we also thank Dr. Yu-Fa Li in the department of pathology and laboratory medicine, Guangdong General Hospital for helping with pathological phenotype analysis of colorectal cancer.

Author contributions

R.C., F.Y. and S.C.M. performed the experiments; Z.Y.D., J.H.Y. and L.L. designed the experiments; Y.L., H.B.Z., Y.W.L., and J.J.K. delineated the ROI of TCGA and Asian-CRC cohorts; Y.Z., W.J.L., T.X.W. and W.M.T tiled the WSIs and preprocessed the patches; Y.L. and W.J.C. collected external validation histopathology images and helped identify samples validated by MSI detection; R.C., F.Y., W.J.H. and S.C.M. contributed to the analysis of the data; Z.Y.D., J.H.Y. and L.L. conceived and directed the project; Z.Y.D., S.C.M., J.H.Y., L.L., Y.Z. and F.Y. wrote the manuscript with the assistance and feedback of all the other co-authors.

Data availability

The source codes are available at https://github.com/yfzon/EPLA. Images were downloaded from the open TCGA database. MSIsensor score was downloaded from the published article (doi: 10.1016/j.cell.2018.03.033). The DNA mutation profile and the mRNA expression profile of TCGA-COAD were retrieved from cBioPortal.

Ethics approval and consent to participate

In order to use these histopathology images, patient consents are obtained, and this study is approved by the Institutional Ethical Review Boards of Nanfang Hospital (NFEC-2020-055).

Competing Interests

F.Y., Y.Z., W.J.L., T.X.W., W.J.H., W.M.T and J.H.Y. are employed by Tencent and W.J.C. is employed by Shanghai Tongshu Biotechnology Co., Ltd.

References

1. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology. 2010; 138: 2073-87.e3.
2. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med. 2003; 348: 919-32.
3. Vilar E, Gruber SB. Microsatellite instability in colorectal cancer-the stable evidence. Nat Rev Clin Oncol. 2010; 7: 153-62.
4. Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, *et al.* Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. J Clin Oncol. 2010; 28: 3219-26.
5. Germano G, Lamba S, Rospo G, Barault L, Magrì A, Maione F, *et al.* Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. Nature. 2017; 552: 116-20.
6. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017; 357: 409-13.
7. Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, *et al.* Feasibility of screening for Lynch syndrome among patients with colorectal cancer. J Clin Oncol. 2008; 26: 5783-8.
8. Prasad V, Kaestner V, Mailankody S. Cancer Drugs Approved Based on Biomarkers and Not Tumor Type-FDA Approval of Pembrolizumab for Mismatch Repair-Deficient Solid Cancers. JAMA Oncol. 2018; 4: 157-8.
9. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, *et al.* ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. Ann Oncol. 2019; 30: 1232-43.
10. Lindor NM, Burgart LJ, Leontovich O, Goldberg RM, Cunningham JM, Sargent DJ, *et al.* Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. J Clin Oncol. 2002; 20: 1043-8.
11. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol. 2019; 20: e253-e61.
12. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, *et al.* Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. Cell Rep. 2018; 23: 181-93.e7.
13. Bhargava R, Madabhushi A. Emerging Themes in Image Informatics and Molecular Analysis for Digital Pathology. Annu Rev Biomed Eng. 2016; 18: 387-412.
14. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018; 24: 1559-67.
15. Zeng Y, Xu S, Chapman WC, Jr., Li S, Alipour Z, Abdelal H, *et al.* Real-time colorectal cancer diagnosis using PR-OCT with deep learning. Theranostics. 2020; 10: 2587-96.
16. Pathania D, Landeros C, Rohrer L, D'Agostino V, Hong S, Degani I, *et al.* Point-of-care cervical cancer screening using deep learning-based microholography. Theranostics. 2019; 9: 8438-47.
17. Cai H, Peng Y, Ou C, Chen M, Li L. Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: a machine learning approach. PLoS One. 2014; 9: e87387.
18. Yamamoto Y, Tsuzuki T, Akatsuka J, Ueki M, Morikawa H, Numata Y, *et al.* Automated acquisition of explainable knowledge from unannotated histopathology images. Nat Commun. 2019; 10: 5642.
19. Wang J, Yang X, Cai H, Tan W, Jin C, Li L. Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning. Sci Rep. 2016; 6: 27327.
20. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, *et al.* Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. 2019; 25: 1054-6.
21. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, *et al.* The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. Theranostics. 2019; 9: 1303-22.
22. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics. 2014; 30: 1015-6.
23. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F, *et al.* Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. Cancer Cell. 2018; 33: 125-36 e3.
24. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, *et al.* ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. Ann Oncol. 2019; 30: 1232-43.
25. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, *et al.* Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst. 2004; 96: 261-8.
26. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med. 2019; 25: 1301-9.
27. Cheng J, Aurélien B, Mark van der L. The relative performance of ensemble methods with deep convolutional neural networks for image classification. J Appl Stat. 2018; 45: 2800-18.
28. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6: pl1.
29. Mandal R, Samstein RM, Lee K-W, Havel JJ, Wang H, Krishna C, *et al.* Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. Science. 2019; 364: 485-91.
30. Meléndez B, Van Campenhout C, Rorive S, Remmelink M, Salmon I, D'Haene N. Methods of measurement for tumor mutational burden in tumor tissue. Transl Lung Cancer Res. 2018; 7: 661-7.
31. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12: 323.
32. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9: 559.
33. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16: 284-7.
34. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015; 160: 48-61.
35. Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, *et al.* Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet. 2017; 389: 67-76.
36. Vickers AJ, Cronin AM, Begg CB. One statistical test is sufficient for assessing new predictive markers. BMC Med Res Methodol. 2011; 11: 13.
37. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J. 2005; 47: 458-72.
38. Xu J, Xue K, Zhang K. Current status and future trends of clinical diagnoses via image-based deep learning. Theranostics. 2019; 9: 7556-65.
39. Jenkins MA, Hayashi S, O'Shea AM, Burgart LJ, Smyrk TC, Shimizu D, *et al.* Pathology features in Bethesda guidelines predict colorectal cancer microsatellite instability: a population-based study. Gastroenterology. 2007; 133: 48-56.
40. Cortes-Ciriano I, Lee S, Park W-Y, Kim T-M, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun. 2017; 8.
41. Liu L, Bai X, Wang J, Tang X-R, Wu D-H, Du S-S, *et al.* Combination of TMB and CNA Stratifies Prognostic and Predictive Responses to Immunotherapy Across Metastatic Cancer. Clin Cancer Res. 2019; 25: 7413-23.
42. Dong Z-Y, Zhong W-Z, Zhang X-C, Su J, Xie Z, Liu S-Y, *et al.* Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma. Clin Cancer Res. 2017; 23: 3012-24.
43. Kikuchi T, Mimura K, Okayama H, Nakayama Y, Saito K, Yamada L, *et al.* A subset of patients with MSS/MSI-low-colorectal cancer showed increased CD8(+) TILs together with up-regulated IFN-γ. Oncol Lett. 2019; 18: 5977-85.
44. Nirschl CJ, Suárez-Fariñas M, Izar B, Prakadan S, Dannenfelser R, Tirosh I, *et al.* IFNγ-Dependent Tissue-Immune Homeostasis Is Co-opted in the Tumor Microenvironment. Cell. 2017; 170: 127-41.e15.
45. Ivashkiv LB. IFNγ: signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. Nat Rev Immunol. 2018; 18: 545-58.

