Future Perspectives on Automated Machine Learning in Biomedical Signal Processing ⋆

Luis Miguel Lopez-Ramos[0000-0001-8072-3994]

[1] SFI Offshore Mechatronics Center, University of Agder
[2] Department of ICT, University of Agder, Grimstad, Norway

Abstract. Recent developemnts in Machine Learning (ML) and Signal Processing (SP) are of paramount importance for the advancement of artificial intelligence in the health sector. The implications of these enabling technologies encompass several broad fields from mathematics to product development in connection with the practitioners and patients. This paper aims at guiding researchers and entrepreneurs in the journey from basic research to application of data-driven technology for the health sector. Two key topics related to automatization of data processing, namely sparsity-based processing, and automated ML, are introduced and discussed in relation to biomedical SP. The discussed topics are exemplified in the context of cardiac signal processing.

Keywords: signal processing · biomedical engineering · interpretability · machine learning · automated machine learning · autonomous information management

1 Introduction

Machine learning (ML) algorithms play a key role in recent technological advances to improve clinical care, design new medicines, and reduce health-care costs [1]. Predictive algorithms are part of the core of enhanced diagnosis, automated treatments, and systematic understanding of disease progression [2]. Under the umbrella term of Artificial intelligence (AI), many different technologies are applied and continuously evolved to increase the quality of computer-aided patient care, monitoring [3], and healthcare system management [4].

Not all algorithms are based on learning from data, but the development of algorithms to efficiently implement *Big-data* analytics [5, 6] has raised a lot of interest in the last years in many sectors, including healthcare. In this context, the availability of enormous amounts of information stems from the rapid and systematic digitalization of hard copies of patient records [7], and the capability of new equipment to easily store data from medical imaging (e.g., magnetic resonance imaging (MRI) [8]) and biomedical signals such as electrocardiography (ECG) [6] and electroencephalography (EEG) [9].

⋆ The work in this paper was supported by the SFI Offshore Mechatronics grant 237896/O30.

An increased capability to store and process large amounts of data at a practical speed allows to have access to data from large numbers of patients[3], but it also allows to store signal data with a high resolution, and reducing the negative impact of compression algorithms. The analysis of signals with high resolution and high dimensionality [10] is subject to a great deal of research activity, where the recent advances in ML are rapidly incorporated and exploited.

Signal processing (SP) is the enabling technology for the transformation and interpretation of real-world sources of information in many forms such as electrical potentials, pressure, light or, more generally, anything that can be measured by a sensor or user interface. Signals are different from bulk data in the fact that they are defined on structured domains, such as time, space (in the case of images or volumetric signals), or graphs (in the case of, e.g. epidemiological [11] or point-cloud data [12]). Whereas SP extracts knowledge from information considering its structure, biomedical SP focuses on physiological activities ranging from gene and protein sequences, to neural and cardiac rhythms, to tissue and organ images [13].

Recent mathematical developments in statistical SP (SSP) are a rigorous foundation for signal interpretation [14, 15] beyond mere prediction of outcomes, enabling theoretical performance bounds [16], or interpretation of numerical parameters of trained models [17], a simple but prototypical example being regression analysis [18]. Interpretability [19], explainability [20], and adaptation to diverse tasks [21] are desirable properties of ML and SP algorithms, especially when they are part of critical systems such as biomedical signal monitoring and decision-making [22, 23]. Modern SP research seeks continuous improvement in terms of adaptivity of signal refining algorithms [24], and accuracy of pathology classification [25]. The main technical challenges in this context are related to the computational complexity [26] and the need for large and representative datasets [27] for the algorithms to be applicable in a wide range of patients and clinical cases.

