IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 5 July 2022, accepted 3 August 2022, date of publication 16 August 2022, date of current version 24 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3198935

SURVEY

Assessing Human Motion During Exercise Using Machine Learning: A Literature Review

FOTOS FRANGOUDES[1,2], MARIA MATSANGIDOU[1], EIRINI C. SCHIZA[1,2], (Member, IEEE), KLEANTHIS NEOKLEOUS[1], AND CONSTANTINOS S. PATTICHIS[1,2,3], (Fellow, IEEE)

[1]CYENS—Centre of Excellence, 1016 Nicosia, Cyprus
[2]Department of Computer Science, University of Cyprus, 1678 Nicosia, Cyprus
[3]Biomedical Engineering Research Centre, University of Cyprus, 1678 Nicosia, Cyprus

Corresponding author: Fotos Frangoudes (frangoudes.fotos@gmail.com)

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

ABSTRACT The World Health Organization promotes healthy living through regular physical activities, such as exercise and sports, as well as access to healthcare and rehabilitation services for people with motor dysfunctions. However, there is a lack of specialized personnel and increased costs associated with such activities. These have led to the increased use of machine learning for the analysis and evaluation of human motion during exercise. To study the latest advancements in this area, a systematic literature review focusing on publications from 2017 to 2021 was performed. As a result, 88 relevant publications were identified, which developed both shallow machine learning and deep learning algorithms. The results indicated that algorithms for human motion assessment should provide personalized and informative assessments, with explainable and interpretable outcomes, that can be computed in real-time or concurrently with the execution of an exercise. Furthermore, they should be easy to adapt based on the needs of applications and should be able to perform with different motion capture systems. This has been challenging because of the usually small amount of collected data, the lack of large open datasets, and the unique characteristics of exercise motions. Based on the above findings, guidelines for the development of such algorithms are proposed and discussed. They relate to the selection of the type of assessment, handling data imbalances, selecting of motion capture technologies, balancing between accuracy and speed, selecting the right algorithm, performing concurrent assessment during an exercise, personalization and scalability, and evaluation.

INDEX TERMS Assistive technologies, deep learning, exercise, healthcare, human action evaluation, machine learning, motion analysis, motion quality assessment, rehabilitation, sports.

I. INTRODUCTION

Human Motion Quality Assessment (HMQA), also referred to as Human Action Evaluation, Human Performance Assessment, or in broader terms, Action Quality Assessment (AQA), has seen an increased interest in the scientific literature, because of the need for unbiased and personalized assessment of human motion in various domains. HMQA is defined as "quantifying the motion quality from a functional point of view by assessing its deviation from an established model" [1]. In other words, it analyzes and quantifies how someone performs an activity. HMQA should not be confused with Human Activity Recognition (HAR), which is another motion analysis technique that focuses on recognizing what activity is being performed [2], [3].

Applications of HMQA have been used in sports training applications [4], athletic event performance scoring systems [5], motor rehabilitation systems [6], [7], medical diagnostics [8], skill assessment and education [9], [10], and to assess ergonomic risks [11]. Such assessments, in the absence of automated processes, are mostly performed by experts in their respective fields and thus rely on their level of expertise

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

and experience [12], [13]. However, there is often a low agreement between different raters performing HMQA of the same activity. In other words, they have a low inter-rater reliability in their evaluations [14], [15]. This can happen for various reasons. For example, time limitations can hinder the performance of comprehensive assessments. In addition, assessments rely on the experience level of raters [16] and are also affected by raters' subjective biases. Lower inter-rater and intra-rater (i.e., the ability of a rater to consistently rate the same activity) reliability scores decrease further when assessing complex motions, and during live performances, compared to offline evaluations of recordings [17]. The limitation in consistently rating activities in real-time is especially detrimental when there is a need to use the assessment results to provide immediate feedback to the performer. Finally, because such assessments are performed by a human expert, they are restricted in their scalability and cannot be performed without the presence of the expert. For example, this is the case in home applications [18], and industrial settings [19].

A. HMQA DURING EXERCISE

Exercise is defined as a "planned, structured, and repetitive bodily movement done to improve or maintain one or more components of physical fitness" [20]. It is a subset of physical activity that includes body conditioning and sports activities [20]. If we decompose the above definition, planned or goal-oriented exercise motions are not reactive to an event but are performed to achieve a specific outcome [21]. Structured or task-oriented exercise motions have common characteristics that are specific to the task at hand [21]. For example, these characteristics may include set starting and ending points for the motion and predefined motion trajectories. Finally, exercise motions are often repetitive and performed multiple times during a session [21]. HMQA is often required during exercise to assess the execution of performed motions. This assessment is usually performed by therapists and medical personnel for healthcare applications (e.g., for diagnostic purposes and rehabilitation), and by coaches and trainers for sports and wellness applications. Various approaches for HMQA during exercise have recently been explored to facilitate and automate the assessment process.

This study presents Machine Learning (ML) processes that have been proposed for HMQA during exercise. A systematic literature review was conducted to identify previous studies in the area that used ML. Although there have been some relevant reviews in individual application domains in the past [4], [5], [6], [22], to the best of our knowledge, no prior review has studied the use of HMQA during exercise across multiple domains, particularly with the use of ML. This review aims to address the following research questions (RQ):

- RQ1: What are the requirements for ML algorithms used for HMQA during exercise?
- RQ2: What are the challenges for ML algorithms used for HMQA during exercise?

The remainder of this paper is structured as follows: Section II provides a high-level description of the process for applying ML for HMQA during exercise. Section III describes the methodology used in the literature review. Section IV presents the results of the review, and Section V discusses the results, answers the above research questions, and proposes a set of guidelines for developing ML algorithms for HMQA during exercise. Finally, Section VI summarizes the process and findings.

II. HMQA USING MACHINE LEARNING

Traditionally, HMQA has been performed using either rule [23], [24], [25] or template-based [26], [27], [28] approaches. Rule-based approaches use predefined conditional statements to evaluate the properties of human motion (e.g., joint angles, range of motion, and relations between joints). Template-based approaches use algorithms such as Dynamic Time Warping (DTW) and Hidden Markov Models (HMM), which perform pattern matching using previously recorded motions. These methods are easy to implement and run in real-time; thus, they are ideal for assessing motion during exercise. However, the number of exercise executions that can be used to create rules/templates is limited. This reduces the scalability of these approaches [29]. Consequently, it is difficult to develop personalized solutions that can match the unique characteristics or impairments of each individual. To overcome the above limitations, there has recently been increased interest in using ML for HMQA. The process for developing such solutions includes several steps. A high-level schematic representation of the overall process is shown in Fig. 1. However, the exact steps that each solution adapts, can vary based on the input data modalities, type of assessment, and algorithm used. Next follows a brief description of each phase of this process.

A. DATA COLLECTION AND PREPROCESSING

The first step is the collection of data using motion capture (MoCap) systems (e.g., inertial-based, optical). Data collection can occur either online, where data are directly used to train/improve ML models (e.g., using reinforcement learning), or offline, where they are stored as a dataset that is used later for training. Raw data captured using MoCap systems are continuous signals, (e.g., (x, y, z) coordinates of the position of joints over time). These time-series often contain noise, which makes them more difficult to analyze, and thus may require preprocessing. Filters are used to remove such unwanted features [30]. The selection of the filter and its parameters varies based on the type of signal and capture characteristics (e.g., the polling rate and signal resolution). Filtering is often used with shallow ML algorithms, whereas Deep Learning (DL) approaches often use unfiltered data.

When capturing data during an exercise, especially in a non-controlled setting (i.e., not in a laboratory or following strict protocols), values are recorded continuously. Therefore, captured motion data include values from motions performed before and after the execution of an exercise (i.e., during preparation and post-exercise actions). These values are irrelevant for the assessment and are discarded, thus, segmenting

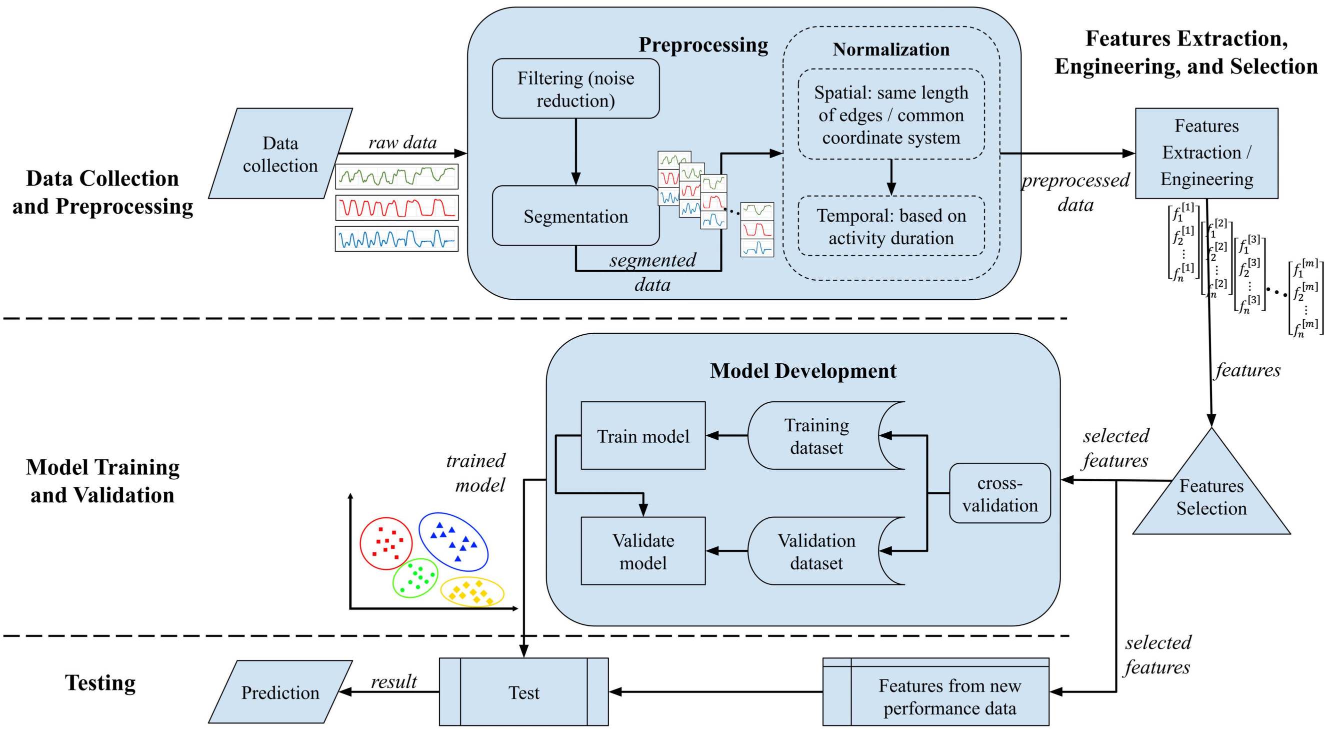

FIGURE 1. High-level schematic representation of the process for HMQA using ML.

the data into smaller time-series [21]. The remaining values include data from only a single exercise execution or repetition, which can be further divided into sub-phases of an exercise (e.g., a tennis serve can be separated into three phases: preparation, acceleration, and follow-through [31]). Segmentation is typically performed based on the specific characteristics of individual signals. For example, in a single-leg squat exercise, analysis of the knee angle can discern repetitions [32].

Next, based on the MoCap system, it may be necessary to normalize the data to ensure their comparability. Normalization can be performed across two different modalities: spatial and temporal. Spatial normalization ensures that all data are represented in the same coordinate system and aligned based on a common plane. If the input data are not spatially normalized, especially when data are represented using a human skeleton, it can result in a lower, and with a greater variance, accuracy of the assessment [33]. Temporal normalization ensures that all segments have the same length, and may be necessary when using certain ML algorithms. For instance, when using Convolutional Neural Networks (CNN) that require data to fit within a fixed-size matrix. However, this alters the time-series (i.e., by discarding or introducing new values) and may yield unexpected results. Overall, preprocessing can improve assessment accuracy by making the input data comparable and easier to analyze. Nonetheless, it is time-consuming [34] and therefore, each application should examine which of these steps to apply.

B. FEATURE EXTRACTION, ENGINEERING, AND SELECTION

This section describes steps for extracting features from the captured signals and engineering new features. The simplest approach for feature extraction is by using the raw data from each component of the captured signals as a feature vector (e.g., for a triaxial accelerometer, the values of each of the three axes (x, y, z) across time can be used as a feature). These feature vectors include all time-series values from an exercise execution and are used as inputs for ML algorithms [35]. Such input vectors are typically used in DL algorithms. A different method is through the application of various statistical/aggregate functions (e.g., minimum, mean, and standard deviation). This produces descriptive statistics for the time-series, which are used as features [36]. These features are often used with shallow ML algorithms.

In addition to the above methods, feature engineering transforms features into different and more efficient representations while maintaining the expressivity of the original features. Many techniques are used for such transformations, each taking advantage of different characteristics of the data. One such approach is the use of dimensionality reduction algorithms [37] (e.g., Principal Component Analysis), which use a set of features to create new features with reduced dimensionality. Another approach is transfer learning, that transforms features into a format that has been successfully used in other domains [38]. For example, a set of feature vectors representing time-series values can be transformed into an RGB image, where each column represents a different feature (e.g., a skeletal joint), rows represent time points, and the RGB values represent the values of that feature at that time point (i.e., (x, y, z) position of a joint) [39]. Another transformation, mostly used with skeletal data, is into a graph that can model the adjacencies of the joints of the human body [40].

Once all features are extracted and/or engineered, it is possible to have many features that will increase the complexity and accuracy of the ML model [41]. Therefore, to reduce complexity, a subset of features may be selected instead of using all of them to develop the model. This may be needed, especially in more complex systems that use multiple input devices and modalities, and can therefore include many features irrelevant to the performed motions [42]. In simple cases, this process can be performed manually based on empirical knowledge. However, various feature selection algorithms have been proposed to select optimal subsets of features. Feature selection algorithms are divided into unsupervised and supervised algorithms [43]. Unsupervised algorithms can use correlation analysis or clustering techniques to identify redundant features. Supervised algorithms, on the other hand, are separated into (1) wrapper techniques (e.g., sequential algorithms and genetic algorithms), (2) filter techniques (e.g., correlation criteria and mutual information), and (3) embedded techniques that select features while training the model (e.g., Random Forest (RF)) [43].

C. MODEL TRAINING, VALIDATION, AND TESTING

The final steps in the process include the training and validation of the ML model and subsequent testing. The selection of the ML algorithm that will be used in the process depends heavily on the type of assessment performed.

1) TYPES OF HMQA

HMQA is divided into three categories based on the type of assessment performed: regression-based, ranking-based, and parametric assessments [44].

a: REGRESSION-BASED ASSESSMENTS

Regression-based assessments characterize the quality of motion using a continuous numerical value (i.e., a score). One approach to developing such assessments is to use only the optimal executions of an exercise to train the ML model [45]. The assessment is then performed by detecting the deviations of a performed motion as compared to the optimal ones. If the two have a high similarity (i.e., only small deviations), the performed motion was without any errors. On the contrary, greater deviations signify a motion of lower quality and with more errors. As only the optimal executions of an exercise are required for the assessment, it can be implemented even with a few representative data samples. However, if there are samples across a wider range of motion quality, a model can also be developed by learning associations of features with different scores [46]. Regression assessment can be applied using a unified regression model across a set of different motions. However, this approach can lead to problems in fitting all the data into a single model and decreasing the overall performance, especially with similar activities [47]. Instead, it is preferable to use an individual assessment model for each motion. However, regression-based assessments are unable to provide more context about their outcomes (e.g., identifying specific erroneous patterns or what causes the detected deviations).

b: RANKING-BASED ASSESSMENTS

Ranking-based assessments use classification to an ordinal value that represents the skill level of the execution (e.g., poor, fair, good [48]). Such assessments classify the performed motions into specific classes that follow a rank but may be difficult to quantify exactly. In its simplest form, this can include two classes [49] (e.g., incorrect or correct motion). However, additional classes can provide a finer distinction between the performed motions [50]. In addition, hybrid approaches have been studied in which ranking-based assessment is followed by regression-based algorithms that use the confidence scores of the classification to compute a motion quality score [3]. Ranking-based assessments have the advantage of offering a better distinction in terms of motion quality. This requires more samples from each class. As such, they can develop a scaling problem if there are many classes or characteristics that increase complexity. Moreover, although they provide more information, the ranking nature of the assessment does not allow the identification of disjoint classes that are not comparable to each other.

c: PARAMETRIC ASSESSMENTS

Parametric assessments evaluate domain-specific exercise characteristics using either a continuous numerical value (e.g., follow-through for table tennis strokes [51]) or by classifying them (e.g., identification of postural errors during a barbell squat exercise [52]). Such methods utilize techniques similar to those of the other two approaches, but with a focus on domain-specific parameters. Therefore, they provide more context in their assessments with relevant information about the motion performed. However, as a result, more samples are required to implement them, and it is harder to reuse them in other applications.

