THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 3; ISSUE 1 (JAN-JUNE); (2017)
WEBSITE: THE COMPUTERTECH

Harnessing the Potential of Artificial Intelligence and Big Data in Healthcare

Divya Sai Jaladi[1*], Sandeep Vutla[2]

[1]Senior Lead Application Developer, SCDMV, 10311 Wilson Boulevard, Blythewood, SC 29016, UNITED STATES

[2]Assistant Vice President, Senior-Data Engineer, Chubb, 202 Halls Mill Rd, Whitehouse Station, NJ 08889, UNITED STATES

*Corresponding Author: divyasaij26@gmail.com

Abstract

Data-driven research, together with its implications in machine learning and the broader domain of artificial intelligence, have the capacity to instigate significant transformations in medicine. Nevertheless, medicine is not a science akin to others. It is intricately and firmly connected to an extensive network of legal, ethical, regulatory, economic, and social interdependencies. Consequently, advancements in science and technology related to information management, processing, and interlinking for decision support and predictive systems must be supported by concurrent modifications in the global environment, including many stakeholders, including citizens and society. What may first appear as an impediment and a factor hindering the advancement of data science may, however, be seen as a significant advantage. Only worldwide adoption can convert the potential of big data and artificial intelligence into significant advancements in health and medicine management. This necessitates the collaboration of science and society, as well as scientists and citizens, for mutual advancement.

Keywords: Artificial Intelligence; Big Data; Healthcare; Ethics; Economics; Medicine

Introduction

The majority of daily news and recently published scientific articles on research, innovations, and applications in artificial intelligence (AI) pertain to machine learning—algorithms that utilise extensive datasets and diverse methodologies to identify patterns, assist in decision-making, generate predictions, or, in the case of deep learning, autonomously discern significant features within data. Nonetheless, artificial intelligence is a multifaceted notion, and the majority of individuals possess limited comprehension of its true nature. Artificial Intelligence was established as an academic subject in 1956 and, despite its relative novelty, possesses a substantial history. Over the last 60 years, artificial intelligence has evolved into a vast domain of research and development, employing diverse strategies to tackle several difficulties, encompassing theoretical frameworks, methodologies, tools, practical applications, risk assessments, and effect evaluations. The definition of AI is dynamic and evolves with advancements in the area. Since its inception, the domain of AI has facilitated the advancement of several methodologies that assist in decision-making and forecasting, traditionally performed by humans. In 1958, it was anticipated that a perceptron might "walk, talk, see, write, reproduce itself, and possess self-awareness," sparking significant scientific debate between neural network and symbolic reasoning methodologies [3].

The domain of AI research encompasses knowledge representation and engineering, rule-based and symbolic reasoning, temporal reasoning and planning, sensing and perception, learning, evolutionary and emergent social behaviours, as well as the capacity for movement and object manipulation, with particular emphasis on deep machine learning and autonomous feature extraction. The research adopts a perspective that recognises a current tendency towards confining AI to autonomous deep machine learning. The expansive terrain of the topic encompasses philosophy, mathematics, information sciences, computer science, psychology, anthropology, and social sciences, sciences, linguistics, and several other disciplines. Some experts and visionaries, including Ray Kurzweil, assert that deep machine learning will facilitate the creation of an artificial general intelligence capable of autonomous self-improvement and the ability to comprehend or acquire any intellectual task that a human can, potentially surpassing human intelligence. However, the majority of experts concur that significant gaps remain, and achieving this goal is still distant, notwithstanding recent notable advancements in quantum computing. A recent white paper released by the European Commission and composed by the High-Level Expert Group on AI offers a concise introduction of AI, detailing its primary capabilities, anticipated applications, and relevant disciplines [7].

Figure 1: Shows the level of Data and Learning

In the realm of artificial intelligence, it is crucial to highlight that AI is currently widely utilised in the field of medicine. Decision assistance via knowledge engineering and rule-based systems is extensively utilised in computerised provider order entry (CPOE) globally. Advanced signal processing is utilised in pacemakers and defibrillators for decision-making, in cochlear implants with human-machine interfaces, and in electrocardiograms for signal analysis and automated diagnostics.

