

Guest Editorial

Special Section on Learning With Multimodal Data for Biomedical Informatics

IN THIS Special Section of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, it is our honor to present emerging advanced machine learning and data analytics algorithms aiming at catalyzing synergies among image/video processing, text/speech understanding, and multi-modal learning in biomedical informatics. Our goals are to 1) introduce novel data-driven models to accelerate knowledge discovery in biomedicine through the seamless integration of medical data collected from imaging systems, laboratory and wearable devices, as well as other related medical devices; 2) promote the development of new multi-modal learning systems to enhance the healthcare quality and patient safety; and 3) promote new applications in biomedical informatics that can leverage or benefits from the integration of multi-modal data and machine learning.

Fast-growing biomedical and healthcare data of multiple modalities have encompassed multiple scales ranging from molecules, individuals, to populations. With increasing sources, bandwidth, depth, and resolution, those data are becoming a key enabler for accelerating basic science discoveries and facilitating evidence-based clinical solutions. Meanwhile, the heterogeneous and increasingly more diverse modalities of the data present major barriers toward their understanding, fusion, and translation into effective clinical actions. For example, electronic health records (EHRs) are representative examples of multimodal/multisource data collections, including not only measurements but also images, videos, audios, and free texts. The diversity of information sources and the increasing amounts of medical data produced by healthcare institutes pose significant challenges for data-driven health monitoring and decision making. Other examples include mobile health for remote patient care, where typical data modalities include patient- or caregiver-generated photos, self-reported symptoms of pain, body temperature, and more.

While biomedical and healthcare research traditionally focuses on the structured measurement data, the growing availability of novel data modalities has created a compelling demand for novel machine learning, image/video/audio/text processing, and multi-modal fusion algorithms that specifically tackle the unique challenges associated with biomedical and healthcare data and allow decision-makers and stakeholders to better interpret and exploit the data.

This proposed Special Section has offered a timely collection of information to benefit the researchers and practitioners working in the broad research fields of biomedical informatics applications, signal and image processing, and machine learning community, including in a series of 12 articles to familiarize readers with this fast-moving research field. On a high level, those articles can be categorized as “image-focused” [A1]–[A5], “medical report-focused” [A6], [A7], and “other multimodal information source and management” [A8]–[A12]. A short description of the contributions brought by each article is next presented.

In [A1], Mo *et al.* studied the segmentation of multimodal MRI, an essential preprocessing step for initial diagnosis, stage differentiation, and post-treatment efficacy evaluation in clinical situations. They proposed a graph learning-based approach to efficiently extract modality-specific features and establish regional correspondence effectively among all modalities. Their framework is validated on two multimodal MRI datasets: our private liver lesion dataset and a public prostate zone dataset.

In [A2], Avola *et al.* proposed a novel end-to-end knowledge-driven classification framework, focusing on multimodal data generated by thyroid ultrasonography. It acts as a computer-aided diagnosis (CAD) system by providing a thyroid nodule classification into the benign and malignant categories. The proposed system leverages cues provided by an ensemble of experts to guide the learning phase of a densely connected convolutional network (DenseNet). The ensemble is composed of various networks pretrained on ImageNet, including AlexNet, ResNet, VGG, and others. The proposed system achieves relevant performances in terms of qualitative metrics for the thyroid nodule classification task.

To facilitate the assessment of a COVID-19 patient’s severity, in [A3], Zhou *et al.* proposed a multi-modality feature learning and fusion model for end-to-end patient severity prediction using the blood test supported electronic medical record (EMR) and chest computerized tomography (CT) scan images. The experimental results demonstrate that the proposed multimodality feature learning and fusion model achieves high performance in realistic scenarios.

