WHITE PAPER

Use of AI/ML for Improving the Quality of RWD

TABLE OF CONTENTS

INTRODUCTION

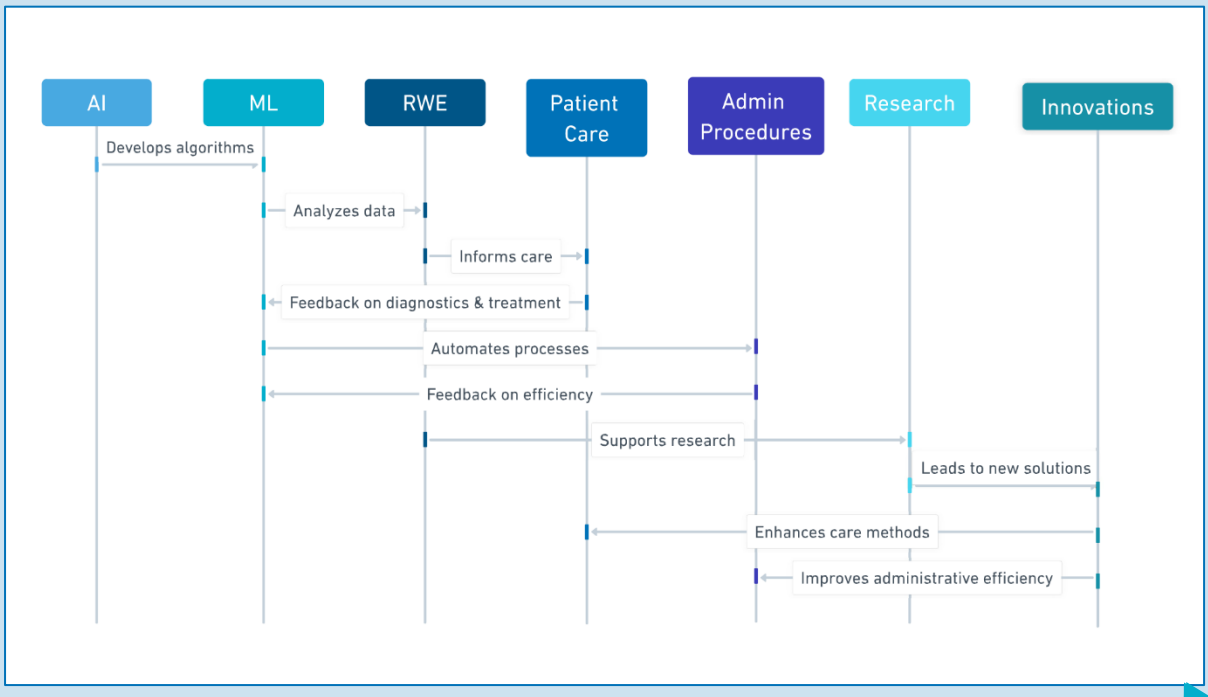
Within the complex and pressing landscape of global health issues, emerging technologies such as artificial intelligence (AI) and machine learning (ML) are positioned to act as transformative drivers that will enhance the precision, accessibility, and economic sustainability of healthcare.

AI and its associated technologies are becoming increasingly prevalent in various business sectors and in society. Real World Evidence Data (RWD), derived from the vast and varied patient data generated in clinical practice, holds the promise of transforming healthcare decision-making, treatment strategies, and policy formulation. Yet, harnessing the full potential of RWE has long been hampered by challenges related to data quality, heterogeneity, and accessibility [1]. Moreover, pandemics like the coronavirus (COVID-19) put a strain on healthcare systems, resulting in a lack of protective equipment, insufficient or erroneous diagnostic tests, overworked physicians, and a lack of information exchange.

As these tools continue to evolve, new opportunities are expected to emerge, the global market for AI in healthcare valued at $15.4 billion in 2022, estimated to grow with a compound annual growth rate (CAGR) of 37.5% from 2023 to 2030 [2]. One of the primary drivers is the increasing availability of extensive datasets containing digital health-related information of patients. Additionally, there is a rising demand for personalized medicine, driven by the need to tailor medical treatments to individual patient characteristics. Another influential factor is the growing emphasis on cost reduction in healthcare delivery.

