Data Science and Management 4 (2021) 19–29

Contents lists available at ScienceDirect

Data Science and Management

journal homepage: www.keaipublishing.com/en/journals/data-science-and-management

Discussion

Machine learning-based approach: global trends, research directions, and regulatory standpoints

Raffaele Pugliese[a,*,1], Stefano Regondi[a,1], Riccardo Marini[b]

[a] NeMO Lab, ASST Niguarda Cà Granda Hospital, Milan, 20162, Italy
[b] CBA Lex, Corso Europa, 20122, Milan, Italy

ARTICLE INFO

Keywords
Machine learning
Artificial intelligence
Research trends
Healthcare
Data governance
Nanotechnology
Cybersecurity

ABSTRACT

The field of machine learning (ML) is sufficiently young that it is still expanding at an accelerating pace, lying at the crossroads of computer science and statistics, and at the core of artificial intelligence (AI) and data science. Recent progress in ML has been driven both by the development of new learning algorithms theory, and by the ongoing explosion in the availability of vast amount of data (often referred to as "big data") and low-cost computation. The adoption of ML-based approaches can be found throughout science, technology and industry, leading to more evidence-based decision-making across many walks of life, including healthcare, biomedicine, manufacturing, education, financial modeling, data governance, policing, and marketing. Although the past decade has witnessed the increasing interest in these fields, we are just beginning to tap the potential of these ML algorithms for studying systems that improve with experience. In this paper, we present a comprehensive view on geo worldwide trends (taking into account China, the USA, Israel, Italy, the UK, and the Middle East) of ML-based approaches highlighting the rapid growth in the last 5 years attributable to the introduction of related national policies. Furthermore, based on the literature review, we also discuss the potential research directions in this field, summarizing some popular application areas of machine learning technology, such as healthcare, cyber-security systems, sustainable agriculture, data governance, and nanotechnology, and suggest that the "dissemination of research" in the ML scientific community has undergone the exceptional growth in the time range of 2018–2020, reaching a value of 16,339 publications. Finally, we report the challenges and the regulatory standpoints for managing ML technology. Overall, we hope that this work will help to explain the geo trends of ML approaches and their applicability in various real-world domains, as well as serve as a reference point for both academia and industry professionals, particularly from a technical, ethical and regulatory point of view.

1. Introduction

In today's world, we are constantly surrounded by data. Everything around us is connected to a data source (i.e., smartphones, social media, personalized advertising, speech and facial recognition, self-driving cars, genome sequencing, energy-efficient buildings, computer interactive games, language translation), and everything in our lives is digitally recorded (Schafer and Jin, 2014; Libbrecht and Noble, 2015; Sainath et al., 2015; Bang et al., 2018; Lopez-de-Ipina et al., 2018; Stilgoe, 2018; Chan and Siegel, 2019; Gao et al., 2019; Gu et al., 2019; Wan et al., 2019; Sajjad et al., 2020; Shahamiri, 2021; Yeong et al., 2021). Data are the new DNA of the 21st century carrying important knowledge, insights, and potential, becoming an intrinsic constituent of all data-driven organisms. Extracting information from data can be used to create various smart applications in different fields, such as science, healthcare, manufacturing, education, financial modeling, cybersecurity, data governance, police and marketing (Sarker, 2021b). Therefore, data management tools that are able to extract useful insights from data quickly and intelligently are urgently needed.

Artificial intelligence (AI), and in particular, Machine Learning (ML), have progressed remarkably in recent years as key instruments to intelligently analyze such data and to develop the corresponding real-world applications (Koteluk et al., 2021; Sarker, 2021b). For instance, ML has emerged as the method of choice for developing practical software for

Peer review under responsibility of Xi'an Jiaotong University.
* Corresponding author.
E-mail address: raffaele.pugliese@nemolab.it (R. Pugliese).
[1] These authors contributed equally.

https://doi.org/10.1016/j.dsm.2021.12.002
Received 25 August 2021; Received in revised form 4 November 2021; Accepted 16 December 2021
Available online 23 December 2021

computer vision, speech recognition, and language processing (Cummins et al., 2018; Hegde et al., 2019; Le Glaz et al., 2021) (Fig. 1). The effect of Machine Learning has also been widely felt across industries dealing with data-intensive issues, such as consumer services, diagnosing failures in complex systems, and controlling supply chains (Schaeffer and Sanchez, 2020).

There has been a similarly broad range of effects across sciences, as ML methods could assist scientists in their discovery of cancer classification through DNA microarray analysis (Tan and Gilbert, 2003; Wang et al., 2005), or that of solving one of the greatest challenges of biology, namely that of determining the three-dimensional structure of a protein starting from the amino acid sequence (Hutson, 2019; Callaway, 2020; Senior et al., 2020; Wu and Xu, 2021). Furthermore, COVID-19 has forced the use of ML for the prediction of SARS-CoV-2 diagnosis based on symptoms and for discovering new candidate drugs and vaccines in silico (Keshavarzi Arshadi et al., 2020; Lalmuanawma and Hussain, 2020; John et al., 2021; Zoabi et al., 2021).

Generally, the fruitfulness and the efficiency of a ML solution depend on the nature and characteristics of data and the performance of the learning algorithms. For these reasons, a diverse array of ML algorithms (*e.g.*, supervised, unsupervised, semi-supervised, and reinforced) has been developed to cover the wide variety of data across different ML problems (Das et al., 2015; Dey, 2016). As a matter of fact, although the past decade has seen increasing interest in these fields, we are just beginning to scratch the surface of the potential of ML algorithms for studying systems that improve with experience. Unfortunately, as reported by Jordan and Mitchell (2015), many questions regarding (1) the precision with which the algorithm can learn from a particular type and volume of data, (2) the algorithm's robustness with respect to errors in its modeling assumptions or errors in training data, (3) the possibility of designing an effective and successful algorithm based on a given learning problem is still open.

Furthermore, the privacy protection that derives from the data processed by ML algorithms must not be neglected. Indeed, trust and transparency are core problems when dealing with personal, and potentially sensitive information, especially when the algorithms in place are hard or even impossible to understand. This can be a major risk for acceptance, not only by end users, but also by experienced engineers who are required to train models (Holzinger et al., 2016, Kieseberg et al., 2016, Holzinger et al., 2018). To cope with these challenges, Google recently proposed the federal learning as a possible solution (Konečný et al., 2016).

On the other hand, it is also true that some gaps have been resolved. For instance, until a few years ago, research on AI was commonly divided into technological concerns (connected to natural sciences and engineering) and social concerns (connected to social sciences and humanities). These two strands were finally connected, which is an important aspect that has finally been overcome because this technology cannot be approached as a neutral object and cannot be separated from the social things. Indeed, as reported by Emma Dahlin: "to better understand AI and ML technology in the context in which it operates, the inseparability of these two concerns needs to be reflected in AI and ML research" (Wanzirah et al., 2015). As well as, to create truly acceptable systems, it is equally important that the social attention is given to the human-AI interaction, thus to improve explainability, transparency, and the black-boxed to their intended users (Castelvecchi, 2016).

For all these reasons, here will be provided a detailed summary on geo worldwide trends (taking into account China, the USA, Israel, Italy, the UK, and the Middle East) of ML-based approaches that can be applied to enhance the intelligence and the capabilities of an application. Furthermore, based on the literature review, we will discuss the potential research directions, challenges, and regulatory standpoints in this field. We hope this review will help and inspire the scientific community and industry professionals to push forward the potential development of ML-based approaches.