The vision that this paper aims to put forth is that the design of signal processing algorithms must be based on rigorous formulation of optimization problemsand judiciously chosen figures of merit [28]. In other words, effective algorithm development is based on rigorously applied mathematics. With the correct formulation, application of novel optimization methods has demonstrated the potential to reduce the computational and large sample requirements [29] of current signal processing techniques. Much of the research effort consists in carefully analyzing the problem formulations and data structures to find the mathematical properties that allow to reduce computation times by orders of magnitude[30]. When this is achieved, the researcher has broken the barrier associated with "intractable" technical problems by actually making them tractable,

[3] It is worth noting that the legal and ethical challenges associated with the automated analysis of personal sensitive data are equally important and closely associated with the technical challenges, motivating the inter-disciplinarity between both research areas.

and creating new opportunities for development of products that will have a direct impact on the health sector.

Overview The goal of this paper is to explain more in detail the interrelation between the multiple research topics that are relevant to the development of new solutions for biomedical signal and data processing, and how this can have an impact in making eHealth accessible to an ample majority of the world population (democratization and reduction of the technological gap). Specifically, several topics in mathematics, computation, ML theory, SP/data analysis, and biomedical applications will be discussed in depth.

2 From mathematical formulas to data-driven biomedical applications

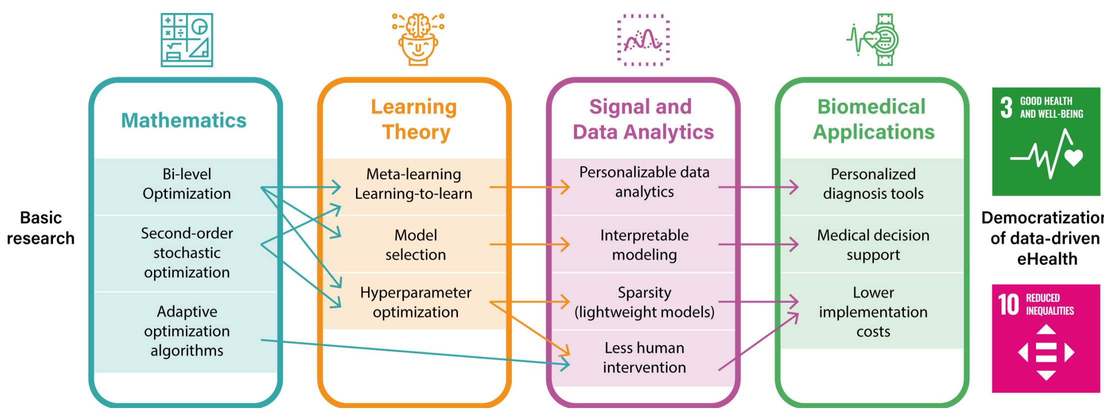

Fig. 1. Summary of the key research topics that, in view of the author, will enable innovation in Biomedical SP research in the near future.

2.1 Mathematical foundations of ML

Optimization: finding models with good data fit. Optimization theory is the basis to design and analyse the properties of adaptive mechanisms in many applications in science and engineering. The best known applications of optimization formulations are in engineering system design in areas such as telecommunications, mechatronics, and computational physics. However, optimization formulations are of paramount importance in ML and SP. With the rigorous formulation of optimization problems where data-fit cost functions are optimized, together with recently developed and powerful optimization algorithms, modern applications such as artificial vision, dimensionality reduction, data fusion, and biomedical engineering enjoy advancements in capabilities that were only achievable by humans in the past

Statistics: reasoning under uncertainty. For various reasons ranging from equipment imperfections to radiation dose reduction, the acquired data may suffer from uncertainty. Many datasets contain bad or missing data, and the need of imputing (inferring) some data entries from the rest of the available data becomes larger as larger the number of variables or data points are. Also, noise is unavoidable in any physical data acquisition system, and the signal processing research continuously provides with novel techniques to mitigate the effect of noise using the structure underlying the data.

