Advancing Artificial Intelligence in Sensors, Signals, and Imaging Informatics

William Hsu[1], Christian Baumgartner[2], Thomas Deserno[3], Section Editors for the IMIA Yearbook Section on Sensors, Signals, and Imaging Informatics

[1] Medical and Imaging Informatics, Department of Radiological Sciences, University of California, Los Angeles, United States of America
[2] Institute of Health Care Engineering with European Testing Center of Medical Devices, Graz University of Technology, Austria
[3] Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany

Summary

Objective: To identify research works that exemplify recent developments in the field of sensors, signals, and imaging informatics.
Method: A broad literature search was conducted using PubMed and Web of Science, supplemented with individual papers that were nominated by section editors. A predefined query made from a combination of Medical Subject Heading (MeSH) terms and keywords were used to search both sources. Section editors then filtered the entire set of retrieved papers with each paper having been reviewed by two section editors. Papers were assessed on a three-point Likert scale by two section editors, rated from 0 (do not include) to 2 (should be included). Only papers with a combined score of 2 or above were considered.
Results: A search for papers was executed at the start of January 2019, resulting in a combined set of 1,459 records published in 2018 in 119 unique journals. Section editors jointly filtered the list of candidates down to 14 nominations. The 14 candidate best papers were then ranked by a group of eight external reviewers. Four papers, representing different international groups and journals, were selected as the best

papers by consensus of the International Medical Informatics Association (IMIA) Yearbook editorial board.
Conclusions: The fields of sensors, signals, and imaging informatics have rapidly evolved with the application of novel artificial intelligence/machine learning techniques. Studies have been able to discover hidden patterns and integrate different types of data towards improving diagnostic accuracy and patient outcomes. However, the quality of papers varied widely without clear reporting standards for these types of models. Nevertheless, a number of papers have demonstrated useful techniques to improve the generalizability, interpretability, and reproducibility of increasingly sophisticated models.

Keywords

Sensors, signals, imaging informatics; medical informatics; artificial intelligence; machine learning

Introduction

No other field has experienced the magnitude of impact of recent developments in artificial intelligence (AI) and machine learning (ML) like sensors, signals, and imaging informatics (SSII). Keeping pace with the rapid evolution in this field is challenging. Readers are encouraged to refer to a number of high quality review articles that have been published in 2018 in the areas of sensors and signals [1, 2], imaging informatics [3, 4], and both [5]. The selected candidate and best papers highlighted in this synopsis provide a representative sampling of noteworthy work being conducted internationally that have advanced aspects of AI/ML in healthcare involving sensors, signals, and imaging informatics data.

Paper Selection Process

The process of searching the literature for candidate best papers by the SSII section remained largely unchanged from prior years. Two large bibliographic indices, PubMed and Web of Science, were searched using a combination of Medical Subject Headings (MeSH) and terms provided by the section editors.

The objective of the search process was to generate a manageable list of papers for section editors to review while encompassing the breadth of work that has been conducted this past year and the diversity of journals. For sensors and signals, terms encompassed computer-assisted processing, physiologic monitoring, and biosensing techniques (e.g., electroretinogram, photoplethysmography). For imaging informatics, terms included imaging modalities (e.g., computed tomography, electrocardiogram), image processing (e.g., segmentation, registration), and machine and deep learning techniques (e.g., convolutional neural networks, generative adversarial networks). The query was further restricted to journal articles written in English, describing original research, including an abstract, and published in 2018. Scripts used to execute the full queries are available upon request from the corresponding author.

The original search was performed during the first week of January 2019. The search result was made of 1,459 papers published in 119 unique journals. Results were imported into BibReview [6], a software tool primarily used to standardize the formatting of citation and abstract formats and to perform an initial review. Citations were then exported into Google Sheets, which allowed real-time concurrent review of the citations by all section editors.

