Multi-Modal Clinical Data Harmonisation and Feature Extraction: AI-Enhanced Systems for Automated Biomedical Data Analysis and Interpretation

Dr. Christos Papachristou, Professor of Electrical and Computer Engineering, University of Cyprus

1. Introduction to AI-Enhanced Systems in Biomedical Data Analysis

Biomedical data are increasing in both size and complexity; however, the capabilities of biomedical data analysis using conventional methods are limited. Artificial intelligence has reshaped the process of biomedical data analysis. In this essay, we present the advanced AI techniques and elaborate on AI-based systems in the field of biomedical and health informatics, specifically focusing on four production points of biomedical data analysis: pre-analytical procedures and diagnostic workflows, data breaches, interpretative report editing, and multi-dimensional data integration. By incorporating data preprocessing, quality control, and knowledge acquisition into the computational analytics, automating biomedical data analysis could improve accuracy and shorten the time required for a diagnosis through the optimized examination frequency. A shortened timeframe would increase research and diagnostic efficiency for all patients and would result in an extended and more relevant lifespan.

Technological advancements in various dimensions, such as data acquisition, data processing, data analysis, and final decision-making, have made significant contributions to work processes and performance during the rapid expansions of the healthcare and information technology sectors. Innovations in the medical and IT sectors are a major driving force behind the improved quality of life for an increasing number of people worldwide. Today, AI and machine learning systems may help scientists and physicians overcome challenges associated with integrating and analyzing complex datasets, ultimately influencing the composition of big data and healthcare. Here, we present AI-based systems for biomedical data analysis, focusing on two domains: AI-integrated systems for data analysis and AI systems for data analysis. This literature review may provide a basis for future studies and can be utilized by researchers in several related fields. Additionally, the use of AI-based systems is anticipated to eliminate obstacles in integrating and interpreting big biomedical data.

1.1. Overview of Biomedical Data Analysis

Analyzing and gaining insights from biomedical data, a process known as biomedical data analysis, is indispensable for decision-making in most aspects of healthcare research. Biomedical data can be categorized into various types, such as clinical records, molecular data, functional data, imaging results, and phenotypic data. This causes heterogeneity challenges for data analysis due to the different data types. For example, image results may be combined with clinical or molecular data analysis to gain more robust decision-making. However, clinical and molecular data are not common forms of data for analysis using traditional image analysis. Analyzing this biomedical data is a very complex process since many parameters play a role in the process. Moreover, manually processing and analyzing seems not feasible since the volumes of data are tremendously large. For example, single-cell RNA data can contain around one million samples.

Analyzing large volumes of clinical records requires a considerable amount of time. In fact, researchers may become overwhelmed during the manual processing. To address these challenges, it is essential to develop various computational methods and techniques for automated data analyses. The process consists of four stages: data collection; data preprocessing; data interpretation; and using these collected, pre-processed, and interpreted data for decision-making. Scientists have long been devoted to creating a computational tool to possess the capacity for the entire process. These tools use artificial intelligence as a reliable and accurate method to drive reliable biomedical decision-making and transform the whole healthcare system. Voluntary data-driven decision-making powered by AI-based biomedical data analytical tools can support a rapid cessation of costly medical errors, saving substantial costs.

1.2. Role of AI and Machine Learning in Data Automation

Machine learning algorithms have been leveraged to improve the speed and accuracy of biomedical data analysis. Such algorithms can analyze large amounts of data and identify hidden patterns and correlations among data points that may escape human analysis and even traditional statistical tests. As such, a natural use of these systems has been to automate the process of identifying key phenotypic or molecular characteristics within large datasets. One of the more commonly cited uses of these systems is to identify molecular subtypes of disease from large, well-annotated clinical and omics datasets. However, the strength of machine learning systems goes much further than choosing the appropriate mathematical model for predicting an existing classification. When trained appropriately, machine learning algorithms can use some biological data and make discoveries that were not readily apparent using traditional methods.