8

Concluding Remarks

The medical image data is increasing at a pace gradually exceeding the addressing capability of clinical professionals, due to the limitations in training and experience, and factors of fatigue. In the future precision medicine, each diagnosis should be personalized based on all known information of the patient, combined with collective experience from billions of patients, which is also beyond the ability of medical experts. With the progress of artificial intelligence techniques and increasing computing power, computer-aided analysis of medical image data has shown its potential to assist clinical providers in diagnosis, treatment planning, and prognosis, etc. In this work, we focus on developing novel deep-learning-based methods for solving challenging medical image analysis tasks including medical image segmentation and classification. Due to the publication-based nature of this thesis, the Chapter 3 to 7 are self-contained and in their original form. This final chapter, therefore, provides a summary as well as a more general discussion of the work, including directions for future research.

8.1 Conclusion

In Chapter 3, we proposed a *knowledge-aided* convolutional neural network (KaCNN) for small organ segmentation with limited training data. we developed a novel knowledge-aided convolutional network architecture to combine the assets of deep-learning based methods and traditional methods for small organ segmentation and it outperformed either of those methods taken individually. We applied cascaded localization and segmentation steps which took advantage of aligning training and test images to be similar to each other by registrations, focusing the field of view on the region of interest, avoiding over-segmentation. Experimental results on the ISBI 2015 VISCERAL challenge dataset demonstrated that the proposed method outperforms deep-learning based U-Net and traditional methods in the segmentation of small organs on the same dataset.

Additionally, in Chapter 4, we leveraged the deep learning method for the detection of prostate cancer lesions in ^{68}Ga-PSMA-11 PET/CT scans. An end-to-end 3D deep supervised residual convolutional neural network was proposed, which encoded richer spatial information and extracted more discriminative representations via the hierarchical architecture trained with 3D samples. ^{68}Ga-PSMA-11 PET/CT scans from three different centers were used to train and evaluate the proposed networks. The preliminary test confined to the pelvic area confirmed the potential of deep learning methods for this task, setting the stage for a more extensive data collection and annotation. Increasing the amount of training data should further enhance the performance of the developed deep learning methods, especially in light of the requirement for whole-body assessments

Moreover, in Chapter 5, we dealt with a challenging clinical task of automatic prediction of lymph node metastasis using histopathological images of colorectal cancer. A deep GCN-based MIL method combined with a feature selection strategy was proposed. Experimental results demonstrated that our approach outperformed other state-of-the-art methods and it benefited from the designed components, including (1) VAE-GAN for instance-level feature extraction, (2) instance-level feature selection and (3) GCN-based MIL for bag representation and bag-level classification.

In Chapter 6, we aim at developing an automated method for the differential diagnosis of idiopathic Parkinson's disease and atypical parkinsonism. A 3D deep residual convolutional neural network was applied. A dataset including 920 patients was collected to develop and evaluate the proposed method. Experimental results demonstrated that the proposed method achieved excellent diagnostic accuracy. The greatest salience was in expected regions of the basal ganglia, but initial findings also implicated high order visual cortex and prefrontal cortex.

Finally, in Chapter 7, we developed a multiple instance learning (MIL) based deep learning model to predict MSI status directly from histopathology images and further designed an interpretable analysis pipeline to link the image phenotypes to genotypes. We first trained and validated our model on 429 patients from The Cancer Genome Atlas (TCGA-COAD), achieving an AUCROC which outperforms the state-of-the-art approach. We also collected 785 patients from an Asian population (Asian-CRC) as an external validation cohort, on which we proposed a transfer learning technique to generalize our model from TCGA-COAD, considering the great heterogeneity across data sets in clinical practice. We found that using only a minority of the new domain data for model fine-tuning can already improve the prediction performance to a satisfactory level, indicating the universality of our identified pathological signatures between the two different patient populations. Furthermore, we performed a comprehensive analysis to associate the identified pathological signatures with the

genomic and transcriptomic profiles for clinical interpretation. Molecular association analysis revealed that our model recognized pathological signatures related to mutation burden, DNA repair pathways and immunity, which helps the pathologists and oncologists to understand the discovery and make the decision on immunotherapy.

