Towards Digital Twins in Rehabilitation Processes - Examples of Publicly Available Datasets

Barbara Jantos[1,*,†], Michał Tomaszewski [1,†]

[1] Opole University of Technology, Prószkowska 76 Street, 45-758 Opole, Poland

Abstract

This article presents the proposed application of digital twins in rehabilitation. Such a system could lead to an enhancement in the quality of rehabilitation care. It highlights the potential benefits of incorporating such a system into therapy and discusses challenges connected to the issue. The main goal of the article is to discuss publicly available datasets that could be utilized to create a digital twin and to point out their advantages and disadvantages. The presented datasets encompass diverse aspects of rehabilitation: from patient characterization with specific conditions to sensors and wearables utilization, exercise performance monitoring, extended reality integration, and characterization of healthy individuals as the target patient state. As no perfect dataset for digital twins was found, further research is suggested as the direction for achieving a serviceable digital twin in rehabilitation.

Keywords

digital twin, dataset, rehabilitation, artificial intelligence

1. Introduction

The field of rehabilitation is undergoing a significant transformation, with information technology (IT) playing an increasingly crucial role in supporting patients on their road to recovery. This trend is driven by several factors, such as improved accessibility and reach, enhanced engagement and motivation, data-driven insights, and personalized care or improved monitoring and feedback.

According to [1,2], traditional rehabilitation often requires frequent in-person visits with therapists, which can be a barrier for patients in remote areas or those with limited mobility. IT solutions like telerehabilitation, which utilizes video conferencing for remote therapy sessions, address this concern by making rehabilitation more accessible. Telerehabilitation involves delivering rehabilitation services via telecommunication networks or the Internet, enabling remote treatments for individuals at home or from a distance. The emergence of COVID-19 has significantly strained the healthcare system, preventing many patients from accessing in-person treatments.

[1]ITTAP'2024: 4th International Workshop on Information Technologies: Theoretical and Applied Problems, November 20–22, 2024, Ternopil, Ukraine, Opole, Poland
* Corresponding author.
† These authors contributed equally.
b.jantos@student.po.edu.pl (B. Jantos); m.tomaszewski@po.edu.pl (M. Tomaszewski) 0009-0001-9557-6787 (B. Jantos); 0000-0001-6672-3971 (M. Tomaszewski)

Additionally, individuals with chronic or long-term health conditions have been unable to maintain their regular follow-up appointments, and healthcare professionals cannot attend all consultations. Telerehabilitation effectively addresses this gap by facilitating and expanding access to rehabilitation services.

Rehabilitation exercises can be repetitive and sometimes tedious. Technologies like virtual reality (VR), augmented reality (AR), and gamification [3] can create immersive and engaging environments that make therapy sessions more enjoyable and motivating for patients, leading to potentially better adherence to exercise programs.

Wearable sensors and motion capture technologies can collect real-time patient progress data by monitoring various parameters such as movement patterns, joint angles, muscle activity, and overall physical performance. These advanced technologies [4] provide detailed and continuous data, enabling therapists to understand the patient's condition comprehensively. This data allows therapists to personalize treatment plans based on the patient's unique needs and progress, ensuring that the interventions are as effective as possible. Additionally, the objective nature of the data collected helps track progress over time with high precision. Therapists can use this information to make datadriven decisions, adjusting the treatment plans as necessary to optimize outcomes. By identifying specific areas needing improvement, therapists can target their interventions more accurately, addressing the precise aspects of the patient's rehabilitation that require attention. This tailored approach not only enhances the efficiency of the rehabilitation process but also improves patient engagement and motivation, as they can see tangible evidence of their progress. Additionally, as stated in [5], IT-based monitoring systems can track patients' adherence to exercise plans, vital signs, and overall well-being. This allows for early intervention in case of setbacks and provides valuable feedback to patients and therapists.

A digital twin (DT) is a virtual representation of physical objects, systems, or processes created using real-time data and simulation models. Digital twins serve as dynamic, realtime digital counterparts of their physical counterparts, enabling continuous monitoring, analysis, and optimization. The concept of digital twins integrates several advanced technologies, including the Internet of Things (IoT), artificial intelligence (AI), machine learning (ML), and big data analytics. The idea of digital twins originated in the manufacturing sector, where it was first introduced by Michael Grieves in 2002 [6]. Initially, digital twins were used to enhance product lifecycle management (PLM) by providing detailed insights into the performance and condition of products. Over time, as technology advanced and the benefits of digital twins became more apparent, their application expanded to various other fields.

In manufacturing, digital twins are used to model and simulate production processes, helping to identify bottlenecks, optimize workflows, and reduce downtime [7]. They also enable predictive maintenance by monitoring machinery and equipment in real-time, predicting failures, and scheduling maintenance proactively, which minimizes disruptions and extends the lifespan of assets [8]. Moreover, engineers and designers use digital twins to virtually test and validate new products, reducing the need for physical prototypes and accelerating the development cycle [9].

Smart cities utilize digital twins for urban planning, simulating various scenarios to help planners and policymakers make informed decisions about infrastructure development, traffic management, and resource allocation [13]. Smart grids use digital twins to optimize energy distribution and consumption, enhancing efficiency and sustainability. Furthermore, digital twins can model potential disaster scenarios, aiding in emergency preparedness and response planning.

In the automotive industry, digital twins are employed in vehicle design and testing, simulating performance under different conditions to reduce the need for physical testing and accelerate the design process [9]. Connected vehicles leverage digital twins for realtime monitoring and diagnostics, improving maintenance and safety. The aerospace and defense sectors benefit from digital twins through aircraft maintenance monitoring and predictive analytics, enhancing safety and reducing operational costs [14]. Defense systems use digital twins to simulate and optimize mission strategies and operations.

