Computers in Biology and Medicine 172 (2024) 108082

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A systematic review of the application of deep learning techniques in the physiotherapeutic therapy of musculoskeletal pathologies

Alberto Nogales [a], Manuel Rodríguez-Aragón [b], Álvaro J. García-Tejedor [a,*]

[a] CEIEC, Research Institute, Universidad Francisco de Vitoria, Ctra. M-515 Pozuelo-Majadahonda Km 1800, 28223, Pozuelo de Alarcón, Spain
[b] Rehabilitation and Technology Department, Adamo Robot SL. Miguel de Villanueva, 6, 26001, Logroño, Spain

A R T I C L E I N F O

Keywords:
Systematic review
Deep learning
Physiotherapy
Musculoskeletal pathologies

A B S T R A C T

Physiotherapy is a critical area of healthcare that involves the assessment and treatment of physical disabilities and injuries. The use of Artificial Intelligence (AI) in physiotherapy has gained significant attention due to its potential to enhance the accuracy and effectiveness of clinical decision-making and treatment outcomes. Nevertheless, it is still a rather innovative field of application of these techniques and there is a need to find what aspects are highly developed and what possible job niches can be exploited. This systematic review aims to evaluate the current state of research on the use of a particular AI called deep learning models in physiotherapy and identify the key trends, challenges, and opportunities in this field. The findings of this review, conducted following the PRISMA guidelines, provide valuable insights for researchers and clinicians. The most relevant databases included were PubMed, Web of Science, Scopus, Astrophysics Data System, and Central Citation Export. Inclusion and exclusion criteria were established to determine which items would be considered for further review. These criteria were used to screen the items during the first and second peer review processes. A set of quality criteria was developed to select the papers obtained after the second screening. Finally, of the 214 initial papers, 23 studies were selected. From our analysis of the selected articles, we can draw the following conclusions: Concerning deep learning models, innovation is primarily seen in the adoption of hybrid models, with convolutional models being extensively utilized. In terms of data, it's unsurprising that body signals and images are predominantly used. However, texts and structured data present promising avenues for ground-breaking work in the field. Additionally, medical tests that involve the collection of 3D images, Functional Movement Screening, or thermographies emerge as novel areas to explore applications within the scope of physiotherapy.

1. Introduction

The field of physiotherapy has proactively incorporated technological advances. The use of various measurement techniques to interpret the signals obtained from different treatments and interventions has become a standard methodology. This practice facilitates the continuous monitoring of patients, as well as the progression and regression of their pathologies. Consequently, a copious amount of data is generated that requires appropriate analytical methods for its interpretation and application, highlighting the imperative need to incorporate emerging technological advances, such as Artificial Intelligence (AI), into the field of physiotherapy. The application of AI offers the possibility of devising more effective and tailored treatment strategies. AI is a field of computer science that studies and interprets the mechanisms responsible for generating intelligent behaviours in humans. Its ultimate goal is to replicate these behaviours in machines, albeit not necessarily using the same mechanisms, [1]. Within AI, the field of machine learning (ML) currently offers the most promising results. ML aims to generalize behaviours by detecting patterns in the information provided through examples and experience [2], using a wide variety of methods and techniques among which Artificial Neural Networks (ANNs) stand out. ANNs are computational systems consisting of interconnected simple processing elements, known as neurons, which behave in a manner determined by the topology and weights of their connections, [3]. These models are adjusted by a process called training (supervised, unsupervised, semi-supervised, or reinforced) that needs large amounts of data, but remained not very efficient until more sophisticated models called deep neural networks were proposed.

* Corresponding author.
E-mail addresses: alberto.nogales@ceiec.es (A. Nogales), manuel.rodriguez@adamorobot.com (M. Rodríguez-Aragón), a.gtejedor@ceiec.es (Á.J. García-Tejedor).

https://doi.org/10.1016/j.compbiomed.2024.108082
Received 22 June 2023; Received in revised form 21 December 2023; Accepted 27 January 2024
Available online 29 January 2024

Deep learning (DL) models are ANNs that are hierarchical and multilayered. These models are capable of learning representations of data with increasing levels of abstraction, starting from the input data, [4]. The impact of DL models can be seen in the improvement that the application of these models has caused in tasks like image classification, and many other domains. Fig. 1 shows the relationship between AI and the different methods and techniques mentioned above.

DL, the latest breakthrough in AI, can provide more effective, personalized, and efficient care, leading to improved patient outcomes. It can analyze vast amounts of patient data and help physical therapists detect patterns and risk factors in injury development and progression. Deep neural models also can be used to plan personalized treatments for each patient, creating specific treatment plans that are tailored to their unique needs. DL models together with sensors and wearables can collect real-time data on patient movement and performance. This can help physiotherapists to measure a patient's progress during treatment and thus rethink and always adapt clinical decisions. Neural models optimize physiotherapists' processes, speed up their decision-making and reduce waiting times for patients. For example, chatbots can answer common patient questions or help them schedule appointments. In summary, DL models can be a very useful tool to improve the efficacy and personalization of physiotherapy treatments, which can positively impact patient recovery.

This wide range of possibilities is just beginning to be discovered and reflected in the scientific literature, which has not yet been systematized. Therefore, the main contribution of this paper is to compile scientific works published in the field of physiotherapy in which DL techniques have been applied. The compilation has been carried out by applying the methods of a systematic review.

There are other works such as [5–7] whose scope of review is the set of all AI techniques or others whose review is applied only in very specific areas such as [8] (distal radial fractures) and [9] (computer vision). During the process, we present a collection of statistical metrics along with graphical representations that aid in comprehending the information. This proves beneficial in several contexts. It enables the identification of the most employed DL models within physiotherapy and the metrics they attain based on various data types. Understanding the commonly used tests and their performance also reveals areas where there is potential for experimental exploration. For professionals in physiotherapy, gaining insights into the frequently studied physiological variables and specific body areas targeted is invaluable. Additionally, an evaluation of the prevalent intervention strategies has been conducted to provide a comprehensive overview of current practices. This paper is organized into the following sections: Section 2 summarizes the DL models used today and introduces the data types from the field of medicine. Section 3 details the methodology used to obtain the articles selected. Section 4 contains the in-depth study of the articles that were selected at the end, setting out their theoretical foundations, contributions, and applications. Finally, Section 5 presents the conclusions obtained from this research study.

2. State of art

Since the inception of AlexNet, the first big milestone in DL in 2012 [10], there has been a steady rise in the number of studies related to deep learning that have been published in medical bibliographic databases. As illustrated in Fig. 2, the annual count of deep learning publications on PubMed has nearly doubled each year since 2015, until 2021. Even then, each year the number of publications has been surpassed. In this, more than 5000 publications have been released and only a quarter of the year has passed.

DL techniques are based on multiple models and architectures, although the effectiveness of all of them is directly related to the nature and quality of the data with which they are trained. This section describes the data types that are commonly used in DL, as well as the architectures and models that best adapt to them.

Data in AI can be divided into the following groups depending on its nature; structured, images, texts and time series, Electronic Health Records (EHRs), and graphs. Structured data can be described as data that adheres to a standard schema or format, as defined by Ref. [11], typically presented in tabular forms, like in Comma Separated Value (CSV), Excel or database files. Structured data typically follows a row and column structure, where the columns consist of headers that define the attributes of the data, and the rows contain the corresponding values for each attribute, of a specific individual or item in question. Medical images are typically acquired from diagnostic tests such as ultrasound scans. Textual data encompasses written information used for patient monitoring, including medical records and reports. Time series data in physiotherapy commonly refer to signals such as electrocardiograms (ECGs) and electroencephalograms (EEGs) that record changes in electrical activity over time. According to Ref. [12], the information in question consists of a series of repeated observations taken at regular intervals over a substantial number of observations, focusing on a single unit or individual. The type of data is directly related to the DL model that will perform better. Following, a taxonomy of most of these models is described.

DL models can be trained in various ways, which are determined by the specific utilization of data and the particular problem they aim to solve. Besides, there is a recent trend in the development of hybrid models, which combine two or more architectures. It is considered an important field of research in deep learning for the near future, [13]. All the models that are described below fall into these categories.

The multilayer perceptron (MLP) represents the simplest form of a deep learning model, consisting of an input layer, multiple hidden layers, and an output layer.

Convolutional neural networks (CNNs) are widely used models in

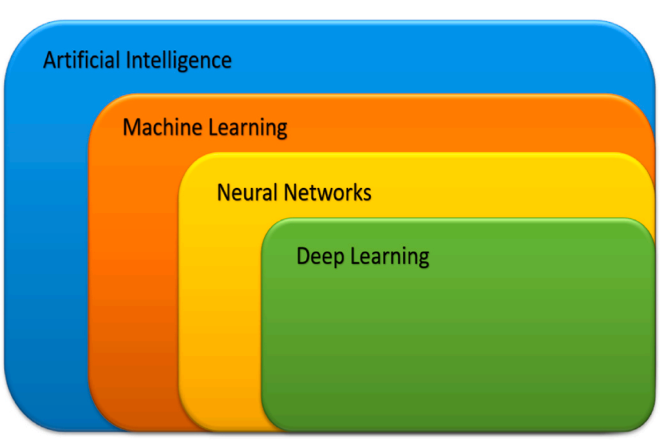

Fig. 1. Relationship between AI approaches.