The strength of automated data processing is threefold: (1) automated data processing is often faster than manual data processing; (2) automated data processing, when model feature weights are available, reduces the possible introduction of human error from the analysis pipeline; and (3) the model and feature weights of the AI system can be used to single out key features and relationships in the data for further follow-up. Continuous, AI-driven advances in the biomedical field are possible through the integration of machine learning into existing data analysis frameworks. As with standardizing the research process, automated data processing facilitates more rapid system learning and breakthroughs in broader models of biological function and disease. Scientists can use such computerized methods as a framework for continuous discovery through the collection of new and expanded data. One consequence of automating data evaluation is the creation of a quicker, more accurate decision-making process. The ability to fine-tune AI and machine learning to accurately evaluate data offers numerous benefits. Long-term drug reactions can be observed, proteins associated with specific types of cancer can be isolated, and procedures involving risk factors may be better understood. This ensures that the accuracy of these procedures is enhanced, which ultimately alters the scope and productivity of research.

2. Challenges in Biomedical Data Analysis and the Need for Automation

Data in biomedical fields are being captured at an ever-increasing rate from a wide range of biological processes using various methodologies and are being shared across different platforms around the world. The diverse nature and different formats of the data, as well as the way data is structured and represented within various entities, are making data-driven analysis more complex. The process of transforming meaningful insights from big data and diverse data in this complex ecosystem is done through traditional hypothesis and/or knowledge-driven methods, which mostly involve a lot of time, experience, resources, manpower, and close supervision, or tedious and time-consuming curation of data prior to the analysis phase.

This is a combinatorial amount of work with some artificial intelligence tools to move forward into an analytical phase. A lack of standards for the data generated by an experiment to follow, the experimental details in the metadata of the file, and the manner in which observational data is documented pose barriers to creating reproducible and valid data analytics. Overall, these issues lead to data integration inefficiencies, inconsistency, and sometimes human errors, as well as the consumption of a large amount of resources in transforming content, validating data, and turning these ideas into structured elements for analytic purposes.

This limitation calls for a rethinking of traditional bench and in silico practices towards the use of AI for enabling rapid assumption-free exploration of known connections in biological research.

2.1. Complexity and Volume of Biomedical Data

1 Introduction

Research in the biomedical field benefits from knowledge discovered or validated automatically from heterogeneous complex datasets. A biomedical dataset can be highly structured, for instance, a dataset describing how metabolites are connected within a metabolic network; semi-structured, such as a diabetes dataset containing tabular or CSV format data; or unstructured, such as different "-omics" data, medical imaging, and pathology report data. Besides, a complex dataset can be multi-modal, consisting of image, tabular, and text or unstructured data. From the analytical perspective, each data type or representation of the dataset will have a higher dimensionality. For example, the feature representation of genomic data can have a very high dimensionality of the order of tens of thousands. One of the grand challenges of biomedical research, to an extent more complex than many other scientific disciplines, is the integration and analysis of data from different fields, like -omics data, proteomics, and metabolomics, together in addition to the clinical data. In genomics, the databases are being filled at an annotative geometry that has grown to more than 20,000 billion nucleotide base pairs in the last five years. The rapid growth of molecular and genetic data is the result of international genome research collaboration. A biologist examines only a small part of the molecular and genetic role of living organisms for applications in agriculture, medicine, biotechnology, and so on. Therefore, a biologist is overwhelmed with the amount of data of different kinds. Thus, there is a need for scalable knowledge discovery systems. The genome research holds the promise of helping to discover cures for a host of common and rare diseases. But first, scientists must sift through the three billion bits of information encoded in the code of human DNA. In some ten years, DNA-trace evidence will be a routine and admissible investigative tool. Cognitive and AI researchers trained in machine learning techniques are beginning to use and improve upon techniques in applied fields to progress scientific projects, which is autonomic experimentation leading towards autogeny.

2.2. Data Integration and Interoperability Issues

Using a single resource is rather restrictive, but little data are completely interoperable for several reasons that make the integration of data a significant issue in the biomedical domain. First, sources of biomedical data are quite heterogeneous and typically not designed to be used together. In this scenario, different kinds of data have to be used in combination to obtain a comprehensive view of, e.g., a patient. The information to take into account may come from electronic patient records, but also from all the available lab results, from the various aspects of the omics sciences, i.e., genomics, proteomics, metabolomics, and so on, providing information from the biggest to the smallest biomedical systems levels before considering the next system levels detailed by the most common analyses conducted by full radiomics and histological images.