2) MODEL DEVELOPMENT

After selecting the ML algorithm, the model is trained. The process starts by separating the dataset into the corresponding training and validation subsets (often by using a cross-validation technique such as k-fold and leave-k-subjects-out). The cross-validation technique used depends on the characteristics of the dataset. For instance, smaller datasets may have higher accuracy with a leave-k-subjects-out approach [3] because the data of individual subjects are not mixed in both the training and validation subsets. The cross-validation technique can also be used during the training phase of the model to facilitate the selection of the optimal parameters of the algorithm.

Once the model is trained, it can be tested for its performance over a set of new data, referred to as the test dataset. This set of data can be held-out from the initial dataset or recorded and assessed in real-time. The final testing is performed by following a similar process to the one during the training by applying preprocessing steps (i.e., filtering, segmentation, and normalization as necessary), and feature extraction and selection. Finally, the data are fed into the trained model for evaluation.

III. LITERATURE REVIEW METHODOLOGY

A systematic literature review was conducted following the PRISMA methodology [53]. The steps in the process are presented in Fig. 2 and discussed in detail below.

A. SEARCH STRATEGY

A search was conducted for publications in the electronic databases ACM Digital Library, IEEE Xplore Digital Library, PubMed, Sage Journals, Taylor & Francis Online, and Scopus. In addition, the databases Academic Search Ultimate, CINAHL, MedLine, and ScienceDirect were searched using EBSCOHost. The results were restricted to publications published in English between January 2016 and December 2021.

The search across the databases was based on the search pattern [Machine Learning] AND [Motion] AND [Human] AND [Assessment]. Each of the four core search terms was expanded to a set of keywords of similar context as follows: Machine Learning: machine learning, deep learning, neural network, convolutional network, memory network, LSTM, SVM, Vector Machine, forest, regression; Motion: exercise, rehab, sport, human action, action quality, movement, fitness; Human: human, patient, athlete, limb; Assessment: assess, quality, evaluate, correct. Additional publications were retrieved and processed based on existing knowledge.

B. SELECTION CRITERIA

The focus of this review is to study methods for the assessment of human performance during exercise using ML algorithms. Therefore, various inclusion criteria were selected to ensure that only relevant publications were included in the analysis. For this review, the term exercise was adopted based on the definition in Section I-A and includes motions performed for body conditioning (e.g., squat, lunge, shoulder flexion) or related to a sport (e.g., strokes in tennis, martial arts moves). Publications that only analyzed motions related to activities of daily living (e.g., walking, drinking from a cup, standing-to-sitting) were excluded. Publications that only included wrist and/or finger exercises were also excluded, as were publications that focused on static postures and did not include motion.

In addition, all included publications used ML algorithms for HMQA. Publications that only recognized different exercises (i.e., only performed HAR), or performed assessments without the use of ML algorithms, were excluded. Regarding the modalities of the captured data, only studies that used human-body-centric data modalities were included. The term human-body-centric data modalities refers to data whose values relate to human body segments (e.g., forearm, shoulder, chest). For example, such data can be captured from internal sensors (i.e., values from an accelerometer, gyroscope, and magnetometer) that are attached at specific body locations or from optical cameras that can discern skeletal data (i.e., joint positions and/or orientations). Publications in which the analysis was only based on other modalities, such as muscle activity, force, and image-based characteristics with no skeletal context, were excluded.

C. SCREENING STRATEGY AND ARTICLE REVIEW

The results from each electronic database were retrieved and imported into Zotero (http://www.zotero.org/), a reference management software. After removing duplicate entries, the initial results were screened based on titles and abstracts against the inclusion criteria. Publications that did not meet all inclusion criteria were excluded. The remaining publications were further screened based on their full text to assess their eligibility for inclusion in the analysis. Simultaneously, additional publications were identified and screened for eligibility as well. These were identified based on references in the screened publications, or from the publications that cited publications that met the inclusion criteria. Once the set of publications for analysis was finalized, the included publications were reviewed in detail, and the extracted data were added to a spreadsheet for analysis.

IV. RESULTS

The search of the electronic databases yielded 26851 publication records (Fig. 2). An additional 28 records were included based on previous knowledge. After duplicates were removed, 12176 unique records were screened based on their titles and abstracts. Of those, 11740 were excluded, and 436 were assessed based on their full text for eligibility. Another 53 publications were identified and included in the full-text assessment based on references and citations of the assessed records, bringing the total number of publications assessed for eligibility based on their full-text to 489. Of these, 401 publications did not meet the inclusion criteria and were excluded from further analysis. Almost three-quarters (297/401) were excluded because they did not include exercise-related motions. Another 59 publications did not perform any assessment of human motion, and 25 publications assessed motion without using ML. Fifteen publications did not use human-body-centric data for the assessment. Finally, five publications were excluded because they did not include details regarding the implementation of the assessment or only proposed the use of the technology for future use. The full text of one publication could not be found. This process yielded 88 publications that were included in the analysis. Tables 1, 2, 3, and 4 list the included publications as related to healthcare, rehabilitation, sports, and wellness applications, respectively. The following subsections

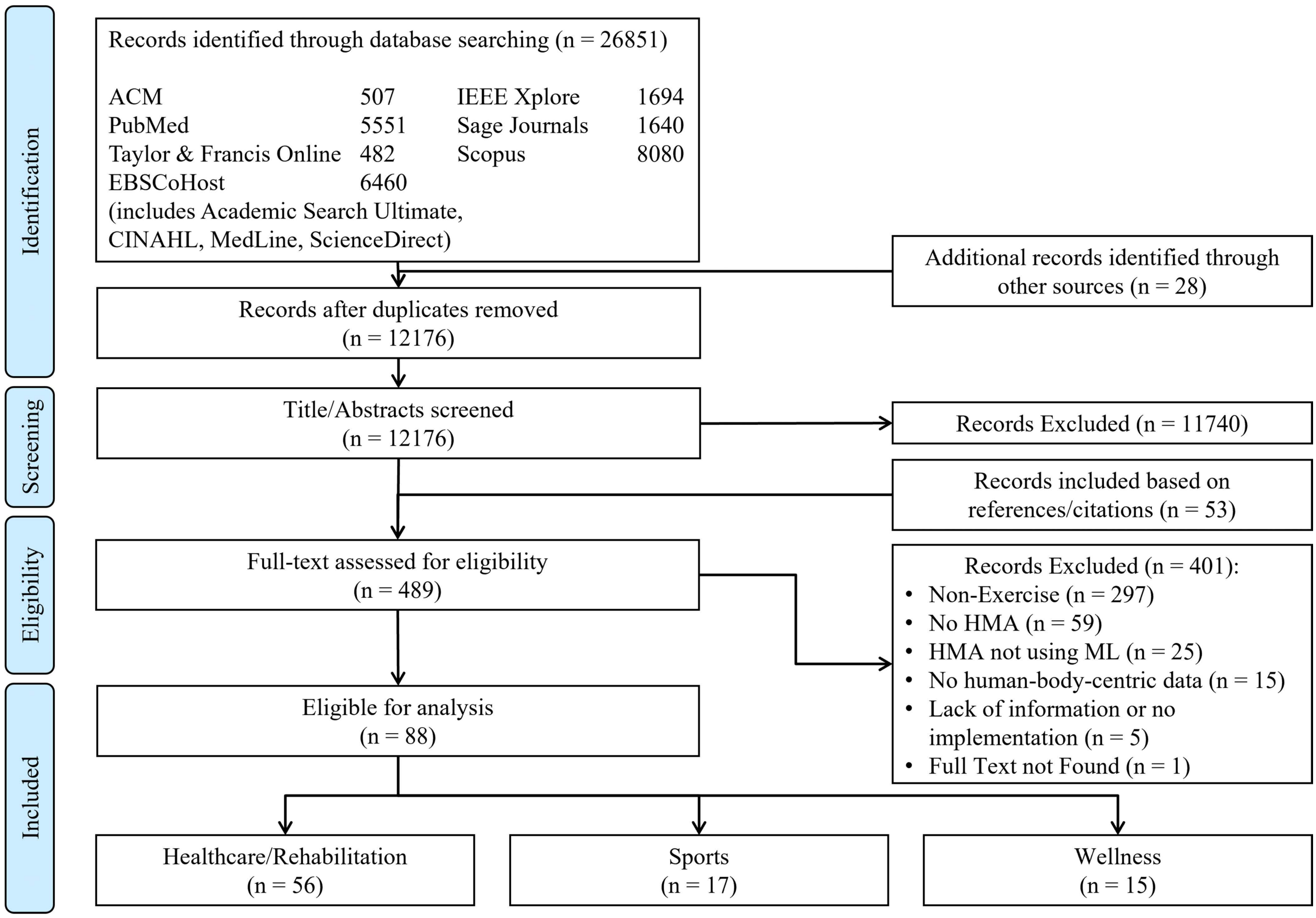

FIGURE 2. Summary of the PRISMA methodology followed for the literature review.

provide information about various aspects of the publications extracted from the analysis.

A. APPLICATION DOMAINS AND TARGET POPULATIONS

The majority of publications (56/88) were related to healthcare and rehabilitation applications (Tables 1 and 2). Another 17 publications were related to sports applications (Table 3), and the remaining 15 were related to wellness applications (Table 4). Subdomains were identified in each application domain based on the purpose of the assessment, as shown in Fig. 3. Of the publications in the healthcare/rehabilitation domain, 17/56 performed a functional motor assessment of patients, while 16 described systems for monitoring exercise during rehabilitation. Another 12/56 related to the development of virtual therapist applications that provide specific feedback to patients. Six publications presented serious games that use motion assessment within the context of the game, and 5/56 were more general with no specific stated use. From publications in the sports domain, 7/17 were designed as coaching systems that provide corrective feedback to athletes based on performed motions. Another seven assessed the skill level of athletes, and three were related to systems for judging sports performances by predicting judged athletic performance scores. Finally, in the wellness domain, 5/15 were related to fitness coaching applications that guide (mostly novice) people during exercise, and one publication performed a functional movement assessment that differentiates between younger and older users. The remaining 9/15 publications were related to general exercise assessment for wellness systems.

The included studies also focused on a wide range of population groups. For example, in the healthcare/rehabilitation domain (Fig. 4 (a)), a large portion of these publications (38/56) were related to neurological disorders, with the majority (29/38) involving stroke survivors. Another 5/38 publications related to people with Parkinson's disease. Of the last four, two were related to people with Alzheimer's disease, one to people with multiple sclerosis, and one to neurological disorders in general. Another population group of interest within the healthcare domain was people with musculoskeletal disorders (11/56 publications). Of those, eight were general not focusing on a specific musculoskeletal disorder, two related to patients recovering from anterior cruciate ligament (ACL) reconstruction, and one with people with knee osteoarthritis. Finally, 2/56 publications in the healthcare/rehabilitation domain were related to people with motor

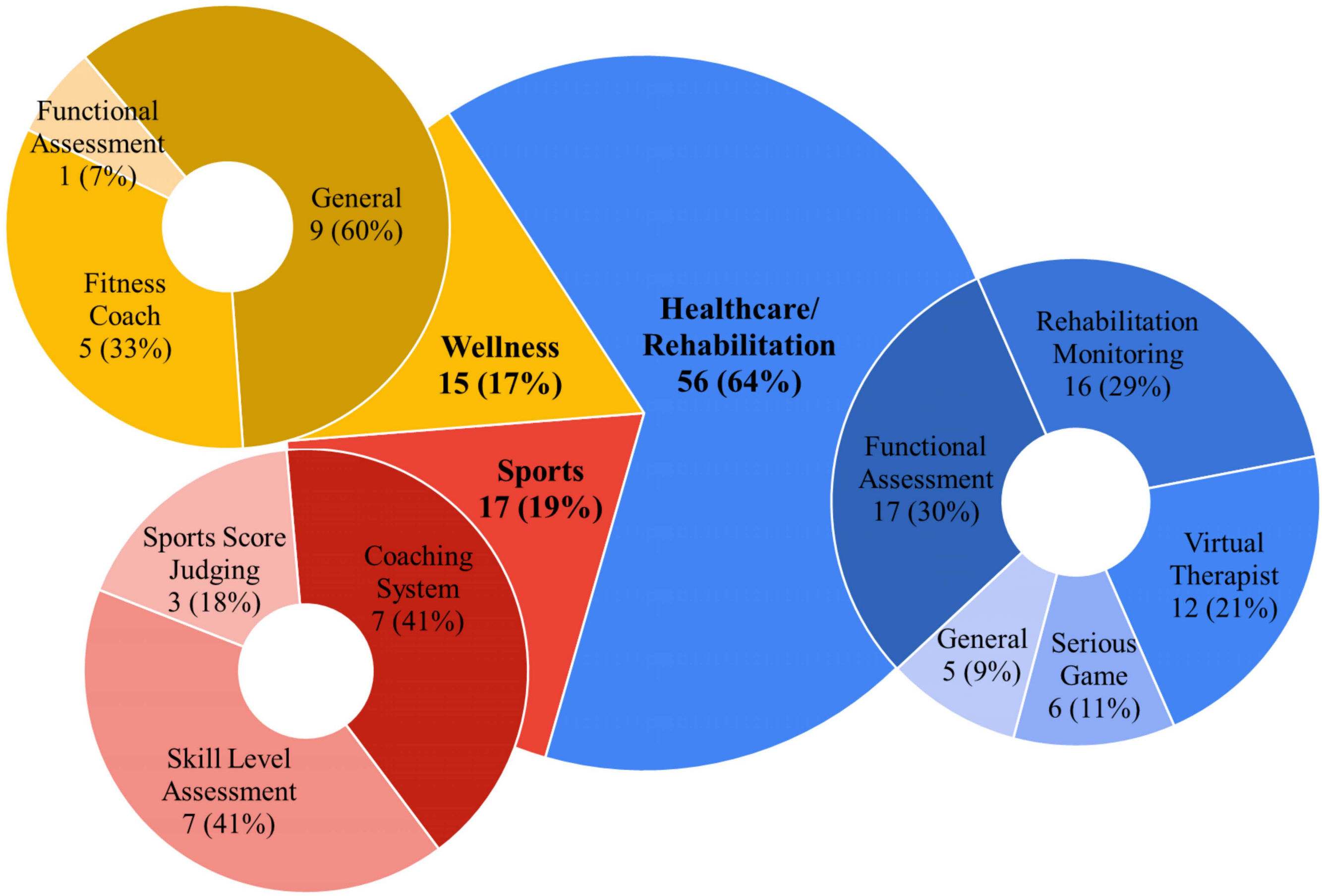

FIGURE 3. Application domains and subdomains based on the purpose of the assessment of identified publications.

deficits, and another 7/56 were more general in the healthcare domain and rehabilitation.

Similarly, various sports were included in publications in the sports domain (Fig. 4 (b)). High interest in HMQA exists in racket sports (8/17 from publications related to sports). Of these, four publications were related to tennis, three to table tennis, and one to badminton. Other sport types of interest included gymnastics, with one publication for floor exercises, and one for the vault, and martial arts, where one publication was related to kick-boxing, and another to Tai Chi. Two publications related to water sports, with one studying canoeing, and one diving. Similar were winter sports, with one publication related to figure skating, and one to skiing. A comparison of athletes from different sports performing various exercises was performed in 3/17 studies. In addition, one publication performed a separate HMQA analysis on motions performed in three different sports (i.e., gymnastics vault, diving, and skiing). Therefore, the total number of publications accounting for individual sports adds to 19, which is two more than the number of publications included in the study. Publications related to the wellness domain aimed toward the general population, with no specific characteristics, to improve overall fitness levels, and promote healthy living.

B. HMQA USING ML

1) DATA COLLECTION AND DATASETS

The first step in the ML process is data collection. The 88 publications included in the review developed ML solutions using 73 unique datasets (i.e., nine datasets were used in more than one publication). Forty-one of the 73 datasets related to the healthcare domain, 17 to sports, and the remaining 15 to wellness (Table 5).

a: STUDIES DESIGN

Almost two-thirds of the datasets (46/73) used a one-group design (Table 5) and all participants were treated as a single group with common characteristics. For these datasets, data collection was performed in two ways. First, the participants performed exercises only in an optimal manner, and the models were developed to recognize deviations from the optimal motions. Second, the participants performed the exercises both correctly and incorrectly. Thus, both were used to train the ML models, which learned to detect variations in the quality between motions. The remaining 27/73 used a multiple-group design where participants were separated into two or more groups (e.g., based on skill level (amateur vs. elite), based on health condition (impaired vs. healthy)). When using these datasets, the developed models

TABLE 1. Included studies related to HMQA during exercise for healthcare applications.

