Guest Editorial
Open and Interpretable AI in Computational Pathology

RECENT years have seen exponential advances in the quality of Artificial Intelligence (AI) techniques in the medical filed. Particularly, AI/ML tools have been widely exploited pathological image analysis to examine and assess the function of human organs and/or to provide trustworthy prediction of diseases. However, black-box nature of classical AI based on deep networks have limited its clinical use. Thus, a growing need to build more transparent and explainable AI tools has opened a new research avenue several known as explainable or interpretable AI. The latter aims to improve human understanding, determine the justifiability of decisions made by the machine, introduce trust and reduce bias. In computational pathology, building trustworthy and explainable AI systems are of immense importance to facilitate high-quality healthcare to patients.

Despite their widespread, most of the existing ML models are becoming less interpretable, i.e., neither provide explanations nor trustworthy for the predictions. On the other hand, trustworthy and explainability in AI tools based on Deep Learning (DL) is an emerging field of research with great promise for increased high-quality healthcare that produce human-comprehensible solutions, i.e., provide explanations and interpretations for disease diagnosis and predictions, as well as recommended actions. The explainable solution enables enhanced prediction accuracy with decision understanding and traceability of actions taken.

The focus of this research topic is on the recent AI-based systems focusing on interpretable AI methods to generate human readable explanations. The target audience for this topic includes engineering and medical researchers, particularly, those with the purpose to (i) improve the trust and reduce analysis bias, (ii) stimulate discussion on the system designs, and (iii) use and evaluate novel explainable AI for improving the accuracy of pathology workflows for disease diagnostic and prognostic. Also, research work dedicated on practical use-cases of trustworthy AI models that discuss adding a layer of interpretability and trust to powerful algorithms.

The topic consists of 13 accepted articles that deal with artificial intelligence/ML algorithms to investigate different problems. A brief summary of each article is provided below.

A study by Ganaie et al. [A1] for Alzheimer's disease (AD) employed a support vector regression (SVR) based models for brain age estimation. They introduced an improved least square twin SVR (ILSTSVR) module to overcome the limitation of twin SVR (TSVR), such as overfitting. Also, the model incorporated the regularization term and formulates different Lagrangian functions such that the inversion of matrix computation is avoided. Evaluation of their formulated ILSTSVR model is conducted on the subjects including cognitively healthy, mild cognitive impairment and AD for brain-age estimation. Experimental evaluation and statistical tests demonstrate the efficiency of the proposed ILSTSVR model for brain-age prediction.

Shao et al. [A2] introduced a pipeline to improve the accuracy and interpretability of predicting cardiovascular and cerebrovascular events (CCEs) based on a novel heart rate variability (HRV) feature extraction method. The designed scheme integrates instantaneous high-order singular entropy (IHSE) and a deep belief network (DBN). The former use 8 HRV features based on the pathogenesis of CCEs. The DBN model incorporate prior knowledge. Evaluation was conducted on public data set with demonstrated advantages over the existing schemes for predicting CCEs (overall accuracy was 90.80%).

The third paper by Nassem et al. [A3] adopted a vision-language transformer that embeds vision (images) and language (questions) features for an interpretable transformer-based Path-VQA (TraP-VQA). In their design transformers' encoder layers are embedded with vision and language features extracted using pre-trained CNN and domain-specific language model (LM), respectively. Evaluation on public PathVQA dataset of 4998 images and 32799 QA pairs demonstrated that TraPVQA outperformed the state-of-the-art comparative. Namely, final prediction for PathVQA was 64.82% accuracy on both open and close ended question types. Grad-Cam and SHAP were used to interpret retrieved answers visually to indicate which area of the image contributed to the predicted answer using Resnet 50 to focus on the region of interest.

A deep learning framework for inferring anticancer synergistic combinations is proposed in [A4] by Alsherbiny et al. The so-called SynPredict model uses five synergy metrics together with the combination sensitivity score (CSS) to provide and effective prediction of synergy. Their model also assessed the impact of multimodal fusion architectures of the input data, including the gene expression data of cancer cells, along with the representative chemical features of drugs in pairwise combinations. Both ONEIL and ALMANAC anticancer combination datasets are employed comparatively and early and intermediate data integration are investigated to highlight the impact of heterogeneous data. Evaluation reveals that Synpredict outperforms existing state-of-the-art predictive models, including

---

Date of current version 5 April 2023.
Digital Object Identifier 10.1109/JBHI.2023.3252839

DeepSynergy, AuDNN synergy, TranSynergy and DrugComb, with up to 74% decline in the mean square error.

To segment gland of colonic histopathological images, Ding et al. developed a two-path feature gland segmentation algorithm of colon pathological image based on local semantic guidance [A5]. They introduced an improved region search algorithm to provide more local feature information of pathological tissue for the model. Additionally, the proposed network integrates a candidate region semantic guidance network and a context feature extraction network with attention. The interpretability of their method is analyzed and compared with the other six mainstream algorithms. Model verification was conducted on Warwick Qu-Dataset, compared against popular segmentation algorithms. Also, the algorithm has demonstrated performance using area metrics (Dice coefficient, F1-score) and surface-to-surface metrics (Hausdorff distance) on different types of test sets.

