Averting Crisis in Health Care Delivery through Knowledge Engineering and Artificial Intelligence

Digital Transformation

By Dr Asoke K Talukder,

Co-founder, Chief Science & Technology Officer,

Vibrant Health Limited

CONTENTS

ABSTRACT

The health care vertical is in crisis globally due to limitations inherent in prevailing Health IT (HIT) systems. These systems are known as EMR (Electronic Medical Records) or EHR (Electronic Health Records). These systems are storage and retrieval silos for medical records. They do not have any intelligence other than structured storage and quick retrieval of medical data. They do not interoperate. These systems are brainless, toothless, and armless systems. The data entered into these systems are not machine understandable; therefore, an EMR system cannot distinguish between correct and incorrect data. This data can be consumed only by human experts. To reduce the disease burden and improve the public health, knowledge engineering and artificial intelligence needs to be added along with functional modules such that these systems can work as independent expert solutions. In this article we present a model of the Next Generation Health Care systems developed by Vibrant Health Sciences. This intelligent system includes advanced care assessment and delivery solutions to accomplish these critical goals, averting the health care crisis through dramatic digital transformation.

INTRODUCTION

The disease demography across the world is changing. Deaths caused by infectious or communicable diseases are generally stable; whereas, deaths due to Noncommunicable diseases (NCD) are increasing. In USA 88% of all deaths are due to NCDs, including heart disease, strokes, cancer, diabetes, and chronic lung disease being at the top [13]. According to WHO (World Health Organization), 70% of all deaths in the world in 2015 were due to NCDs [14] out of which 42% are premature and avoidable [15]. Because of aging population trends, geriatric disease penetration including cancer is experiencing a steep growth rate. According to recent studies, 1 in 2 people in the UK will get cancer sometime in their life [1]. According to another study, cancer will affect 1 in 2 men and 1 in 3 women in the United States [17].

NCDs are caused by genetic or lifestyle related pathogenicity. NCDs present themselves in various overlapping and confusing symptoms (phenotypes). Unlike communicable diseases, which are caused by external agents and generally have distinguishable markers, NCDs are difficult to diagnose. This results into late onset of NCD recognition and medical errors are common. Researchers in Mayo Clinic in the US found that 88% of patients were misdiagnosed at the primary care center. This includes more than 21% of patients wrongly diagnosed and 66% of patients incorrectly diagnosed [12]. These medical errors are sometimes fatal. Research suggests that medical error is the third major cause of death in the US [8]. To make things even worse, there are many health care service encounters that are not required. About one third of health care spending in the US is unwarranted – this includes unnecessary medical tests, excess medication, unnecessary procedures and unnecessary health care [4].

The OECD (The Organisation for Economic Co-operation and Development) average of per-capita spending for health care is USD $3,484. Table 1 presents the health of the health care system in some high income countries (HIC) and middle & low income countries (MLIC). A majority of the countries in Africa have less than 0.1 physician density per thousand and health care spending less than USD $100 per-capita. Globally health care is becoming unaffordable and most importantly unsustainable [2]. WHO's global business case for NCDs showed that spending of an additional USD $1.27 per person per year in low-and lower-middle-income countries can save 8.2 million lives and generate $350 billion by 2030, a 7:1 return on investment (ROI) [6]. The additional fund of $1.27 per-capita additional funding recommended is to be spent on (a) ***digital transformation****, (b)* ***training of human***

resource *(physicians) and (c)* ***promoting P6 medicine*** *(Participatory, Personalized, Proactive, Preventive, Predictive and Precision) Medicine [10].*

Table 1: The health of healthcare

Country	**USA**	**Switzerland**	**Luxembourg**	**Norway**	**Germany**	**Austria**	**Greece**	**Australia**	**China**	**Cuba**	**India**
Per-capita Spending (in USD)	9,892 (2016)	7,919 (2016)	7,463 (2016)	6,647 (2016)	5,551 (2016)	5,227 (2016)	2,366 (2016)	4,708 (2016)	480 (2012)	817 (2014)	157 (2012)
Doctor density (per 1k)	2.568 (2014)	4.248 (2016)	2.921 (2016)	4.385 (2015)	4.191 (2015)	5.23 (2016)	6.255 (2013)	3.496 (2015)	1.812 (2015)	7.519 (2014)	0.758 (2016)