The identification of candidate best papers proceeded in three phases. In the first phase, each section editor was assigned a set of papers to identify relevant papers

based on their titles and/or abstracts. The assignments were made such that each paper was reviewed by two section editors. Papers were rated on a Likert scale between 0 to 2, where 0 was assigned to papers that were irrelevant and should be removed from consideration, and 2 was given to papers that were highly relevant and should be considered. Papers with a cumulative score of 2 or above (n=24) were kept. In the second phase, every section editor independently reviewed these remaining papers and again assigned a score. The section editors discussed the papers based on the results of the scoring and nominated 14 candidate best papers for the SSII section. In the third phase, the candidate best papers were sent to a group of eight external reviewers, comprised of other Yearbook editorial members and representative researchers from the SSII community who were not involved in the production of the Yearbook. Each paper was considered by at least five external reviewers and assigned an overall score based on criteria such as scientific impact, quality of content, originality, and clarity. Section editors nominated five best papers taking the external scores into account while selecting papers that are representative of the diverse research, institutions, and journals in the field. Finally, the Yearbook editorial board discussed the nominations and identified any potential overlap with other sections, and settled on four best papers for the SSII section (see Table 1).

Selected Best Papers and Their Implications

The paper by Lee *et al*., [7] investigates variants of deep convolutional neural networks to classify the presence of intracranial hemorrhage (ICH) and if present, its subtype. The work is noteworthy for its attempt to visually explain how the model is processing input images. As models become increasingly complex, the uninterpretable "black box" nature of ML algorithms needs to be addressed [8]. Class activation maps and activation maximization are techniques that help spatially visualize features or regions where the networks are focusing to make the subtype prediction

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2019 in the section 'Sensors, Signals, and Imaging Informatics'. The articles are listed in alphabetical order of the first author's surname.

Section
Sensors, Signals, and Imaging Informatics
▪ Bandeira Diniz JO, Bandeira Diniz PH, Azevedo Valente TL, Corrêa Silva A, de Paiva AC, Gattass M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Comput Methods Programs Biomed 2018 Mar;156:191-207.
▪ Lee H, Yune S, Mansouri M, Ki M, Tajmir SH, Guerrier CE, Ebert SA, Pomerantz SR, Romero JM, Kamalian S, Gonzalez RG, Lev MH, Do S. An explainable deep-learning algorithm for the detection of acute intracranial hemorrhage from small datasets. Nat Biomed Eng 2019 Mar;3(3):173-82.
▪ Samad MD, Ulloa A, Wehner GJ, Jing L, Hartzel D, Good CW, Williams BA, Haggerty CM, Fornwalt BK. Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning. JACC Cardiovasc Imaging 2019 Apr;12(4):681-9.
▪ Vasilakakis MD, Iakovidis DK, Spyrou E, Koulaouzidis A. DINOSARC: color features based on selective aggregation of chromatic image components for wireless capsule endoscopy. Comput Math Methods Med 2018 Sep 3;2018:2026962.

[5]. Lee *et al*., [7] provided an effective demonstration of how attention maps are able to help radiologist readers understand regions that contributed to the model's prediction of ICH subtype. They found a 78.1% overlap rate when interpreting regions where the model focused its attention versus segmentations of ICH regions done by a radiologist. Moreover, the authors investigated variants of the model architecture (i.e., VGG16, ResNet-50, Inception-v3, Inception-ResNet-v2) and the use of various data augmentation and preprocessing techniques to generate the best performing ensemble model.

The paper by Samad et al., [9] predicts five-year mortality using a combination of clinical, ejection function, and echocardiographic measurements. Authors showed that a random forest model was able to achieve high performance (AUC = 0.89) as compared to existing clinical risk scores with value demonstrated from the inclusion of echocardiography-derived features. The study's dataset is impressive in scope. The cohort consisted of 171,510 patients who underwent 331,317 echocardiograms and were extracted from the authors' institutional medical record system. However, despite the sheer size of the dataset, missing values remained a practical challenge. The authors discussed techniques such as multivariate imputation by chained equations to address this issue. They also utilized the "mean decrease impurity" approach [10] to characterize the relative importance of features in the random forest model. In summary,

the authors described a large-scale study leveraging ML framework to generate accurate outcome predictions using real-world clinical cardiology data.