In this paper, we aim to elucidate the profound impact of AI in elevating the quality of real-world evidence data. We will delve into the fundamental role of AI impact, the challenges faced, and the best practices to improve quality of RWD data.

FIGURE 1 | Index Prescription Medications for Asthma Symptom Control

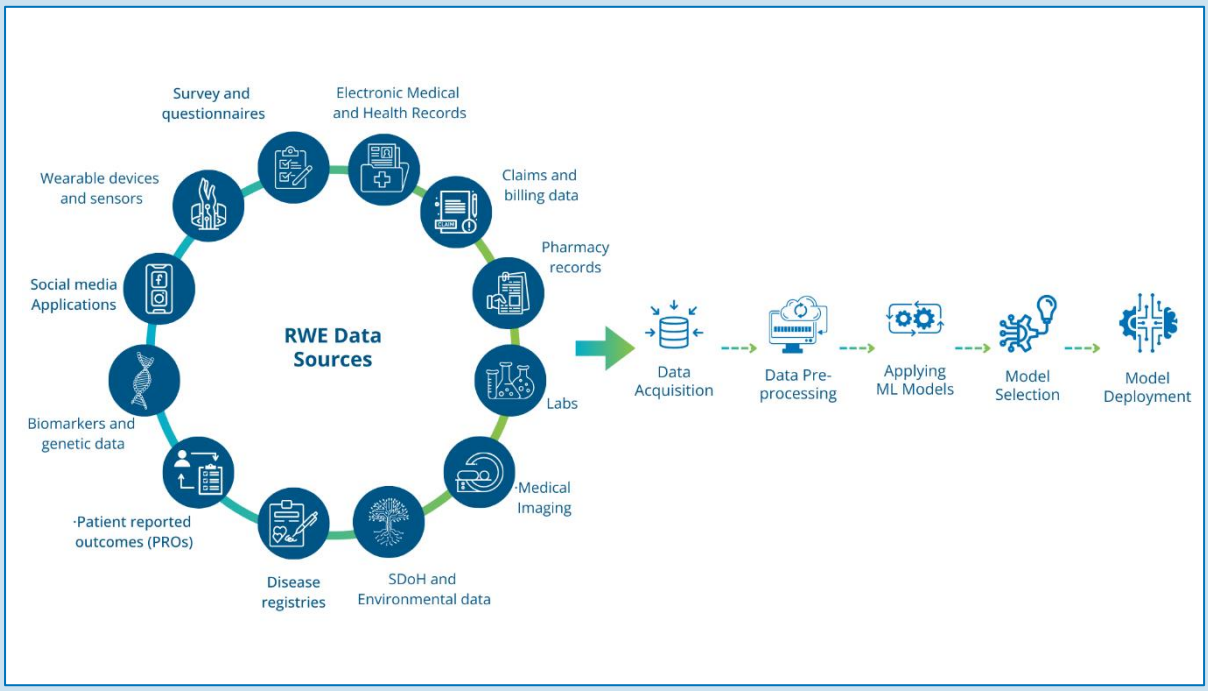


UNDERSTANDING AL/ML AND POTENTIAL AI OPPORTUNITIES IN HEALTHCARE

What is AI and ML?

Artificial Intelligence (AI) and Machine Learning (ML) are two closely related fields in computer science that involve the development of algorithms and systems capable of performing tasks that typically require human intelligence, such as understanding natural language, recognizing patterns, learning from vast amounts of data, making predictions, and assist in clinical decision-making.

FIGURE 2 | RWE Data Sources



How AI is used today in healthcare?

In this era of technological advancement, artificial intelligence (AI) stands as a transformative force, poised to revolutionize how we collect, manage, and utilize RWE. These technologies hold the promise of revolutionizing multiple facets of patient care and streamlining administrative procedures within healthcare providers, insurance companies, and pharmaceutical organisations [3][4][5][13].