2.2 ML theory and applications

Optimization formulations aiming at finding the model with the best data fit are challenged by a problem known as overfitting, which can be described intuitively as the behaviour when a machine "memorizes" the data with a low generalization capability, so that it will perform poorly when used in cases not covered by the training data. To avoid overfitting, optimization problems in ML are often regularized, to express a priori information about the desired properties, and/or promote simple solutions.

Data engineers often need to define a penalty (regularization) function based on application-dependent information. Current practice can benefit from more systematic choice methods. Most often, a regularization technique involves the addition of one or more additional parameter (a.k.a. hyperparameter) to control the intensity of the regularization and, thus, the trade-off between the quality of data fit and the model simplicity. Procedures to systematically adjust hyperparameters have only recently been proposed [31], and they have not been fully analysed yet despite their importance.

Meta-learning techniques (and more specifically, model-agnostic meta-learning, MAML[32]) have the power to produce personalized data analysis tools as reported in [33]. On the other hand, meta-learning has proven to enhance the clinical risk prediction with limited electronic health records [34]. The latter two results motivate further studies to exploit the potential of meta-learning.

Meta-learning and hyperparameter optimization rely heavily on bi-level optimization [31]. Several techniques are available, and some of them use second-order information (Hessian) of the inner objective. In ML applications though, due to the high dimensionality, models are optimized through first-order stochastic methods, and second-order information is often not available. Recent development of second-order stochastic optimization has a clear potential to extract the necessary second-order information for the bi-level optimization problems that this project aims to formulate and solve.

2.3 Signal and Data Processing

Recent advances in SP are based on ML and optimization algorithms. Empirical risk minimization (fitting a model to the available data) is not always enough though, as it does not take into account the problems associated with excessive complexity of the models.

2.4 Biomedical Applications

One of the multiple knowledge needs in biomedical engineering and eHealth is related to the efficiency of the data processing algorithms. Improvements in such efficiency have an economic meaning but, more importantly, direct benefits for individual and public health. For the reasons argued above, aiming this research action at developing interpretable SIP algorithms specifically for biomedical/e-Health applications has a huge potential to turn recent mathematical developments into a direct and tangible benefit for the society.

The aforementioned results in Meta-learning [33, 34] motivate future studies where meta-learning is used to produce a system for personalized clinical risk prediction tool, but no research work has addressed this task yet.

Cardiac pathologies. In order to focus our discussion on a specific example, Sec. 5 will focus on biomedical SP in cardiology. The classification of cardiac pathologies from electrocardiographic (ECG) signals will be discussed. The widespread use of ECG recordings relies heavily on the development of SIP tools for denoising and interference suppression, feature extraction, and diagnosis of cardiac diseases.

3 Sparsity: automated simplicity

Sparse data structures can be easily compressed as they contain many zero entries. Sparsity is discussed and exploited in many fields of mathematics: mainly to improve the efficiency at solving large systems of equations. In sparse data, having the information concentrated in a few data entries helps interpretability. Data structures such as vectors, matrices, tensors, graphs and neural networks, the sparser they are, the more efficiently they can be stored and processed, because all the zero entries do not need to be written/loaded in memory.

One of the most relevant aspects of sparsity is that, apart from being useful for mathematical procedures, it can be enforced in the data structures that are obtained by inference algorithms that are present in engineering, statistics, measurement equipment, data analytics, and network science. The ability of the aforementioned algorithms to obtain data representations that (apart from accurately describing the data) are as sparse as possible, has a direct impact on the scalability of the data-processing systems.

Lengthy descriptions are generally not human-readable, and patients and doctors need to access the available information at different levels of detail. One of the advantages of sparsity-aware SP is that it is naturally endowed with the ability to regulate the level of detail of a model inferred from data.

Natural signals can often be processed in a way such that relevant information lies in a few points of the signal domain (e.g., frequencies), as opposed to complex representations that are not human-readable. To this end, sparsity-aware signal processing (SP) algorithms can detect the structure that enhances interpretability. This is important in health-related applications as health issues are generally explained by a restricted set of variables [35].

