Artificial Intelligence - A Game Changer in Unifying Clinical Data

Introduction

The rapid growth of healthcare data is fueled by EHRs, Wearables, and SMAC Technologies.

A huge amount of clinical data is generated today through EHR (Electronic Health Records)/EMR (Electronic Medical Records), clinical trial systems, lab reports, wearables, remote monitoring devices, and the likes. Proliferation of digital health apps and SMAC are driving an unprecedented growth in healthcare data. As per a report by RBS Capital markets (1) every second, an exponential amount of healthcare data is generated and mined for valuable insights. Today, approximately 30% of the world's data volume is being generated by the healthcare industry. By 2025, the annual compound growth rate of data for healthcare will reach 36%. That's 6% faster than manufacturing, 10% faster than financial services, and 11% faster than media & entertainment.'

However, this humongous and valuable data sits in silos, fragmented across various systems. Besides residing in disparate systems, it is non-uniform; uses varied metadata; stored in different formats with varied levels of details; and embeds natural language and images. This paper focuses on the use of Artificial Intelligence (AI) Technology in unifying clinical data across systems for making clinical trials more effective, safer and more efficient, which can bring revolutionary progress in healthcare.

This paper focuses on the use of Artificial Intelligence (AI) Technology in unifying clinical data across systems for making clinical trials more effective, safer and more efficient, which can bring revolutionary progress in healthcare.

Clinical trial challenges

Clinical trials often face significant challenges due to inconsistent metadata across systems, even within the same sponsor or CRO organization.

During Clinical Trials, even within the same sponsor or CRO organization, multiple systems use different metadata. For example, the schema of the data captured by the electronic data capture (EDC) systems is different from that captured in the safety systems. Sites input almost the same data repeatedly on different systems, increasing the load on sites. Another example is that the analytics on site performance generated by a study monitoring system are often not fed into the feasibility systems for site rankings. Pharma companies struggle to integrate clinical and omics data, and many of them integrate clinical data from multiple studies – even if these are from the same therapeutic area (TA).

This siloed data leads to many missed opportunities in gaining holistic insights, which would be very useful in bringing adverse events to the fore, improving efficiencies in R&D and healthcare in general. Today if patients go from one healthcare provider to another, they themselves will have to carry their health records. The diagnosis and treatment for rare diseases remain localized. In the absence of consolidated data, it is difficult to identify diverse eligible patients who could benefit from participating in clinical trials. Due to the fragmented data, the safety signals are not detected in time. There are many missed opportunities.

Standards

Bridging the Gap in Clinical Data Interoperability industry standards

Compare this to other industries where data interoperability is a basic requirement. We may be having an account with a certain bank, but we can use our credit and debit cards across ATMs of all banks and across merchants globally. The financial systems talk to each other seamlessly using SWIFT standards. The same is the case with supply chain systems of the world, which use ISO standards to communicate and have required access to other systems for transactions and analytics seamlessly.

CDISC (Clinical Data Interchange Standards Consortium) has done tremendous work over the years in standardizing data structures thus enabling easy interchange and consolidation of clinical data. Data submission to the health authorities is now largely streamlined. However, the CDISC standards are not fully adopted globally owing to challenges related to initial investments, differences in requirements of various health authorities and not enough ROI for small and medium sized pharma companies. The integration with legacy systems, and integration of clinical trials' data with real world sources remains a challenge. The data standards used by providers and pharma sponsors are different and so difficult to integrate and consolidate.

AI in clinical data

Unification of clinical data is on the verge of a huge transformation with Artificial Intelligence technology. AI is driving a paradigm shift in deciphering, mapping, interchange and interoperability of health data. AI has the potential to consolidate data to create a unified view - something that earlier technologies have not been able to effectively implement at scale owing to the dependence on implementation of standard metadata and structures.

1. **AI brings the following advantages over traditional technology: Convert natural language to structured data:** The AI models understand natural language easily. They can further be trained to understand clinical data based on context just like humans. AI can read and understand natural language and can convert unstructured data to a structured format.