Source: https://en.wikipedia.org/wiki/List_of_countries_by_total_health_expenditure_per_capita

Source: Density of physicians (total number of physicians per 1000 people in the population, latest available year) (http://www.who.int/gho/health_workforce/physicians_density/en/)

While the world as a whole has embraced ICT (Information and Communication Technology) in business and daily life, health care is lagging far behind. In the US, policies and regulations required hospitals to implement advanced capabilities of certified electronic health records (EHRs). This has led to accelerated implementation of health information technologies (HIT) in health care settings [16]. In HIT, the health records are being transformed into electronic or digital forms such that, it will be easy to retrieve a health record. However, these electronic records in HIT are useful for human consumption only – they are not machine understandable – computer applications cannot use them. Many of these records (clinical notes and prescriptions) are in scanned image format and cannot even be read by computers. This condition is like a species without teeth, arms, and most importantly, without a brain. To ensure a healthy health care system, it is necessary to add the "brain" power, and other important parts to the HIT system through (a) ***artificial intelligence*** *(AI), (b)* ***knowledge engineering*** *(KE), and (c)* ***mobile applications****. Digital transformation is the only way to make HIT a complete solution that will reduce the disease burden and counter the medical crisis.*

To take full advantage of digital transformation, the medical domain has to overcome many complex challenges. They are,

1. ***Normalized Input data****: Collating all disease related medical data of a person from different data sources like EHR, Pathology (Laboratory, Radiology etc.), Genomics (DNA, transcriptomic etc.), IoT (Internet of Things) and then normalize the same in a unified way to obtain the person's disease history. There are many challenges in this step. There are duplicate data, missing data, and erroneous data in the databases. Moreover, the data is fragmented and distributed – stored in many clinics, hospitals, and diagnostics centers with proprietary protocols and incompatible storage formats. There were attempts though to create HIE (Health Information Exchange) to address this gap with limited success.*

2. ***Machine Understandable Biomedical information****: Being electronic, health data is expected to be machine readable. Health data is unstructured and non-numeric descriptive texts. The data does not have*

any meaning outside of the context. To take advantage of automation, health care data needs to be converted not only into machine readable but also machine understandable formats. We need to utilize NLP (Natural Language Processing) based artificial intelligence to resolve this challenge.

3. ***Knowledge Engineering****: The Intelligence in synthetic systems like banking, retail, or airline reservation systems are statically built in consultation with domain experts. Domain experts ensure that the correct business knowledge is rightly added inside the IT (Information Technology) solution. Once built these domain knowledge rarely change. In medical science however there are about 250 overlapping symptoms for about 10,000 diseases. In cancer alone there are more than 700 different subtypes (or diseases). Moreover, the understanding of disease is constantly maturing; therefore, medical knowledge cannot be statically added into any medical application. The intelligence in a medical system therefore must be dynamically supplied through a knowledge engineering module at runtime. This will ensure that we have up-to-date and just-in-time accurate knowledge. The knowledge here will be a combination of (a)* ***clinical knowledge*** *that will help identification of clinical signs and symptoms, and (b)* ***biological knowledge*** *that will help understand the underlying biology that is driving the disease.*

4. ***Clinical Reasoning:*** *In clinical reasoning, a clinician collects, processes, and interprets a patient's information to develop a hypothesis. A decision is derived from this hypothesis based on the concordance of multiple knowledge bodies that are generally assumed to be true. The therapeutics part however will need prediction of the likely outcome. To formulate the best outcome, Bayesian probability is used. For digital adaptation of this phase, there will be two distinct stages of digital transformation: (a)* ***probabilistic outcome*** *computation, and (b)* ***deductive and inductive reasoning*** *suggestions.*

In this article we present the use case of making a health care system intelligent, showing how and where we are going far beyond the human consumable health records of today's HIT systems. In this digital transformation of current health care systems that we built at Vibrant Health Sciences, we will interpret the medical text created by human specialists, make smart robotic decisions to facilitate accurate medical decisions, adding health care intelligence and reducing disease burden to facilitate improved public health.

