Artificial Intelligence Applied to Kidney Disease or The Challenge of Decision Support in Complex Patients*

Jose Ibeas[1,2]

[1]Nephrology Department. Parc Taulí Hospital Universitari. Universitat Autònoma de Barcelona. Sabadell, Spain
[2]Clinical, interventional and computational nephrology (CICN) research group. Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Spain

Abstract
Artificial intelligence has reached the health sector as one of those included in the Industry 4.0 revolution. Its capacity to generate a huge amount of data continuously and its possibility of being analyzed in real-time generates an ideal environment to be able to apply all the technologies related to Big Data analysis. This scenario has generated the beginning of a paradigm shift both in research and in the generation of evidence since it means going from talking about Evidence-Based Medicine to Data-Driven Medicine. This is inevitably associated with a series of new challenges: Data management, its new methods of analysis, the required training of professionals, the project roadmap, and some new scenarios in the area of ethics and regulation. The exponential proliferation of research publications in this area in recent years, however, does not go hand in hand with the number of tools approved by regulatory entities, which reflects that although this new environment is very promising, there is still a long way to go to standardize their use in clinical practice. In this context, the world of nephrology, that is, the environment of kidney diseases, which is characterized by a high complexity in the data generated by having different sources of origin, associated with significant complexity in decision-making, is one of the fields where it is precisely possible to observe the full potential of artificial intelligence as well as the associated challenges.

Keywords
Artificial Intelligence, Machine Learning, Nephrology, Dialysis, Kidney Disease, Big Data

1. Introduction

Like the previous industrial revolutions, the fourth, which is based on the arrival of artificial intelligence, Big Data and robotics, has brought a substantial change in all activities related to industrial processes. The health environment, although in a later way, has not been an exception. And it is in this field that this technology is probably going to have one of its most important repercussions in the immediate future. Multiple technological characteristics define this 4.0 revolution, but the main ones on which its usefulness in the health environment is based are the use of massive data, its potential use from a cloud environment and, above all, its implementation

Proceedings of Artificial Intelligence Governance Ethics and Law (AIGEL) Reviewed, Selected Papers, November 02 - December 19, 2022, Barcelona, Spain
jibeas@tauli.cat (J. Ibeas)
0000-0002-1292-7271 (J. Ibeas)

under the infrastructure of the Internet of things, in this case medical things (IoMT). It is precisely in healthcare environments where the greatest volume of data is being generated today. The enormous generation of information at the expense of analytical data, images, text, etc. of millions of patients makes it one of the environments that has seen the progression of terminology related to the definition of data volume grow more closely. In a context in which Industry 4.0 is based on the generation of algorithms that extract information from data in search of patterns and that are capable of learning on their own, in the health environment it can translate into enormous potential in all the spheres that go from the diagnostic processes to the evaluation of the response to the treatments.

The objective of this review is to explain what the arrival of artificial intelligence means in the health environment and specifically in one of its most complex models such as kidney disease. It will go through 4 main points. The first is from evidence-based medicine to data-based medicine; the second is the challenges of artificial intelligence in health; and the third place, talks about the use of algorithm modelling and knowledge extraction for predictive models, finally, after having reviewed these 3 main points, put nephrology as an example of how artificial intelligence can contribute to complex patient management.

2. From Evidence-Based Medicine to Data-Driven Medicine

Evidence-based medicine consists of 3 fundamental aspects. The first is clinical experience, the second is having the best evidence, and the third, and no less important, are the patient's values. From the methodological point of view, the architecture of evidence-based medicine (EBM) is based on the scaling of different types of studies, from the least robust to the one that has the most translation into the recommendation in clinical care practice from the clinical point of view. The most basic would be based on the opinion of experts, to then progress through the case series, studies, control cases, and cohort studies and then to reach those with the most robust generation of evidence, such as clinical trials or especially systematic reviews or meta-analyses [1].

