

PROFESSOR AMELIA LICARI (Orcid ID : 0000-0002-1773-6482)

DR SALVATORE FASOLA (Orcid ID : 0000-0002-5378-5505)

PROFESSOR GIAN LUIGI MARSEGLIA (Orcid ID : 0000-0003-3662-0159)

Article type : Review

Artificial intelligence in the diagnosis of pediatric allergic diseases

Authors: Giuliana Ferrante^{1§}, Amelia Licari^{2§}, Salvatore Fasola³, Gian Luigi Marseglia², Stefania La Grutta³

[§]Giuliana Ferrante and Amelia Licari should be considered joint first authors

Affiliations:

¹Department of Health Promotion Sciences, Maternal and Infant Care, Internal Medicine and Medical Specialities "G. D'Alessandro", University of Palermo, Palermo, Italy

²Department of Pediatrics, Fondazione IRCCS Policlinico San Matteo, University of Pavia, Pavia, Italy

³Institute for Biomedical Research and Innovation (IRIB), National Research Council (CNR), Palermo, Italy

Running title: Artificial intelligence in pediatric allergy

Correspondence:

Amelia Licari,

Fondazione IRCCS Policlinico San Matteo, Department of Pediatrics, University of Pavia, piazzale C. Golgi 19, 27100 Pavia, Italy

Email: amelia.licari@unipv.it

Word count: 3670

Number of figures: 3

Conflict of interests: All authors have no competing interests to declare.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/PAI.13419](#)

This article is protected by copyright. All rights reserved

Financial support: none.

Abstract:

Artificial intelligence (AI) is a field of data science pertaining to advanced computing machines capable of learning from data and interacting with the human world. Early diagnosis and diagnostics, self-care, prevention and wellness, clinical decision support, care delivery, and chronic care management have been identified within the healthcare areas that could benefit from introducing AI.

In pediatric allergy research, the recent developments in AI approach provided new perspectives for characterizing the heterogeneity of allergic diseases among patients. Moreover, the increasing use of electronic health records and personal healthcare records highlighted the relevance of AI in improving data quality and processing and setting-up advanced algorithms to interpret the data.

This review aims to summarize current knowledge about AI and discuss its impact on the diagnostic framework of pediatric allergic diseases such as eczema, food allergy, and respiratory allergy, along with the future opportunities that AI research can offer in this medical area.

Keywords: artificial intelligence, children, diagnosis, allergy, eczema, food allergy, respiratory allergy

Introduction

In the upcoming era of Precision Medicine, the novel concept of customized patient care is established to couple clinical features of a disease with molecular profiling to develop diagnostic, prognostic and therapeutic strategies specific to each patient's need.¹ The relevant progress obtained in increasing the comprehension of cellular and molecular mechanisms involved in immune regulation opened innovative avenues to diagnose allergic diseases. The recent availability of "omics" technologies has provided unique opportunities in understanding the pathophysiology of allergic diseases by discovering novel biomarkers that can help classify patients and improve the diagnostic accuracy. Furthermore, a multi-omics approach would be more useful to characterize disease endotypes.² The integration of multi-omics and clinical data requires large-scale databases and strong computational power.

The application of Artificial intelligence (AI) technologies in the field of allergy is thought to contribute significantly to expanding knowledge and providing new insights for disease detection, risk profiling, and decision support. Advanced computational tools may capture the complex relationships linking multi-omics and clinical data, supporting physicians in the diagnostic process identifying risk trajectories, planning early intervention, and delivering personalized healthcare.³

This review aims to summarize current knowledge about AI and discuss its impact on the diagnostic framework of pediatric allergic diseases such as eczema, food allergy, and respiratory allergy, along with the future opportunities that AI research can offer in this medical area.

Artificial Intelligence: sub-disciplines and related fields

Adopting the European Parliament's definition, AI is "the capability of a computer program to perform tasks or reasoning processes that we usually associate with intelligence in a human being".⁴ The term "AI" is often used interchangeably with "Data Science", which is a multidisciplinary subject (including Statistics, Mathematics, and Programming) referring to the whole set of scientific methods, processes, algorithms, and systems aiming to extract knowledge and insights from data.⁵ While a Data Scientist carries out multiple processes such as transformation, analysis, interpretation, and visualization of data, AI is more focused on computer algorithms able to work autonomously (emulating cognition and human understanding) and to develop predictive models. For a Data Scientist, AI can be seen as a tool or a procedure. Similarly, not all AI is Data Science since not all computer programs make use of data.

There are various types of AI. “Assisted Intelligence” is concerned with the development of processes aiming to perform simple, common, and well-defined tasks, therefore “replacing” human intelligence (e.g., an algorithm for calculating the arithmetic mean). “Augmented Intelligence” is concerned with the development of processes designed to “enhance” human intelligence and capabilities, allowing people to do things they could not otherwise do (e.g., an algorithm for the minimization of a complex loss function). The most advanced form of AI is “Autonomous Intelligence”, that is concerned with the development of processes that allow machines and systems to “generate” human intelligence and to act on their own (e.g., processes able to identify the most appropriate algorithm to use for solving a decision problem).

Machine Learning (ML) can be defined as the field of AI providing dynamic systems designed for automatically learning from experience and improve the accuracy of the outcome predictions without being explicitly programmed,⁶ and can be placed under the umbrella of Augmented Intelligence. One aspect that separates ML from other AI fields, such as Symbolic Artificial Intelligence (providing static systems that explicitly define knowledge and rules for the behavior of computer programs), is its ability to modify itself when exposed to more data. ML has close similarities with Computational Statistics, which can be defined as a subset of Data Science concerned with the formulation of statistical models and their estimation using computer algorithms to make inference about the relationships between variables. Indeed, the main difference is that the primary goal of ML is to make predictions as accurate as possible, while statistical models are designed to infer something about variable relationships.

ML can be “unsupervised” if concerned with the task of identifying data patterns in the absence of any pre-existing data label (the outcome is not known, for example when the aim is to identify clusters of observations), or “supervised” if concerned with the task of learning a function that maps an input to an output based on example input-output pairs (the outcome is known, as in regression).