The main problems in seamless data integration are due to two kinds of heterogeneity: on one hand, the heterogeneity due to the use of different formats and metadata in different databases (which requires conversions and mappings); on the other hand, the heterogeneity at the semantic level, due to the use of different terminologies and controlled values in different databases. The second heterogeneous level is particularly important because the data coming from different systems can then be much more easily converted if it is possible to employ simple tools for data format transformations and metadata mappings management. It is then increasingly important to use a number of common data-sharing practices, i.e., to employ standards, provide access to metadata, make available good-quality technical documentation, profile systems able to discover, access, and use available data, and provide mechanisms for secure and auditable data handling. The lack of the mentioned activities heavily affects both the quality of the research missions performed and the good continuity of real applications in the clinical sector. A first example of this kind might be retrieved from a European open-access project aiming to set the basis of a collective virtual professional network in the interventional radiology scenario by the design and experimental validation of new methods for fast and automatic dynamic 3D-2D angiographic image registration, virtual embolization, virtual aorta flow sensing, and endoleak treatment procedure simulation. In this project, a significant feature is represented by the involvement of some end users in the needs definition, both concerning the new methods and the final tool for medical usage. Some possible future applications are evaluated through a prototype design and simulated performance related to validation tests.

3. Machine Learning Techniques for Data Extraction

Biomedical data of any nature, be it research data or digital health data, contain important information that should be extracted for various purposes such as developing predictive models for different diseases, examining public health trends, guideline development, and many others. This data is available in unstructured and structured forms, and thus using machine learning techniques is necessary in order to extract, understand, and analyze such data. Knowing a patient’s condition, diagnosis, and potential outcomes is mandatory to ensure patient well-being by managing treatment options so that healthcare professionals can make informed decisions efficiently and effectively. Methods for data extraction can involve a variety of categories such as named entity recognition, relation extraction, and event extraction. The named entity recognition methods use classification techniques to identify different entities in a document that can help in further analysis and processing. Relation extraction is performed between entities to extract valid relations, although they are of little use as of now. Event extraction is employed to model different molecular events in a given sentence. The use of machine learning, and in particular, deep learning techniques can help to automate the data extraction process with good capability to boost accuracy. Data extraction can be done using different machine learning mechanisms such as supervised learning, unsupervised learning, and rule-based learning. Each of these cognitive approaches has diverse techniques, models, and algorithms with varying complexities and methodologies. Choosing the right machine learning technique for data extraction is determined by the data type, purpose, and the way it is used in cognitive research.

3.1. Supervised Learning Methods

Supervised learning is a fundamental form of machine learning that is used to predict numerical or categorical outcomes of unseen examples. Once it is built and trained on labeled data, the model can extract patterns from available data and consequently provide predictions, representation, categorization, or quantification for new or unseen cases. Consequently, supervised learning algorithms are widely used in biomedical analysis to extract clinical, pharmacogenomics, GWAS, or OMICs insights from large or high-dimensional datasets.

The extraction of information based on machine learning is the process of data representation and transformation where the input data are represented by the machine learning algorithms' output data model. This process is further divided into stages. The first stage is algorithm training (tuning), which depends on a set of comprehended features that are called training data. The second stage is the evaluation and validation of the output algorithm, which is described, then, by another set of data occurrences that are referred to as the testing set. As a consequence, it is possible to use the machine learning approach to build several types of algorithms from supervised and unsupervised learning approaches, used for key predictive modeling in large datasets. Supervised learning, in particular, can produce accurate quantification and categorization of extracted information.

There are several algorithms used in biomedical datasets to generate insights from health records, such as decision trees and random forests. There are also several algorithms optimized to extract insights from genomic data analysis, including regression methods, Naïve Bayes, and k-nearest neighbors. The published case studies significantly produced biomedical insights using supervised learning algorithms. In the biomedical field, there are different challenges that need to be addressed in the machine learning process, including missing and noisy data. A crucial issue is also the use of large amounts of representative datasets.

In biomedical profiling data analysis, an association rule was used to discover interaction relationships among Hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). The decision tree method was applied to identify risk factors associated with childhood obesity. Another study surveyed prospective cohort data of 11,118 European adults and analyzed BMI trajectories using generalized additive mixed models with P-splines and cubic B-splines. Epidemiologic analyses presented in this case study aim to fill that knowledge gap by assessing how large genetic datasets combined with risk factor data can be used to identify genetic predictors of insulin resistance and glucose dysregulation. Generally, the supervised approach is a cornerstone technique, suitable for carrying out automatic biomedical data analysis in combination either with structured health data or produced OMICs data. However, validated supervised learning-based algorithms for data profiling are relatively few. Decisions concerning the evaluation of individual genes must be replicated from additional studies and the available scientific literature.