Thus, the EBM is based on the systematic review of published evidence based on structured clinical questions. These questions, called PICO (problem, intervention, comparator, outcome), generate the systematic search for evidence in the literature in a methodologically standardized way that allows, step by step, to search for what the research has generated and that can be translated into clinical recommendations. The problem that this system has is that even though methodologically many studies may appear to have a very adequate and even very robust level, the methodology they have used may be biased and this means that the real quality of this evidence may not be good enough to translate into reliable clinical recommendations. In other words, as important as the research methodology is the quality with which it has been done, which translates into the level of confidence. The second important aspect is that, once the evidence is generated, it is with this that recommendations must be generated and this inevitably depends on the experts. This would be the second aspect that could generate another source of bias. That is the subjectivity of the recommendations derived from the interpretation of the evidence by the expert. An example of this is the fact that there are clinical guidelines on the same subject made with the same literature and by experts who may participate in several of these guides and which in the end generate different recommendations. The conclusion of all this is that EBM, in short, depends on the evidence itself reflected in the literature in the first place, on clinical experience in the second place, which is subject to knowledge that can be derived from different specialities, geographical or even economic environments. and, finally, the values and circumstances of the patient. And that's how EBM works. It is the best that the clinic has today to make decisions, but, as can be seen, it is a situation not free of biases.

In a context in which we are talking at EBM about a pyramid in which the lowest quality scientific studies appear from the bottom and end at the top of the pyramid with the most robust ones, in a parallel reality the pyramid based on DDM begins to be drawn. A pyramid in which the base has all the data sources and as it scales can be found data preparation, data transformation, data mining, knowledge extraction, and finally decision making. We are therefore talking about a paradigm shift. In other words, we find that while on the one hand, we are working in the usual environment of EBM, on the other, the architecture of data-based evidence is beginning to take shape.

If we compare the aspects that support EBM with DDM, the aspects that we would have to highlight would be concerning the first one: it represents the basis of the usual clinical practice of medicine that today we still understand as modern, it is based on observations that come from studies population clinical studies, the gold standard is the double-blind controlled prospective study, the studies are based on a hypothesis that is generated by a group of experts and from here the experiment is tested, a sample of patients is selected, subjected to the test, the evolution is observed, and finally an expert or a group of experts postulates a hypothesis or a conclusion. Quite obviously, each of these steps may be associated with a cognitive bias. On the other hand, if we are talking about the DDM and analysing with artificial intelligence, the points through which it would go through in a basic way would be the use of already existing data and not arbitrarily selected to generate algorithms with machine learning and that the accuracy and usefulness generated of these algorithms would be fundamentally based on the data that have motivated the learning. There is therefore no particular bias. Patterns can be discovered in this data that can lead to diagnoses or predictions. All this can be done in real-time, and in the end, precise and personalized recommendations are generated, which do not come from population studies derived from arbitrary selections of a determined population. In summary, the potential biases of EBM can be overcome by DDM, since all available data can be analyzed to generate knowledge in an observer-independent way [2].

3. The Challenges of Artificial Intelligence in Health

The challenges of artificial intelligence in health come first of all from the possibility of incorporating massive data from different sources simultaneously to generate knowledge. This data can be structured data such as analytics and numeric variables or unstructured data such as image-based and free text. All of this data comes from multiple patients in different conditions. These conditions can be contextualized from different perspectives associated with different types of professionals. These patients can be population grouped associated with hospital centres, health networks, regions or even countries. All this data may be hosted on specific servers in health centres or, on the other hand, in the cloud. And in turn, all these data networks can be connected. What does this mean? The translation of all this is that we are talking about massive data that can be analyzed with Big Data analytics technology [3].

The fundamental aspects of Big Data analytics are, on the one hand, the characteristics of Big Data and, on the other, its analysis. Big Data is traditionally defined by its ”Vs”, which can reach up to 9. But we will talk about the 5 fundamental ”Vs”. These are volume, variety, velocity, variability, and value [4]. These characteristics are what define Big Data and, as we can see, health data sources meet these characteristics. On the other hand, when we talk about Big Data analytics, we can talk about analytics 1.0, 2.0 or 3.0 [5] [6]. Analytics 1.0, also called Business Intelligence, was the classic work method in the 90s. It was characterized by low speed, static data, structured data, and results formulated in reports. Analytics 2.0, also called Big Data Analytics, was born in the 2000s and already incorporated large datasets, was capable of working in real-time, incorporating unstructured data, and was capable of generating predictive models. But it is with the analytics 3.0 that we currently have, with which the possibility of combining conventional Business Intelligence, with big data and above all with the internet of things has arrived. This allows working with artificial intelligence models in real-time at the exact point where we will have to make the decision and even with distributed computing, that is, in the cloud.