In the era of Big Data (large and multifaceted volumes of information coming from multiple sources and growing at constantly-increasing rates) and Cloud Computing (on-demand availability of high-performance computer resources, such as platforms for Big Data storage and distributed parallel computing), the mining potential of ML has increased substantially. In this regard, Deep Learning (DL) has developed as modern sub-field of ML based on Artificial Neural Networks (ANN) with many layers (deep) that mimic the mechanisms of the human brain in processing complex and high-dimensional data such as images, video, or text.⁷ Indeed, training deep ANN is

highly computationally demanding. However, their potential applications for scientific research and healthcare field can have vital significance, including biological omics data processing, biological image processing, and biomedical diagnosis. Deep learning has close connections with Natural Language Processing (NLP), a subset of AI concerned with the development of processes for analyzing large amounts of human natural language data, with the aim to perform speech recognition, text classification, and natural language generation.

A glossary of all the aforementioned AI terms, sub-disciplines, and related fields (in hierarchical order) is provided in Table 1. A more detailed description of the AI tools⁸ mentioned in the remainder of this paper is provided in Table 2.

Artificial intelligence in healthcare and pediatric allergy research

The AI role in healthcare is focused on improving patient care through big data processing and cloud computing in view of developing innovative medical management information systems and more efficient administrative applications.⁹ Early diagnosis and diagnostics, along with self-care, prevention and wellness, clinical decision support, care delivery, and chronic care management, have all been identified within the healthcare areas with a direct impact on the patient that could benefit from introducing AI.¹⁰ Moreover, the increasing use of electronic health records (EHRs) and personal healthcare records, which may link longitudinal and genomic data, highlighted the relevance of AI in improving data quality and processing, and to set-up advanced algorithms to interpret the data.¹¹

With respect to the past, the increasing availability of digital data and their cloud storage and computing, coupling with the innovative use of ML applications for data interpretation, represent the main sources of potential benefits for the health system (Figure 1). Indeed, the main advantage of ML approaches is to overcome the limits of using conventional modelling techniques, through which the underlying complexity of the data may not be accurately accounted for.¹² In the last years, AI tools increasingly gained attention among researchers, encouraging them to develop mutual collaborations to explore alternative routes to access adequate computational resources. Of note, the positive and fair contribution of AI advances, in terms of smart tools and analytical techniques, requires multidisciplinary research involving social scientists who can make such tools acceptable, usable, and ethical. Currently, the contribution of AI is relevant to the development of sophisticated, personalized models that can enable clinicians to plan the diagnostic workflow based on a range of possible, and increasingly accurate, predicted outcomes.¹³ This transformation

generated new data storage and processing technologies, offering new tools for understanding the complexity of diseases, such as allergic diseases.¹⁴

In pediatric allergy research, the recent developments in ML and DL approaches provided new perspectives for characterizing the heterogeneity of allergic diseases among patients.¹⁵⁻¹⁷ While the transformative impact of AI on biomedical research in the field of pediatric allergy is undeniable (Figure 2), physicians' poor knowledge of such a promising approach may be a barrier for a full application in clinical practice.

AI tools for allergy test interpretation

Advancements in the field of allergy diagnosis by using recombinant allergens led to the development of the so-called Component-Resolved Diagnosis (CRD).¹⁸ CRD can help to depict the patient's immunoglobulin E (IgE) profiles and clarify whether specific IgE derives from primary sensitization or cross-reactivity. It is helpful especially in polysensitized subjects, those with an unclear clinic and/or sensitization pattern, and patients who do not respond to treatment.¹⁹

In the last decades, the implementation of AI in the diagnosis of allergic diseases have produced innovative instruments for the interpretation of allergy tests, known as Computer-Supported Diagnosis (CSD) tools. CSD is based on Expert Systems (ESs), i.e., knowledge-based systems developed to support clinical routine acting as human experts for delivering advice and suggestions. ESs are advanced informatics tools used in clinical practice to support human experts' replacement with appropriate decision-making systems.²⁰ For example, interpreting the results of allergy tests based on microarray technology may be quite complex, particularly for clinicians not dealing with such a test in their clinical routine. Therefore, the availability of a computer-assisted tool trained with data from international guidelines and experts' opinions may play a significant role in supporting the diagnostic process. In recent years, an ES called Allergenius® was developed to support the interpretation of allergy tests based on microarray technology (ImmunoCAP ISAC®). The study validated this ES using artificial and real-life cases and compared its reports to those of clinicians with experience in molecular diagnostics, suggesting that Allergenius® could be a promising tool to improve the understanding of ISAC® results and to obtain a comprehensive interpretation of the patient's IgE profile. Indeed, Allergenius® was found to be able to distinguish cross- and co-sensitization, and to identify sensitization to genuine and cross-reactive components. A perfect concordance between the

allergists and the ES was found in 10/18 issues. Of note, allergists did not identify issues that Allergenius® was not able to detect.²¹

The understanding of the clinical implications of CRD may be improved by applying clustering methods to classify variables based on the similarity or dissimilarity of the data. A very recent study on children and adults with mite-sensitized respiratory allergic disorders evaluated the co-sensitization and cross-reactivity of 17 allergen components using the CRD method. Der p 1 (81.48%), Der f 2 (77.78%), Der f 1 (74.07%), Der p 2 (66.20%) and Der p 23 (54.63%) were the main sensitized components, and the 17 components were divided into five sensitization clusters through hierarchical cluster analysis.²² The allergen component in each cluster showed a high proportion of co-sensitization and cross-reactivity. Notably, similar sensitization profiles were found for each component in the same cluster, regardless of age, total IgE levels, and disease type. Overall, these data suggest that the application of advanced informatics and statistical tools may help to support clinicians in interpreting CRD results, allowing to accurately identify the sensitization profile and cross-reactions between component proteins, and suggesting closer attention to allergy components that are strictly related to each other.²²

AI tools for diagnosing eczema

Eczema is a complex inflammatory skin disease which comprises highly heterogeneous sub-phenotypes/endotypes. Advanced computational approaches have been increasingly used for capturing this heterogeneity, learning from existing data, and making predictions on new data for distinguishing between disease states and for identifying complex relationships between multi-omics and clinical data.²³ In this context, Ghosh et al. have previously shown that a Support Vector Machine (a supervised method that performs classification by mapping the inputs into high-dimensional spaces using linear and non-linear transformations) successfully discriminated eczema from healthy skin with 98% predictive accuracy. Using human microarray datasets, the authors identified 89 eczema-related signature genes that may represent useful biomarkers to classify patients from healthy controls, and that could also serve for therapeutic stratification.²⁴