Author Year	Application Subdomain	Population	Inertial	RGBD	Optical	Features Type	Algorithms	Assessment Type	Evaluation Metric	Evaluation Value
Richter *et al.* (2019) [54]	FA	ACL			SP	AV	LR	O	Acc	0.52-0.81
Yu *et al.* (2020) [55]	FA	AD		SPO		TR	GCN	O	SD	0.808
Yu *et al.* (2021) [56]	FA	AD		SPO		TR	GCN	S	SD	0.933
Albert *et al.* (2021) [30]	FA	PD,S		SP		TM	CNN	N	F1s	0.81
Jiang *et al.* (2017) [57]	FA	S	A			AV	AB	O	Acc	0.9925
Julianjatsono *et al.* (2017) [58]	FA	S	AG	SO		AV	NNR	S	CC	0.58
Jung *et al.* (2017) [59]	FA	S	O			AV	LR	O	F1s	0.9119
Eichler *et al.* (2018) [60]	FA	S		SP		AV	RF	O	Acc	0.9318
Eichler *et al.* (2018) [61]	FA	S		SP		AV	RF	O	Acc	0.9318
Jung *et al.* (2018) [62]	FA	S	O			AV	MT	O	F1s	0.79
He *et al.* (2019) [63]	FA	S	AGM			AV	SVM	O	Acc	0.95-0.99
Sarwat *et al.* (2020) [64]	FA	S	AG			AV	XGB	O	Acc	0.9556
Sarwat *et al.* (2020) [65]	FA	S	AG			AV	SOM	O	Acc	0.918
Wang *et al.* (2020) [66]	FA	S			SO	DR	SVM	O	F1s	0.95
Chen *et al.* (2021) [67]	FA	S		SP		TM	GCRNN	S	CoC	0.89
Formstone *et al.* (2021) [68]	FA	S	AGO			DR	LightGBM	O	F1s	0.17-0.91
Lee *et al.* (2021)a [69]	FA	S	A			AV	GPC	O	Acc	0.866
Li *et al.* (2018) [70]	MA	G			SO	TM	DCGAN	S	SAD	0.793-1.99
Sadawi *et al.* (2019) [71]	MA	G		SP		DR	k-NN	O	Acc	0.9467
Miron *et al.* (2021) [72]	MA	G		SP		TM	Res-TCN	O	Acc	0.712
Whelan *et al.* (2017) [52]	MA	MC	A			AV	RF	N	Acc	0.65
Whelan *et al.* (2017) [73]	MA	MC	A			AV	RF	N	Acc	0.78

Application Domains: FA: Functional Assessment, MA: Movement Assessor
Population: ACL: ACL Reconstruction, AD: Alzheimer's Disease, G: General, MC: Musculoskeletal Conditions, PD: Parkinson's Disease, S: Stoke
Input Modalities: A: accelerometer, G: gyroscope, M: magnetometer, O: orientation, SP: skeletal joint position, SO: skeletal joint orientation
Features Type: AV: Aggregate Values Feature Vector, DR: Dimensionality Reduction Feature Vector, TM: Time-series Feature Vectors, TR: Transformed Feature Vectors
Algorithms: AB: Adaptive Boosting (AdaBoost), CNN: Convolutional Neural Network, DCGAN: Deep Convolutional Generative Adversarial Network, GCN: Graph Convolutional Neural Network, GCRNN: Group-Constraint Convolutional Recurrent Network, GPC: Gaussian Process Classifier, k-NN: k-Nearest Neighbors, LightGBM: Light Gradient Boosting Model, LR: Logistic Regression, MT: Model Tree, NNR: Neural Network Regression, Res-TCN: Temporal Convolutional Network, RF: Random Forest, SOM: Self-Organizing Maps, SVM: Support Vector Machine
Assessment Type: N: Nominal, O: Ordinal, S: Score
Evaluation Metric: Acc: Accuracy, CC: Confidence of Determination, CoC: Correlation Coefficient, F1s: F1-Score, SAD: Sum of Absolute Differences, SD: Separation Degree

were designed to identify the group to which the participant performing the exercise belonged to. The latter approach was more frequently used in the sports domain, while in healthcare/rehabilitation and wellness domains a one-group design was more often preferred (Table 5).

b: SAMPLE SIZE

Regarding the number of participants (Table 6), the datasets in healthcare/rehabilitation and wellness domains had similar sizes with a combined median number of 21 participants (25th percentile = 10, 75th percentile = 36). In sports, the median number of participants was less than half that of the other two (see Table 6), mostly because of the number of datasets with only a single participant. In contrast, the number of included exercises across all domains was more uniform (Table 6), with a combined median number of three exercises (25th percentile = 1, 75th percentile = 6). Finally, the number of exercise executions included in each dataset (i.e., number of repetitions) varied between the three domains (Table 6), with sports having the fewest and wellness the most. The combined number of samples across all three domains had a median of 815 (25th percentile = 174, 75th percentile = 1760).

c: INPUT DEVICES AND DATA MODALITIES

The captured motion data in all the datasets were in a human-body-centric format based on the inclusion criteria of the review. As such, motion data were mostly recorded using inertial sensors and/or optical cameras, from which a skeletal representation was extracted. Both types of technologies were used in 39/73 datasets (i.e., some datasets used both technologies).

Fig. 5 shows the combinations of types of inertial sensors used in the datasets. Accelerometers and gyroscopes (i.e., inertial sensors with six degrees of freedom (DOF)), were used most often (16/39 datasets). The two sensors together with magnetometers (i.e., nine DOF) were used in another 15 datasets. Individual sensors (i.e., three DOF) were used less frequently. Inertial sensors were placed at various locations based on the exercises performed in each dataset. Fig. 6 shows the locations on the body sensors were attached, along with the number of datasets that placed sensors at each location. Sensors were attached most often at the wrists for the upper extremities (18 and 17 placements at the right and left wrist, respectively), and at the thighs for the lower extremities (14 and 13 placements at the right and left thigh, respectively). For the core body, sensors were most often attached

TABLE 2. Included studies related to HMQA during exercise for rehabilitation applications.

Author Year	Application Subdomain	Population	Inertial	RGBD	RGB	IR	Features Type	Algorithms	Assessment Type	Evaluation Metric	Evaluation Value
Decroos *et al.* (2019) [34]	RM	A		SP			AV	XGB	N	Acc	0.74
Kianifar *et al.* (2017) [32]	RM	ACL	AG				AV	SVM	O	Acc	0.66-0.96
De Villa *et al.* (2021) [36]	RM	E	AG				AV	SVM	O	PR	0.97-0.99
Rao *et al.* (2019) [74]	RM	G		SP			DR	LSTM	S	Acc	0.96
Du *et al.* (2021) [75]	RM	G		SP			TR	ST-GCN, LSTM	S	MAE	0.02
Akshaya Devi *et al.* (2021) [76]	RM	KO	A				AV	RF	MN	F1s	0.79-1.0
Zhu *et al.* (2019) [77]	RM	MC	A				TM	LSTM	O	Acc	0.91
Liao *et al.* (2020) [78]	RM	MC		SO			DR	CNN, LSTM	S	MAD	0.03
Zhang *et al.* (2020) [35]	RM	MC	A				TM	CNN	O	Acc	0.99
Chowdhury *et al.* (2021) [40]	RM	MC		SP			TR	GCN, LSTM	S	RMSE	0.19
Whitford *et al.* (2021) [79]	RM	MC	AO		2DS		AV	RF	O	Acc	0.96-0.98
Palomares-Pecho *et al.* (2020) [49]	RM	MD			2DS		AV	IF	O	Acc	0.96
Raihan *et al.* (2021) [80]	RM	PD,S		SP			DR	CNN	S	MAD	0.14
Zhi *et al.* (2018) [81]	RM	S		SP			TM	LSTM	N	F1s	0.88
Khoramdel *et al.* (2021) [82]	RM	S		SP			TM	TR	N	Acc	0.85
Miao *et al.* (2021) [83]	RM	S	AGO	SO			TM	LSTM	O	Acc	0.96
Esfahlani *et al.* (2019) [50]	SG	MS	AG	SO			AV	FFNN, SOM, NAR	O	Acc	0.93
Chytas *et al.* (2021) [84]	SG	ND				SP	AV	DCNN	O	Acc	0.92
Avola *et al.* (2018) [85]	SG	PD		SPO			TM	LSTM	S	Acc	0.95
Avola *et al.* (2019) [45]	SG	S		SPO			TM	GRU RNN	S	Acc	0.79
Bai *et al.* (2019) [86]	SG	S		SP			TR	THCNN	O	Acc	0.87
Lee *et al.* (2021) [87]	SG	S		SP			AV	MLP	O	Acc	0.93
Richter *et al.* (2017) [88]	VT	G		SP			AV	SVM	MN	BA	0.84
Richter *et al.* (2017) [33]	VT	G		SP			AV	SVM	MN	BA	0.87
Wei *et al.* (2018) [89]	VT	PD		SP			AV	SVM	MN	Acc	0.85-0.97
Wei *et al.* (2019) [90]	VT	PD		SP			AV	SVM	MN	Acc	0.86-0.94
Lee *et al.* (2018) [91]	VT	S	AG				AV	k-NN, RF	N	Acc	0.84
Lee *et al.* (2019) [92]	VT	S		SP			AV	DT, ANN, TM	O	F1s	0.91
Mortazavi *et al.* (2019) [21]	VT	S		SP			AV	SVM	MN	Acc	0.84-0.87
Lee *et al.* (2020) [93]	VT	S		SP			AV	RL, ANN	O	F1s	0.81
Lee *et al.* (2020) [94]	VT	S		SP			AV	RL, ANN, E	O	F1s	0.91
Lee *et al.* (2020) [95]	VT	S		SP			AV	ANN, E	O	F1s	0.82
Kanade *et al.* (2021) [96]	VT	S		SPO			DR	CNN	S	MSE	0.02
Miao *et al.* (2021) [97]	VT	S	AGO				AV	DTW, k-NN	O	Acc	0.67-0.86

Application Domains: RM: Rehabilitation Monitoring, VT: Virtual Therapist, SG: Serious Game
Population: A: Athletes, ACL: ACL Reconstruction, E: Elderly, G: General, KO: Knee Osteoarthritis, MC: Musculoskeletal Conditions, MD: Motor Deficits, ND: Neuromuscular Disorders, PD: Parkinson's Disease, S: Stoke
Input Modalities: A: accelerometer, G: gyroscope, O: orientation, SP: skeletal joint position, SO: skeletal joint orientation, 2DS: 2D skeletal position
Features Type: AV: Aggregate Values Feature Vector, DR: Dimensionality Reduction Feature Vector, TM: Time-series Feature Vectors, TR: Transformed Feature Vectors
Algorithms: ANN: Artificial Neuronal Network, CNN: Convolutional Neural Network, DCNN: Dynamic Convolutional Neural Network, DT: Decision Tree, DTW: Dynamic Time Warping, E: Ensemble, FFNN: Feed-Forward Neural Network, GCN: Graph Convolutional Neural Network, GRU RNN: Gated Recurrent Unity, IF: Isolation Forest, k-NN: k-Nearest Neighbors, LSTM: Long Short-Term Memory, MLP: Multi-Layer Perceptron, NAR: AutoRegressive Neural Network, RF: Random Forest, RL: Reinforcement Learning, SOM: Self-Organizing Maps, ST-GCN: Spatial-Temporal Graph Convolution Network, SVM: Support Vector Machine, THCNN: Two-headed Convolutional Neural Network, TM: Threshold Model, TR: Transformer, XGB: eXtreme Gradient Boosting (XGBoost)
Assessment Type: MN: Multiple Nominal, N: Nominal, O: Ordinal, S: Score
Evaluation Metric: Acc: Accuracy, BA: Balanced Accuracy, F1s: F1-Score, MAD: Mean Absolute Deviation, MAE: Mean Absolute Error, MSE: Mean Squared Error, PR: Precision, RMSE: Root Mean Squared Error

to the chest, and the lumbar spine area (10 placements each).

Another 39/73 datasets used an optical camera for MoCap. More specifically, 24 datasets used depth sensor technologies (i.e., an RGBD camera), mainly the 1st and 2nd generation Kinect (https://en.wikipedia.org/wiki/Kinect) by Microsoft (https://www.microsoft.com/). Other input technologies that were used included high-end marker-based optical systems, such as OptiTrack (https://optitrack.com/) and Vicon (https://www.vicon.com/), which were used in eight datasets total. Traditional RGB cameras were used in six datasets, and one dataset used only an Infrared (IR) camera.

d: DATASETS AVAILABILITY

The majority of the datasets used were private and were developed and used only by the respective groups of each publication (60/73). Publicly available datasets comprised 17.8% of the total (13/73). The most commonly used public datasets were the UI-PRMD dataset [121] used in six publications [55], [56], [70], [75], [78], [122], and the KIMORE dataset [123] used in five publications

TABLE 3. Included studies related to HMQA during exercise for sports applications.

Author Year	Application Subdomain	Population	Inertial	RGBD	Optical	RGB	Features Type	Algorithms	Assessment Type	Evaluation Metric	Evaluation Value
Yang (2017) [31]	CS	T	AG				AV	LNR	S	CC	0.70
Bačić (2018) [98]	CS	T			SP		TR	RBFN	O	Acc	0.85-0.95
Bačić et al. (2018) [99]	CS	T			SP		TM	ECF	O	Acc	0.81
Hülsmann et al. (2018) [100]	CS	TC			SPO		TR	SVM	N	F1s	0.65-0.74
Lim et al. (2018) [101]	CS	TT	AG				DR	LSTM	O	F1s	0.93
Hegazy et al. (2021) [102]	CS	TT	AG	SP			DR	RNN	MN	Acc	0.99
Tabrizi et al. (2021) [51]	CS	TT	AGMO				TM	LSTM, SVR	MS	RMSE	3.39
Ghosh et al. (2020) [103]	SL	B	AGM				TM	CNN	O	Acc	0.86
Liu et al. (2020) [104]	SL	C	SO				AV	XGB	S	Acc	0.99
Soekarjo et al. (2018) [105]	SL	KB			SP		DR	k-NN	O	Acc	0.73
Makino et al. (2019) [106]	SL	T	A				TM	LSTM	O	F1s	0.52
Ross et al. (2018) [107]	SL	V			SP		DR	LDA	O	Acc	0.71-0.83
Boyle et al. (2020) [108]	SL	V			SP		DR	SVM	O	Acc	0.59
Ross et al. (2020) [109]	SL	V	AG		SP		DR	LDA	O	F1s	0.77
Brock et al. (2017) [110]	SSJ	S	AGO				TR	CNN	MN	Acc	0.8-0.92
Khan et al. (2020) [3]	SSJ	GF	A				AV	SVM	MS	F1s	0.88-0.98
Lei et al. (2020) [47]	SSJ	D,FS,GV				2DS	TR	SVR	S	CC	0.41-0.52

Application Domain: SSJ: Sports Score Judging, CS: Coaching System, SL: Skill Level Assessment
Population: B: Badminton, C: Canoeing, D: Diving, FS: Figure Skating, GF: Gymnastics Floor, GV: Gymnastics Vault, KB: Kick Boxing, S: Ski, TT: Table Tennis, TC: Tai Chi, T: Tennis, V: Various
Input Modalities: A: accelerometer, G: gyroscope, M: magnetometer, O: orientation, SP: skeletal joint position, SO: skeletal joint orientation, 2DS: 2D skeletal position
Features Type: AV: Aggregate Values Feature Vector, DR: Dimensionality Reduction Feature Vector, TM: Time-series Feature Vectors, TR: Transformed Feature Vectors
Algorithms: CNN: Convolutional Neural Network, ECF: Evolving Clustering Function, k-NN: k-Nearest Neighbors, LDA: Linear Discriminant Analysis, LNR: Linear Regression, LSTM: Long Short-Term Memory, RBFN: Radial Basis Function Network, RNN: Recurrent Neural Network, SVM: Support Vector Machine, SVR: Support Vector Regression, XGB: eXtreme Gradient Boosting (XGBoost)
Assessment Type: MN: Multiple Nominal, MS: Multiple Scores, N: Nominal, O: Ordinal, S: Score
Evaluation Metric: Acc: Accuracy, CC: Correlation Coefficient, F1s: F1-Score, RMSE: Root Mean Squared Error

TABLE 4. Included studies related to HMQA during exercise for wellness applications.