8.2 Outlook

8.2.1 Interpretability

Currently, increasing concern regarding the interpretability of deep neural networks has been raised. A clinical interpretable model, which can reveal the working mechanism and improve the credibility of the model, is especially demanded in the clinical community. Saliency-map methods are an increasingly used tool for the interpretation of deep neural networks which highlight input features relevant to the result of a learned model. However, saliency-map based methods are still under development and many existing approaches are lack of locational accuracy for the saliency regions. Moreover, several existing saliency methods is independent both of the model and of the data generating process. Consequently, they have very limited ability to explain the model [137, 138, 139]. Nowadays, authors in [140] proposed a full-gradient saliency map method considering both the input importance (indicating the contribution of individual input voxels) and neuron importance (reflecting the contribution of groups of voxels with specific structural information), which is sharper and more tightly confined to object regions compared to other existing methods. However, the current interpretation methods cannot meet the requirements of the clinical community and are an important area worthy of further researches.

8.2.2 Weakly- and Semi-Supervised Learning

Machine learning methods are data-driven, data plays a critical role in the machine learning algorithms, especially for the deep neural network. However, it is hard to collect sufficient huge medical datasets due to the restriction of sharing and protection of patient privacy. Moreover, the annotation medical data is time-consuming, tedious, and requires expert domain knowledge. Weakly- and semi-supervised learning is a promising method to alleviate the impact of data limitation. Authors in [141] summarized the current Weakly- and semi-supervised learning methods. Although the semi-supervised and weak-supervised learning methods have made considerable progress, the experimental results show that the current methods still have shortcom-

ings, and many aspects need further research. Besides, self-supervised learning [142] is another promising way to reduce the requirement of annotated data.

8.2.3 Multi-modality Learning

In clinical practice, there exist medical data of multiple modalities including radiology images such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), histology slides, blood biomarker, Genetic information, demographic as well as clinical features, etc. Researches [143] have demonstrated that employing multi-modality data usually outperforms using only one individual modality. Multi-modality learning is gaining increasing attention. Future directions to improve multi-modality performance include how to exploit the latent relationship between different modalities and how to design multimodal networks to efficiently integrate the multi-information.

Appendices

A

List of Publications

The following publications were written *during this thesis*.

Peer-reviewed Journal Articles

- **Y. Zhao**, H. Li, S. Wan, A. Sekuboyina, X. Hu, G. Tetteh, M. Piraud, and B. Menze. "Knowledge-aided convolutional neural network for small organ segmentation." In: *IEEE journal of biomedical and health informatics* 23.4 (2019), pp. 1363–1373. DOI: 10.1109/JBHI.2019.2891526.
- R. Cao*, F. Yang*, S.-C. Ma*, L. Liu*, Y. Zhao*, Y. Li*, D.-H. Wu, T. Wang, W.-J. Lu, W.-J. Cai, H.-B. Zhu, X.-J. Guo, Y.-W. Lu, J.-J. Kuang, W.-J. Huan, W.-M. Tang, K. Huang, J. Huang, J. Yao, and Z.-Y. Dong. "Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer." In: *Theranostics* 10 (2020), pp. 11080–11091. DOI: 10.7150/thno.49864.
- X. Hu, R. Guo, J. Chen, H. Li, D. Waldmannstetter, Y. Zhao†, B. Li, K. Shi, and B. Menze. "Coarse-to-Fine Adversarial Networks and Zone-based Uncertainty Analysis for NK/T-cell Lymphoma Segmentation in CT/PET images." In: *IEEE Journal of Biomedical and Health Informatics* (2020). DOI: 10.1109/JBHI.2020.2972694.
- L. Xu, G. Tetteh, J. Lipkova, **Y. Zhao**, H. Li, P. Christ, M. Piraud, A. Buck, K. Shi, and B. H. Menze. "Automated whole-body bone lesion detection for multiple myeloma on ^{68}Ga-Pentixafor PET/CT imaging using deep learning methods." In: *Contrast media & molecular imaging* 2018 (2018). DOI: https://doi.org/10.1155/2018/2391925.

Peer-reviewed Conference Proceedings

- **Y. Zhao**, F. Yang, Y. Fang, H. Liu, N. Zhou, J. Zhang, J. Sun, S. Yang, B. Menze, X. Fan, et al. "Predicting Lymph Node Metastasis Using Histopathological Images Based on Multiple Instance Learning With Deep Graph Convolution." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 4837–4846.
- **Y. Zhao**, A. Gafita, G. Tetteh, F. Haupt, A. Afshar-Oromieh, B. Menze, M. Eiber, A. Rominger, and K. Shi. "Deep Neural Network for Automatic Characterization of Lesions on ^{68}Ga-PSMA PET/CT Images." In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 951–954. DOI: 10.1109/EMBC.2019.8857955.
- **Y. Zhao**, P. Wu, J. Wang, H. Li, N. Navab, I. Yakushev, W. Weber, M. Schwaiger, S.-C. Huang, P. Cumming, et al. "A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes on ^{18}F-FDG PET Images." In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 3531–3534. DOI: 10.1109/EMBC.2019.8856747.

Peer-reviewed Workshop Proceedings

- X. Hu, H. Li, **Y. Zhao**, C. Dong, B. H. Menze, and M. Piraud. "Hierarchical multi-class segmentation of glioma images using networks with multi-level activation function." In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 116–127.
- L. Xu, G. Tetteh, M. Mustafa, J. Lipkova, **Y. Zhao**, M. Bieth, P. Christ, M. Piraud, B. Menze, and K. Shi. "W-Net for Whole-Body Bone Lesion Detection on ^{68}Ga-Pentixafor PET/CT Imaging of Multiple Myeloma Patients." In: *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*. Springer, 2017, pp. 23–30.

