Clinical Entity Recognition and Longitudinal Patient Summary Generation: AI-Driven Natural Language Processing Systems for Automated Medical Record Analysis

Dr. Stefan Wagner, Associate Professor of Computer Science, Graz University of Technology, Austria

1. Introduction

Artificial intelligence (AI) is increasingly playing a transformative role in the healthcare sector. In addition to many other possibilities, one of the first considerations is its opportunity to boost diagnostic accuracy and quicker, more informed decisions by care providers, allowing for significant advancements in resource use and patient care. In terms of hard data, the efficiency and resources of healthcare forecasting are significantly threatened. However, filtering vast amounts of medical records requires automatic reasoning systems. In order to mimic human learning and problem-solving with complex or rapid data transfer, AI grasps discrete information from underdeveloped and disorganized statistics. Extracting knowledge from minable information hidden in large databases has long been a major contributor to machine learning.

Machine learning is a subset of artificial intelligence that focuses on the development of algorithms that predict outputs from valuable knowledge. This paper explores the comparable work of pioneers in a computing-based category. A few other artificial intelligence categorization systems are briefly discussed and may even be used in the future to draw comparisons. Healthcare is also used to illustrate the complexities that data-mining applications need to address due to the range of opportunities that these developments foster. AI-driven systems face widespread confusion when introduced to healthcare. Although medical science has undergone paradigm changes over time, its expansion has significantly improved in the last few decades. The desire to build machine learning applications in medicine is growing day by day in the face of the rising

complexity of healthcare. Conducting research and planning solutions in medical science remains a common desire for most researchers.

1.1. Overview of AI in Healthcare

Artificial intelligence in healthcare is increasingly becoming popular due to the rapid advancements in deep learning algorithms and various high-performance computing resources being available to train such algorithms. Unlike traditional rule-based programming, artificial intelligence (AI) modeling can be trained on a large dataset to learn complex patterns and relationships and provide more robust outputs. AI and related technologies have been successfully applied for diverse clinical applications, such as predictive analytics, diagnosis, prognosis, and personalized medicine. Moreover, healthcare data such as medical images, electronic health records, omics data, and text records have vastly increased over the years, therefore demanding more sophisticated techniques to efficiently analyze and utilize the data. Relevant studies have demonstrated that the potential use of AI modeling as an aiding tool for clinicians can significantly enhance diagnostic accuracy and provide valuable insights for clinical decisions.

However, the great utilization of AI still faces significant challenges, especially regarding the integration of AI with current healthcare information technology systems. In addition, regulatory agencies are currently developing regulations to govern the development and use of artificial intelligence for clinical decision-making. The future of AI in healthcare is still unclear and, hereafter, it is critical to determine whether AI can revolutionize the healthcare industry as a standalone technology or needs an active interdisciplinary collaboration to benefit from the vast information that this sector stores. Despite the challenges, there are many successful examples where AI technology has been beneficial for improving the quality and efficiency of patient care: an analysis of chest X-ray images by such systems was able to predict pneumonia along with two other radiologists, virtual health assistants help individuals in monitoring their blood sugar levels, and family medicine physicians, among others. These developments continue to lure many healthcare organizations to seriously contemplate AI-driven operations over the next five years.

1.2. Importance of Automated Medical Record Analysis

In contemporary healthcare systems, automated medical record analysis can be an extremely important component utilized by hospitals and other healthcare organizations. Clinicians are faced with a number of time and effort issues. Anything that can minimize the load would be appreciated. Moreover, clerical and information management activities represent the majority of time-consuming undertakings. Clinical evaluation is necessary; without a doubt, simply managing and processing a large amount of patient details will take a fair amount of time. In large chunks, seeing records is a new method of shortening total treatment time in itself. High-tech, ubiquitous decision support systems will not be successful if they operate on old data and stale material. Accurate and immediate decision-making rests on fast and efficient data access. Moreover, one in two hospitals contains substantial data input errors; for example, in privacy and contact details. These errors can cost up to 10 percent per record.

