Future of Large Language Models and Digital Twins in Precision Healthcare: A Symmetric Literature Review

[1]Neel Shah, [2]Dr. Nirav Bhatt, [3]Dr. Nikita Bhatt

[1][2] Department of Artificial Intelligence And Machine Learning, Chandubhai S. Patel Institute of Technology, CHARUSAT University, Changa, India
[3] Department of Computer Engineering,
Chandubhai S. Patel Institute of Technology, CHARUSAT University, Changa, India

Abstract — Digital twin and large language model technologies have been increasingly applied in precision healthcare and patient applications in recent years. This publication fills the research gap by providing an overview of the recent advances, applications, and challenges of digital twins and large language models in precision healthcare. It also proposes a state-of-the-art technology that combines a large language model and a digital twin that can be used to create models specific to patients to help with diagnosis, treatment planning, therapy planning, checking the effectiveness of drugs on individuals, and many other cases. And with this proposed technology, the healthcare and pharmaceutical industries can be revolutionized.

Keywords — Artificial Intelligence, Large Language Models, Digital Twins, Precision Healthcare, Medicine

1. INTRODUCTION

Healthcare is entering a new era where abundant biomedical data plays an increasingly important role. In this context, precision medicine, for example, attempts to "ensure that the right treatment is given to the right patient at the right time" by taking into account several aspects of patient data, including variability in molecular characteristics, environment, health records, and lifestyle [18].

Generating actionable insights and knowledge from high-dimensional, complex, and heterogeneous biomedical data remains a major challenge in healthcare reform. In modern biomedical research, complex, heterogeneous, low-descriptive, and generally unstructured electronic health records have emerged with a variety of data, including imaging, sensor data, and text. Traditional statistical learning approaches, such as data mining and machine learning, usually require first engineering to extract effective and reliable features from the data and then building predictive models or ensembles on top of that data. Both steps have many challenges in complex data scenarios and a lack of sufficient domain knowledge [18].

The solution to the problems mentioned above is deep learning, such as high performance, end-to-end learning schemes and complex feature learning, and the ability to handle complex and multimodal data. Deep learning enables computational models composed of multiple layers of processing to explore data representations with multiple levels of abstraction. This technique has greatly improved the state of the art in speech recognition, visual object recognition, object detection, and many other areas such as drug discovery and genomics. Deep learning uncovers complex structures in large data sets using a multi-propagation algorithm to show how to change the internal parameters used to calculate the representation in each layer from the representation in the previous layer [11].

A. Large Language Models (LLMs)

Large language models (LLMs) are deep learning models that are trained on extremely large datasets of text and are capable of multiple natural language processing tasks, such as translation, summarization, report-making, and grammar correction. By learning which word (or token) is most probable to appear after a sequence of preceding words in a self-supervised manner, the LLM is able to predict the next single word and is therefore described as generative.

Diverse applications of LLMs have appeared in the healthcare industry, including facilitating clinical documentation, creating discharge summaries, generating clinic, operation, and procedure notes, obtaining insurance pre-authorization, summarizing research publications, or working as a chatbot to answer questions for patients with their specific data and concerns. LLMs can also assist physicians in diagnosing conditions based on medical records, images, and laboratory results and suggesting treatment options or plans. At the same time, patients can potentially become more autonomous than with prior search methods by obtaining an individualized assessment of their data, symptoms, and concerns.

B. Digital Twins (DTs)

A virtual replica of a product or system throughout its life cycle is called a "digital twin” [1]. Digital twins provide learning, reasoning, and dynamic recalibrating for improved decision-making using real-time data and other sources. They are intricate computer models that can be modified, changed, and updated in real-time and are twins, or exact reproductions, of real-world things.

Digital twins can be used for various purposes, such as patient care, clinical studies, better personal health, diagnostics and medical training, health forecasting, better clinical research methods, identifying the best treatments, enhancement of medical innovation, best course of therapy, possibilities for effective treatment and diagnosis, therapeutic trials, digital monitoring of the human body, and many more [1].

Beside this comprehensive emulation and being equipped with AI, the DT is able to uncover information, including system descriptions, hidden patterns, and unknown correlations. The ability to record, control, and monitor the conditions and changes of the physical system allows for the application of AI predictive and prescriptive techniques for forecasting failures, testing the outcome of possible solutions, and activating self-healing mechanisms. This brings us to the approach called predictive maintenance, where collapses or failures are anticipated and changes can be simulated to avoid errors or find optimal solutions [3].

Both large language models and digital twins are powered by deep learning. Large language models use deep learning to learn the patterns and relationships in language, while digital twins use deep learning to learn the patterns and relationships in the physical world.

Some of the state-of-the-art deep learning models in the healthcare industry with their brief descriptions and limitations are listed in Table I.

Table I. State-Of-The-Art Deep Learning Models In The Healthcare Industry

Title	**Authors**	**Year**	**Brief Description**	**Limitations**
SYNDEEP: a deep learning approach for the prediction of cancer drug synergy. [9]	Torkamannia et al.	2023	This study presented a state-of-the-art approach for predicting synergy in drug combinations. Various physicochemical, genomic, protein-protein interaction and protein-metabolite Interaction information was used to predict the synergistic effects of the combinations of different drugs.	The paper only classifies the synergy for different drug combinations, but the model used in the paper does not provide any personalized recommendations for improving the synergy or an area where different combinational experiments can be made. And they have also used no technology through which one can ask questions regarding the model's prediction.
An Ensembled Framework for Human	Mustafa et al.	2023	This study presented a state-of-the-art approach	The paper only classifies the survivability, but the model

Breast Cancer Survivability Prediction Using Deep Learning [28]			for cancer survivability prediction. It classifies survivability into two categories: long-term survivability and short-term survivability.	used in the paper does not provide any reason or personalized recommendations for the given output, and there is also no virtual space where experimentation with different conditions can be made or for making the survivability longer. And they have also used no technology through which one can ask questions regarding the model's prediction.
Deep learning-based dose prediction in radiotherapy planning for head and neck cancer [27]	Teng et al.	2023	This study presented a state-of-the-art approach for deep learning-based automatic dose distribution prediction in radiotherapy planning for head and neck cancer.	The paper only revolves around dose distribution, But the model used in the paper does not provide any kind of personalized suggestions for making a change in the angle or direction of the beam used for radiating radiation. And they have also used no technology through which one can ask questions regarding the model's prediction.
Robust deep learning model for prognostic stratification of pancreatic ductal adenocarcinoma patients [12]	Ju et al.	2021	This study presented a state-of-the-art deep learning model for prognosis-correlated subtyping to identify subtypes (from two subtypes) of pancreatic ductal adenocarcinoma (PDAC), a type of pancreatic cancer, and predict patient prognosis (the likelihood of a disease developing and the chances of recovery).	The paper only revolves around the identification of two distinct subtypes with different survival outcomes, corresponding to DNA damage repair and immune response. But the model used in the paper does not provide any kind of personalized suggestions for curing the cancer or slowing down its growth based on the person’s immune system and other factors. And they have also used no technology through which one can ask questions regarding the model's prediction.

Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography [38]	Alves et al.	2021	This study presented a state-of-the-art, fully automatic deep learning framework for pancreatic ductal adenocarcinoma (PDAC) detection on contrast-enhanced computed tomography (CE-CT) scans.	The paper only revolves around the detection of pancreatic cancer, but the model used in the paper does not provide any kind of personalized suggestions regarding any type of test by which the cancer can be confirmed. It also does not recommend any of the medicines based on the specific human immune system and based on the experimentation of different medicines for their curation. And they have also used no technology through which one can ask questions regarding the model's prediction.

The solution to all the above-mentioned limitations is the proposed state-of-the-art technology, which is a combination of the digital twin (DT) and large language model (LLM). Large language models and digital twins are two emerging technologies with the potential to revolutionize many industries. A brief introduction to the large language model and digital twin is given below.

C. Proposed State-of-the-art Technology

In this paper, we introduce a state-of-the-art technology that is the combination of a large language model and a digital twin. The motivation for this state-of-the-art technology is a technological gap in precision healthcare where there is no technology that can communicate with the digital twin or the virtual part of the living beings and know what effect of any therapy, medicine, vaccine, or any other thing is occurring or can occur on them. By using this new technology, a huge boost can be given to the medical and pharmaceutical research that currently occurs by experimenting on living beings or by performing different tests.

The proposed state-of-the-art technology is a combination of state-of-the-art works [6], [40], [41]. In ref. [6] they created plausible virtual proxies of human behavior that can power interactive applications with the help of the LLM. In ref. [40], they connected the LLM-like architecture to the chest X-ray and generated the chest radiology report automatically. In ref. [41], they created a view-based, natural language processing-based video captioning model that describes routine second-trimester fetal ultrasound scan videos, in which the model created captions similar to those of the words spoken by sonography experts.