2. Types of machine learning techniques

ML involves the development and deployment of algorithms that, rather than being programmed to assign certain outputs (namely actions) in response to specific inputs from the environment, analyze the data and

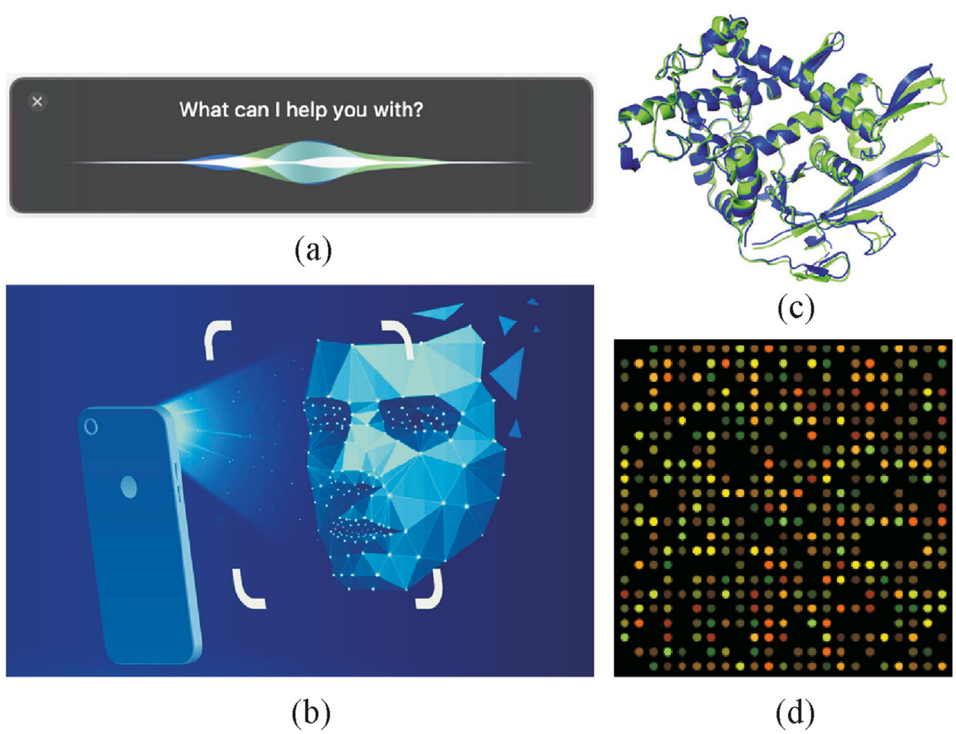

Fig. 1. Real-world applications of ML. ML is having substantial effects on many areas of technology and science; examples of recent applied success stories include: a) speech processing; b) facial recognition; c) determination of the final three-dimensional structure of proteins; d) cancer classification through DNA microarray analysis.

its properties, and determine the action by using statistical tools. Usually, ML algorithms are dynamic and improve or "learn" as more data is introduced (Duda et al., 2001; Bishop, 2006). ML algorithms can be broadly classified into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning, as depicted in Fig. 2.

2.1. Supervised learning

Supervised learning relies on ML tasks to learn a function that maps an input to an output based on sample input-output pairs. Hence, this learning process is based on comparing the calculated output and predicted output, that is, learning refers to computing the error and adjusting the error for achieving the expected output. Examples of such algorithms include Naïve Bayes classification, linear and logistic regression, Support Vector Machines (SVMs) (Table 1). Examples of applied supervised learning are automatic answering of incoming messages (useful in case of large companies), or face recognition which is useful as security measures at an ATM, surveillance areas, closed circuit cameras, criminal justice system, and image tagging in social networking sites like Facebook. Another prominent example of supervised learning is the epileptic seizure detection able to construct patient-specific detectors capable of detecting seizure onsets quickly and avoiding the risk of sustaining physical injuries and death with high accuracy (Kharbouch et al., 2011). The authors classified a feature vector as representative of seizure or non-seizure activity using a Support-Vector Machine (SVM). Since the seizure and non-seizure classes are often not linearly separable, they generated non-linear decision boundaries using an RBF kernel.

2.2. Unsupervised learning

Unsupervised learning analyzes unlabeled datasets without human interference. In unsupervised learning, the algorithm optimally separates the samples into different classes on the basis of the features of the training data alone, without corresponding labels. The unsupervised algorithms are *k*-means clustering, principal component analysis, and autoencoders. The most common example of unsupervised learning is the automatic identification of friends for a user in social media channels like Facebook or Google, or the discerning of the maximum number of mails sent to a particular person and categorized into collective groups. Furthermore, computational biology (also known as bioinformatics) is developing unsupervised algorithms from biological data to establish relations among various biological systems, collecting lots of useful data about gene sequences, DNA sequences, and gene expression, thus providing a much better understanding of the human genome (Mahendran et al., 2020; Ledesma et al., 2021). Bayesian networks, neural trees, and radial basis function (RBF) networks are used for the analysis of these datasets.

2.3. Semi-supervised learning

Semi-supervised learning can be defined as a hybridization of the above-mentioned supervised and unsupervised methods, as it employs both labeled and unlabeled data (Zhou and Belkin, 2014). The ultimate goal of a semi-supervised learning model is to provide a better outcome for prediction than that produced using the labeled data alone from the model. Such a method is widely used in machine translation, fraud detection, labeling data, and text classification.

2.4. Reinforcement learning

Reinforcement learning lies on a family of algorithms that typically operate sequentially to automatically evaluate the optimal behavior in a particular environment to improve its efficiency, i.e., an environment-driven approach (Buşoniu et al., 2010). At each step, a reinforcement algorithm, also referred as "agent", acts and predicts the features at a future step on the basis of past and present features, and a reward or penalty is assigned on the basis of the prediction. Therefore, it is a

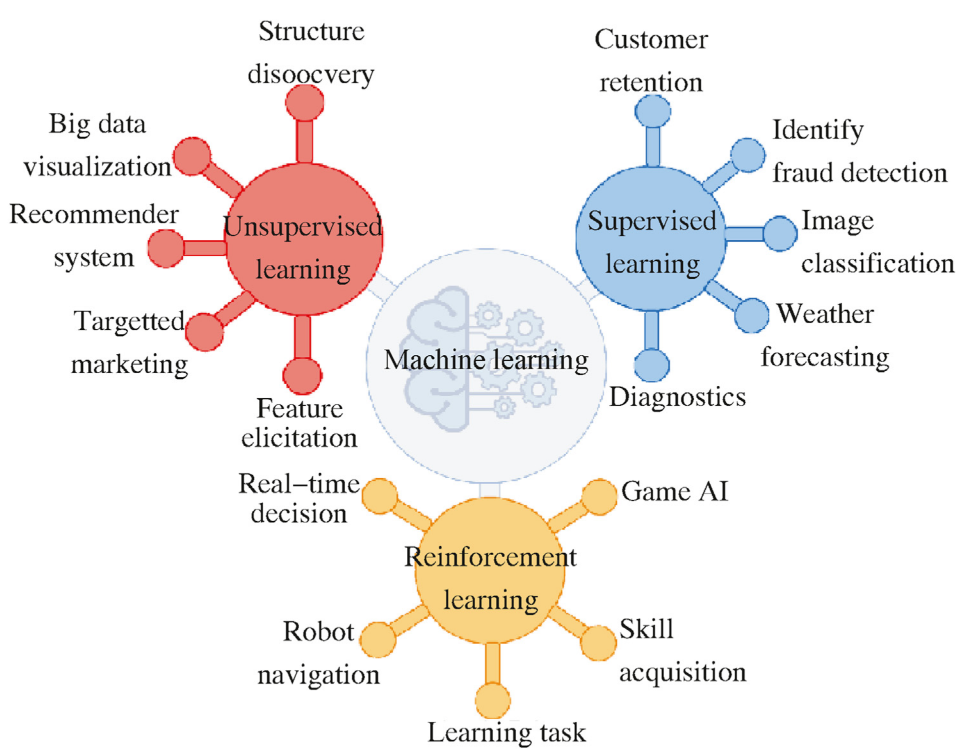

Fig. 2. Machine learning algorithms classification. Classification of the main machine learning techniques, namely supervised learning, unsupervised learning, and reinforcement learning with some examples.

Table 1
Types of machine learning algorithms and real-world application examples.