Author Year	Application Subdomain	Inertial	RGBD	RGB	Features Type	Algorithms	Assessment Type	Evaluation Metric	Evaluation Value
Ebert et al. (2017) [29]	MA	AG			AV	RF	O	Acc	0.99-1.0
Leightley et al. (2017) [111]	FA		SP		AV	SVM	MO	Acc	0.94
O'Reilly et al. (2017) [112]	FC	A			AV	RF	N	Acc	0.78
O'Reilly et al. (2017) [113]	MA	A			AV	RF	N	Acc	0.7
Tekriwal et al. (2017) [114]	MA	A			AV	ANN	O	-	-
Hagelbäck et al. (2019) [46]	MA		SP		AV	RF	S	CC	0.83
Jain et al. (2019) [44]	MA			2DS	DR	LSTM	S	MSE	0.12
Dajime et al. (2020) [48]	MA		SP		AV	MLR	O	Acc	0.74-0.85
Hsiao et al. (2020) [115]	FC	AG			DR	LSTM	N	F1s	0.88
Lee et al. (2020) [116]	FC	AG			TM	CNN, LSTM	N	Acc	0.81-0.92
Zhang et al. (2020) [117]	MA			2DS	TM	LSTM	O	Acc	0.97
Hannan et al. (2021) [118]	FC	G			TM	k-NN	N	Acc	0.89
Kim et al. (2021) [39]	MA		SP		TR	ResNet	O	Acc	0.81-0.99
Müller et al. (2021) [119]	FC	AG			AV	MC	N	F1s	0.96
Zhang et al. (2021) [120]	MA			2DS	TM	LSTM	O	Acc	0.96-0.99

Application Domain: FA: Fitness Coach, FC: Functional Assessment, MA: Movement Assessor
Input Modalities: A: accelerometer, G: gyroscope, SP: skeletal joint position, 2DS: 2D skeletal position
Features Type: AV: Aggregate Values Feature Vector, DR: Dimensionality Reduction Feature Vector, TM: Time-series Feature Vectors, TR: Transformed Feature Vectors
Algorithms: ANN: Artificial Neuronal Network, CNN: Convolutional Neural Network, k-NN: k-Nearest Neighbors, LSTM: Long Short-Term Memory, MC: Meta-classifier, MLR: Multiclass Logistic Regression, ResNet: Residual Network, RF: Random Forest, SVM: Support Vector Machine
Assessment Type: N: Nominal, O: Ordinal, S: Score
Evaluation Metric: Acc: Accuracy, CC: Correlation Coefficient, F1s: F1-Score, MSE: Mean Squared Error

[30], [40], [78], [80], [96]. Table 7 provides information about all publicly available datasets, including a reference to the link from which each one can be accessed. Of the publications included in the review, 64/88 used a private dataset, while 22 used public datasets. Another two used both private and public datasets.

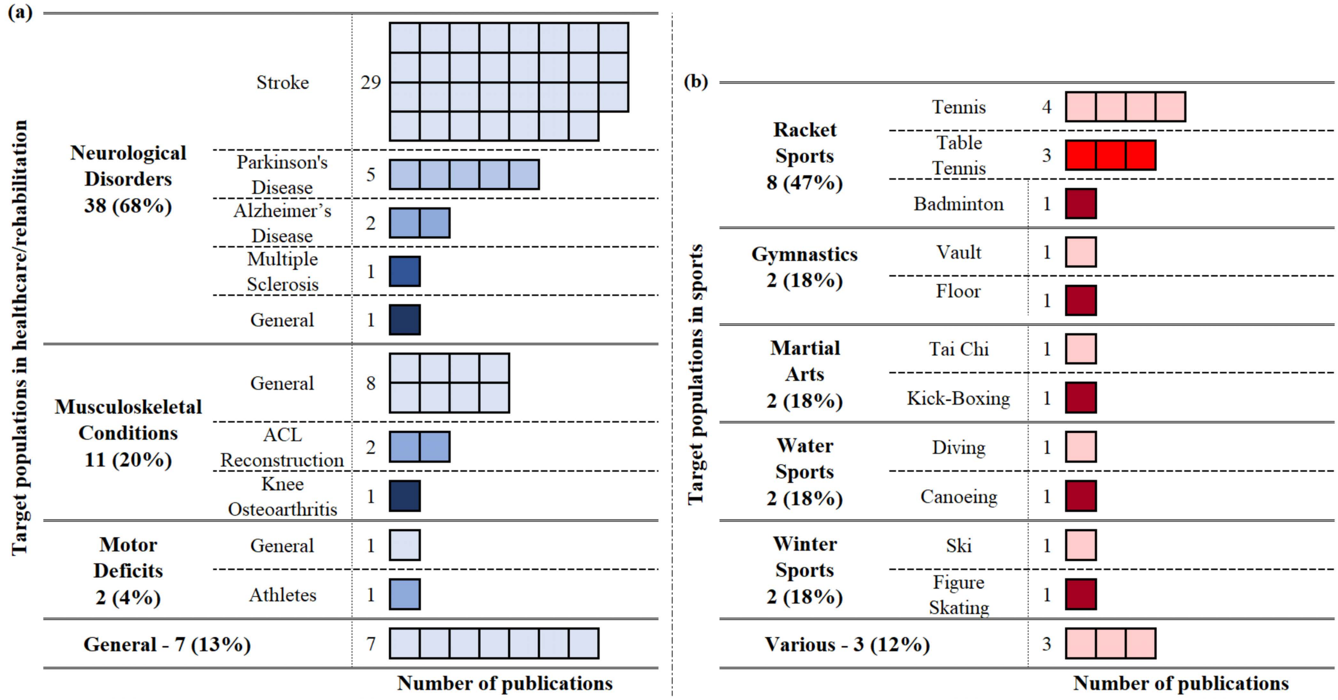

FIGURE 4. Target populations for publications related to healthcare/rehabilitation applications (a), and sports applications (b).

TABLE 5. Datasets per application domain, and study design method used.

	Application Domains			
Dataset Study Design	Healthcare/ Rehabilitation	Sports	Wellness	Total
one-group design	26 (36%)	7 (10%)	13 (18%)	**46 (63%)**
multiple-group design	15 (21%)	10 (14%)	2 (3%)	**27 (37%)**
Total	**41 (56%)**	**17 (23%)**	**15 (21%)**	**73**

TABLE 6. Sample size of datasets per application domain.

					Percentiles	
Metric	Domain	Mdn	Min	Max	25th	75th
Number of participants	Healthcare	21	1	218	9	34
	Sports	10	1	542	3	30
	Wellness	23	5	135	11	50
Number of exercises	Healthcare	3	1	28	1	6
	Sports	3	1	12	2	7
	Wellness	3	1	13	1	5
Number of exercise executions	Healthcare	843	60	35440	282	1685
	Sports	382	14	4769	138	1191
	Wellness	1164	54	8576	163	2058

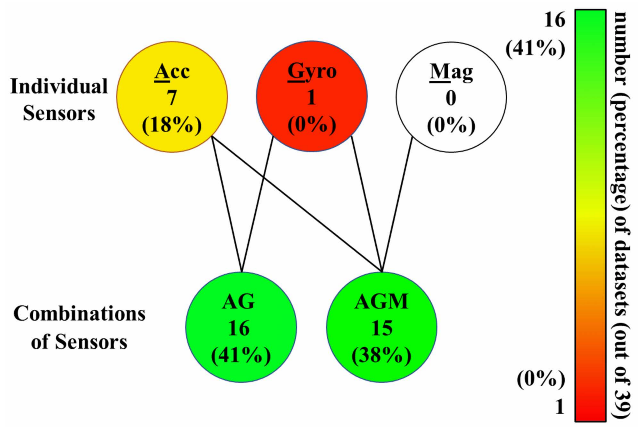

FIGURE 5. Types of inertial sensors used in datasets in the included publications. The colors and displayed numbers indicate the number of datasets that used each type. The lower half indicates the combinations of sensors (i.e., AG = accelerometer and gyro, AGM = accelerometer, gyro, and magnetometer). The percentages are based on 39 datasets that used inertial sensors.

2) PREPROCESSING

Once the data have been collected and based on their modality and the ML approach that will be used (see Section II-A), preprocessing steps might be required before meaningful features can be extracted. The first step is the application of filters to remove noise from sensor values. Of the 88 publications, 29 applied a filter to their data (Table 8), with 16 in the healthcare domain, eight in the sports domain, and five in the wellness domain. The most common filter was the Butterworth filter, which was used in 18 publications (16 low-pass [29], [30], [32], [52], [54], [68], [69], [73], [76], [86], [105], [107], [108], [109], [112], [113], one high-pass [84], and one band-pass [50]), and then the moving average used in five publications [46], [92], [93], [94], [95]. Filters are

TABLE 7. Publicly available datasets.

Name	Input Device	Exercises
Healthcare		
AMIE [34][a]	RGBD	(1) squat, (2) lower lunge, (3) side lunge
IRDS [72][b]	RGBD	bilateral (1-2) elbow flexion, (3-4) shoulder flexion, (5-6) shoulder abduction, (7) shoulder forward elevation, bilateral (8-9) side tap
SBU [124][c]	RGBD	bilateral (1) shoulder abduction/adduction, (2) shoulder horizontal abduction/adduction, (3) shoulder flexion/extension, (4) elbow extension/flexion
TRSP [125][d]	RGBD	(1) reach forward, (2) reach side
Sports		
Hülsmann [100][e]	Optical	(1) squat, (2) Tai Chi push
MIT Olympic Scoring [126][f]	RGB	(1) diving, (2) figure skating, (3) gymnastics vault
UNLV Olympic Scoring [127][g]	RGB	(1) diving, (2) figure skating, (3) gymnastics vault
Tabrizi [128][h]	Inertial	table tennis forehand strokes: (1) topspin, (2) push, (3) basic
Wellness		
Ebert [129][i]	Inertial	(1) crunch, (2) lunge, (3) jumping jack, (4) bicycle crunch, (5) squat, (6) mountain climber, (7) Russian twist, (8) push-up
HSiPu2 [120][j]	RGB	(1) push-up, (2) pull-up, (3) sit-up
K3Da [130][k]	RGBD	(1-2) balance open/closed eyes, (3) chair stand, (4-5) jump low/maximum power, (6-7) one leg balance closed/open eyes, (8) semi tandem balance, (9) tandem balance, (10-11) walk forth and back, (12) timed stand up and go, (13) hopping (one-leg)
KIMORE [123][l]	RGBD	(1) lifting of the arms, (2) lateral tilt of the trunk with the arms in extension, (3) trunk rotation, (4) pelvis rotations on the transverse plane, (5) squat
UI-PRMD [121][m]	Optical and RGBD	(1) deep squat, (2) hurdle step, (3) inline lunge, (4) side Lunge, (5) sit to stand, (6) standing active straight leg raise, (7) standing shoulder abduction, (8) standing shoulder extension, (9) standing shoulder internal-external rotation, (10) standing shoulder scaption

[a] https://dtai.cs.kuleuven.be/software/amie
[b] https://zenodo.org/record/4610859#.YlHw4shBwmI
[c] doi.org/10.6084/m9.figshare.6741713
[d] https://www.kaggle.com/datasets/derekdb/toronto-robot-stroke-posture-dataset
[e] https://pub.uni-bielefeld.de/record/2930611
[f] https://www.csee.umbc.edu/ hpirsiav/quality.html
[g] http://rtis.oit.unlv.edu/datasets.html
[h] https://data.mendeley.com/datasets/b7bc9y232m/3
[i] https://github.com/andrebert/body-weight-exercises
[j] https://github.com/mindayao/HSiPu2
[k] https://filestore.leightley.com/k3da/
[l] https://univpm-my.sharepoint.com/:f:/g/personal/p008099_staff_univpm_it/EiwbKIzk6N9NoJQx4J8aubIBx0o7tIa1XwclWp1NmRkA-w?e=F3jtBk
[m] https://webpages.uidaho.edu/ui-prmd/

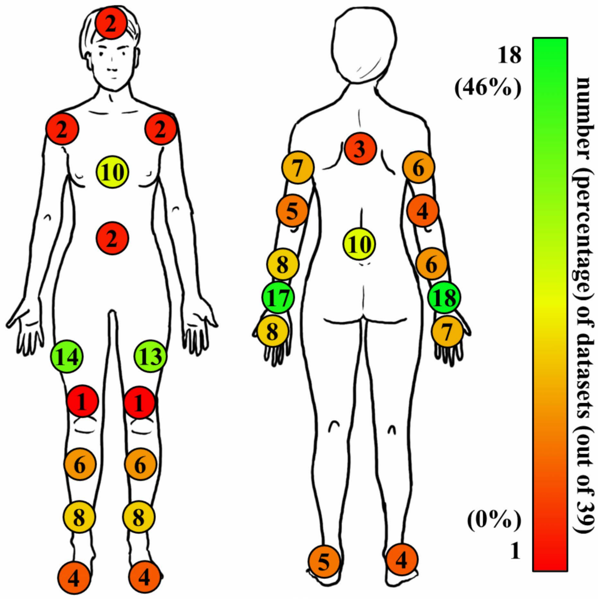

FIGURE 6. Placement of inertial sensors used in datasets in the included publications. The colors and displayed numbers indicate the number of datasets in which a sensor was placed at the location. The percentages are based on 39 datasets that used inertial sensors.

TABLE 8. Filter used per MoCap technology based on the publications that used a filter (percentages are based on 29/88 publications).

MoCap technology	Filter	Number of publications
Inertial	low-pass Butterworth	10 (34%)
	median	2 (7%)
	band-pass Butterworth	1 (3%)
	Kalman	1 (3%)
RGBD	moving average	5 (17%)
	low-pass Butterworth	2 (7%)
	Savitzky-Golay	1 (3%)
High-end Optical	low-pass Butterworth	4 (14%)
	polynomial curve fitting	1 (3%)
RGB	Discrete Cosine Transformation	1 (3%)
IR	high-pass Butterworth	1 (3%)

most commonly applied to data captured from inertial sensors, with 14/39 publications that used inertial sensors applying them [29], [32], [50], [52], [68], [69], [73], [76], [102], [103], [109], [112], [113], [119]. Among the publications that used optical cameras, 8/39 that used RGBD data applied filters [30], [46], [81], [86], [92], [93], [94], [95], and 5/9 of the ones that used high-end optical based systems [54], [98], [105], [107], [108].

Related to the segmentation of data into executions/repetitions, 34/88 used an automated technique to segment them. The other 54/88 used set time windows or triggers when they recorded the data, or manually performed segmentation. Of those that used an automated approach, 16/34 performed segmentation using peak detection of specific sensor values (e.g., the velocity of a specific angle or values from one axis of the accelerometer) [21], [30], [31], [32], [33], [58], [63], [66], [76], [79], [88], [91], [105], [115], [116], [119]. Other approaches included using the Euclidean distance between a body segment and a predefined reference position [34], [102], or through pattern-matching algorithms such as HMM [89], [90], and DTW [100].

Another step that may be desired during preprocessing, especially in non-controlled conditions, is the normalization of the data in either the spatial or the temporal domain. The normalization in the spatial domain relates to and can be applied only to data that represent a human skeleton using the positions of the joints. In such cases, two different types of normalization can be applied.

The first attempts to transform the positions of the joints so that their relative coordinates use a common reference point, which was performed by 12/49 publications of those that included such data [30], [33], [44], [47], [49], [81], [82], [86], [88], [107], [108], [109]. To perform this normalization, a body segments' property (e.g., pelvis height) is selected as reference, and its transformation properties (e.g., position, scale, magnitude) are transformed to match a specific value. Then, the transformation matrix used for the reference body segment is applied to the rest of the joints. The most common reference body segments used in the publications were the head (i.e., based on the subjects' height) [44], [107], [108], [109], the center point between the two shoulders [49], [81], [82], [86], and the pelvis [33], [47], [88].

The other spatial normalization relates to the reference orientation of the values, that is, the direction of the human skeleton. This type of realignment was applied by 17/49 publications that included applicable data representations (i.e., a human skeleton) [33], [45], [46], [47], [48], [49], [60], [61], [81], [82], [85], [86], [88], [107], [108], [109], [110]. Realignment is usually performed by transforming the human skeleton such that the center point between either the shoulders or hips is facing forward.

Another type of normalization is in the temporal domain. More specifically, it is often necessary to transform a given time-series to a specific length. This type of transformation was performed in 35/88 publications. Most publications accomplished this via resampling [131]. Resampling can be performed either by down-sampling longer timer-series, or up-sampling shorter ones using interpolation algorithms [131] (e.g., the most common in the included publications was through the use of bilinear interpolation [35], [77]). This interpolation can be performed naively, or by detecting key events within the movement, and ensuring that these events occur at set time points [21], [100], [110] (e.g., on a cyclic motion reaching the half-point of the movement can be at the half-point of the time-series). Another method used for up-sampling a time-series with a shorter length is by padding the time-series vector at the end with zeros to fill the missing values [44], [122]. Other approaches for temporal normalization include the use of signal analysis algorithms such as DTW or Multi Event-Class Synchronization (MECS) [132].