With the help of this new technology, medical and pharmaceutical research can be made more efficient, reliable, cost-effective, safe, and fast as compared to current ones. Further, this technology can also be extended for the treatment of critical diseases in patients. A demonstration of our proposed technology in precision healthcare is shown in Fig. 1.

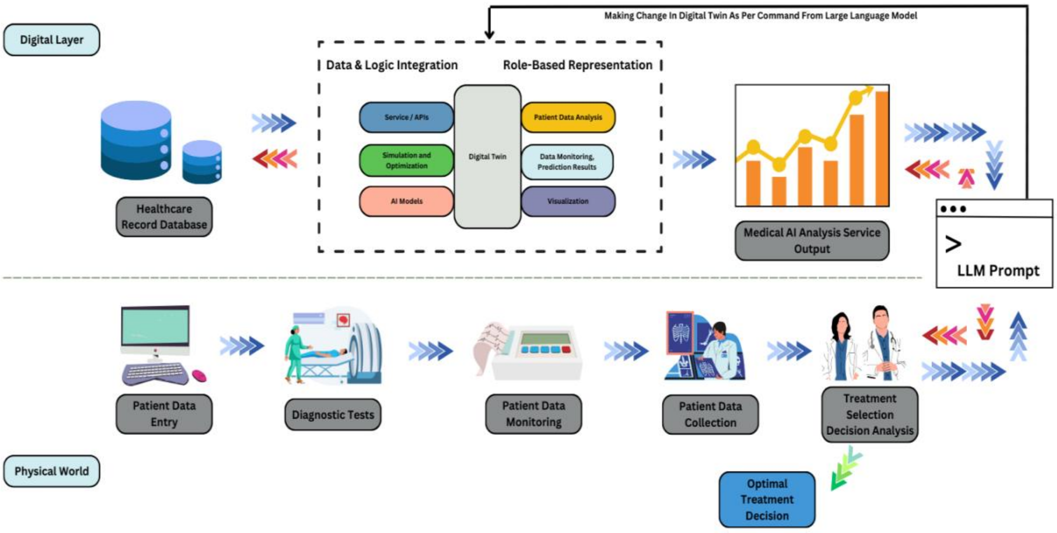
Making Change In Digital Twin As Per Command From Large Language Model
Digital Layer
Data & Logic Integration
Role-Based Representation
Service / APIs
Simulation and Optimization
AI Models
Digital Twin
Patient Data Analysis
Data Monitoring, Prediction Results
Visualization
Healthcare Record Database
Medical AI Analysis Service Output
LLM Prompt
Patient Data Entry
Diagnostic Tests
Patient Data Monitoring
Patient Data Collection
Treatment Selection Decision Analysis
Optimal Treatment Decision
Physical World

Figure 1. Workflow of proposed technology in the precision healthcare

The paper is organized such that there are 8 sections, in which Section II covers an overview, Section III covers the challenges in the healthcare industry, Section IV covers the methodology used for finding research papers, Section V covers proposed state-of-the-art technology, Section VI covers applications of large language models, digital twins, and proposed technology, Section VII covers future scopes, and the paper concludes by discussing the conclusion of the paper.

2. Overview

A. Large Language Models (LLMs)

Ref. [30] has provided many valuable insights regarding large language models. The language models can mainly be divided into four categories:

i. *Statistical language models.* The main idea in the statistical language model is to create a word prediction model based on Markov prediction. predict the next text based on the last word. The statistical language model whose word length is fixed is termed as the n-gram language model. Some of the examples of the n-gram language models are the trigram and bigram language models. Statistical language models have been extensively used to improve the performance of problems in obtaining information and natural language processing [13], [14], [15], [16].

ii. *Neural language models.* Neural language models takes the current sequence of words as input and pass it through an artificial neural network, such as recurrent neural networks. The network then outputs a probability distribution over all possible next words. The neural language model then selects the word with the highest probability and outputs it as the next word in the sequence [17], [18], [19].

iii. *Pre-trained language models.* Pre-trained language models are language models that are pre-trained on large-scale corpora in a self-supervised fashion. These pre-trained language models have fundamentally changed natural language processing approaches. Further enhancement of pre-trained language models was done by adding a transformer architecture with self-attention and multi-head attention mechanisms [23]. BERT (Bidirectional Encoder Representations from Transformers) was also proposed by pre-training bidirectional language models [24].

iv. *Large language models.* It has been found that scaling up pretrained language models, i.e., increasing data size and model parameters, often leads to an increase in model capacity in downstream problems as per the scaling law mentioned in [25]. Although scaled mainly in model size and architecture similar to pre-trained language models, these large pre-trained language models show different behavior and show surprising capabilities in solving different complex problems [26] [26].

Some of the representative emergent abilities of the LLMs are described below:

- *In-context learning: It* is a method of prompt engineering that allows large language models to learn new tasks without further training. In-context learning demonstrations of the task are provided to the model as part of the prompt in natural language. The model can then learn from a few examples in the context.
- *Instruction fine-tuning:* It is a specialized technique to tailor large language models to perform specific tasks based on explicit instructions. It is a form of supervised learning where the model is trained on a dataset of input instructions and corresponding desired outputs. The instructions can be in any form, such as natural language, code, or mathematical equations.

Prompt Engineering

Prompt engineering is the most crucial aspect of utilizing LLMs effectively and is a powerful tool for customizing the interactions, like with ChatGPT. It involves crafting clear and specific instructions or queries to elicit the desired responses from the language model.

There are various types of methods included in prompting engineering. Some of them are mentioned below:

i. *Zero-Shot Prompting*: The zero-shot strategy involves the LLM generating an answer without any examples or context. This strategy can be useful when the user wants a quick answer without providing additional detail or when the topic is so general that examples would artificially limit the response.

ii. *One-Shot Prompting*: The one-shot strategy involves the LLM generating an answer based on a single example or piece of context provided by the user. This strategy can guide ChatGPT-like response and ensure it aligns with the users' intent. The idea here would be that one example would provide more guidance to the model than none.

iii. *Few-Shot Prompting*: The few-shot strategy involves the LLM generating an answer based on a few examples or pieces of context provided by the user. This strategy can guide ChatGPT-like response and ensure it aligns with the users' intent.

Ref. [48] Flan-PaLM, a large language model that achieved state-of-the-art performance on the different medical LLM benchmarks, was developed by combining chain of thought, few-shot, and other prompting technologies.

Mainstream Architectures of LLMs

Due to its excellent parallelizability and capacity, the Transformer architecture has become the de facto backbone for developing various LLMs [22]. In general, the mainstream architectures of existing LLMs can be roughly categorized into three major types, namely encoder-decoder, causal decoder, and prefix decoder.

i. *Encoder-decoder architecture*: The transformer's pure form consists of two blocks, i.e., the encoder and decoder [22]. The encoder block takes the input text and produces a sequence of the hidden states, while the decoder block takes the encoded hidden states and produces the output text. Both the encoder and decoder are made up of a series of self-attention layers. Some of the examples of the encoder-decoder architecture that have demonstrated their effectiveness in some of the natural language processing problems are T5 and BERT [73], [24].

ii. *Causal Decoder Architecture:* Causal decoders are a type of decoder architecture that ensures that each input token will only attend to itself and past tokens. This is done using a unidirectional attention mask, which sets the attention weights for future tokens to zero. This is in contrast to the non-casual decoder architecture, which allows each input token to attend to all tokens in the sequence.

iii. *Prefix Decoder Architecture:* The prefix decoder architecture, also known as the non-causal decoder, It is a type of decoder that allows bidirectional attention on prefix tokens and unidirectional attention on generated tokens [29]. This enables the decoder to learn long-range dependencies in the input sequence and to generate more accurate and fluent outputs. The prefix decoder architecture is typically used in transformer-based language models, such as BART and GPT-3.

Scaling of LLMs

The scaling of large language models is a topic of active research in the field of artificial intelligence. There are a number of factors that contribute to the scaling of the LLMs, including the number of parameters, the size of the training dataset, the compute budget, and the network architecture. Some of them are discussed in depth below.

i. *Size of the training dataset:* The size of the training dataset is also an important factor in the scaling of LLMs. Larger datasets provide more training data for the model to learn from, which can lead to improved performance.

ii. *Number of parameters*: The number of parameters in an LLM is a measure of its complexity. Large models have more parameters, which allows them to learn more about complex patterns in the data.

iii. *Compute budget:* The compute budget is the amount of computational resources available to train an LLM. Larger compute budgets allow for the training of larger models and larger datasets, which can lead to improved performance.

The most important one is the *scaling law*. A number of scaling laws have been proposed to describe the relationship between the performance of an LLM and the number of parameters, the size of the training dataset, and the compute budget [52], [38]. These scaling laws can be used to predict the performance of an LLM, given a particular set of resources.