Sparsity-aware signal processing uses recent advances in mathematical optimization and statistics to exploit sparsity in signals and networks in both natural and technical contexts.

The recent mathematical developments in statistical learning with sparsity [36] are a rigorous foundation for interpretation of sparse vectors of coefficients when performing, e.g., a regression analysis. In order to discover a sparse structure (which is also associated with good generalization of ML models), regularization is a critical element. If this is realized correctly, the impact in the application side will be huge.

Optimization formulations are often regularized to promote simple solutions such as sparse structures, i.e., where information is concentrated in a few entries while most of the rest are zero. Sparsity, as a form of simplicity, is one of the key features for interpretability. However, sparsity-promoting penalties are sensitive to hyperparameters.

Recent developments in sparsity from a rigorous mathematical point of view [37] characterize the properties of statistical learning with sparsity. Beyond the mathematical analysis, engineering researchers and practitioners are actively finding new ways to exploit sparsity to improve the applicability and impact of their algorithms.

4 Automated Machine Learning (AutoML)

Adaptation to diverse tasks is a desirable propertiy of ML algorithm. In this context, the recently developed paradigm of automated machine learning (AutoML) [38] is leading to a higher degree of autonomous information management as it allows to adjust ML models with reduced human intervention

In the context of AutoML, learning-to-learn techniques [32] allow to tailor ML model training for specific classes of problems, to process data even more efficiently than traditional ML approaches. This project aims at using these tools to bring personalization of data analysis to tackle the diversity of health-related needs across individuals and population groups. Recent works about meta-learning (learning general features for classes of tasks) have motivated increased interest for bi-level optimization techniques. Additionally, algorithms that automatically optimize hyperparameters [39] have a huge potential for developing signal processing (SP) algorithms.

The meta-learning framework breaks the “one machine-one task” pattern present in ML, one of the obstacles towards a general artificial intelligence (AI).

Instead of processing one dataset associated with one task, a meta-learning algorithm processes a collection of datasets (related to a class of tasks) extracting general information useful for learning a new task of the same class with a minimal amount of data. Meta-learning is related to transfer learning, with the additional advantage that it can extract general information pertaining a class of tasks. This is particularly useful when many of the tasks have a small amount of data associated.

Conventional ML approaches aim at training a single classifier from the data from multiple patients, trying to minimize the prediction error averaged over the whole set of patients represented in the dataset. This generally produces one-size-fits-all models that apply the same rules to all signals regardless of the patient. However, signal patterns may have different meanings if they are produced by different underlying clinical conditions. Consequently, patients with minoritarian clinical conditions may be affected by larger errors.

On the other hand, a meta-learning approach, formalizes signal classification for each patient as a separate task, removing the constraint of using a single classifier aimed at all patients. The output of a meta-learning algorithm is a learning entity that can use patient-specific training data to complete their training and improve the classification accuracy for each individual. An example of such an entity is a computer program consisting of: i) a deep neural network (DNN) with specific initial neural weights, and ii) a training algorithm. The specific initial weights are determined using data from all patients in the dataset using model-agnostic meta-learning (MAML) and the training is completed for each patient by running a few learning epochs with the patient-specific data. The result is a personalized classifier for each patient, leading to a higher overall accuracy. Alternative meta-learning approaches aim at automatically designing the training algorithm to maximize the capacity of learning from the data of a single patient.

5 Cardiac Signal Enhancement and Interpretation

The external ECG is traditionally obtained non-invasively by attaching a set of electrodes to the chest and limbs, although alternative wearable and wireless devices to acquire ECG have been recently developed [40]. On the other hand, intracavitary or intracardiac ECGs (referred to as electrograms (EGMs)) are obtained by setting one or more electrodes in direct contact with the inner surface of the heart; this technique is used for the identification of ventricular tachycardias and classification of other types of arrythmias. The processing of EGMs has the additional associated challenge of using a multivariate signal, and the exploitation of its spatio-temporal correlations is a currently active research field. However, currently existing algorithms lack the scalability required to process a large set of time series while limiting the model complexity.