3) FEATURE EXTRACTION, ENGINEERING, AND SELECTION

The raw features extracted from the captured data are represented as time-series, where each data point characterizes the state of a human body part at a given time point. These raw features are either data values from inertial sensors placed at body parts, as shown in Fig. 6, or the positions and/or orientations of skeletal joints extracted from images recorded by optical cameras using pose estimation algorithms (there are no raw features in the form of images, as they were excluded based on the human-body-centric inclusion criteria in the screening process). The input data modalities used in each publication are listed in Tables 1-4. More specifically, of the 39 publications that used inertial sensors, 34 extracted features from accelerometers, 22 from gyroscopes, and four from magnetometers. Seven of the 39 publications used these values to compute the orientation of the respective joints. Finally, one publication [104] fused values from multiple inertial sensors and computed the orientation of the joints for the full-body human skeleton of a subject. Fifty-four publications used optical-based systems and extracted a full-body human skeleton from the captured data. Thirty-six out of the 54 publications used as features only the position of the joints, while six publications used the orientation of the joints [50], [58], [66], [70], [78], [83]. Another six publications used features of both modalities together [45], [55], [56], [85], [96], [100]. Finally, six publications recorded subjects with conventional RGB cameras and then used pose-recognition algorithms (e.g., OpenPose [133]) to compute a 2D representation of the human skeleton (i.e., the positions of joints were in a 2-dimensional coordinate space) [44], [47], [49], [79], [117], [120].

Of the 88 publications, 72 used the raw features described above directly to train their ML algorithms (Fig. 7). Furthermore, raw features were used to engineer new features that provide additional details related to motion during a performed exercise. Geometric features provide information about the relationship between joints in space using linear algebra operations (e.g., angles and distances). These features were used in 21 publications [34], [45], [46], [54], [59], [60], [61], [62], [68], [76], [84], [85], [89], [90], [92], [93], [94], [95], [99], [111], [118]. Kinematic features, which were used in 20 publications, describe the motion of joints [21], [32], [45], [54], [58], [59], [60], [61], [62], [66], [69], [84], [85], [86], [87], [91], [92], [93], [94], [95]. For example, these can be the velocity, acceleration, or jerkiness of a joint. Jerkiness is the derivative of acceleration and describes a motion's smoothness. Event-related features were used in 12 publications and characterize exercise-specific events [29], [31], [48], [50], [59], [60], [61], [62], [65], [84], [87], [91]. These can include the duration of an exercise or a sub-phase of an exercise (e.g., during the loading and release phases of a serve in tennis [31]). Finally, different signal analysis algorithms were used in ten publications [3], [33], [49], [57], [68], [69],

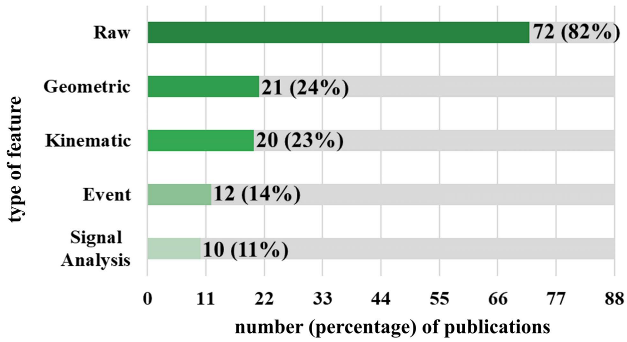

FIGURE 7. Types of features extracted/engineered.

TABLE 9. Modality of the final feature vectors for publications using shallow ML and DL algorithms (indicated percentages are column-wise).

Feature Vector Modality	Shallow ML	DL	Total
Aggregate Values	39 (76%)	5 (14%)	**44 (50%)**
Dimensionality Reduction	7 (14%)	6 (16%)	**13 (15%)**
Time-series	2 (4%)	19 (51%)	**21 (24%)**
Transformed	3 (6%)	7 (19%)	**10 (11%)**
Total	**51 (58%)**	**37 (42%)**	**88**

[76], [88], [97], [104]. Such algorithms can engineer features relating motion trajectories to predefined templates or extract frequency domain characteristics of the signals.

All the features, both the extracted (raw values) and engineered ones, were then used to create feature vectors that were provided as input for the ML algorithms (Table 9). The modalities of the final feature vectors used in each publication are listed in Tables 1-4. The most frequently used method to create feature vectors, used in 44/88 publications, was to apply statistical functions to extract contextual information regarding the performed motion. These functions compute aggregate values that describe a time-series as a whole and are then combined into a single feature vector. This approach was primarily used in publications that implemented shallow ML algorithms (39/44 publications that used the technique implemented shallow ML algorithms (Table 9)). The statistical function used most often by 35/44 publications was the mean function. Other frequently-used functions, with uses in at least ten publications each are maximum (in 24 publications), standard deviation (in 22 publications), minimum (in 19 publications), range (in 18 publications), variance (in 12 publications), and zero-crossings (in 10 publications). Other functions that were used in at least five publications each were the root means square, skewness, and kurtosis, each used in nine publications, the median and interquartile range used in seven publications each, and energy used in six publications.

Publications using DL approaches primarily used input feature vectors created from time-series of raw and/or engineered features. For example, a publication that used the raw values from a triaxial accelerometer, created three feature vectors, one for each axis. Each feature vector includes the values from that axis for each time point of the series. This approach was used in 21/88 publications, of which 19 implemented DL algorithms (Table 9).

Another approach for the creation of feature vectors is to transform the extracted and/or engineered features into a different modality, which was used in 13 publications. Such transformations can provide additional contextual information that is relevant to the target modality or take advantage of existing approaches optimized using this modality. The most frequently used technique in this category was remapping data into a graph-based representation [40], [55], [56], [75]. Each node in such a graph represents the values recorded from a human joint, whereas the edges of the graph represent joints that are connected (e.g., the elbow to the shoulder). Therefore, the graph can provide information about the hierarchy of the human skeleton, starting from the pelvis as the root of the graph and expanding to the extremities (i.e., hands and feet), which represent the leaves of the graph. Similarly, another type of features transformation is into images [39], [47], [86], [110]. The rows of the image represent the different joints, and the columns represent the time points of the time-series. Each pixel in the image stores the triaxial values of the joint at that time point as an RGB value. This approach was popularized because it can benefit from existing approaches used in image analysis that use the same features representation.

Finally, feature vectors were created using dimensionality reduction in 13/88 publications. This approach is similar to the transformation of features; however, the final feature vector can be orders of magnitude smaller. The primary method of dimensionality reduction used in seven of the publications was Principal Component Analysis [66], [68], [105], [107], [108], [109], [115]. Other techniques used include autoencoders [78], [96], discrete wavelet transform [74], and transformation to numeric/word sequences [44], [71], [80].

During the training of ML models, it may be beneficial to select only a subset of the available features, instead of using all of them, to optimize the overall performance of the model. Particularly for systems supporting multiple exercises or full-body skeleton-based input devices, many features may be irrelevant and could reduce the accuracy of the assessment. Therefore, various methods have been used for feature selection from all categories of feature selection techniques (see Section II-B). Of the 88 publications, 19 used feature selection methods to reduce the number of features (Table 10). The most common category of methods used was embedded, with nine publications using it. These studies used ML algorithms such as RF [46], [52], [60], [61], [73], [100], [113], and regression [31], [48] to dynamically select features while training the ML model. Another popular category, with six publications using such algorithms, was filter methods. Primarily, they included statistical techniques, such as Pearson's Correlation and ANOVA, that were used to identify features with the highest correlation to the expected output [49], [69], [70], [84], [91]. Wrapper techniques were used in two publications by applying RL for feature learning [93], [94], and one publication that used forward feature selection [118].

Finally, one publication made use of several techniques from all three categories of feature selection and made the final selection of features based on those that were ranked the highest across all the methods [32].

TABLE 10. Feature selection algorithms from the publications that used one (percentages are based on 19/88 publications).

Feature selection category	Feature selection method	Number of publications
Embedded[a]	Random Forest	7 (37%)
	Regression	2 (11%)
Filter[a]	Statistical	5 (26%)
	Neighborhood Components Analysis	1 (5%)
Wrapper[a]	Reinforcement Learning	2 (11%)
	Forward Feature Selection	1 (5%)

[a]One additional publication made use of 18 different feature selection methods from all three categories and selected features ranked higher by multiple methods

4) MODEL TRAINING

Various ML approaches were proposed in the identified publications for HMQA during exercise. Three different categories of algorithms were explored. Fifty-one publications used shallow ML models, 35 used DL models, and two publications used RL models [93], [94]. Shallow ML algorithms (Fig. 8), can be separated into different subcategories based on their approach. The most commonly used subcategory was ensemble techniques, which combine multiple learning algorithms (primarily decision trees), with 17/51 publications using them [29], [34], [46], [49], [52], [57], [60], [61], [64], [68], [73], [76], [79], [104], [112], [113], [119]. RF was used in 10/17 publications using ensemble techniques [29], [46], [52], [60], [61], [73], [76], [79], [112], [113], and eXtreme Gradient Boosting (XGBoost) was used in three publications [34], [65], [104]. Support Vector Machines (SVM) were also widely used, with 14/51 publications applying them, for both classification [3], [21], [31], [32], [33], [36], [63], [88], [89], [90], [100], [108], [111] and regression [3], [47]. Other algorithms that were used include k-Nearest Neighbours (k-NN), which was used in four publications [71], [97], [105], [118], Logistic Regression used in three publications [48], [54], [59], and variations of Artificial Neural Networks (ANN) used in five publications [58], [87], [95], [98], [114]. There were also hybrid approaches that combined different algorithms into a hierarchical structure. These include the use of k-NN with RF [91], SVM with an Ensemble classifier [66], and ANN with a Decision Tree and a Threshold Model [92].

DL algorithms (Fig. 9), which were used in 35 publications, can be separated into discriminative and generative algorithms [134]. Discriminative algorithms, given an input x and a label y, attempt to learn the probability of the output label y based on the given input directly. Generative algorithms, on the other hand, try to understand how a dataset is generated, and therefore learn the joint probability of a pair (x, y) and how it can be used to predict the output label y [134]. The vast majority, with 31 publications using them, were the discriminative algorithms. Of these, 12 publications used variations of CNN. More specifically, six publications used traditional CNNs [30], [35], [77], [86], [103], [110], two publications used Graph CNNs [55], [56]. Dynamic CNN [84], Temporal CNN (Res-TCN) [72], Residual Network (ResNet) [39], and a combination of a Genetic Algorithm and a CNN [80] were all used in one publication each. The other big category of algorithms used under discriminative models was Recurrent Neural Networks (RNN), which were used in 11/31 publications. Long Short-Term Memory (LSTM) RNN, in particular, were used in nine publications, making it the most used ML algorithm in the publications [44], [51], [74], [81], [83], [85], [101], [106], [115]. A traditional RNN [102], and a Gated Recurrent Unity Network (GRU) [45] were used in one publication each. CNNs and RNNs were used together in eight publications, seven of which used a CNN with an LSTM [40], [75], [78], [96], [116], [117], [120], and another used a CNN with an RNN [67]. Finally, four publications used generative DL algorithms. The algorithms used were Generative Adversarial Network (GAN) [70], Self-Organizing Map (SOM) [65], Transformer [82], and a combination of a Feed-forward neural network, SOM, and Nonlinear AutoRegressive Neural Network (NAR) [50]. Moreover, two publications used RL algorithms for their assessment, using a combination of Deep Q-network with Double Q-Learning (DQ), Markov Decision Process (MDP), and Artificial Neural Network [93], [94].

As it was stated in Section II-C1, motion assessment can be performed through regression, rank-based classification, and parametric assessment. Of the included publications, the majority (42/88 publications) used rank-based classification, whereas 15 used regression. The remaining 31/88 publications used parametric-based assessment, of which 27 were based on classification and the other four regression. Fig. 10 shows the type of assessment used in the publications. Of the ones that used classification, most (21/88) used binary classification [36], [39], [49], [55], [59], [60], [61], [65], [66], [71], [72], [79], [84], [101], [102], [105], [106], [111], [114], [117], [120]; that is, datasets were assessed using a dichotomous variable (e.g., normal or abnormal motion). Other often-used types of assessments were using 3-point ordinal scale values evaluated from rank-based classification approaches [35], [48], [54], [62], [63], [64], [68], [77], [83], [86], [87], [97], [98], [99], [103], [107], [108], [109], and score-based assessments evaluated using regression methods [31], [40], [44], [45], [46], [47], [56], [58], [67], [70], [74], [75], [78], [80], [85], [96], [104]. Of the parametric approaches, the most commonly used was the classification to a set of nominally-labeled variables, where each of the labels described a different error that can occur during a motion execution [34], [52], [73], [81], [82], [91], [100], [112], [113], [115], [116], [118], [119]. Of the publications using this approach, ten assessed the motion using the label that had the highest probability, and another five publications returned all possible

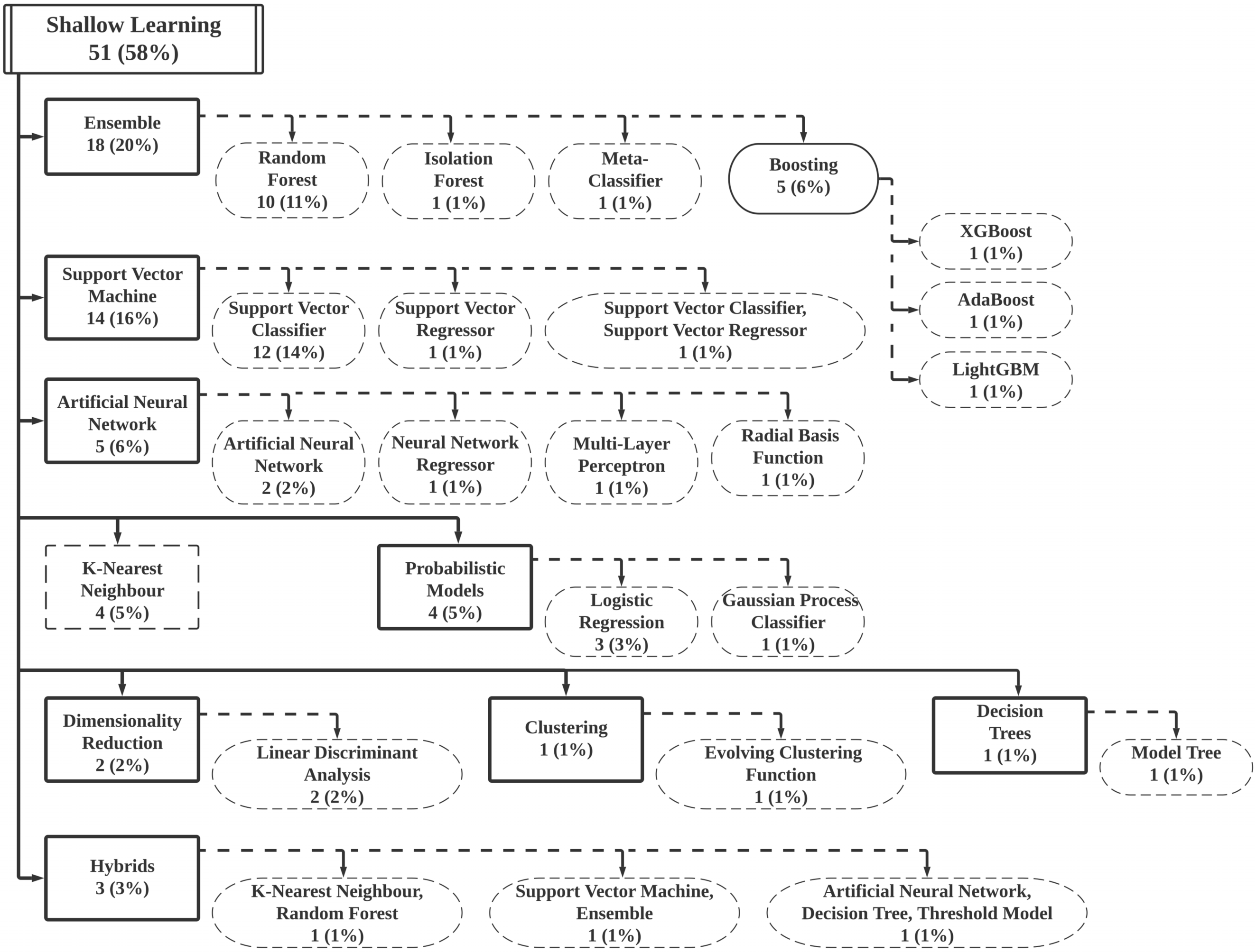

FIGURE 8. Taxonomy of shallow ML algorithms for HMQA used in the reviewed publications. Numbers indicate the number of publications that used a category of algorithms (rectangular shapes) or a specific algorithm (shapes with dashed lines). The percentages are based on the total number of publications (i.e., 88).

labels that were identified. Other publications in this category provided assessments using either multiple nominal variables [21], [33], [76], [88], [89], [90], [110] or multiple scores [3], [51], providing multiple specific assessments of the domain-specific characteristics of the performed motions.