All of this has generated the infrastructure of what we call the Process of Knowledge Discovering. This is defined by some fundamental points: the integration of the data from the sources that can be very diverse and that generate a dataset for later, once the data is selected after established criteria, perform the preprocessing, transformation, and data mining to be able to generate the algorithms that will allow the generation of knowledge once have interpreted, evaluated and validated these algorithms.

But how does machine learning work? To explain it in a way that is easy to understand, if we start from the basis that traditional research systems based on classical statistics depend on software that incorporates data to generate a result, unlike this, what machine learning does is incorporate data of which the output is known to generate a program based on an algorithm. This program is capable of learning from the data continuously in such a way that the following data that comes from a patient will allow the algorithm to generate the outcome of this patient. What we have just defined is what we call supervised systems because the data we enter is “labelled” so the algorithm is capable of generating a learning system aimed at recognizing these outcomes. But on the other hand, machine learning is also capable of working without labelled data. That is, we can incorporate data without any type of label, in such a way that the algorithms look for patterns among the data that are capable of generating knowledge. This is what we call unsupervised systems. In between these systems, there are other systems with characteristics of both that we call reinforcement [7] [8].

What we are talking about when we want to put it into practice, therefore, requires an organizational system in different phases that allow its applicability. It is noteworthy here that, unlike traditional research systems, what we are talking about now is innovation. And innovation intrinsically has characteristics associated with a value chain that begins with the research itself but ends with the generation of a sanitary product, in this case, a decision support system, which, in addition to being implemented in clinical practice, generates value. And the channeling of this value uses the usual language of innovation, and this is the one associated with entrepreneurship or business models. Therefore, these phases would first comprise the strategic roadmap, the business model, the necessary resources, the definition of the roles and responsibilities of those involved, the data architecture, the choice of appropriate tools and technologies, governance of data and standards, regulation, executive support, the organizational system, the analytical process and generation of reports, to finally generate an interpretation that allows decision-making [9].

In this context, a fundamental point that should not be forgotten is the ethical aspects. What do we mean when we talk about ethics in data science? This is a branch of ethics that deals with how to handle data in a private environment associated with decision-making. There are 3 types of ethics: data ethics, that is, those related to its generation, collection, use, ownership, security and transfer. The ethics of intelligence, that is, the product or results of the model generated by the data that will be used for decision-making. And lastly, the ethics of the practice, that is the morality of the innovation systems can generate concerns about how to handle these new tools [5].

There is a key aspect, which is privacy and security. It is probably one of the fundamental points of data management when we talk about artificial intelligence. If this aspect is already important with the usual research activity, let's imagine when we talk about big data. They are the same aspects but multiplied concerning the volume of data and therefore its associated risks when management is not adequate. This requires taking into account with special care everything related to Data Protection laws, security systems, encryption techniques, data access control systems and finally security systems associated with Big Data technology [10].

All of this gives rise to the generation of open questions from which concerns arise which, depending on how they are interpreted, may be challenges to overcome or risks to not forget to consider. These would be those associated with access to data, its security, its storage, or its transfer. We could summarize these challenges in four fundamental aspects: those related to the data and its processing, those related to the professionals who are going to work with this data, those related to the domain we are talking about, in this case, health and finally those related to the organizational system.

The first of the challenges, that is to say, the one related to the data and its processing, is associated with a series of points linked between them in which each one supposes an aspect that in itself is of great importance to take into account. These would include data acquisition, storage, interoperability, quality assessment, complexity, and security and privacy. The second challenge, associated with professionals, is a basic aspect that is often underestimated and this is so because it must be noted that today it is difficult to find talent in this area, and when talent is found it is difficult to retain it, and when we have it it is not easy to coordinate as it requires doing so in multidisciplinary teams that handle different types of experience and knowledge. The third aspect associated with the domain of health is highly relevant because health data is very complex. This complexity is derived from the origin of the data from very different sources. They are associated with different time windows because they require interpretation associated with contexts that interrelate areas of medicine that can be very different and, lastly, because their interpretation that requires a high level of experience. Finally, the organizational challenges would be derived from a series of circumstances that are intrinsic to the needs of the organization's management: the objectives to be defined in the artificial intelligence context require specific knowledge, the possible technological disparity, one must know how to choose the appropriate tools to generate the algorithms, that is, therefore have previous knowledge from a practical point of view, one must know how to deal with resistance in organizational systems to new technologies, one must assume costs that are usually not budgeted and finally when you have large sources of data, you must not forget that you must know how to share them to help generate knowledge in the scientific community [6] [11].