A ML approach was recently used to test allergens' predictive ability in relation to allergic diseases in children participating in a population-based birth cohort. The discriminative ability of specific IgE patterns in relation to clinical outcomes was analyzed through different ML models such as decision trees (that grouped study population based on allergen thresholds and combinations) and Bayesian networks (exploring causal dependencies of variables), to explore the

association between the information provided by the microarray allergy test and clinical symptoms. Unlike symptoms of asthma, rhino-conjunctivitis, wheeze, and airway hyper-reactivity, the prediction models showed poor performance for eczema symptoms, suggesting that CRD may not be a reliable diagnostic tool in children with eczema, reasonably for the more relevant role of the disrupted skin barrier function than IgE mediated mechanisms in the pathogenesis of eczema, or for the poor availability on the chip of allergens with high predictive ability.²⁵

Nonetheless, a very recent study, using an innovative method of ML (a cluster analysis based on dissimilarity measures derived through random forests) to investigate complex relationships among various chemokines and eczema in a cohort of schoolchildren, highlighted the benefits of such an approach in characterizing eczema endotypes. The authors reported that children whose signs and symptoms fulfilled the criteria of eczema tended to show higher levels of the chemokine ligands, CCL17, CCL22, and CCL27, which are useful for diagnosing eczema.²⁶

Overall, despite recent developments in computational analysis highlight the applicability and the promising contribution of these methods for eczema endotyping and classification, few studies have been conducted in children so far. Different approaches will need to be integrated and validated in large-scale and longitudinal cohorts to achieve a better definition of pathophysiologic mechanisms and to identify novel biomarkers for diagnosis.²⁷

AI tools for diagnosing food allergy

Food allergy (FA) is emerging as a complex and heterogeneous disease encompassing different reaction phenotypes, defined in terms of both clinical and molecular characteristics (i.e., endotypes). The actual classifications of FA are mainly based on clinical characteristics and known immunologic mechanisms. The potential of AI to improve FA diagnosis and to correlate endotypes with clinical phenotypes is considerable.

ML techniques have been integrated into a medical algorithm to identify patients with FA. A predictive probability score for eosinophilic esophagitis (EoE), named p(EoE), was developed using esophageal mRNA transcript analyses with ML strategies (random forest classification following weighted factor analysis).²⁸ The P(EoE) demonstrated high sensitivity and specificity (0.91 and 0.93, respectively) for diagnosis in histologically unequivocal patients, while it diagnosed EoE in 84.6% of the equivocal patients in their first biopsies. Furthermore, p(EoE) was able to identify a subgroup of EoE patients characterized by IgE-mediated tissue allergy.²⁸ Other ML approaches (decision tree, logistic regression, radial basis function and multi-layer perceptron

neural networks) were used to create a diagnostic model for FA, using methylation data from 58 patient samples collected from both infants with egg or peanut allergy and infant only sensitized to one of those allergens. With this computational approach, 18 highly discriminating Cytosine-phosphate-Guanine (CpGs) dinucleotides were identified, and presented high accuracy in differentiating between food sensitization and clinical reactivity; also, the 18 CpGs mapped to 13 genes, most of which involved in the immune system and previously associated with FA.²⁹ Furthermore, the combination of the results of a peptide microarray immunoassay and ML methods (decision tree and Support Vector Machine) allowed to improve the diagnosis in 62 peanut-allergic children and adolescents.³⁰

In this way, another step has been recently taken by the application of unbiased, data-driven analyses to peanut allergy. In particular, k-means clustering has been performed to identify patterns of mast cell activation test (MAT) responses in a cohort of 42 peanut-sensitized patients.³¹ With this approach, five different clusters of patient phenotypes with similar MAT responses were identified, ranging from peanut-tolerant patients to patients who reacted to a low threshold of peanut exposure with a tendency toward systemic reactions.³² This exploratory, unsupervised machine-based learning approach could also be extended to other allergens and could potentially be useful in identifying FA phenotypes if related to the clinical reactivity and the outcome of diagnostic tests from individual patients.

Accurate documentation of patients with food allergies is vital to patient management and quality of care and may also serve as a source for clinical and epidemiological research. A NLP approach has been applied to process, encode, and group a large quantity of food allergen terms and any adverse sensitivities reported within large datasets of EHRs in the United States.³³ This innovative approach revealed several limitations in how food allergies have been documented, including the frequently missing documentation of reaction mechanism and type, and the lack of a comprehensive terminology of food allergens. Furthermore, these results highlighted the need to redesign the collection and use of allergy information within EHRs.³³

Several studies suggest that the number or function of T-cells is impaired in allergic patients. ANN, a family of powerful prediction methods that mimic biological neural networks, have been used to investigate the expression patterns of different immune cells in children with cow's milk allergy (CMA). In particular, Savilahti et al. identified different expression profiles of both T helper cell type 2 (Th2)- and T regulatory cell (Tregs)-related genes in children with persistent CMA, children with tolerance development, and non-atopic subjects. After searching through all

surface marker combinations, the combined expression of Tregs and Th2 cell markers forkhead box Protein 3 (FoxP3), nuclear factor of activated T-cells (Nfat)-C2, interleukin (IL)-16, and trans-acting T-cell-specific transcription factor (GATA)-3 distinguished children with persisting CMA most accurately from other study groups and revealed novel expression patterns of T-cells in CMA.³⁴

Overall, these studies highlight the ability of AI to improve diagnostic precision and to improve patient care in FA, disclosing new perspectives in the implementation of AI in clinical practice.