Learning type	Description	Model building	Examples	Pros & cons
Supervised learning	Data are labeled with classes and outcomes (task-driven approach).	Models learn from labeled data. This type of learning comprises classification, regression, Naïve Bayes classification, random forests, and neural networks.	E-mail data (i.e., automatic answering of incoming messages, mail organization into folders, spam filtering); face and speech recognition; information retrieval; Data Center Optimization.	Pros: exact idea about the classes in the training data; find out exactly how many classes are there before giving the data for training; helpful in classification problems; predict a numerical target value from given data and labels. Cons: limited in a variety of sense so that it cannot handle some of the complex tasks in machine learning; it cannot cluster or classify data by discovering its features on its own; training and classification need a lot of computation time, especially if the data set is very large, which will test the machine's efficiency.
Unsupervised learning	Data are not labeled. The algorithm optimally separates the samples into different classes on the basis of the features of the training data alone, without corresponding labels (data-driven approach).	Models learn from unlabeled data. This type of learning comprises *k*-means clustering, hierarchical clustering, and Principal Component Analysis (PCA).	Organizing large computer clusters; social network analysis; DNA classification; computational biology; market segmentation.	Pros: It can detect what human eyes cannot understand; the potential of hidden patterns can be powerful for the business as it can detect fraud detection; output can determine the unexplored territories and new ventures for businesses; the outcome of an unsupervised task can yield an entirely new business vertical or venture. Cons: It is expensive as it might require human intervention to understand the patterns and correlate them with the domain knowledge; it is not always certain that the obtained results will be useful since there is no label or output measure to confirm its usefulness; it is heavily dependent on the model and in turn, on the machine, and the results often have lesser accuracy.
Semi-supervised learning	The algorithm works with labeled and unlabeled data.	Models are built using combined labeled and unlabeled data. This type of learning includes both classification and clustering.	Text document classifier; text filtering; semantic scene classification; fraud detection.	Pros: It is easy to understand; it reduces the amount of annotated data used. Cons: Iteration results are not stable; it is not applicable to network-level data; it has low accuracy.
Reinforcement learning	The algorithm operates sequentially to automatically evaluate the optimal behavior in a particular context or environment to improve its efficiency (environment-driven approach).	Models are based on reward or penalty. This type of learning uses classification.	Traffic forecasting service; computer games; machinery applications; autonomous driving tasks; medicine; surgery.	Pros: It can be used to solve complex problems that cannot be solved by other techniques; it is preferred to achieve long-term results, which are very difficult to achieve; its model is very similar with the learning of human beings; the model can correct the errors that occurred during the training process. Cons: Too much reinforcement learning can lead to an overload of states, which can diminish the results; it needs a lot of data and a lot of computation.

powerful tool for training AI models that could optimize the operational efficiency of sophisticated systems, such as robotics, autonomous driving tasks, manufacturing, and supply chains. The reinforcement algorithms are TD (lambda) with function approximation, gradient temporal difference learning, and least-squares method. An outstanding example of an application of reinforced learning is the algorithm that can automatically generate the appropriate tension and optimal direction for a given cutting trajectory of either a laparoscopic surgeon or an automated cutting instrument. In addition, various papers have proposed the reinforcement learning for self-driving cars (AbuZekry et al., 2019, Rafi Omar Al-Nima et al., 2019; Cao et al. 2021), where reinforcement algorithms help to trajectory optimization, motion planning, dynamic paths, and controller optimization.

2.5. Federated learning

Google proposed the concept of federated learning in 2016 (Konečný et al., 2016). The main idea is to build ML models based on data sets that are distributed across multiple devices while preventing data leakage. Google prompted up such federated mechanisms as an effective solution to allow knowledge to be shared without compromising user privacy and security (Yang et al., 2019). Federated learning (also known as collaborative learning) is the ML technique that allows to train an algorithm through the use of decentralized devices or servers that hold data, without sharing them, thus addressing critical issues such as data privacy, data security, data access rights and access to heterogeneous data. This type of ML approach can be divided into centralized, decentralized and heterogeneous learning. In centralized federated learning methods, a central server is in charge of managing the different steps for the algorithms used and coordinating the all-participating nodes in the learning process. Moreover, the central server is responsible for choosing the nodes at the beginning of the process and aggregating the received model updates (Kairouz et al., 2021). In decentralized federated learning methods, nodes have the ability to coordinate themselves in order to achieve the global model. This technique allows overcoming the issue of centralized approaches since the nodes are able to exchange model updates without the coordination of a central server. Since an increasing number of application domains involve a large set of heterogeneous clients (i.e., mobile phones and IoT devices), recently, an heterogeneous federated learning framework (namely HeteroFL) was developed to address heterogeneous clients equipped with very different computation and communication capabilities (Diao et al., 2020). The HeteroFL technique enables the training of heterogeneous local models with dynamically varying computation complexities while still producing a single global inference model. Examples of such federated algorithms include deep neural networks, federated stochastic gradient descent (FedSGD), and federated averaging (FedAvg).

Overall, as reported by Yang et al., (2019), it is expected that in the near future, federated learning would break the barriers between industries and establish a community where data and knowledge could be shared together with safety. The benefits would be fairly distributed according to the contribution of each participant.

Thus, it is not surprising that by analyzing global datasets over the past five years (collected by Google Trends), real-world interest and application of reinforcement learning (Fig. 3, in green) have dramatically increased, peaking in 2019 with a popularity index of 63.5, compared to supervised learning (in red) and unsupervised learning (in blue), which have an index of 30.13 and 36.75, respectively. On the contrary, semi-supervised learning (in yellow) that uses either labeled or unlabeled data has not shown any growth (popularity index of 5.6 during the last five years).

Based on our knowledge, we believe that this growing interest in reinforcing algorithms is due to the fact that the latter, unlike supervised

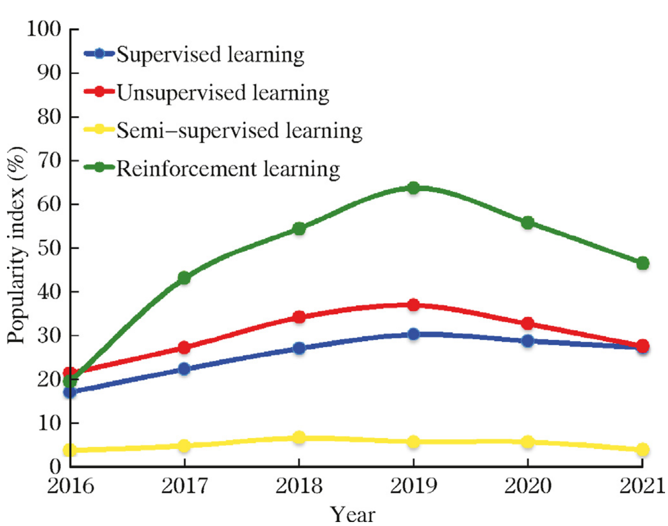

Fig. 3. Machine Learning popularity index. The worldwide popularity score of various types of machine learning algorithms (supervised, unsupervised, semi-supervised, and reinforcement) in a range of 0 (minimum) to 100 (maximum) over time of five years. All data were collected from Google Trends.

and unsupervised learning, is based on interaction with the environment, which can be used to solve different real-world problems in various fields, such as game theory, control theory, operations analysis, information theory, simulation-based optimization, manufacturing, supply chain logistics, swarm intelligence, aircraft control, robot motion control, laparoscopic surgery, traffic forecasting service, smart cities development, and so on.

3. Global trend: AI vs ML

Worldwide, ML is driving changes in technologies and their applications in the real world. According to Google Trends, the interest over time in "AI" (Fig. 4a) and "ML" (Fig. 4b) had a tremendous increase over the past five years. Although we are aware that such data do not reflect the full picture, the Google search highlights important insights in global AI and ML access across countries (i.e., Italy, China, the USA, Israel, the UK, and the Middle East).

Fig. 4a and b show the average values of the timestamp information in terms of years (x-axis) and the corresponding popularity in the range from 0 to 100 (y-axis). In particular, we observed that the indicative popularity values of these areas were around 30 in 2016 for Italy (in red), the UK (in cyan), and the USA (in yellow), while they exceeded 70 in 2020, which is more than double in terms of the increase in popularity. Instead, the popularity indication values of these technologies in China (in blue) and Israel (in green) remained almost unchanged over time with a popularity index of 35 and 38, respectively. Interestingly, in Saudi Arabia (in orange), the indicative popularity values of both AI and ML were under 6 in 2016 and reached 37 in 2020, six times more in terms of increased popularity.

Going deeper, we observed a greater push of use/popularity of the ML (average value 77%) compared to AI (average value of 23%) in all regions of the analyzed countries (Fig. 4 c–h). Furthermore, we noted that all countries recorded the highest popularity index value where universities, industries, research centers, start-ups, governments, etc. are concentrated, i.e., Massachusetts, Washington, and California in the USA; Lombardy, Liguria, and Lazio in Italy; Tel Aviv and Jerusalem in Israel; Beijing, Sichuan, Shanghai in China; England, Scotland, Ireland in the UK; Al-Sharqiyya, Al-riyad, and Mecca in Saudi Arabia (for further details see Table 2).