The topic of artificial intelligence is inherently aspirational and is anticipated to make substantial contributions to medicine, encompassing both research and patient-centered healthcare. Machine learning and deep learning have facilitated the most recent significant advancements in artificial

intelligence, including sound (speech and music) recognition, picture (face, radiology, pathology, dermatology, etc.) recognition, and gaming. Recently, picture recognition has attained a degree of maturity that enables its use and development by non-experts in AI [8,9]. Nonetheless, the excitement around AI in recent years has generated elevated expectations and correspondingly significant apprehensions. Few autonomous deep learning systems have gained widespread adoption in the commercial arena.

The realm of AI may be succinctly encapsulated in three successive and overlapping phases:

Q 1: Humans instruct machines in the management of data and information.

Q 2: Humans impart knowledge to machines.

Q 3: Humans instruct robots to autonomously acquire knowledge.

Challenge was to contracting the influenza virus? What is the likelihood of diabetes in an individual with elevated blood sugar levels? To exemplify the Bayesian pitfall, consider a straightforward case—a pregnancy test. This is a basic examination; it may yield a positive or negative result. Consider a classical test with 99% sensitivity and 95% specificity. If 100 tests are conducted on 100 individuals and 5 provide positive results, the inquiry is to ascertain how many of these individuals are pregnant women. This inquiry pertains to ascertaining the positive predictive value of a positive test, indicating the likelihood that a positive result accurately reflects the presence of the element being assessed by the test. To address this inquiry, it is essential to ascertain the previous chance of pregnancy among the examined population. To comprehend this, envision a scenario where 100% of the subjects tested are men; in this instance, none of the five positive tests would pertain to a pregnant individual. If all individuals tested are pregnant women, then all five positive tests would pertain to pregnant women. If the prior probability is around 1%, then utilising Bayesian principles indicates that the likelihood of becoming pregnant, given a positive test result, is roughly 17%. This indicates that around 80% of tests have false positive results. If the prior probability is around 20% (i.e., a woman exhibiting many indicators of a prospective pregnancy), the likelihood of a positive test being a true positive exceeds 80%. Consequently, less than one in five tests yields a false positive. The example illustrates the significant ramifications of the prior probability in Bayesian contexts.

These are the ramifications of artificial intelligence. The models must consider the prior probability within the population to which they are applied. This should be more comprehensively appreciated, particularly when disseminating results in the literature, which typically focusses solely on specificity and sensitivity. A further consequence emerges only when many concentrated and almost flawless systems are employed concurrently in intricate scenarios. For instance, possessing several systems, each with its distinct false-positive rate, may result in consolidated systems that aggregate the total of all false positives. Decision support systems in Computerised Physician Order Entry (CPOE) have demonstrated a significant incidence of false-positive alerts, particularly among patients undergoing intricate pharmacological treatments [10].

Regulatory Maze

Most diagnostic and therapeutic methods now employed in medicine must navigate intricate regulatory systems to obtain market approval. The regulatory authorities primarily ground their choices in safety, evidence, and additional value. Furthermore, Health authorities frequently utilise medicoeconomic evaluations. Numerous issues must be confronted in the domain of artificial intelligence as it pertains to medicine. Many of them are not exclusive to medicine and health; yet, their inclusion substantially complicates the achievement of the objectives.

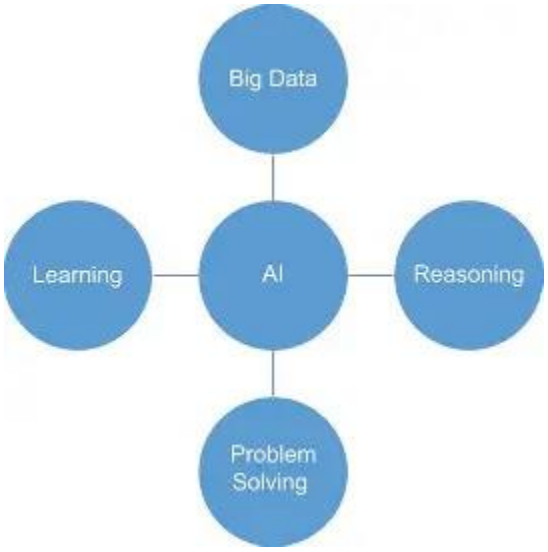

Figure 2: Shows the AI related fields

Bayesian Pitfall

Medicine and health determinants are often defined by their essential Bayesian characteristic. In the Bayesian probability framework, a prior probability is necessary to assess the robustness of the prediction.