The paper by Bandeira Diniz et al., [11] describes a pipeline to automate the detection of masses in mammograms. Using a set of models based on convolutional neural networks, the authors first classified whether the mammogram showed dense or non-dense breast tissue. The images were then registered, breast parenchyma was segmented, and candidate masses were extracted. Two separate models tailored to dense versus non-dense breasts were used to perform the final determination of whether a candidate was a mass. The manuscript was notably detailed in the description of the method developed, outlining the entire pipeline. The work was also noteworthy in its use of domain knowledge to facilitate classification tasks. While the field has largely relied on the inherent ability of deep learning techniques to discover salient features directly from the data, incorporating domain knowledge as part of the process of modeling training can improve overall performance. The authors demonstrated that the use of bilateral analysis (comparing mammograms of each breast to each other with the assumption that they should be symmetrical) and knowledge of breast density (dense versus non-dense) would produce better models.

The paper by Vasilakakis *et al*., [12] presents an approach to extract salient features from a large set of color images taken during wireless capsule endoscopy (WCE). During

this procedure, a pill containing a color camera passes through a patient's bowel, collecting images with the goal of detecting abnormalities such as bleeding, polyps, and ulcers. The reliable extraction of salient features from a high dimensional dataset is a critical aspect of ML. The paper describes an automated approach to generate color-based salient points that are then used to generate local and global image descriptors that characterized potential abnormalities. The authors showed that their approach, distances on selective aggregation of chromatic image components (DINOSARC), generated more relevant points in a more efficient manner as compared to other detectors such as scale-invariant feature transform (SIFT).

Discussion and Conclusion

Identifying the best papers for this year's Yearbook afforded an opportunity to gain a high-level characterization of the current trends in the broad field of SSII. The use of deep learning methods remains a trending area of research. However, as researchers attempt to bridge the gap from experimentation to clinical translation of these methods, papers are addressing issues such as model interpretability [7].

Evaluations that involve comparing the performance of human readers against AI/ML are increasingly being reported [13, 14]. Studies have also moved towards training and testing on larger datasets [9, 11, 14, 15]. While a majority of papers utilized deep learning techniques, several of the represented papers apply traditional machine learning techniques (e.g., logistic regression, random forest, fuzzy support vector machines) [9, 16], traditional signal processing [17], and structural co-occurrence matrix [18]. Moreover, papers touched upon the topics that are agnostic to the modeling approach, including the detection of salient features from a high dimensional feature space [12], and the integration of data from multiple heterogeneous data sources [9].

The diversity of institutions and countries represented in the 14 candidate best papers has increased with the greatest number

of papers coming from institutions in the United States, the United Kingdom, and China. Nevertheless, the field continues to see a wide range of heterogeneity in how models and their results are reported in the literature. The availability of datasets used to train models or the models themselves is scarce, making fair comparisons of model performance difficult. One candidate paper proposed a standardized pipeline to facilitate model comparisons, which is a step towards this direction [19].

In conclusion, SSII are rapidly evolving fields with a growing number of successful examples of how AI/ML can improve sensor, signal, and imaging data. The next horizon for AI in SSII is the continued development, refinement, and translation of these algorithms in an ethical, interpretable, and reproducible manner.

References

1. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. Comput Methods Programs Biomed 2018;161:1-13.
2. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 2019;115:213-37.
3. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer 2018;18(8):500-10.
4. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, et al; Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. Can Assoc Radiol J 2018 May;69(2):120-35.
5. Yu K-H, Beam AL, Kohane IS. (2018) Artificial intelligence in healthcare. Nat Biomed Eng 2018;2(10):719.
6. Bouaud J, Jaulent MC, Kerdelhué G, Griffon N, Séroussi B, Lamy JB. Toward a Formalization of the Process to Select IMIA Yearbook Best Papers. Methods Inf Med 2018;54(2):135-44.
7. Lee H, Yune S, Mansouri M, Kim M, Tajmir SH, Guerrier CE, et al. (2019) An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nat Biomed Eng 2019;3(3):173-82.
8. Hsu W, Elmore JG. Shining Light Into the Black Box of Machine Learning. J Natl Cancer Inst 2019 Jan 10.
9. Samad, MD, Ulloa A, Wehner GJ, Jing L, Hartzel D, Good CW, et al. Predicting Survival From Large