Clinical Trials

Traditional Challenges

- Recruiting patients for clinical trials is a significant problem, with about 80% of trials worldwide failing to meet enrolment goals on time [6]. This issue leads to delays and substantial financial strains on trials
- Traditional clinical trials face a 30% dropout rate, influenced by factors like

inconvenience, visit frequency, complex protocols, inadequate support, and lack of expense reimbursement
- Unstructured and lack of quality data from clinical settings

AI Solutions

- Patient recruitment and identification: AI can help identify and recruit potential target cohort by extensively analysing and linking patient records from various sources including EHR, claims, imaging, labs, etc. This further promote streamlining data using NLP and improving quality
- Decentralised Trial solutions: AI offers a solution by enabling remote health assessments and decentralised clinical trials which can notify medical professionals and patients about available trial opportunities, minimising unnecessary patient screenings. Furthermore, AI can simplify entry criteria, making trials more accessible to potential participants
- Digital Twin Technology: Integration of digital twin in clinical trial enable real-time observation of patients while simulating and predicting various clinical outcomes

Benefits

- Reduce delays and financial strain on healthcare sites
- Empowering researchers to design trials that are more patient-centric and efficient by identifying suitable candidates and improved participation
- Data accuracy, faster access and enhance patient safety

Diagnosis and Early Detection

Traditional Challenges

- Ensuring accurate and timely diagnosis can be challenging due to the complexity of many diseases and misdiagnosis
- Delayed or incorrect diagnoses can lead to ineffective treatments, longer hospital stays, and increased healthcare costs
- Limited access to healthcare inefficiencies and healthcare disparities can lead to ineffective treatments, longer hospital stays, and increased healthcare costs

AI Solutions

- AI can aid in diagnosing diseases by analysing patient data including imaging, genomic, symptoms, and medical histories. Machine learning models can identify patterns and assist in early diagnosis and disease detection
- AI can help in predicting and risk stratification of disease outbreaks, disease progression and patient readmissions, allowing healthcare providers to allocate resources more efficiently

Benefits

- Timely and accurate disease identification
- Improved prognosis and treatment planning
- Potential for early intervention and better patient outcomes

Drug Discovery

Traditional Challenges

- A significant portion of global GDP, around 6% to 7%, is allocated to healthcare

expenses annually, and the cost of bringing new medicine to the market is an expensive and time-consuming process, with costs exceeding $1 billion and taking up to 14 years [7]
- Low success rates in drug development make investments in the pharmaceutical industry highly risky, leading to inflated drug prices to compensate for the failures

AI Solutions

- AI accelerates drug discovery by analysing vast datasets to identify potential drug candidates, predict their effectiveness, and optimise their chemical properties.
- Drug repurposing is facilitated by AI, which can identify existing drugs that may be effective against new diseases

Benefits

- Proactive treatment, reduced adverse events and better patient outcomes
- Reduce the time and cost involved in bringing new medications to market
- Reduce resource utilisation and financial burden
- Accuracy and save time in risk stratification

Precision Medicine

Traditional Challenges

- The global burden of chronic diseases is escalating rapidly, affecting a significant portion of the population
- Chronic diseases are intricate, involving various underlying mechanisms and manifesting with diverse symptoms and comorbidities
- Existing diagnostic methods rely primarily on clinical symptoms and conventional biomarkers, often lacking precision and detecting diseases at advanced stages

AI Solutions

- AI analyses genetic, clinical, and lifestyle data to tailor treatments and medications to individual patients, improving treatment efficacy and minimising side effects.
- AI-driven decision support systems help physicians make treatment recommendations based on a patient's unique profile

Benefits

- Effective treatments and patient adherence with higher success rates
- Reduces the risk of adverse effects and minimises harm to patients
- Improve outcomes by addressing diseases in their earliest stages
- Lead to cost savings by avoiding ineffective treatments and hospitalisations, as well as by optimising drug use

Remote Monitoring and Wearables

Traditional Challenges

- The emergence of COVID-19 prompted a global response, emphasising social isolation as the primary strategy to curb its spread
- This crisis, coupled with shortages of personal protective equipment, has compelled healthcare stakeholders to reconsider the necessity of face-to-face medical encounters

•

AI Solutions

- AI-enabled wearables and remote monitoring devices collect real-time health data, allowing healthcare providers to monitor patients outside of clinical settings and intervene when necessary
- Telemedicine platforms use AI for remote consultations, making healthcare more accessible, especially in underserved areas