AI tools for diagnosing respiratory allergic diseases

ML approaches quickly attracted scientists so that the number of successful applications (e.g., Bayesian ML modelling framework, Markov chain probability models, hierarchical clustering, Bayesian networks) in the field of pediatric allergic respiratory diseases rapidly increased in the last years.^{25,35-38} All these encouraging results are due to the 'data revolution' through which the ML background was created. In this context, a data-driven approach to identify structures within the data adopting unsupervised learning of latent variables has been proposed to yield a cutting-edge way of capturing the heterogeneity in longitudinal patterns of distinct symptom categories within individual patients.^{25,35-38}

Diagnosing an allergic disease based on the subjective answer to standardized questions asked by the physician or by fulfilling a self-administered questionnaire may be affected by recall bias, thereby posing a risk for the precise identification of the time in which the disease occurred, the so-called "index data". Similarly, the lack of a clear identification of temporal aspects may reduce the possibility of characterizing the relationship between predictors and outcomes. The application of novel AI methods such as NLP systems may offer advantages in determining the index data or incidence date of outcome events, based on their capability of taking into account the occurrence of outcome events throughout time.³⁹⁻⁴¹ The feasibility of diagnosing asthma by a NLP algorithm application demonstrated very good concordance with chart review of electronic medical records (EMRs) by abstractor (e.g., criterion validity) and an association with the known risk factors for asthma (e.g., construct validity).⁴² Recently, Kaur et al. developed and validated an NLP algorithm to identify patients that meet Asthma Predictive Index (API) criteria, providing evidence that asthma diagnosis by NLP-API has the potential to enhance asthma care in practice.⁴³

Within the umbrella of AI tools, the application in pediatric allergy of deep learning tools (Convolutional Neural Network (CNN), ANN) allowed to adapt and learn from a massive quantity

of data with multiple levels of abstraction in order to enhance physicians in the diagnostic process.⁴⁴ After using CCN methods for the nasal cellularity image processing and image segmentation techniques, Dimauro et al. have successfully gained the automatic identification and cellular classification of rhinocytogram, providing an innovative, compelling diagnostic support to practical clinical care in making an accurate differential diagnosis.⁴⁵ The relevance of these results is undeniable since nasal cytology is a diagnostic technique characterized by simplicity and low invasiveness, qualities that allow its wide diffusion.

Perspectives and conclusions

The promise of AI in the field of pediatric allergic diseases offers the opportunity to dramatically improve the diagnostic accuracy and to deliver personalized healthcare, even though several aspects need to be addressed before such an innovative approach could be included in routine clinical practice (Figure 3).

It is mandatory to test the potential of new computational algorithms and deep learning in large-scale studies. In this context, randomized clinical trials are especially needed for obtaining measures of clinical benefit, facilitating causal inference, given their paucity in this field of research.⁴⁶ Further, unmet needs like cost-efficacy and safety issues should be considered before translating AI technologies from research to clinical settings. Indeed, appropriate regulatory processes are required to guarantee security measures for sensitive data storage and management.

To guarantee a fair and inclusive AI-based healthcare represents another critical issue. Healthcare AI should be trained and validated on population-representative data to achieve generalizable performance levels.⁴⁷ Indeed, social and health disparities could leave out groups of children who traditionally have poor access to care. Consequently, collecting data predominantly from majority ethnic groups or high socioeconomic-level patients would risk biasing AI performance, as the computer will learn diagnostic patterns deriving from these over-represented groups.⁴⁸

Another relevant issue concerns the development of educational programs to support clinicians in using and interpreting AI products and services via continuing education. Indeed, although AI can have an important role in improving the diagnostic process, it may entail pitfalls in misleading diagnosis and inaccurate risk assessment.

Finally, structured coordination amongst healthcare systems, the research community, and industry will be particularly important to address all the critical aspects of AI and to guarantee its transferability from academic research towards real life.

Physicians will likely work closely alongside AI-guided computers in the near future. By discovering patterns in data obtained from different diagnostic tests, AI will assist clinicians in the work-up of an individual patient, improving clinical management in a revolutionary fashion.⁴⁹ In this scenario, the contribution of healthcare professionals working in multidisciplinary teams will be fundamental. Indeed, the application of AI algorithms involves the collection and the analysis of large amounts of data, which have to be adequately interpreted within a specific clinical context. It is, therefore, in the space between "autonomous intelligence" and "assisted intelligence" that the collaboration between clinicians and computers will succeed. Within this space called "augmented intelligence", humans and machines will work in synergy, sharing decisions and collaborating to optimize the diagnostic workflow.

Authors' contribution statement: Giuliana Ferrante, Amelia Licari: writing – original and final draft, review and editing (equal). Salvatore Fasola: writing – original and final draft, review and editing (equal). Gian Luigi Marseglia: Conceptualization (supporting); Supervision (supporting); Writing – review and editing (equal). Stefania La Grutta: Conceptualization (lead); Supervision (lead); Writing – review and editing (equal).

Acknowledgments: None.

Impact statement: Artificial intelligence may offer the opportunity to improve the diagnostic accuracy and to deliver personalized healthcare in the field of pediatric allergic diseases.

References

1. Seyhan AA, Carini C. Are innovation and new technologies in precision medicine paving a new era in patients centric care? *J Transl Med.* 2019;17:114.
2. Mersha TB, Afanador Y, Johansson E, Proper SP, Bernstein JA, Rothenberg ME, et al. Resolving Clinical Phenotypes into Endotypes in Allergy: Molecular and Omics Approaches. *Clin Rev Allergy Immunol.* 2020 May 6.
3. Rider NL, Srinivasan R, Khoury P. Artificial intelligence and the hunt for immunological disorders. *Curr Opin Allergy Clin Immunol.* 2020 Sep 29.
4. Artificial intelligence: Potential benefits and ethical considerations, European Parliament Legal Affairs briefing, Policy Department C: Citizens' Rights and Constitutional Affairs, PE 571.380,2016,
[https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI\(2016\)571380_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_EN.pdf) Accessed May 24, 2020
5. Dhar, V., 2013. Data science and prediction. *Communication of the ACM.* Volume 56, 12, pp. 64-73, O'Neil, C. and Schutt, R. 2014. *Doing Data Science.* O'Reilly Media, CA
6. Arthur L. Samuel. Some studies in machine learning using the game of checkers, in *IBM Journal of research and development*, 1959
7. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444
8. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer.
9. Transforming healthcare with AI: The impact on the workforce and organizations.
https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey_Transforming-Healthcare-with-AI.pdf Accessed May 24, 2020
10. Colclough G, Dorling G, Riahi F, Ghafur S, Sheikh A. Harnessing Data Science and AI in Healthcare: From policy to practice. Doha, Qatar: World Innovation Summit for Health, 2018 <http://www.wish.org.qa/wp-content/uploads/2018/11/IMPJ6078-WISH-2018-Data-Science-181015.pdf> Accessed May 24, 2020
11. Artificial intelligence and health. Summary report of a roundtable held on 16 January 2019. The Academy of Medical Sciences. National Institute for Health Research (NIHR)
12. Bishop CM. Model-based machine learning. *Philos Trans A Math Phys Eng Sci.* 2012;371:20120222.