The summary of all this is that the profile of the professional approach and the management of health data is changing. This is explained because we are used to understanding that there are 3 groups of professionals who depend on 3 large areas of expertise: Computer Science, Mathematics and Statistics and finally being an expert in a specific subject. It is easy to understand that the combination between computer science and a subject matter expert leads to the generation of specific software, it is easy to understand that the relationship between traditional statistics and a subject matter expert gives rise to traditional research, and finally, we can understand that there is a new environment derived from computer Science related to mathematics and statistics, which is machine learning. But if we wanted to look for an expert with knowledge in computer science, mathematics and statistics, and experience in any subject, this is what is usually called a "unicorn". And if it is called a unicorn because it does not exist.. In other words, if we want to find an adequate relationship between the 3 different fields, there is no other option than to work in a multidisciplinary way. And this is possibly the most important challenge when facing the management of Big Data, artificial intelligence, and knowledge extraction in the health area [12].

4. Modelling algorithm optimization and Knowledge extraction from predictive models

When talking about the generation and optimization of algorithms to be able to extract knowledge and thus take it to clinical practice, what we are talking about is organizing all the previously mentioned steps from a structured point of view to implement them. This is therefore specifying the problem, preparing the data, choosing the appropriate learning method, applying it, evaluating the method and the results, optimizing it, and finally generating a report to draw conclusions and put them into practice.

There are four types of learning systems. These are derived from different ways of converting data into information. These would be the descriptive, diagnostic, predictive, and finally prescriptive methods. The latter are the ones that would make it possible to assess the response to a given intervention. From a technical point of view, any learning system must go through a series of steps in which both the data expert and the subject matter from which the data is generated must know its details. These would be the preparation of the dataset, its cleaning, the feature engineering, the model training and its validation. The different types of machine learning must be known because depending on the type of variables, objectives, or type of study, the project must be adapted accordingly. And finally, it must be clear to both the technical expert and the subject matter expert that the combination of knowledge of both in a specific subject is what will help to guarantee success in achieving the objectives [6].

When systematic searches are carried out in the bibliographic sources to seek experience in the generation of algorithms in health, it can be easily discovered that the increase in the projects published today is not linear but exponential, which translates that the revolution 4.0

has arrived in the health environment. Two more objective proofs of this are that investment in the artificial intelligence sector in health is increasing year after year by percentages that show that this environment is one of the most interesting for investors and that approvals by regulatory entities such as the FDA begin to increase progressively when a few years ago there was practically no authorized model [13].

5. Nephrology as an example of what AI can contribute to the complex patient

If you want to give an example in the health area of how we can go through each of the previously exposed steps and above all exemplify how artificial intelligence can benefit a complex health environment, this is the area of nephrology. Nephrology is the area that treats kidney diseases. Kidney diseases are characterized by requiring the analysis of data from very diverse sources to make decisions: analysis of laboratory data, images, data from dialysis machines, biopsies, special diagnostic or treatment techniques, complex treatments, the relationship with other specialties, etc. This creates a very favorable field to demonstrate the speed with which artificial intelligence can benefit areas of high complexity and difficult analysis. Only 5 years ago the first robust results began to be published, demonstrating how promising the use of artificial intelligence could be in prevalent pathologies for which a large number of data were available. An example of this was pathologies derived from cardiovascular risk or imaging such as diabetic retinopathy [14] [15]. The progressive proliferation of this type of work even led to the generation of systematic reviews that began to show that artificial intelligence was at least as powerful as medical analysis [16]. In parallel, databases began to appear in nephrology that allowed any user in the area to train models [17]. And so progressively until the last 2 years, papers on specific pathologies published by "traditional" experts began to appear, showing more than promising results [18]. And it is in these last 3 years that the proliferation of works both with analytical data, as well as with biopsies or even with genetics or variables from minority diseases began to generate the impression that any type of nephrological pathology could be susceptible to being implemented with this new technology [19]. A curious fact, for example, is research into arterial hypertension where the last 30 years have been characterized by the proliferation of clinical trials to generate new drugs. However, in the last 3 years, all the investment has been channeled into the study of new drugs to be replaced by genomic sequencing research in this area [20]. Point something unimaginable a few years before.