Peer-reviewed Conference Abstract Proceedings

- G. Tetteh, A. Gafita, L. Xu, **Y. Zhao**, C. Dong, A. O. Rominger, K. Shi, C. Zimmer, B. Menze, and M. Eiber. "Fully Convolutional Neural Network

to Assess Skeleton Tumor Burden in Prostate Cancer Using ^{68}Ga-PSMA-11 PET/CT: Preliminary Results." In: *European journal of nuclear medicine and molecular imaging* 45.S1 (2018), S41–S41.

Bibliography

[1] A. Rajkomar, J. Dean, and I. Kohane. "Machine learning in medicine." In: *New England Journal of Medicine* 380.14 (2019), pp. 1347–1358.

[2] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*. Vol. 1. MIT press Cambridge, 2016.

[3] D. Shen, G. Wu, and H.-I. Suk. "Deep learning in medical image analysis." In: *Annual review of biomedical engineering* 19 (2017), pp. 221–248.

[4] A. Maier, C. Syben, T. Lasser, and C. Riess. "A gentle introduction to deep learning in medical image processing." In: *Zeitschrift für Medizinische Physik* 29.2 (2019), pp. 86–101.

[5] A. S. Lundervold and A. Lundervold. "An overview of deep learning in medical imaging focusing on MRI." In: *Zeitschrift für Medizinische Physik* 29.2 (2019), pp. 102–127.

[6] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás. "3D Deep Learning on Medical Images: A Review." In: *arXiv preprint arXiv:2004.00218* (2020).

[7] J. Ker, L. Wang, J. Rao, and T. Lim. "Deep learning applications in medical image analysis." In: *Ieee Access* 6 (2017), pp. 9375–9389.

[8] N. Dimitriou, O. Arandjelović, and P. D. Caie. "Deep learning for whole slide image analysis: An overview." In: *Frontiers in Medicine* 6 (2019).

[9] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. "Dermatologist-level classification of skin cancer with deep neural networks." In: *nature* 542.7639 (2017), pp. 115–118.

[10] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng. "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network." In: *Nature medicine* 25.1 (2019), p. 65.

[11] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. C. Corrado, A. Darzi, et al. "International evaluation of an AI system for breast cancer screening." In: *Nature* 577.7788 (2020), pp. 89–94.

[12] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, et al. "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography." In: *Nature medicine* 25.6 (2019), pp. 954–961.

[13] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. W. K. Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs. "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images." In: *Nature medicine* 25.8 (2019), pp. 1301–1309.

[14] L. Papp, C. P. Spielvogel, I. Rausch, M. Hacker, and T. Beyer. "Personalizing medicine through hybrid imaging and medical big data analysis." In: *Frontiers in Physics* 6 (2018), p. 51.

[15] A. Elnakib, G. Gimel'farb, J. S. Suri, and A. El-Baz. "Medical image segmentation: a brief survey." In: *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*. Springer, 2011, pp. 1–39.

[16] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." In: *IEEE transactions on medical imaging* 34.10 (2014), pp. 1993–2024.

[17] L. Wang, G. Li, F. Shi, X. Cao, C. Lian, D. Nie, M. Liu, H. Zhang, G. Li, Z. Wu, et al. "Volume-based analysis of 6-month-old infant brain MRI for autism biomarker identification and early diagnosis." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 411–419.

[18] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, et al. "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 415–423.

[19] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, et al. "ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI." In: *Medical image analysis* 35 (2017), pp. 250–269.

[20] M. Bieth, L. Peter, S. G. Nekolla, M. Eiber, G. Langs, M. Schwaiger, and B. Menze. "Segmentation of skeleton and organs in whole-body CT images via iterative trilateration." In: *IEEE Transactions on Medical Imaging* 36.11 (2017), pp. 2276–2286.

[21] H. Li, G. Jiang, J. Zhang, R. Wang, Z. Wang, W.-S. Zheng, and B. Menze. "Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images." In: *NeuroImage* 183 (2018), pp. 650–665.

[22] M. Sezgin and B. Sankur. "Survey over image thresholding techniques and quantitative performance evaluation." In: *Journal of Electronic imaging* 13.1 (2004), pp. 146–166.

[23] S. Hojjatoleslami and F. Kruggel. "Segmentation of large brain lesions." In: *IEEE Transactions on Medical Imaging* 20.7 (2001), pp. 666–669.

[24] S.-Y. Wan and W. E. Higgins. "Symmetric region growing." In: *IEEE Transactions on Image processing* 12.9 (2003), pp. 1007–1015.

[25] J. C. Bezdek, L. Hall, L. Clarke, et al. "Review of MR image segmentation techniques using pattern recognition." In: *MEDICAL PHYSICS-LANCASTER PA-* 20 (1993), pp. 1033–1033.

[26] D. W. Scott. *Multivariate density estimation: theory, practice, and visualization.* John Wiley & Sons, 2015.

[27] J. Franklin. "The elements of statistical learning: data mining, inference and prediction." In: *The Mathematical Intelligencer* 27.2 (2005), pp. 83–85.

[28] P. Yan, S. Xu, B. Turkbey, and J. Kruecker. "Discrete deformable model guided by partial active shape model for TRUS image segmentation." In: *IEEE Transactions on Biomedical Engineering* 57.5 (2010), pp. 1158–1166.

[29] J. Liu and J. K. Udupa. "Oriented active shape models." In: *IEEE Transactions on medical Imaging* 28.4 (2008), pp. 571–584.

[30] A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky. "Mutual information in coupled multi-shape model for medical image segmentation." In: *Medical image analysis* 8.4 (2004), pp. 429–445.

[31] J. Yang, L. H. Staib, and J. S. Duncan. "Neighbor-constrained segmentation with level set based 3-D deformable models." In: *IEEE Transactions on Medical Imaging* 23.8 (2004), pp. 940–948.