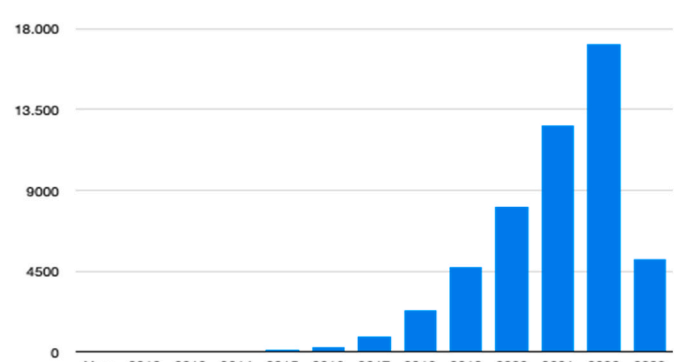

Fig. 2. Distribution by the publication year of the deep learning articles indexed in PubMed from 2012 to 2023 (n = 46,230).

computer vision with various applications. Their key ability is to detect spatially invariant patterns, allowing them to learn specific patterns in one image that may appear in different locations in other images. More recently, a specific type of CNN called Graph Convolutional Neural Networks (GCNN) has emerged, that can handle graphs. In their 2016 paper [14], introduced GCNN as a method that encodes both the structure of a graph and its nodes' features using a specialized type of CNN.

Recurrent neural networks (RNNs) were first introduced by Ref. [15] as models that use input vectors of variable length and apply a transition function recursively to their internal hidden state vector, ht. RNNs are well-suited for analyzing time-series data, such as EEGs. Among RNNs, a specific type called Long Short-Term Memory (LSTMs) or Gated Recurrent Units (GRUs) has gained popularity. LSTMs were proposed by Ref. [16] to address the challenge of working with noisy or ambiguous input data without losing important information.

Deep autoencoders (DAEs) employ unsupervised learning and were originally defined by Ref. [17]. A distinguishing feature of DAEs is that both the input and output layers have the same, and the model consists of two processing structures. The first is the encoder, which takes the input data and reduces its size to a smaller representation containing its main characteristics. The second part is the decoder, which aims to reconstruct the original input by upsampling the reduced representation until it reaches the original input size. Restricted Boltzmann Machines (RBMs) were introduced by Ref. [18] that can learn a probability distribution. In deep learning, RBMs were used to implement Deep Boltzmann Machines (DBMs) by [19].

Generative adversarial networks (GANs) belong to the generative class of models. GANs consist of two neural models, the generator and the discriminator, which work together in an adversarial training process, as introduced by Ref. [20]. The architecture of GANs aims to learn and imitate a given data distribution. The generator is responsible for producing synthetic instances of the input data, while the discriminator evaluates these instances and decides whether they are similar enough to the input data or not. The discriminator assigns a probability of authenticity to each instance, indicating whether it is from the input distribution or generated by the generator. Through repeated iterations of this process, the generator learns to create synthetic data that better resembles the input distribution.

3. Methodology

The process employed to identify research works that pertain to a specific field or address particular research questions is referred to as a Systematic Literature Review (SLR), also known as a systematic review. Numerous protocols are available for conducting an SLR. A good example is the methodology outlined by Keele et al. [21], encompassing a series of structured steps to ensure a comprehensive and unbiased review. These steps initiate with the assessment of the necessity for the review, followed by the definition of precise research questions to guide the search. The development of a detailed protocol precedes the identification of pertinent research papers, ensuring a systematic approach to data collection and analysis. Criteria for inclusion and exclusion are defined to streamline the selection of relevant studies. Each selected paper is then subjected to an in-depth analysis of its key features, culminating in a well-organized presentation of the review findings in a scholarly format. This structured approach ensures that the SLR is conducted with rigour, transparency, and replicability, contributing to the enhancement of the existing body of knowledge. During the process of choosing these papers, we apply the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), [22]. Next, we analyze both the datasets and the papers in which they are utilized. Finally, we present a comprehensive account of all this information in our paper. This process is depicted in Fig. 3.

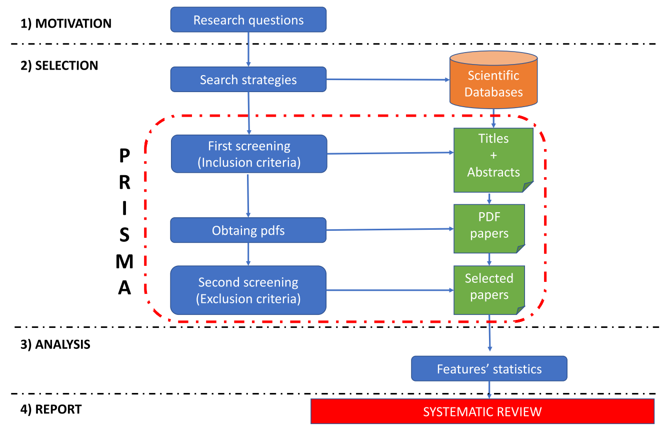

Fig. 3. Workflow diagram to achieve the systematic review.

3.1. Research questions

Since the primary objective of this systematic review is to assemble scientific papers in physiotherapy that have utilized DL models, we can extract useful insights by analyzing the characteristics of these works. Accordingly, the following research questions, compiled in Fig. 4, are formulated to provide a clear understanding of the review's purpose and usefulness.

3.2. Search strategies

Different search strategies have been considered using keywords obtained in the case of physiotherapy from the Medical Subject Headings (MeSH[1]) thesaurus. The searches follow the schema of combining a term from the deep learning field with a term of physiotherapy. As there are different interesting terms in both fields these are combined by using the logic operator OR. The search strategies have the following schema described in Fig. 5.

The search and collection of papers include everything published until February 8, 2023. The following sources were used to make the searches: Scopus,[2] PubMed,[3] Astrophysics Data System,[4] Web of Science (WOS),[5] and Central[6] Citation Export.

3.3. PRISMA process

As a way to obtain a set of high-quality papers a set of criteria has been established by a multidisciplinary team of researchers composed of computer scientists, biomedicine engineers and physiotherapists. During this stage, we have applied the PRISMA guidelines as described following.

For the first screening, after the searches, all the titles and abstracts of the papers were read by two researchers with opposite profiles (biomedical engineer and physiotherapist) to check if they accomplished the inclusion criteria of being a paper that "applies a deep learning model to solve a use case in physiotherapy". In case of disagreement, a third researcher with a background in computer science took the final decision to decide if the paper should pass to the next stage. During the application of these criteria, we are covering the perspective of both fields: computer science and physiotherapy. In fact, biomedical engineers not only know computer science but also can

[1] https://www.ncbi.nlm.nih.gov/mesh/.
[2] https://www.scopus.com/.
[3] https://pubmed.ncbi.nlm.nih.gov/.
[4] https://ui.adsabs.harvard.edu/.
[5] https://www.webofscience.com/.
[6] https://www.cochranelibrary.com/central/citation-export.

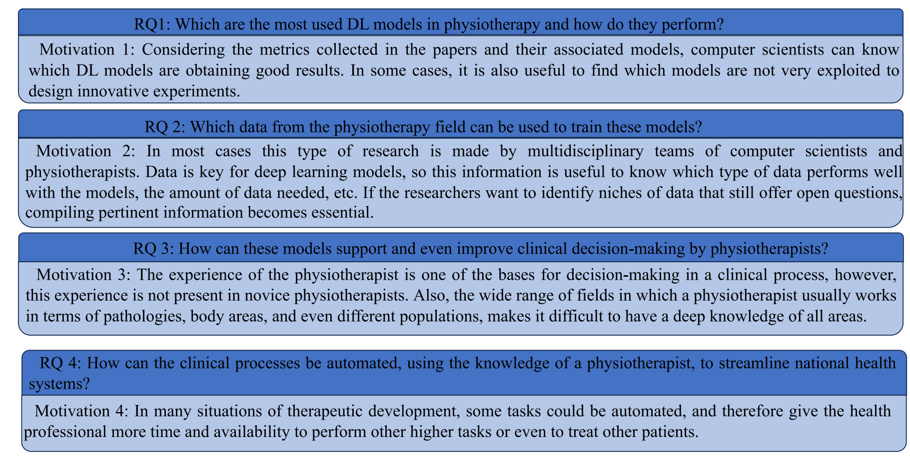

Fig. 4. Research questions.

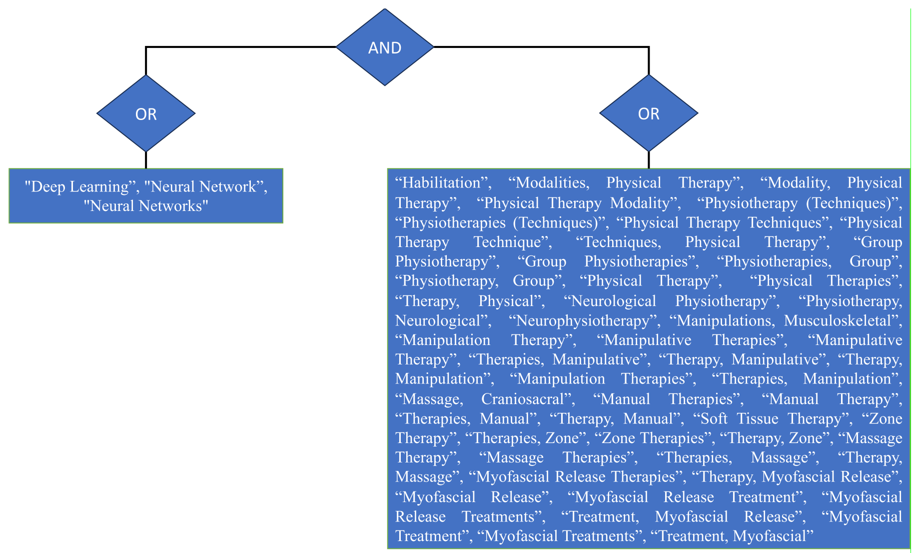

Fig. 5. Research strategies.

understand some terms of the physiotherapy field. As the application of deep learning is a complex field, the third reviewer is a researcher with extensive knowledge of the field that has been reflected in the publication of more than 10 papers. We considered that at this point if the title and abstract do not reflect that the research is applying deep learning models in the field of physiotherapy is because this is not important enough to be considered a relevant paper.