It can be summarized that the overall performances of Italy, China, the USA, Israel, the UK, and Saudi Arabia have demonstrated the rapid growth in AI and ML outputs in the last five years. This may be attributed to their substantial research foundation and evolution of ML technology, as well as their introduction of related national policies (Berkel et al., 2020). Undeniably, such occurrences may have also been found in other countries or regions, which have not been considered in this work.

Given this growing popularity of ML that has rapidly developed in recent years from a niche topic to one that has transformed the entire areas of technology, we wondered how "dissemination of research" in the ML scientific community is evolving and what the emerging trends and research directions are in this field.

To ensure integrity, as many publications as possible were selected for review in this work choosing from the following databases: PubMed, Web of Science, and ScienceDirect. Since ML appeared in the 1990s, all published documents (i.e., journal papers, reviews, conference papers, preprints, code repositories and more) related to this field from 1990 to 2020 have been selected, and specifically, within the search fields, the following keywords were used: "machine learning" OR "machine learning-based approach" OR "machine learning algorithms". Fig. 5a shows the number of papers about ML published each year since 1990, providing an overview of the history about ML-related publications. In the early 1990s and till 1998, there is a period of stagnancy in ML, since during this period, ML was mainly used for logistics, medical diagnostics, and industry (Graham et al., 1990; King and Sternberg, 1990; Kosko, 1990). Instead, from the early 21st century, ML research entered a period with numerous research outputs, with the first peak in 2016

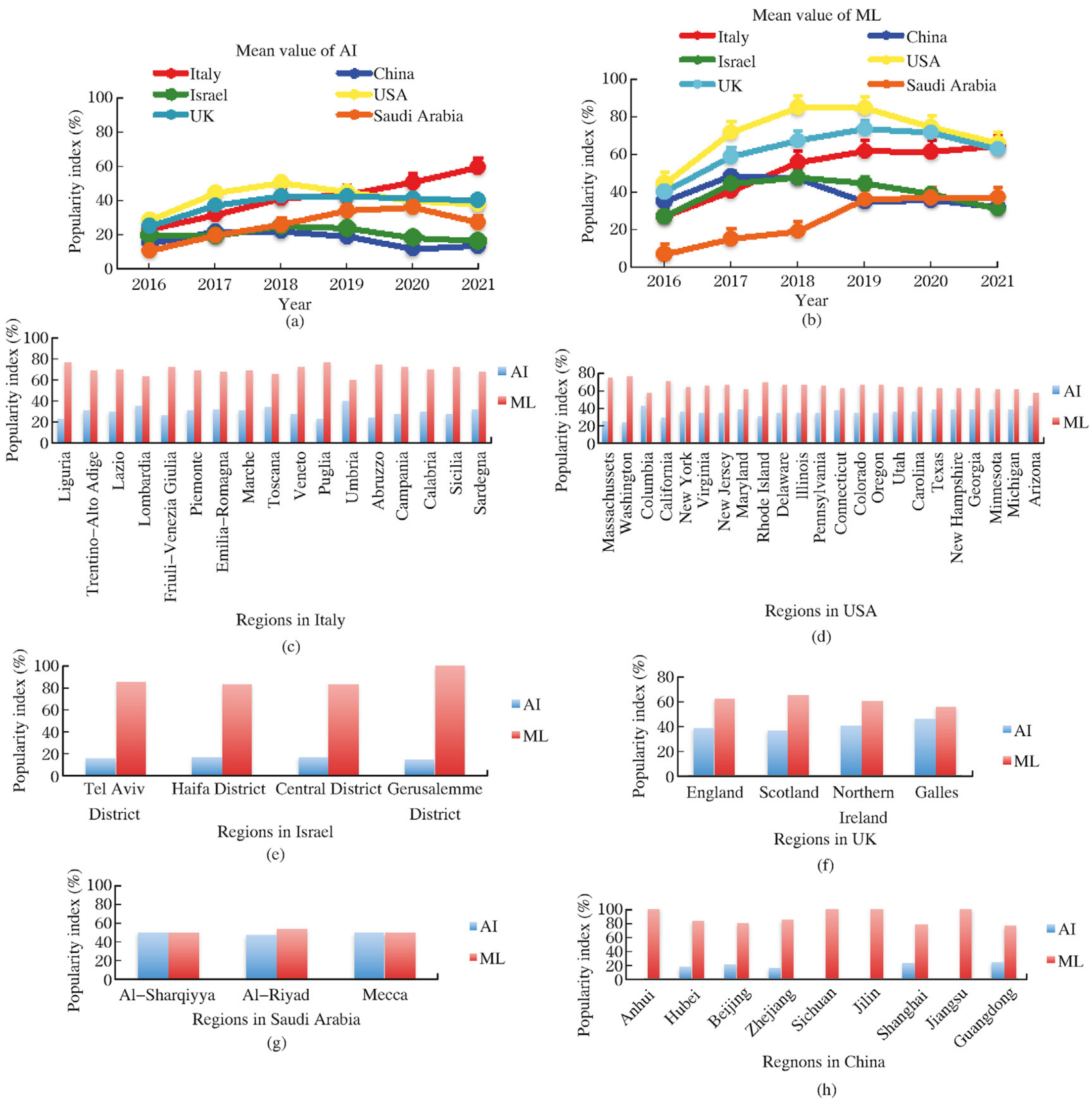

Fig. 4. Geo Trends of AI vs ML. Global popularity values of a) artificial intelligence and b) machine learning according to Google Trend, over the past five years. c–h) popularity interest by region of AI and ML in all analyzed countries.

with 3,886 publications. It is interesting to note that in the time range of 2018–2020, ML-related publications have undergone an exceptional growth reaching a value of 16,339 publications. Among these 16,339 publications, journal articles (14,272) and reviews (1,552) were the most common types, followed by letters (196) and clinical trials (129) as reported in Table 3.

The success of such dissemination of ML research could be attributed to several factors: ML is widely used across many fields, including medicine, manufacturing, education, logistics, financial, agriculture, nanotechnology, as well as the development of new learning algorithms theory, the ongoing explosion of "big data" and low-cost computation. Therefore, given this huge volume of available publications, based on the bibliographic data of 2020, we investigated the trends and the current situation of the ML literature (Fig. 5b). Among these publications, 59% of all articles were about medical field, 11% were related to engineering, 3% were related to agriculture and financial fields respectively, and 1% were about nanotechnology. Despite this growing interest in applying ML, there are still concerns regarding preventive measures to hinder disastrous and unknown consequences that are difficult to address. One prominent example is the loss of human life if an AI medical algorithm goes wrong, or the national security is threatened, if an adversary feeds disinformation to a military AI system, and so are significant challenges for organizations, from reputational damage and revenue losses to regulatory backlash, criminal investigation, and diminished public trust.

3.1. Medical field

In particular, we took the publications concerning medicine (Fig. 6a) into consideration. As the "future medicine" or preventive medicine has the task of anticipating disease, it not only relies on a better knowledge of the biological and molecular processes that underlie the development of pathologies, but also on the analysis of a large amount of data for the formulation of predictive algorithms. ML-based computer decision support systems have been found to be employed in cancer management, surgical interventions, cardiovascular disease treatment, pandemic prediction, and drug discovery, as they have the potential to perform complex tasks that are currently assigned to specialists to improve the diagnostic accuracy, increase the process efficiency, thereby improving clinical workflow, reducing human resource costs, and improving treatment choices (Goldenberg et al., 2019; Checcucci et al., 2020; Davoudi et al., 2021; Dong et al., 2021; Hirschprung and Hajaj, 2021; Nikolaou et al., 2021; Shuhaiber and Conte, 2021; Smole et al., 2021; Tunthanathip and Oearsakul, 2021; Zhan et al., 2021). In this application field, the main ML research methods are SVMs, Bayesian networks, neural trees,

Table 2
Summary of use/popularity of AI and ML in the regions of analyzed countries. All data were extrapolated using Google Trend.