A significant portion of medical practices, particularly in diagnosis, adheres to the Bayesian method. For instance, if an individual presents with a fever, what is the likelihood of it occurring? Based on many aspects, including quality-adjusted life years and illness load, utilising metrics such as disability-adjusted life years [4]. Consequently, these actions include economic and legal ramifications, including accountability. The function of regulatory bodies is examined, particularly concerning market entry concerns, such as image recognition [5]. The Federal Drug Administration has initiated a request for contributions about "Artificial Intelligence and Machine Learning in Software as a Medical Device." The assistance of regulatory agencies is crucial, as it fosters trust among healthcare practitioners in using medical equipment and encourages corporations to invest in solid market-ready goods. Nonetheless, this necessitates the establishment of a definitive regulatory framework and suitable evaluation processes and benchmarking tools that do not impede innovation [7].

Pedagogy and Application Discrepancy

Medicine is a discipline including a variety of instruments and apparatus, including stethoscopes, scalpels, microscopes, scanners, protocols, and recommendations. Many of these technologies and

gadgets need education and often particular certification processes for the care workers that utilise them, in addition to requiring substantial expertise. This should also apply to software, algorithms, and other decision support systems. Nonetheless, this is not the situation. Instruction in utilising software and comprehending systems, such as computerised health records, is frequently inadequate. Education about big data and AI is generally inadequate, often nonexistent. Only a limited number of medical colleges instruct future health professionals in the use of AI. Artificial intelligence should be mandated as a core subject in all medical schools globally as a priority. Experts have been questioning this matter for two decades, however it has just garnered significant attention recently [2]. In 5-10 years, as today's young students commence their clinical activities, machine learning, grounded in data science, will be integrated into numerous tasks, devices, and software. Its application, misuse, and overuse, along with the resultant implications for patients and accountability, will hinge on the proficiency of its users.

Data Integrity Chiasm

The quality of data is a persistent subject of discourse about big data and analytics. A defining aspect of the big data age is that data are frequently utilised for purposes distinct from those that prompted their collection. This is a significant divergence from conventional hypothetic-deductive methodologies in medicine, wherein a hypothesis informs the design of a technique, subsequently resulting in the collecting of particular facts. In the era of big data, the principal objective of data-generating operations frequently diverges entirely from the potential applications of the data. It is noteworthy to highlight that long-term clinical cohorts and biobanks have analogous issues. Developing long-term cohorts and establishing a metadata framework together with standard operating procedures for biobanking presents significant hurdles, since they must anticipate applications that may arise years after the initial design.

These enquiries have resulted in subsequent work focussing on data quality and the secondary use of clinical data. Nonetheless, the majority of this research is to delineate aspects that can evaluate the "intrinsic" or "absolute" quality of data [8]. Another strategy may involve implementing a "fit-for-purpose" methodology that focusses solely on the quantitative and descriptive attributes of data, facilitating further processing. The "qualitative" attributes of each dataset may only be evaluated in relation to a particular secondary application. This indicates that the identical dataset will be suitable for addressing certain scientific enquiries while being unsuitable for others. The data are not inherently “good” nor “bad”; their value is determined by the precise context in which they are utilised, particularly in the “fit-for-purpose” evaluation. This is a primary purpose of the FAIR data effort, which seeks to ensure "a posteriori" data usefulness (see below).