After discarding repeated items, conference papers surveys, or arxiv papers, the final selection of works has been made to apply more restrictive criteria. As the methodology to obtain the set of high-quality papers cannot be done automatically, a set of exclusion criteria in the field of computer science and physiotherapy has been established so that the papers will not be discarded in this second screening, all the following criteria must be fulfilled.

1. The description of the deep learning models is sufficiently detailed so readers can replicate them. Deep learning models are composed of multiple hyperparameters as the number of layers, number of neurons in each layer, activation functions, optimizer, etc. A paper that does not include this information could not be considered a high-quality paper as this information is one of the most important in this type of research.
2. During the training stage, data has been separated into training a test set which is a good practice to measure the performance of the model in terms of under and overtraining. When a deep learning model is trained, three situations could happen: the model is undertrained, overtrained or performs accurately. The way to measure that is by using the bias-variance trade that needs to compare accuracy at training and test stages. This task needs to split the initial dataset normally at rates of 80 and 20 %. Thus, providing information regarding the training status of a model is mandatory in a research paper to ensure the credibility and reliability of the presented findings and analyses.
3. The final performance of the models is compiled by using metrics such as accuracy, loss, sensitivity, etc. This point is linked with the previous one, as a consequence of applying a training-test pipeline we obtained the performance of the model which will tell us if the method is improving human performance or previous baselines doing the same task. In this way, different metrics correspond to classification or regression problems. Also, some metrics let us understand if the model is having problems with false positives or false negatives which is meaningful information in cases that involve human health.

4. Non-human studies. Physiotherapy primarily addresses the needs of human patients, ensuring that the results and conclusions are directly applicable to physiotherapy in a healthcare context.
5. Studies that do not focus on functionality/disability and/or pain. By including studies containing these variables, the aim is to strengthen direct application to the clinical domain, enhance individuals' interaction with their environment, alleviate their pain, and improve their daily task performance.

A flow diagram depicting the PRISMA can be found in Fig. 6. The diagram provides a summary of the paper selection process, which resulted in the datasets reported in this study.

Following the consultations made on the March 22, 2023 (Scopus 52, PubMed 76, Astrophysics Data System 22, WOS 61, CENTRAL 3), a total of 214 papers were acquired, of which 136 were used to begin the first screening. In this phase, 86 papers were eliminated because they did not satisfy the criterion of "being a deep learning study developed on physiotherapy," We were unable to acquire complete texts for 4 of the 47 papers in the second screening, so 43 were analyzed, of which 20 studies met the exclusion criteria for review. So, a total of 23 papers formed the final selection.

3.4. Data extraction, classification of studies, and statistical analysis

Once there is a set of selected papers some important features are obtained to analyze the development of deep learning models in the field of physiotherapy. The study focuses on the year of publication, deep learning models used, performance measurement metrics, type of data, performed test, type of physical intervention performed, physiological variable measured, and area of the body assessed.

4. Results

Statistics were gathered based on the previously mentioned characteristics of the papers. This information was compiled by developing Python scripts and using Matplotlib, a library to generate a set of charts [23]. Additionally, we created a VOSviewer[7] (Visualization Of Similarities) using the keywords from the selected papers to confirm their relationship to the keywords used in the original searches. Fig. 7 displays this VOSviewer with the aforementioned information.

As seen in the Figure above, most of the keywords demonstrate that the selected papers are related to the search strategies. We can see "deep learning" which is one of the main characteristics of the papers we want to compile in this review. Other words like "machine learning", "cnn" and "deep neural network" also belong to this field. There is another set of words related to physiotherapy: "exergaming", "hand exoskeleton design" or "balance evaluation".

4.1. Summary of papers

The selected papers have been compiled in Table 1 including some important characteristics such as the deep learning model, which metrics have been used and its values, type of data, tests used to compile the data, type of physical intervention performed, physiological variable measured, and area of the body assessed and year of publication.

4.2. Statistics and analysis of the included studies

The first relevant information is the distribution of articles by year, shown in Fig. 8.

It should be noted that the data for 2023 correspond only to the first quarter (January–March) when the publication databases were consulted.

Another element to consider is the number of participants in each study. Most of the articles feature participant numbers ranging from 5 to 48. However, there are outliers; one study includes a notably larger sample of 1424 subjects, while two studies are characterized by a single participant each. In these instances, with only one participant, the data collected comprises a series of exercises executed by that individual. Excluding the outliers (1 and 1424), the average participant count across the studies is 27.95, with a standard deviation of 19.64, indicating a moderate variation in sample sizes. These values can be used to decide the number of subjects to be recruited when starting new research.

One of the most interesting things, we can study in this review is which DL model has been applied in the different studies. For this purpose, we have made a pie chart showing the percentages of use, Fig. 9.

DL models are directly related to the type of data used for training them. As we stated above there are 4 main types of data: structured, signals, images, and texts. In the following Fig. 10, a pie chart describes how the 4 types of data are distributed in percentages.

In the field of physiotherapy, we find another feature related to the type of data which is the test used to compile the information to analyze. This information can be used by both computer scientists and physiotherapists to find niches where specialized data is not widely used. Fig. 11 shows a pie chart showing the distribution of different ways to compile the data in physiotherapy.

After training a model and depending on the data, we can measure their performance using different metrics. We have classification problems that mostly use metrics like accuracy, precision, sensitivity, specificity, and f1-score or regression problems that use error metrics like Mean Square Error (MSE), Percentage Root Mean Square Error (PRMSE), or Mean Error (ME). In the following 4 Figs. 12, 13, 14 and 15, we described how many times these metrics have been used in the different papers (bar diagrams) and the distribution of the values, these metrics obtain in the different models used in the papers (boxplot).

Other metrics have been used only a few times or are not very well-known, so they have not been included in the graphs. Table 2 compiles this information related to the Area Under the Receiver Operating Characteristic (AUROC), Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), Cumulative Match Curve (CMC), and coefficient of determination R^2.

Within the realm of physiotherapy, several interesting characteristics are noteworthy. Initially, our analysis focused on identifying the primary physiological variable examined in the literature. Physiotherapy encompasses the management of a myriad of pathologies, where assessment and follow-up are occasionally conducted through subjective tests. The incorporation of deep learning tools facilitates a transition to a more objective approach to patient assessment and monitoring. Consequently, it becomes essential to identify the prevalent physiological variables and explore the available data collection systems. Fig. 16 presents a bar diagram, compiling and illustrating the distribution of this information across various studies.

Another noteworthy aspect is the examination of body parts in the various studies. In the same way that it is of interest to know the most collected physiological values, and taking into account the different specialities that physiotherapy covers, knowing which parts of the body are the most focused on will allow new proposals, fields of study and trends to be established to broaden and improve the objectives of the research. In Fig. 17 we described how this characteristic is distributed in the papers using a bar diagram.

The intervention performed describes the therapeutic approaches utilized for treating various pathologies. This is of interest to support, speed up, and in some cases improve the clinical decision of the therapist and optimize patient follow-up. Fig. 18 shows a pie chart with the percentage of interventions performed regarding the total of selected papers.

Another way to measure the impact of deep learning in physiotherapy is by obtaining graphs that combine features from both fields. Following, we show some heatmaps describing different use cases of this

[7] https://www.vosviewer.com/.

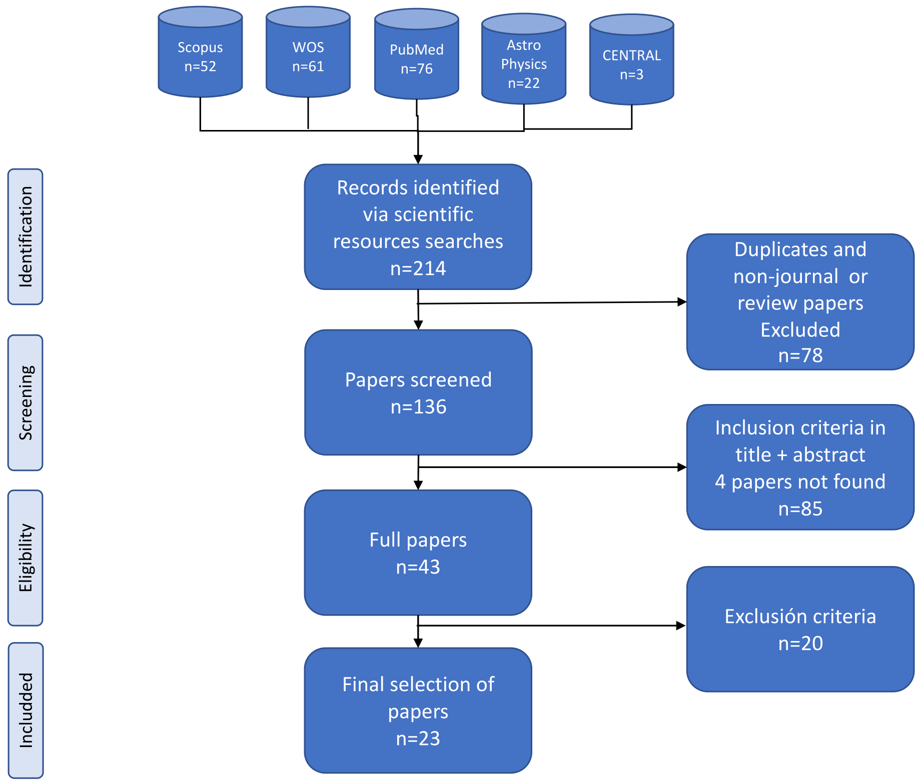

Fig. 6. PRISMA diagram of the systematic review.