In [A4], Gao *et al.* studied cross-resolution face recognition (CRFR), a classical topic in biometric forensics which recently demonstrated utility in physical and mental health monitoring. The authors proposed to fully exploit the multilevel deep convolutional neural network (CNN) feature set for robust CRFR, via adaptive cross-layer fusion, feature set-based

representation learning, and hierarchical output fusion. In [A5], Das *et al.* proposed a novel spatio-temporal framework for apathy classification, which is streamlined to analyze facial dynamics and emotion in videos. Using a gated recurrent unit (GRU) architecture, they show that fusion of characteristics such as emotion and facial dynamics obtains a high accuracy of 95.34% in apathy classification.

In [A6], Karthikeyan *et al.* developed an OCR post-correction approach based on the robustly optimized bidirectional encoder representations from transformers for effective medical reports processing. Specifically, they propose to add a post-processing component to the OCR engine to identify incorrect words and improve accuracy in the OCR outputs. This method is evaluated on two datasets: the U.K. NHS medical reports which consist of scanned documents including clinical letters, reports, and discharge summaries, and the publicly available MiBio dataset. The experimental results demonstrate that this approach is applicable to domain-specific datasets for the effective reduction of word error rate in OCR outputs.

In [A7], Momoki *et al.* proposed a novel machine-learning algorithm to fuse text data from radiology reports with image data from computed tomography (CT) for effective detection and characterization of pulmonary nodules. They build an image classifier that is trained by pseudo-labels extracted by a text classifier from radiology reports to cope with the challenge of the lack of training data. This approach does not require the manual image labeling process, which is extremely time-consuming and requires medical domain knowledge. The experimental results show that the proposed method is capable of effectively fusing information from both the radiology reports and CT images by achieving the same level of detection performance as the one trained by CT images with labels annotated by radiologists.

The fast-growing techniques of measuring and fusing multimodal biomedical signals enable advanced motor intent decoding schemes of lower limb exoskeletons. In [A8], Yi *et al.* proposed an ahead-of-time continuous prediction of lower limb kinematics, with the prediction of knee angles during level walking as a case study. They utilized electromyography (EMG) signals of nine muscles and knee angles calculated from inertial measurement unit (IMU) signals recorded from subjects, and showed successful prediction of knee angles. This is the first study of continuously predicting lower limb kinematics in an ahead-of-time manner based on the electromechanical delay (EMD).

In [A9], Cheng *et al.* studied the precise prediction of deleterious SMS. They explored multimodal features across four groups including functional score, conservation, splicing, and sequence features, and then trained eight conceptually different machine learning classifiers for each of them, resulting in 32 base classification models to make ensembles. The results suggested that their ensembling solution achieved better performance compared to other state-of-the-art predictors on the training and independent test datasets.

In [A10], Delvigne *et al.* studied attention deficit hyperactivity disorder (ADHD), the most prevalent neurodevelopmental disorder among children. ADHD affects patients' lives

in many ways: inattention, difficulty with stimuli inhibition, or motor function regulation. The authors present a novel framework to record physiological signals in specific attention states and able to estimate the corresponding attention state. They propose a database composed of electroencephalography signals (EEG), and an eye-tracker labeled with a score representing the attention span. Different features are derived from the signals and new machine learning (ML) algorithms are proposed, exhibiting high accuracy for attention estimation.

In [A11], Cao *et al.* studied approximate nearest neighbor (ANN), a canonical problem in large-scale multimodal retrieval. The authors analyzed the importance of various hashing projected dimensions, distributing them into several groups and quantizing them with two types of values which can both better preserve the neighborhood structure among data. They validated the proposed framework by conducting experiments on five common public data sets containing up to one million vectors. In [A12], Zhou studied feature selection in small-data regimes, with multi-source transfer where the privileged information from another data source or modality—only available during the training phase—is exploited to improve the performance of feature selection.

Finally, we would also like to thank the Editor-in-Chief Feng Wu and the Associate Editor-in-Chief Zhu Li for their support in guiding the development of this Special Section from its initial proposal to its final publication. We are also deeply indebted to the reviewers for their quick and insightful responses to the manuscripts submitted.