- Accepted Article
13. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. *Per Med*. 2013;10:565-576.
 14. Bousquet J, Smit H, Keil T, Roll S. Sex-specific incidence of asthma, rhinitis and respiratory multimorbidity before and after puberty onset: individual participant meta-analysis of five birth cohorts collaborating in MeDALL. *BMJ Open Respir Res*. 2019;6:e000460.
 15. Wu J, Prosperi MC, Simpson A, Hollams EM, Sly PD, Custovic A, Holt PG. Relationship between cytokine expression patterns and clinical outcomes: two population-based birth cohorts. *Clin Exp Allergy*. 2015;45:1801-1811.
 16. Brew BK, Chiesa F, Lundholm C, Örtqvist A, Almqvist C. A modern approach to identifying and characterizing child asthma and wheeze phenotypes based on clinical data. *PLoS One*. 2019;14:e0227091.
 17. Virkud YV, Kelly RS, Croteau-Chonka DC, Celedón JC, Dahlin A, Avila L, et al. Novel eosinophilic gene expression networks associated with IgE in two distinct asthma populations. *Clin Exp Allergy*. 2018;48:1654-1664.
 18. Sastre J. Molecular diagnosis in allergy. *Clin Exp Allergy*. 2010;40:1442-1460.
 19. Stringari G, Tripodi S, Caffarelli C, Dondi A, Asero R, Di Rienzo Businco A, et al.; Italian Pediatric Allergy Network (I-PAN). The effect of component-resolved diagnosis on specific immunotherapy prescription in children with hay fever. *J Allergy Clin Immunol*. 2014;134: 75–81.
 20. Hopgood AA. The state of artificial intelligence. *Advances in computers*, M. Zelkowitz Ed. Elsevier. 2005;7:1–75.
 21. Melioli G, Spenser C, Reggiardo G, Passalacqua G, Compalati E, Rogkakou A, et al. Allergenius, an expert system for the interpretation of allergen microarray results. *World Allergy Organ J*. 2014;7:15.
 22. Huang Z, Zou X, Chen H, Liao C, Hu H, Luo W, et al. Identifying Potential Co-Sensitization and Cross-Reactivity Patterns Based on Component-Resolved Diagnosis. *Int Arch Allergy Immunol*. 2020;181:81-93.
 23. Ghosh D, Bernstein JA, Khurana Hershey GK, Rothenberg ME, Mersha TB. Leveraging Multilayered "Omics" Data for Atopic Dermatitis: A Road Map to Precision Medicine. *Front Immunol*. 2018;9:2727.

24. Ghosh D, Ding L, Sivaprasad U, Geh E, Biagini Myers J, Bernstein JA, et al. Multiple Transcriptome Data Analysis Reveals Biologically Relevant Atopic Dermatitis Signature Genes and Pathways. *PLoS One*. 2015;10:e0144316.
25. Prosperi MC, Belgrave D, Buchan I, Simpson A, Custovic A. Challenges in interpreting allergen microarrays in relation to clinical symptoms: a machine learning approach. *Pediatr Allergy Immunol*. 2014;25:71-79.
26. Yamamoto-Hanada K, Kawakami E, Saito-Abe M, Sato M, Mitsubuchi H, Oda M, et al. Exploratory analysis of plasma cytokine/chemokine levels in 6-year-old children from a birth cohort study. *Cytokine*. 2020;130:155051.
27. Eyerich K, Brown SJ, Perez White BE, Tanaka RJ, Bissonette R, Dhar S, et al. Human and computational models of atopic dermatitis: A review and perspectives by an expert panel of the International Eczema Council. *J Allergy Clin Immunol*. 2019;143:36-45.
28. Sallis BF, Erkert L, Moñino-Romero S, Acar U, Wu R, Konnikova L, et al. An algorithm for the classification of mRNA patterns in eosinophilic esophagitis: Integration of machine learning. *J Allergy Clin Immunol*. 2018;141:1354-1364.e9.
29. Alag A. Machine Learning Approach Yields Epigenetic Biomarkers of Food Allergy: A Novel 13-gene Signature to Diagnose Clinical Reactivity *PLoS One*. 2019;14(6):e0218253.
30. Lin J, Bruni FM, Fu Z, Maloney J, Bardina L, Boner AL, et al. A bioinformatics approach to identify patients with symptomatic peanut allergy using peptide microarray immunoassay. *J Allergy Clin Immunol*. 2012;129:1321-1328.e5.
31. Bahri R, Custovic A, Korosec P, Tsoumani M, Barron M, Wu J, et al. Mast cell activation test in the diagnosis of allergic disease and anaphylaxis. *J Allergy Clin Immunol*. 2018;142:485-496.e16.
32. Chong KW, Ruiz-Garcia M, Patel N, Boyle RJ, Turner PJ. Reaction phenotypes in IgE-mediated food allergy and anaphylaxis. *Ann Allergy Asthma Immunol*. 2020;124:473-478.
33. Plasek JM, Goss FR, Lai KH, Lau JJ, Seger DL, Blumenthal KG, et al. Food Entries in a Large Allergy Data Repository. *J Am Med Inform Assoc*. 2016 Apr;23:e79-87.
34. Savilahti EM, Karinen S, Salo HM, Klemetti P, Saarinen KM, Klemola T, et al. Combined T regulatory cell and Th2 expression profile identifies children with cow's milk allergy. *Clin Immunol*. 2010;136:16-20.