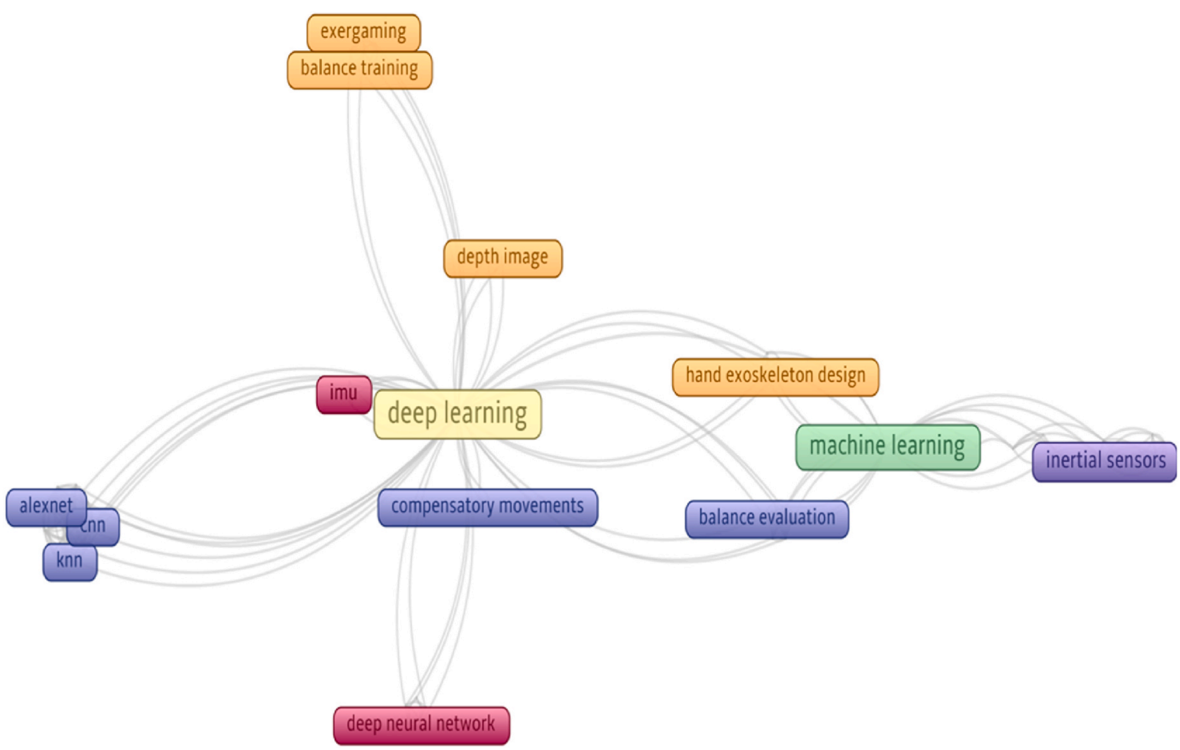

Fig. 7. VOS viewer showing the keywords used in the papers.

type. These graphs are made by visualizing 2D data as a matrix of coloured cells, with each colour representing the magnitude of the corresponding data value. Data is prepared in a table-like format where the value of each cell represents the number of instances that happen when crossing the values of each feature represented in the axes.

In Fig. 19, we show a heatmap that faces the deep learning models that have been used with the different tests to compile the data. We can see a non-mentioned model such as Convolutional Recurrent Neural Network (CRNN). This is very useful for computer scientists to know which deep learning models perform better with the different ways of

Table 1
Summary of the main papers reviewed.

Paper	Subjects	DL model	Performance metrics	Medical test	Data type	Physiological feature	Body parts	Intervention performed	Exercise	Year published
[24]	1424	CNN	100 %, 99.78 % Precision	Plantar pressure	Image	Foot progression angle	Foot	No intervention	Not mentioned	2022
[25]	1	MPL + VSM	95.58 % Accuracy	EEG	Signal	Movement	Hands	MI	Not mentioned	2021
[26]	11	CNN	66.36 % Accuracy	EEG	Signal	Movement	Hands	MI	Hand grip	2021
[27]	20	CNN + LSTM	91.3 % Accuracy, 90.57 % Precision, 90.42 % Recall, 91.00 % f1-score	Videos	Image	Movement	Neck, trunk, upper limb, lower limb	Exercise	Does not specify which exercises were used with patients.	2022
[28]	20	CNN + LSTM	77.75 % f1-score	FMN	Image	Movement	Trunk, upper limb, lower limb	Exercise	Deep Squat (DS), Hurdle Step (HS), Inline Lunge (IL), and Trunk Stability Pushup (TSP)	2022
[29]	15	CNN	84.2 % Accuracy, 83.8 AUROC	EEG	Signal	Robotic movement	Upper limb	Exercise	Bobath's upper-limb rehabilitation exercise	2021
[30]	32	CNN + LSTM	97.1 %, 99.9 % Recall, 94.3 %, 99.8 % Precision, 98 %m 99.8 % f1-score	Plantar pressure	Signal	Falls and physical activity	Foot	Falls and physical activities simulation	Simulated falls during stand-to-sit, walking (backward, forward, lateral)/Physical activities: jump, walk (in place, backward, up and downstairs), jogging, sitting, body swing	2023
(W.-H [31])	30	CNN	90.2 %, 90.7 %, 84.8 %, 82.4 % Accuracy	Sensors	Signal	Walking patterns	Ankle, foot	Walk	Level walking (LW), descent (DC), and ascent (AC) slope walking, as well as downstairs (DS) and upstairs (US) walking	2020
[32]	10	CNN	2.3, 5.6, 3.4, 6.5, 1.8, 4.7 RMSE	Sensors	Signal	Joint angles	Lower limb	Running	Running at five different speeds	2020
[33]	78	GCNN	0.971, 1.855, 0.621 MAD, 1.993, 3.822, 1.180 MA PE, 3.081, 6.810, 1.591 RMSE	Videos	Image	Movement	Full body joints	Exercise	Describes only squats and arm lifts out of 10 exercises	2022
[34]	5	CNN	99.1 % Accuracy	Sensors	Signal	Movement	Hands	Hand actions	Hand flexion-extension and grip	2022
[35]	25	CNN + LSTM, CNN, CNN + GRU	98 %, 99 %, 100 % Accuracy	Videos + Sensors	Image	Movement	Upper limb	Exercise	(1) Left flexion of shoulder (2) Left abduction of shoulder (3) Left elbow flexion (4) Left median rotation of shoulder (5) Left internal shoulder rotation (6) Right flexion of shoulder (7) Right abduction of shoulder (8) Right elbow flexion (9) Right median rotation of shoulder (10) Right internal shoulder rotation	2023
(K.-Y [36])	1	LSTM, MLP, CNN	0.22, 0.29, 0.33 RMSE, 2.3, 3.5, 3.0 PRMSE, 99.5 %, 99.4 %, 94.3 % CMC	3D images	Image + Signal	Movement	Upper limb	Exercise	Arm reaching	2022
[37]	71	CNN, LSTM	89 % Accuracy	EEG	Signal	Pain, Movement	Trunk	Massage, MI and motion execution	Trunk left bending or trunk right bending	2021
[38]	20	RNN	80 % f1-score	Videos	Image	Hand function	Hands	Hand activities	Hand–object interactions	2022
[39]	12	LSTM	10.7 RMSE, 0.97, 0.77 R^2	Sensors	Image	Balance/ground reaction forces	Full body	Exercise	Exergame for balance training	2022
[40]	21	CNN	99.5 %, 96.3 % Accuracy	Videos	Image	Movement	Low back, shoulders	Exercise	McKenzie therapy, and exercises for shoulder motor control and strength	2022
[41]	30	CNN	97 % Precision, 93 % Recall	Thermography	Image	Movement	Hands	Exercise	Arm lifting and hand exercises	2020
[42]	61	CNN	96.83 %, 96.73 % Precision, 97.73 %, 97.43 % Recall, 92.27 %, 97.04 % f1-score	Sensors	Signal	Exercise recognition	Upper body, lower body	Exercise	Bicep Curls (BC), Frontal Raise (FR), Lateral Raise (LR), Triceps Extension Right arm (TER), Pec Dec (PD), and Trunk Twist (TT), along with four lower-body exercises: squats (SQ), lunges alternating sides (L), leg lateral raise (LLR), and standing bicycle crunches (SBC)	2020
[43]	20	CRNN	00.4 %, 88.9 % Accuracy	Sensors	Signal	Adherence to physiotherapy program	Shoulders	Exercise	Pendulum (PEN), abduction (ABD), forward elevation (FEL), internal rotation (IR), external rotation (ER),	2018

(continued on next page)

Table 1 (*continued*)

Paper	Subjects	DL model	Performance metrics	Medical test	Data type	Physiological feature	Body parts	Intervention performed	Exercise	Year published
									trapezius extension (TRAP), and upright row (ROW)	
[44]	22	CNN	11.96 % Mean error	Videos	Image	Squat angle	Knee	Exercise	Squats	2019
[45]	41	CNN	87.7 % Sensitivity, 94.2 % Specificity	Videos + Sensors	Image	Balance	Full body	Exercise	Pose and balance	2020
[46]	15	MLP, CNN, CNN + LSTM	98.58 %, 99.58 %, 99.38 % Accuracy	Videos	Image	Pose correction	Full body	Exercise	Yoga	2022

compiling the information.