Several supervised algorithm performance metrics and validation procedures must be used to avoid overfitting and the false generalization of implemented algorithms. Typical validated supervised-based methods for structured health data and OMICs data include decision trees, random forests, and Naïve Bayes. Further approaches used in biological health data profiling in general include (a) general algorithms such as regression methods, (b) clustering algorithms, and (c) bandit algorithms. Such performance evaluation in supervised learning will provide a valid performance assessment of the learning model.

3.2. Unsupervised Learning Methods

Another important trend in machine learning, called unsupervised learning, is rooted in extracting information from datasets not annotated with output labels. Instead of learning the relationship between input and output variables, unsupervised learning focuses on the underlying structure inside an unlabeled dataset. It can help identify possible patterns and structures in real-world data without any prior knowledge or guidance imposed on the model. Among its key algorithms, one can find representations like clustering, dimensionality reduction methods, and association rule learning and anomaly detection. In the healthcare domain, unsupervised learning applications help extract unknown groups of similar patterns in patient records with varied clinical features and identify possible abnormalities without giving exhaustive instructions on how to represent and learn the expected behavior in the hospital database. For example, unsupervised techniques can identify clusters of comorbidities or disease trends or the hidden classes people belong to. Moreover, one can utilize association rules to discover, demonstrate, and give interpretations of strong associations like treatment effectiveness and other embedded patterns in clinical databases. In general, unsupervised learning approaches turn out to be profound in knowledge acquisition from large and complex datasets in use cases with no labeled training examples. The broad use of unsupervised learning approaches in the data mining community is demonstrated by their utilization in a wide variety of application fields. The unsupervised techniques are particularly useful when dealing with complex data with multimodal properties or lacking labeled training examples for calibration of the model. All in all, unsupervised learning methods are thus a valuable complement to supervised learning approaches, which can enhance the process of biomedical data understanding or extraction.

4. Enhancing Data Integration through AI

Most biomedical research today is assisted by advanced computational techniques; however, substantial effort is exerted to collect data from diverse sources such as clinical records, molecular biological assays, functional genomics, imaging, and so on. The seamless integration of this data is key in maximizing the capabilities of biomedical analytics to develop novel insights and, in a wider context, to advance our understanding of the complex web of interactions coded within cellular, tissue, and organ systems. Ontologies and semantic web technology are instrumental for representing data and relationships in an unambiguous manner. By employing these technologies, it becomes possible to resolve the conceptual mismatch between integrated data sources, facilitating their interoperability and adopting original procedures and tools for further data exploration and scientific conclusions. Several strategies can be employed in integrating data in the biomedical research domain, including semantic mediation, ontology-based data integration, integration of ontologies, entity resolution, dataset linkage, and others. Data integration approaches generally lead to more reliable and diverse insights when analyzing the properties of data types and connecting their respective attributes or properties. AI is involved in performing tasks constituting the core of data integration, such as enabling easier data representation, acting as a substitute for logical mediation, and so forth. By developing rules and patterns necessary to enable such data integration, AI frees human experts from the laborious process of assigning meaning to such relationships across disparate data sources with varying schemas. In doing so, AI potentially can help predict and anticipate effects and, in turn, suggest novel approaches to healthcare or in discovering new biology. Moreover, computer-based AI systems can quickly analyze large numbers of potential options, allowing entire ecosystems of hypotheses to be proposed and tested. Importantly, data analysis is enhanced as well by assessing the validity and reliability of the biomedical data being reviewed, thereby adopting the necessary preprocessing steps needed to develop robust insights. Therefore, artificial intelligence offers a way for data analysis to be more reliable. Hence, in the future, these capabilities will represent the basic infrastructure for systems that provide biomedical research with both novel tools to reveal previously unappreciated science and clinical diagnostics.