Echocardiography and Electronic Health Record Datasets Optimization With Machine Learning. JACC Cardiovasc Imaging 2018;12(4):681-9.
10. Louppe G, Wehenkel L, Sutera A, Geurts P. Understanding variable importances in forests of randomized trees. Adv Neural Inf Process Syst 2013:431-9.
11. Diniz JOB, Diniz PHB, Valente TLA, Silva AC, de Paiva AC, Gattass M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Comput Methods Programs Biomed 2018;156:191-207.
12. Vasilakakis MD, Iakovidis DK, Spyrou E, Koulaouzidis A. (2018) DINOSARC: Color Features Based on Selective Aggregation of Chromatic Image Components for Wireless Capsule Endoscopy. Comput Math Methods Med 2018:1-11.
13. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson 2018;20(1):65.
14. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. (2018) Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15(11):e1002686.
15. Wu J, Tan Y, Chen Z, Zhao M. Decision based on big data research for non-small cell lung cancer in medical artificial system in developing country. Comput Methods Programs Biomed 2018;159:87-101.
16. Huang H, Bruschini C, Antfolk C, Enz C, Li T, Justiz J, Koch VM. Automatic hand phantom map generation and detection using decomposition support vector machines. Biomed Eng Online 2018;17(1):74.
17. Khavas ZR, Asl BM. (2018) Robust Heartbeat Detection Using Multimodal Recordings and ECG Quality Assessment with Signal Amplitudes Dispersion. Comput Methods Programs Biomed 2018;163:169-82.
18. Filho PPR, Peixoto SA, da Nóbrega RVM, Hemanth DJ, Medeiros AG, Sangaiah AK, et al. Automatic histologically-closer classification of skin lesions. Comput Med Imaging Graph 2018;68:40-54.
19. Samper-González J, Burgos N, Bottani S, Fontanella S, Lu P, Marcoux A, et al; for the Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle flagship study of ageing. (2018) Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. Neuroimage 2018;183:504-21.

Correspondence to:
William Hsu
Medical and Imaging Informatics
Department of Radiological Sciences
University of California, Los Angeles
United States of America
E-mail: WHsu@mednet.ucla.edu

Appendix: Content Summaries of Selected Best Papers for the 2019 IMIA Yearbook, Section Sensors, Signals, and Imaging Informatics

Bandeira Diniz JO, Bandeira Diniz PH, Azevedo Valente TL, Corrêa Silva A, de Paiva AC, Gattass M

Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks

Mammographic imaging is a critical tool for detecting breast cancers early. However, 13% of cancers go undetected. Computational techniques are being explored to improve the early detection rate of suspicious findings associated with breast cancer. In this paper, the authors describe a computer-aided diagnosis methodology to assist breast radiologists with identifying mass regions in 2D mammographic images. The approach consists of several modules: a preprocessing step (*e.g.*, resizing, cropping), a density classification model (*i.e.*, classifying dense versus non-dense breasts), an image and segmentation step, the selection of candidate masses, and a final classifier to identify actual masses. Separate final classifiers are trained for dense and non-dense breast images. In addition, the authors utilize bilateral analysis to identify differences between contralateral breasts as a way to assist in identifying candidate masses. Their system was trained using a large public dataset called Digital Database for Screening Mammography, which contains more than 2,500 labeled exams of digitized film mammograms. They showed that the method achieved 91% accuracy in classifying masses in non-dense breast tissue and 95% accuracy in detecting masses in dense breasts. This work demonstrates the utility of incorporating domain knowledge such as breast density and bilateral analysis to improve the performance of mass detection. The authors provide a comprehensive expla-

nation of their methodology and comparison of their approach's performance with other state-of-the-art methods.

Lee H, Yune S, Mansouri M, Ki M, Tajmir SH, Guerrier CE, Ebert SA, Pomerantz SR, Romero JM, Kamalian S, Gonzalez RG, Lev MH, Do S

An explainable deep-learning algorithm for the detection of acute intracranial hemorrhage from small datasets