Similar to the diagram above, we have obtained another heatmap that combines the physiological variable with the body parts studied in the papers. Just as there are different types of specialities, physiotherapy also deals with the whole musculoskeletal system, i.e., upper limbs, lower limbs, trunk, etc. The treatment and analysis of some of the body areas are more prevalent depending on the pathology and functionality required. This information is of interest for physiotherapists for two reasons: to have the best starting point of knowledge if they work with the most studied areas and to detect body areas and pathologies that are yet to be exploited in research.

Finally, we have created a heatmap to describe the relationship between the evaluated body parts and the intervention performed. In this case, a heatmap graph can be seen in Fig. 21. This last vision allows us to set the basis of what are the results of the interventions that have been studied with DL to complete the clinical process and select the best treatments according to the area of the body studied.

5. Discussion

Next, we present the conclusions derived from the study findings, aiming for a comprehensive analysis that integrates annual publication trends, deep learning model preferences, data types utilized, and performance metrics, among others.

5.1. Annual publication trends

Our analysis, as illustrated in Fig. 8, underscores a consistent rise in publications from 2018 to 2023, highlighting the growing interest in combining deep learning with physiotherapy. Since 2018 is relatively recent, even the earliest publication can still be considered current. Notably, a pronounced increase in publications was observed in 2022, and the trend appears to be continuing into the current year (2023) since having completed just one quarter, two publications have already been identified. This indicates the possibility that the total count may surpass that of the previous year. These findings highlight the influence of deep learning models in the field of physiotherapy and emphasize the ample scope for further research in this area.

5.2. Predominance of CNN architectures

Fig. 9 attests to the predominance of CNNs, which are employed in half of the reviewed works, suggesting their effectiveness and adaptability in this interdisciplinary domain. Given the widespread development and proven effectiveness of CNNs within DL paradigms, this prevalence is not unexpected. It also prompts an exploration into whether this preference for CNN models correlate with the most frequently used data types.

5.3. Hybrid models and data types

An emerging trend in the reviewed studies is the adoption of hybrid models, specifically three combinations: CNN + LSTM (Convolutional Neural Network + Long Short-Term Memory), CNN + GRU (Convolutional Neural Network + Gated Recurrent Unit), and MLP + SVM (Multi-Layer Perceptron + Support Vector Machine). Among these, the CNN + LSTM hybrid stands out, ranking second in terms of use and featured in 17 % of the surveyed works. This indicates an exciting potential for the innovative application of such hybrid models in physiotherapy research. As we speculated earlier, an alignment between the choice of DL models and the data types is observed, with images and signals being predominant, as depicted in Fig. 10. Most of the data used falls under two categories - images (52 %) and signals (44 %). This distribution aligns logically with the most prevalent DL models: CNNs, CNN + LSTM, and LSTM. CNNs are flexible and can process both signals and images depending on whether 1D or 2D convolutional filters are applied. An additional point of interest is the simultaneous use of both image and signal data within a single study, suggesting a useful approach for future research in the field. Conversely, text and structured data have yet to find their place in this context, pointing to an untapped potential or

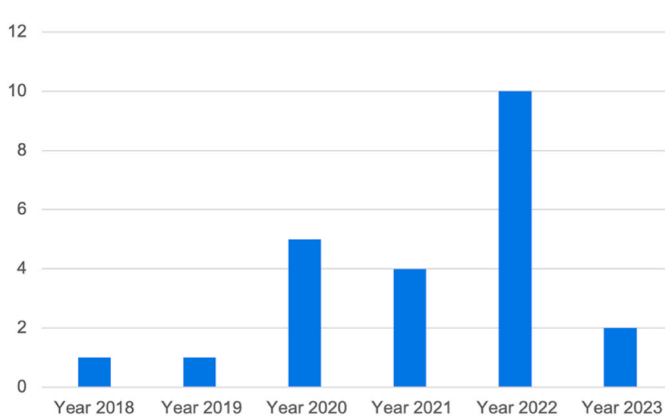

Fig. 8. Distribution by the publication year of the papers obtained without duplicates (n = 23).

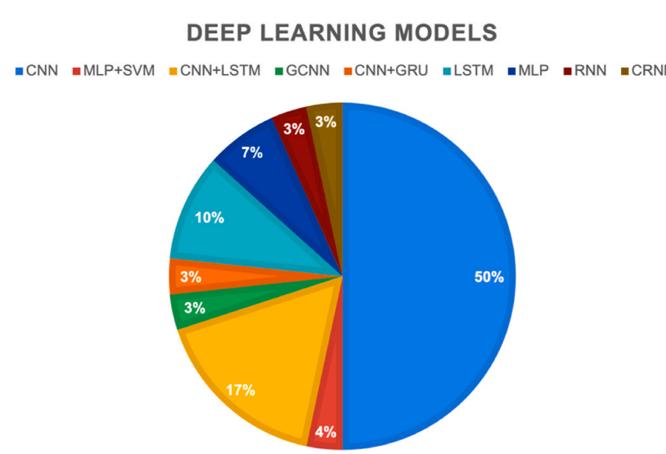

Fig. 9. Percentage of used deep learning models.

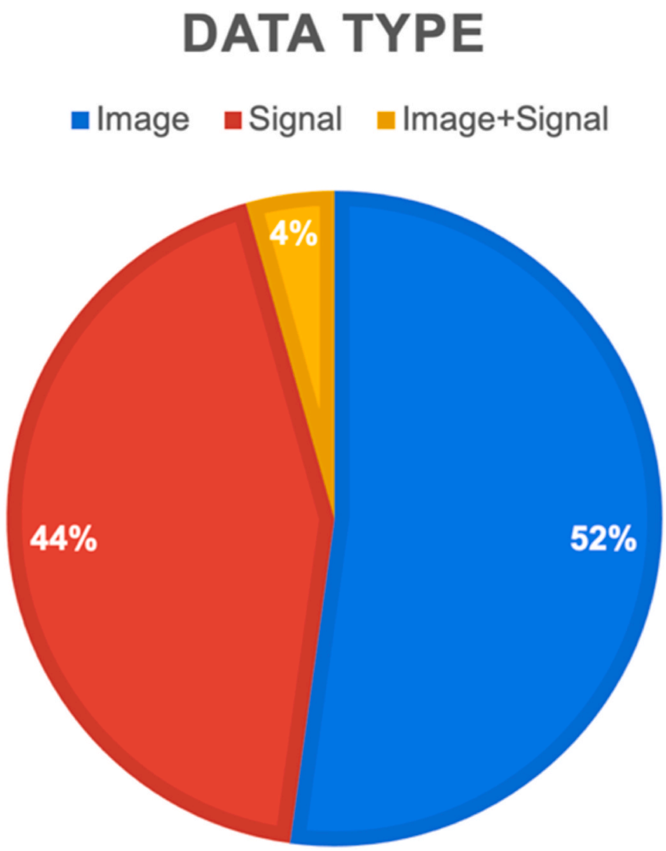

Fig. 10. Percentage of used types of data.

niche within physiotherapy.

5.4. Data sourcing and emerging techniques

The data, predominantly sourced from sensors and video recordings capturing movement (such as egocentric videos or posture analysis), aligns with the model preferences, as indicated in Fig. 11. EEG data also rank high in usage. Notably, a small number of studies have begun exploring innovative data collection methods such as thermographic imaging, 3D imaging, and Functional Movement Screening (FMS). Thus, it's reasonable to suggest that these nascent methods represent potential avenues for future deep-learning applications in physiotherapy.

5.5. Performance metrics

The choice of the DL model and data type can significantly influence performance metrics. From our analysis, we discerned four distinct categories, segmented into classification and error metrics. Within classification metrics, accuracy is the most prevalent, nearly twice as common as the second-most used metric. However, it performs rather modestly in terms of average output, not significantly outperforming other metrics. Sensitivity and specificity demonstrate consistent results, while the F1-score fluctuates noticeably. Precision also deserves mention due to two outlier instances where model performance was suboptimal. Although accuracy is a suitable initial performance indicator, it is not comprehensive as it fails to consider false positives and negatives. Lastly, our observations from Figs. 14 and 15 provide insights into error metrics. Root Mean Squared Error (RMSE) is the most frequently utilized, its values spanning from near zero to ten. PRMSE shows stable results but relies only on three experimental bases. These findings call for expanded use of diverse metrics to ensure comprehensive model performance evaluation.

5.6. Focus on movement

Fig. 16 offers a summary of the physiological features explored across the reviewed papers. The predominant focus is clearly on the study of movement, overshadowing other features significantly. This dominance suggests that while movement remains a crucial aspect, there is a multitude of other features that remain largely unexplored, thereby presenting potential opportunities for innovation. The prominence of movement-related features may be attributed to most of the studies focusing on populations with movement disorders, such as neurological patients. Turning our attention to Fig. 17, we observe a range of body parts evaluated in the included studies. Given the demographic targeted in these studies, it is particularly notable that the upper limbs and hands emerge as the most frequently analyzed regions for the application of DL. This focus reflects the nature of common physiotherapeutic interventions, often concentrated on these body parts.

The intent behind physiological assessment is to propose relevant interventions, as evidenced in Fig. 18. Here, we find that most interventions are related to exercise, which directly addresses issues of movement. This aspect is further reinforced by the data in Fig. 20, which highlights movement-related physiological variables and exercise-based interventions (as outlined in Fig. 21) as the most popular research targets. Given the frequent assessment of movement in the upper limb and hand, it may seem intuitive to propose an exercise intervention corresponding to these areas. Interestingly, exercise interventions have been applied more frequently to the lower limbs and lower back. This disparity suggests that other methods, such as motor imagery, may be more prevalent for the upper limb and hand, as depicted in Fig. 21. Regarding the explored exercises in the studies. There is a wide range of them such as deep squats, hurdle steps, inline lunges, and trunk stability pushups, which are traditionally used in physical therapy. However, a notable observation is that a significant portion of the studies did not provide detailed descriptions of the exercises performed. This lack of exercise documentation raises concerns about the reproducibility of the interventions and hinders the ability to compare and analyze the effectiveness of deep learning models in the context of physiotherapy. Furthermore, a substantial number of these studies conducted experiments on healthy individuals rather than patients with musculoskeletal conditions or those in need of rehabilitation.