Country	Region	AI (%)	ML (%)
Italy	Liguria	23	77
	Trentino-Alto Adige	31	69
	Lazio	30	70
	Lombardia	36	64
	Friuli-Venezia Giulia	27	73
	Piemonte	31	69
	Emilia-Romagna	32	68
	Marche	31	69
	Toscana	34	66
	Veneto	28	72
	Puglia	23	77
	Umbria	40	60
	Abruzzo	25	75
	Campania	28	72
	Calabria	30	70
	Sicilia	28	72
	Sardegna	32	68
China	Anhui	0	100
	Hubei	17	83
	Beijng	21	79
	Zhejiang	16	84
	Sichuan	0	100
	Jilin	0	100
	Shanghai	23	77
	Jiangsu	0	100
	Guangdong	24	76
Israel	Tel Aviv District	15	85
	Haifa District	17	83
	Central District	17	83
	Jerusalemme District	14	100
USA	Massachusetts	25	75
	Washington	24	76
	Columbia	43	57
	California	29	71
	New York	36	64
	Virginia	35	65
	New Jersey	34	66
	Maryland	39	61
	Rhode Island	31	69
	Delaware	34	66
	Illinois	34	66
	Pennsylvania	35	65
	Connecticut	37	63
	Colorado	34	66
	Oregon	34	66
	Utah	36	64
	Hawaii	41	59
	Florida	48	52
	Wyoming	58	42
	Nevada	48	52
	Arkansas	46	54
UK	England	38	62
	Scotland	36	64
	Northern Ireland	40	60
	Galles	45	55
Saudi Arabia	Al-Sharqiyya	50	50
	Al-Riyad	47	53
	Mecca	50	50
	Medina	0	100

and radial basis function (RBF) networks, classification, regression, clustering, and principal component analysis (PCA).

3.2. Financial field

Attention to governance, security administration, human rights, and intellectual property is also rapidly growing. Indeed, a series of academic studies emerge in such disciplines (Fig. 6b). In this field, investing in ML offers tremendous benefits, as it has the potential to help organizations work efficiently, manage costs, and make dramatic progress in decision quality.

It should be noted that the New York University and Stanford University have published the report "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies" (Engstrom et al., 2020) commissioned by the Administrative Conference of the United States (ACUS). The research was conducted by a working group composed of jurists, computer scientists, and social scientists on a sample of 142 US federal agencies and bodies in order to map the current use of AI and ML technologies in the administrative field and identify possible lines of development. According to the report, AI and ML promise to transform the way how government agencies do their work, even if they have to face privacy and security issues, compatibility with legacy systems and evolving workloads, as well as the correct design of algorithms and user interfaces, and the boundaries between public actions and private procurement. Overall, the authors point out that rapid developments in AI and ML have the potential to reduce the cost of core governance functions, improve decision quality, and unleash the power of administrative data, thereby making government performance more efficient and effective.

3.3. Cybersecurity field

Cybersecurity is another hot topic that is receiving enormous attention due to the growing reliance on the Internet of Things (Li et al., 2015). cybercrime, malware attack, data breach, etc., are causing devastating financial losses not only to organizations and industries, but to individuals as well. It is estimated that the global annual cybercrime cost 400 billion USD (Fischer, 2014). Therefore, to address this issue, numerous researchers are developing ML techniques to build cybersecurity models useful for detecting and protecting data, with minimal human intervention. Through the literature, we observed a significant increase in publications concerning the cybersecurity area in the last ten years, reaching a total of 1,268 scientific publications (Fig. 6c).

Among those, Sarker and co-workers have reported an ML-based approach (namely Intrusion Detection Tree - "IntruDTree"), which is capable of detecting cyber intrusions by first classifying security features based on their importance, and then building a tree-based generalized intrusion detection model based on selected important characteristics (Sarker et al., 2020). The authors showed that "IntruDTree" is effective in terms of forecast accuracy by conducting a range of experiments on cybersecurity datasets, thus minimizing the security issues and reducing the computational cost and time.

For readers who are particularly interested in this research field, we recommend the recent literature reviews (Handa et al., 2019; Suryotrisongko and Musashi, 2019; Meng, 2019; Sarker et al., 2021; Dixit and Silakari, 2021; Sarker, 2021a).

3.4. Nanotechnology field

Nanotechnologies are no longer just a buzzword in the field of materials science, but rather a tangible reality. Just think that today there are more than 3,000 different types of commercial nanoproducts available all over the world in different sectors (RamosCampos, 2021). This reflects that nanotechnology is everywhere and is a daily practice. To emphasize, the fact that we have two vaccines that use nanoparticles carrying mRNA to produce SARS-CoV-2 viral proteins (i.e., Pfizer/BioNTech and Moderna) has destroyed all the skepticism about the use of nanoproducts in humans and this marks a new era in nanotechnology and nanomedicine.

Along this line, there is a growing interest in the use of ML techniques for predictive modeling and the design of nanoproducts (Fig. 6d). As reported by Talebian et al. (2021), ML can boost and reshape the *de novo* design of nanodelivery systems, thus generating new challenges for the next generation of smart drugs. Furthermore, Whitelam and Tamblyn (2021) developed an ML algorithm based on the principles of evolution

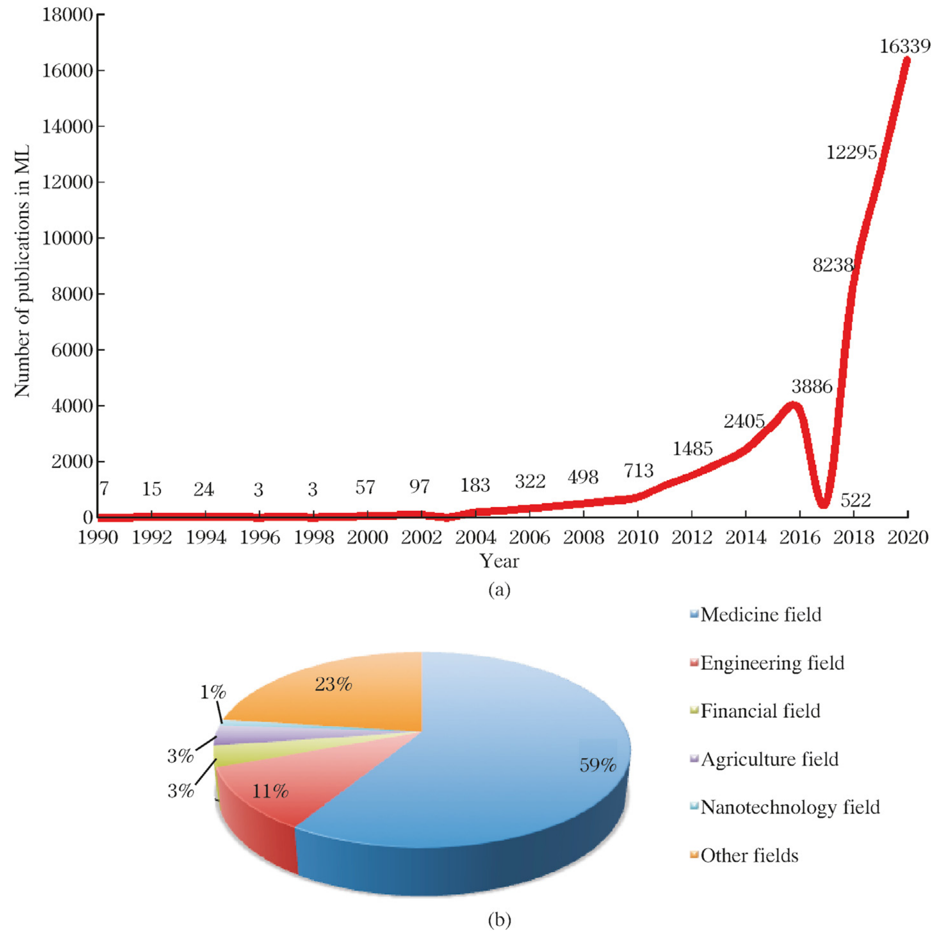

Fig. 5. Temporal evolution of machine learning-related publications. a) Temporal patterns of machine learning-related articles; b) relative percentage estimation in 2020.

Table 3
Number of papers related to machine learning published in 2020.

Types of publications	Number of publications
Journal article	14,272
Review	1,552
Book	5
Clinical trial	129
Meta analysis	54
Case report	9
Letter	196
Preprints	42
Other	80

for molecular simulations to self-assemble nanomaterials with user-defined properties. In order to allow a thorough exploration of the self-assembly behavior accessible to this class of model, the authors expressed the inter-particle potential and time-dependent assembly protocol as arbitrary functions, encoded by neural networks, and optimized via evolutionary methods. The authors showed that this evolutionary approach furthering progress toward automated materials discovery or "synthesis by design", which is a challenging problem that requires significant human input and trial and error.