DETAILS OF THE PAPER

*Various factors are pushing medical care towards a medical crisis. At the center of this crisis is the change in disease patterns combined with the cost of care and the scarcity of qualified and trained medical professionals. In the 21st Century, following the Human Genome project, and the availability of the many algorithms and technologies, the growth in the scientific understanding of diseases in the population has been unprecedented. It is humanly impossible for a unaided health care professional today to deliver timely medical care with the efficacy, consistency, safety, and accuracy – that the full range of advanced knowledge could support. The aid a medical professional will need is "just in time" knowledge and evidence. They will need filtering out of unnecessary noise data. Moreover, multiple layers of artificial intelligence are required to be built over and above the EHR. EHR is a system without a brain; an AI system therefore will function like the neocortex. This layer of artificial intelligence will bring in transparency, accuracy, increased productivity, and reduce risks, and eliminate wastage. Fig. 1 shows the architecture of such a health system that we call "***Proof of Health*** *(PoH)" system.*

The PoH system comprises of many functional modules. These modules are:

1. *Spatial and temporal health data*

2. *Cognitive Knowledge base (Biomedical Evidence)*
3. *Proof of Disease*
4. *Data Storage*
5. *Deductive & Inductive Reasoning*

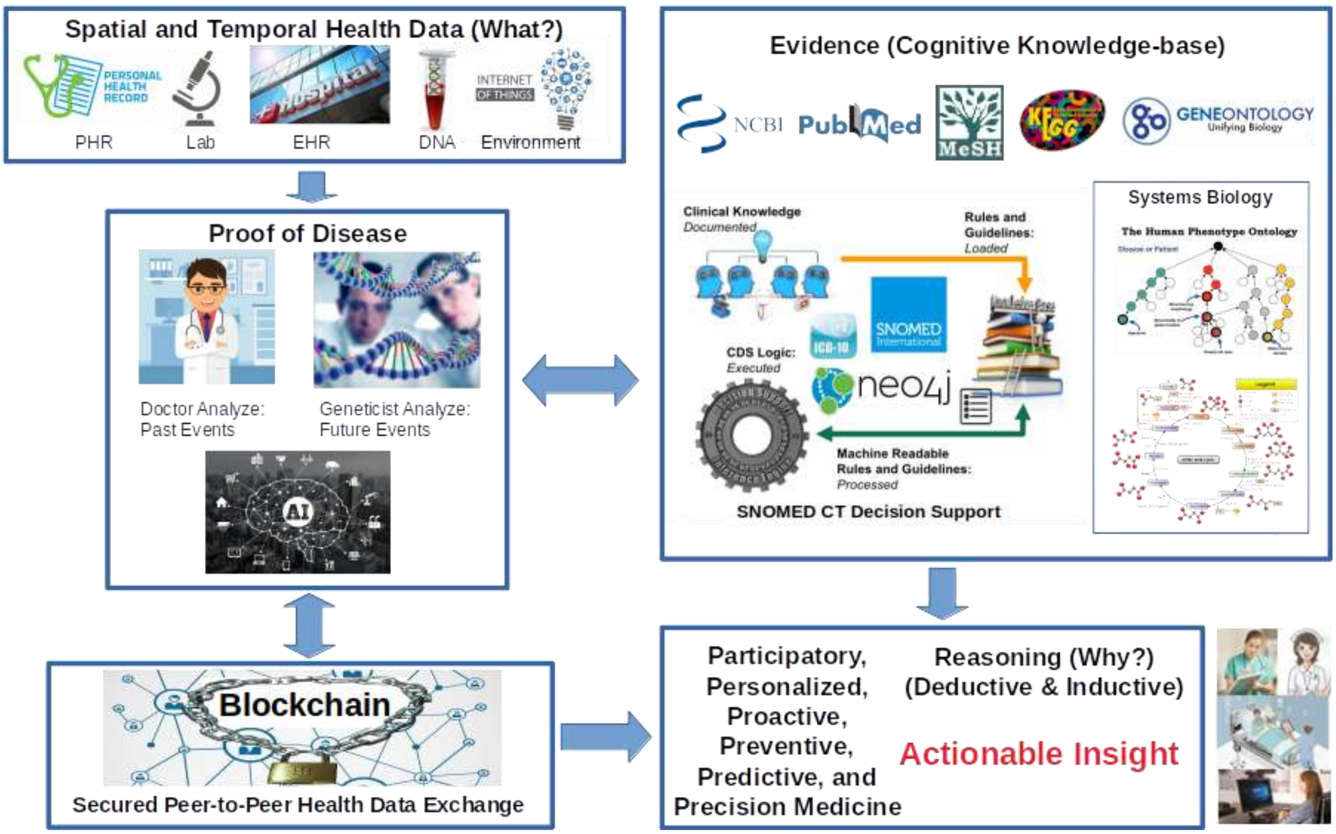

Fig 1: Architecture of the ***Proof of Health*** *(PoH) system. This is a future ready health care system architecture that will add decision theory based* ***artificial intelligence –*** *driven by* ***semantic knowledge network*** *created from Snomed CT, ICD, Gene Ontology, hospital data, and many biological databases. It will lead an expert to arrive at an* ***actionable insight*** *quickly and accurately.*

Patient Data

This is the functional area where health care data is generated. Health care data has many flavors like,

1. ***Clinical note, discharge note, progress notes****: These are data related to the complete documentation of events of a medical encounter of a patient. These data are unstructured texts written in natural languages.*

A synthetic system (like a banking system) is created by humans. Synthetic systems are created with adapted structures to fit into computer systems' capabilities using tools like UML (Unified Modeling Language). In contrast, biomedical systems are unstructured natural systems. To understand a natural system, a computer system has to adapt itself to the natural properties. This is the first step of the complex automation. The National Library of Medicine's (NLM) in 1986 created Unified Medical Language System (UMLS), which is a compendium of many controlled vocabularies in the biomedical sciences. The purpose of UMLS is to facilitate the development of computer systems that behave as if the computer system "understands" the meaning of the language of biomedicine and health [11].