On the other hand, the extraction of detailed patient descriptions from outlets is done using custom-designed algorithms. The emphasis is on obtaining information on comorbidities, background, weaknesses, diagnostics, therapies, and outcomes. The latest information on emerging discoveries, long-term studies, deviations, adverse effects, risk evaluations, logistical data, effectiveness, and 'real-world' information could all be made available on a continuous basis to improve risk, expense reduction, policy evaluation, resource allocation, and value-based payment. Both of these issues pertain to the trend towards mechanization and scrutiny of healthcare delivery. Delivering beneficial care in the twenty-first century is complex. Much more data and facts are available, and we have a greater understanding of health and illness.

2. Machine Learning Fundamentals

The field of machine learning is built on a core idea: providing systems with the skills needed to learn and make predictions based on data. Essentially, these systems can be trained to find patterns and signals in an input data set, then generate outputs when presented with new data. Significantly, the process of teaching a model to accomplish this can be done in two main ways: through supervised and unsupervised learning. Supervised learning is a method of training where the objective or goal is known and the algorithm learns by finding the patterns between the input data and the output data.

This is generally done through labeled data, which identifies what specific patterns correspond to which outputs. Inputs without associated outputs comprise a testing set, on which the model can make predictions. As opposed to supervised learning, unsupervised learning is driven by data where the outcomes are not known in advance; its objective is to mine the data for patterns.

In practice, algorithms are used to expedite these learning methods. In the context of machine learning, algorithms refer to processes built into the system dynamics that allow learning to occur and transform data into predictions. These algorithms give machines the power to learn on their own, which is central to using machine learning in applications such as medical record analysis. Extracting insights from records is a complex task that involves understanding a vast amount of data, test results, comorbidities, and perhaps even narrative descriptions of patient symptoms, among many other possible variables. Not only can this data shape patient care decisions, but it can also be useful for research and outcome analysis. Machine learning has the potential to reveal not just descriptive insights from medical and health data, but can also predict the likelihood of outcomes and make recommendations based on available evidence. In addition, as we have a huge number of possible inputs to consider, machine learning changes focus from a rule-based approach, where decision-support systems guide by instructing which data points to consider, to a data-driven approach, where the algorithms consider hundreds of thousands of data points and grade their relevance to make a prediction. This can be particularly useful in cases where diseases have multiple manifestations or other clinical characteristics, and there are no firm guidelines to help clinicians make decisions.

2.1. Supervised Learning

In supervised learning, the algorithm is provided with a dataset that includes pairs of inputs and corresponding outputs. The algorithm learns to map these inputs to the specific outputs based on the supplied examples. It is essential to split the available datasets into two parts: a training dataset is used to teach the model and a testing dataset serves as an unseen set that may be used to confirm how good the model is for real-world usage. Special attention is needed to develop a good model that can generalize well to new, unseen inputs, without solely memorizing examples seen during training, also known as overfitting. This subsection presents an overview of common

algorithms used widely in clinical decision-making systems and utilized heavily in the healthcare sector.

In a clinical setting, the trained models can be used to predict diseases and outcomes that could be useful and informative for helping doctors design patient-tailored care plans. Generally, these trained models provide probabilistic values for determining whether a person has certain diseases or comorbid factors based on inputs and decisions made by trained physicians. In medical research, supervised learning has the benefits of having evaluation metrics that can be used to assess an algorithm's performance. However, acquiring labeled patient data for training an algorithm has several challenges.