An unforeseen result of the "data quality chiasm" is its impact on altering acquisition techniques, particularly in clinical settings. Statements like “the quality of clinical data is insufficient for research” are frequently seen. Consequently, there is an ongoing impetus to transition towards more systematic data gathering methodologies. The RECIST (Response Evaluation Criteria In Solid Tumours) guidelines aim to standardise radiologic evaluation criteria in oncological trials involving solid tumours. This has been successfully developed for testing. The application of RECIST necessitates substantial expertise to mitigate interobserver variability, which may reach 20% [9]. This evaluation has been modified to account for alterations in radiological response, such as in

immunotherapies where tumour size may grow despite a favourable treatment outcome [2]. Regrettably, there is increasing push to broaden the application of RECIST and analogous staging recommendations outside clinical trials for all radiological evaluations to enhance the ability to utilise conventional clinical care for treatment assessment. This results in significant time constraints on the operational operations of radiology departments and a growing number of unskilled individuals utilising these staging methods. As natural interfaces, including voice recognition and natural language processing, advance and become more prevalent across various devices, I advocate for minimising artificial structuring in data acquisition processes. It is preferable to retain data in its most organic form, leveraging more intuitive interactions such as voice and text, while developing robust natural language processing tools that can subsequently generate structured information during a postprocessing phase. This will facilitate the reprocessing of all tales as necessary using newly organised resources.

Search for Veracity

Numerous facets of the artificial intelligence ecosystem need a clear understanding of truth. Knowledge engineering constructs the graph of the "known" or the "relevant," as exemplified by SNOMED CT (Systematised Nomenclature of Medicine - Clinical Terms) or the Open Biological and Biomedical Ontology Foundry. Similarly, rule-based techniques or symbolic reasoning must articulate rules, representing truth in a formalised manner, as do supervised machine learning approaches, which need training sets that convey truth, at least in a probabilistic sense. These methodologies entail numerous expectations, particularly when integrated [6]; however, all, with the exception of unsupervised deep machine learning, necessitate sources of truth, thereby raising the fundamental inquiry regarding the identification of such sources in life sciences and the evidential rigour underpinning that truth. At first view, it appears to be a minor inquiry. Nevertheless, the "truth" is frequently "obscured in text" as the majority of sources depend on intricate narratives that contextualise the ideas they provide. Furthermore, the "truth" is significantly diluted. For instance, with over 2500 articles indexed daily in Medline/PubMed [7], it is virtually hard for an expert to monitor all publications within their study domain. Ultimately, science is inherently dynamic, and hence, scientific "truths" that were formerly accepted may no longer hold validity today. For instance, it was evident until recently that there are two varieties of lymphocytes—B cells and T cells. A recent article by Rizwan et al. [38] delineates a novel lymphocyte type exhibiting traits of both B and T cells, potentially contributing to the pathogenesis of autoimmunity in certain conditions, including diabetes [8]. The characterisation of sources of truth, including their quality of evidence and context of usage, is becoming increasingly significant. This should be accessible to everyone, akin to Cochrane [9], encompassing all domains of life sciences; consistently updated; and in a machine-readable format.