Catastrophic Forgetting

While scaling the large language model, the problem of catastrophic forgetting is common. Catastrophic forgetting is a phenomenon that occurs when a machine learning model is trained on a new task and subsequently loses its ability to perform well on previously learned tasks. This is a problem for large language models (LLMs) because they are typically trained on massive datasets of text and code. As a result, LLMs can be very good at performing a wide range of tasks, but they are also susceptible to catastrophic forgetting [50].

There are a number of reasons why catastrophic forgetting occurs in LLMs. One reason is that LLMs are typically trained using a technique called backpropagation. Backpropagation is a method for updating the parameters of a neural network based on the errors it makes on a training dataset. However, backpropagation can also cause the model to forget previously learned information. This is because the updates to the parameters are made in a way that minimizes the error on the new task, even if it means sacrificing performance on the old tasks.

Another reason why catastrophic forgetting occurs in LLMs is that they are typically trained on very large datasets. This means that the model has a lot of parameters, and it is difficult to update all of these parameters without affecting the performance of the old tasks.

There are a number of techniques that can be used to mitigate catastrophic forgetting in LLMs. One technique is to use a technique called *elastic weight consolidation (EWC)* [51]. EWC works by penalizing updates to the parameters of the model that are important for the old tasks. This helps prevent the model from forgetting the old information.

Another technique that can be used to mitigate catastrophic forgetting is to use a technique called *progressive neural networks (PNNs).* One example of this technique is *low-rank adaptation (LoRa),* which works by gradually adding new layers to the model as it is trained on new tasks [49]. This helps to prevent the model from forgetting the old information because the new layers are not trained on the old tasks.

There are also a number of other techniques that can be used to mitigate catastrophic forgetting in LLMs. These techniques are still under development, but they show promise in reducing the problem of catastrophic forgetting.

Model Optimization For Development

There are mainly three techniques for model optimization:

i. *Distillation:* It is a technique that focuses on having a larger teacher model to train a smaller student model. The student model learns to statistically mimic the behavior of the teacher model, either in the final prediction layer or in the model's hidden layer as well [42], [43].

ii. *Quantization:* It is a technique in which the precision of the model weight is reduced. For example, if the model has data in the form of a 32-bit floating point, that data is reduced to a 16-bit floating point. This technique is applied to the model weight and the activations [44], [45].

iii. *Pruning:* It is a technique involved in the removal of redundant or unnecessary data from the model's training dataset. This can help reduce the size of the model and improve its performance [46], [47].

Some of the state-of-the-art large language models in the healthcare industry with their brief descriptions and limitations are mentioned in the Table II.

Table II. State-Of-The-Art Large Language Models In The Healthcare Industry

Title	**Authors**	**Year**	**Brief Description**	**Limitations**
BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining [34].	Luo et al.	2022	This study presented a state-of-the-art approach for a domain-specific generative transformer language model pre-trained on large-scale biomedical literature.	In this paper, researchers included an ambiguous medical dataset based on information from the Wikipedia web, which may introduce biases or limitations in terms of the quality and accuracy of the data.
X-ray GPT: Chest Radiographs Summarization using Medical Vision-Language Models [66].	Thawaket al.	2023	This study presented a state-of-the-art approach for a conversational medical vision-language model that can analyze and answer open-ended questions about chest radiographs. It aligns a medical visual encoder with a fine-tuned large language model to enable exceptional visual conversation abilities grounded in a deep understanding of radiographs and medical domain knowledge.	In this paper, though XrayGPT aligns a medical visual encoder with a fine-tuned language model to enhance its visual conversation abilities, the specific limitations or challenges faced by this model are not explicitly mentioned in the provided sources.

MSQ-BioBERT: Ambiguity Resolution to Enhance BioBERT Medical Question-Answering [36].	Guo et al.	2023	This study represented a state-of-the-art approach called Multiple Synonymous Questions BioBERT integrates question augmentation to enhance the performance of BioBERT on medical question-answering tasks.	In this paper, performance was evaluated on web-based constructed medical datasets and open biomedical datasets, which may not fully represent the range of real-world medical question-answering scenarios.

Medical LLM Benchmarks

To evaluate how well LLMs encode clinical knowledge and assess their potential in medicine, researchers consider medical question answering. This task is challenging providing high-quality answers to medical questions requires comprehension of the medical context, recall of appropriate medical knowledge, and reasoning with expert information. Existing medical question-answering benchmarks are often limited to assessing classification accuracy or automated natural language generation metrics (e.g., BLEU [67]) and do not enable the detailed analysis required for real-world clinical applications [33]. This creates an unmet need for a broad medical question-answering benchmark to assess LLM's response factuality, use of expert knowledge in medical and scientific reasoning, helpfulness, precision, health equity, and potential harm to humans accepting model outputs as facts.

To address this, the authors of ref. [10] introduced state-of-the-art MultiMedQA, a benchmark comprising seven medical question-answering datasets, including six existing datasets: MedQA [53], MedMCQA [54], PubMedQA [55], LiveQA [56], MedicationQA [57], and MMLU clinical topics [58]. And they named it HealthSearchQA, which consists of commonly searched health questions.

B. Digital Twins (DTs)

AI in Digital Twins

The integration of AI models (of physical objects) and big data analytics for processing IoT data motivates one of the latest and probably one of the most important advancements in the field of technology, which is the digital twin [15], [16], [17]. DT models are gaining more and more interest for their potential and strong impact in application fields such as manufacturing, aerospace, healthcare, and medicine.

The original use of the digital twin was promoted by tuegel et al. in [68], ref. [54], [53] proposed a conceptual model of how DT can be used as a virtual sensor to predict the lifetime of an aircraft structure and ensure its structural integrity. All the aforementioned research led to the definition of the DT airframe, a computational model of individual aircraft. This model had the potential to improve the way US Air Force aircraft are managed throughout their life cycle by creating individualized structural management plans. DTs can provide configuration checks for each aircraft in its inventory through computational simulation. They can act as a virtual health monitor and predict future maintenance needs for each aircraft [3].

It is important to understand that while a digital twin is an intelligent system, it is not necessarily completely autonomous [3]. Indeed, AI-based applications and digital twins continue to be widely used by many people, specifically in scenarios that require intervention to test new features, change physical properties, or provide answers such as diagnosis or treatment.

DT technology includes continuous advances in artificial intelligence. It refers to unsupervised and supervised learning algorithms whose predictive competencies are refined by processing continuous sensory data obtained from physical twins and their encircled environment. This virtual mind uses predictive, descriptive, and prescriptive algorithms to carry out a sequence of tasks as a brilliant product [3].

There are mainly three types of digital twins [5]:

i. *Product twinning:* It provides a virtual physical connection to analyze how a product performs under various conditions and make adjustments in the virtual world to ensure that the physical product will perform exactly as planned in the field.
ii. *Process twinning:* It is used to improve processes and workflows by allowing managers to tweak inputs and see how outputs are affected without the risk of upending existing workflows.
iii. *System or performance digital twins:* It capture, analyze, and act on operational data, providing insights for informed decisions to maintain effective interactions among the components of the system at the system level.

There are three forms of transmission channels that should be anticipated for digital twins [3]:
i. Among physical and virtual twins
ii. Among ambient DT and isolated DT
iii. Among the DT and the domain specialists who engage with and control the DT

All data exchanged needs to be stored in a data storage machine so that it is easily accessible by the digital twin. In addition to dynamic data, the data storage machine also contains historical data that reflects the physical twin's memory and records historical data provided by human specialists and past actions, in addition to descriptive static data that describes essential characteristics of the physical twin that should not be modified with time [3].

Self-adaptation and self-parametrization

DTs have self-adaptation and self-parameterization abilities and can resemble physical twins during their life cycle [32], [33]. This will be carried out effortlessly by development of highly parameterized and modular DT. Modularity ensures that changes in a single module do not affect other modules. Parameterization ensures easy changes in DT.

DTs need to be able to cope with high-dimensional data and therefore require powerful techniques for decoding and analyzing high-dimensional data, in addition to techniques used for integrating multiple data sources to produce more accurate, consistent, and useful information. Therefore, it must have data fusion algorithms that are better than those provided by individual data sources.

DTs makes use of predictive analytics [55], [56] to anticipate future states and essential modifications like failures in the product life cycle. DTs uses the output of predictive and descriptive techniques as input to prescriptive analytics [57] to make choices applicable to its own destiny via computationally determining alternative actions or choices given a complex set of goals, necessities, and constraints. Ultimately, it makes use of optimization algorithms to acquire the best end result at the same time as handling the uncertainty inside the records [86].