2.2. Unsupervised Learning

In unsupervised learning, unlike supervised learning which requires labeled data, the algorithms are presented with data patterns in which the algorithms themselves are required to identify and extract the hidden patterns and correlations within data for human clarification. Common applications of unsupervised learning include the following: (1) constructing a feature representation of the input data, (2) segmenting a dataset into subgroups, (3) identifying abnormal patterns that do not conform to the dataset distribution, and (4) reducing the dimensions of the data, which is very important for data visualization. Many times, we need to perform exploratory data analysis in the healthcare domain in which dataset labels are not known. Using unsupervised learning methods, we can formulate the hidden patterns and correlations within the dataset to generate new hypotheses from data for further analysis. Common applications of unsupervised methods for healthcare datasets include the following: (1) clustering of patients based on clinical and phenotypic features, (2) dimensionality reduction of high-dimensional data, and (3) anomaly detection of clinical notes and sequences. Unsupervised learning methods uncover overall structure in the data, particularly in clinical notes, and they excel in identifying subpopulations in the datasets. Consequently, these methods are employed in the development of personalized and precise treatment options for patients. However, these methods inherently lack interpretability and require noisy labels and a series of parameters for their applications. Consequently, domain knowledge should always validate any

unsupervised outputs. Overall, unsupervised methods complement supervised learning methods, enriching the analytic toolbox of healthcare data.

3. Medical Record Data Preprocessing

After the general outlook and state-of-the-art overview, a systematization of the major methods and techniques proposed in the literature is provided. The systematic analysis starts with a discussion about medical record data preprocessing. Before getting to the core of data analysis, a preliminary and crucial step is the data preprocessing. The main techniques for cleaning raw data and feature engineering are summarized. Medical records, like any human-generated data, may be subject to different levels of data entry errors; indeed, they sometimes incur inconsistencies, missing data, or misalignments. Hence, it is crucial to address some feature selection and feature construction techniques, such as dimensionality reduction or other advanced methods, in order to optimize performance. Information overload may force some drastic reductions, for instance, considering only summary statistics; however, this directly reduces the granularity of available information, hence affecting the potential outcome of knowledge extraction.

The primary function of data preprocessing is to complete the initial phase in the process of extracting insights from the inputs. Among the wide range of methodologies available, there is no "one-size-fits-all" solution. Rather, the keyword is "trade-off". Clearly, a record consisting of hundreds of poorly designed, uncorrelated parts is neither efficient nor effective, from a computational point of view, necessitating more complex methods that require substantially longer computational time. Of course, increasing the data granularity can lead to more accurate forecasting. The proposed trade-offs are essential for improving the performance of the domain experts' decision-making support, since the collected data satisfies the requirements of AI-driven predictive models.

3.1. Data Cleaning

Just as in information retrieval, data cleaning is a critical step when dealing with medical records. Medical data may suffer from inaccuracies, posing difficulties during the data analysis phase. Common sources of inaccuracies include data from wearables suffering from transmission issues, or professionals inputting heart rates or time of arrival quickly and not always accurately. To create useful and reliable insights from this noisy data,

dedicated processing is essential. Data cleaning consists of low-level techniques such as deduplication, which removes two near-identical records with small errors, or filling in missing values. It is also crucial to assess data consistency, completing missing parts as well as ensuring data completeness, or getting rid of an incomplete patient's record. Another often-forgotten step is validation from a medical standpoint, ensuring that blood pressure and heart rate are in the correct range for each other. Of crucial importance is the cleaning phase in clinical application, as it can impact the patient's treatment. Software bugs, server issues, and rare highly pathological events can corrupt the data and lead, for instance, to biased classifiers if not taken care of. Severe issues can be encountered in some patient databases, such as missing sex for the patient, or no date for hospital admission and release, rendering the record unusable without manual input. The importance of cleaning medical data in data mining is also addressed. Missing values can affect large fractions of records in early disease stages where tests might not be requested. When missing data limit analysis to either a low amount of records or, most importantly, records with no missing data, the sample becomes biased due to the exclusion of patients that were either most noticed or requested fewer tests. Moreover, medical advances such as liquid biopsies possess limited validation data in databases, constituting a small fraction of the information to manage and leading to incomplete files as researchers wait for the clinical outcome. Around 25-44% of data in electronic hospital databases is cosmological, generally from taking the mean for the missing day. Finally, human error in manual clinical record input still makes human control of tables or files necessary. In this respect, AI tools can also remedy both ongoing data management in hospital records and data cleaning prior to data preparation for researchers with less expertise. In accordance with real-time data production and medical events, data should be continually cleaned.