4.1. Ontologies and Semantic Web Technologies

Ontologies and semantic web technologies play a pivotal role in data coming from a variety of sources belonging to the same domain, allowing for easy integration and interoperability. An ontology defines a set of terms used to provide a structured framework for capturing and representing knowledge, supporting consistent interpretation of inferences and reasoning about that body of knowledge. In the biomedical field, ontologies are typically used to structure domain-specific knowledge, consisting of terminologies and the relationships between terms, and are usually available in a standard format consumable by both humans and computers. In this regard, the utilization of semantic web technologies is particularly advantageous; to work with knowledge present on the web, they rely on a shared format across: 1. Various ontologies and the relationships between terms to interpret domain knowledge. This provides a means to organize and distribute domain knowledge. 2. With a shared format, aligned terminologies enable the easy integration of knowledge present in diverse databases originating from various entities such as institutions and companies, effectively using Semantic Terminological Resources. Biomedical data can then be queried, combined, visualized, and shared via different databases that understand and exploit such semantics. 3. Using the above points, workflows can be enriched. Use cases depicting how the implementation of ontologies and terminologies based on semantic web technologies have contributed to the elimination of naive Bayes bottlenecks, harmonized NLP-based natural language, supported conversational agents, facilitated data retrieval based on standards, and helped visualize patient data are also available. The creation of semantic web technologies, whose primary use is deriving insights from data in a timely fashion to improve outcomes and utilize the right resources in terms of personnel and infrastructure, continues. While ontology development progresses with the exponential growth in biomedical knowledge and ongoing research, changes to terminologies must be maintained and constantly updated. To address such limitations, controlled vocabulary support implementation is essential, emphasizing harmonization and standardization in data collection, processing, sharing, and accessibility pathways among stakeholders. Several groups have developed multiple ontologies for varied domains along with data-determined models to specifically manage prioritized research based on stable resources.

5. Interpreting Biomedical Data with Machine Learning

A promising application of machine learning (ML) and artificial intelligence (AI) techniques has been in the domain of biomedical data analysis. Biomedical data is traditionally complex, arising from a wide variety of sources including clinical trials, patient records, and genomic studies across a range of species and platforms. Pattern detection in this data is the key to gaining insight, whether in patient stratification, determining a specific effect of drug dosage in a trial, or establishing gene functions from a large-scale genetic screen. Applications that fall into the ML category and that are of particular interest in biomedical research include, but are not limited to, supervised and unsupervised clustering methods for pattern detection in data, where predictive or explanatory models are based on some given function, and classification tools allow efficacy or outcomes of drug treatments to be predicted. Predictive analytics, in the early stages of drug discovery or precision medicine, has received significant attention due to the potential for improved drug efficacy predictions, increased likelihood of further drug development, and risk reduction in clinical trials.

Several successful case studies may be found in predictive biomarker discovery or correlating this with clinical outcomes: machine learning based on gene expression data enabled the discovery of early prognostic biomarkers for osteosarcoma. In another study leveraging large databases of patient records covering a range of different diseases, unsupervised and supervised techniques combined with text mining to stratify patients and determine disease outcomes. An unmet need was identified as a classification problem to identify response to treatment: using a random forest model, a 71% area under the curve (AUC) was achieved in predicting resistance to a certain type of treatment. An emerging challenge is that of explaining predictive models in machine learning, particularly how to interpret models and provide explanations in the context of an ethical and trust framework. An interpretable approach for drug sensitivity prediction derived a biologically plausible pathway from gene expression data to potential drug targets for personalized treatment. Importantly, the gold standard for an interpretable model, beyond publishable validation statistics, is the transition from predicting population-level measurements to in vivo predictions with clear biological insights. Given the high quality and depth of biological understanding, performing these analyses in vivo and correlating predictive biomarkers with, for example, patient response, is critical. Given that most of the current successes in biomedical applications of machine learning have been in the model development pipeline, it is evident that to produce more biologically interpretable models, data scientists must collaborate with domain experts such as medical and biological researchers.

5.1. Predictive Analytics in Drug Research

One key application of predictive analytics in the long journey of bringing a new drug to market is drug research. Drugs, like other commercial products, have passed their most expensive production phase; that is, the research required for their discovery and development. Based on a number of deductive and speculative predictions about drug action and effect, institutes and pharmaceutical companies direct research to relevant treatment facilities. The gulf in market value extends to a loss of hundreds of billions of dollars, and current products and processes are plagued by a propensity to resist standards. The synthesis of metazoan-like cells of human origin, also called biological systems, as experimental equipment, represents an essential current industry reaction. Such biotechnological representations are component-based, and enzymes or receptors for environmental signals are reintegrated, supported by high-throughput screening systems in each of thousands of solutions, to mimic realistic physiological activation. Nevertheless, the outcomes have been primarily targeted at drug visual and drug safety, as suggested by pharmacological results.