[32] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich. "Multi-atlas segmentation with joint label fusion." In: *IEEE transactions on pattern analysis and machine intelligence* 35.3 (2013), pp. 611–623.

[33] C.-C. Chang and C.-J. Lin. "LIBSVM: A library for support vector machines." In: *ACM transactions on intelligent systems and technology (TIST)* 2.3 (2011), pp. 1–27.

[34] E. Geremia, B. H. Menze, O. Clatz, E. Konukoglu, A. Criminisi, and N. Ayache. "Spatial decision forests for MS lesion segmentation in multi-channel MR images." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2010, pp. 111–118.

[35] M. P. Heinrich and M. Blendowski. "Multi-organ segmentation using vantage point forests and binary context features." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 598–606.

[36] T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, D. Rueckert, A. D. N. Initiative, et al. "Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling." In: *NeuroImage* 76 (2013), pp. 11–23.

[37] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning." In: *nature* 521.7553 (2015), p. 436.

[38] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[39] D. Nie, L. Wang, Y. Gao, and D. Shen. "Fully convolutional networks for multi-modality isointense infant brain image segmentation." In: *2016 IEEE 13Th international symposium on biomedical imaging (ISBI)*. IEEE. 2016, pp. 1342–1345.

[40] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng. "3D deeply supervised network for automatic liver segmentation from CT volumes." In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2016, pp. 149–157.

[41] L. Yu, X. Yang, H. Chen, J. Qin, P.-A. Heng, et al. "Volumetric convnets with mixed residual connections for automated prostate segmentation from 3D MR images." In: *AAAI*. Vol. 17. 2017, pp. 36–72.

[42] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation." In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[43] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 424–432.

[44] F. Milletari, N. Navab, and S.-A. Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." In: *2016 fourth international conference on 3D vision (3DV)*. IEEE. 2016, pp. 565–571.

[45] G. Lin, A. Milan, C. Shen, and I. Reid. "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1925–1934.

[46] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng. "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes." In: *IEEE transactions on medical imaging* 37.12 (2018), pp. 2663–2674.

[47] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis. "Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal." In: *IEEE journal of biomedical and health informatics* 24.2 (2019), pp. 568–576.

[48] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation." In: *IEEE transactions on medical imaging* 39.6 (2019), pp. 1856–1867.

[49] E. National Academies of Sciences, Medicine, et al. *Improving diagnosis in health care*. National Academies Press, 2015.

[50] B. Y. Reis, I. S. Kohane, and K. D. Mandl. "Longitudinal histories as predictors of future diagnoses of domestic abuse: modelling study." In: *Bmj* 339 (2009).

[51] A. Rajkomar, J. W. L. Yim, K. Grumbach, and A. Parekh. "Weighting primary care patient panel size: a novel electronic health record-derived measure using machine learning." In: *JMIR medical informatics* 4.4 (2016), e29.

[52] A. L. Beam and I. S. Kohane. "Big data and machine learning in health care." In: *Jama* 319.13 (2018), pp. 1317–1318.

[53] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, et al. "Clinically applicable deep learning for diagnosis and referral in retinal disease." In: *Nature medicine* 24.9 (2018), pp. 1342–1350.

[54] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen. "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans." In: *Scientific reports* 6.1 (2016), pp. 1–13.

[55] K. Doi. "Current status and future potential of computer-aided diagnosis in medical imaging." In: *The British journal of radiology* (2014).

[56] N. Sharma and L. M. Aggarwal. "Automated medical image segmentation techniques." In: *Journal of medical physics/Association of Medical Physicists of India* 35.1 (2010), p. 3.

[57] Z. Wang, K. K. Bhatia, B. Glocker, A. Marvao, T. Dawes, K. Misawa, K. Mori, and D. Rueckert. "Geodesic patch-based segmentation." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2014, pp. 666–673.

[58] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng. "3D deeply supervised network for automated segmentation of volumetric medical images." In: *Medical image analysis* 41 (2017), pp. 40–54.

[59] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert. "Automated abdominal multi-organ segmentation with subject-specific atlas generation." In: *IEEE transactions on medical imaging* 32.9 (2013), pp. 1723–1730.

[60] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. "Deep neural networks segment neuronal membranes in electron microscopy images." In: *Advances in neural information processing systems*. 2012, pp. 2843–2851.

[61] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle. "Brain tumor segmentation with deep neural networks." In: *Medical image analysis* 35 (2017), pp. 18–31.

[62] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár. "Learning to refine object segments." In: *European Conference on Computer Vision*. Springer. 2016, pp. 75–91.

[63] M. P. Heinrich and O. Oktay. "BRIEFnet: Deep Pancreas Segmentation using Binary Sparse Convolutions." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2017, pp. 329–337.

[64] Y. Song, G. Wu, K. Bahrami, Q. Sun, and D. Shen. "Progressive multi-atlas label fusion by dictionary evolution." In: *Medical image analysis* 36 (2017), pp. 162–171.

[65] G. Sanroma, O. M. Benkarim, G. Piella, O. Camara, G. Wu, D. Shen, J. D. Gispert, J. L. Molinuevo, M. A. G. Ballester, A. D. N. Initiative, et al. "Learning non-linear patch embeddings with neural networks for label fusion." In: *Medical image analysis* 44 (2018), pp. 143–155.

[66] G. Wu, M. Kim, G. Sanroma, Q. Wang, B. C. Munsell, D. Shen, A. D. N. Initiative, et al. "Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition." In: *NeuroImage* 106 (2015), pp. 34–46.

[67] G. Wu, Q. Wang, D. Zhang, F. Nie, H. Huang, and D. Shen. "A generative probability model of joint label fusion for multi-atlas based brain segmentation." In: *Medical image analysis* 18.6 (2014), pp. 881–890.