3.2. Feature Engineering

Feature engineering is a critical part of the medical record analysis process. It refers to the process of converting raw data into insightful features, also known as predictors or independent variables. This process is particularly crucial for the medical domain as the data is typically large, multidimensional, and high precision. The key to effective feature engineering in healthcare lies in molding the data into features that are clinically meaningful. Correctly identifying and including the most important or relevant features

will result in successful machine learning models that can make meaningful predictions based on the medical record data.

Domain expertise is essential for understanding the data and identifying the most relevant and important clinical features. In healthcare, the high-dimensional nature of omics data poses specific challenges including the increased likelihood of overfitting if a model is trained on the original data. Consequently, most models built on this data will incorporate dimensionality reduction techniques, such as selecting the most relevant features, before the modeling process even begins.

Effective feature engineering has been shown to greatly enhance the performance and predictive accuracy of machine learning models across a broad range of healthcare applications. In many cases, the development of new, successful, or novel methods can have a major caring impact. For example, using natural language processing, a model that was developed to predict which patients would become no-shows at outpatient clinics in the obstetrics and gynecology department achieved an AUC-ROC of 0.95 when trained using a newly developed word embedding data field. This method was created through feature creation and engineering, including word-level features, using patient notes. Key to the success of the model was iteration, testing, and validation of features. The highest performing features were the final inputs for the model.

4. AI Applications in Medical Record Analysis

AI can be used to support the accurate analysis of medical records. In terms of disease diagnosis, machine learning and deep learning technologies can leave intelligence within medical record analysis to speed up the process and find new insights that were previously difficult and time-consuming for manual clinicians. Predictive models can be created to analyze patient histories from related databases, which can help identify populations or people at risk of a certain disease or those who already have it. AI can suggest a suspected diagnosis that leads healthcare professionals to interpret the diagnosis with a broader perspective and, in turn, suggest more related ancillary tests to ensure the diagnosis. Treatment recommendation systems can propose possible generic therapies to use for the type of disease a patient has based on the medical record of prior clinical trials. Treatments can be more personalized instead of generic for any patient as decisions are based on a patient's specific type of presentation with related symptoms. Evidence-based decision-making is far more informed and intelligent as an exception

handling process for AI than current clinical practices and largely intangible medicine. AI could potentially support a volume of perhaps more consistent and equitable care plans that would help balance the ever-increasing demands of health services under cost constraints. Unfortunately, many of the offers do require the breaking of ethical principles and confidentiality regarding a patient's private data. A key to AI-based healthcare technology is to build the AI model around the known way physicians want to work rather than implement new stand-alone technology that demands a new model of prior physicians' work. Successful adoption of AI in healthcare may inspire the transformation more rapidly. Once proven, clinicians are far more likely to adopt an innovation that also improves their effectiveness beyond other conflicting criteria.

4.1. Disease Diagnosis

The automatic analysis of medical records may assist in various applications, the first of which we explore in this subsection being disease diagnosis. There is potential for AI models to not only analyze current records but also to identify patterns present in past records of given symptoms and responses to treatment. This may improve accuracy, diagnosis speed, and ultimately have a positive impact on patient outcomes. Indeed, research has demonstrated that human clinicians are very limited in their ability to effectively analyze large volumes of ad-hoc patient data in a reasonable timeframe. In contrast, machine learning approaches are capable of analyzing vast amounts of current patient data in a relatively short time. These systems may be able to find existing co-occurrences or dependencies between certain diseases and their symptoms that are unclear to human experts.