Renal disease is characterized because it can be either acute onset or chronic. In the latter, the possibility of associated complications of different kinds is very high. These can be associated with the appearance of anemia, or alteration of mineral metabolism, cardiovascular pathology, and especially mortality. Mortality is possibly one of the major problems in this disease since it can reach 15

If it is observed in this context that, on the one hand, traditional statistical methods, when they seek to generate predictive models of renal failure progression, the appearance of complications, or mortality, are very limited, and that, on the other hand, renal pathology is characterized by having massive data from different forts in different conditions, we are faced with a scenario in which, precisely because of its complexity, this pathology can facilitate the generation of knowledge with the application of algorithms based on artificial intelligence.

Aware of the complexity of kidney disease on the one hand and the challenge posed by the application of these new tools on the other, our group has spent the last few years generating various lines of research in this context. The two main areas to highlight would be predictive models of mortality on the one hand and progression of renal failure on the other. Using both hospital data and massive data from the Catalan health system, it has been possible to demonstrate very robust results that allow generating a high degree of confidence in this technology as a tool for short-term clinical use [21] [22]. Although, indeed, we must never forget that any type of generated algorithm always requires a validation process before being put into clinical practice. This is why there is a great disparity between the thousands of research-oriented articles published during the use of artificial people as opposed to the few projects that have obtained authorization from regulatory agencies to be implemented in clinical practice.

The translation of this scientific activity in recent years in the area of medicine to generate innovation at the expense of artificial intelligence must be recognized that although it is very promising, on the other hand, it has to face the reality of the health system. In other words, one must be aware that although industry 4.0 technology has arrived in medicine, and some groups are beginning to demonstrate its usefulness, it is necessary to adapt the health system and overcome organizational resistance, train health professionals, inform the end user who is the patient of this change process, and finally adapting all the fundamental points associated with the security and privacy, regulatory, and ethical part in a context that the system is still not used to.

6. Conclusions

We are facing a promising future at the expense of the paradigm shift from evidence-based medicine to data-based medicine that lies in the possibility of making personalized medicine on the one hand and avoiding the associated biases on the other. But great challenges have to be overcome. The first of these is knowing how to work in a multidisciplinary way between engineers and clinicians to be able to make these new dynamics efficient in a new ecosystem that requires knowing how to manage new resources with specific technical knowledge and the ability to handle this type of project. On the other hand, just as important as the potential usefulness of these tools is the need for their validation to be implemented in clinical practice. And finally, the responsibility for the success of any health product based on artificial intelligence will depend on the evaluation of its applicability and this lies in the arena of the health professional, so the health ecosystem has no other option but to enter the match and play for many resistances that exist.

References

[1] Je Andrews, Gordon Guyatt, Andrew D Oxman, Phil Alderson, Philipp Dahm, Yngve Falck-Ytter, Mona Nasser, Joerg Meerpohl, Piet N Post, Regina Kunz, et al. Grade guidelines: 14. going from evidence to recommendations: the significance and presentation of

recommendations. Journal of clinical epidemiology, 66(7):719-725, 2013.

[2] KP Arjun and K Sampath Kumar. Machine learning-a neoteric medicine to healthcare. International Journal on Emerging Technologies, 11(3):195-201, 2020.

[3] Blagoj Ristevski and Ming Chen. Big data analytics in medicine and healthcare. Journal of integrative bioinformatics, 15(3), 2018.

[4] Harshawardhan S Bhosale and Devendra P Gadekar. A review paper on big data and hadoop. International Journal of Scientific and Research Publications, 4(10):1-7, 2014.