In the context of recording depression-related emotional vital signs, Ahmed et al. [A6] developed attention networks that utilized covert levels of self- attention to of detect sadness in social media by focusing on terms in personal remarks. The network assigns each node in a neighborhood its weight without performing costly matrix operations (e.g., similarity or network architecture knowledge) since nodes/words can express the properties/emotions of their neighbors. The latent representation of the synonym is created using this lexicon to expand the synonym according to its context of use. The attention-based method trains the input text and categorizes it based on the learned representation. Data was collected from an Internet forum, web-site, and social networking site while using Amazon Mechanical Turk3 to label the 500 texts. The method achieved a 0.87 area under the receiver operating characteristics (ROC).

For a study of autism spectrum disorder (ASD), Wadhera et al. [A7] presented a framework to study network topology using the node degree concept. Their pipeline included brain networks constructions using a visibility, new metric that encode edge weights between nodes to compute the participation of the brain ROIs to a complex network, and an explainable machine learning (XML) classifier. Their introduced new metric is called weighted hierarchical complexity (WHC) compute neural hierarchical complexity of the brain. An XML step using SVM classification highlighted the role WHC effectiveness in classifying ASD with an overall accuracy of 98.76%. The study concluded that, graph-theory weaker long-range connections and stronger intra-connections are markers of ASD. Namely, the whole architecture of the intrinsic brain activity can be characterised by chain-like structures organised in a hierarchical modular arrangement.

One of the challenges of deep neural networks (DNNs) trainings is the scenario of noisy-labels. To achieve more sufficient-learning on clean samples, Wang et al. proposed an adaptive learning algorithm, termed as meta-probability weighting (MPW), to assign different weights to the components of output probability and more robustness to label noise [A2]. Their method is inspired by the sample-weighting strategy and have the advantages of (1) the ability to adaptively learn the probability weight from data without a complicated weight function design and (2) utilizing easy general optimization algorithm in which well-known gradient-descent-based networks can be utilized. Using both real and synthetic data, MPW achieves competitive performance with other state-of-the-art methods.

Chen et al. [A9] introduced a contrastive-based learning method for AD prediction using 18F-FDG PET images. In their analysis approach, PET slices were augmented with filtering and cropping and subjects' fuzzy labels were used as supervised information to expand (narrow) inter-class (intra-class) feature distances and to construct contrastive losses. To increase the feature aware field range of the network, their architecture adopted a prediction network based on a double convolution hybrid attention mechanism with the advantage of suppressing redundant information from PET slices as well. Finally, multi-correlation analysis on the predicted PET results is integrated with the results of clinical neuropsychological assessment for comprehensive AD prediction. The experimental results documented the advantage of the method in effectively predicting AD in comparison with the state-of-the-arts work on brain 18F-FDG PET images, in addition to its satisfactory computational performance

A local histogram feature (LHF) extraction algorithm that is specific-to-cervical-grading problem is proposed by Calik et al. [A10]. The LHF includes RGB spectral values from upper, central and basal regions of epithelial tissue. Consequently, the proposed LHF vectors have discriminative and explainable representation inspired by pathologists' interpretation of the tissue grading problem. The authors also introduced two classification methods, namely, statistical local histogram classification (S-LHC), and neural network-based LHC (NN-LHC). A dataset of 957 SEP images consisting of 4 categories (Normal, CNI1. CNI2, CNI3) is used for evaluation and an overall of 78.69% accuracy was reported, while morphology-based methods were 69.07% and convolutional neural network (CNN) patch-based algorithms were 75.77%.

Habib et al. [A11] utilized a CNN with sinc-convolution as a constrained first-layer in an effort to establish transparency in DL model's inner-working and decision-making. The proposed technique aims at using prominent sinc frequency-bands, as a means of CNN-based model optimisation. Particularly, sinc kernels are band-pass filter-like kernels with lower and upper cut frequencies that are optimized via backpropagation. The effects of sinc-kernels was visualised by means of explanation vector of joint time-frequency representation of time-series signal. A step that help in identifying comparatively significant frequency-bands, that may provide domain-specific interpretation, for the given task. The proposed architecture was validated using ECG beat-classification (five-class classification task), and R-peak localisation (sample-wise segmentation task).

A two-stage pipeline by Shao et al. [A12] was introduced for intracranial aneurysms detection. Namely an unsupervised representation learning on 3D point clouds of vessel segments is proposed and consists of unsupervised pre-training and downstream tasks. They introduced an effective augmentation method for generating pairs of each vessel segment. As for the downstream tasks, a dual-branch contrastive network with an encoder for each branch and a subsequent common projection head to facilitate training with contrastive loss. The effectiveness of their method id highlighted using limited label situation with comprehensive experiments and comparison with state-of-the-art supervised techniques. In total, classification accuracy of 90.79% on ModelNet40 and 95.01 on ModelNet10 over IntraA public dataset of 2025 vessel segments.