2. ***Laboratory and pathology data**: These data relate to various biochemical tests of body fluids like blood etc. For diseases like cancer, these could be histopathology reports or tissue biopsy. If this is a radiology data, it will include the report in DICOM (Digital Imaging and Communications in Medicine) standard. This data is more structured compared to clinical notes or discharge summary.*

3. ***Genomic data**: These are unstructured big data relating to the genomic tests for mutations (variations) in the DNA. Many diseases are passed on by parents to their children through the genetic material children inherit – these are known as germline mutations. Age, lifestyle, and environmental factors also cause the human genome to change causing many diseases over time that are called somatic mutations. Such mutations produce pathogenic disease causing proteins in the body. The NGS (Next Generation Sequencing) technology takes the human genome and then smears it onto small random fragments. The nucleotides in these small random fragments are read by the NGS machine to generate digital data. A human genome is about 3 billion nucleotides long whereas the protein coding region within a genome is about 65,000 nucleotides. The **exome** genomic data generated by NGS machine of only the protein coding region of one person is of the order of 12GB (Giga Bytes). The small random fragments of genomic data an NGS machine generates is about 300 nucleotides long. Complex computational and statistical algorithms are required to reconstruct the genome inside a computer followed by discovery of the mutations through a series of complex analysis of big-data. The details process of discovery of cancer in patients through the big-data genomic analytics of exome data is described in one of our previous work [10].*

4. ***IoT and Environmental data** – these data relate to the environment and lifestyle captured through the IoT devices. Some of these data will capture the disease state and diseases status. These data will also include the workout status.*

Evidence

The majority of NCDs are systemic diseases and not limited to just an organ. Systemic disease affects the entire body as a system. Therefore a proper decision about an NCD demands support from population data and external knowledge bodies. A study with general internists in the US showed that only 5.8% difficult cases could be diagnosed correctly with the present "brainless" EMR systems. The study also revealed that physicians are relatively insensitive towards the diagnostic accuracy and examine all cases through the same benchmark tools (intuition). Physicians are quite confident about their diagnosis and unlikely to reexamining difficult cases or look for evidence where their diagnosis may be incorrect [3].