5) MODEL VALIDATION, AND TESTING

Related to the testing and validation of the algorithms, more than one-third (35/88) of the publications only used holdout validation, separating their datasets into either two subsets, one for training and one for testing, or three subsets for training, validation, and testing. The remaining 53/88 publications used different types of cross-validation to iteratively validate their models using different subsets from their datasets. Of these, 32/53 publications used k-fold cross-validation, with most splitting their datasets into 10 random subsets (i.e., 10-fold cross-validation). Another 20 publications used leave-one-out cross-validation, where usually the data of one of the subjects were left out for validation and the remaining data were used for training, repeating for all subjects. One publication used stratified shuffle split cross-validation.

Various metrics can be used to evaluate how well a proposed model assesses actual data. These metrics depend on the type of assessment that the model performs, with classification and regression algorithms requiring different metrics. The primary evaluation metric and evaluation results of each publication are listed in Tables 1-4. Of the included publications that used classification-based algorithms, 60/69 primarily evaluated their results using the accuracy of the model, which assessed the number of correctly classified patterns of each class. Other metrics used, included recall (or sensitivity) used in 22/69 publications, F1-score used in 21 publications, precision used in 13 publications, and specificity used in ten publications.

The variety in the metrics used to evaluate ML algorithms makes the comparison between different approaches challenging. Fig. 11 shows a box and whisker plot

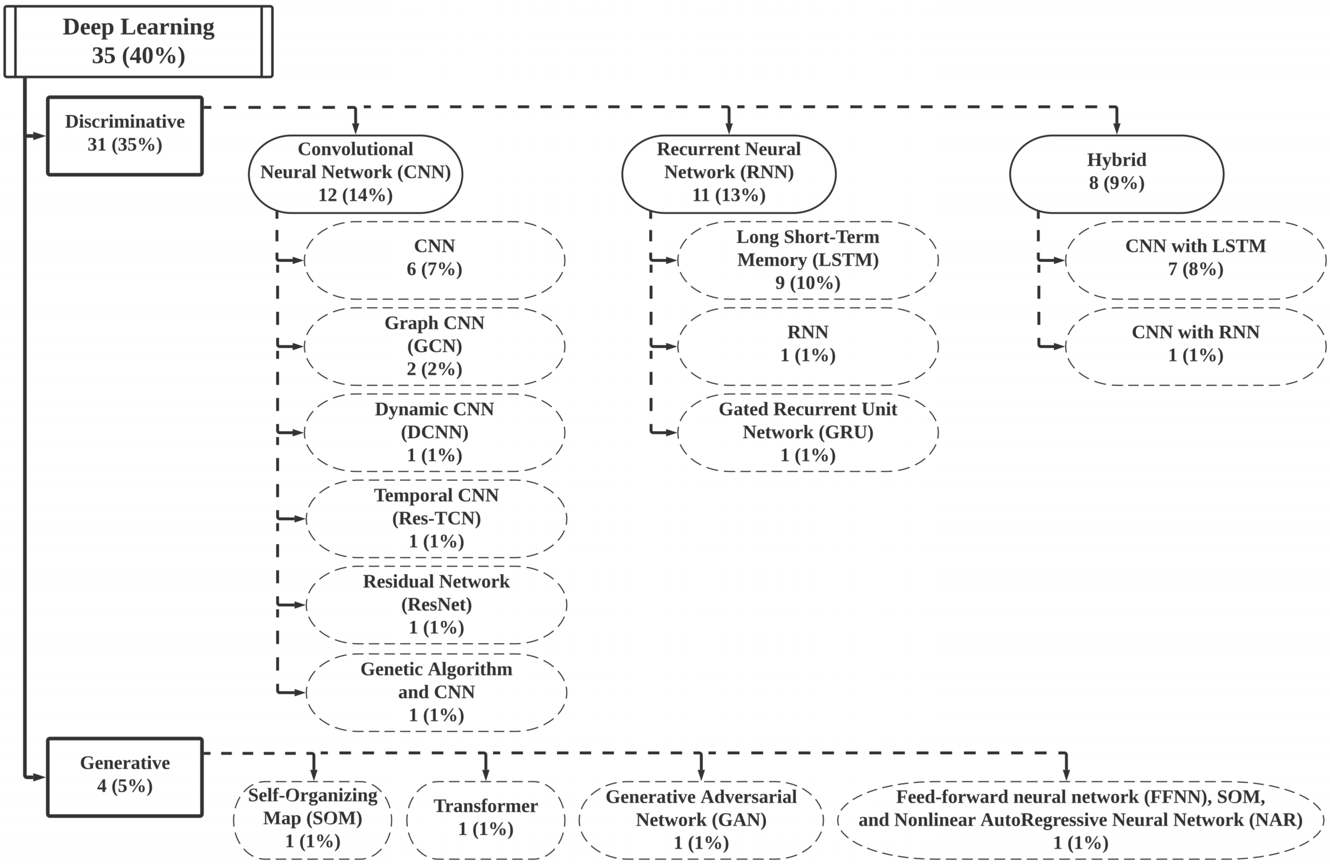

FIGURE 9. **Taxonomy of the DL algorithms for HMQA used in the reviewed publications. Numbers indicate the number of publications that used a category of algorithms (rectangular shapes) or a specific algorithm (shapes with dashed lines). The percentages are based on the total number of publications (i.e., 88).**

with the reported accuracies of the algorithms used for classification-based assessments from the 60 publications that used the accuracy metric. The plot displays only algorithms that were used in at least two publications and separates the shallow and DL approaches. The algorithms are presented in order based on their median values. Of the shallow algorithms, Boosting approaches had the highest median value of 95.56% (25th percentile = 75%, 75th percentile = 98.91%). Following are the ANN, SVM, and k-NN algorithms. Regarding DL algorithms, the combination of CNNs with RNNs had the highest reported median accuracy of 97.27% (25th percentile = 97.14%, 75th percentile = 97.49%), but using a very small sample of just two publications. RNNs followed after that with median accuracy comparable to that of Boosting approaches, with a median value of 95.5% (25th percentile = 89.64%, 75th percentile = 96.69%). CNNs had the lowest median accuracy values among DL algorithms.

Regarding the publications that used regression algorithms, the results were reported using different metrics. In total, 23 publications reported regression-based results. Five out of the 23 publications reported their results using the correlation coefficient, four used Root Mean Square Error (RMSE), and Mean Squared Error (MSE) and Mean Absolute Error (MAE) were used in three publications each. Results were also reported using a Mean Absolute Percentage Error (MAPE), Mean of Absolute Deviations (MAD), separation degree, confidence of separation, and Sum of Absolute Deviations (SAD). The small number of instances per metric does not allow meaningful reporting of any aggregate values for these metrics.

Overall, the evaluation of the algorithms across different applications, populations, input data modalities, and assessment types provided relatively consistent results. However, there were some noteworthy exceptions. For example, although ensemble techniques had high accuracies across most cases and configurations, their accuracy was lower when dealing with classification into multiple (i.e., more than 3-4) different categories [34], [52], [73], [76], [112], [113]. On the contrary, SVM algorithms were evaluated with higher accuracies when performing classification to dichotomous values (i.e., binary assessments) [32], [36], [66], [111].

In addition to the above metrics, nine publications reported the time required for a trained model to assess new data [3], [34], [46], [83], [90], [101], [102], [111], [118]. These times ranged from a few milliseconds to several seconds.

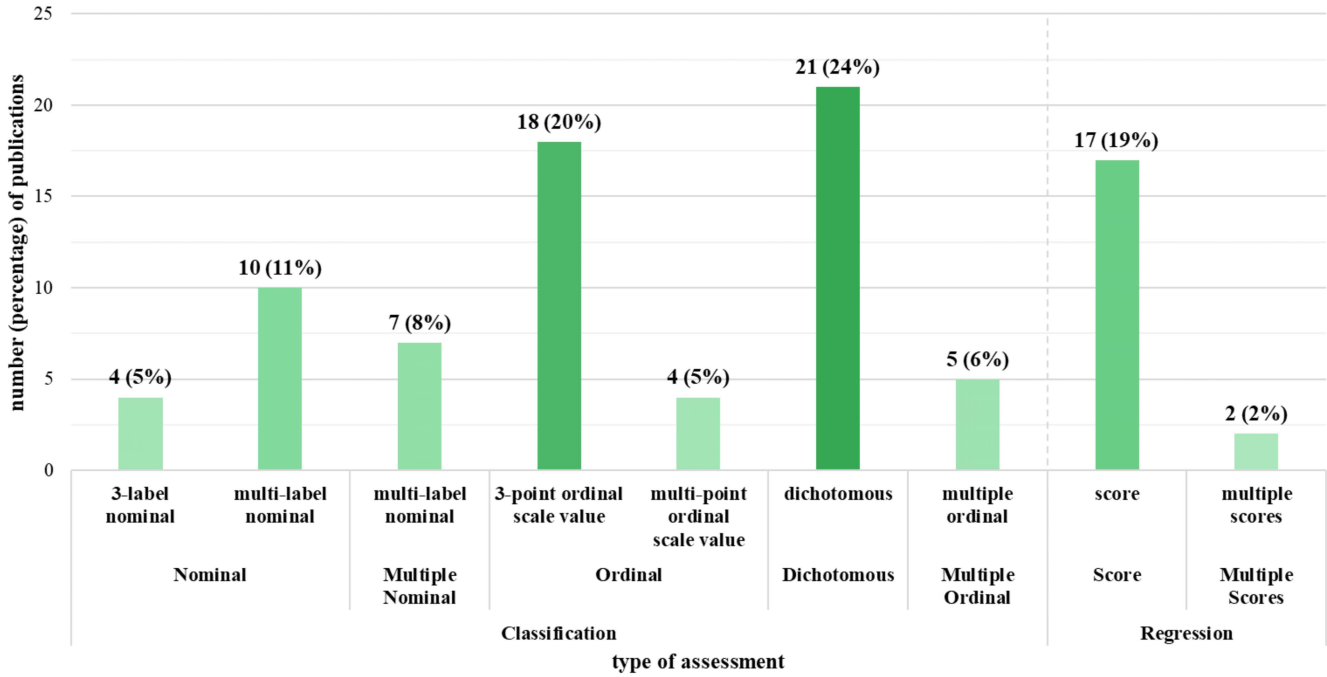

FIGURE 10. Types of assessment used in the publications, and the types of variables used for the assessment.

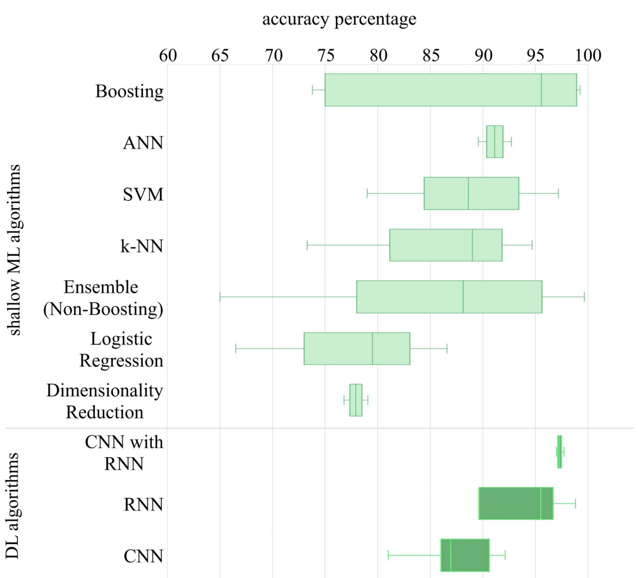

FIGURE 11. Box and whisker plot of the accuracies of ML algorithms of publications that used classification and the 'accuracy metric'. The boxes include the 25th to 75th percentiles range and the line indicates the median value. Listed are algorithms used by at least two publications each. The algorithms are separated into shallow ML and DL approaches and sorted based on the median value of accuracy.

V. DISCUSSION

In this study, the current state of research was explored in the field of HMQA during exercise using ML. In the last five years, there have been several publications in this area, which is indicative of high research interest and its applicability in different domains. Public interest has also been high, with various commercial applications using relevant technologies, including mobile-based [135], [136], [137] and standalone applications. For example, Tempo Studio [138] (Tempo, United States) is a home gym system and virtual trainer, that uses Azure Kinect DK (Microsoft, United States) to capture the motion of users during exercise, analyze it using ML, and provide corrective feedback. In the following subsections, the research questions posed at the end of Section I-A will be addressed. More specifically, we discuss the requirements for ML algorithms used for HMQA during exercise (subsection V-A), what challenges they face (subsection V-B), and finally, provide some guidelines for developing them (subsection V-C).

A. RQ1. REQUIREMENTS FOR ML ALGORITHMS USED FOR HMQA DURING EXERCISE

According to the identified publications, HMQA during exercise is applicable in three application domains: healthcare/rehabilitation, sports, and wellness. The three domains introduce several requirements that a system for HMQA should have. Some are unique to each domain based on the characteristics of the target populations and the intended use of the systems, and others are common across all three domains. The main requirements derived from the applications included in the publications are presented next.

1) INFORMATIVE ASSESSMENTS

The main purpose of all applications using HMQA during exercise is the analysis of the performed motions and extraction of meaningful information. For example, in the healthcare domain, therapists often perform functional assessments of their patients (Table 1). Thus, they require a detailed image of the patient's performance during the various exercises. Moreover, applications designed for patients, such as virtual therapists or gamified experiences (Table 2), must generate corrective feedback based on how they perform. The World Health Organization (WHO) also highlights the need for research in the development of telerehabilitation services so that everyone can have access to the same care, as part of its Rehabilitation 2030 initiative [139]. Similarly in the other two domains, athletes and trainees expect feedback on their exercise execution from coaching systems to improve their technique (Tables 3 and 4).

In general, research has shown that providing feedback can positively affect motor learning [140]. Visual feedback can improve physical reach and exercise performance in general [141]. This is achieved by providing a more engaging experience during exercise and by immersing users in the activity [142]. Feedback can also help with the psychological state of users by improving their motivation and mood [143], [144]. However, to accomplish this the feedback generation component of such systems requires receiving a well-structured assessment from the HMQA component, which is both of high quality and provides comprehensive information. This means that the assessment should provide quantitative metrics of the quality of motion during exercise and any erroneous characteristics. This can help with the generation of appropriate feedback that is easy for the user to understand [142].

2) EXPLAINABILITY AND INTERPRETABILITY

Expanding on the previous section, assessments beyond being informative should also be intuitive. The logic between the performed exercise data and the end assessment itself should be easily understood by users [100]. In ML, explainability signifies the ability to explain *how* an algorithm works internally [145]. Interpretability, on the other hand, provides an understanding of *why* a specific outcome is observed and how it will change based on different inputs [145]. Both of these principles should be considered to ensure that the end user of such systems can understand both how and why an assessment was made. Research has shown that the lack of transparency in such systems prevents users from trusting them, and therefore are hesitant to adopt them [146]. Thus, more recent studies have explored how to develop systems whose decision-making processes are easier to understand. In one such example [94], users can manually select the features to be considered by the assessment algorithm, and the system can visualize how these affect the assessment outcome. Extending the above study, the same authors evaluated their system with therapists [93] by comparing it with the traditional assessment methods. Their results revealed significant differences across various metrics related to the usefulness of the explainability of the system and how it can better assist therapists in monitoring patients. In another study, Hülsmann *et al.* [100] developed a dynamic visualization method within a Virtual Reality-based coaching system to improve the interpretability of their assessments. Their solution, which identified several error patterns that can occur during squats, provided users with easy-to-understand feedback based on each error. However, the visualizations were manually assigned to each of the errors and were not extracted directly from the ML model. In another approach, Leightley *et al.* [111] tried to provide pinpointed information about the assessment outcome and why it was made by grouping different joints together (i.e., each arm, and leg bilaterally), and analyzing each group. To accomplish this, however, they developed a separate SVM model for each group of joints, instead of using a single unified model.