Early in the 19th century, pharmacologists began to explore how drugs operate physiologically to see how they can be fine-tuned by personal norms. In the past 30 years, significant progress in data mining and exchange has been achieved, and the potential to establish predictive analyses of very large data sets that can help decision-making is becoming a reality. Modern computers and machine learning techniques have the capacity to analyze big data to derive information. There has also been a significant increase in private investment in these systems. It is believed that in related sectors, such innovations will soon be the norm. For the concept of healthcare, such organizations promise individualized and preventive approaches to disease. Potential side effects, utilities, and so on are some of the uses that have emerged. Experimental and clinical data can be used to study molecular interactions and their association with illness. Based on molecular interactions of disease, one can determine how new medications are probably going to be effective. Regulatory agencies are starting to acknowledge these new developments and aim to evolve them.

6. Future Direction

The advancements in machine learning, genomics, and biotechnology are trends that will shape the future of AI-enhanced systems for biomedical data analytics. In particular, we foresee further developments and improvements in machine learning techniques that enable them to handle large-scale, high-dimensional, feature-rich biomedical data. As a system matures, the sophistication of tools it employs evolves from investigating relationships based on a single data source to combinations of data sources, to differing scales and depths of big data analytics. For example, single-cell RNA sequencing analysis is gaining prominence as the ultimate way to describe cellular states within a tissue and can be integrated with bulk transcriptomic data. AI can significantly enhance the field with potentially enormous computational capacity to support big data analytics in understanding the entire path of information flow from DNA to RNA to protein. When used in a clinical setting, integrated with a platform, it could enhance understanding of the mechanisms of disease and provide better predictive modeling about treatments, especially the emergence and magnitude of adverse responses.

The rise of digital and smart healthcare brings a lot of potential data that drive AI platforms. Research at this stage is to integrate comprehensive information in an operating manner to support the development of personalized medicine. In addition, transparency and ethical considerations are rapidly becoming prominent issues in the use of AI in healthcare. AI systems are becoming increasingly more generalizable as new and different data are used to train them. The production of data changes as technology changes, and the analytical tools also improve. In real-time analysis, real-time data are important in the decision-making on patient care. The results of data analytics will affect all healthcare professionals and patients because clinical choices are captured from the data. Public opinion is important because a national healthcare payer will be mandated for the process. Regulatory agencies may cave in the face of public disapproval. Ethical considerations and human satisfaction are also important. Models are always numbers such as weights, averages, power, and p-values. They are not people. Knowing the importance of the data and its analysis by physicians would be mandatory in real-time applications. Policymakers and healthcare professionals are invited to examine analyses to determine if the methods are both conducive and preferred. In summary, the data AI-enhanced system in the future will include AI capacity to further focus on the depth and heterogeneous features within in vivo and in vitro elements. It will improve the scope of the enhancement of the use of this data to provide comprehensive, ethical outcomes with the opportunity for iterative improvements. Continuous learning and adaptation of new advancements in biomedical data are important.

7. Conclusion

This review addresses the necessity of effectively managing complex biomedical data and advances in the implementation of automation in the field. Driven by an ever-growing dataset, advanced technologies are being used to reduce the gap of biomedical knowledge, focusing on analytical tools, machine learning methods, and novel computation methods. However, dataset complexity and scale hinder their human-sided analysis; thus, integrating intelligent systems into biomedical research is crucial. The need is that the generated solutions must have validity, credibility, and the highest possible confirmability. To achieve these four challenging targets, we need to generate proper tools and demonstrate the effectiveness of that novel advance.

The development of machine learning techniques and methods can, in principle, address the complexity of biological data. AI-enhanced systems developed from this research on laboratory tests were solely used for preclinical enrichment phases and substratification markers intended for patient selection. Moreover, the adoption of a multimodal approach in radiomics allows such signatures to be enhanced with other types of variables useful to validate pathological assessment, refine outcome prediction, and gain new pathophysiological insights. Future applications will be increasingly patientoriented and will aim to guide therapeutic strategies in a more precise and personal way.

To conclude, from its beginning in segmentation issues up to more consolidated predictive analytics to help physicians, the adoption of AI in clinical practice is very appealing, especially with the development of decision support systems. At present, any medical domain can take advantage of machine learning AI applications in medicine: from specialist communities to general practitioner level. However, it is extremely important that these tools are implemented in accordance with the ethical codes of medicine and according to clinical evidence. It is therefore desirable that AI is approved as an adaptive technology able to learn over time and not to replace the clinician. The vision is one where machine learning systems greatly enhance human decision-making and drive improved practice. In this way, biodesign and medicine in general can collect added value in representing the bridge for the dissemination of AI towards healthcare providers.