3.5. *Agriculture field*

ML techniques are also bringing new opportunities in agricultural production systems, and the scientific research related to this sector is increasing exponentially (Fig. 6e). As reported by Liakos et al. (2018), and by Benos et al. (2021), ML algorithms in farm management systems provide insightful advice and information on (1) crop management, (2) yield prediction, (3) the identification of possible diseases and weed species, (4) livestock management and welfare, (5) water and soil management, (6) the level of soil moisture, seeding and harvesting dates, and the phenostages of crops.

Such Technologies are useful in the agricultural sector, as they help farmers optimize their operations, improve their crops, and increase profitability, even in the midst of challenges such as climate change, over-cultivation and pollution, thus creating a smart and sustainable agro-technology sector, which leads to more accurate and faster decision-making, and improves today's agricultural practices to feed the growing global population in the future as well.

Also in this case, for readers who are particularly interested in this research field, we recommend the recent literature reviews (Balducci, 2018; Liakos et al., 2018; Storm et al., 2020; Sader et al., 2020; Benos et al., 2021; Liu, 2021; Zhao, 2021).

4. Regulatory standpoint and challenges

In recent years, ML and AI have also come under the spotlight of authorities and law scholars around the world, which are willing to create a regulatory environment capable of balancing the needs for protecting, *inter alia*, consumers harmed by ML and/or AI tools.

The nature of ML and AI raises challenging issues for law scholars who are making efforts to outline the main features of such environments, including "opaqueness" (an external observer may not be able to identify the potentially harmful characteristics of ML and AI) and

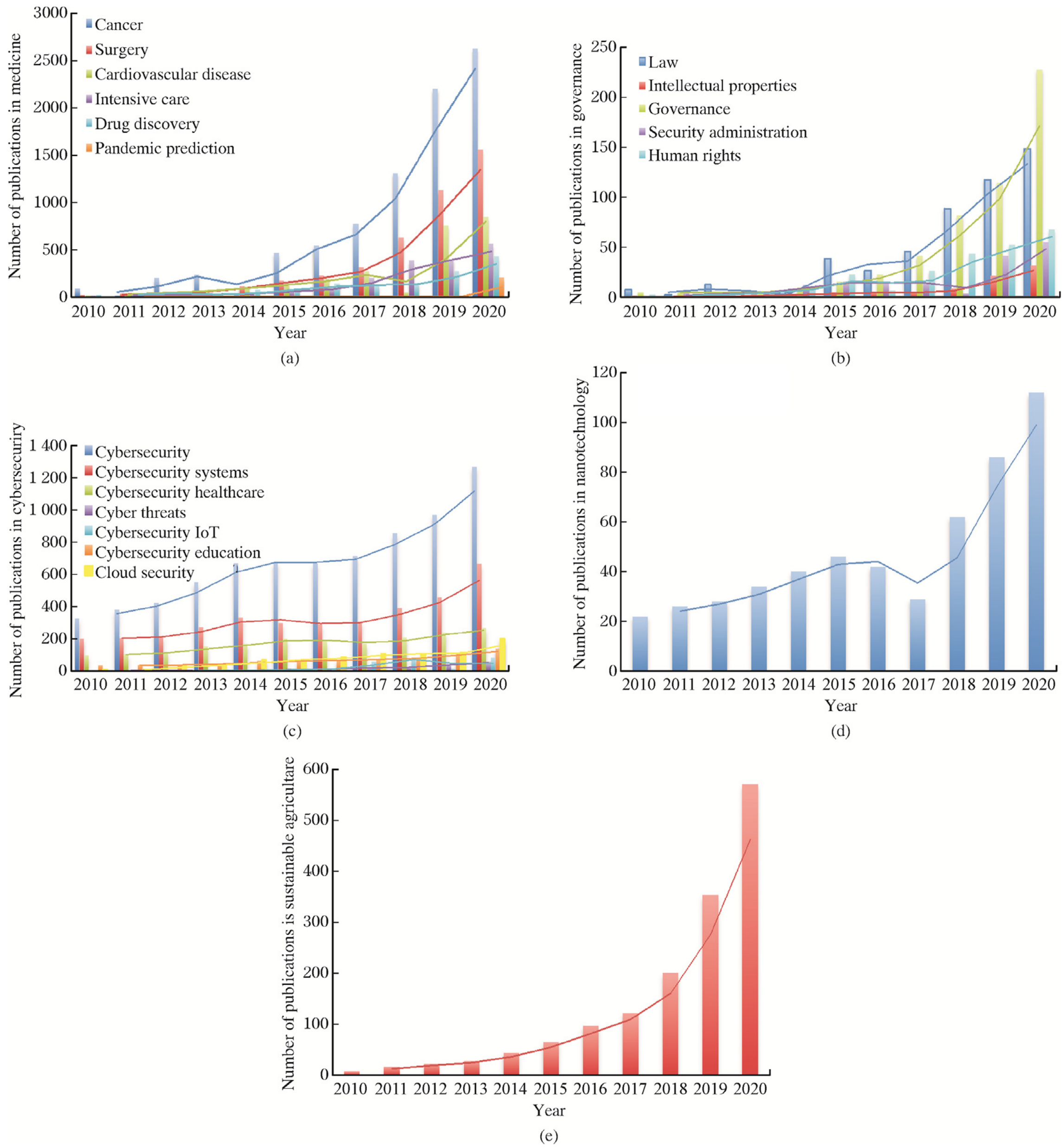

Fig. 6. Growth trend of ML articles in a) medicine, b) governance, c) cybersecurity, d) nanotechnology, and e) sustainable agriculture.

"unforeseeability" (ML and AI learn from "their experience" and, as a consequence, their "conduct" is potentially unpredictable). Such specific features make it particularly complex to establish effective rules (Scherer, 2016; Sun et al., 2021).

When attempting to regulate the matter at hand, the first challenge is to provide a correct and flexible definition of AI and ML (White Paper on Artificial Intelligence: a European approach to excellence and trust, 2020). Unfortunately, there is not an accepted definition of AI that could be used from a regulatory standpoint, probably because, as stated by McCarthy, there is not a definition of intelligence that does not depend on its relation with human intelligence (McCarthy, 2007). Nevertheless, it is uncontroversial that a clear definition of AI and ML is necessary to create an effective legal framework. Such a need for a definition has been recently acknowledged at the European level, as expressed in the proposal for a regulation of the European Parliament and of the Council "Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts" (EU Commission, 2020). In the proposal, an AI system is defined as a "software that is developed with one or more of the techniques and approaches and can, for a given set of human-defined objectives, generate outputs, such as content, predictions, recommendations, or decisions influencing the environments they interact with". It should be noted that the definition proposed by the EU Commission is just one of the possible definitions of AI (the Commission reserves the right to amend such a definition in order to keep it up to date with AI developments – see also Art. 4 of the proposal) and that, in fact, more than 50 definitions of AI have been proposed over the last 50 years.

The challenges in providing a definition could be considered an example of the magnitude of the issues connected to ML and AI. Also, lawmakers shall face the matter of "liability" with respect to ML and AI. In fact, the aforementioned "opaqueness" and "unforeseeability" make it difficult to establish who should be responsible for the damages caused by ML or AI tools. As mentioned above, their conduct is potentially unpredictable and, sometimes, unavoidable. Moreover, the liability issue should also be carefully addressed in light of the fact that ML and AI tools could be applied with reference to high-risk activities (*e.g.*, self-driving cars, medical/paramedical tools, etc.) that could cause serious damages to final users. In other words, the legal framework should be capable of guaranteeing that final users are given a way to compensate for any damages suffered. Despite the fact that each country has its own rules and policies on liability for damages, these should probably be amended and adapted to such particular "products" that are capable of self-learning. Several countries are dealing with this issue, and therefore different approaches could be taken (strict liability or fault-based approach, which could even imply, as suggested by certain scholars, a direct responsibility of AI and ML for damages).

The European Union, the US, China, etc. are tackling the question, to a certain extent, in different ways. (Center for Data Innovation, 2017). A general approach to liability that seems to be acknowledged by scholars and lawmakers could possibly lead to classifying the risk involved in the application of AI and ML on a case-by-case basis (*e.g.,* the application of AI and ML in the medical field could be considered to involve a high risk), thereby establishing different rules and liabilities for each risk level (high risk could impose strict liability on operators, while a lower risk level could be subject to a fault-based approach). This view appears to be shared by the European Union Commission that, in its proposal, AI systems are classified according to certain risk levels (the Proposal mainly deals with high-risk AI systems, requires the compliance of such systems with certain requirements and lays down certain obligations and duties on providers). It should be noted that the classification could be made if ML and AI were "transparent" to the relevant authorities competent for assessing/verifying the relevant risk level.