3) REAL-TIME AND CONCURRENT ASSESSMENTS

As noted above, generating informative assessments can be used to provide appropriate feedback in various types of systems, including virtual trainers in the wellness domain [112], [115], [116], [118], [119], and coaching systems for sports [31], [51], [98], [99], [100], [101], [102]. Similarly, in rehabilitation applications such as virtual therapists [21], [33], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97] and serious games [45], [50], [84], [85], [86], [87] patients need to be informed their performance. However, this type of feedback should also be provided in real-time to improve the effectiveness of such systems. This was highlighted in a user study of a serious game used for rehabilitation [45]. Therapists who participated in the evaluation underlined the importance of providing real-time assessments while monitoring exercise execution. Therefore, HMQA algorithms must be designed to provide real-time assessments that can be used for immediate feedback generation. It should also be considered that in such systems, assessment is only part of the feedback generation process. For example, Decroos *et al.* [34] reported in their results that the time required to provide a user with feedback using their system was 0.28s. However, the majority of the time was used for data transfer, segmentation of the data, and feature extraction, whereas the actual assessment took less than a millisecond.

Feedback can be classified as intrinsic, which is done by providing sensory-perceptual information, or extrinsic, which is provided by an external source (i.e., a coach/trainer or a therapist), in the form of verbal remarks [147]. Extrinsic feedback can either provide information related to the outcome of an activity, referred to as knowledge of results feedback, or information about the characteristics of the activity, referred to as knowledge of performance feedback [147], [148]. As such, knowledge of results feedback is terminal, meaning it is given after the execution, while knowledge of performance can also be concurrent and provide real-time feedback during the execution of the activity [147]. In general, the feedback provided using knowledge of performance is more effective than knowledge of results [140], [149]. The dynamic nature of exercising further underscores the importance of providing feedback. Therefore, it is worth exploring ways to provide concurrent assessments during the execution of an exercise, not just at the end of the exercise.

4) ADAPTABILITY AND INTERCHANGEABILITY OF INPUT MODALITIES

An important factor related to the population that often uses such systems is access to MoCap technologies and their usability. In particular, in healthcare and rehabilitation domains, the cost of such assistive systems should be considered. Patients are usually incurred with the costs of healthcare services, and the high cost of rehabilitation assistive systems could be a prohibitive factor for their use [150], [151], [152], [153], [154]. This suggests that the use of low-cost MoCap technologies, such as inertial sensors and low-cost RGB(D) cameras, may be preferred. Algorithms for HMQA will thus have to operate under these conditions and adapt based on the type of MoCap technology available in a specific application. It is therefore important to study how the reduction in the number of inertial sensors used affects the accuracy of the algorithms for HMQA during exercise [31], [52], [73], [112], [113], [116], [119], as well as the use of more cost-effective alternatives, such as conventional RGB cameras with pose recognition algorithms [44], [47], [49], [79], [117], [120].

At the same time, the use of the systems by elderly populations, which are less familiar with the technologies, can make the use of wearables and other devices more difficult [115] (e.g., through the misplacement of sensors that can introduce noise in the data). HMQA algorithms should be able to manage such occurrences, adapt, and make consistent assessments, regardless of the setup of the MoCap system.

5) PERSONALIZATION

Personalization is another characteristic ML algorithms for HMQA during exercise should consider. This term means that the algorithm should be able to consider the unique physiological characteristics of the user who is performing an exercise [155]. Personalization is especially important in the healthcare domain, where subjects, beyond their primary condition or impairment, can have other co-morbidities that could prevent them from being easily matched to more generic models [94].

B. RQ2. CHALLENGES FOR ML ALGORITHMS USED FOR HMQA DURING EXERCISE

Through this review, various challenges of ML approaches for HMQA during exercise have been analyzed.

1) SMALL NUMBER OF DATA SAMPLES IN DATASETS, AND BIASES

The accuracy and performance of ML algorithms, particularly for the analysis of human movement, are heavily dependent on the data used for training and validation [156]. Unlike other fields where established datasets exist to train ML models and in agreement with previous studies [157], it was shown that there is a lack of data related to the assessment of the quality of motion during exercise (Table 6). In addition, the small size of the datasets results in few samples per exercise, and in the case of datasets with multiple classes (e.g., for different qualities of motion and types of erroneous movement), there are often only a few available samples per class. This leads to unbalanced classes in the datasets [91]. These two characteristics often reduce the scope of proposed solutions. For example, Whitford *et al.* [79], because of a lack of data, merged the data of three different types of errors into a single incorrect category, instead of using all three error types separately and provide a finer assessment.

Similar to the use of ML in other areas, another challenge that can be even greater because of the small number of samples, is the introduction of biases because of the unbalanced nature of the training data [90], [119], [158]. This can be especially true when analyzing human motion, as people have different characteristics affected by medical conditions or other personal traits. However, limited research has been conducted on this topic. For example, Albert *et al.* [30], studied neurological disorders and were able to identify when exercises were performed by healthy subjects, people with Parkinson's Disease, and stroke survivors. In another study, Kianifar *et al.* [32], showed that different biomechanical characteristics between sexes can affect the relevant features of the performed exercises. Therefore, the dataset composition can directly affect the accuracy of the developed models.

2) LACK OF LARGE, OPEN DATASETS

In addition to the previous challenge, there are a limited number of publicly available datasets (see Table 7). Moreover, the available datasets have relatively small sample sizes. The median number of participants in these datasets is 19 (25th percentile = 12, 75th percentile = 42) and the median number of samples is 580 (25th percentile = 216, 75th percentile = 1950). This mandates the collection of new data before ML algorithms can be developed. However, this can be a difficult task, particularly in the healthcare domain, where access to specific populations can be difficult [30]. Overall, the limitation of available open datasets leads to difficulty in developing new and better approaches, as well as in comparing the accuracy of the ones proposed.

Moreover, most of the available data are often collected in very controlled conditions within laboratories, and thus might not necessarily be representative of data recorded in a natural setting, such as in a home rehabilitation or home training session. Another characteristic of the publicly available datasets that was identified, is the prevalence of optical-based MoCap technologies. Of the 13 datasets identified, only two captured data using inertial sensors [128], [129]. Others mostly used depth-based cameras (e.g., Kinect), conventional RGB cameras, or high-end marker-based optical cameras. This is a clear limitation, especially in the sports domain, where motion assessment is often performed in the field, an environment that is more friendly to the use of wearables with inertial sensors [159], [160].

3) EXERCISES CHARACTERISTICS

One characteristic of motion data from exercising involves possible variations in the execution of activities, also known as intra-class variations [21], [100]. In other words, an exercise can be performed in different ways, and characteristics of the users could introduce execution variations. As a result, the accuracy of the assessments may decrease, even if the executions are correct. This occurs especially in the case of rehabilitation, as mentioned above, in which patients often have different functional abilities based on their disability. In addition, data from exercises may also display inter-class similarities [161]. Thus, motion data from an erroneous exercise execution may only have subtle differences compared with a correct exercise execution, making it difficult to differentiate between the two [112].

An additional challenge is the differences in the duration between the executions of an exercise [51], [77], [100], [132], which could prevent the use of certain approaches. The variability in the temporal domain also extends to the spatial domain, where some errors can be found at different time points, or even across multiple joints [100], making it difficult to identify them. These problems become even more challenging when coupled with other external conditions that can affect the data, such as noise in the signals and occlusion of limbs in vision-based approaches [77]. These problems are inherent to motion data, and exercise in particular, and different approaches should be explored, both relating to feature engineering and the implementation of ML algorithms, to minimize their impact.

Finally, exercises are often complex motions where they are performed by the synergies of different body joints and include several phases. Most of the included publications focused on simple motions, where only a single joint was responsible for movement. More complex motions are usually analyzed by first decomposing them into more basic motions, followed by more traditional approaches. This is often implemented using HMM [89], [90]. However, as assessments are performed in more areas, the need for such solutions will also increase. Therefore, a study of other approaches to achieve this can be beneficial for future scalability.

C. SUGGESTED GUIDELINES FOR DEVELOPING ML SOLUTIONS FOR HMQA DURING EXERCISE

Next, a set of general guidelines is proposed to be considered when developing ML solutions for HMQA during exercise. The guidelines stem from approaches adapted in the reviewed publications and relate to the requirements and challenges outlined above. The guidelines cover the full design and development of such solutions, from the conception of the solution to the implementation and final evaluation of the algorithms. A summary of these guidelines is provided in Table 11.

TABLE 11. Summary of guidelines for the development of ML solutions for HMQA during exercise.

Topic	Guidelines
Selecting type of Assessment	• generate informative assessments that will assist the end-user • selection between classification, regression, and parametric • selection between single and multiple outcomes • full body, or separation or grouping of joints
Handling data imbalances	• selection of algorithms that work well with small samples • transfer learning to take advantage of models trained using other data • data augmentation to generate new samples
Selection of Mo-Cap Technologies	• optical: easy to setup, non-invasive, possible occlusions (may consider multiple synchronized cameras), lower polling rate, longer time to process high-quality image • inertial: ubiquitous technology, minimally invasive, may require strict placement, drift problems • use of both technologies through data fusion
Balance between accuracy and speed	• further exploration of GRU RNN as a lightweight DL algorithm • data preprocessing and feature engineering on a needs-basis • use of feature selection to minimize used features
Selecting the right ML Algorithm	• combination of different algorithms may provide improved results • RNN with CNN implementations and Ensemble algorithms have shown higher performances
Concurrent Assessment during Exercise	• generate full exercise execution time-series from partial data (online) using data augmentation algorithms
Personalization and Scalability	• development of personalized models is desirable but can be challenging to achieve • use of generic models with online customization options • use of reinforcement learning • development of robust models easier to scale, or to transfer knowledge to new domains
Evaluation	• use of multiple evaluation metrics for presenting results • use of proper cross-validation techniques based on available data

1) SELECTING TYPE OF ASSESSMENT

The first step in designing a ML algorithm is to select the type of assessment that the algorithm will perform. This includes whether the assessment will be classification-based or regression-based and whether it will include a single or multiple output values. In the case of HMQA, this decision should be guided based on the information that the end-user needs to receive. The requirement to provide informative assessments highlights the essential role feedback can play in applications that use HMQA during exercise, and how that can affect the type of assessment. For example, the assessment can relate to biomechanical characteristics, such as the deviation of the joints' angle or the distance between two joints [92]. Therefore, a dichotomous classification of correct or incorrect motions based on a threshold, or a score from a regression-based assessment may be preferred. It can also be used to describe the quality of the performed motion based on the flow of motion using motion trajectories [21], [101], etc. Parametric-based variables from the HMQA can also provide domain-specific knowledge of performance feedback. For example, Tabrizi *et al.* [51] assessed forehand strokes in table tennis based on different characteristics (e.g., forward swing, follow-through, and appropriate speed of the racket movement), that required multiple outputs for the assessment. Several studies have also used classification to identify various errors during exercise. In one such publication, Wei *et al.* [90], as part of a virtual physical therapist application, identified errors patients performed during a forward lunge (e.g., bending of the back knee, not keeping the front shank vertical, etc.). Therefore, these characteristics and the analysis of the expected outcome of the targeted applications can help to define the proper assessment result types that can generate the required information.

The majority of the proposed solutions performed HMQA using classification into different classes, with only 19 publications using regression. In general, this approach is less scalable, as there is a limited number of classes that can be represented while ensuring the efficient and real-time assessment of data. On the other hand, regression-based approaches can identify more subtle changes in the performed motions and thus provide more specific feedback on their execution [46]. This, of course, could also be perceived as a limitation in some cases, especially when dealing with more complex motions, since a single score may be insufficient to provide sufficient context about the performed quality of the exercise. However, variations of the above are also noteworthy. For example, some studies [3], [51] used regression to assess the multiple domain-specific characteristics of an exercise. This allows the capture of subtle deviations in the execution of exercises across different relevant features and providing a more informative assessment.

Another hybrid approach is the use of classification for exercises into different classes, and then using the confidence scores from the classification to generate a regression-based assessment [3], [56], [94], [162]. Thus, the training set includes clusters of the correct execution and one of each error pattern. To evaluate a new exercise execution, the cluster to which the execution is closest is computed. Finally, the confidence scores for each of the clusters (i.e., the probability that the execution belongs to a cluster) are used with a regression-based algorithm. Lim *et al.* [101], on the other hand, first used a classifier to identify a motion based on each class, and then augmented the classifier using a probabilistic inference model, which then computed the latent trajectory of each class that was given back to the user as feedback. Further research should be performed on such methods, to ensure that the assessment can provide sufficient contextual information.

In addition, relevant is the decision whether the assessment should characterize the full human body, or whether it is prudent to separate individual joints or group joints together (i.e., upper limbs, lower limbs, torso) [163]. Such assessments can identify erroneous patterns missed by other approaches. However, they result in a more complex assessment model, risking having a slower assessment. At the same time, they are unable to identify errors related to movement dynamics, such as synergies between different body parts, or other spatio-temporal characteristics [164]. Overall, both the type of assessment and the reference body for the assessment can affect how the results of HMQA are understood by the end-users. Therefore, careful consideration must be made during the development of any HMQA algorithm to ensure its effective use.

2) HANDLING DATA IMBALANCES

When collecting data, limited access to the target population, and challenges with data capturing can often lead to data imbalances in the recorded datasets. Various solutions have been proposed to address this problem. For example, some have tried using algorithms that can work well with a relatively small set of samples [68], [94], [98]; however, this usually limits the type of assessments that can be derived. Other approaches augment existing datasets by adding variations based on previously recorded values or by generating new samples using GANs [30], [70], [109]. Khoramdel *et al.* [82] used a modified version of the focal loss function when training with their dataset to overcome the problem. Finally, some research has been performed using transfer learning, by utilizing existing models trained with other types of data, and then using them in the relevant domain(s) [30], [39].

3) SELECTION OF MoCap TECHNOLOGIES

Capturing motion data can be achieved using various technologies. Optical-based systems, both RGB-based and depth-based (i.e., RGBD-based), are easy to set up, and non-invasive, with no need to wear any sensors. Their main limitation is their static position, which limits the view angle, making them prone to occlusion, especially when dealing with complex motions [21], [77], [86], [104]. Several approaches have been developed toward using multiple synchronized cameras. This is implemented in high-end marker-based optical systems [54], [66], [70], [98], [99], [100], [105], [107], [108], as well as by using conventional cameras [60], [61], [117], [120]. However, this may lead to a longer preprocessing time to extract a single skeletal representation of the user. Depth-based optical systems are also often limited by the available polling rate of the data they record. For example, the average polling rate of the RGBD devices in the included publications was 30 frames per second, compared to inertial-based sensors, which had an average polling rate of 83 recordings per second. Moreover, the recorded data are high-resolution images and thus require more processing power for analysis. This often leads to video frames being skipped to improve resource management and achieve real-time performance [5].

With the expansion of wearables, inertial sensors, have become a common go-to solution for exercise tracking as a minimally invasive device [142], [160], [165], [166], [167]. Common hardware devices used in such cases include personal smartphones, smart-watches, custom-made sleeves, or other commercial kits. However, these solutions often require strict placement of devices at specific positions and orientations, making them less robust [168], even though approaches have been studied to overcome such problems [169], [170]. Inertial sensors also suffer from drift problems that need to be compensated during preprocessing to ensure that the recorded data are meaningful [171], [172].

Limited work has been conducted to compare algorithms using different input data modalities. Both optical-based and inertial devices have their advantages and disadvantages; however, their usage has mostly been driven by usability factors. The two technologies were compared in two studies [109]. Both studies showed that inertial data may provide a higher assessment accuracy. However, the comparisons had shortcomings, with one using simulated inertial data derived from skeletal data recorded using an optical system [109], and the other comparing inertial data with 2D skeletal data computed from an RGB camera [79]. In the former of the two, Ross *et al.* [109], however, also noted that despite having higher accuracies, feedback generated from the inertial data was harder to interpret and provide meaningful feedback that could lead to improvements during exercise. In addition, the fusion of data from both technologies has been used in some publications [50], [58], [79], [83], [102]. A comparison between the use of fused data and separate input modalities has shown higher accuracy from algorithms trained using fused data [79], [173].

Some studies have worked toward designing HMQA models to work across different technologies. This can be achieved by transforming the input data into a common format, regardless of the type of raw input data. One such format is a representation that uses joint angles. For example, Decroos *et al.* [34] transformed the skeletal joint positions into joint angles and used those for the assessment. Similarly, joint angles can be computed from inertial data using various filtering algorithms, as reported in some publications [51], [104]. The ability to create ML models that can operate with different MoCap technologies without compromising the overall accuracy or violating the time requirements of an assessment is an area that merits further research.