[68] D. Zikic, B. Glocker, and A. Criminisi. "Atlas encoding by randomized forests for efficient label propagation." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2013, pp. 66–73.

[69] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua. "BRIEF: Computing a local binary descriptor very fast." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.7 (2012), pp. 1281–1298.

[70] M. Bieth, E. Alberts, M. Schwaiger, and B. Menze. "From Large to Small Organ Segmentation in CT Using Regional Context." In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2017, pp. 1–9.

[71] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." In: *CA: a cancer journal for clinicians* 68.6 (2018), pp. 394–424.

[72] T. Maurer, M. Eiber, M. Schwaiger, and J. E. Gschwend. "Current use of PSMA–PET in prostate cancer management." In: *Nature Reviews Urology* 13.4 (2016), pp. 226–235.

[73] N. Howlader, A. Noone, M. Krapcho, N. Neyman, R. Aminou, S. Altekruse, C. Kosary, J. Ruhl, Z. Tatalovich, H. Cho, et al. "SEER cancer statistics review, 1975–2009 (vintage 2009 populations)." In: *Bethesda, MD: National Cancer Institute* (2012), pp. 1975–2009.

[74] K. D. Bernacki, K. L. Fields, and M. H. Roh. "The utility of PSMA and PSA immunohistochemistry in the cytologic diagnosis of metastatic prostate carcinoma." In: *Diagnostic cytopathology* 42.7 (2014), pp. 570–575.

[75] K. Fizazi, L. Faivre, F. Lesaunier, R. Delva, G. Gravis, F. Rolland, F. Priou, J.-M. Ferrero, N. Houede, L. Mourey, et al. "Androgen deprivation therapy plus docetaxel and estramustine versus androgen deprivation therapy alone for high-risk localised prostate cancer (GETUG 12): a phase 3 randomised controlled trial." In: *The Lancet Oncology* 16.7 (2015), pp. 787–794.

[76] M. Weineisen, M. Schottelius, J. Simecek, R. P. Baum, A. Yildiz, S. Beykan, H. R. Kulkarni, M. Lassmann, I. Klette, M. Eiber, et al. "68Ga-and 177Lu-labeled PSMA I&T: optimization of a PSMA-targeted theranostic concept and first proof-of-concept human studies." In: *Journal of Nuclear Medicine* 56.8 (2015), pp. 1169–1176.

[77] A. Afshar-Oromieh, T. Holland-Letz, F. L. Giesel, C. Kratochwil, W. Mier, S. Haufe, N. Debus, M. Eder, M. Eisenhut, M. Schäfer, et al. "Diagnostic performance of 68 Ga-PSMA-11 (HBED-CC) PET/CT in patients with recurrent prostate cancer: evaluation in 1007 patients." In: *European journal of nuclear medicine and molecular imaging* 44.8 (2017), pp. 1258–1268.

[78] C. Kratochwil, F. Bruchertseifer, F. L. Giesel, M. Weis, F. A. Verburg, F. Mottaghy, K. Kopka, C. Apostolidis, U. Haberkorn, and A. Morgenstern. "225Ac-PSMA-617 for PSMA-targeted α-radiation therapy of metastatic castration-resistant prostate cancer." In: *Journal of Nuclear Medicine* 57.12 (2016), pp. 1941–1944.

[79] C. Kratochwil, K. Schmidt, A. Afshar-Oromieh, F. Bruchertseifer, H. Rathke, A. Morgenstern, U. Haberkorn, and F. L. Giesel. "Targeted alpha therapy of mCRPC: Dosimetry estimate of 213 Bismuth-PSMA-617." In: *European journal of nuclear medicine and molecular imaging* 45.1 (2018), pp. 31–37.

[80] K. Rahbar, M. Schmidt, A. Heinzel, E. Eppard, A. Bode, A. Yordanova, M. Claesener, and H. Ahmadzadehfar. "Response and tolerability of a single dose of 177Lu-PSMA-617 in patients with metastatic castration-resistant prostate cancer: a multicenter retrospective analysis." In: *Journal of Nuclear Medicine* 57.9 (2016), pp. 1334–1338.

[81] A. Afshar-Oromieh, U. Haberkorn, C. Zechmann, T. Armor, W. Mier, F. Spohn, N. Debus, T. Holland-Letz, J. Babich, and C. Kratochwil. "Repeated PSMA-targeting radioligand therapy of metastatic prostate cancer with 131 I-MIP-1095." In: *European Journal of Nuclear Medicine and Molecular Imaging* 44.6 (2017), pp. 950–959.

[82] M. S. Hofman, J. Violet, R. J. Hicks, J. Ferdinandus, S. P. Thang, T. Akhurst, A. Iravani, G. Kong, A. R. Kumar, D. G. Murphy, et al. "[177Lu]-PSMA-617 radionuclide treatment in patients with metastatic castration-resistant prostate cancer (LuPSMA trial): a single-centre, single-arm, phase 2 study." In: *The Lancet Oncology* 19.6 (2018), pp. 825–833.

[83] M. Eiber, W. P. Fendler, S. P. Rowe, J. Calais, M. S. Hofman, T. Maurer, S. M. Schwarzenboeck, C. Kratowchil, K. Herrmann, and F. L. Giesel. "Prostate-specific membrane antigen ligands for imaging and therapy." In: *J Nucl Med* 58.Suppl 2 (2017), 67S–76S.

[84] M. Bieth, M. Krönke, R. Tauber, M. Dahlbender, M. Retz, S. G. Nekolla, B. Menze, T. Maurer, M. Eiber, and M. Schwaiger. "Exploring new multimodal quantitative imaging indices for the assessment of osseous tumor burden in prostate cancer using 68Ga-PSMA PET/CT." In: *Journal of Nuclear Medicine* 58.10 (2017), pp. 1632–1637.