A typical ECG SP pipeline is compound of several processing blocks, including 5 fundamental steps: filtering to remove noise and interferences, waveform delineation, feature extraction, ECG compression, and classification of the patient's status. In the context of ECG denoising, the spectral properties (frequencies) of the signal are the main features used to identify and separate the different components (QRS complex, P-wave, T-wave, noise, and artifacts). For instance, baseline wander and power line interference have strong structure that can be exploited to filter them out. On the other hand, electrode noise and electromyographic noise have frequential components that overlap the characteristic frequencies of some of the waveforms of interest [LOT2020]. The frequency overlap between some types of noise and the signals of interest calls for the development of nonlinear SP algorithms, which have not been analysed as much as linear time invariant (LTI) systems and are object of current research.

The properties of LTI filters allow further development into adaptive filtering algorithms. Adaptive filters are solutions where the filter's configuration (response coefficients) varies with time depending on the input data, which is particularly useful when the noise is nonstationary. LTI and adaptive filters have been proposed and compared for the removal of different types of noise and artifacts in cardiac signals. Their flexibility and applicability motivate further research in adaptive nonlinear filtering and extended Kalman filtering [41, 42]. Surprisingly, the benefits of hyper-parameter optimization and meta-learning in the context of univariate or multivariate adaptive filters have not been studied yet.

Denoising and classification of ECG can also be designed via regression-based methods. These can be further divided into parametric and nonparametric (fully data-driven) methods. Parametric methods based on wavelets express the signal as the addition of many instances of "basic waveforms" (called wavelets because of their short duration), and its main advantages are their flexibility and interpretability. In this basis, a complex and long ECG can be expressed by listing the location and shape of the wavelets in a sparse data structure. A few recent works (see [43] and references therein) have developed sparse regression techniques where each coefficient corresponds to a heartbeat, which enjoys high interpretability. However, such a method has only been tested on synthetic data and its adaptation to real signals requires its hyperparameters to be optimized.

The use of sparsity-aware SP appears also in the detection of atrial fibrillation, a pathology of high risk that is also a high-complexity electrical pattern in the heart [35].

6 Conclusion

Signal processing algorithms are designed to extract knowledge from biomedical information, based on rigorous mathematical formulations and with a solid practical motivation. The interpretability of the information structures produced by ML algorithms is valuable when they are applied to biomedical signals and patient records, and this motivates future research in sparsity-aware signal processing. Future research will aim at improving the efficiency and learning capability of the aforementioned SP techniques. Researchers and innovation-oriented business face the challenge of identifying specific biomedical practical problems where the newly developed schemes can bring a qualitative improvement. The resulting algorithms will be a key piece in medical equipment software and end-user eHealth applications, with the aim of making them more dynamic and easily reprogrammable. An important step in this process is to demonstrate the usability and reduced need for human intervention of the newly developed algorithms and product prototypes.

Acknowledgement: The author wants to thank Ayan Chatterjee, Martin Wulf Gerdes, David Luengo, Rune Fensli, and Baltasar Beferull-Lozano for fruitful discussions during the preparation of this survey.