Zhangyang Wang, *Guest Editor*

Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712 USA
e-mail: atlaswang@utexas.edu.

Vishal Patel, *Guest Editor*

Department of Electrical and Computer Engineering
Johns Hopkins University
Baltimore, MD 21218 USA
e-mail: vpa-tel36@jhu.edu

BING YAO, *Guest Editor*

Department of Industrial Engineering and Management
Oklahoma State University
Stillwater, OK 70047 USA
e-mail: bing.yao@okstate.edu

STEVE JIANG, *Guest Editor*

Division of Medical Physics and Engineering
University of Texas Southwestern Medical Center
Dallas, TX 75390 USA
e-mail: steve.jiang@utsouthwestern.edu

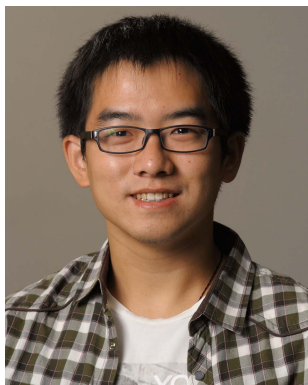
HUIMIN LU, *Guest Editor*

Department of Mechanical and Control Engineering
Kyushu Institute of Technology
Kitakyushu, Fukuoka 804-8550, Japan
email: dr.huimin.lu@ieee.org

YANG SHEN, *Guest Editor*
 Department of Electrical and Computer Engineering
 Texas A&M University
 College Station, TX 77843 USA
 e-mail: yshen@tamu.edu

APPENDIX: RELATED ARTICLES

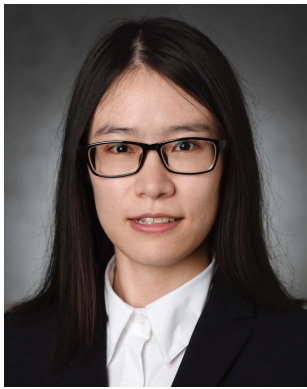
- [A1] S. Mo *et al.*, "Mutual information-based graph co-attention networks for multimodal prior-guided magnetic resonance imaging segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2512–2526, May 2022.
- [A2] D. Avola, L. Cinque, A. Fagioli, S. Filetti, G. Grani, and E. Rodola, "Multimodal feature fusion and knowledge-driven learning via experts consult for thyroid nodule classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2527–2534, May 2022.
- [A3] J. Zhou *et al.*, "Cohesive multi-modality feature learning and fusion for COVID-19 patient severity prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2535–2549, May 2022.
- [A4] G. Gao, Y. Yu, J. Yang, G.-J. Qi, and M. Yang, "Hierarchical deep CNN feature set-based representation learning for robust cross-resolution face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2550–2560, May 2022.
- [A5] A. Das *et al.*, "A spatio-temporal approach for apathy classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2561–2573, May 2022.
- [A6] S. Karthikeyan, A. G. S. de Herrera, F. Doctor, and A. Mirza, "An OCR post-correction approach using deep learning for processing medical reports," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2574–2581, May 2022.
- [A7] Y. Momoki, A. Ichinose, Y. Shigeto, U. Honda, K. Nakamura, and Y. Matsumoto, "Characterization of pulmonary nodules in computed tomography images based on pseudo-labeling using radiology reports," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2582–2591, May 2022.
- [A8] C. Yi *et al.*, "Continuous prediction of lower-limb kinematics from multi-modal biomedical signals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2592–2602, May 2022.
- [A9] N. Cheng *et al.*, "An ensemble framework for improving the prediction of deleterious synonymous mutation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2603–2611, May 2022.
- [A10] V. Delvigne, H. Wannous, T. Dutoit, L. Ris, and J.-P. Vandeborre, "PhyDAA: Physiological dataset assessing attention," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2612–2623, May 2022.
- [A11] Y. Cao, S. Chen, J. Gui, H. Qi, Z. Li, and C. Liu, "Hash learning with variable quantization for large-scale retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2624–2637, May 2022.
- [A12] J. T. Zhou, "Feature selection with multi-source transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2638–2646, May 2022.