35. Simpson A, Tan VY, Winn J, Svensén M, Bishop CM, Heckerman DE, Buchan I, Custovic A. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med*. 2010;181:1200–1206.
36. Prosperi MC, Sahiner UM, Belgrave D, Sackesen C, Buchan IE, Simpson A, Yavuz TS, Kalayci O, Custovic A. Challenges in identifying asthma subgroups using unsupervised statistical learning techniques. *Am J Respir Crit Care Med*. 2013; 188: 1303–1312.
37. Lazic N, Roberts G, Custovic A, Belgrave D, Bishop CM, Winn J, Curtin JA, Hasan Arshad S, Simpson A. Multiple atopy phenotypes and their associations with asthma: similar findings from two birth cohorts. *Allergy*. 2013; 68: 764–770.
38. Belgrave DC, Granell R, Simpson A, Guiver J, Bishop C, Buchan I, Henderson AJ, Custovic A. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med*. 2014;11(10):e1001748.
39. Dalan D. Clinical data mining and research in the allergy office. *Curr Opin Allergy Clin Immunol*. 2010;10:171-177.
40. Goss FR, Plasek JM, Lau JJ, Seger DL, Chang FY, Zhou L. An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. *AMIA Annu Symp Proc*. 2014;2014:580-588.
41. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *J Allergy Clin Immunol*. 2020;145:463-469.
42. Wi CI, Sohn S, Rolfes MC, Seabright A, Ryu E, Voge G, et al. Application of a Natural Language Processing Algorithm to Asthma Ascertainment. An Automated Chart Review. *Am J Respir Crit Care Med*. 2017;196:430-437.
43. Kaur H, Sohn S, Wi CI, Ryu E, Park MA, Bachman K, et al. Automated chart review utilizing natural language processing algorithm for asthma predictive index. *BMC Pulm Med*. 2018;18:34.
44. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
45. Dimauro G, Ciprandi G, Deperte F, Girardi F, Ladisa E, Latrofa S, et al. Nasal cytology with deep learning techniques. *Int J Med Inform*. 2019;122:13-19.
46. Angus DC. Randomized Clinical Trials of Artificial Intelligence. *JAMA*. 2020 Feb 17. doi: 10.1001/jama.2020.1039.

47. Matheny ME, Whicher D, Thadaney Israni S. Artificial Intelligence in Health Care: A Report From the National Academy of Medicine JAMA. 2019 Dec 17; 10.1001/jama.2019.21579.
48. Bergl PA, Wijesekera TP, Nassery N, Cosby KS. Controversies in diagnosis: contemporary debates in the diagnostic safety literature. *Diagnosis (Berl)*. 2020;7:3–9.
49. Ferrante G, Licari A, Marseglia GL, La Grutta S. Artificial intelligence as an emerging diagnostic approach in paediatric pulmonology. *Respirology*. 2020 May 10. doi: 10.1111/resp.13842.

Tables

Table 1. Glossary of Artificial Intelligence terms, its sub-disciplines, and related fields (in hierarchical order)

<i>Term</i>	<i>Definition</i>	<i>Sub-discipline of</i>	<i>Connections with</i>
Artificial Intelligence	The capability of a computer program to perform tasks or reasoning processes that we usually associate with intelligence in a human being.	-	Data science
Data Science	Multidisciplinary subject referring to the set of scientific methods, processes, algorithms and systems aiming to extract knowledge and insights from data.	-	Artificial intelligence
Assisted Intelligence	A subset of Artificial Intelligence concerned with the development of processes aiming to perform simple, common, and well-defined tasks (replacing human intelligence).	Artificial intelligence	Data Science
Augmented Intelligence	A subset of Artificial Intelligence concerned with the development of processes designed to augment human capabilities (enhancing human intelligence).	Artificial intelligence	Data Science
Autonomous Intelligence	A subset of Artificial Intelligence concerned with the development of processes that allow machines and systems to act on their own, independent of human intervention (generating human intelligence).	Artificial intelligence	Data Science

Symbolic Artificial Intelligence	A subset of Artificial Intelligence providing static systems that explicitly define knowledge and rules (typically if-then rules) for the behaviour of computer programs.	Artificial intelligence	Data science
Computational Statistics	A subset of Data Science concerned with the formulation of statistical models and their estimation using computer algorithms, with the goal of making inference about the relationships between variables.	Data science	Machine Learning
Machine Learning	A subset of Artificial Intelligence providing dynamic systems designed for automatically learning from experience and improve the accuracy of the outcome predictions without being explicitly programmed.	Artificial intelligence	Data science
Unsupervised Machine Learning	A subset of Machine Learning concerned with the task of identifying data patterns in the absence of any pre-existing data label.	Machine learning	Data science
Supervised Machine Learning	A subset of Machine Learning concerned with the task of learning a function that maps an input to an output based on example input-output pairs.	Machine learning	Data science
Deep Learning	A subset of Machine Learning based on artificial neural networks with many layers (deep) that mimic the mechanisms of the human brain in processing complex and high-dimensional data such as images, video, or text.	Machine learning	Data science
Big Data	Large and multifaceted volumes of information coming from multiple sources and growing at constantly-increasing rates.	-	Deep learning
Cloud Computing	On-demand availability of high performance computer resources, such as	-	Deep

	platforms for Big Data storage and distributed parallel computing.		learning
Natural Language Processing	A subset of Artificial Intelligence concerned with the development of processes for analyzing large amounts of human natural language data, with the aim to perform speech recognition, text classification, and natural language generation.	Artificial intelligence	Deep learning

Table 2. Artificial Intelligence tools and their application in pediatric allergy.