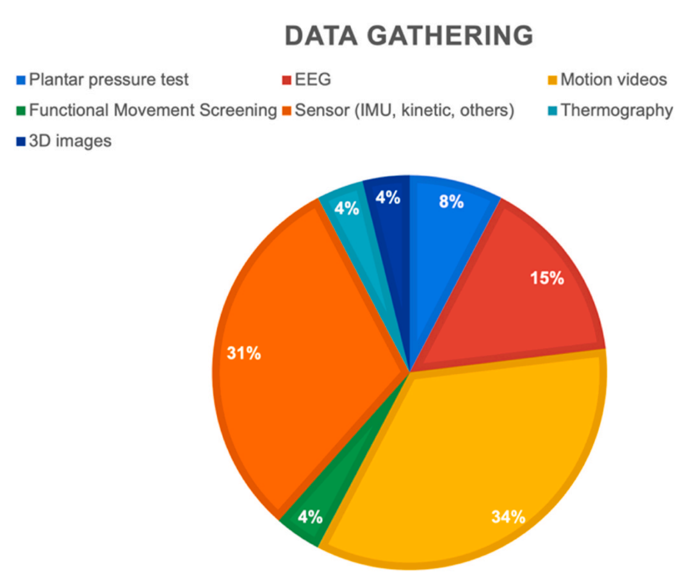

Fig. 11. Percentage of methods for data gathering.

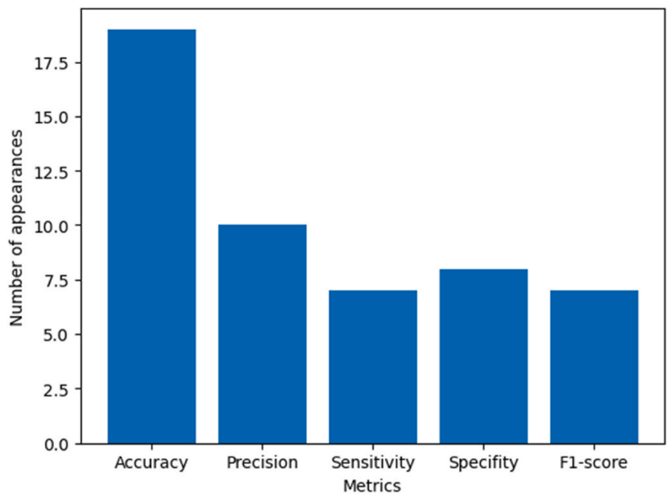

Fig. 12. Diagram bar for classification metrics.

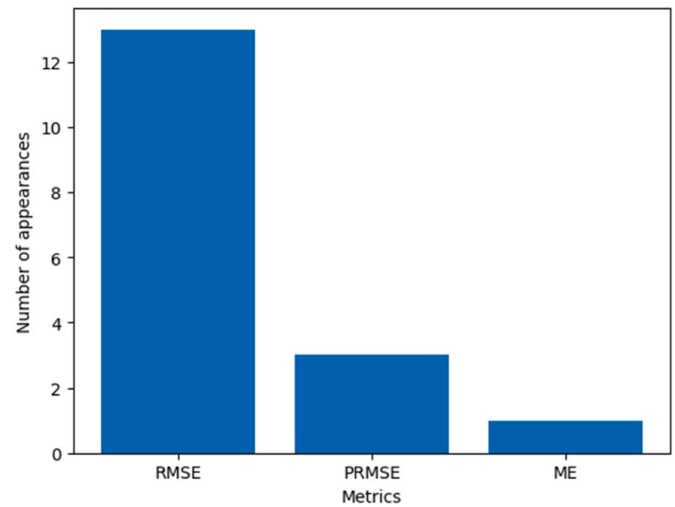

Fig. 14. Diagram bar for error metrics.

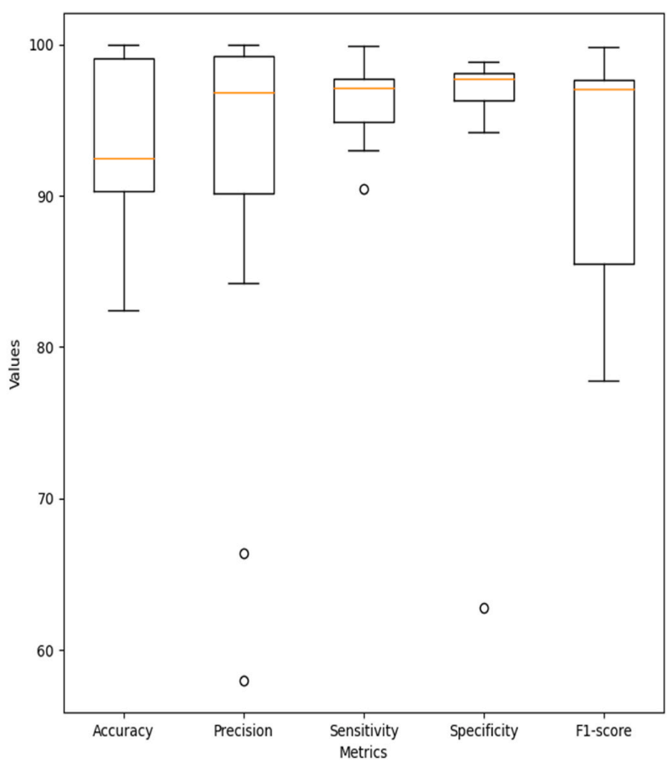

Fig. 13. Boxplot for the values of classification metrics.

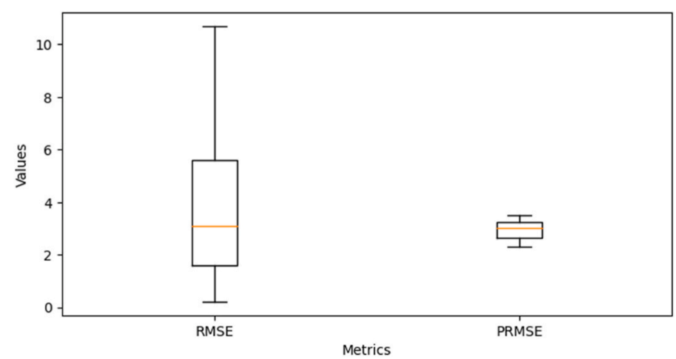

Fig. 15. Boxplot for error metrics.

Table 2
Summary of the used metrics.

Metric	Values
AUROC	83.9 %
MAD	0.971, 1.855, and 0.621
MAPE	1.993, 3.822, and 1.180
CMC	0.995, 0.994, and 0.943
R^2	0.97 and 0.77

5.7. *Diagnostic and treatment enhancement*

The two primary areas for conducting rehabilitation are patient diagnosis/tracking and treatment decision-making and execution. Concerning the physiotherapeutic diagnostic and tracking aspect, the study underscores a palpable gap in integrating DL with diagnostic tools like ultrasound, algometry and thermography. The potential for DL to refine diagnoses, personalize treatments, and enhance the precision of physiotherapy is emphasized. These tools aim not only to assess the range of motion but also to determine the state of the involved tissue during these movements or to evaluate pain and the patient's perception of such movements. DL could have the capacity to support these tools in personalizing and refining diagnostic conclusions, improving reevaluation, and even providing prognosis estimates that would enable physiotherapists to adapt each clinical decision to the patient's current condition. It would be of great interest to have DL-supported decision-making that suggests the best type of intervention for each case, its intensity, and even its duration and repetition, all to achieve precision physiotherapy.

5.8. *Expansion to diverse pathologies*

A call for the diversification of DL applications beyond neurological issues to musculoskeletal pathologies and ageing-associated challenges is made. There is a need for a holistic approach to expanding knowledge and application to musculoskeletal pathologies, which are prevalent both globally and socially, such as back pain, joint injuries, ligamentous or tendinous issues, among others. This expansion should not be limited to prevalent conditions but should also encompass ageing populations and their associated challenges in Western society, which pose a significant challenge to national healthcare systems.

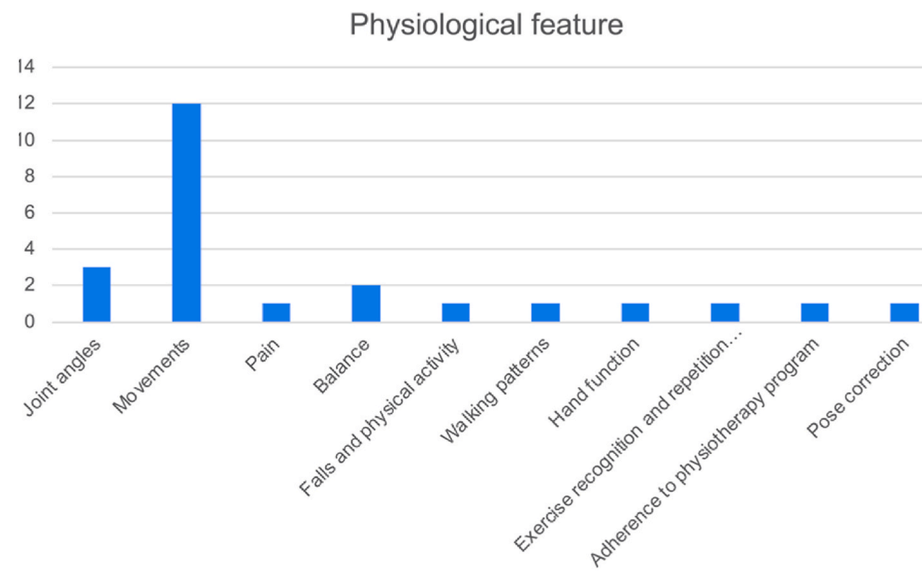

Fig. 16. Diagram of bars for the physiological features.