Finally, interpretability of tumor mutation burden (TMB) was investigated in [A13] Liu et al. using a two-step prognostic framework using whole-slide images (WSIs). Firstly, their framework adopted a deep residual network to encode the WSIs' phenotype and classifies patient-level TMB. Secondly, stratification of patients' prognosis is conducted using the TMB-related information obtained during the classification model development. The study focused on clear cell renal cell carcinoma (ccRCC) using local data sets of 295 Haematoxylin & Eosin stained WSIs, and prognostic biomarkers evaluation is is performed on The Cancer Genome Atlas-Kidney ccRCC (TCGA-KIRC) project with 304 WSIs. The study findings documented the feasibility of mining TMB-related information from WSI to achieve stepwise prognosis prediction. This has been confirmed with the higher AUC for TBM classification and significant stratification of patients' overall survival ($p < 0.05$), outperforming the original TMB signature in risk stratification of patients with advanced disease.

In total, all thirteen papers tackle different but extremely relevant to interpretable machine learning in computational pathology and focusing on development of better diagnostic tools to analyze medical images. We believe this Special Issue will raise awareness in the scientific community, through presenting and highlighting the advances and latest novel and emergent technologies, implementations, applications concerning the early diagnosis of life-threatening disease. In closing, we would like to thank all the authors who submitted their research work to this special issue. We would also like to acknowledge the contribution of many experts in the field who have participated in the review process, and provided helpful suggestions to the authors to improve the contents and presentations of the articles. We would in particular like to thank Professor Dimitrios I. Fotiadis, the Editor-in-Chief, and the publishing team for their support and very helpful suggestions and comments during the delicate stages of concluding the special issue.

IMRAN RAZZAK, *Guest Editor*
School of Computer Science and Engineering
University of New South Wales Sydney
Sydney NSW 2052, Australia
imran.razzak@unsw.edu.au

MUHAMMAD KHURRAM KHAN, *Guest Editor*
Center of Excellence in Information Assurance
King Saud University
Riyadh 11653, Saudi Arabia
mkhurram@ksu.edu.au

GUANDONG XU, *Guest Editor*
School of Computer Science
University of Technology
Sydney, NSW 2008, Australia
guandong.xu@uts.edu.au

FAHMI KHALIFA, *Guest Editor*
Electrical and Computer Engineering
Department
School of Engineering
Morgan State University
Baltimore MD 21251 USA
fahmi.khalifa@morgan.edu

APPENDIX
RELATED ARTICLES

[A1] M. A. Ganaie, M. Tanveer, and I. Beheshti, "Brain age prediction with improved least squares twin SVR," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3147524.

[A2] S. Shao, T. Wang, A. Mumtaz, C. Song, and C. Yao, "Predicting cardiovascular and cerebrovascular events based on instantaneous high-order singular entropy and deep belief network," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3162894.

[A3] U. Naseem, M. Khushi, and J. Kim, "Vision-language transformer for interpretable pathology visual question answering," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3163751.

[A4] M. Alsherbiny et al., "Trustworthy deep neural network for inferring anticancer synergistic combinations," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2021.3126339.

[A5] S. Ding, H. Wang, H. Lu, M. Nappi, and S. Wan, "Two path gland segmentation algorithm of colon pathological image based on local semantic guidance," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3207874.

[A6] U. Ahmed, J. C.-W. Lin, and G. Srivastava, "Multi-aspect deep active attention network for healthcare explainable adoption," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3204633.

[A7] T. Wadhera and M. Mahmud, "Brain functional network topology in autism spectrum disorder: A novel weighted hierarchical complexity metric for electroencephalogram," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3232550.

[A8] Z. Wang, S. Jin, L. Li, Y. Dong, and Q. Hu, "Meta-probability weighting for improving reliability of DNNs to label noise," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2023.3237033.

[A9] Y. Chen et al., “Contrastive learning for prediction of Alzheimer’s disease using brain 18F-FDG PET,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3231905.

[A10] N. Calik et al., “Classification of cervical precursor lesions via local histogram and cell morphometric features,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3218293.

[A11] A. Habib, C. Karmakar, and J. Yearwood, “Interpretability and optimisation of convolutional neural networks based on sinc-convolution,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3185290.

[A12] D. Shao, X. Lu, and X. Liu, “3D intracranial aneurysm classification and segmentation via unsupervised dual-branch learning,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2022.3180326.

[A13] X. Liu et al., “Development of prognostic biomarkers by TMB-guided WSI analysis: A two-step approach,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, Apr. 2023, doi: 10.1109/JBHI.2023.3249354.