<i>Tool</i>	<i>Description</i>	<i>Applications in pediatric allergy diseases</i>
Hierarchical clustering	Unsupervised algorithm for grouping similar objects into clusters, by sequentially merging elementary units (agglomerative) or sequentially splitting sets of units starting from the whole sample (divisive).	To identify clusters of atopic sensitizations. ²¹ To identify asthma phenotypes. ³²
k-means clustering	Supervised algorithm aiming to partition observations into (k) clusters, so that each observation will belong to the cluster with the nearest mean (cluster centroid), serving as a prototype of the cluster.	To identify peanut allergy phenotypes. ²⁸
Factor analysis	Unsupervised method used to describe the variability among several observed, correlated variables, in terms of a potentially lower number of unobserved variables called factors.	To identify a subgroup of patients with eosinophilic esophagitis characterized by IgE-mediated tissue allergy. ¹²
Logistic regression	Supervised approach used to model the probability of a certain class or event to occur as a function of explanatory variables.	To differentiate between food sensitization and clinical reactivity. ²⁶ To assess whether asthma severity is associated with particular disease phenotypes. ³²

Decision trees	Supervised method in which the data are sequentially splitted (throughout the branches) according to several input thresholds to obtain final groups (represented in the leaves) that are informative about the target value.	To differentiate between food sensitization and clinical reactivity. ²⁶ To identify patients with symptomatic peanut allergy. ²⁷
Random forests	Multitude of decision trees (forest) trained by taking random decisions about the samples and the variables to include each time (for reducing overfitting), and outputting predictions based on the mode (classification) or the mean (regression) of the predictions of the individual trees.	To characterize eczema endotypes. ¹³
Support Vector Machine	Supervised method that performs classification by mapping the inputs into high-dimensional spaces using linear or non-linear transformations.	To identify biomarkers for eczema diagnosis. ²³
Artificial Neural Networks	Supervised method based on a collection of connected units or nodes (from the input nodes to the output nodes, through several layers of hidden nodes) called artificial neurons, vaguely inspired by the neurons in a biological brain. Each artificial neuron receives a signal, processes it according to some activation function, and signals neurons connected to it.	To identify different immunologic profiles in children with cow's milk allergy. ³⁰
Radial Basis Function	An artificial neural network that uses radial basis functions as	To differentiate between food sensitization and

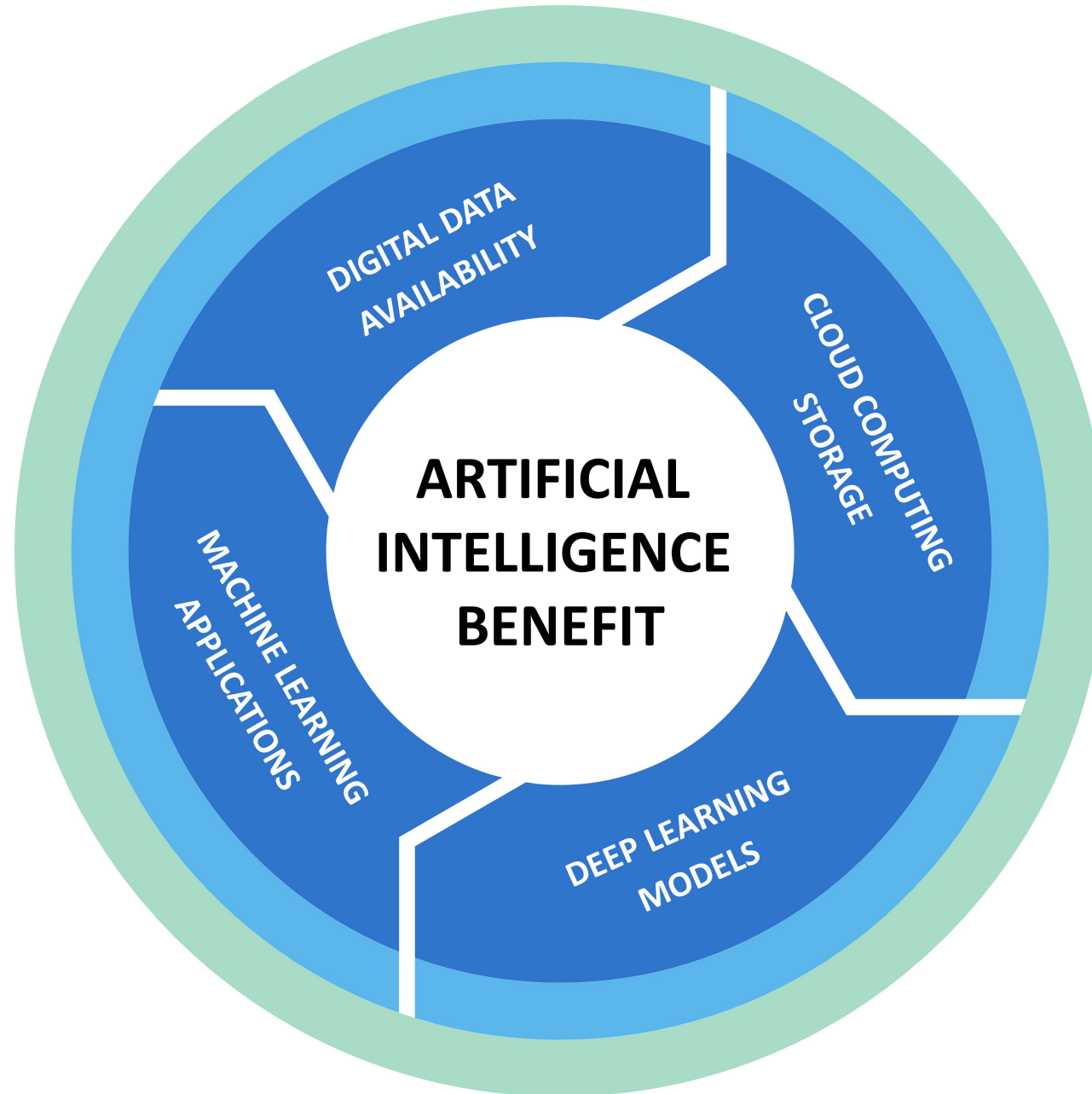
neural networks	activation functions.	clinical reactivity. ²⁶
Multi-Layer Perceptron neural networks	A deep artificial neural network with multiple layers and multiple nodes characterized by a non-linear activation function.	To differentiate between food sensitization and clinical reactivity. ²⁶
Convolutional Neural Network	An Artificial Neural Network where the particular input data structure (for example the matrix of color intensities for an image) enhances the importance of spatial neighborhood to catch locally invariant responses, through the introduction of locally equal weights.	To classify rhinitis through nasal cytology. ⁴¹
Markov chain probability models	Supervised models describing a sequence of possible states, where the probability of each state depends only on the previously attained state.	To identify phenotypes of eczema, wheeze, and rhinitis over time. ³⁴
Bayesian network	Supervised, probabilistic graphical model that represents a set of variables and their conditional dependencies through a directed acyclic graph.	To assess whether asthma severity is associated with particular disease phenotypes. ³²