According to [15] in the energy sector, digital twins are applied in oil and gas industries to monitor equipment, optimize production, and ensure safety. Renewable energy sources, such as wind farms and solar power plants, use digital twins to maximize energy generation and improve maintenance practices.

In healthcare, digital twins facilitate personalized medicine by simulating and predicting individual treatment responses, enabling precise medical interventions tailored to each patient [10]. Surgeons use digital twins to practice and plan complex procedures, improving accuracy and outcomes [11]. Additionally, digital twins support remote monitoring of patients with chronic conditions, allowing for timely interventions and better management of health issues [12]. Digital twins in rehabilitation involve creating detailed and dynamic digital models of patients using real-time data from sources like wearable sensors, motion capture systems, medical imaging, and health records. These digital replicas provide continuous, real-time feedback that allows healthcare providers to tailor and enhance rehabilitation protocols precisely.

2. Digital Twins in Rehabilitation

Digital twins in rehabilitation refers to creating highly detailed and dynamic digital models of patients undergoing rehabilitation. These digital replicas are built using realtime data gathered from various sources, such as wearable sensors, motion capture systems, medical imaging, and patient health records. By continuously updating with live data, digital twins provide a comprehensive and evolving representation of the patient’s physical and physiological state. The core concept of digital twins in rehabilitation involves integrating advanced IT technologies. These technologies allow the digital twin to process and analyze the collected data, facilitating a deeper understanding of the patient's progress and needs. This continuous, real-time feedback loop enables healthcare providers to tailor rehabilitation protocols precisely, enhancing the effectiveness and efficiency of treatment plans.

It can be distinguished by plenty of application areas of DT in rehabilitation. The main ones include:

- Personalized therapy planning - Digital twins customise rehabilitation programs to suit each patient's unique needs [16]. By simulating different treatment scenarios and predicting outcomes, therapists can design and adjust personalized therapy plans that maximize recovery. This approach helps identify each individual's most effective exercises and interventions, ensuring optimal results.
- Real-time monitoring, long-term monitoring, and feedback (f.e. enhanced recovery tracking) - According to [17], digital twins offer immediate feedback on a patient’s performance

during rehabilitation exercises through continuous data collection and realtime analysis. This allows therapists to monitor progress closely and make necessary adjustments on the fly. Real-time feedback also empowers patients to perform exercises correctly, improving adherence to therapy protocols and reducing the risk of injury or incorrect movements. Digital twins can continuously monitor patients' health (for example, in post-Covid therapy [18]), tracking their progress over extended periods and providing data on vital signs, respiratory function, and physical performance. This enables personalized adjustments to rehabilitation plans as patients recover. Digital twins offer detailed insights into the recovery trajectory of post-Covid patients, helping to identify patterns and predict long-term outcomes. This can guide public health strategies and resource allocation for ongoing pandemic management.

- Remote rehabilitation and telehealth - The Covid-19 pandemic has introduced new challenges for healthcare systems, particularly in the realm of rehabilitation. Patients recovering from Covid-19 often experience prolonged symptoms and require extensive rehabilitation to regain their pre-illness health status. Digital twins facilitate remote rehabilitation by enabling healthcare providers to monitor and guide patients from a distance. Patients can perform their exercises at home while their digital twin provides continuous data to their therapist. This capability is particularly valuable for patients with difficulty accessing in-person therapy sessions due to geographical, mobility, or pandemicrelated constraints.
- Simulation, predictive analytics, and outcome forecasting - Review [19] shows that digital twins can be used to create virtual simulations for training patients and healthcare providers. These simulations can help patients understand and practice complex movements in a controlled virtual environment. For healthcare providers, digital twins offer a platform to train and refine their skills in diagnosing and treating various rehabilitation scenarios without risk to real patients. Digital twins can leverage predictive analytics to forecast the likely outcomes of different rehabilitation strategies. By analyzing historical and current data, they can predict future patient progress, identify potential setbacks, and suggest proactive measures. This capability helps set realistic goals and timelines for recovery, enhancing patient and therapist expectations and planning.
- Enhanced patient engagement and motivation - Engagement and motivation are critical to successful rehabilitation. Digital twins can provide visualizations and simulations that help patients understand their progress and the impact of their efforts. By visualizing improvements and future potential outcomes, patients are more likely to stay motivated and committed to their rehabilitation programs.
- Economic importance and resource optimization - By utilizing digital twins, healthcare providers can optimize the use of limited resources, focusing on patients needing the most intensive support while enabling others to engage in self-guided rehabilitation under remote supervision [20].