[5] Nancy S Burns and Pamela W Miller. Learning what we didn't know|the sprint data analysis challenge. New England Journal of Medicine, 376(23):2205-2207, 2017.

[6] Juan Jose Beunza Nuin, Enrique Puertas Sanz, and Emilia Condies Moreno. Manual prefactico de inteligencia artificial en entornos sanitarios. Elsevie-, 2020.

[7] SA Senthilkumar, Bharatendara K Rai, Amruta A Meshram, Angappa Gunasekaran, and S Chandrakumarmangalam. Big data in healthcare management: a review of literature. American Journal of Theoretical and Applied Business, 4(2):57-69, 2018

[8] Yogesh Kumar, Kanika Sood, Surabhi Kaul, and Richa Vasuja. Big data analytics and its benefits in healthcare. In Big Data Analytics in Healthcare, pages 3-21. Springer, 2020.

[9] Yogesh Kumar, Kanika Sood, Surabhi Kaul, and Richa Vasuja. Big data analytics and its benefits in healthcare. In Big Data Analytics in Healthcare, pages 3-21. Springer, 2020.

[10] Raghav Toshniwal, Kanishka Ghosh Dastidar, and Asoke Nath. Big data security issues and challenges. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(2), 2015.

[11] Uthayasankar Sivarajah, Muhammad Mustafa Kamal, Zahir Irani, and ishanthWeerakkody. Critical analysis of big data challenges and analytical methods. Journal of Business Research, 70:263-286, 2017.

[12] Je Hale. The Most In-Demand Skills for Data Scientists. Toward Data Sci 2018. https://towardsdatascience.com/the-most-in-demand-skills-for-data-scientists-4a4a8db896db. Accessed: 2023-07-14.

[13] Michaela Miller, Technology & Analytics Practice Lead, U.S. MedTech, IQVIA Oct 29, 2021

[14] Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J. 2017 Feb 14;38(7):500-507.

[15] Al-Jarrah MA, Shatnawi H. Non-proliferative diabetic retinopathy symptoms detection and classification using neural network. J Med Eng Technol. 2017 Aug;41(6):498-505. doi: 10.1080/03091902.2017.1358772. Epub 2017 Aug 8. PMID: 28786703.

[16] Al-Jarrah MA, Shatnawi H. Non-proliferative diabetic retinopathy symptoms detection and classification using neural network. J Med Eng Technol. 2017 Aug;41(6):498-505..

[17] Thongprayoon C, Kaewput W, Kovvuru K, Hansrivijit P, Kanduri SR, Bathini T, Chewcharat A, Leeaphorn N, Gonzalez-Suarez ML, Cheungpasitporn W. Promises of Big Data and Artificial Intelligence in Nephrology and Transplantation. J Clin Med. 2020 Apr 13;9(4):1107

[18] Schena FP, Anelli VW, Trotta J, Di Noia T, Manno C, Tripepi G, D'Arrigo G, Chesnaye NC, Russo ML, Stangou M, Papagianni A, Zoccali C, Tesar V, Coppo R; members of the VALIGA study. Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy. Kidney Int. 2021 May;99(5):1179-1188.

[19] Becker JU, Mayerich D, Padmanabhan M, Barratt J, Ernst A, Boor P, Cicalese PA, Mohan C, Nguyen HV, Roysam B. Artificial intelligence and machine learning in nephropathology. Kidney Int. 2020 Jul;98(1):65-75. doi: 10.1016/j.kint.2020.02.027. Epub 2020 Apr 1. PMID: 32475607; PMCID: PMC8906056.

[20] Mueller FB. AI (Artificial Intelligence) and Hypertension Research. Curr Hypertens Rep. 2020 Aug 27;22(9):70.

[21] E Macias, A Morell, J Serrano, JL Vicario, J Ibeas. Mortality prediction enhancement in end-stage renal disease: A machine learning approach, Informatics in Medicine Unlocked, Volume 19

[22] O Gallés, N MONILL-RAYA, A Morell, J Serrano, D Rexach, J López-Vicario, J Comas, E Martinez, J Ibeas. Deep learning-based prediction for mortality in patients with chronic kidney disease: a new model developed with data from 10.000 patients over the last 11 years