In addition to use of prescriptive and predictive algorithms, DT encodes the calculated prescriptions and optimization scheme by using proper methodologies to encode high-dimensional features. This allows remarks to be sent both to the physical twin and to other DTs throughout the surrounding area. Alternatively, intended users can use the interactive interface to leverage the calculated information and check the status of the DT.

The life cycle of digital twins

Ref. [3] describes two possible DT life cycles, from their design to their disposal.

- The first life cycle refers to an entity that does not yet exist, and in this scenario, the design workflow concurrently creates each of the physical twin and its digital twin.
- The second lifecycle refers to an entity that already exists but has no DT in place, and in this scenario, the design workflow focuses on extending the entity to be attached.

Both life cycles share a common timeline. That is, first the design stage, then the development stage, the exploitation stage, and finally the disposal stage.

In this *first case*, the DT begins to exist earlier than the physical entity as a prototype and is utilized by the designer in the design stage of the prototype entity [3]. At the beginning of the design stage, the prototype is utilized as if it were an actual entity, simulating, testing, modifying, and finally validating the design choices until a satisfactory results is determined. At some stage in this design cycle, designers use the following things:

i. *Historical data*: The data that prototype obtains from different existing DTs associated with similar entities
ii. *Static data*: The data describing the past state of a DT; information about other connected DTs
iii. The outcomes of simulations performed by the prototype, the prediction results calculated by the prototype, and its recommendations and optimization schema

When an entity ceases to be used because of obsolescence or some other cause, it has to be dismantled, and the dismantling stage begins. The stored historical data of the product DT is backed up and made available to other DTs and domain experts; in this way, designers or any other area expert may be able to use the collected facts to optimize the production of future devices.

The *second lifecycle* is different in that the entity is already implemented and in use but has no DT attached. In this scenario, the design stage involves the development of new prototypes that are tested, refined, and finally validated. The development stage considers the development of connections between physical entities and the DT prototype, and the operational stage considers the operational life of the prototype. The digital twins continue until they are dismantled in the dismantling stage.

Some of the state-of-the-art digital twins in the healthcare industry with their brief descriptions and limitations are listed in Table III.

Table III. State-Of-The-Art Digital Twins In The Healthcare Industry

Title	**Authors**	**Year**	**Brief Description**	**Limitations**
Toward an artificial intelligence-assisted framework for reconstructing the digital twin of vertebra and predicting its fracture response [79].	Ahmadiet al.	2022	The paper presents a state-of-the-art AI-assisted framework called ReconGAN for creating a realistic digital twin of the human vertebra and predicting the risk of vertebral fracture (VF). The paper demonstrates the applicability of digital twins generated using this AI-assisted framework to predict the risk of VF in a cancer patient with spinal metastasis through a feasibility study.	The paper does not discuss the limitations of the ReconGAN framework or the accuracy of the predictions made by the digital twin model. And they have also used no technology through which one can ask questions regarding the model's prediction.
Development and Verification of a Digital Twin Patient Model to Predict Specific Treatment Response During the First 24 Hours	Lal et al.	2020	The paper presents state-of-the-art, verified digital twin models of critically ill patients using an artificial intelligence approach to predict the response to specific treatment during the first 24 hours of sepsis. Directed acyclic graphs were used to define the	The study focused on the first 24 hours of sepsis and did not assess the long-term response to treatment or the impact on patient outcomes beyond this timeframe

of Sepsis [59].			causal relationship among organ systems and specific treatments.	And they have also used no technology through which one can ask questions regarding the model's prediction.
The Living Heart Project: A robust and integrative simulator for human heart function [64].	Baillarget al.	2014	The paper presents a state-of-the-art proof-of-concept simulator for a four-chamber human heart model using computer topography and magnetic resonance images. In which they performed visualization of the electrical potential and mechanical deformation across the human heart throughout its cardiac cycle.	The paper does not explicitly mention the limitations of the study or any potential drawbacks in the methodology used. And they have also used no technology through which one can ask questions regarding the model's prediction.
An In Silico Subject-Variability Study of Upper Airway Morphological Influence on the Airflow Regime in a Tracheobronchial Tree [65].	Feng et al.	2017	The paper employed a state-of-the-art computational fluid-particle dynamics (CFPD) model to simulate airflow patterns in three different human lung-airway configurations. The paper focused on identifying morphological parameters that significantly influence the airflow field and nanoparticle transport in the respiratory system.	The paper does not explicitly mention the limitations of the study or any potential drawbacks in the methodology used. And they have also used no technology through which one can ask questions regarding the model's prediction.

3. Challenges In Healthcare

There are many challenges in the field of healthcare related to large language models and digital twins; some of them are listed below.

Huge Gap Between Generated Content vs. Real Content: Generated content refers to information that is produced by an artificial intelligence model without being directly sourced from real-world data or human input. It is generated based on patterns and knowledge learned from training on vast amounts of existing data, while the real content, on the other hand, refers to information that is derived from authentic, verified, and reliable sources, such as medical literature, clinical guidelines, or direct input from healthcare professionals.

Low Accuracy of Answers: One of the biggest challenges with AI algorithms in healthcare is their accuracy and reliability. They are trained on massive datasets, but these datasets may contain errors or biases. This could lead to the generation of inaccurate or misleading information, which could have serious consequences for patients.

Base Less Outputs: Another challenge is the interpretability of AI algorithms. It can be difficult to understand why a model generates a particular output, which can make it difficult to trust the information that the model provides. This is especially important in healthcare, where decisions about patient care need to be made based on sound evidence.

Diverse Data for Algorithm Training: To ensure the algorithms used in healthcare applications are representative and unbiased, it is necessary to train them on diverse healthcare data. This includes data from different demographics, geographic regions, and medical conditions. By including a wide range of data sources, healthcare organizations can reduce the risk of biased or skewed outcomes that may disproportionately impact specific patient populations.

Proper Usage of AI Algorithms: AI algorithms, such as those used in medical decision-making, should follow proper guidelines and best practices. This includes rigorous testing, validation, and peer review before deployment in a clinical setting. Collaborating with healthcare professionals, domain experts, and regulatory bodies can help establish guidelines for the appropriate use of AI algorithms. Implementing robust governance frameworks and regular auditing processes can also help identify and rectify any incorrect or harmful medical decisions made by these algorithms.

Lack of Standardization: Artificial intelligence algorithms rely on consistent data structures and standardized protocols for effective communication and interoperability. However, the lack of universal standards in healthcare can pose challenges to integrating these technologies across different healthcare systems and organizations. This can result in data inconsistencies and compatibility issues and hinder seamless collaboration.

Dynamic and Evolving Patient Data: Patient health data is constantly evolving, with new data points being generated regularly. Incorporating real-time, dynamic data into AI algorithms can be complex, as it requires continuous data updating and integration. Ensuring that the AI algorithm accurately reflects the most recent patient information in real time is a challenge.

Model Calibration and Validation: AI algorithms rely on mathematical models that simulate patient physiology and behavior. Calibrating and validating these models to accurately represent individual patients or patient populations can be challenging. Ensuring the accuracy and reliability of these models across a diverse range of patient characteristics and medical conditions is crucial for effective decision-making.

Limited Accessibility and Inclusivity: AI algorithms often require advanced technological infrastructure and availability to high-quality internet connections, which may not be available in all healthcare settings or regions. Limited accessibility to AI algorithms can create disparities in healthcare delivery and access to personalized care, particularly in resource-constrained areas.

Psychological and Ethical Considerations: Interacting with a digital representation of oneself can have psychological implications for patients. It is important to consider the emotional and ethical aspects of using AI algorithms in healthcare, such as patient autonomy, informed consent, and the potential impact on the doctor-patient relationship.

Integration with Clinical Workflows: Incorporating AI algorithms seamlessly into existing clinical workflows and decision-making processes can be challenging. Healthcare professionals may face difficulties integrating the insights provided by AI algorithms into their practice, potentially leading to resistance or limited adoption of this technology.

Regulatory and Legal Frameworks: As AI algorithms evolve, existing regulatory and legal frameworks may struggle to keep pace. Ensuring that regulatory bodies and legal frameworks adapt to address the unique challenges and considerations posed by AI algorithms in healthcare is necessary to ensure compliance and patient safety.

4. Methodology Used For Finding Research Papers

At the start of the work, the researchers identified three unique research questions that guided the entire review:

RQ1: Which are the state-of-the-art LLMs in the field of precision healthcare, and what are their applications?
RQ2: Which are the state-of-the-art digital twins in the field of precision healthcare, and what are their applications?
RQ3: Is there any literature that relates to the combinational use of LLMs and digital twins in the field of precision healthcare?