[85] R. Segal, K. Miller, and A. Jemal. "Cancer statistics, 2018." In: *Ca Cancer J Clin* 68.1 (2018), pp. 7–30.

[86] L. L. Gunderson, D. J. Sargent, J. E. Tepper, N. Wolmark, M. J. O'Connell, M. Begovic, C. Allmer, L. Colangelo, S. R. Smalley, D. G. Haller, et al. "Impact of T and N stage and treatment on survival and relapse in adjuvant rectal cancer: a pooled analysis." In: *Journal of Clinical Oncology* 22.10 (2004), pp. 1785–1796.

[87] K. S. Chok and W. L. Law. "Prognostic factors affecting survival and recurrence of patients with pT1 and pT2 colorectal cancer." In: *World journal of surgery* 31.7 (2007), pp. 1485–1490.

[88] H. Wang, X.-Z. Wei, C.-G. Fu, R.-H. Zhao, and F.-A. Cao. "Patterns of lymph node metastasis are different in colon and rectal carcinomas." In: *World Journal of Gastroenterology: WJG* 16.42 (2010), p. 5375.

[89] G. J. Chang, M. A. Rodriguez-Bigas, J. M. Skibber, and V. A. Moyer. "Lymph node evaluation and survival after curative resection of colon cancer: systematic review." In: *Journal of the National Cancer Institute* 99.6 (2007), pp. 433–441.

[90] Z.-H. Zhou. "A brief introduction to weakly supervised learning." In: *National Science Review* 5.1 (2018), pp. 44–53.

[91] P. Tang, X. Wang, A. Wang, Y. Yan, W. Liu, J. Huang, and A. Yuille. "Weakly supervised region proposal network and object detection." In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 352–368.

[92] F. Wan, C. Liu, W. Ke, X. Ji, J. Jiao, and Q. Ye. "C-mil: Continuation multiple instance learning for weakly supervised object detection." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2199–2208.

[93] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan. "Salient object detection via multiple instance learning." In: *IEEE Transactions on Image Processing* 26.4 (2017), pp. 1911–1922.

[94] G. Xu, Z. Song, Z. Sun, C. Ku, Z. Yang, C. Liu, S. Wang, J. Ma, and W. Xu. "Camel: A weakly supervised learning framework for histopathology image segmentation." In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 10682–10691.

[95] D. Pathak, P. Krahenbuhl, and T. Darrell. "Constrained convolutional neural networks for weakly supervised segmentation." In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1796–1804.

[96] J. Wu, Y. Yu, C. Huang, and K. Yu. "Deep multiple instance learning for image classification and auto-annotation." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3460–3469.

[97] M. Ilse, J. M. Tomczak, and M. Welling. "Attention-based deep multiple instance learning." In: *35th International Conference on Machine Learning, ICML 2018*. International Machine Learning Society (IMLS). 2018, pp. 3376–3391.

[98] S. Manivannan, C. Cobb, S. Burgess, and E. Trucco. "Subcategory classifiers for multiple-instance learning and its application to retinal nerve fiber layer visibility classification." In: *IEEE transactions on medical imaging* 36.5 (2017), pp. 1140–1150.

[99] J. Amores. "Multiple instance classification: Review, taxonomy and comparative study." In: *Artificial intelligence* 201 (2013), pp. 81–105.

[100] H. Lee, A. Battle, R. Raina, and A. Y. Ng. "Efficient sparse coding algorithms." In: *Advances in neural information processing systems*. 2007, pp. 801–808.

[101] Q. Zhang and S. A. Goldman. "EM-DD: An improved multiple-instance learning technique." In: *Advances in neural information processing systems*. 2002, pp. 1073–1080.

[102] S. Andrews, I. Tsochantaridis, and T. Hofmann. "Support vector machines for multiple-instance learning." In: *Advances in neural information processing systems*. 2003, pp. 577–584.

[103] G. Liu, J. Wu, and Z.-H. Zhou. "Key Instance Detection in Multi-Instance Learning." In: *Asian Conference on Machine Learning*. 2012, pp. 253–268.

[104] E. R. Dorsey, A. Elbaz, E. Nichols, F. Abd-Allah, A. Abdelalim, J. C. Adsuar, M. G. Ansha, C. Brayne, J.-Y. J. Choi, D. Collado-Mateo, et al. "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016." In: *The Lancet Neurology* 17.11 (2018), pp. 939–953.

[105] A. J. Hughes, S. E. Daniel, Y. Ben-Shlomo, and A. J. Lees. "The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service." In: *Brain* 125.4 (2002), pp. 861–870.

[106] A. J. Hughes, Y. Ben-Shlomo, S. E. Daniel, and A. J. Lees. "What features improve the accuracy of clinical diagnosis in Parkinson's disease: a clinicopathologic study." In: *Neurology* 42.6 (1992), pp. 1142–1142.

[107] K. L. Chou, M. S. Forman, J. Q. Trojanowski, H. I. Hurtig, and G. H. Baltuch. "Subthalamic nucleus deep brain stimulation in a patient with levodopa-responsive multiple system atrophy: case report." In: *Journal of neurosurgery* 100.3 (2004), pp. 553–556.

[108] V. Lambrecq, E. Krim, W. Meissner, D. Guehl, and F. Tison. "Deep-brain stimulation of the internal pallidum in multiple system atrophy." In: *Revue neurologique* 164.4 (2008), pp. 398–402.

[109] A. J. Stoessl, S. Lehericy, and A. P. Strafella. "Imaging insights into basal ganglia function, Parkinson's disease, and dystonia." In: *The Lancet* 384.9942 (2014), pp. 532–544.

[110] C. C. Tang, K. L. Poston, T. Eckert, A. Feigin, S. Frucht, M. Gudesblatt, V. Dhawan, M. Lesser, J.-P. Vonsattel, S. Fahn, et al. "Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis." In: *The Lancet Neurology* 9.2 (2010), pp. 149–158.