Benefits

- Increased engagement and motivation to make healthier lifestyle choices.
- Early intervention, potentially avoiding costly medical procedures and hospital stays.
- Enhances the quality of remote care and enables timely decision-making.
- Lead to improved patient outcomes, better quality of life, and increased patient satisfaction

TABLE 1 | AI Challenges, Solutions and Benefits in Healthcare

Potential Opportunity	Traditional Challenges	AI Solutions	Benefits
Clinical Trials	• Difficulty in patient recruitment, leading to trial delays. • High dropout rates due to various factors. • Unstructured and poor-quality data.	• AI for patient recruitment and identification. • Decentralised trial solutions with remote health assessments. • Integration of digital twin technology.	• Reduced delays and financial strain. • More patient-centric and efficient trials. • Enhanced data accuracy and patient safety.
Diagnosis and Early Detection	• Challenges in accurate and timely diagnosis. • Risks of misdiagnosis leading to ineffective treatments. • Limited access and healthcare disparities.	• AI for analysing patient data for early diagnosis. • AI in predicting disease outbreaks and progression.	• Improved disease identification and prognosis. • Early intervention and better patient outcomes.
Drug Discovery	• High costs and time consumption in new drug development. • Low success rates in drug development.	• AI for accelerating drug discovery and repurposing.	• Reduced time and cost in medication development. • Improved treatment efficacy and reduced adverse events.
Precision Medicine	• Rising global burden of chronic diseases. • Complexity in chronic disease mechanisms. • Conventional diagnostic methods often lack precision.	• AI analyses for tailored treatments based on genetic, clinical, and lifestyle data. • AI-driven decision support systems for treatment recommendations.	• Effective treatments with higher success rates. • Reduced risk of adverse effects. • Cost savings by avoiding ineffective treatments.

Potential Opportunity	Traditional Challenges	AI Solutions	Benefits
Remote Monitoring and Wearables	• Need for social isolation during COVID-19. • Shortages of personal protective equipment and reconsideration of face-to-face medical encounters.	• AI-enabled wearables for real-time health data collection. • AI in telemedicine platforms for remote consultations.	• Increased patient engagement and healthier lifestyle choices. • Early intervention avoiding costly procedures. • Enhanced quality of remote care and timely decision-making. • Improved patient outcomes and satisfaction.

SIGNIFICANCE OF RWD QUALITY

Real-World data quality is vital for enhancing patient care, ensuring safety, and optimising healthcare operations. It enables healthcare providers to make informed decisions, preventing medical errors and streamlining administrative tasks. Moreover, quality data supports medical research, helping researchers identify trends, conduct clinical trials, and develop new treatments.

It also aids policymakers in making data-driven decisions about resource allocation and healthcare policies, ultimately improving access to care and reducing costs. Patient engagement is enhanced through access to their own health information, and interoperability allows for seamless information sharing among healthcare providers, improving coordination of care. In summary, quality healthcare data is a linchpin in the effective functioning of the healthcare system, benefiting both patients and healthcare professionals.

What are the challenges and shortcomings in RWE data quality?

Real-world evidence (RWE) has traditionally been used to address post-approval requirements and uncertainties about medical product risks. However, its use has expanded to include capturing clinical outcomes in pragmatic trials and providing disease history data, especially in situations where randomised trials are not feasible, such as in rare diseases or unmet medical needs. RWE has been increasingly employed to support regulatory decisions, including label extensions and accelerated approvals.

Despite its growing importance, RWE's use for demonstrating medical product benefits is debated due to concerns about research design, data quality, and biases in non-randomised studies. Challenges include standardisation issues, data privacy concerns, and a lack of consistent guidelines [9].

How is RWD quality measured?

As the utilisation of real-world data (RWD) has expanded, there has been a concurrent growth in regulatory and policy guidance concerning its use. In recent years, various frameworks and guidance documents have been published by authoritative bodies such as the FDA, EMA, NICE,

Duke-Margolis Health Policy Centre, and the Patient-Centred Outcomes Research Institute, emphasising the critical importance of data quality in this context.