The researchers defined three different word strings to be searched in Google Scholar: “precision healthcare and large language model”, precision healthcare and digital twin, and “digital twin and large language model in precision healthcare” to address the above three research questions. The researchers selected Google Scholar to keep away from bias in prefer of any particular scholarly publisher, as recommended through [31]. The search was performed in *August 2023*, and researchers have not specified the time ranges for the search.

The researchers excluded replications or extensions of previously published work from their results. The researchers then performed a snowball calculation on the collection of articles they found. The researchers used the reference lists of all publications to identify new publications for potential inclusion in the review. Once more, researchers applied the same inclusion and exclusion criteria as previously mentioned. When no new publications were found, the researchers stopped their work. The researchers produced an initial definitive set of 13 publications in which 12 belonged to RQ1 and 1 to RQ2, they are listed in Table IV. And to further expand the study, the researchers then included all of the above research questions for healthcare as well, And statistics related to that are shown in the Fig. 2.

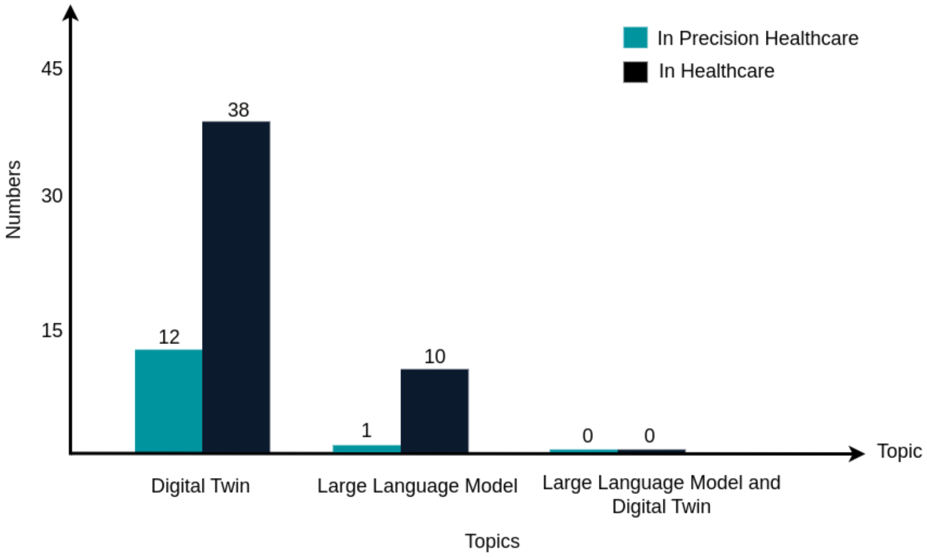

Figure 2. Workflow of proposed technology in precision healthcare

Table IV. List Of Publications Related To Precision Healthcare

Title	**Authors**	**Year**	**Related RQ**	**Brief Description of paper**
Digital twins for precision healthcare [87].	Asalemisalemi et al.	2020	RQ1	The paper discusses cyber-security-related challenges and ethical implications related to DT in precision healthcare.
Digital twins will revolutionize healthcare. Digital twin technology has the potential to transform healthcare in a variety of ways, improving the diagnosis and treatment of patients, streamlining preventative care and facilitating new approaches for hospital planning [88].	James et al.	2021	RQ1	The paper discusses the potential of digital twin technology to transform healthcare by improving diagnosis and treatment.

The ‘Digital Twin’ to enable the vision of precision cardiology [89].	Corral-Acet al.	2020	RQ1	The paper discusses the challenges and opportunities ahead in developing the digital twin in cardiovascular medicine.
Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation [90].	Venkatesh et al.	2022	RQ1	The paper discusses the main considerations for implementing digital twin research into clinical practice, including the computational requirements.
The health digital twin: advancing precision cardiovascular medicine [91].	Coorey et al.	2021	RQ1	The paper discusses both the challenges and opportunities of precision cardiovascular medicine.
Digital twins to enable better precision and personalized dementia care [92].	Wickramasinghe et al.	2022	RQ1	The paper discusses the potential benefits of incorporating digital twins in dementia care, ensuring greater precision and personalization.
Digital twins for predictive oncology will be a paradigm shift for precision cancer care [93].	Hernandez-Boussard et al.	2021	RQ1	The paper discusses the potential application of digital twins in modeling cancer patients’ outcomes.
The “virtual digital twins” concept in precision nutrition [94].	Gkouskou et al.	2020	RQ1	The paper discusses the potential of incorporating genetic information, longitudinal metabolomics, immune parameters, and bioclinical variables to create a "virtual digital twin" model.
Deep digital phenotyping and digital twins for precision health: time to dig deeper [95].	Fagherazzi et al.	2020	RQ1	This paper discusses the idea of combining real-world digital data with clinical data and omics features to identify someones’ digital twin.
The future of digital twins in precision dentistry [96].	Saghiri et al.	2023	RQ1	This paper discusses the usage of digital twins in dentistry.
A Proposed Framework for Digital Twins-Driven Precision Medicine Platform: Values and Challenges [85].	Elshaier et al.	2022	RQ1	The paper gives an idea of the usage of digital twins in healthcare.
Design of Precision Medicine Web-service Platform Towards Health Care Digital Twin [71].	Kolekar et al.	2023	RQ1	The paper proposes a state-of-the-art digital twin-based integrated precision medicine web-services platform that offers multidisciplinary precision medicine services and can be easily implemented in hospital organization interfaces.
Precision Health in the Age of Large Language Models [70].	Poon. et al.	2023	RQ2	This paper discusses the concept of utilizing digital twins in dentistry, which is still in its early stages.

5. Combination Of Large Language Model And Digital Twin (Proposed Technology)

During our search on Google Scholar to find the answer to RQ3, researchers found there exists no publication or work related to the combinational use of LLM and DT in precision healthcare and healthcare. Researchers also tried other search engines like PubMed and ScienceDirect, but they found no work or publication related to this.

So, researchers decided to propose state-of-the-art technology in the fields of precision healthcare and artificial intelligence by connecting a large language model with the digital twin, which is related to precision healthcare. With the help of the digital twin model, one can ask the questions, and that model can answer those questions. This type of model can bring a revolution to the entire healthcare and pharmaceutical industries.

An example application of proposed technology can be like connecting the digital twin of heart to a large language model like chatGPT, And by which one can make different questions to the heart model like what is the condition of the any part of the heart, Or is there any problem to the heart which is transferring the blood from one part of heart to another, Or how are the heart’s valves are performing and one can also ask the heart to produce an report on condition of heart like is there any anomalous activity occurring in heart like tightening or leaking in the heart’s valve(Structure that allow the blood to flow from one chamber to another chamber or blood vessel), Or Is there any problem of blockages in any part of the heart?

In a manner similar to the above application, proposed new state-of-the-art technology can be implemented in any part of the human body or any living thing.

And the researchers took this idea from state-of-the-art work [6], in which they created plausible virtual proxies with human behavior that can power interactive applications from immersive environments to human communication testbeds to prototyping tools. They introduced generative agents, which were similar to the digital twins in our case, which are computational software agents that stimulate believable human behavior. In our case, the behavior of different body parts or other things related to healthcare

In ref. [6], generative agents wake up, eat, and depart for work. Artists draw, authors write, and students go to college. They notice each other and also initiate conversations; they reminisce and reflect on the past as they plan for the next day. To operate or give command to the generative agents, they describe an architecture which extends a large language model to maintain a detailed record of the agent's experiences via natural language, to represent these memories at a higher level over time, and to manage them dynamically. They created an interactive sandbox environment where end users or the last user can communicate with a small town of agents with the help of natural language.

Another state-of-the-art work from which researchers got the idea of connecting the LLM to a digital twin in medical-related from [40], in which they connected the LLM-like architecture to the chest X-ray and generated the chest radiology report automatically. And in [41], they created a view-based, natural language processing-based video captioning model that describes routine second-trimester fetal ultrasound scan videos, in which the model created captions similar to those of the words spoken by sonography experts.