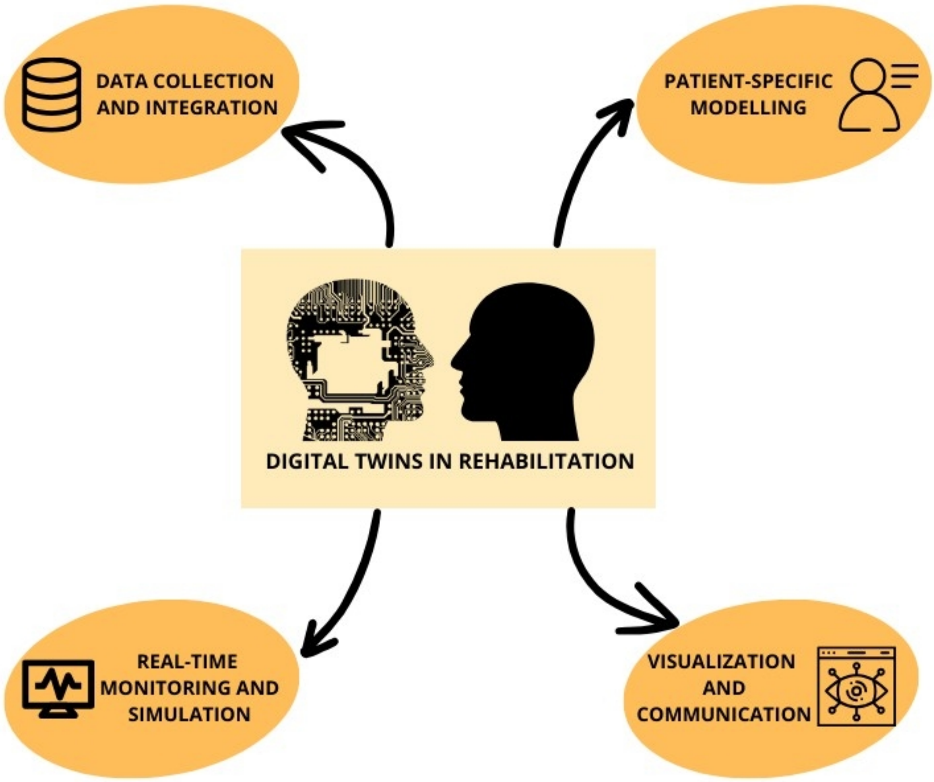

Figure 1: Elements necessary to create a Digital Twin in the field of rehabilitation.

The literature review carried out as part of this article indicates that several key elements are crucial to constructing a digital twin for post-COVID rehabilitation effectively. The main ones are shown in Fig. 1 and described below:

- Data collection and integration:
  - Clinical data: Comprehensive clinical data, including patient demographics, medical history, COVID-19 severity, current symptoms, and laboratory findings, must be gathered. This data can be obtained from electronic health records (EHRs), patient interviews, and physical examinations.
  - Biomedical data: Integration of physiological data such as heart rate, respiratory rate, oxygen saturation, blood pressure, and sleep patterns is essential. This data can be collected using wearable sensors, home monitoring devices, and hospital telemetry systems.
  - Functional data: The assessment of the patient's functional abilities, including mobility, balance, strength, and cognitive function, is crucial. This can be done through standardized assessments, physical therapy evaluations, and selfreported questionnaires.
- Patient-specific modeling:
  - Physiological models: Computational models representing the patient's physiological systems, such as the cardiovascular, respiratory, and immune systems, should be developed. These models must incorporate the patient's specific characteristics and COVID-19 history.
  - Functional models: Models capturing the patient's functional abilities, including musculoskeletal dynamics, neuromuscular control, and cognitive performance, need to be constructed. These models should be personalized based on the patient's assessment results.

◦ Integration of patient data: The collected clinical, biomedical, and functional data should be integrated into the patient-specific models. This integration will allow the digital twin to reflect the patient's current state and respond to changes in their condition.

- Real-time monitoring and simulation:

◦ Continuous data stream: A continuous data stream from wearable sensors, home monitoring devices, and other sources must be established to update the digital twin in real-time. This ensures that the twin accurately reflects the patient's current physiological and functional status.

◦ Simulation of interventions: The effects of potential rehabilitation interventions, such as exercise programs, medication adjustments, and alternative therapies, should be simulated on the digital twin. This allows clinicians to predict how patients respond to different treatment options.

◦ Predictive analytics: Machine learning algorithms should be utilized to analyze the data and predict potential complications, such as secondary infections, long-term sequelae, or relapses. This enables proactive interventions and personalized risk management.

- Visualization and communication:

◦ Interactive dashboards: An interactive dashboard that visualizes the patient's digital twin, including their physiological parameters, functional status, and predicted outcomes, should be developed. This dashboard must be accessible to clinicians, therapists, and patients themselves.

◦ Real-time alerts: Real-time alerts notifying clinicians and patients of any significant changes in the patient's digital twin, indicating potential concerns or the need for intervention, should be implemented.

◦ Personalized communication tools: Tools facilitating personalized communication between clinicians, therapists, and patients using the digital twin as a shared reference point should be developed. This promotes collaborative care and patient engagement.

The literature review highlights that the initial step in creating a digital twin for rehabilitation purposes is the collection of appropriate datasets. Comprehensive data collection is fundamental as it provides the essential input needed to build accurate and reliable digital models. This article, therefore, identifies and presents examples of publicly available datasets that are particularly relevant for this purpose. These datasets encompass a wide range of information, including physiological, clinical, and functional data, which are critical for constructing a detailed and dynamic digital twin. By utilizing these resources, researchers can lay a solid foundation for the development and testing of a digital twin prototype specifically designed for the rehabilitation of post-COVID patients.

Physiological datasets might include measurements of heart rate, respiratory rate, oxygen saturation, and other vital signs that can be continuously monitored using wearable sensors. Clinical datasets could provide insights into patient demographics, medical history, the severity of COVID-19 infection, and current symptoms, which are crucial for tailoring the digital twin to individual patient profiles. Functional datasets may offer information on mobility, balance, strength, and cognitive function, all of which are important for assessing and improving rehabilitation outcomes.

The rest of the article delves into these datasets in detail, discussing their origins, contents, and how they can be effectively used in the context of digital twin development. Additionally, it explores the potential of these datasets to support various stages of the digital twin lifecycle, from initial modeling to real-time updates and predictive analytics.