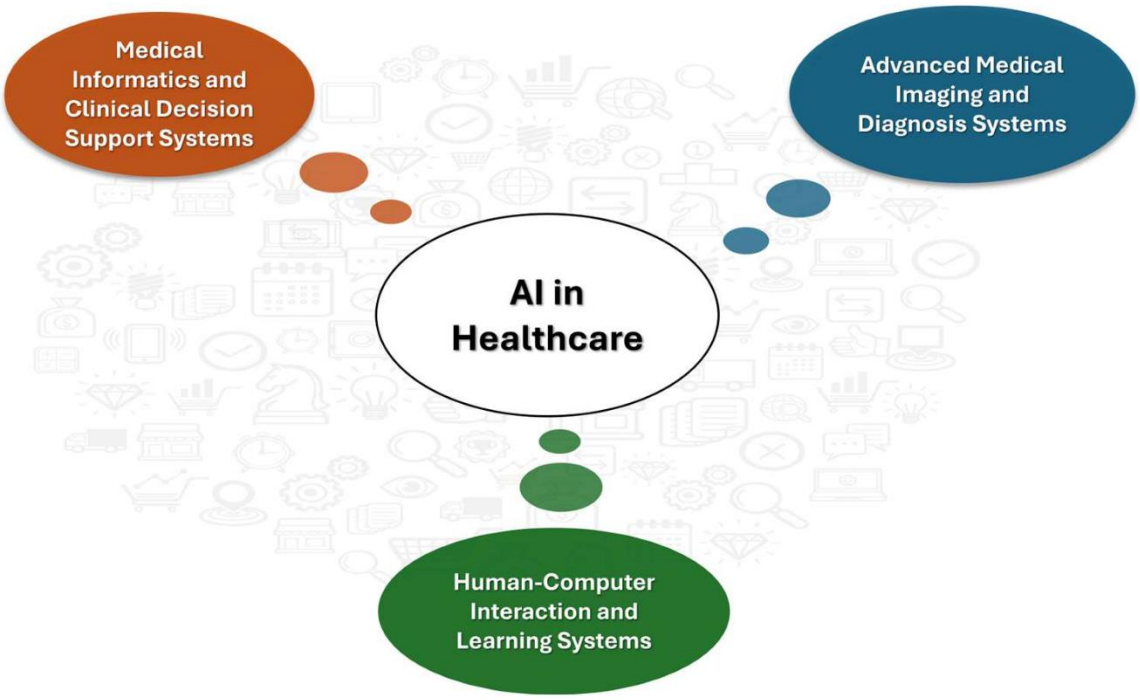

Figure 3: Shows AI in healthcare

Establishing Trust

In science, trust is closely associated with the establishment of proof. Trust is essential not just inside the scientific community but also broadly, to foster adoption, garner political support, and ensure public acceptance. In the summer of 2019, a study released by the Pew Research Centre indicated a favourable public perception of science, however accompanied by apprehensions over integrity, transparency, and prejudice. In all, 86% of Americans express at least "a fair amount" of faith in scientists. A issue is the frequent conflation of scientific dependability with trustworthiness [4]. While scientific evidence may be robust, particularly with immunisation or online health information, the level of trust might be significantly diminished [3]. Numerous elements have been articulated in establishing confidence in science, although they may be encapsulated in three concepts: one pertaining to scientists and organisations, one concerning the subjects of study, and one related to the procedures involved. Integrity is paramount and encompasses scientific integrity, finance, conflicts of interest, among other aspects. Transparency is essential for the motivation, results, and procedures. Ultimately, the approaches employed to manage the processes must be rigorous and resilient. Constructing evidence necessitates consideration of several characteristics, including bias, generalisability, repeatability, and explainability. Certain issues are more formidable with big data and artificial intelligence.

Appropriate Regulation

The collecting and flow of data is typically more challenging than in conventional controlled research. The result is that the data possess distinct characteristics, which are not consistently well-handled, including selection biases. Occasionally, the assumptions limiting the application of

analytical methods, such as homoscedasticity in several statistical tests, are not thoroughly comprehended. Moreover, deep machine learning encounters obstacles related to exact repeatability and explainability. The latter is presently the focus of various studies aimed at comprehending the intermediate representation of data in neural networks that can anticipate and elucidate their behaviour. Explainability and interpretability are frequently utilised as synonyms. Interpretability refers to the degree to which one can anticipate the system's behaviour in response to alterations in input or computational parameters. Conversely, explainability refers to the degree to which the internal mechanisms of the deep learning system can be comprehended and articulated. Molnar [4] provided an excellent explanation of the issue in an open-access book available on GitHub. Nonetheless, explainability may not be the most effective means to enhance worldwide trust in deep machine learning methodologies, particularly when the explanations are inherently complex. Alternative aspects, such openness, repeatability, or qualifications of uncertainty, may prove to be more beneficial [5]. In 2018, Hutson published in Science a review of 400 artificial intelligence papers given at prominent conferences, revealing that just 6% included code for the algorithms and 30% provided test data, thus constraining replication opportunities.