   This capability is very useful in extracting relevant clinical data from the massive text available in-patient interfacing apps, doctors' notes, social media, published articles, to mention a few. AI can summarize and extract relevant information that too in a structured format, from lengthy clinical documents. It can convert all this information into standardized medical codes.

   For example, from an adverse event description given by a patient in natural language, AI can identify severity, causality, product name and concomitant medicine and other important information and structure in a standard form.

2. **Fuzzy matching:** AI is highly effective in handling inexact matches – especially where there is a need to **understand** the context before matching. AI can map data points, which are **different looking** but have **similar meaning** due to its ability to establish fuzzy or inexact mapping. AI

models can match phrases that are semantically similar but have different words. This helps in solving the challenges related to quality of data, local variations in terminology, and contexts.

Fuzzy matching can help in reconciliation of drugs (e.g. acetaminophen and paracetamol), diseases (e.g. rubeola and measles) and adverse events (e.g. difficulty in breathing and breathlessness), dosage regimes (e.g. 100 mg every 12 hours or 12mg twice a day, or 100mg 2x/day) which have the same meaning. Fuzzy matching can effectively map words with spelling errors or colloquial terms.

3. **Data extraction sans template:** AI does not depend on pre-defined templates or field-labels or metadata to **understand** the data. This is like how humans can look at data in a form (such as gender, age, address… etc.,) without looking at the field labels and understand very easily what that data represents. OCR (Optical Character Recognition) and pre-trained models can examine special arrangements, field labels, tables and graphical elements to understand form's layout and content. For example, information in the lab reports from various labs can be read, mapped and consolidated to create a cohesive view.

4. **Text summarization**: AI tools can summarize the text or extract **relevant** information from the text and match it with structured data or another text. This allows combining two different data types (e.g. text and International Classification of Diseases [ICD] code), with structured data to supplement each other and create a more insightful conclusion. Clinical notes often have detailed descriptions of patient's history, symptoms and treatments. AI can extract core medical details and align them to standardized formats. This feature also is very useful in accessing relevant information from RWE.

5. **Learnings from historic data:** Now this feature is the most revolutionizing feature of AI. AI models can be trained on vast historic data, which has been processed by traditional technology or by humans and create synonyms and mappings for future use. AI models are very good at understanding the rules applied on the data previously and the context in which those were applied. These learnings can then be applied on any unprocessed or data silos to process the data using the same contextual rules. This feature can help in improving patient recruitments, predicting adverse events, improve selection of sites, predicting efficacy patterns, identifying potential non-compliance there are numerous use cases.

6. **Pattern recognition for text and image mapping**: AI models can recognize patterns in non-textual data, such as images with similar visual features. It can help in merging image data (e.g. lesions size and colour, tumour size changes). It can compare images taken from different machine types with different protocols, resolution and provide consistent inferences. AI can also label and annotate the data which helps in comparison with similar images taken in the past, and across patients and sites.

Key Success factor

AI has the potential to revolutionize healthcare by unifying clinical data, accelerating trials, enhancing safety, and improving success rates, but requires collaboration, data sharing, and regulatory support to create a seamless digital ecosystem.

While there are multiple AI applications being implemented by pharma sponsors to improve efficiency, accuracy and compliance, using AI for unification of clinical data has a global strategic significance. It will have a huge impact on accelerating clinical trials, making them safer, making trial more patient centric, and overall improving the success rate of clinical trials. AI is the biggest game changer which will revolutionize healthcare.

Collaboration and willingness to share data in an organization or ecosystem, while complying with data privacy guidelines, is a critical requirement for technology to succeed. Once the initial pilots are successful, health authorities and consortiums must take steps to create a digital ecosystem with seamless data interchange among all stakeholders.

About the Author

Rachna Malik, Global head, TCS ADD™, TCS

Rachna Malik is the Vice President and Global Head of the TCS ADD Platform at Tata Consultancy Services, a global leader in IT services, consulting, and business solutions. The TCS ADD Platform offers innovative technology solutions to pharma companies for digitizing clinical trials. The platform, powered by artificial intelligence, accelerates the clinical R&D value chain and makes clinical trials more agile and safe.