References

1. K. Weintraub, "Giving medicine a dose of ai," https://www.technologyreview.com/2019/04/24/135725/giving-medicine-a-dose-of-ai/, accessed: 2021-06-21.
2. K. Advisory, "Living in an ai world 2020 report: Healthcare insiders," https://advisory.kpmg.us/content/dam/advisory/en/pdfs/2020/healthcare-living-in-an-ai-world.pdf, accessed: 2021-06-21.
3. S. Uckun, "Intelligent system in patient monitoring and therapy management," *International Journal of Clinical Monitoring and Computing*, vol. 11, no. 4, p. 241, 1994.
4. M. Chen and M. Decary, "Artificial intelligence in healthcare: An essential guide for health leaders," in *Healthcare management forum*, vol. 33, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2020, pp. 10–18.
5. K. Slavakis, G. B. Giannakis, and G. Mateos, "Modeling and optimization for big data analytics:(statistical) learning tools for our era of data deluge," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 18–31, 2014.
6. J. M. Lillo-Castellano, I. Mora-Jiménez, R. Moreno-González, M. Montserrat-García-de Pablo, A. García-Alberola, and J. L. Rojo-Álvarez, "Big-data analytics for arrhythmia classification using data compression and kernel methods," in *2015 Computing in Cardiology Conference (CinC)*. IEEE, 2015, pp. 661–664.
7. W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health information science and systems*, vol. 2, no. 1, pp. 1–10, 2014.
8. V. Edupuganti, M. Mardani, S. Vasanawala, and J. Pauly, "Uncertainty quantification in deep mri reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 239–250, 2020.
9. F. Yang, J. Elmer, and V. I. Zadorozhny, "Smartprognosis: Automatic ensemble classification for quantitative eeg analysis in patients resuscitated from cardiac arrest," *Knowledge-Based Systems*, vol. 212, p. 106579, 2021.
10. H.-B. Xie, P. Zhou, T. Guo, B. Sivakumar, X. Zhang, and S. Dokos, "Multi-scale two-directional two-dimensional principal component analysis and its application to high-dimensional biomedical signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1416–1425, 2015.
11. G. Cutura, B. Li, A. Swami, and S. Segarra, "Deep demixing: Reconstructing the evolution of epidemics using graph neural networks," *arXiv preprint arXiv:2011.09583*, 2020.
12. A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

13. H.-H. Chang and J. M. Moura, “Biomedical signal processing,” *Biomedical engineering and design handbook*, vol. 2, pp. 559–579, 2010.
14. R. Shiavi, *Introduction to applied statistical signal analysis: Guide to biomedical and electrical engineering applications.* Elsevier, 2010.
15. Q. T. Nguyen, “Contributions to statistical signal processing with applications in biomedical engineering,” Ph.D. dissertation, Télécom Bretagne, Université de Bretagne Occidentale, 2012.
16. S. Yazdanfar, C. Yang, M. V. Sarunic, and J. A. Izatt, “Frequency estimation precision in doppler optical coherence tomography using the cramer-rao lower bound,” *Optics express*, vol. 13, no. 2, pp. 410–416, 2005.
17. W. Luo, D. Phung, T. Tran, S. Gupta, S. Rana, C. Karmakar, A. Shilton, J. Yearwood, N. Dimitrova, T. B. Ho *et al.*, “Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view,” *Journal of medical Internet research*, vol. 18, no. 12, p. e323, 2016.
18. E. Vittinghoff, D. V. Glidden, S. C. Shiboski, and C. E. McCulloch, *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models.* Springer, 2006.
19. R. Kumar, R. Srivastava, and S. Srivastava, “Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features,” *Journal of medical engineering*, vol. 2015, 2015.
20. E. Tjoa and C. Guan, “A survey on explainable artificial intelligence (xai): Toward medical xai,” *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
21. M. Javanmardi and T. Tasdizen, “Domain adaptation for biomedical image segmentation using adversarial training,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* IEEE, 2018, pp. 554–558.
22. H.-Y. Tsao, P.-Y. Chan, and E. C.-Y. Su, “Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms,” *BMC bioinformatics*, vol. 19, no. 9, pp. 111–121, 2018.
23. G. Valdes, J. M. Luna, E. Eaton, C. B. Simone, L. H. Ungar, and T. D. Solberg, “Mediboost: a patient stratification tool for interpretable decision making in the era of precision medicine,” *Scientific reports*, vol. 6, no. 1, pp. 1–8, 2016.
24. C. Chandrakar and M. Kowar, “Denoising ecg signals using adaptive filter algorithm,” *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 1, pp. 120–123, 2012.
25. J. Lillo-Castellano, I. Mora-Jimenez, R. Santiago-Mozos, F. Chavarría-Asso, A. Cano-González, A. García-Alberola, and J. L. Rojo-Álvarez, “Symmetrical compression distance for arrhythmia discrimination in cloud-based big-data services,” *IEEE journal of biomedical and health informatics*, vol. 19, no. 4, pp. 1253–1263, 2015.
26. M. Blanco-Velasco, F. Cruz-Roldán, F. López-Ferreras, A. Bravo-Santos, and D. Martinez-Munoz, “A low computational complexity algorithm for ecg signal compression,” *Medical engineering & physics*, vol. 26, no. 7, pp. 553–568, 2004.
27. R. Bellazzi, “Big data and biomedical informatics: a challenging opportunity,” *Yearbook of medical informatics*, vol. 9, no. 1, p. 8, 2014.
28. Z.-Q. Luo and W. Yu, “An introduction to convex optimization for communications and signal processing,” *IEEE Journal on selected areas in communications*, vol. 24, no. 8, pp. 1426–1438, 2006.
29. F. Guo and G. Qian, “Sample-efficiency-optimized auxiliary particle filter,” in *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005.* IEEE, 2005, pp. 393–398.