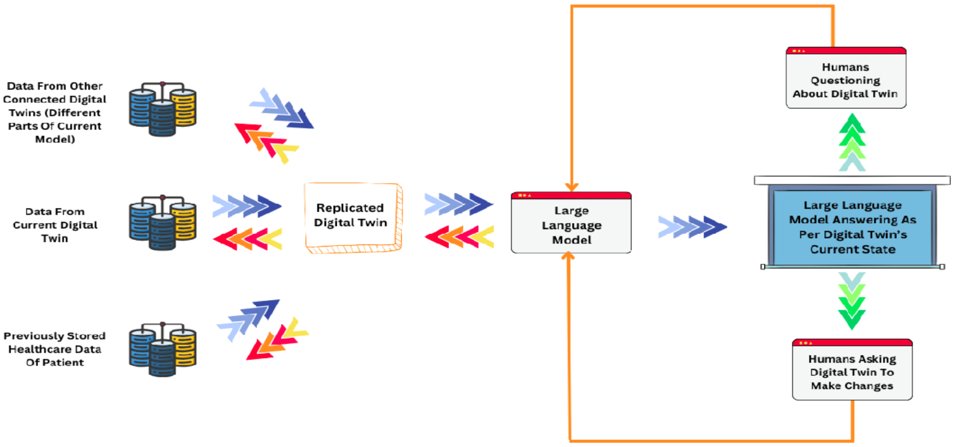
Data From Other Connected Digital Twins (Different Parts Of Current Model)
Data From Current Digital Twin
Previously Stored Healthcare Data Of Patient
Replicated Digital Twin
Large Language Model
Humans Questioning About Digital Twin
Large Language Model Answering As Per Digital Twin's Current State
Humans Asking Digital Twin To Make Changes

Figure 3. Pictoritical representation of the proposed technology

The proposed structure of the combination technology of a large language model and a digital twin for precision healthcare can be easily understood from the Fig. 3.

By leveraging the proposed technology, one can also make any organ of the body an environment, and different parts of that organ can be like generative agents. One can also simulate the whole process with the help of the digital twin, where each part of that organ can communicate with another part, and an external user can also interact with the environment with the help of the language model.

Further work on this idea is required to be done by the research community. This technology can transform the lives of millions of people, and the proposed mechanism can help in the prediction of any disease occurring in the body prior, to check the effect of any medicine or therapy on humans, and there are many other endless ideas where it can be helpful.

6. APPLICATIONS

A. Applications Of LLMS

There are many applications of LLMs in various fields, like chatbots and virtual generation, in research, education, entertainment, content generation, healthcare, and many more. Some of the applications related to precision healthcare are listed below:

- LLMs can help with the creation of medical reports based on keywords or authorized letters. Documentation takes a quarter of a doctor's time and a fifth of a nurse's time [2], [4]. The use of LLM has been shown to reduce the time spent by clinicians or other health professionals in producing documentation materials [7], [8]. This can free up doctors and nurses to care for patients, which can help improve care efficiency and reduce care costs.
- Recent studies have shown that LLM can think in the medical field and can answer medical questions with greater accuracy [20], [21].
- LLMs such as BioGPT [34], which focus on the generation of biomedical texts and post-training mining of PubMed articles, may have significant implications for the future of medicine and medical research. Other examples of similar models are SciBERT [35], BioBERT [36], PubMedBERT [37], and ScholarBERT [39].
- LLMs are used to summarize complex academic publications, allowing authors to understand complex concepts in manuscripts or automatically generate abstracts. In addition, they can help you convert your article into a format suitable for various journal publications. There are references to academic writings to improve the clarity and coherence of this text and give suggestions.
- LLMs have the ability to collect large amounts of clinical data, for example, in order to convert clinical notes into short summary statements for patient training.
- In another study, they generated the impact part of the chest radiology report by connecting a pretrained BERT-based language model [40].
- In another study, a gaze-assisted NLP-based video captioning model was proposed to describe routine second-trimester fetal ultrasound scan videos in the vocabulary of sonography experts. [41].

B. Applications Of Digital Twins

There are many applications of the digital twins in manufacturing, aviation, and healthcare. Some of the applications related to precision healthcare are listed below:

- The emergence of approaches to the treatment and prevention of diseases in precision healthcare includes the use of new diagnostic and therapeutic approaches that identify the patient's needs based on their genetics, biomarkers, phenotypic, physical, or psychosocial characteristics [61].
- DTs are used to model the patient's spine and anticipate fracture risk using ReconGAN, which is a generative adversarial network trained on quantitative microcomputed tomography images of the spine [79]. DT is used to anticipate the risk of spine fracture based on patient-specific imaging data.

- In addition to physical situations, DTs are also being studied as a tool to assist human beings in managing intellectual fitness issues such as despair and stress [84].
- Ref. [63] implemented another great idea that is the AnyBody Modeling System, which allows the human body to move in accordance with its environment. The AnyBody model allows users to run advanced simulations to calculate: 1) muscle forces 2) joint contact forces and moments. 3) Metabolism. 4) Elastic energy of veins 5) opposite muscle movement.
- One of the human DT applications in precision healthcare is the Virtual Physiological Human (VPH) [69]. A complete computer model designed to "study the human body as a single and complex system together". By customizing VPH for each patient, researchers and doctors create a platform to test any treatment protocol. VPHs can be a "virtual human laboratory" and help, for example, in clinical trials and in silico experiments [62].
- Another use case for DTs was developed by French startup Sim&Cure for patient-based virtualization of aneurysms and surrounding blood vessels (https://sim-and-cure.com/). An aneurysm is a bulge in a blood vessel caused by weakening of the artery wall. They are found in 2% of the population. These tiny but scary parts of an aneurysm can cause blood clots, strokes, and death.
- Another important use case of DTs is the anticipation of treatment response for patients with sepsis in the first 24 hours after identification. This is achieved by building a DT that uses graphical structures that represent the causal relationship between the organ system and the medication or therapy used. Methods such as agent-based modeling, Bayesian networks, and event simulation were used to anticipate how specific medications or therapies would affect organ systems [59].
- DTs are also demonstrated to be a precious predictive tool in oncology, the study of most cancers. Many research have been conducted using DTs to better apprehend cancer development and its effects [60], [70], [71], [72]. DTs are also used to model cancer remedies [73]. as an example, research uses modeling and digital truth techniques to create DTs of radiotherapy systems and examine their reliability and person-friendliness [74].
- DTs have additionally been used to monitor patient behavior in therapeutic techniques, including postural correction in patients with Parkinson's sickness, a neurodegenerative ailment that impairs movement [80], [81].
- Similar studies are underway to use Deep Q networks and DTs to optimize the treatment tiers of head and neck cancer patients based on patient information, inclusive of age and tumor grade [75]. Those deep Q networks were skilled at offering choice-making tools for making planing of 3-level treatments for head and neck cancer patients.
- One other is also developed by the French software company Dassault Systèmes, That is live Heart which was made available in 2015 for research [64]. This is the first DT of an organ that covers all aspects of organ function, including circulation, mechanics, and electrical impulses. This software requires a 2D scanner input, which is converted into a faithful 3D model of the organ. With the furnace model, doctors can run hypothetical scenarios like adding a pacemaker or changing heart chambers to anticipate patient outcomes and make decisions.
- DTs can also pick out therapies for illnesses via multilayer modules, wherein more than one form of molecule is mapped onto a protein-protein interaction (PPI) network. Scientists can perceive which genetic, protein, or cellular problems are causing a patient's sickness and respond for this reason [78].
- DTs can also be used to reverse diabetes by predicting in real-time how blood sugar levels will react by inputting foods the patient wants to eat [82].
- In ref. [65] authors simulated several different eventualities of aerosol particle movement with distinctive parameters such as particle diameter, inhalation flow rate, and initial position of the drug inside the aerosol. Those simulations show that designing targeted and patient-specific drug delivery methods that limit the particle size and surface area of the active drug within the aerosol instead of dispersing it uniformly throughout the spray can maximize drug deposition efficiency by 90%.

- DTs are also are used to diagnose and treat illnesses by modeling the affected person's colon and comparing the outcomes of pharmaceuticals on the patient [83].
- Every other application of DTs in cardiology is the anticipation of excessive blood pressure. Ref. [76], [77] advanced an evidence-of-concept model of the usage of patient information and data to create a mathematical version of blood movement that may be used as a DT to anticipate hypertension.
- One study created a DT of nuclear energy plant employees to monitor their health and make certain safe and effective shifts [86].

C. Applications Of Proposed Technology

There are many applications related to the proposed technology in the precision healthcare and pharmaceutical industries. Some of the applications are listed below:

- The proposed technology can be utilized in pharmaceutical research for experimenting with different medicines and vaccines and getting feedback based on the effects on the DT.
- The proposed technology can also be utilized in precision medicine to set the dosage of any medicine according to the patient’s immune system. With the help of this technology, one can know how a medicine is affecting the patient and can change the dosage of that medicine instead of using the current methodology in which these types of experiments are performed on the patient itself.
- The proposed technology can also be utilized in the healthcare industry for the curation of life-threatening diseases such as Parkinson's, cancer, Alzheimer's, etc.
- The proposed technology can also be utilized for studying the workings of the brain, nervous system, immune system, and many other important parts of the human body and other living organisms.

7. Future Scope

In this work, researchers have presented a state-of-the-art technology that is a combination of a digital twin and a large language model. Future research direction can be in designing a new novel architecture for proposed technology, such as that used in ref. [6] and implementing it for different ethical use cases for the benefit of society.