The philosophy of Evidence Based Medicine (EBM), is to change this. EBM is an approach whereby collection of latest and best knowledge is used as reference to make the best and accurate medical decision. The greatest achievements of EBM have been the development of systematic reviews and meta-analyses of biomedical research. Through this process practitioners identify multiple studies on a topic, separate the best ones and then critically analyze them to come up with a summary of the best available decision [7].

In our model, we constructed a cognitive biomedical knowledge base as evidence. This has mainly three components; viz., (a) ***Systems Biological knowledge base***, (b) ***Clinical knowledge base***, and (c) ***Research and experimental knowledge base***.

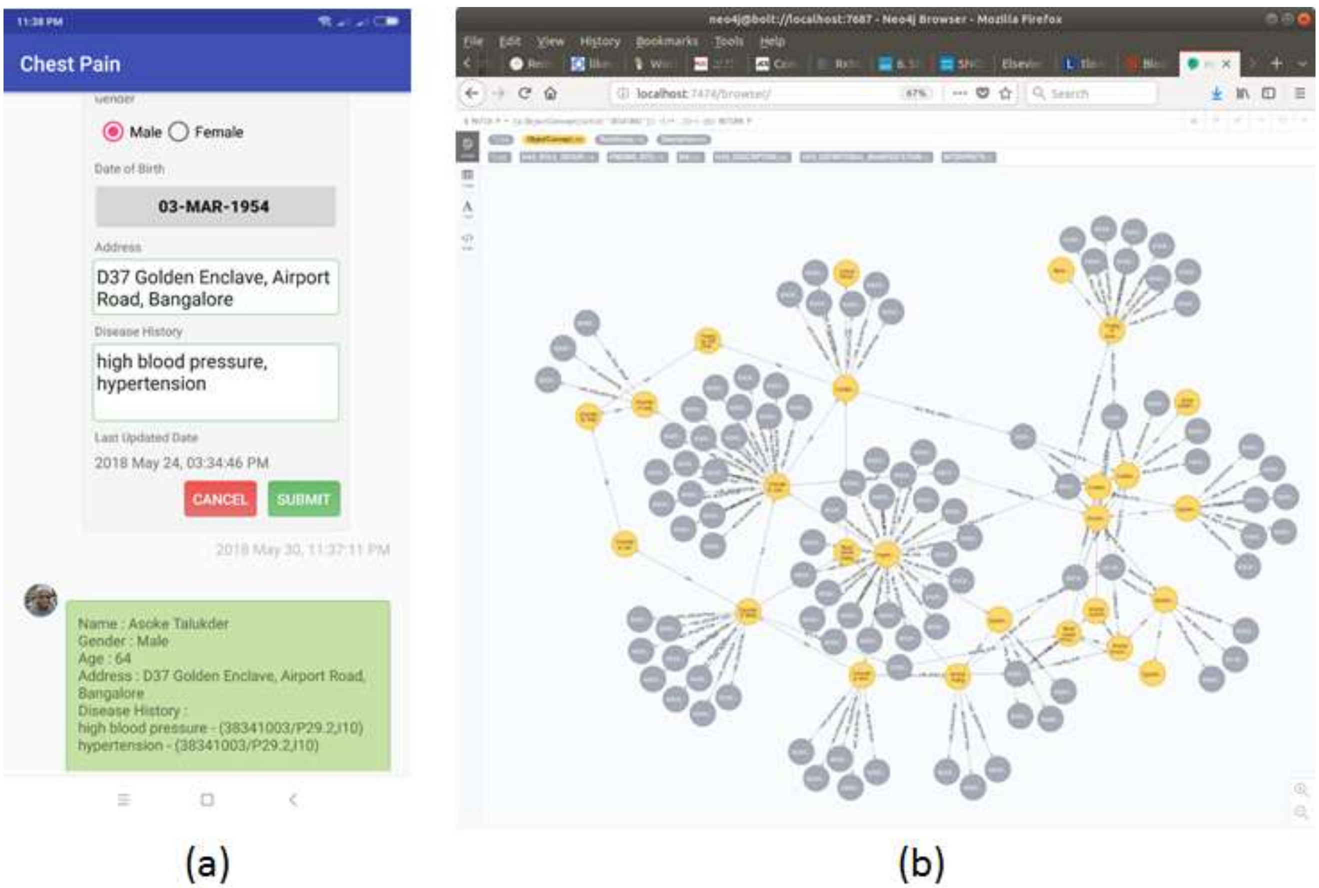

Fig. 2: Machine understandable disease ontology

Proof of Disease

*The architecture of **Proof of Disease** (PoD) module is shown in Fig 1 and how it fits into the overall scheme of P6 Medicine. In this functional module experts arrive at a consensus about the presence of a disease. Mobile applications or desktops are used as user devices to collect the symptoms and phenotypic details (Fig. 2). The patient uses the user application (mobile or desktop) to enter pathophysiological details of the disease in simple English or local languages. This is a multilingual application where disease related information is collected through autonomous robotic AI software. The vital information and the disease history is taken from the EMR/EHR. In case*

EMR/EHR data is not available (or offline), vital information is entered through user facing applications. Fig 2 shows an example of digital transformation. In this mobile application, as an example, we entered two diseases viz., "high blood pressure" and "hypertension". For a computer software, "high blood pressure", and "hypertension" are two different character strings and they are unequal. However, any knowledgeable person in the medical domain, can easily say that