As expressed, the purpose of this regulatory focus is to introduce some of the challenges that ML and AI will impose on scholars and lawmakers around the world, by describing two of the main issues among many (*e.g.,* ethical implications, data protection, privacy, etc.) in a fast-evolving field.

5. Conclusion

In this paper, we have conducted an overview of ML algorithms for intelligent data analysis and applications. We have briefly discussed how various types of ML methods can be used for making solutions to various real-world issues, highlighting that a successful ML technique depends on both the data and the performance of the learning algorithms. Then we turned to discuss the global interest of AI and ML across different countries such as Italy, China, the USA, Israel, the UK, and the Middle East, highlighting a greater push of ML use over the past five years, ascribable to the evolution of ML technology, to the "dissemination of research" in the ML scientific community, and the introduction of the relevant national policies regarding AI and ML. Finally, we summarized and discussed the potential research opportunities and future directions in the ML area, as well as the regulatory challenges to be faced.

Given the above analyses and research efforts done in this field, we believe that machine learning-based solutions are opening up a promising direction worldwide and can be used as a reference guide for different real-world applications in both short and long terms, although we still have to pay close attention to the value and management of the available data.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

References

AbuZekry, A., et al., 2019. Comparative study of NeuroEvolution algorithms in reinforcement learning for self-driving cars. Eur. J. Eng. Sci. Technol. 2 (4), 60–71.
Balducci, F., et al., 2018. Machine learning applications on agricultural datasets for smart farm enhancement. Machines 6 (3), 38.
Bang, J., et al., 2018. Adaptive data boosting technique for robust personalized speech emotion in emotionally-imbalanced small-sample environments. Sensors 18 (11), 3744.
Benos, L., et al., 2021. Machine learning in agriculture: a comprehensive updated review. Sensors 21 (11), 3758.
Berkel, N.V., Papachristos, E., Giachanou, A., et al., 2020. A systematic assessment of national artificial intelligence policies: perspectives from the Nordics and beyond. ACM digital library. Available at: https://doi.org/10.1145/3419249.3420106.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer-Verlag, New York.
Buşoniu, L., Babuka, R., Schutter, B.D., 2010. Multi-agent Reinforcement Learning: An Overview. Innovations in Multi-Agent Systems and Applications - 1. Springer, Heidelberg, pp. 183–221.
Callaway, E., 2020. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. Nature 588 (7837), 203–204.
Cao, Z., et al., 2021. Confidence-aware reinforcement learning for self-driving cars. IEEE Trans. Intell. Transport. Syst. PP(99), 1–12.
Castelvecchi, D., 2016. Can we open the black box of AI? Nature 538 (7623), 20–23.
Center for Data Innovation, 2017. How government are preparing for artificial intelligence. Available at: https://datainnovation.org/2017/08/how-governments-are-preparing-for-artificial-intelligence/.
Chan, S., Siegel, E.L., 2019. Will machine learning end the viability of radiology as a thriving medical specialty? Br. J. Radiol. 92 (1094), 20180416.
Checcucci, E., Autorino, R., Cacciamani, G.E., 2020. Artificial intelligence and neural networks in urology: current clinical applications. Minerva Urol. Nefrol. 72 (1), 49–57.
Cummins, N., Baird, A., Schuller, B.W., 2018. Speech analysis for health: current state-of-the-art and the increasing impact of deep learning. Methods 151 (Dec.), 41–54.
Das, S., Dey, A., Pal, A., Roy, N., 2015. Applications of artificial intelligence in machine learning: review and prospect. Int. J. Comput. Appl. 115 (9), 31–41.
Davoudi, A., Ahmadi, M., Sharifi, A., et al., 2021. Studying the effect of taking statins before infection in the severity reduction of COVID-19 with machine learning. BioMed Res. Int. 2021 (1), 9995073.
Dey, A., 2016. Machine learning algorithms: a review. Int. J. Comput. Sci. Inf. Technol. Available at: https://dx.doi.org/10.21275/ART20203995.
Diao, E., Ding, J., Tarokh, V., 2020. HeteroFL: computation and communication efficient federated learning for heterogeneous clients arXiv: 2010.01264.
Dixit, P., Silakari, S., 2021. Deep learning algorithms for cybersecurity applications: a technological and status review. Comput. Sci. Rev. 39 (4), 100317.
Dong, M., Tang, C., Ji, J., et al., 2021. Transmission trend of the COVID-19 pandemic predicted by dendritic neural regression. Appl. Soft Comput. 111 (14), 107683.
Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification. Wiley, New Jersey.
Engstrom, D.F., Ho, D.E., Sharkey, C.M., et al., 2020. Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. NYU School of Law, Public Law Research Paper. Available at: https://doi.org/10.2139/ssrn.3551505.
EU Commission, 2020. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. COM. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.
Fischer, E., 2014. Cybersecurity Issues and Challenges: in Brief. Congressional Research Service, Washington D.C.
Gao, W., Alsarraf, J., Moayedi, H., et al., 2019. Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. Appl. Soft Comput. 84 (Nov.), 105748.
Goldenberg, S.L., Nir, G., Salcudean, S.E., 2019. A new era: artificial intelligence and machine learning in prostate cancer. Nat. Rev. Urol. 16 (7), 391–403.
Graham, A.R., Paplanus, S.H., Bartels, P.H., 1990. A diagnostic expert system for colonic lesions. Am. J. Clin. Pathol. 94 (4 Suppl. 1), S15–S18.
Gu, Y., Lyu, X., Sun, W., et al., 2019. Mutual correlation attentive factors in dyadic fusion networks for speech emotion recognition. Proc. ACM Int. Conf. Multimed. 2019 (Oct.), 157–166.
Handa, A., Sharma, A., Shukla, S.K., 2019. Machine learning in cybersecurity: a review. Data Min. Knowl. Discov. 9 (1), e1306.
Hegde, S., Shetty, S., Rai, S., et al., 2019. A survey on machine learning approaches for automatic detection of voice disorders. J. Voice. 33 (6), 947 e11–947 e33.
Hirschprung, R.S., Hajaj, C., 2021. Prediction model for the spread of the COVID-19 outbreak in the global environment. Heliyon. 7 (7), e07416.
Holzinger, A., et al., 2018. Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: from Machine Learning to Explainable AI. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction.
Holzinger, K., Mak, K., Kieseberg, P., et al., 2016. Can we trust machine learning results? Artificial intelligence in safety-critical decision support. ERCIM News. 112.
Hutson, M., 2019. AI protein-folding algorithms solve structures faster than ever. Nature. Available at: https://doi.org/10.1038/d41586-019-01357-6.
John, C.C., Ponnusamy, V., Chandrasekaran, S.K., et al., 2021. A survey on mathematical, machine learning and deep learning models for COVID-19 transmission and diagnosis. IEEE Rev. Biomed. Eng. Available at: https://doi.org/10.1109/RBME.2021.3069213.
Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349 (6245), 255–260.
Kairouz, P., McMahan, H.B., Avent, B., et al., 2021. Advances and open problems in federated learning arXiv: 1912.04977.
Keshavarzi Arshadi, A., Webb, J., Salem, M., et al., 2020. Artificial intelligence for COVID-19 drug discovery and vaccine development. Front. Artif. Intell. Available at: https://doi.org/10.3389/frai.2020.00065.
Kharbouch, A., Shoeb, A., Guttag, J., et al., 2011. An algorithm for seizure onset detection using intracranial EEG. Epilepsy Behav. 22 (Suppl. 1), S29–S35.