“Real-world evidence (RWE) data quality is defined as fitness for purpose for users’ needs in relation to health research, policy making, and regulation and that the data reflect the reality, which they aim to represent”. ---EMA

It is crucial for ensuring that data collected from real-world sources, such as electronic health records, claims data, and patient registries, can be trusted and used effectively for healthcare research and decision-making. To assess the quality of RWE data, various metrics and dimensions are considered.

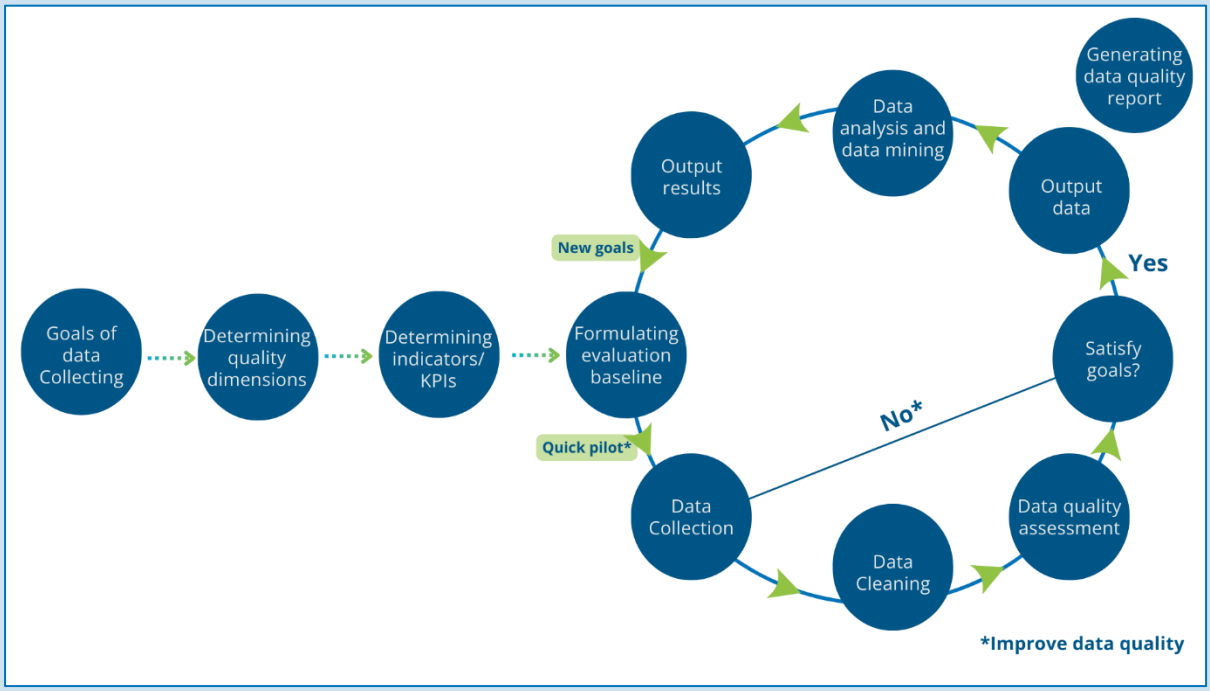
Key quality metrics for RWE data quality:

1. **Reliability**- Reliability in data quality assessment refers to how well data accurately represent reality and can be trusted for their intended purpose. It encompasses precision (the level of approximation), accuracy (the degree of discrepancy), and plausibility (the likelihood of information being true). In the context of the "fit for purpose" definition of data quality, reliability encompasses the correctness and trustworthiness of the data.
2. **Extensiveness**- Extensiveness dimension in data quality assessment combines Completeness and Coverage, addressing how much data is available and whether it's sufficient for the intended purpose. Determinants of Extensiveness depend on how data are collected and specified in the data collection process.
3. **Coherence**- Coherence in the context of data quality refers to the degree of consistency and uniformity in both representation and meaning across different parts of a dataset. It assesses whether a dataset can be effectively processed as a whole, considering factors like consistent data formats, uniform precision, and proper linking of references to entities
4. **Timeliness**- Timeliness in data quality pertains to having data available when needed for regulatory decision-making. It involves collecting and providing data within an acceptable timeframe to ensure it accurately reflects the reality at the intended point in time.
5. **Relevance**- Relevance in data quality refers to the dataset's suitability for addressing a specific research question or purpose. It assesses whether the dataset contains the necessary data elements for the intended analysis. Determining relevance is context-specific and requires linking it to a particular regulatory question or purpose. In some cases, predefined questions can be used to establish relevance within specific domains or areas of use.