both "high blood pressure" and "hypertension" mean the same. Using NLP and medical corpus medical concepts are identified. Then we use medical thesaurus to normalize the concepts. In the above example using such techniques we resolved both "high blood pressure" and "hypertension" to the same Snomed CT code "38341003" (Hypertensive disorder, systemic arterial (disorder)) and ICD10 code "I10" (Essential (primary) hypertension).

Fig. 2b shows the Snomed CT semantic subgraph for "hypertension" for Snomed CT code 38341003 with the diameter of 3. SNOMED CT (Systematized Nomenclature of Medicine -- Clinical Terms) is the most comprehensive and precise clinical health terminology product in the world, owned and distributed by SNOMED International (https://www.snomed.org/). In this subgraph, yellow nodes are concept node and gray nodes are attributes and symptoms. This is created through Snomed CT [5] in Neo4j (https://neo4j.com/) Graph database. Fig 2b shows a multi-partite graph of hypertension in Neo4j.

During the PoD phase doctors look at the the patient's current encounters and details of signs, symptoms, and findings of the current disease episode. In PoD phase doctors also look at disease history of all previous episodes and encounters already stored in the blockchain P2P (Peer-to-Peer) database. Blockchain being a P2P network, disease data is ubiquitous and accessible universally through a security approval. The P2P data includes all types of medical data that comprises of pathophysiology and genomic data in machine understandable ontology codes. In this blockchain database as the diseases are stored in ICD and Snomed code, it is possible to examine the patient holistically quickly instead of just a disease or an episode. From all this findings doctors construct the disease trajectory and decide about the best possible line of treatment. For complex diseases like cancer, the PoD will comprise of a panel of oncologists or a Tumor Board that will examine all clinical, medical, radiological, and genomic data.