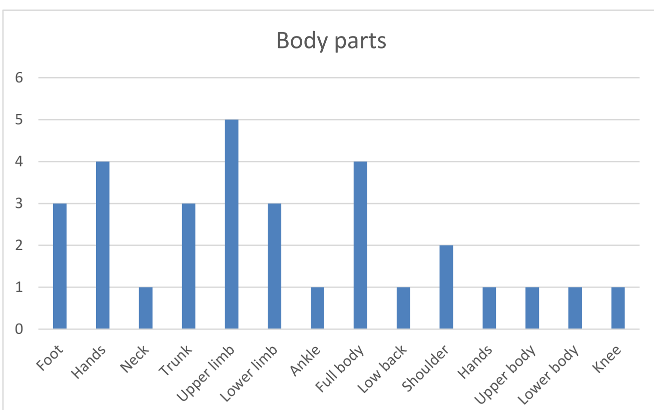

Fig. 17. Diagram of bars for the studied body parts.

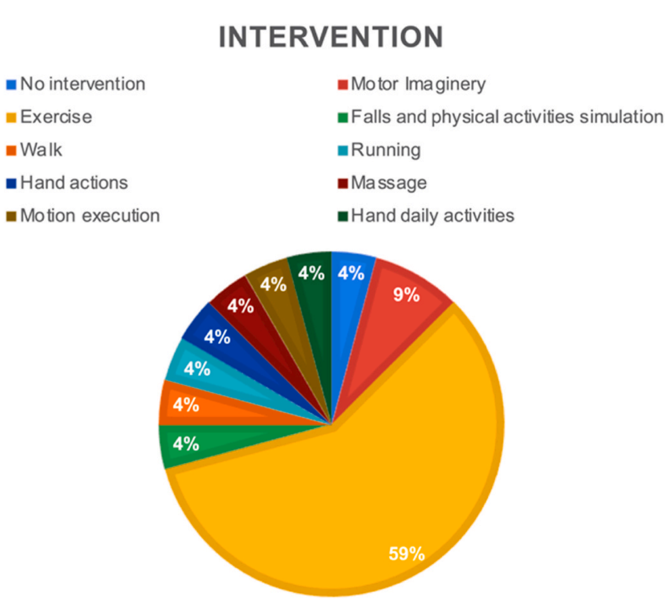

Fig. 18. Pie chart of interventions performed.

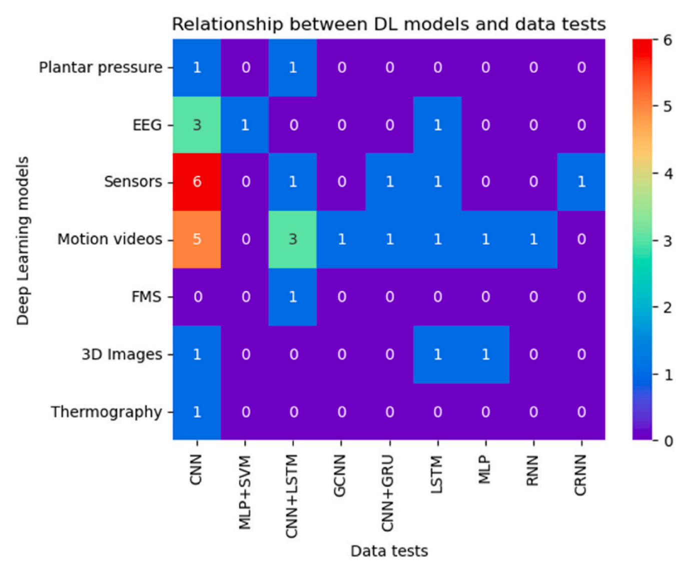

Fig. 19. Heatmap: deep learning models vs. data tests.

5.9. Evidence-based physiotherapy

The importance of aligning technological advancements with evidence-based physiotherapy, incorporating professional experiences and patient preferences, is underscored. In this context, not only does knowledge generated by scientific studies apply, but the experiences of healthcare professionals who must implement it and the preferences of patients receiving therapy become critically important. Therefore, there should be a concerted effort to promote these systems, resulting in education and knowledge expansion among healthcare professionals, and enhancing their expertise. Furthermore, the study of DL should encompass other aspects of the patient beyond the physical, such as the psychoemotional sphere, which can influence their recovery. This approach will foster comprehensive rehabilitation.

6. Conclusions and future works

Our study conducts a comprehensive review of the current advancements in applying deep learning techniques in the field of physiotherapy. The intersection of both fields underscores promising trends, significant opportunities, and impending challenges. Results and discussions were obtained following a structured methodology to select a representative series of papers (23) from which we derived comprehensive, quantitative insights.

Our study yields several distinct conclusions about the emerging field that encompasses deep learning models with physiotherapy. The publications reviewed are recent, ranging from 2018 to 2023, underscoring the infancy and evolving nature of this interdisciplinary research area.

Convolutional Neural Networks (CNNs) are predominantly featured, accounting for 50 % of the selected studies. This prevalence is due to their reliable performance in processing image and signal data. However, the proliferation of CNNs also highlights a potential lack of innovative approaches. In contrast, the emergence of hybrid models, though currently less common, is indicative of a trend that could potentially catalyze innovation and enhance the application of deep learning in physiotherapy.

Data types and sourcing are pivotal aspects elucidated in our analysis. Images (52 %) and signals (44 %) dominated the analyzed studies. Concurrent usage of images and signals within a single study was observed, underscoring the potential benefit of multi-modal data integration. Furthermore, text and structured data, absent in the reviewed physiotherapy studies, represent a yet unexplored niche.

Concerning data sourcing, conventional methods, notably sensors and video recordings capturing movement, are predominant. However, underexplored techniques such as thermographic imaging, 3D imagery, and Functional Movement Screening (FMS) present interesting perspectives for future exploration. From a purely physiotherapeutic perspective, the assessment of movements featured prominently in the reviewed studies. The upper limbs and hands were the most frequently analyzed body parts. Most proposed interventions largely focused on exercise-based treatments for movement-related issues.

However, it's important to acknowledge that a significant limitation of the current body of research lies in the absence of studies that systematically compare intervention groups to control groups, hindering a comprehensive understanding of how deep learning performs in clinical settings, particularly when dealing with patients presenting actual musculoskeletal pathologies and undergoing real therapeutic interventions.

Considering the current gaps in the literature, future research endeavours could explore the application of deep learning methodologies in physiotherapy with a broader scope. This would involve expanding beyond the realm of neurorehabilitation and venturing into areas such as musculoskeletal, respiratory, sports, and other specialized fields of physiotherapy. Additionally, there's an intriguing prospect in utilizing deep learning techniques to enhance the assessment process by integrating them with established physiotherapeutic tests, thereby enhancing the overall reliability and effectiveness of patient evaluations. Projecting into the future, the review points out extensive opportunities for augmenting the integration of deep learning models in physiotherapy. Hybrid models stand out as potentially instrumental, attributed to their capacity to process diverse data types. An illustrative prospect could encompass the mixed use of patient imagery with biosignal data, fostering comprehensive analyses and personalized interventions. Additionally, the exploration of varied data collection methodologies could instigate the inception of innovative tools and techniques, propelling the field into new frontiers of efficacy and precision.

In essence, the synthesis of our findings not only underscores the progress being made but also points out the expansive potential that remains untapped. The exploration of this potential, particularly focusing on diversifying deep learning models, data types, and collection techniques, is set to be fundamental in optimizing the contributions of DL to physiotherapy.

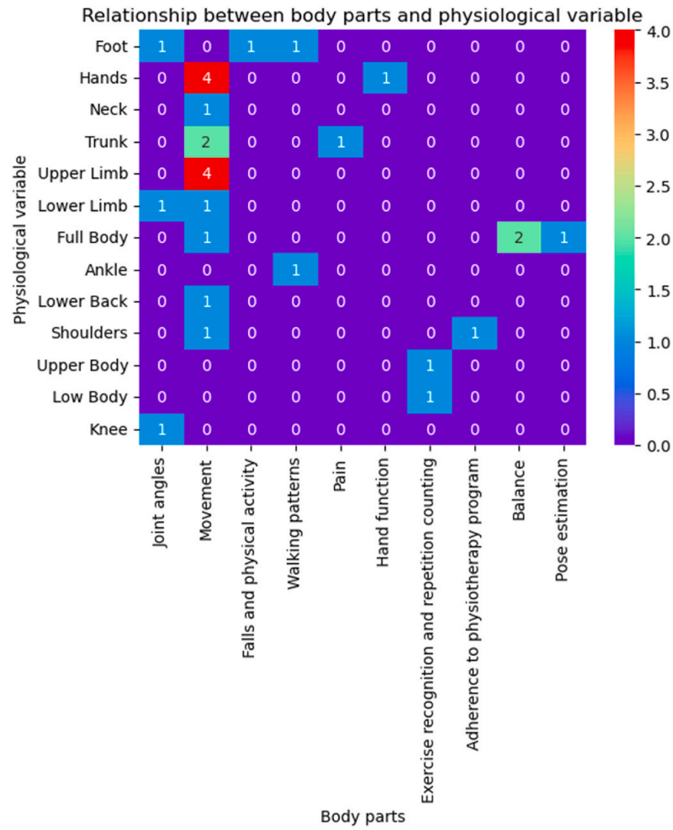

Fig. 20. Heatmap: physiological variables vs. body parts.