3. Leveraging Public Datasets for Digital Twin Construction in Rehabilitation

Finding publicly available data suitable for digital twin creation is a significant challenge due to the sensitive nature of such data. Perfect data for digital twin should be gathered per one patient as on observation for a longer time with carefully annotated timestamps. The data should be comprehensive and encompass various diseases, interventions, therapies, rehabilitation processes, medications, exercises, and dietary patterns. Additionally, it should capture the patient's lifestyle factors, which significantly impact their rehabilitation and recovery journey. Unfortunately, no suitable dataset has been identified that meets the specified criteria. Instead, we present samples of publicly available datasets worth considering in prototyping a digital twin. Datasets found during the literature review could be divided into the following categories, each one could be beneficial for constructing DT in different scopes:

- Clinical data containing non-specific conditions, excluding mental health, children, and fetal data, in this category most of the datasets focused on ECG records and EEG signals.
- Clinical - general - category for datasets which includes a lot of general patient data like whole electronic health records data. Such a dataset could be crucial in terms of creating a digital twin.
- Clinical - imaging - datasets with images, mostly in DICOM format, or datasets with image annotations.
- Clinical - specific condition - data focused on subjects with specific conditions like stroke patients or diabetes patients.
- Gait, posture and movement - category for datasets that describe body position, movement specific, or pose 3D points. It could be beneficial when it comes to asses exercise quality or detecting the impact of the movement.
- VR and AR - category for datasets that concern the usage of extended reality methods. Only one dataset was found.
- Wearables + sensors - a category that gathers datasets focused especially on wearables and their ability to measure human health signals or exercise performance. Creating a digital twin requires a variety of datasets. In a perfect-world scenario, such datasets will mirror real patients and their reactions to therapy. However, it is challenging to prepare such a dataset, so it is an emerging task in such a domain. To highlight current possibilities, sample available datasets are presented that could be useful in the flow of digital twins in rehabilitation.

The presented diagram shows elements of digital twins in rehabilitation. Red squares were used for existing sample datasets that could be useful in creating a DT. Blue squares represent domains that lack such datasets. Even if some data exists, it is unsuitable for DT in rehabilitation or unavailable in digitalized versions in open access.

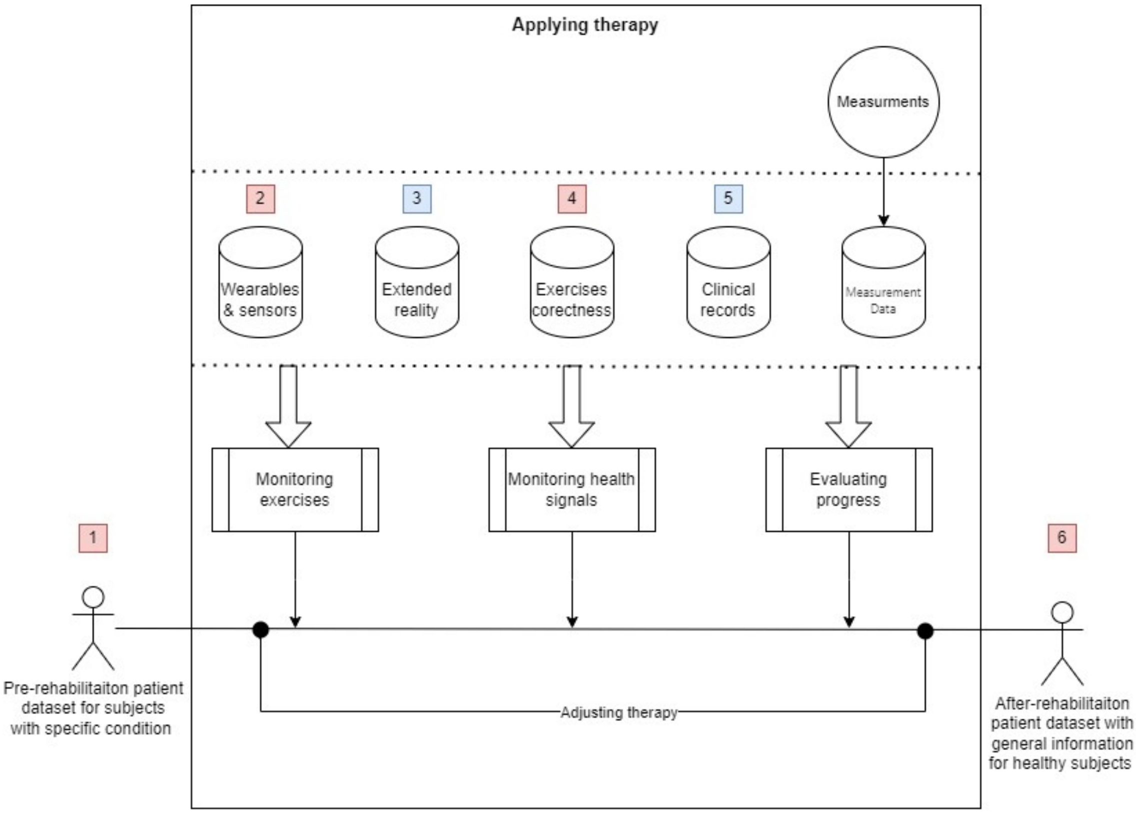

Figure 2: Datasets in proposed digital twin in rehabilitation.

The first and fifth stages of the proposed system consist of datasets dedicated to the description of specific medical conditions per patient. Data should be especially focused on features that should be carefully monitored in particular situations to reflect those characteristics in digital twins.

The second element could focus on the element of therapy connected to monitoring exercises and body reactions. The third could harness possibilities of extended reality as a way to visualize exercises adjusted to the patient. The next part could be dedicated to exercising correctness so every subject will be allowed to check if their posture and movement are performed well. Those elements need clinical monitoring and medical evaluation, and they could be partially gathered by electronic systems but still, they have to be assessed by medical professionals.

All those parts need a reference point, which could be a digital twin of a healthy subject. To achieve that it is needed to gather a variety of data, presenting people in different stages of life and with diverse lifestyles.

3.1. Pre-rehabilitation Patient Dataset for Subjects with Specific Condition

As an entry point, the Patient-level dataset to study the effect of COVID-19 in people with multiple sclerosis [22] could be utilized as it is focused on patients with specific medical conditions. The dataset consists of 1141 people and is fully de-identified. It covers both subjects' demographics and symptoms. Some of the variables have missing values. The data was gathered through a questionnaire delivered by the central platform of the COVID-19 and MS Global Data Sharing Initiative. Data is in tabular format with columns as presented in table 1. It could be even more resourceful for COVID-19 patient digital twins if it would cover also a detailed description of clinical records and completed therapy.