Another area of research is the development and improvement of existing protocols, regulations, and ethical guidelines for the use of the proposed technology in precision healthcare. Protecting patients’ rights and privacy is important. This includes developing standards for information security and data sharing.

Another research direction would be to integrate reinforcement learning with the proposed technology, which can learn by itself, so that the technology would also be able to deal with circumstances that did not happen earlier. While using the proposed technology with reinforcement learning, it will also not require historical data, as in reinforcement learning, the model makes the data on its own. But while developing this kind of model or technology, precautions should also be taken so that the model does not learn the wrong information that would be harmful to society.

Furthermore, the development of digital twin tools for healthcare requires the participation of stakeholders consisting of patients, healthcare providers, and governments. Finally, the healthcare providers and patients perceptions of the use and interpretation of the proposed technology are important for the development and adoption of the proposed technology in healthcare.

8. Conclusion

In this review, answer are given to the three research questions: What are the applications of LLMs in precision healthcare? What are the applications of the digital twins in precision healthcare? Is there any literature that uses the combination of LLMs and digital twins in healthcare? And tried to help researchers fill this research gap by proposing state-of-the-art technology.

The main contribution of this publication is a new state-of-the-art technology that contains the characteristics of both digital twins and large language models. This can be a revolutionary idea in the field of precision healthcare and pharmaceutical industry, and the lives of millions can be saved by the help of this proposed technology. This proposed technology can help patient-specific models aid in diagnosis and treatment planning, therapy planning, drug effect checking on individuals, and many other cases.

In this study, the researchers highlighted the potential of language models and digital twins in the field of precision healthcare. But there are nevertheless challenges that need to be addressed to completely exploit the capacity of large language models and digital twins in precusion healthcare applications. The researcher hopes that this publication will fucntion as inspiration for other researchers to innovate and create new approaches using digital twins, large language models, and proposed technology. Researchers encourage other researchers in the healthcare and AI domains to learn from this publication and use it as a springboard to further explore other possibilities in healthcare.

References

[1] Haleem, Abid, et al. "Exploring the revolution in healthcare systems through the applications of digital twin technology." Biomedical Technology 4 (2023): 28-38.

[2] Clynch, Neil, and John Kellett. "Medical documentation: part of the solution, or part of the problem? A narrative review of the literature on the time spent on and value of medical documentation." International journal of medical informatics 84.4 (2015): 221-228.

[3] Barricelli, Barbara Rita, Elena Casiraghi, and Daniela Fogli. "A survey on digital twin: Definitions, characteristics, applications, and design implications." IEEE access 7 (2019): 167653-167671.

[4] Henry, T. "Do you spend more time on administrative tasks than your peers." American Medical Association (2018).

[5] Elkefi, Safa, and Onur Asan. "Digital Twins for Managing Health Care Systems: Rapid Literature Review." Journal of medical Internet research 24.8 (2022): e37641.

[6] Park, Joon Sung, et al. "Generative agents: Interactive simulacra of human behavior." arXiv preprint arXiv:2304.03442 (2023).

[7] Shen Y, Heacock L, Elias J, et al. ChatGPT and other Large Language Models are double-edged swords. Radiology. 2023:230163.

[8] Jeblick, Katharina, et al. "Chatgpt makes medicine easy to swallow: An exploratory case study on simplified radiology reports." arXiv preprint arXiv:2212.14882 (2022).

[9] Torkamannia, Anna, Yadollah Omidi, and Reza Ferdousi. "SYNDEEP: a deep learning approach for the prediction of cancer drugs synergy." Scientific Reports 13.1 (2023): 6184.

[10] Singhal, Karan, et al. "Large language models encode clinical knowledge." arXiv preprint arXiv:2212.13138 (2022).

[11] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.

[12] Ju, Jie, et al. "Robust deep learning model for prognostic stratification of pancreatic ductal adenocarcinoma patients." Iscience 24.12 (2021).

[13] Liu, Xiaoyong, and W. Bruce Croft. "Statistical language modeling for information retrieval." Annu. Rev. Inf. Sci. Technol. 39.1 (2005): 1-31.

[14] Zhai, ChengXiang. "Statistical language models for information retrieval a critical review." Foundations and Trends® in Information Retrieval 2.3 (2008): 137-213.

[15] Thede, Scott M., and Mary Harper. "A second-order hidden Markov model for part-of-speech tagging." Proceedings of the 37th annual meeting of the Association for Computational Linguistics. 1999.

[16] Brants, Thorsten, et al. "Large language models in machine translation." (2007).

[17] Bengio, Yoshua, Réjean Ducharme, and Pascal Vincent. "A neural probabilistic language model." Advances in neural information processing systems 13 (2000).

[18] Miotto, Riccardo, et al. "Deep learning for healthcare: review, opportunities and challenges." Briefings in bioinformatics 19.6 (2018): 1236-1246.

[19] Kombrink, Stefan, et al. "Recurrent Neural Network Based Language Modeling in Meeting Recognition." Interspeech. Vol. 11. 2011.

[20] Singhal, Karan, et al. "Large language models encode clinical knowledge." arXiv preprint arXiv:2212.13138 (2022).

[21] Liévin, Valentin, Christoffer Egeberg Hother, and Ole Winther. "Can large language models reason about medical questions?." arXiv preprint arXiv:2207.08143 (2022).

[22]Matthew, E. "Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer. Deep contextualized word representations." Proc. of NAACL. Vol. 5. 2018.

[23] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[24] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[25] Kaplan, Jared, et al. "Scaling laws for neural language models." arXiv preprint arXiv:2001.08361 (2020).

[26] Wei, Jason, et al. "Emergent abilities of large language models." arXiv preprint arXiv:2206.07682 (2022).

[27] Teng, L et al. *Nan fang yi ke da xue xue bao = Journal of Southern Medical University* vol. 43,6 (2023): 1010-1016. doi:10.12122/j.issn.1673-4254.2023.06.17

[28]Mustafa, Ehzaz, et al. "An Ensembled Framework for Human Breast Cancer Survivability Prediction Using Deep Learning." Diagnostics 13.10 (2023): 1688.

[29] Wang, Thomas, et al. "What language model architecture and pretraining objective works best for zero-shot generalization?." International Conference on Machine Learning. PMLR, 2022.

[30] Zhao, Wayne Xin, et al. "A survey of large language models." arXiv preprint arXiv:2303.18223 (2023).

[31] Wohlin, Claes. "Guidelines for snowballing in systematic literature studies and a replication in software engineering." Proceedings of the 18th international conference on evaluation and assessment in software engineering. 2014.

[32] Talkhestani, Behrang Ashtari, et al. "Consistency check to synchronize the Digital Twin of manufacturing automation based on anchor points." Procedia Cirp 72 (2018): 159-164.

[33] Schleich, Benjamin, et al. "Shaping the digital twin for design and production engineering." CIRP annals 66.1 (2017): 141-144.

[34] Luo, Renqian, et al. "BioGPT: generative pre-trained transformer for biomedical text generation and mining." Briefings in Bioinformatics 23.6 (2022): bbac409.

[35] Beltagy, Iz, Kyle Lo, and Arman Cohan. "SciBERT: A pretrained language model for scientific text." arXiv preprint arXiv:1903.10676 (2019).

[36] Guo, Muzhe, et al. "MSQ-BioBERT: Ambiguity Resolution to Enhance BioBERT Medical Question-Answering." Proceedings of the ACM Web Conference 2023. 2023.

[37] Gu, Yu, et al. "Domain-specific language model pretraining for biomedical natural language processing." ACM Transactions on Computing for Healthcare (HEALTH) 3.1 (2021): 1-23.

[38] Alves, Natalia, et al. "Fully automatic deep learning framework for pancreatic ductal adenocarcinoma detection on computed tomography." Cancers 14.2 (2022): 376.

[39] Hong, Zhi, et al. "The diminishing returns of masked language models to science." Findings of the Association for Computational Linguistics: ACL 2023. 2023.

[40] Cai, Xiaoyan, et al. "Chestxraybert: A pretrained language model for chest radiology report summarization." IEEE Transactions on Multimedia (2021).

[41] Alsharid, Mohammad, et al. "Gaze-assisted automatic captioning of fetal ultrasound videos using three-way multi-modal deep neural networks." Medical Image Analysis 82 (2022): 102630.

[42] Gu, Yu, et al. "Distilling large language models for biomedical knowledge extraction: A case study on adverse drug events." arXiv preprint arXiv:2307.06439 (2023).

[43] Hsieh, Cheng-Yu, et al. "Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes." arXiv preprint arXiv:2305.02301 (2023).

[44] Liu, Peiyu, et al. "Do emergent abilities exist in quantized large language models: An empirical study." arXiv preprint arXiv:2307.08072 (2023).