Data Storage

The data storage in PoH system is hybrid database in nature. For example, in synthetic systems (like banking or airline ticket reservation), we box the data through normalization (1NF, 2NF, and 3NF) where main focus is ease of data access. In such systems objects grow faster compared to relationships. In natural systems like social networks or biomedical system the relationships grows faster than the objects. RDBMS is not suitable for such natural systems. Graph database is ideal for such applications. Moreover, in health care a patient will talk to a doctor, pathologist, advocacy groups etc – it demands a secured peer-to-peer network. We use three data storage access methods for three major functional areas of health sciences,

1. ***Relational Database (MySQL)**: Relational database store the structured part of the data like patient detail, drug detail, hospital equipments, hospital administration, human resource, inventory etc. It will store the unstructured patient and disease related documentations.*
2. ***Graph Database (Neo4j)**: The cognitive knowledge base in our model stores knowledge that is used as evidence. Knowledge is represented through controlled vocabulary concepts and relationships amongst them that can formally describe diseases, symptoms and findings. The knowledge in our model is stored in Neo4j graph database through a semantic network of concept (Fig 2b). It is the formalism of the biomedical knowledge that is used for decision making. This cognitive knowledge is represented through a semantic network with semantic relationships amongst concepts. This includes knowledge from Snomed CT, Kegg*

(https://www.genome.jp/), Gene Ontology (http://www.geneontology.org/), OMIM (https://www.omim.org/), and many more Knowledge represented through knowledge network.

3. ***Distributed Database (Blockchain)****: This is an immutable, secured, peer-to-peer network. Health data needs many to many peer-to-peer access and interaction without any intermediary. Blockchain is the most appropriate database for these function. Security and accessibility will be resolved through ERC20 compatible smart contracts. We have developed peer reviewed blockchain architecture for integration with our next generation health care solutions already developed.*

Reasoning

*This is a functional module that relies on decision theory. In this module the final decision is made by a doctor. Our model system supplies all the necessary information required for an accurate and informed decision. Here we use both deductive and inductive reasoning supported by Bayes probability. Through normative process the system will determine the highest utility under conflict and constraints. We started with clinical data where a patient is presented with various clinical features. In a way data relating to a clinical encounter describes one instance (one episode) of the "****what****" – a single instance of fact. We use the detail of previous episodes from the blockchain distributed database to get a systemic view of the person. We use the biomedical knowledge base for deductive (or inductive) reasoning. Through reasoning, we derive the answer to the question "****why****"? Why do we see these clinical features in the patients? High probable answers with high utility are presented to the doctor. The doctor chooses the best answer. Once the doctor knows the answer to the question why, he can decide about the right treatment and therapeutics to optimize the outcome. This combination of AI and human intelligence is optimal for real world implementation success.*

Project Management Challenges

The project management challenges in health care transformation are daunting. Highly skilled and interdisciplinary teams are needed where people from various disciplines like ***Biological sciences****,* ***Medical sciences****,* ***Data sciences****,* ***Computer sciences****,* ***Mathematical*** *and* ***Statistical sciences*** *will be working together. Moreover, the team members will be of different age group – young to old with varied academic degrees from MD to PhD. The technology domains will also be very wide – it will include hybrid databases like noSQL, RDBMS, Graph, Blockchain distributed databases, knowledge engineering, and artificial intelligence. Such a team requires expertise in* ***cloud computing****,* ***unstructured Genomic Big-data****,* ***NLP****,* ***AI*** *&* ***Data mining****,* ***Knowledge Engineering****,* ***Graph theory*** *etc.*

These interdisciplinary teams will achieve outcomes through:

1. ***Shared Vision***
2. ***Trust and Communication***
3. ***Leadership***

Some other project and team management qualities worth considering in such complex project are,

1. ***Have a good understanding of academic culture***
2. ***Do not micromanage***
3. ***Serve as a mentor***

4. ***Build confidence and trust through mutual respect***

CONCLUSION

*Digital Transformation and use of knowledge engineering along with artificial intelligence solutions in the Health Care Sector will bring in transparency, repeatability, and accuracy. In the First Generation of Health Care systems, medical records were handwritten that could be read and understood by human experts (doctors). In the Second Generation Health Care system we have seen electronic medical records but we still need human experts to understand and interpret these records. In the Next Generation (3rd Generation) Health Care system, however, health records will be machine readable and machine understandable. This machine interpretable health records will be realized through a layer of actionable artificial intelligence. The objective of this health care transformation solutions will be interoperability to ensure better and affordable public health. In this article we have presented an architecture that is developed by Vibrant Health Sciences. This architecture combines knowledge engineering and artificial intelligence to solve the medical care challenges – one of the most complex and rewarding problem computer science has ever encountered. We showed how such an integrated system will transform the health care practice. Here we presented the architecture and the information flow of a patient doctor interaction that we call "**Proof of Health (PoH)**". In PoH the patients narrates the story of the disease or the ailment he or she has – the "**what?**". The computer system looks at the knowledge body and presents answer to a doctor in the form of "**why?"** with the help of artificial intelligence. The doctor picks up the best answer with highest utility. This architecture adds value to the EMR/PHR that will help in following ways*