The models generated by these systems can be used to automatically pre-screen patients, making it faster for clinics to operate within their demand and ensuring that there are enough spaces for all patients. In this capacity, though, it is important that the system's decisions are always double-checked by a human. It would also be useful to involve a human being in the decision-making process that will activate, deactivate, or modify treatment options. Machine learning algorithms have the potential to deal with ambiguity in a dataset in order to achieve this. Hence, as it stands, the deployment of AI-driven systems in clinical practice arenas is not straightforward. Some of the complex conditions clinicians face in the clinic are difficult to learn from, their pattern not reflected in the available data. Hence, diagnosis of disease in a clinic is highly complex

to automate, although progress is being made to this end. It is also difficult to formulate AI-driven systems that do not overlook features in data diagnosis. A robust system should be developed that is capable of generalizing to patient data from various clinics. The system should display high reliability across different clinical sites. The system should also extensively collaborate with clinicians in the various stages of development. Appropriate validation inputs may be required, such as input from real patients, in addition to the anonymized datasets used during AI development.

A key requirement for any system developed for automated disease diagnosis is therefore detecting features that a clinician alone may fail to detect. Indeed, research has demonstrated that machine learning-based solutions for image diagnosis have matched human expert capability to varying extents and almost outperformed them in the case scenario of diabetic retinopathy. This approach was essential, as the rate at which ground truth labels, and thus required image numbers for training, were available is low. AI-driven systems have been able to overcome this shortcoming through proactive self-training. AI-driven systems provide accelerated diagnosis in many cases. In the field of sepsis, for example, the introduction of a support vector machine with a complex hyperplane changed the secure diagnosis by 12 hours and lowered hospital mortality. Exploiting a data-driven approach, the system accurately analyzed up to 160 data fields to provide diagnosis. Here, the previous column documents that adept clinicians were not able to quickly diagnose.

4.2. Treatment Recommendation

AI supports automatic treatment recommendation systems. In clinical practice, having support for treatment recommendations is a key factor for decision-making. AI models for treatment recommendations obtain the patient's clinical data from electronic medical records to suggest individual treatments. There is a current demand for AI-driven systems that focus on providing evidence-based treatments. The evidence for recommendations can be the outcome of hypotheses or trial results that identify the treatment subpopulation for which the most benefit is obtained. An additional line of exploration is to analyze extensive demonstration data to identify treatment pathways that, on average, have the same level of efficiency and consider patient preferences or clinical guidelines. Patients stratified by different phenotypes can share the same disease and treatment recommendations for care personalization.

Our research explores the design of an automatic treatment grooming system to integrate evidence-based algorithms, patient preference integrators, and domain experts, such as physicians. The use of AI techniques to recommend treatments has some considerations. An increasingly effective solution requires knowledge of the clinical history of a patient. Treatment decision support systems could also be enhanced by a system that, from a multi-professional team perspective, not only suggests treatments based on evidence and algorithmically obtained data but also recommends diagnostic exams to perform, timing, and monitoring. It is also important to address data privacy and governance to develop machine learning and deep learning systems. A treatment recommendation system can also have biases in suggesting a particular treatment, which should be managed, for example, to minimize biases and to promote health equity. The system must be designed to recommend treatments automatically, allowing clinicians to evaluate the specific case and have the final say based on their domain expertise.

5. Challenges and Ethical Considerations

AI-driven systems for automating medical record analysis bring a series of challenges. First, the analyses are performed on large-scale electronic medical records stored in databases, and legal and ethical implications must be considered when sharing and using such information. It is fundamental to protect patients' privacy; hence, several algorithms have been developed to guarantee that no sensitive patient-related information is disclosed. Moreover, the outputs from such analyses must comply with the requirements of well-designed legal and ethical frameworks.