Zhangyang Wang (Senior Member, IEEE) received the Ph.D. degree from the University of Illinois at Urbana–Champaign, under the supervision of Prof. Thomas Huang. He was an Assistant Professor at Texas A&M University from 2017 to 2020. He is currently the Jack Kilby/Texas Instruments Endowed Assistant Professor with the Department of Electrical and Computer Engineering, The University of Texas at Austin. His research interests include machine learning, computer vision, optimization, and their interdisciplinary applications. Most recently, he studies automated machine learning (AutoML), learning to optimize (L2O), robust learning, efficient learning, and graph neural networks.



Vishal Patel received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2010, working under Prof. Rama Chellappa. He is currently an Associate Professor with the Department of Electrical and Computer Engineering (ECE), Johns Hopkins University (JHU). Prior to joining JHU, he was an A. Walter Tyson Assistant Professor with the Department of ECE, Rutgers University, and a member of the Research Faculty at the University of Maryland Institute for Advanced Computer Studies (UMIACS). His current research interests include signal processing, computer vision, pattern recognition, and multi-modal fusion, with applications in biometrics and medical imaging.



Bing Yao received the B.S. degree in physics from the University of Science and Technology of China in 2012 and the Ph.D. degree in industrial engineering and operations research from The Pennsylvania State University in 2019. She is currently an Assistant Professor with the School of Industrial Engineering and Management, Oklahoma State University. Her research focuses on the physical–statistical modeling and optimization of complex systems. Specifically, her research interests include machine learning, data mining, and multi-modal signal processing, in particular for applications in spatiotemporal complex systems, cardiac electrodynamics, smart health management, and advanced manufacturing.



Steve Jiang received the Ph.D. degree in medical physics from the Medical College of Ohio in 1998. After completing his post-doctoral training at Stanford University, he joined the Massachusetts General Hospital and Harvard Medical School, as an Assistant Professor of radiation oncology, in 2000. In 2007, he was recruited to the University of California at San Diego as a Tenured Associate Professor to build the Center for Advanced Radiotherapy Technologies, for which he was the Founding and Executive Director. He was then promoted to a Full Professor with tenure in 2011. In October 2013, he joined the University of Texas Southwestern Medical Center, as a Tenured Full Professor, a Barbara Crittenden Professor in Cancer Research, the Vice Chair of the Radiation Oncology Department, and the Director of the Medical Physics and Engineering Division. His research interest includes the development and deployment of artificial intelligence and image processing technologies to solve medical problems. He is a fellow of the Institute of Physics and the American Association of Physicists in Medicine.



Huimin Lu (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Kyushu Institute of Technology, Japan, in 2014. From 2013 to 2016, he was a JSPS Research Fellow (DC2, PD, and FPD) at the Kyushu Institute of Technology. He is currently an Associate Professor with the Kyushu Institute of Technology and an Excellent Young Researcher of the Ministry of Education, Culture, Sports, Science and Technology. His research interests include artificial intelligence, machine vision, the Internet of Things, and robotics.



Yang Shen (Member, IEEE) received the B.E. degree from the University of Science and Technology of China and the Ph.D. degree in systems engineering from Boston University. He received post-doctoral training in biological engineering at the Massachusetts Institute of Technology. He is currently an Associate Professor with the Department of Electrical and Computer Engineering and the Center for Bioinformatics and Genomic Systems Engineering, Texas A&M University. His research interests are in optimization and learning algorithms for modeling biological molecules, systems, and data. Applications include protein docking, protein and drug design, and systems and synthetic biology. He is a recipient of the Maximizing Investigators' Research Award (MIRA) from the National Institute of General Medical Sciences of the National Institutes of Health.