Table 1
Attributes of Patient-level dataset to study the effect of COVID-19 on people with Multiple Sclerosis

Attribute	Description	Has missing values
secret_name	Identifier of patient	No
report_source	Source of data. It can be either 'clinicians' for data reported by medical professionals or 'patients' for data reported directly by the patients themselves.	No
age_in_cat	Category of age: 0 for age (0,18); 1 for <18,50>; 2 for <51, 70>, 3 for > 70	No
bmi_in_cat2	“not_overweight” for patients with BMI <= 30; “overweight” for BMI > 30	Yes
covid19_admission_hospital	“Yes” if patient was admissioned to the hospital due to COVID-19	No
covid19_confirmed_case	“Yes” for positive and confirmed COVID-19 diagnosis; “No” otherwise	No
covid19_diagnosis	One of values “not_suspected”, “suspected”, “confirmed” - diagnosis of the subject	No
covid19_has_symptoms	“No” if no COVID-19 symptom was observed, “Yes” otherwise	Yes
covid19_icu_stay	“Yes” if patient was in ICU as a result of COVID-19, otherwise “No”	Yes
covid19_self_isolation	“Yes” or “No”	Yes
covid19_sympt_chills	“Yes” if chills were present, “No” otherwise	Yes
covid19_sympt_dry_cough	Yes” if dry cough was present, “No” otherwise	Yes
covid19_sympt_fatigue	Yes” if fatigue was present, “No” otherwise	Yes

covid19_sympt_ fever	Yes” if fever was present, “No” otherwise	Yes
covid19_sympt_loss _smell_taste	Yes” if loss of smell or taste was present, “No” otherwise	Yes
covid19_sympt_ nasal_congestion	Yes” if nasal congestion was present, “No” otherwise	Yes
covid19_sympt_ pain	Yes” if pain was present, “No” otherwise	Yes
covid19_sympt_ pneumonia	Yes” if pneumonia was present, “No” otherwise	Yes
covid19_sympt_ shortness_breath	Yes” if shortness of breath was present, “No” otherwise	Yes
covid19_sympt_sore _throat	Yes” if sore throat was present, “No” otherwise	Yes
covid19_ventilation	Yes” if ventilator unit was used during hospital stay, “No” otherwise	Yes
current_dmt	Disease-Modifying Therapy (DMT) Status - this variable captures a patient's current status regarding disease-modifying therapy at the time of data entry. It has three categories: Yes: Patient is currently receiving DMT. No: Patient is not currently receiving DMT. Never Treated: Patient has never received DMT.	No
dmt_glucocorticoid	“Yes” if patient is taking glucocorticoid, “No” otherwise	Yes
edss_in_cat2	Expanded Disability Status Scale (EDSS) Category: 0: EDSS score between 0 and 6 (no to moderate disability). 1: EDSS score above 6 (severe disability).	Yes
Pregnancy	“Yes” if subject is pregnant, “No” otherwise	Yes
Sex	The biological sex of the patient. “Male” for males and “Female” for females.	No

ms_type2	Multiple Sclerosis (MS) Phenotype: “relapsing_remitting”: Relapsing-Remitting: “progressive_MS”: characterized by steady worsening of symptoms (includes secondary progressive and primary progressive). “other”: Includes clinically isolated syndrome (CIS), missing data, or patient/clinician uncertainty.	No
current_or_former_ smoker	“Yes” if subject is or was a smoker, “No” otherwise	No
dmt_type_overall	Category of Dimethyltryptamine on which is subject	Yes
duration_treatment _cat	0 - for subjects whose treatment is less than 11 years, 1 otherwise	Yes
stop_or_end_date_ combined	Date of stopping DMT	Yes
covid19_outcome_ levels_2	0 - for subjects with COVID-19 but not hospitalized 1 - for subjects with COVID-19 and hospitalized 2 - for subjects hospitalized, in ICU and/or in a ventilation facility	No
has_comorbidities	“Yes” if any comorbidities are present, “No” otherwise	No
com_cardiovascular _disease	“Yes” if any cardiovascular comorbidities are present, “No” otherwise	Yes
com_chronic_kidney _disease	“Yes” if any chronic kidney diseases are present, “No” otherwise	Yes
com_chronic_liver_	“Yes” if any chronic liver diseases are present, “No”	Yes
disease	otherwise	
com_diabetes	“Yes” if diabetes is present, “No” otherwise	Yes
com_hypertension	“Yes” if hypertension is present, “No” otherwise	Yes
com_immunodeficie ncy	“Yes” if immunodeficiency is present, “No” otherwise	Yes
com_lung_disease	“Yes” if any lung disease is present, “No” otherwise	Yes
com_malignancy	“Yes” if any malignancy is present, “No” otherwise	Yes
com_neurological_ neuromuscular	“Yes” if any neurological and/or neuromuscular comorbidities are present, “No” otherwise	Yes
comorbidities_other	Names of other comorbidities	Yes

The different dataset that could be used to prepare a digital twin for subjects with medical conditions is eICU Collaborative Research Database Demo [23]. This dataset does not focus on only one disease as it contains records from Intensive Care Unit admissions for 2520 patients in the demo version. On the one hand, it allows building a more comprehensive database for digital twin modeling, on the other it could miss conditionspecific characteristics. The database is available in SQLite database and CSV file format. The dataset is completely de-identified and was developed with the Philips Healthcare telehealth system. It covers a wide range of patient data presented in tables as stated in Table 2. Creating a digital twin it lacks results of treatment per patient at the end of ICU treatment and reactions for applied therapy.