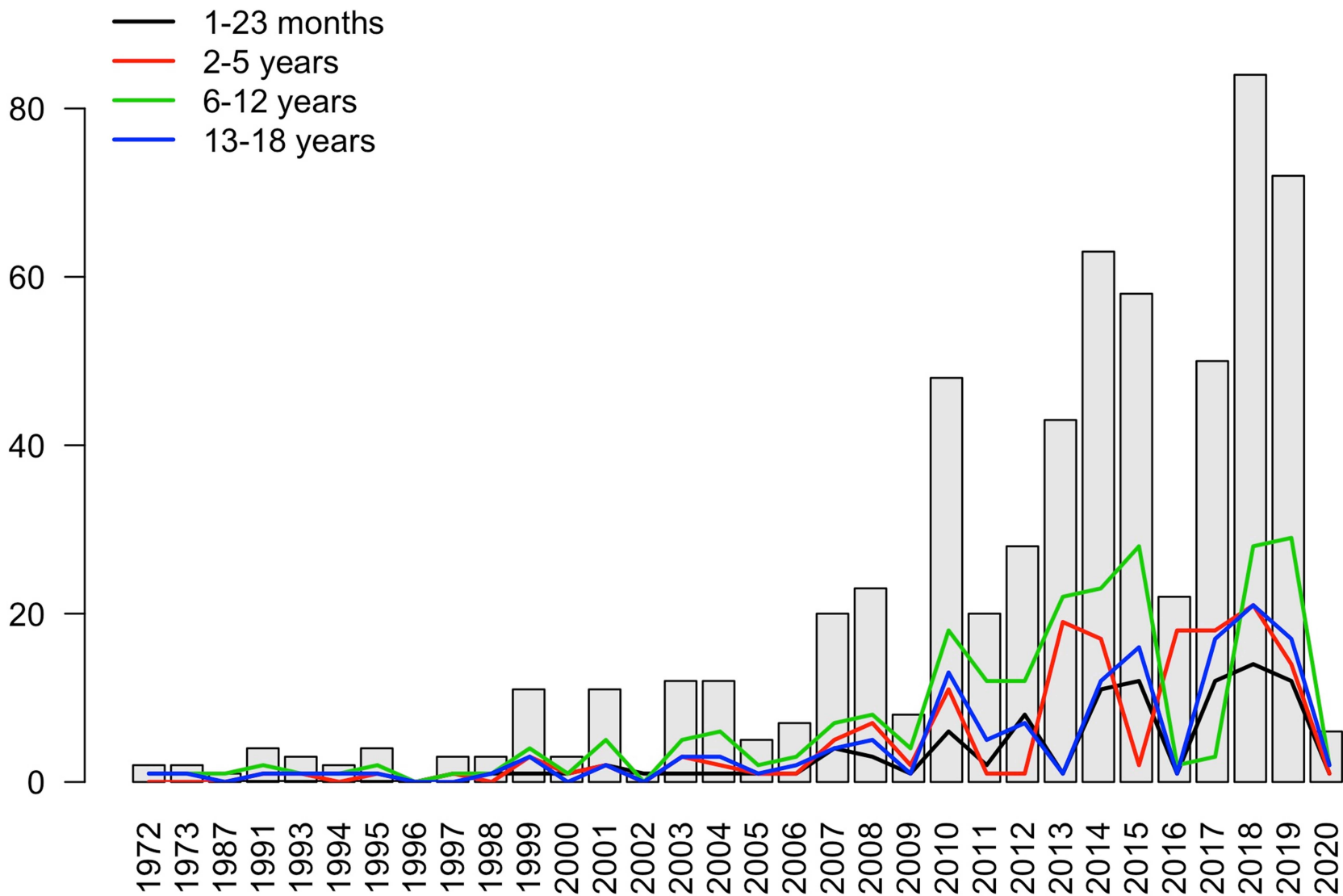
Figure Legends:

Figure 1. Potential benefit of artificial intelligence for the health system.

Figure 2. Artificial intelligence and diagnosis of pediatric allergic diseases: growing impact of scientific literature. Electronic search was undertaken in Medline from the date of inception to May 28, 2020, using the following key-words: (artificial intelligence OR machine learning OR principal component analysis OR deep learning OR natural language processing OR cluster analysis OR latent class analysis OR directed acyclic graph OR Neural Networks or data mining) AND (allergy) AND (children) AND (diagnosis). The number in each bar represents the total records found by the initial query that were published during that year. The four lines show the number of records by publication year when the following age groups are considered: 1-23 months (gray line); 2-5 years (red line); 6-12 years (green line); 13-18 (blue line).

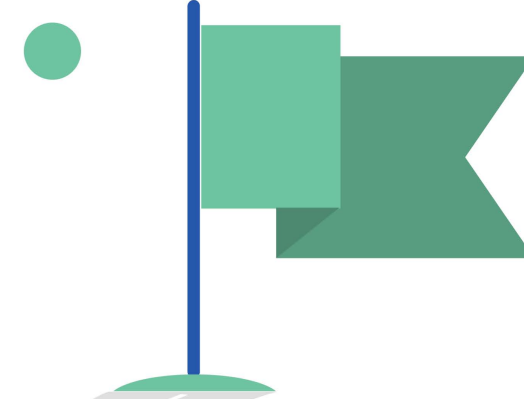
Figure 3. The roadmap of artificial intelligence in healthcare.





AI in routine clinical practice

Improvement of diagnostic accuracy and delivery of personalized healthcare



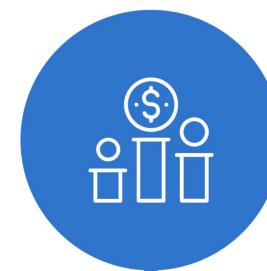
Ensuring equitable and inclusive AI-based healthcare

Achieving performance levels suitable for all populations
Reducing social and health disparities



Addressing unmet needs

Identifying and solving safety issues
Evaluation of cost-effectiveness



Translating AI from research to clinical settings

Testing the efficacy of AI technologies in clinical trials
Education and support of clinicians