1. ***Increase productivity of physicians and clinics through automation –*** *that will increase the footprint of medical care*
2. ***Reduce the cost of medical care***
3. ***Improve Outcomes***
4. ***Increase accuracy in medical decision making and reduce the error in medicine***
5. ***Reduce the wastage by reducing the over-treatment***
6. ***Avert the Medical Crisis the World is expected to encounter***

REFERENCES

[1] 1 in 2 people in the UK will get cancer. Cancer Research UK. http://www.cancerresearchuk.org/about-us/cancer-news/press-release/2015-02-04-1-in-2-people-in-the-uk-will-get-cancer

[2] Cost of Care: Health Care Costs Rising To Unsustainable Levels – http://www.genfkd.org/cost-care-health-care-costs-rising-unsustainable-levels

[3] Ashley N. D. Meyer, et al., Physicians' Diagnostic Accuracy, Confidence, and Resource Requests A Vignette Study. JAMA Intern Med. 2013

[4] Atul Gawande. (2015). Overkill. Analysis of Health Care, The Newyorker, May 11, 20 issue. (https://www.newyorker.com/magazine/2015/05/11/overkill-atul-gawande)

[5] Campbell WS, et al.. An alternative database approach for management of SNOMED CT and improved patient data queries. J Biomed Inform. 20 15 Oct;57:350-7

[6] Nishtar Sania et al., Time to deliver: report of the WHO Independent High-Level Commission on NCDs. The Lancet, 2018

[7] Masic I, Miokovic M, Muhamedagic B. (2008). Evidence based medicine - new approaches and challenges. Acta Inform Med. 2008;16(4):219-25. doi: 10.5455/aim.2008.16.219-225.

[8] Medical error—the third leading cause of death in the US. (2016). BMJ 2016;353:i2139. https://doi.org/10.1136/bmj.i2139

[9] Talukder AK, Ravishankar S, Sasmal K, Gandham S, Prabhukumar J, Achutharao PH, et al. (2015) XomAnnotate: Analysis of Heterogeneous and Complex Exome- A Step towards Translational Medicine. PLoS ONE 10(4): e0123569. https://doi.org/10.1371/journal.pone.0123569

[10] Talukder AK. (2017). Big Data Analytics Advances in Health Intelligence, Public Health, and Evidence-Based Precision Medicine. LNCS, volume 10721, Springer

[11] About the UMLS - https://www.nlm.nih.gov/research/umls/about_umls.html

[12] Van Such M, Lohr R, Beckman T, Naessens JM. (2017). Extent of diagnostic agreement among medical referrals. J Eval Clin Pract. (4):870-874

[13] WHO. Noncommunicable diseases country profiles 2014. - http://www.who.int/nmh/countries/en/

[14] WHO. NCD mortality and morbidity – http://www.who.int/gho/ncd/mortality_morbidity/en/

[15] WHO. Noncommunicable diseases prematurely take 16 million lives annually, WHO urges more action – http://www.who.int/mediacentre/news/releases/2015/noncommunicable-diseases/en/

[16] Yen PY, McAlearney AS, Sieck CJ, Hefner JL, Huerta TR. (2017). Health Information Technology (HIT) Adaptation: Refocusing on the Journey to Successful HIT Implementation. JMIR Med Inform. 2017 Sep 7;5(3):e28. doi: 10.2196/medinform.7476.

[17] Zosia Chustecka. Cancer Strikes 1 in 2 Men and 1 in 3 Women (2007). https://www.medscape.com/viewarticle/551998