Table 2
Tables available in eICU Collaborative Research Database Demo

Table	Description
admissionDrug	Admission diagnosis
admissionDx	Medication taken prior to admission
allergy	Patient allergies
apacheApsVar	Acute Physiology Score (APS) III for patients
apachePatientResults	Predictions made by the APACHE score
apachePredVar	Variables underlying the APACHE predictions
carePlanCareProvider	Intervention category of the care provider
carePlanEOL	End of life care
carePlanGeneral	General plan of care
carePlanGoal	Treatment goals
carePlanInfectiousDise ase	Care plan part for infectious diseases
customLab	Unmapped in tool lab tests
diagnosis	Patient diagnosis
hospital	Hospital data
infusionDrug	Details of drug infusions
intakeOutput	Patients' intakes and outputs
lab	Standard lab tests
medication	Patients' active medication
microLab	Microbiology data
note	Notes
nurseAssessment	Nurse assessment documentation
nurseCare	Nurse care documentation
nurseCharting	Semi-structured data entered by the nurse
pastHistory	Patient's relevant past medical history
patient	Patient demographics and admission data
physicalExam	Results of a physical exam
respiratoryCare	Data of respiratory care such as vent start, airway type and others
respiratoryCharting	Respiratory chart data

treatment	Active treatments
vitalAperiodic	Invasive vital sign data at irregular intervals
vitalPeriodic	Invasive vital sign data at regular intervals (5 minute median)

3.2. Wearables and Sensors Datasets

The second part of the proposed environment - wearables and sensors could be crucial to creating digital twins dedicated to rehabilitation and physiotherapy, especially in remote contexts. One of the datasets that could power the digital twin model is ScientISST MOVE: Annotated Wearable Multimodal Biosignals recorded during Everyday Life Activities in Naturalistic Environments [24]. Data was gathered with three wearable devices: a chestband, an armband, and the Empatica E4 wristband worn by 17 healthy people. The authors provided a subject info file that contains age, sex, and clinical history info data - it is worth highlighting that 11 of 17 subjects recovered from COVID-19, so this dataset could be thematically linked to the previously mentioned dataset. Data was collected for baseline state, lifting a chair, greetings, gesticulation, jumping, walking before running, running outside, and walking after running. Wearables allow to acquire signals for:

- electrocardiogram (ECG),
- electrodermal activity (EDA),
- photoplethysmogram (PPG),
- electromyogram (EMG),
- skin temperature (TEMP),
- chest acceleration (C-ACC), • wrist acceleration (W-ACC).

The dataset would be more valuable if the experiments included more exercises performed by a larger number of volunteers. Additionally, not every routine was performed by every participant and some of the data were excluded from the dataset because of poor data quality. Despite its imperfections, it remains a relevant resource for modeling the body's exercise response with sensors. Such sensors could be useful in monitoring muscle activity during rehabilitation or in gathering athlete’s fitness data for improving their performance [25].

3.3. Extended Reality Datasets

The next part of the suggested schema is incorporating VR and AR, these technologies are increasingly being incorporated into the healthcare processes [26]. However, a review of the literature found no datasets specifically focused on rehabilitation and VR or AR. The only dataset Body Sway When Standing and Listening to Music Modified to Reinforce Virtual Reality Environment Motion [27] covers the topic of the impact of music in virtual reality on body sway. Collected data is not suitable for digital twins in rehabilitation, nonetheless, the general direction of using virtual reality is worth exploring. It could be used to simulate exercises by highlighting body points that need special attention during exercises. Extended reality is a great opportunity to provide feedback based on wearables and sensors for the patient, who is representing the physical twin in the digital twin schema.

3.4. Exercises Correctness Datasets

A digital twin could further enhance patient support by incorporating a physiotherapy exercise dataset within its structure - corresponding to the fourth element of the diagram. An example of such a dataset is the Classification of Physiotherapy Exercises Dataset [28]. In this dataset, the data of 7 exercises (knee-rolling, bridging, pelvic tilt, "The Clam", repeated extension in lying, prone punches, superman) performed by 30 subjects is presented. It contains data from three sensors:

- Obbrec Astra Depth Camera,
- Sensing Tex Pressure Mat,
- Axivity AX3 3-Axis Logging Accelerometer (placed on the wrist and the thigh).

Although this dataset provides details about performing exercises, it does not describe the subjects and the exact conditions of the test. Those could be found in not linked article [29]. As the authors state – such data could show the differences between proper execution and patient execution. In terms of digital twins, it could be extended to model the outcomes of every exercise with its risks and opportunities.

The mentioned dataset is relatively small, although some ML algorithms achieve promising results even with a small number of cases [30]. It could be useful in terms of personal trainer assistance, however, it is not enough to build a DT alone. To be more useful it should cover also the impact of exercises on subjects over a longer period, about specific conditions. Another challenge is to create a vast dataset with a variety of performed exercises with multiple sensors executed by subjects with different possibilities – all of them should be supervised by a specialist to ensure high-quality data.

3.5. Clinical Records and Measurements

The proposed system's fifth component lies at the intersection of medical expertise and electronic medical data collection. With the development of electronic health records, more and more decision processes could be digitized, however, in such sensitive areas as rehabilitation, it is necessary to include human doctors who will be responsible for making decisions, evaluating results, and designing further therapy. A whole digital twin cannot exist without specialized, human assessment – it could help in the process, and provide better communication and monitoring but it is not a replacement for professional care.

3.6. After-rehabilitation Patient Datasets with General Information for Healthy Subjects

The final sixth element of the proposed schema includes a dataset encompassing healthy subject data. A well-suited example is 'Autonomic Aging: A dataset to quantify changes of cardiovascular autonomic function during healthy aging [31]. This comprehensive dataset features resting ECG recordings and continuous non-invasive blood pressure measurements from 1,121 subjects across 15 age groups (18-92 years). Additionally, it provides sex and BMI information. While offering a broad age range, the dataset could be further enriched with detailed data on mild activity parameters and subjects' medical history, including past illnesses and therapies.