[111] R. Gautam and M. Sharma. "Prevalence and Diagnosis of Neurological Disorders Using Different Deep Learning Techniques: A Meta-Analysis." In: *Journal of Medical Systems* 44.2 (2020), p. 49.

[112] S. Kiryu, K. Yasaka, H. Akai, Y. Nakata, Y. Sugomori, S. Hara, M. Seo, O. Abe, and K. Ohtomo. "Deep learning to differentiate parkinsonian disorders separately using single midsagittal MR imaging: a proof of concept study." In: *European radiology* 29.12 (2019), pp. 6891–6899.

[113] C. R. Boland and A. Goel. "Microsatellite instability in colorectal cancer." In: *Gastroenterology* 138.6 (2010), pp. 2073–2087.

[114] H. T. Lynch and A. De la Chapelle. "Hereditary colorectal cancer." In: *New England Journal of Medicine* 348.10 (2003), pp. 919–932.

[115] E. Vilar and S. B. Gruber. "Microsatellite instability in colorectal cancer—the stable evidence." In: *Nature reviews Clinical oncology* 7.3 (2010), p. 153.

[116] Y. Zhao, H. Li, S. Wan, A. Sekuboyina, X. Hu, G. Tetteh, M. Piraud, and B. Menze. "Knowledge-aided convolutional neural network for small organ segmentation." In: *IEEE journal of biomedical and health informatics* 23.4 (2019), pp. 1363–1373.

[117] Y. Zhao, F. Yang, Y. Fang, H. Liu, N. Zhou, J. Zhang, J. Sun, S. Yang, B. Menze, X. Fan, et al. "Predicting Lymph Node Metastasis Using Histopathological Images Based on Multiple Instance Learning With Deep Graph Convolution." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 4837–4846.

[118] Y. Zhao, A. Gafita, G. Tetteh, F. Haupt, A. Afshar-Oromieh, B. Menze, M. Eiber, A. Rominger, and K. Shi. "Deep Neural Network for Automatic Characterization of Lesions on ^{68}Ga-PSMA PET/CT Images." In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 951–954.

[119] Y. Zhao, P. Wu, J. Wang, H. Li, N. Navab, I. Yakushev, W. Weber, M. Schwaiger, S.-C. Huang, P. Cumming, et al. "A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes on ^{18}F-FDG PET Images." In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 3531–3534.

[120] R. Cao*, F. Yang*, S.-C. Ma*, L. Liu*, Y. Zhao*, Y. Li*, D.-H. Wu, T. Wang, W.-J. Lu, W.-J. Cai, H.-B. Zhu, X.-J. Guo, Y.-W. Lu, J.-J. Kuang, W.-J. Huan, W.-M. Tang, K. Huang, J. Huang, J. Yao, and Z.-Y. Dong. "Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer." In: *Theranostics* 10 (2020), pp. 11080–11091.

[121] S. K. Zhou, H. Greenspan, and D. Shen. *Deep learning for medical image analysis*. Academic Press, 2017.

[122] Y. Bengio, A. Courville, and P. Vincent. "Representation learning: A review and new perspectives." In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), pp. 1798–1828.

[123] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors." In: *nature* 323.6088 (1986), pp. 533–536.

[124] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[125] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition." In: *arXiv preprint arXiv:1409.1556* (2014).

[126] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going deeper with convolutions." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[127] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[128] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. "Densely connected convolutional networks." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.

[129] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. "Deformable convolutional networks." In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 764–773.

[130] R. Zhang, F. Zhu, J. Liu, and G. Liu. "Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis." In: *IEEE Transactions on Information Forensics and Security* 15 (2019), pp. 1138–1150.

[131] F. Yu and V. Koltun. "Multi-scale context aggregation by dilated convolutions." In: *arXiv preprint arXiv:1511.07122* (2015).

[132] M. D. Zeiler and R. Fergus. "Visualizing and understanding convolutional networks." In: *European conference on computer vision*. Springer. 2014, pp. 818–833.

[133] V. Dumoulin and F. Visin. "A guide to convolution arithmetic for deep learning." In: *arXiv preprint arXiv:1603.07285* (2016).

[134] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. "Focal loss for dense object detection." In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988.

[135] S. Jadon. "A survey of loss functions for semantic segmentation." In: *arXiv preprint arXiv:2006.14822* (2020).

[136] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed. "Boundary loss for highly unbalanced segmentation." In: *International conference on medical imaging with deep learning*. 2019, pp. 285–296.

[137] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim. "Sanity checks for saliency maps." In: *Advances in Neural Information Processing Systems*. 2018, pp. 9505–9515.

[138] A. Alqaraawi, M. Schuessler, P. Weiß, E. Costanza, and N. Berthouze. "Evaluating saliency map explanations for convolutional neural networks: a user study." In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 2020, pp. 275–285.

[139] A. Borji. "Saliency prediction in the deep learning era: Successes and limitations." In: *IEEE transactions on pattern analysis and machine intelligence* (2019).

[140] S. Srinivas and F. Fleuret. "Full-gradient representation for neural network visualization." In: *Advances in Neural Information Processing Systems*. 2019, pp. 4124–4133.

[141] M. Zhang, Y. Zhou, J. Zhao, Y. Man, B. Liu, and R. Yao. "A survey of semi-and weakly supervised semantic segmentation of images." In: *Artificial Intelligence Review* (2019), pp. 1–30.

[142] L. Jing and Y. Tian. "Self-supervised visual feature learning with deep neural networks: A survey." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[143] T. Zhou, S. Ruan, and S. Canu. "A review: Deep learning for medical image segmentation using multi-modality fusion." In: *Array* 3 (2019), p. 100004.