Kieseberg, A., Weippl, E., Holzinger, P., 2016. Trust for the doctor-in-the-loop. ERCIM News. Available at: https://ercim-news.ercim.eu/en104/special/trust-for-the-doctor-in-the-loop.
King, R.D., Sternberg, M.J., 1990. Machine learning approach for the prediction of protein secondary structure. J. Mol. Biol. 216 (2), 441–457.
Konečný, J., Mcmahan, H.B., Ramage, D., et al., 2016. Federated Optimization: Distributed Machine Learning for On-Device Intelligence arXiv: 1610.02527.
Kosko, B., 1990. Unsupervised learning in noise. IEEE Trans. Neural Network. 1 (1), 44–57.
Koteluk, O., Wartecki, A., Mazurek, S., et al., 2021. How do machines learn? Artificial intelligence as a new era in medicine. J. Personalized Med. 11 (1), 32.
Lalmuanawma, S., Hussain, J., 2020. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. Chaos Solit. Fractals. 139 (C), 110059.
Le Glaz, A., Haralambous, J., Chhakchhuak, L., 2021. Machine learning and natural language processing in mental health: systematic review. J. Med. Internet Res. 23 (5), e15708.
Ledesma, D., Symes, S., Richards, S., et al., 2021. Advancements within modern machine learning methodology: impacts and prospects for biomarker discovery. Curr. Med. Chem. 28 (32), 6512–6531.
Li, S., Li, D., Zhao, S., 2015. The internet of things: a survey. Inf. Syst. Front 17 (2), 243–259.
Liakos, K.G., Patrizia, B., Dimitrios, M., et al., 2018. Machine learning in agriculture: a review. Sensors 18 (8), 2674.
Libbrecht, M.W., Noble, W.S., 2015. Machine learning applications in genetics and genomics. Nat. Rev. Genet. 16 (6), 321–332.
Liu, Y., 2021. Intelligent analysis platform of agricultural sustainable development based on the Internet of Things and machine learning. Acta Agric. Scand. 71 (8), 718–731.
Lopez-de-Ipina, K., Martinez-De-Lizarduy, U., Ca Lvo, P., et al., 2018. Advances on automatic speech analysis for early detection of alzheimer disease: a non-linear multi-task approach. Curr. Alzheimer Res. 15 (2), 139–148.
Mahendran, N., Durai, R., Srinivasan, K., et al., 2020. Machine learning based computational gene selection models: a survey, performance evaluation, open issues, and future research directions. Front. Genet. 11, 1468.
McCarthy, J., 2007. What is Artificial Intelligence? Stanford University. Available at: http://www-formal.stanford.edu/jmc/whatisai.pdf.
Meng, X.-L., 2019. Data Science: an Articial Ecosystem. Harvard Data Science Review. Available at: https://doi.org/10.1162/99608f92.ba20f892.
Nikolaou, V., Massaro, S., Garn, W., et al., 2021. The cardiovascular phenotype of Chronic Obstructive Pulmonary Disease (COPD): applying machine learning to the prediction of cardiovascular comorbidities. Respir. Med. 186 (Sep.), 106528.
Rafi Omar Al-Nima, R., Han, T., Chen, T., 2019. Road Tracking using Deep Reinforcement Learning for Self-driving Car Applications. Progress in Computer Recognition Systems. Springer, Berlin, pp. 106–116.
RamosCampos, E.V., 2021. Commercial nanoproducts available in world market and its economic viability. Adv. Nano-Fert. Nano-Pestic. Agric. Available at: https://doi.org/10.1016/B978-0-12-820092-6.00023-9.
Sader, S., Husti, I., Daróczi, M., 2020. Enhancing failure mode and effects analysis using auto machine learning: a case study of the agricultural machinery industry. Processes 8 (2), 224.
Sainath, T.N., Kingsbury, B., Saon, G., et al., 2015. Deep convolutional neural networks for large-scale speech tasks. Neural Network. 64 (Apr.), 39–48.
Sajjad, M., Khan, S.U., Khan, N., et al., 2020. Towards efficient building designing: heating and cooling load prediction via multi-output model. Sensors. 20 (22), 6419.
Sarker, I.H., 2021a. Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. SN Comput. Sci. 2 (5), 377.
Sarker, I.H., 2021b. Machine learning: algorithms, real-world applications and research directions. SN Comput. Sci. 2 (3), 160.
Sarker, I.H., et al., 2021. Cybersecurity data science: an overview from machine learning perspective. J. Big Data 7 (1), 1–29.
Sarker, I.H., Abushark, Y.B., Alsolami, F., et al., 2020. IntruDTree: a machine learning based cyber security intrusion detection model. Symmetry 12 (5), 754.
Schaeffer, S.E., Sanchez, S.V.R., 2020. Forecasting client retention — a machine-learning approach. J. Retailing Consum. Serv. 52 (Jan.), 101918.1–101918.9.
Schafer, P.B., Jin, D.Z., 2014. Noise-robust speech recognition through auditory feature detection and spike sequence decoding. Neural Comput. 26 (3), 523–556.
Scherer, M.U., 2016. Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. Harv. J. Law Technol. 29 (2), 354–400.
Senior, A.W., Evans, R., Jumper, J., et al., 2020. Improved protein structure prediction using potentials from deep learning. Nature 577 (7792), 706–710.
Shahamiri, S.R., 2021. Speech vision: an end-to-end deep learning-based dysarthric automatic speech recognition system. In: IEEE Trans Neural Syst Rehabil Eng. IEEE, pp. 852–861.
Shuhaiber, J.H., Conte, J.V., 2021. Machine learning in heart valve surgery. Eur. J. Cardio. Thorac. Surg. Available at: https://doi.org/10.1093/ejcts/ezab324.
Smole, T., Unkovi, B., Piulin, M., et al., 2021. A machine learning-based risk stratification model for ventricular tachycardia and heart failure in hypertrophic cardiomyopathy. Comput. Biol. Med. 135 (Aug.), 104648.
Stilgoe, J., 2018. Machine learning, social learning and the governance of self-driving cars. Soc. Stud. Sci. 48 (1), 25–56.
Storm, H., Baylis, K., Heckelei, T., 2020. Machine learning in agricultural and applied economics. Eur. Rev. Agric. Econ. 47 (3), 849–892.
Sun, L., Zhang, H., Fang, C., et al., 2021. Data security governance in the era of big data: status, challenges,and prospects. Data Sci. Manag. 2 (Jun.), 41–44.
Suryotrisongko, H., Musashi, Y., 2019. Review of cybersecurity research topics, taxonomy and challenges: interdisciplinary perspective. In: 2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA). IEEE, pp. 162–167.
Talebian, S., Rodrigues, T., Neves, J.D., et al., 2021. Facts and figures on materials science and nanotechnology progress and investment. ACS Nano 15 (10), 15940–15952.
Tan, A.C., Gilbert, D., 2003. Ensemble machine learning on gene expression data for cancer classification. Appl. Bioinf. 2 (3 Suppl. l), S75–S83.
Tunthanathip, T., Oearsakul, T., 2021. Application of machine learning to predict the outcome of pediatric traumatic brain injury. Chin. J. Traumatol. 24 (6), 350–355.
Wan, N., Weinberg, D., Liu, T.Y., et al., 2019. Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. BMC Cancer 19 (1), 832.
Wang, Y., Tetko, I.V., Hall, M.A., et al., 2005. Gene selection from microarray data for cancer classification–a machine learning approach. Comput. Biol. Chem. 29 (1), 37–46.
Wanzirah, H., Tusting, L.S., Emmanuel, A., et al., 2015. Mind the gap: house structure and the risk of malaria in Uganda. PLoS One 10 (1), e0117396.
White Paper on Artificial Intelligence: A European Approach to Excellence and Trust 2020. European Commission.
Whitelam, S., Tamblyn, I., 2021. Neuroevolutionary learning of particles and protocols for self-assembly. Phys. Rev. Lett. 127 (1), 018003.
Wu, F., Xu, J., 2021. Deep template-based protein structure prediction. PLoS Comput. Biol. 17 (5), e1008954.
Yang, Q., Liu, Y., Chen, T., et al., 2019. Federated Machine Learning: Concept and Applications. arXiv, 1902.04885.
Yeong, J., Velasco-Hernandez, G., Barry, J., et al., 2021. Sensor and sensor fusion technology in autonomous vehicles: a review. Sensors. 21 (6), 2104.
Zhan, X., Humbert-Droz, M., Mukherjee, P., et al., 2021. Structuring clinical text with AI: old versus new natural language processing techniques evaluated on eight common cardiovascular diseases. Patterns (N Y) 2 (7), 100289.
Zhao, H., 2021. Futures price prediction of agricultural products based on machine learning. Neural Comput. Appl. 33 (7), 837–850.
Zhou, X., Belkin, M., 2014. Semi-Supervised Learning. Academic Press Library in Signal Processing. Elsevier, Amsterdam, pp. 1239–1269.
Zoabi, Y., Deri-Rozov, S., Shomron, N., 2021. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit. Med. 4 (1), 3.