The goal healthy digital twin also should be equipped with a measure that will quantify how far from an ideal healthy person is a physical twin – the patient. It could be useful to predict at the very beginning of the therapy and during the process estimated level of returning to health for the individual. Such a possibility would have a direct impact on health expenses, therapy planning, and chances of getting back to work for a person. This measure could be remodeled with digital twins depending on the version of therapy without the necessity of applying each one to the patient. In a perfect world scenario, this will save time for medical staff and patients and it will significantly improve life quality without the need for exposure to ineffective therapies.

4. Conclusion

The adoption of digital twin technology is transforming industries by enabling more informed decision-making, enhancing operational efficiency, and driving innovation. As technologies like AI, IoT, and big data continue to evolve, the capabilities of digital twins are expected to expand further. Future advancements may include more sophisticated simulations, greater integration with augmented and virtual reality, and broader application across emerging fields such as quantum computing and space exploration. Digital twins represent a pivotal shift towards a more connected and data-driven world, where physical and digital realms converge to unlock new possibilities and drive progress across diverse domains.

Creating a digital twin for the rehabilitation of post-COVID patients therefore requires the synergy of various technologies and approaches that together create a comprehensive and dynamic patient model that allows for precise monitoring and optimization of rehabilitation processes. A key element is the availability of specific public datasets that will allow the conducting of appropriate research experiments. For many reasons mentioned in the article, such collections are still missing. While illustrative datasets have been demonstrated, none are sufficient on their own. To prepare a better dataset for digital twins it is necessary to gather and process a significant amount of health records that describe a variety of patients during a long period, each described by specialists. This poses a challenge due to the ongoing slow digitization of medical records, the need for data anonymization, and the adherence to the highest data protection standards.

Another obstacle to creating a useful digital twin is designing a connection between the physical patient and the digital representation. It should be resistant to possible noise from gathered data but sensitive enough to alarm medical staff in case of emergency. his has the potential to significantly enhance the quality of telemedicine and expand patient access to professional rehabilitation services. Furthermore, such a system could potentially reduce therapy costs, shorten therapy duration, and assist specialists in therapy planning. To get closer to achieving a digital twin, additional studies are required in collaboration with medical professionals.

References

[1] S. Rutkowski, Ladislav Batalik, J. Papathanasiou, and M. Sacco, Editorial: Enhancing the rehabilitation process with digital technologies - solutions for public health, *Frontiers in public health*, vol. 12, no. 12, Feb. 2024, doi: https://doi.org/10.3389/fpubh.2024.1366077.

[2] A. A. Seid, S. B. Aychiluhm, and A. A. Mohammed, Effectiveness and feasibility of telerehabilitation in patients with COVID-19: a systematic review and metaanalysis, *BMJ Open*, vol. 12, no. 10, p. e063961, Oct. 2022, doi: https://doi.org/10.1136/bmjopen-2022-063961.

[3] A. Hernandez, L. Bubyr, P. Archambault, J. Higgins, M. F. Levin, and D. Kairy, VRbased rehabilitation as a Feasible and Engaging Tool for the Management of Chronic Post-Stroke Upper Extremity Function Recovery: A Randomized Controlled Trial, *JMIR Serious Games*, vol. 10 (3), Feb. 2022, doi: https://doi.org/10.2196/37506.

[4] R. De Fazio, V. M. Mastronardi, M. De Vittorio, and P. Visconti, Wearable Sensors and Smart Devices to Monitor Rehabilitation Parameters and Sports Performance: An Overview, *Sensors*, vol. 23, no. 4, p. 1856, Feb. 2023, doi: https://doi.org/10.3390/s23041856.

[5] H. K. Hughes, B. W. Hasselfeld, and J. A. Greene, Health Care Access on the Line — Audio-Only Visits and Digitally Inclusive Care, *New England Journal of Medicine*, vol. 387, no. 20, Nov. 2022, doi: https://doi.org/10.1056/nejmp2118292.

[6] M. Grieves and J. Vickers, Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems, in *Transdisciplinary Perspectives on Complex Systems*, Springer International Publishing, 2017, pp. 85–113. doi: https://doi.org/10.1007/978-3-319-38756-7_4.

[7] F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, Digital twin driven prognostics and health management for complex equipment, *CIRP Annals*, vol. 67, no. 1, pp. 169– 172, 2018, doi: https://doi.org/10.1016/j.cirp.2018.04.055.

[8] W. Hu, T. Zhang, X. Deng, Z. Liu, and J. Tan, Digital twin: a state-of-the-art review of its enabling technologies, applications and challenges, *Journal of Intelligent Manufacturing and Special Equipment*, vol. 2, no. 1, pp. 1–34, Jul. 2021, doi: https://doi.org/10.1108/jimse-12-2020-010.

[9] E. Negri, L. Fumagalli, and M. Macchi, A Review of the Roles of Digital Twin in CPS-based Production Systems, *Procedia Manufacturing*, vol. 11, pp. 939–948, 2017, doi: https://doi.org/10.1016/j.promfg.2017.07.198.

[10] K. Bruynseels, F. Santoni de Sio, and J. van den Hoven, Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm, *Frontiers in Genetics*, vol. 9, no. 31, p. 31, 2018, doi: https://doi.org/10.3389/fgene.2018.00031.

[11] N. G. Kurakova, L. A. Tsvetkova, and Yu. V. Polyakova, Digital twins in surgery: achievements and limitations, *Khirurgiya. Zhurnal im. N.I. Pirogova*, no. 5, p. 97, 2022, doi: https://doi.org/10.17116/hirurgia202205197.