AI/ML IN RWE DATA QUALITY ENHANCEMENT

In urgent medical situations, RWE is recognized as valuable, given limitations in traditional evidence gathering. With increasing access to real-world data and evolving analytical methods, RWE's role in clinical development and regulatory decisions is expected to grow. However, these efforts have often been hindered by the limited availability of high-quality, large-scale, and timely data.

FIGURE 3 | Fit-for-purpose data quality framework [10]



Data Governance Framework

Formulating data governance policies and procedures is crucial for maintaining a uniform and standardised approach to data management within an organisation. These policies encompass various aspects of data handling, such as data collection, storage, sharing, and quality control. AI technologies play a pivotal role in achieving consistency and interoperability across different healthcare systems by standardising the format and coding of medical data. Machine learning algorithms can be leveraged to translate between different coding schemes, such as converting from ICD-9/10, to ensure the maintenance of data consistency [12].

Data Aggregation and Cleansing

- **Quality control for missing data**

The problem of missing data in health datasets is a critical challenge, where essential information is often absent. To address this issue, a range of classical and advanced techniques have been developed. Classical methods, like complete case analysis and simple imputation, have been used, but they may lead to a loss of valuable data or introduce biases in the analysis.

Advanced approaches have harnessed the power of deep learning, such as Generative Adversarial Imputation Nets (GAIN) for static data and models like GRU-D, M-RNN, BRITS, and various GAN-based solutions for time series data. These advanced methods have demonstrated remarkable improvements in imputation performance [11].

- **Quality control for tabulation errors**

Structured tabular data errors can encompass incorrect values, typos, inconsistencies in coding systems, or unit-related mistakes. Classical methods for handling tabulation errors involve quantitative and qualitative approaches. Quantitative error detection identifies outliers, while qualitative error detection specifies logical patterns or relational constraints and identifies violations.

Advanced quantitative approaches including RandNet, Deep Autoencoding Gaussian Mixture Model (DAGMM), and various Generative Adversarial Network (GAN) based models have shown impressive outlier detection capabilities, outperforming classical methods. Qualitative error detection has also seen progress with algorithms like HoloClean, PIClean, and MLNClean, which automatically identify, and repair errors based on user-defined integrity constraints and external knowledge [11]. These advanced techniques offer efficient and effective means of handling tabulation errors in healthcare data, ultimately improving data quality and analysis accuracy.

- **Quality control for data duplication**

ML models use record linkage techniques, such as probabilistic matching, to identify and merge duplicate patient records based on common identifiers. When merging duplicate patient records, the system ensures that data for the same patient is consolidated accurately, reducing redundancy.

Data Standardization and Integration

Data standardisation in healthcare involves aligning data elements with established standards, encompassing terminology, coding, and formatting. Overcoming standardisation challenges in electronic health records (EHRs) and medical image data requires the adoption of standards and the application of pre-processing techniques, accounting for both syntactic and semantic transformations. While existing standards like Fast Healthcare Interoperability Resources (FHIR) and Digital Imaging and Communication in Medicine Grayscale Standard Display Function (DICOM GSDF) are beneficial, additional efforts may be needed for comprehensive standardisation. Recent advancements in data processing and machine learning offer solutions to these challenges.

- **Federated Learning (FL)** enables collaborative model training across distributed clients without sharing sensitive data, simplifying data harmonisation across institutions.
- **Image-to-Image Translation**, like Conditional Generative Adversarial Networks (cGANs) and CycleGANs, aids in standardising medical images from different sources, even without paired training examples, by transforming images from one domain to another.
- **Medical Image Standardisation** methods, including CyTran and STAN-CT, employ GAN-based techniques to address issues like contrast and texture variations in medical images.
- **Synthetic image generation**, using GANs to create artificial datasets, helps in training machine learning models when diverse or limited real-world data is available. These advancements enhance data standardisation and image consistency in healthcare.