Another relevant aspect to consider in building and training models over health-related data is bias management, or fairness in the developed algorithms. If not carefully designed, the outcomes of the implemented models could be discriminatory across different communities or social groups. In addition to general-purpose algorithms aimed at promoting fairness, recent advances in the AI field have allowed the design of models that directly mitigate healthcare disparities due to bias. Model validation is an additional issue to consider: the lack of calibration and overfitting of a predictive model might lead to incorrect results and jeopardize patient safety. In medicine, accountability for decision-making is a fundamental requirement, and clinicians using AI predictions need to understand how the algorithm works and how the results were obtained. Lastly, implementing AI and ML-based systems in clinical practice raises issues of transparency

and interpretability of their decision-making processes. AI must therefore be able to justify its interpretation or course of action to engender trust and acceptance.

5.1. Data Privacy

The field of AI in healthcare is partially bound by data privacy. While hospital and medical records are required for the formation of evidence-based models and disease detection, data privacy is a concern. Since patient data is typically sensitive, the collection, storage, and processing of it could potentially be done without real consent. Furthermore, the legally accepted ethical use of data and informed consent are intertwined and must be carefully managed. One set of legal guidelines for data privacy in the US was established by the Health Insurance Portability and Accountability Act. The primary function of this law is to define the acceptable use and disclosure of identifiable patient-related data and the necessary safeguards.

Good ethical, legal, and secure practices for patient privacy are essential because nonadherence to them could potentially lead to a data breach. A data breach is the intentional or accidental release of secure personal, confidential, or protected data to an untrusted environment. This may be a release to a third party, as in the case of a leak or theft. Data breaches are becoming more common around the world and are becoming increasingly sophisticated. Not only do they inform the victim that their information was or could have been breached, but the victim is also provided with a host of free data protection services, from basic credit monitoring to identity theft insurance and identity recovery-related services. This breach response and managed protection approach are necessary to safeguard both victim data and the general population by extension. The nature of leaks and breaches can have a profound impact on a patient's level of trust in and safety when seeking medical help. Technological methods such as encryption and the establishment of highly secure access controls or automatic regular data redaction can be used to anonymize and increase data privacy from non-validated backend and front-end users. However, these technological methods are not standalone cures, and the organization implementing them would have to put in considerable time, resources, and effort to ensure they are continuously up to date and secure. Machine learning and AI have been extensively used in techniques developed to protect data privacy. Healthcare organizations can meet their legal obligations to patients using AI, provided that the difficult hurdle of data privacy regulations is met and compliant data governance

structures are created. Since patient privacy and HIPAA privacy protection regulations are at odds with AI data requirements, the balance in technique design is ultimately achieved by those who work in the industry. To our knowledge, the field is continuously evolving as techniques are developed to protect patient data privacy as far as possible while still learning and specializing in individual patient medicine.

5.2. Bias and Fairness

Unsurprisingly, AI in healthcare is not immune to the problem of bias either. In fact, using biased data for training machine learning models creates the risk of producing biased outcomes. For instance, if a machine learning model trained on biased training data is used in a clinical setting, it may make decisions that unfavorably affect a certain group of people or result in stalled improvements in the health outcomes of marginalized populations. In the context of automated MRA tool development, fairness needs to be achieved by ensuring that the output of the system is unbiased and does not lead clinicians in healthcare to make decisions that discriminate between different groups. This should be met using training data that is as diverse as possible and ideally include an expert representation of all groups in the labeled training data. These datasets should be publicly available and error-free to avoid the negative impact of bias and ultimately contribute to improving safety issues in related medical software.

There are several strategies and techniques to detect and mitigate bias and unfairness in an AI system. Among these are post-processing techniques utilizing conventional statistical methods, retraining machine learning models with adjusted training data, generating training or testing data synthetically, ensuring that a fair representation of the data from all groups of interest is accounted for at the time of constructing the data, and fair representations of data features reflecting key population characteristics. All these approaches aim to ultimately prevent any unanticipated adverse events resulting from the deployment of machine learning systems when they go into routine use as an active part of the healthcare system. It is also important to involve patients and non-AI experts in ethical discussions on the topic at an early stage of development and establish interdisciplinary and multi-expertise working groups in the whole research process.