30. S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.
31. L. Franceschi, P. Frasconi, S. Salzo, R. Grazzi, and M. Pontil, "Bilevel programming for hyperparameter optimization and meta-learning," in *International Conference on Machine Learning.* PMLR, 2018, pp. 1568–1577.
32. C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning.* PMLR, 2017, pp. 1126–1135.
33. A. Wijekoon and N. Wiratunga, "Personalised meta-learning for human activity recognition with few-data," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence.* Springer, 2020, pp. 79–93.
34. X. S. Zhang, F. Tang, H. H. Dodge, J. Zhou, and F. Wang, "Metapred: Meta-learning for clinical risk prediction with limited patient electronic health records," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2487–2495.
35. S. Monzón, T. Trigano, D. Luengo, and A. Artes-Rodriguez, "Sparse spectral analysis of atrial fibrillation electrograms," in *2012 IEEE International Workshop on Machine Learning for Signal Processing.* IEEE, 2012, pp. 1–6.
36. J. Friedman, T. Hastie, R. Tibshirani *et al.*, *The elements of statistical learning.* Springer series in statistics New York, 2001, vol. 1, no. 10.
37. T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the lasso and generalizations.* Chapman and Hall/CRC, 2019.
38. F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning: methods, systems, challenges.* Springer Nature, 2019.
39. L. M. Lopez-Ramos and B. Beferull-Lozano, "Online hyperparameter search interleaved with proximal parameter updates," in *2020 28th European Signal Processing Conference (EUSIPCO).* IEEE, 2021, pp. 2085–2089.
40. S. Majumder, L. Chen, O. Marinov, C.-H. Chen, T. Mondal, and M. J. Deen, "Noncontact wearable wireless ecg systems for long-term monitoring," *IEEE reviews in biomedical engineering*, vol. 11, pp. 306–321, 2018.
41. A. K. Ziarani and A. Konrad, "A nonlinear adaptive method of elimination of power line interference in ecg signals," *IEEE transactions on biomedical engineering*, vol. 49, no. 6, pp. 540–547, 2002.
42. M. Mneimneh, E. Yaz, M. Johnson, and R. Povinelli, "An adaptive kalman filter for removing baseline wandering in ecg signals," in *2006 Computers in Cardiology.* IEEE, 2006, pp. 253–256.
43. D. Luengo, J. Vía, and T. Trigano, "Efficient iteratively reweighted lasso algorithm for cross-products penalized sparse solutions," in *2020 28th European Signal Processing Conference (EUSIPCO).* IEEE, 2021, pp. 2045–2049.