[12] S. D. Okegbile, J. Cai, C. Yi, and D. Niyato, Human Digital Twin for Personalized Healthcare: Vision, Architecture and Future Directions, *IEEE Network*, pp. 1–7, 2022, doi: https://doi.org/10.1109/mnet.118.2200071.

[13] M. Batty, Digital twins, *Environment and Planning B: Urban Analytics and City Science*, vol. 45, no. 5, pp. 817–820, Sep. 2018, doi: https://doi.org/10.1177/2399808318796416.

[14] E. Glaessgen and D. Stargel, The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles, *Semantic Scholar*, 2012. https://www.semanticscholar.org/paper/The-Digital-Twin-Paradigm-for-Future-NASA-and-U.S.-Glaessgen-Stargel/733d40c04be482d38dbd21f82b81e0fc890b6669

[15] J. V. S. do Amaral, C. H. dos Santos, J. A. B. Montevechi, and A. R. de Queiroz, Energy Digital Twin applications: A review, *Renewable and Sustainable Energy Reviews*, vol. 188, p. 113891, Dec. 2023, doi: https://doi.org/10.1016/j.rser.2023.113891.

[16] M. Cellina *et al.*, Digital Twins: The New Frontier for Personalized Medicine?, *Applied Sciences*, vol. 13, no. 13, 2023, doi: https://doi.org/10.3390/app13137940.

[17] R. Riascos, E. Ostrosi, J.-C. Sagot, and J. Stjepandić, Conceptual Approach for a Digital Twin of Medical Devices, *Advances in Transdisciplinary Engineering*, vol. 28, pp. 320–329, Oct. 2022, doi: https://doi.org/10.3233/atde220661.

[18] A. Khan, M. Milne-Ives, E. Meinert, G. E. Iyawa, R. B. Jones, and A. N. Josephraj, A Scoping Review of Digital Twins in the Context of the Covid-19 Pandemic, *Biomedical Engineering and Computational Biology*, vol. 13, p. 117959722211021, Jan. 2022, doi: https://doi.org/10.1177/11795972221102115.

[19] E. Katsoulakis *et al.*, Digital twins for health: a scoping review, *npj Digital Medicine*, vol. 7, no. 1, pp. 1–11, Mar. 2024, doi: https://doi.org/10.1038/s41746024-01073-0.

[20] S. Elkefi and O. Asan, Digital Twins for Managing Health Care Systems: Rapid Literature Review, *Journal of Medical Internet Research*, vol. 24, no. 8, p. e37641, Aug. 2022, doi: https://doi.org/10.2196/37641.

[21] A. L. Goldberger *et al.*, PhysioBank, PhysioToolkit, and PhysioNet, *Circulation*, vol. 101, no. 23, Jun. 2000, doi: https://doi.org/10.1161/01.cir.101.23.e215.

[22] H. Khan , L. Geys , P. Baneke , G. Comi, and L. Peeters, Patient-level dataset to study the effect of COVID-19 in people with Multiple Sclerosis, *PhysioNet*, 2024. https://doi.org/10.13026/77ta-1866

[23] A. Johnson , T. Pollard , O. Badawi , and J. Raffa, eICU Collaborative Research Database Demo (version 2.0.1), *PhysioNet*, 2021. https://doi.org/10.13026/4mxk-na84

[24] J. Areias Saraiva, M. Abreu, A. S. Carmo , H. Plácido da Silva, and A. Fred, ScientISST MOVE: Annotated Wearable Multimodal Biosignals recorded during Everyday Life Activities in Naturalistic Environments (version 1.0.1), *PhysioNet*, 2024. https://doi.org/10.13026/hyxq-r919

[25] Z. Johnson and M. J. Saikia, Digital Twins for Healthcare Using Wearables, *Bioengineering*, vol. 11, no. 6, p. 606, Jun. 2024, doi: https://doi.org/10.3390/bioengineering11060606.

[26] D. Mikolajewski, A. Bryniarska, P.M. Wilczek, M. Myslicka, A. Sudol, D. Tenczynski, M. Kostro, D. Rekawek, R. Tichy, R. Gasz, M. Pelc, J. Zygarlicki, M. Koziol, R. Martinek, R. Kahankova Vilimkova, D. Vilimek, and A. Kawala-Sterniuk, The Most

Current Solutions using Virtual-Reality-Based Methods in Cardiac Surgery - A Survey, *Computer Science*, 25, 1, 2024, doi: https://doi.org/10.7494/csci.2024.25.1.5633

[27] J. Streepey and S. Dent, Body Sway When Standing and Listening to Music Modified to Reinforce Virtual Reality Environment Motion (version 1.0.0), *PhysioNet*, 2021. https://doi.org/10.13026/x32c-cz47

[28] Classification of Physiotherapy Exercises Dataset, *www.kaggle.com.* https://www.kaggle.com/datasets/rabieelkharoua/classification-ofphysiotherapy-exercises-dataset

[29] A. Wijekoon, N. Wiratunga, and K. Cooper, MEx: Multi-modal Exercises Dataset for Human Activity Recognition, *arXiv (Cornell University)*, Jan. 2019, doi: https://doi.org/10.48550/arxiv.1908.08992.

[30] M. Tomaszewski, P. Michalski, J. Osuchowski, Evaluation of Power Insulator Detection Efficiency with the Use of Limited Training Dataset. *Appl. Sci.* 2020, 10, 2104, https://doi.org/10.3390/app10062104

[31] A. Schumann and K. Bär, Autonomic Aging: A dataset to quantify changes of cardiovascular autonomic function during healthy aging (version 1.0.0), *PhysioNet*, 2021. https://doi.org/10.130 Patient-level dataset to study the effect of COVID-19 in people with Multiple Sclerosis 26/2hsy-t491