Currently, the issue of fairness in AI systems in healthcare has not yet been widely researched, and there are high hopes that by addressing the fairness of this new method of clinical care, it can ultimately help to reduce the social and racial healthcare

disparities so prominent in the world today. While there are relatively few clinical reports where biased AI has actually been found to directly impact patient outcomes, the fact that these defects in AI arise relies on the current development practices of health AI models. Applying AI-driven technologies to health records for automating medical diagnosis, prognosis, and treatment plans has been found to classify patients based on financial status, race, and social group rather than the clinical needs of the patients. This is avoidable by ensuring that hypothesis testing of algorithms is done after removing model-dependent variables strongly collinear with social group. Aside from these serious known clinical cases, these biased and unfair AI algorithms may actively but unwittingly continue to maintain racial and social inequality. Ethical and fairness issues have gained significant attention in the presence of controversies and numerous reports.

6. Future Challenges

Emerging technologies and rapid advances in the AI/ML field ensure that many currently facing challenges and limitations will be solved to some extent. The needed infrastructure to ensure secure, scalable, and robust AI services does not exist today, and real-time deployment of AI services with inference and sometimes control needs close attention. A decline in quality, for example, of an educational AI service may cause large societal turbulence. Consequently, AI/ML will be integrated into healthcare ICT over time. AI/ML is only a small part of a whole when it comes to the assessment of, for example, the impact of an AI-driven assessment system. However, because AI is so dynamic, the AI/ML part of the system is more complex to assess than currently deployed ICT systems. Thus, future research must address the technical transferability of AI/ML algorithms to other institutions, consumers, or patients, together with an economic and patient safety impact.

Finally, the area of AI/ML for healthcare has been scrutinized for bias towards, for example, gender, ethnicity, and socioeconomic class. Remedies exist that can decrease these ethical problems, depending on the ethical challenge, by a combination of using less clinical information about possibly biased groups, reduced decision support capabilities for these individuals, pre- and post-deployment reporting or assessment, or regular equity measuring processes. Note, however, that there must be continuous progress in these subjects because the requirements from stakeholders will change over time, including a shift in the technology area, as exemplified by a decreasing focus on

uncontrolled black-box AI technology without decision transparency to emphasizing the explanatory power of the algorithms and AI technology robustness.

7. Conclusion

Medical record data is expected to be increased by 48% in the next two years, and by 2020 the digitalization of healthcare was considered completed in the more developed countries. AI-driven systems are expected to already be actively deployed in automating the tasks related to analysis and reasoning over data that appear in the patients' records. The reported results suggest that machine learning can be effectively utilized to extract meaningful insights from the available data for the expansion of ICUs in a robust and accurate manner. Even though the need for trained medical professionals remains valid, the application of machine learning is deemed valuable for clinical medicine and healthcare management, as it may reenable its potential on a much larger scale. The filtering capabilities of machine learning can significantly boost clinical decision-making processes. In as much as these systems have gained attention over the last decade, the potential to take advantage of the growing quantities of health-related big data is limited in part due to ethical constraints, leaving substantial scope for ongoing cultivation of research and innovation. Several ethical considerations are identified as pivotal for initiating the deployment and acceptance of AI systems in medical data analysis, the most important of which are data privacy and algorithmic bias. To address these concerns, practitioners propose actions such as examining minimally sufficient data requirements, using only de-identified medical data, and demonstrating the ethical implications of the AI-driven algorithm after examining and perhaps auditing its data factor. Finally, the potential of using the European health research resources in multi-lateral collaboration and, concisely, the potential of integrating initiatives in the AI-on-medical-data landscape are pointed out. The capturing of decentralized healthcare databases to develop globally applicable solutions for improved patient management and clinical research and the promotion of responsible AI and data privacy accountability, security and fairness coincide as logical directions.