Privacy – New Deal

In the era of big data, privacy requires special attention. Usual paradigms of limiting access to deidentified information are becoming less effective to protect privacy. Increasing heterogeneous data sources and richness of data about each of us, associated with data linkage techniques, strongly increases the possibility of reidentification, including anonymized data [5-7]. The challenge and potential impacts are even bigger for genetic information [5-6]. There is no good technical solution that can harmonize the challenge of preserving privacy and answering the increasing need of data-driven science for accessing large genomic et phenotypic datasets, and there are many ongoing ethical and legal discussions [1-6]. Interestingly, this is not restricted to science, and the same applies to patients' needs for health information [6]. There is a need for better global education about implications and risks of privacy, citizen, policy makers, students, research community, and all stakeholders. A recent scoping review has shown that the understanding of anonymization and de-deidentification is heterogeneous in the scientific community [8]. Discrimination is one of the major risks in privacy breaches, and disclosing privacy information can have many consequences [1], including in reimbursement and insurance coverage [3]. It is important to find the right path between naïve positivism and irrational paranoia. An important step forward is to improve awareness and education of all stakeholders about privacy, technical limitations to protect it, and building regulatory barriers to avoid discrimination.

Conclusions

Respecting human rights to enable responsible genomic data sharing, or the European Union General Data Protection AI and big data in medicine are only in their childhood stages; they grow up fast. Whether they grow up well is still an open question that the future will answer. However, they will not grow up well without actively helping them do so. There are several important initiatives that contribute to this, such as the Global Alliance for Genomics and Health (GA4GH), an organization setting a policy and technical framework for Regulation (GDPR) that sets a completely novel privacy regulation for the European Union. Such initiatives are converging

toward building a landscape that enables science while building trust in improving protection of individual rights. The current study invite the readers to visit the JMIR Open Access collections available on the Web on the following topics: “Big Data,” “Decision Support for Health Professionals,” and “Artificial Intelligence”.

References

[1] Pasham, S.D. (2017) AI-Driven Cloud Cost Optimization for Small and Medium Enterprises (SMEs). The Computertech. 1-24.

[2] Shakibaie-M, B. (2013). Comparison of the effectiveness of two different bone substitute materials for socket preservation after tooth extraction: a controlled clinical study. International Journal of Periodontics & Restorative Dentistry, 33(2).

[3] Gopinath, S., Janga, K. C., Greenberg, S., & Sharma, S. K. (2013). Tolvaptan in the treatment of acute hyponatremia associated with acute kidney injury. Case reports in nephrology, 2013(1), 801575.

[4] Shilpa, Lalitha, Prakash, A., & Rao, S. (2009). BFHI in a tertiary care hospital: Does being Baby friendly affect lactation success?. The Indian Journal of Pediatrics, 76, 655-657.

[5] Singh, V. K., Mishra, A., Gupta, K. K., Misra, R., & Patel, M. L. (2015). Reduction of microalbuminuria in type-2 diabetes mellitus with angiotensin-converting enzyme inhibitor alone and with cilnidipine. Indian Journal of Nephrology, 25(6), 334-339.

[6] Alam, K., Mostakim, M. A., & Khan, M. S. I. (2017). Design and Optimization of MicroSolar Grid for Off-Grid Rural Communities. Distributed Learning and Broad Applications in Scientific Research, 3.

[7] Agarwal, A. V., & Kumar, S. (2017, November). Unsupervised data responsive based monitoring of fields. In 2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 184-188). IEEE.

[8] Mishra, M. (2017). Reliability-based Life Cycle Management of Corroding Pipelines via Optimization under Uncertainty (Doctoral dissertation).

[9] Agarwal, A. V., & Kumar, S. (2017, October). Intelligent multi-level mechanism of secure data handling of vehicular information for post-accident protocols. In 2017 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 902-906). IEEE.

[10] Malhotra, I., Gopinath, S., Janga, K. C., Greenberg, S., Sharma, S. K., & Tarkovsky, R. (2014). Unpredictable nature of tolvaptan in treatment of hypervolemic hyponatremia: case review on role of vaptans. Case reports in endocrinology, 2014(1), 807054.