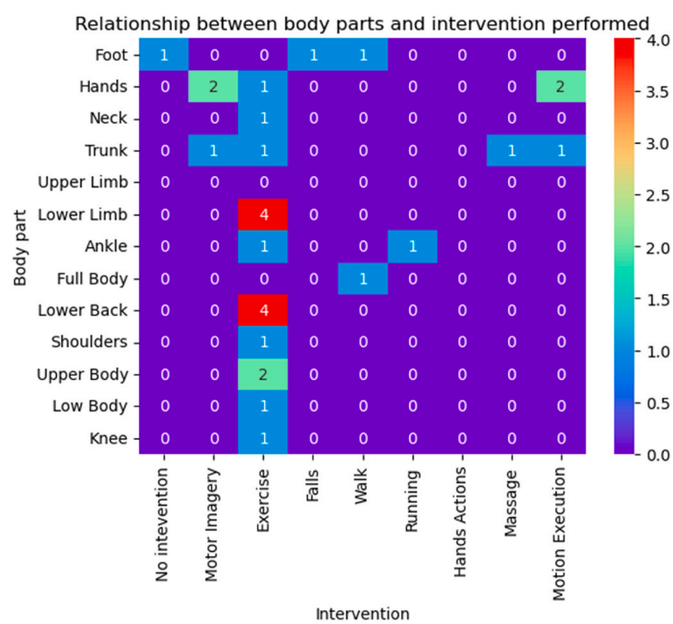

Fig. 21. Heatmap: body parts vs. intervention performed.

CRediT authorship contribution statement

Alberto Nogales: Conceptualization, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Manuel Rodríguez-Aragón:** Conceptualization, Investigation, Validation, Writing – review & editing. **Álvaro J. García-Tejedor:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Call 2022 of the Innovation Check program of the Community of Madrid co-financed by the European Regional Development Fund (ERDF). File No.: 09/835799.9/22.

References

[1] Stuart J. Russell, Norvig Peter, Artificial Intelligence: A Modern Approach. Malaysia, 2016.
[2] Arthur L. Samuel, Some studies in machine learning using the game of checkers, IBM J. Res. Dev. 3 (3) (1959) 210–229.
[3] Reinhard Diestel, Schrijver Alexander, Seymour Paul, Graph theory, Oberwolfach Rep. 4 (2) (2008) 887–944.
[4] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[5] Ravali, Ravula Sahithya, et al., A systematic review of artificial intelligence for pediatric physiotherapy practice: past, present, and future, Neuroscience Informatics (2022) 100045.
[6] Christopher Tack, Artificial intelligence and machine Learning| applications in musculoskeletal physiotherapy, Musculoskeletal Science and Practice 39 (2019) 164–169.
[7] Surabhi Naik, et al., A Literature Review of Artificial Intelligence in Physiotherapy Practice, 2022.
[8] Tasneem Burhani, M Naqvi Waqar, Impact of artificial intelligence in the physiotherapy rehabilitation of distal radial fracture patients: a review, Journal of Pharmaceutical Research International 33 (60B) (2021) 1982–1988.
[9] Bappaditya Debnath, Mary O'brien, Motonori Yamaguchi, Ardhendu Behera, A review of computer vision-based approaches for physical rehabilitation and assessment, Multimed. Syst. 28 (1) (2022) 209–239.
[10] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097–1105.
[11] Xiuli Bi, et al., Computer-aided alzheimer's disease diagnosis by an unsupervised deep learning technology, Neurocomputing 392 (2020) 296–304.
[12] Yang Song, et al., Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI, J. Magn. Reson. Imag. 48 (6) (2018) 1570–1577.
[13] Stefan Kurz, et al., Hybrid modeling: towards the next level of scientific computing in engineering, Journal of Mathematics in Industry 12 (1) (2022) 1–12.
[14] Kipf, N. Thomas, Max Welling, Semi-Supervised Classification with Graph Convolutional Networks, 2016 *arXiv preprint arXiv:1609.02907*.
[15] Jeffrey L. Elman, Finding structure in time, Cognit. Sci. 14 (2) (1990) 179–211.
[16] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[17] Dana H. Ballard, Modular Learning in Neural Networks, AAAI, 1987, pp. 279–284.
[18] Paul Smolensky, Chapter 6: Information Processing in Dynamical Systems: Foundations of Harmony Theory, in: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, 1985.
[19] Ruslan Salakhutdinov, Geoffrey Hinton, Deep Boltzmann machines, in: *Proceedings Of the Twelfth International Conference On Artificial Intelligence And Statistics*, Proceedings of Machine Learning Research, Eds. David Van Dyk and Max Welling. Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA: PMLR, 448–55, 2009, in: https://proceedings.mlr.press/v5/salakhutdinov09a.html.
[20] Ian Goodfellow, et al., Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[21] Staffs Keele, others, Guidelines for Performing Systematic Literature Reviews in Software Engineering, 2007.
[22] David Moher, et al., Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009) e1000097.
[23] Paul Barrett, et al., Matplotlib-A portable Python plotting package, Astronomical Data Analysis Software and Systems XIV (2005) 91.
[24] Peter Ardhianto, et al., A deep learning method for foot progression angle detection in plantar pressure images, Sensors 22 (7) (2022) 2786.
[25] Ming Gao, Jie Mao, A novel active rehabilitation model for stroke patients using electroencephalography signals and deep learning technology, Front. Neurosci. 15 (2021) 780147.
[26] Fangzhou Xu, et al., A transfer learning framework based on motor imagery rehabilitation for stroke, Sci. Rep. 11 (1) (2021) 19783.
[27] Zia Ur Rahman, et al., Automated detection of rehabilitation exercise by stroke patients using 3-layer CNN-LSTM model, Journal of Healthcare Engineering 2022 (2022).
[28] Andreas Spilz, Michael Munz, Automatic assessment of functional movement screening exercises with deep learning architectures, Sensors 23 (1) (2022) 5.
[29] Akshay Kumar, Pirogova Elena, Seedahmed S. Mahmoud, Qiang Fang, Classification of error-related potentials evoked during stroke rehabilitation training, J. Neural. Eng. 18 (5) (2021) 56022.
[30] Hsiao-Lung Chan, et al., Deep neural network for the detections of fall and physical activities using foot pressures and inertial sensing, Sensors 23 (1) (2023) 495.
[31] Wei-Han Chen, et al., Determining motions with an IMU during level walking and slope and stair walking, J. Sports Sci. 38 (1) (2020) 62–69.
[32] Mohsen Gholami, Christopher Napier, Carlo Menon, Estimating lower extremity running gait kinematics with a single accelerometer: a deep learning approach, Sensors 20 (10) (2020) 2939.
[33] Swakshar Deb, Md Fokhrul Islam, Shafin Rahman, Sejuti Rahman, Graph convolutional networks for assessment of physical rehabilitation exercises, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 410–419.
[34] Kang Xia, et al., Hand exoskeleton design and human-machine interaction strategies for rehabilitation, Bioengineering 9 (11) (2022) 682.
[35] Vishwanath Bijalwan, Vijay Bhaskar Semwal, Ghanapriya Singh, Tapan Kumar Mandal, HDL-PSR: modelling spatio-temporal features using hybrid deep learning approach for post-stroke rehabilitation, Neural Process. Lett. 55 (1) (2023) 279–298.
[36] Kai-Yu Chen, et al., Human motion tracking using 3d image features with a Long short-term memory mechanism model—an example of forward reaching, Sensors 22 (1) (2022) 292.
[37] Huihui Li, et al., Massage therapy's effectiveness on the decoding EEG rhythms of left/right motor imagery and motion execution in patients with skeletal muscle pain, IEEE Journal of Translational Engineering in Health and Medicine 9 (2021) 1–20.
[38] Andrea Bandini, et al., Measuring hand use in the home after cervical spinal cord injury using egocentric video, J. Neurotrauma 39 (23–24) (2022) 1697–1707.
[39] Vonstad, Elise Klæbo, et al., Performance of machine learning models in estimation of ground reaction forces during balance exergaming, J. NeuroEng. Rehabil. 19 (1) (2022) 1–12.
[40] Colin Arrowsmith, et al., Physiotherapy exercise classification with single-camera pose detection and machine learning, Sensors 23 (1) (2022) 363.
[41] Chien-Hua Huang, et al., Real-time rehabilitation exercise performance evaluation system using deep learning and thermal image, in: In 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2020, pp. 1–6.
[42] Ghanashyama Prabhu, Noel E. O'connor, Kieran Moran, Recognition and repetition counting for local muscular endurance exercises in exercise-based rehabilitation: a comparative study using artificial intelligence models, Sensors 20 (17) (2020) 4791.
[43] David M. Burns, et al., Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch, Physiol. Meas. 39 (7) (2018) 75007.
[44] Mohd Asyraf Zulkifley, Nur Ayuni Mohamed, Nuraisyah Hani Zulkifley, Squat angle assessment through tracking body movements, IEEE Access 7 (2019) 48635–48644.
[45] Wenchuan Wei, Carter Mcelroy, Sujit Dey, Using sensors and deep learning to enable on-demand balance evaluation for effective physical therapy, IEEE Access 8 (2020) 99889–99899.
[46] Thoutam Anand, Vivek, et al., Yoga pose estimation and feedback generation using deep learning, Comput. Intell. Neurosci. 2022 (2022).