[45] Yao, Zhewei, et al. "ZeroQuant-V2: Exploring Post-training Quantization in LLMs from Comprehensive Study to Low Rank Compensation." arXiv preprint arXiv:2303.08302 (2023).

[46] Wang, Ziheng, Jeremy Wohlwend, and Tao Lei. "Structured pruning of large language models." arXiv preprint arXiv:1910.04732 (2019).

[47] Sun, Mingjie, et al. "A Simple and Effective Pruning Approach for Large Language Models." arXiv preprint arXiv:2306.11695 (2023).

[48] Chung, Hyung Won, et al. "Scaling instruction-finetuned language models." arXiv preprint arXiv:2210.11416 (2022).

[49] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." arXiv preprint arXiv:2106.09685 (2021).

[50] Kenneweg, Philip, et al. "Intelligent Learning Rate Distribution to Reduce Catastrophic Forgetting in Transformers." International Conference on Intelligent Data Engineering and Automated Learning. Cham: Springer International Publishing, 2022.

[51] Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." Proceedings of the national academy of sciences 114.13 (2017): 3521-3526.

[52] Kaplan, Jared, et al. "Scaling laws for neural language models." arXiv preprint arXiv:2001.08361 (2020).

[53] Gockel, Brian, et al. "Challenges with structural life forecasting using realistic mission profiles." 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA. 2012.

[54] Tuegel, Eric. "The airframe digital twin: some challenges to realization." 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA. 2012.

[55] Abbott, Dean. Applied predictive analytics: Principles and techniques for the professional data analyst. John Wiley & Sons, 2014.

[56] Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." International journal of information management 35.2 (2015): 137-144.

[57] Poornima, Shakthi, and Mullur Pushpalatha. "A journey from big data towards prescriptive analytics." ARPN J. Eng. Appl. Sci 11.19 (2016): 11465-11474.

[58] Barricelli, Barbara Rita, Elena Casiraghi, and Daniela Fogli. "A survey on digital twin: Definitions, characteristics, applications, and design implications." IEEE access 7 (2019): 167653-167671.

[59] Lal, Amos, et al. "Development and verification of a digital twin patient model to predict specific treatment response during the first 24 hours of sepsis." Critical care explorations 2.11 (2020).

[60] Greenspan, Emily, et al. "CAFCW 113 Digital Twins for Predictive Cancer Care: an HPC-Enabled Community Initiative." (2019).

[61] Jameson, J. Larry, and Dan L. Longo. "Precision medicine—personalized, problematic, and promising." Obstetrical & gynecological survey 70.10 (2015): 612-614.

[62] Viceconti, Marco, Adriano Henney, and Edwin Morley-Fletcher. "In silico clinical trials: how computer simulation will transform the biomedical industry." International Journal of Clinical Trials 3.2 (2016): 37-46.

[63] Clapworthy, Gordon, et al. "The virtual physiological human: building a framework for computational biomedicine I. Editorial." Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences 366.1878 (2008): 2975-2978.

[64] Baillargeon, Brian, et al. "The living heart project: a robust and integrative simulator for human heart function." European Journal of Mechanics-A/Solids 48 (2014): 38-47.

[65] Feng, Yu, et al. "An in silico subject-variability study of upper airway morphological influence on the airflow regime in a tracheobronchial tree." Bioengineering 4.4 (2017): 90.

[66] Thawkar, Omkar, et al. "Xraygpt: Chest radiographs summarization using medical vision-language models." arXiv preprint arXiv:2306.07971 (2023).

[67] Papineni, Kishore, et al. "Bleu: a method for automatic evaluation of machine translation." Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002.

[68] Tuegel, Eric J., et al. "Reengineering aircraft structural life prediction using a digital twin." International Journal of Aerospace Engineering 2011 (2011).

[69] Marshall, T. "VPH: The Ultimate stage before your own medical Digital Twin." (2019).

[70] Poon, Hoifung, et al. "Precision Health in the Age of Large Language Models." Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023.

[71] Ahmadi-Assalemi, Gabriela, et al. "Digital twins for precision healthcare." Cyber defence in the age of AI, Smart societies and augmented humanity (2020): 133-158.

[72] HamlAbadi, Kamran Gholizadeh, et al. "Digital Twins in cancer: State-of-the-art and open research." 2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE, 2021.

[73] Ahmadian, Hossein, et al. "A digital twin for simulating the vertebroplasty procedure and its impact on mechanical stability of vertebra in cancer patients." International Journal for Numerical Methods in Biomedical Engineering 38.6 (2022): e3600.

[74] Rod'ko, I. I., et al. "The concept of a digital twin of a radiotherapy system." Biomedical Engineering 53 (2020): 421-424.

[75] Tardini, Elisa, et al. "Optimal treatment selection in sequential systemic and locoregional therapy of oropharyngeal squamous carcinomas: deep Q-learning with a patient-physician digital twin dyad." Journal of medical Internet research 24.4 (2022): e29455.

[76] Bodin, Oleg N., et al. "Visualization of a Digital Twin of the Heart." 2021 IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM). IEEE, 2021.

[77] Golse, Nicolas, et al. "Predicting the risk of post-hepatectomy portal hypertension using a digital twin: A clinical proof of concept." Journal of Hepatology 74.3 (2021): 661-669.

[78] Björnsson, Bergthor, et al. "Digital twins to personalize medicine." Genome medicine 12 (2020): 1-4.

[79] Ahmadian, Hossein, et al. "Toward an artificial intelligence-assisted framework for reconstructing the digital twin of vertebra and predicting its fracture response." International journal for numerical methods in biomedical engineering 38.6 (2022): e3601.

[80] Liu, Ying, et al. "A novel cloud-based framework for the elderly healthcare services using digital twin." IEEE access 7 (2019): 49088-49101.

[82] Shamanna, Paramesh, et al. "Type 2 diabetes reversal with digital twin technology-enabled precision nutrition and staging of reversal: a retrospective cohort study." Clinical Diabetes and Endocrinology 7.1 (2021): 1-8.

[83] Schütt, Michael, et al. "Simulating the hydrodynamic conditions of the human ascending colon: a digital twin of the dynamic colon model." Pharmaceutics 14.1 (2022): 184.

[84]Maes, Michael. "Precision nomothetic medicine in depression research: a new depression model, and new endophenotype classes and pathway phenotypes, and a digital self." Journal of personalized medicine 12.3 (2022): 403.

[85] James, Lindsay. "Digital twins will revolutionise healthcare: Digital twin technology has the potential to transform healthcare in a variety of ways–improving the diagnosis and treatment of patients, streamlining preventative care and facilitating new approaches for hospital planning." Engineering & Technology 16.2 (2021): 50-53.

[86] Baranov, Leonid I., et al. "Experience of Creating Digital Twins as a Result of Monitoring the Health of Industrial Personnel." International Scientific and Practical Conference Strategy of Development of Regional Ecosystems "Education-Science-Industry"(ISPCR 2021). Atlantis Press, 2022.

[87] Kolekar, Shivani Sanjay, Haoyu Chen, and Kyungbaek Kim. "Design of Precision Medicine Web-service Platform Towards Health Care Digital Twin." 2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN). IEEE, 2023.

[88] Elshaier, Yaseen AMM, et al. "A Proposed Framework for Digital Twins Driven Precision Medicine Platform: Values and Challenges." Digital Twins for Digital Transformation: Innovation in Industry. Cham: Springer International Publishing, 2022. 67-86.

[89] Saghiri, Mohammad Ali, Julia Vakhnovetsky, and Ali Mohammad Saghiri. "The future of digital twins in precision dentistry." *Journal of Oral Biology and Craniofacial Research* 13.1 (2023): 19.

[90] Fagherazzi, Guy. "Deep digital phenotyping and digital twins for precision health: time to dig deeper." *Journal of medical Internet research* 22.3 (2020): e16770.

[91] Gkouskou, Kalliopi, et al. "The “virtual digital twins” concept in precision nutrition." *Advances in Nutrition* 11.6 (2020): 1405-1413.

[92] Hernandez-Boussard, Tina, et al. "Digital twins for predictive oncology will be a paradigm shift for precision cancer care." *Nature medicine* 27.12 (2021): 2065-2066.

[93] Wickramasinghe, Nilmini, et al. "Digital twins to enable better precision and personalized dementia care." *JAMIA open* 5.3 (2022): ooac072.

[94] Coorey, Genevieve, et al. "The health digital twin: advancing precision cardiovascular medicine." *Nature Reviews Cardiology* 18.12 (2021): 803-804.

[95] Venkatesh, Kaushik P., Marium M. Raza, and Joseph C. Kvedar. "Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation." *NPJ digital medicine* 5.1 (2022): 150.

[96] Corral-Acero, Jorge, et al. "The ‘Digital Twin’to enable the vision of precision cardiology." *European heart journal* 41.48 (2020): 4556-4564.