Natural Language Processing (NLP)

NLP techniques can be used to extract structured data from unstructured clinical notes, improving the overall quality and usability of EHR data. NLP algorithms can detect patterns and relationships within this data, offering valuable information about patient diagnoses, treatment effectiveness, and healthcare utilisation.

Data Quality and Validation

Data quality monitoring employs rule-based or machine learning-based approaches to continuously assess data quality. AI models can automatically validate data against predefined rules, thresholds, or statistical distributions and analyse large datasets for patterns and anomalies.

Multi-Stakeholder Engagement and Education Programmes

To ensure the acceptance and practical implementation of medical AI tools in the real world, it's crucial to involve a wide range of stakeholders throughout the entire lifecycle of these algorithms. This will help ensure that AI tools are designed, validated, and deployed in a way that aligns with the diverse needs and real-world contexts [13].
To ensure the successful adoption of medical AI tools and reduce potential errors, it is crucial to provide comprehensive training to future healthcare professionals in medical AI. This training should cover the advantages, limitations, and risks associated with AI.

CONCLUSION

Over the past decade, healthcare records have undergone a significant digital transformation, primarily aimed at improving efficiency and, in some cases, streamlining billing and reimbursement processes within healthcare systems. The coming decade is poised to shift the focus towards harnessing the insights and value derived from these digital assets. Artificial intelligence is increasingly making its presence felt in the healthcare sector, promising transformative changes in the way healthcare is delivered. It offers personalised, precise, predictive, and portable solutions that can enhance patient care and streamline administrative tasks.

Despite its many advantages, healthcare leaders must address several critical issues when considering the integration of AI into healthcare to take on a more significant role in the medical world [13]-

- **Establishing Ethical and Responsible Data Access Processes:** Healthcare data is exceptionally sensitive, fragmented, and not inherently optimised for machine learning development, evaluation, implementation, and adoption
- **Access to Domain Expertise and Prior Knowledge:** Expertise is indispensable for interpreting data and formulating rules for applying datasets to generate essential insights
- **Access to Sufficient Computing Power:** The ability to make real-time decisions is rapidly evolving, thanks to advancements in cloud computing
- **Research into Implementation Challenges:** It is vital to thoroughly examine and understand the issues that arise when AI algorithms are applied in real-world settings and to ensure the seamless integration of trusted AI algorithms into appropriate workflows

Looking ahead, the development of medical AI algorithms should embrace a co-creation process. This approach entails nurturing a robust and ongoing collaboration between AI developers, clinical end-users, and domain experts will enable the development of AI algorithms that better align with the needs and cultural nuances of healthcare professionals while mitigating potential risks. As AI continues to evolve, it is anticipated to play an increasingly significant role in the medical world, offering improved efficiency, accuracy, and convenience to both healthcare providers and patients.

REFERENCES

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8339486/
2. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/
4. https://osf.io/mjthd/ - Applications of Artificial Intelligence (AI) in healthcare A review.pdf
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7325854/
6. https://www.clinicaltrialsarena.com/comment/transforming-clinical-trials-ai/?cf-view
7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10302550/
8. https://atm.amegroups.org/article/view/48651/html
9. https://cioms.ch/wp-content/uploads/2020/03/CIOMS-WG-XIII_6June2023_Draft-report-for-comment.pdf
10. https://www.thoughtworks.com/en-in/insights/blog/data-strategy/data-quality-framework-part-two
11. https://ieeexplore.ieee.org/abstract/document/9926151
12. https://atlan.com/data-quality-in-healthcare/
13. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8285156/

ABOUT MAXIS CLINICAL SCIENCES

As a boutique firm specializing in clinical research consulting, Maxis Clinical Sciences is committed to optimizing clinical research and development (R&D) processes in the pharmaceutical and life sciences industry. We deliver strategic consulting that drive innovation, efficiency, and improved patient outcomes. With a deep understanding of the complex challenges faced by our clients, we provide comprehensive solutions that encompass clinical trials design, development, real world evidence (RWE) solutions, data analytics, all geared towards providing care that is tailored to individual patient requirements.

510 Thornall Street, Suite 180 Edison, NJ 08837
T+1 (732) 889-2444
www.maxisclinical.com
info@maxisclinical.com