The rapid development of machine and deep learning algorithms have yielded systems that are capable of improving diagnostic accuracy and optimizing the delivery of healthcare. However, a major hurdle towards adoption is the increasing difficulty in fully explaining or understanding the basis of these models' predictions. The authors demonstrate an approach that assists human users with understanding the basis of a model's prediction using a combination of heatmap-based visualizations. The overarching goal of their work is to automatically predict whether intracranial hemorrhage (ICH) is present in a non-contrast head computed tomography (CT) study and if so, classify its subtype. The optimal model architecture was investigated by using multiple pre-trained deep convolutional neural network architectures (VGG16, ResNet-50, Inception-v3, Inception-ResNet-v2) and various data preprocessing techniques. The authors used the class activation mapping technique to highlight important regions of the image for the prediction to a target label. They also generated a radiology atlas of ICH by ranking all activation maps at all blocks for each label based on relevance count and selecting the top 5% of them to best represent the image assigned with the label. The authors had access to an imbalanced dataset of 625 ICH-positive and 279 ICH-negative cases of which 100 positive and 100 negative cases were set aside as a validation set. A retrospective dataset of 100 ICH-positive and 100 ICH-negative cases and a prospective dataset of 79 ICH-positive and 117 ICH-negative cases collected consecutively at a single institution were used to evaluate model performance. Model performance was compared against five radiologists per

dataset (three residents and two board-certified attendings). The system achieved similar performance to that of radiologists in both datasets, achieving a sensitivity of 98% and 92% in the retrospective and prospective datasets, respectively. The work is notable for its effective use of heatmaps to visualize model regions and its comprehensive evaluation of model performance across datasets and against human readers.

Samad MD, Ulloa A, Wehner GJ, Jing L, Hartzel D, Good CW, Williams BA, Haggerty CM, Fornwalt BK

Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning

Over 10 million echocardiograms are performed annually in United States Medicare patients alone. When interpreted alongside patient demographic information, these imaging data are useful in diagnosing and prognosticating a patient's status. However, physicians face the challenge of relating the sheer number of measurements that are generated from echocardiography to the dozens of plausible diagnoses that could be coded from these images. In this paper, the authors present an analytical framework for predicting one- and five-year mortality survival using electronic health record data, including echocardiograms, from a large patient population seen at their institution. Their approach combines clinical variables (age, sex, height, weight, heart rate, blood pressure, cholesterol, smoking status, and 90 ICD-10 codes (International Classification of Diseases 10th revision)) along with physician-reported left ventricular ejection fraction, and 516 different echocardiographic measurements. Missing data were imputed using multivariate imputation by chained equations (MICE). Linear (logistic regression) and nonlinear (random forest) models were trained and evaluated using a 10-fold nested cross-validation design. The authors had access to data from over 171,510 patients and 331,317 electrocardiograms. Machine learning models achieved higher prediction accuracy (AUC

> 0.82) over common clinical risk scores such as Framingham (AUC between 0.61 to 0.79). The nonlinear model outperformed the linear model. The most common informative features across all predictions were found to be age, tricuspid regurgitation jet maximum velocity, heart rate, and left ventricular ejection fraction. This work demonstrated the feasibility of harnessing a large clinical population and combining clinical and imaging information to significantly improve the performance of mortality predictions.

Vasilakakis MD, Iakovidis DK, Spyrou E, Koulaouzidis A

DINOSARC: color features based on selective aggregation of chromatic image components for wireless capsule endoscopy

Comput Math Methods Med 2018 Sep 3;2018:2026962

Wireless Capsule Endoscopy (WCE) is a swallowable pill containing a miniaturized camera that generates hundreds of thousands of color images of the digestive tract. While an innovative and non-invasive way to detect abnormalities such as ulcers, polyps, and bleeding, interpreting WCE images is time-consuming and error-prone. Salient point detection, an unsupervised process of extracting image features that are most associated with abnormalities, is increasingly being used to facilitate discrimination between normal and abnormal WCE images. In this study, the authors developed a salient point and region detection algorithm to estimate local and global image descriptors that are predictive of abnormalities. The algorithm consists of several components including a color-based salient point detector (based on a narrow color range that is usually located on the margins of the overall color range of an image), a salient region detector based on superpixels (using the simple iterative linear clustering algorithm), and a generator of local and global color image descriptors that are extracted from the superpixel regions. A public dataset previously released by the authors consisting of images obtained from a MiroCam endoscope with 360x360 pixels was used to evaluate the approach. The developed algorithm, DINOSARC, achieved the highest percentage of true positive salient points over other state-of-the-art algorithms. It also demonstrated the highest overall area under the curve using its generated local and global image descriptors to classify normal versus abnormal WCE images. This work is important in its focus on salient feature detection and the identification of local and global image descriptors in an efficient and unsupervised manner. It is also demonstrating how novel sensors, combined with appropriate imaging informatics, can be made applicable for a clinical setting.