4) BALANCE BETWEEN ACCURACY AND SPEED

As discussed above, ML algorithms for HMQA during exercise should generate accurate assessments in real-time. Beyond accuracy, however, the need to create informative assessments also adds another layer of complexity, which make real-time assessments more difficult. Therefore, determining the correct balance between the two can be challenging. For example, shallow algorithms that may run faster usually require longer preprocessing time [34]. In contrast, DL approaches usually require no or minimal preprocessing, but their assessments, depending on the size of the data, can be slower. Simultaneously, new approaches in DL algorithms, allow for faster training and inference times. For example, LSTM has been a popular algorithm for human activity analysis, and the same was evident from this review, as it was the most used algorithm across all publications. However, its structure can be complicated, making it slower, particularly for real-time applications. Recently, GRU RNN has been used as an alternative because it achieves similar results to LSTM, albeit using fewer gates [174], [175]. Of the publications included in this review, two publications implemented this approach. Albert *et al.* [30] showed that their proposed algorithm outperformed other approaches that used, among others, LSTM, while simultaneously it reduced training time by 3%. Khoramdel *et al.* [82] also used GRU-based model to detect compensatory movements during upper-limb rehabilitation exercise. Their implementation using a 1-layer GRU RNN had higher accuracies compared to other RNN solutions (including LSTM) with data from both healthy participants and patients.

Data preprocessing, feature extraction, and selection methods also play important roles in this process. Several publications used raw values from the input devices as features, in a time-series format, as input to the ML algorithms, compared with aggregation-based variables that provide more general characteristics of a recorded sample. These features are typically used with DL systems or template-based algorithms. Because of their complexity, even though they can often detect more subtle fluctuations in the input signals, they are computationally more demanding and require more time for analysis [102]. One way to address this problem is to transform the data into other types with either a reduced dimensionality or a more efficient format. Another alternative is the use of feature selection algorithms, which can reduce the number of features used. Nonetheless, any preprocessing of data can incur time penalties that may overshadow the gains from their use.

5) SELECTING THE RIGHT ML ALGORITHM

The identified publications used a vast array of algorithms, ranging from traditional, shallow ML techniques to DL approaches, such as LSTM and CNN. Hybrid approaches have also been proposed to exploit the characteristics of different approaches. One example is the combination of GCNs with LSTMs presented in two publications [40], [75]. The proposed solutions used GCNs to engineer features from raw signals that convey spatial information based on the human skeleton. These new features were then used with LSTMs. Even though this representation performs well with movements related to connected joints and their relation, it can perform poorly when the relevant joints are not connected, such as when symmetry between limbs is studied. Therefore, different representations can be explored based on the analyzed movements [176]. Similarly, the combination of RNN and CNN algorithms has shown a higher median accuracy across all included publications [40], [75], [78], [96], [116],

[117], [120]. With respect to shallow ML algorithms, Ensemble models have shown better accuracies in most scenarios, as have SVMs in simpler assessments. However, this area merits further exploration.

6) CONCURRENT ASSESSMENT DURING EXERCISE

The ability to provide feedback in real-time while performing an exercise is important to prevent injuries and correct prolonged erroneous exercise. However, most current approaches only analyze motion once a repetition has been completed and the full time-series has been extracted. The few exceptions usually rely on template-based methods such as Incremental DTW [33], [88], [100], and HMM [90] to segment an exercise into smaller sub-phases, which are in turn analyzed based on the current completion state of the exercise. However, this approach can generate complex models, because a preprocessing phase is required to first segment the repetition, and then multiple models need to be trained. This can be even more difficult if only a limited amount of available data is considered. An alternative is to use data augmentation techniques to generate full exercise repetition data based on partial data. Such approaches, which use GANs [30], [70], [177], can generate complete exercise execution data that can be used with existing algorithms for HMQA. This can be an important step toward real-time feedback generation during exercises.

7) PERSONALIZATION AND SCALABILITY

The need for the development of personalized models for human activity analysis has been long discussed [178], [179]. This can be even more important for HMQA during exercise that may deal with patients, such as in the healthcare/rehabilitation domain, where each individual may have their own disabilities. Several studies have shown encouraging results using personalized models. At the same time, however, they frequently fail to perform as well when dealing with new data [33], [47], [73], [79]. In addition, the use of personalized models can be more difficult for end-users [73], [88], and more difficult to scale appropriately to a larger number of exercises and errors, as separated samples by every individual user will be required for each possible class.

Interestingly, some publications combined generic models with the ability of users to personalize them online by selecting the desired features to be used. More specifically, Lee *et al.* [92] proposed the use of binary mask vectors for the inclusion or exclusion of features unique to each user by a therapist. In a follow-up study [94], they also used a reinforcement learning approach by applying a Deep Q-network with Double Q-Learning along with Markov Decision Process during future selection to identify relevant features programmatically. Then, they combined this with the vector mask introduced above by the therapist to create a hybrid model. In another approach, Lei *et al.* [47] used weight vectors to describe the features that are relevant for each class. This approach can help customize the models and optimize them for different subjects without compromising their complexity.

As already discussed, HMQA for exercise is applicable to a variety of domains and applications. This also extends to the possible users of the applications using this technology. In the healthcare domain, for example, assessment can help with many conditions, including neurological disorders, musculoskeletal conditions, and other motor deficits. Across the sports and wellness domains, there are a variety of different activities that can use such systems, and each has unique characteristics. The included publications, almost entirely, focused on a single area and dealt with a very specific problem. Often, domain-specific variables are used for assessment. This can be beneficial for one application and can provide more relevant feedback. However, it can also lead to difficulty in generalizing the proposed algorithms to other applications, requiring repeating the development process [2] whenever a new area needs to be evaluated. The same problem exists when dealing with the addition of new types of motions in an existing model or the identification of new deviations for incorrect performances. This highlights the need to create robust models that can be easily adapted and extended without compromising accuracy or performance.

8) EVALUATION

Regarding the analysis of the algorithms and presentation of the results, only the average accuracy was provided in most publications. However, accuracy alone can be misleading and does not present a complete picture of how an algorithm performs [68]. Ideally, various metrics, such as recall, specificity, and F1-score, should be shared for all classification-based results. In regression-based algorithms, various metrics were used, such as MAE, RMSE, correlation coefficient, MAD, and MAPE. Because different metrics provide different insights, a combination of metrics can provide a better understanding of how well a model performs. Therefore, RMSE should be presented, because it provides more information related to larger deviations in the results and MAE, which provides more interpretable results.

A common limitation of current publications stemming from the small sample size of the datasets is the inability to properly validate and test the models used. Different methods of cross-validation should be used for the proposed solution, based on the number of subjects and samples available, to prevent overfitting and expand the ability of the models to generalize to data from new users. If there are enough subjects with samples of all different classes, leave-n-subjects out should be preferred to ensure greater generalizability.

VI. CONCLUSION

HMQA can be an essential component of personal home-based fitness and rehabilitation applications and various sport-related applications. In this study, we performed a systematic literature review of publications published from January 2017 to December 2021 related to HMQA during exercise using ML. The search yielded 88 publications related to three different domains: healthcare/rehabilitation, sports, and wellness. As reported through the analysis of the results and the research questions posed, there are several considerations when deciding which algorithm should be used and what type of assessment should be performed. First, the decision on the input data modalities and the type of output of the HMQA should be made based on how the assessment will be used as part of the overall system, and how the generation of meaningful feedback can be achieved. This is closely related to the application domain of the system and the target population. In addition, the assessment should be performed in real-time, which requires both fast inference times and the ability to potentially augment incomplete exercise data to generate data for full repetitions. At the same time, assessments should be easy for the user to interpret, while simultaneously avoid biases based on the training data.

The range of application domains that use HMQA for exercise also merits better ways to generalize the developed models, providing the ability to transfer knowledge between domains without having to retrain the models. However, to achieve this, there is a need for larger public datasets to be made available in the field. Current public datasets are small and have various limitations. Solutions that develop HMQA that can work with various MoCap technologies could also help expand the available data that can be used to train the models. Through this review, a set of guidelines for developing ML solutions for HMQA have been proposed, a summary of which can be seen in Table 11.

REFERENCES

[1] L. Tao, A. Paiement, D. Damen, M. Mirmehdi, S. Hannuna, M. Camplani, T. Burghardt, and I. Craddock, "A comparative study of pose representation and dynamics modelling for online motion quality assessment," *Comput. Vis. Image Understand.*, vol. 148, pp. 136–152, Jul. 2016, doi: 10.1016/j.cviu.2015.11.016.

[2] A. Khan, S. Mellor, E. Berlin, R. Thompson, R. McNaney, P. Olivier, and T. Plötz, "Beyond activity recognition: Skill assessment from accelerometer data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 2015, pp. 1155–1166, doi: 10.1145/2750858.2807534.

[3] A. Khan, S. Mellor, R. King, B. Janko, W. Harwin, R. S. Sherratt, I. Craddock, and T. Plötz, "Generalized and efficient skill assessment from IMU data with applications in gymnastics and medical training," *ACM Trans. Comput. Healthcare*, vol. 2, no. 1, pp. 1–21, Jan. 2021, doi: 10.1145/3422168.

[4] A. Rajšp and I. Fister, "A systematic literature review of intelligent data analysis methods for smart sport training," *Appl. Sci.*, vol. 10, no. 9, p. 3013, Apr. 2020, doi: 10.3390/app10093013.

[5] Q. Lei, J.-X. Du, H.-B. Zhang, S. Ye, and D.-S. Chen, "A survey of vision-based human action evaluation methods," *Sensors*, vol. 19, no. 19, p. E4129, Sep. 2019, doi: 10.3390/s19194129.

[6] Y. Liao, A. Vakanski, M. Xian, D. Paul, and R. Baker, "A review of computational approaches for evaluation of rehabilitation exercises," *Comput. Biol. Med.*, vol. 119, Apr. 2020, Art. no. 103687, doi: 10.1016/j.compbiomed.2020.103687.

[7] E. D. O. Simbaña, P. S.-H. Baeza, A. J. Huete, and C. Balaguer, "Review of automated systems for upper limbs functional assessment in neurorehabilitation," *IEEE Access*, vol. 7, pp. 32352–32367, 2019, doi: 10.1109/ACCESS.2019.2901814.

[8] G. J. Luvizutto, G. F. Silva, M. R. Nascimento, K. C. S. Santos, P. A. Appelt, E. de Moura Neto, J. T. de Souza, F. C. Wincker, L. A. Miranda, P. T. H. Filho, L. A. P. S. de Souza, R. P. Simoes, E. I. de Oliveira Vidal, and R. Bazan, "Use of artificial intelligence as an instrument of evaluation after stroke: A scoping review based on international classification of functioning, disability and health concept: AI applications for stroke evaluation," *Topics Stroke Rehabil.*, vol. 29, no. 5, pp. 1–16, Jun. 2021, doi: 10.1080/10749357.2021.1926149.

[9] P. Castillo-Segura, C. Fernández-Panadero, C. Alario-Hoyos, P. J. Muñoz-Merino, and C. D. Kloos, "Objective and automated assessment of surgical technical skills with IoT systems: A systematic literature review," *Artif. Intell. Med.*, vol. 112, Feb. 2021, Art. no. 102007, doi: 10.1016/j.artmed.2020.102007.

[10] P. Parmar, J. Reddy, and B. Morris, "Piano skills assessment," in *Proc. IEEE 23rd Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2021, pp. 1–5, doi: 10.1109/MMSP53017.2021.9733638.

[11] E. Stefana, F. Marciano, D. Rossi, P. Cocca, and G. Tomasoni, "Wearable devices for ergonomics: A systematic literature review," *Sensors*, vol. 21, no. 3, p. 777, Jan. 2021, doi: 10.3390/s21030777.

[12] J. W. Cuchna, M. C. Hoch, and J. M. Hoch, "The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis," *Phys. Therapy Sport*, vol. 19, pp. 57–65, May 2016, doi: 10.1016/j.ptsp.2015.12.002.

[13] J. M. Stanek, J. Smith, and J. Petrie, "Intra- and inter-rater reliability of the selective functional movement assessment (SFMA) in healthy participants," *Int. J. Sports Phys. Therapy*, vol. 14, no. 1, pp. 107–116, Feb. 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4004125/

[14] J. Ressman, W. J. A. Grooten, and E. R. Barr, "Visual assessment of movement quality in the single leg squat test: A review and meta-analysis of inter-rater and intrarater reliability," *BMJ Open Sport Exercise Med.*, vol. 5, no. 1, Jun. 2019, Art. no. e000541, doi: 10.1136/bmjsem-2019-000541.

[15] D. Augustovicova, J. Argajova, L. Rupcik, and E. Thomson, "Development of a reliable and valid kata performance analysis template," *J. Phys. Educ. Sport*, vol. 20, no. 6, pp. 3553–3559, Dec. 2020, doi: 10.7752/jpes.2020.06479.

[16] H. Gulgin and B. Hoogenboom, "The functional movement screening (FMS)™: An inter-rater reliability study between raters of varied experience," *Int. J. Sports Phys. Therapy*, vol. 9, no. 1, pp. 14–20, Feb. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3924604/

[17] R. W. Moran, A. G. Schneiders, K. M. Major, and S. J. Sullivan, "How reliable are functional movement screening scores? A systematic review of rater reliability," *Brit. J. Sports Med.*, vol. 50, no. 9, pp. 527–536, May 2016, doi: 10.1136/bjsports-2015-094913.

[18] X. Song, S. Chen, J. Jia, and P. B. Shull, "Cellphone-based automated fugl-meyer assessment to evaluate upper extremity motor function after stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2186–2195, Oct. 2019, doi: 10.1109/TNSRE.2019.2939587.

[19] P. Maurice, A. Malaisé, C. Amiot, N. Paris, G.-J. Richard, O. Rochel, and S. Ivaldi, "Human movement and ergonomics: An industry-oriented dataset for collaborative robotics," *Int. J. Robot. Res.*, vol. 38, no. 14, pp. 1529–1537, Dec. 2019, doi: 10.1177/0278364919882089.

[20] C. J. Caspersen, K. E. Powell, and G. M. Christenson, "Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research," *Public Health Rep.*, vol. 100, no. 2, pp. 126–131, 1985. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1424733/

[21] F. Mortazavi and A. Nadian-Ghomsheh, "Continues online exercise monitoring and assessment system with visual guidance feedback for stroke rehabilitation," *Multimedia Tools Appl.*, vol. 78, no. 22, pp. 32055–32085, Nov. 2019, doi: 10.1007/s11042-019-08020-2.

[22] I. Boukhennoufa, X. Zhai, V. Utti, J. Jackson, and K. D. McDonald-Maier, "Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103197, doi: 10.1016/j.bspc.2021.103197.

[23] A. Ozturk, A. Tartar, B. E. Huseyinsinoglu, and A. H. Ertas, "A clinically feasible kinematic assessment method of upper extremity motor function impairment after stroke," *Measurement*, vol. 80, pp. 207–216, Feb. 2016, doi: 10.1016/j.measurement.2015.11.026.

[24] A. E. F. Da Gama, T. M. Chaves, L. S. Figueiredo, A. Baltar, M. Meng, N. Navab, V. Teichrieb, and P. Fallavollita, "MirrARbilitation: A clinically-related gesture recognition interactive tool for an AR rehabilitation system," *Comput. Methods Programs Biomed.*, vol. 135, pp. 105–114, Oct. 2016, doi: 10.1016/j.cmpb.2016.07.014.

[25] V. Fernandez-Cervantes, N. Neubauer, B. Hunter, E. Stroulia, and L. Liu, "VirtualGym: A Kinect-based system for seniors exercising at home," *Entertainment Comput.*, vol. 27, pp. 60–72, Aug. 2018, doi: 10.1016/j.entcom.2018.04.001.

[26] R. de Souza Baptista, A. P. L. Bo, and M. Hayashibe, "Automatic human movement assessment with switching linear dynamic system: Motion segmentation and motor performance," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 628–640, Jun. 2017, doi: 10.1109/TNSRE.2016.2591783.

[27] R. Baptista, E. Ghorbel, A. E. R. Shabayek, F. Moissenet, D. Aouada, A. Douchet, M. André, J. Pager, and S. Bouilland, "Home self-training: Visual feedback for assisting physical activity for stroke survivors," *Comput. Methods Programs Biomed.*, vol. 176, pp. 111–120, Jul. 2019, doi: 10.1016/j.cmpb.2019.04.019.

[28] Z. Zhang, Q. Fang, and X. Gu, "Objective assessment of upper-limb mobility for poststroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 859–868, Apr. 2016, doi: 10.1109/TBME.2015.2477095.

[29] A. Ebert, M. T. Beck, A. Mattausch, L. Belzner, and C. Linnhoff-Popien, "Qualitative assessment of recurrent human motion," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 306–310, doi: 10.23919/EUSIPCO.2017.8081218.

[30] J. Albert, P. Glockner, B. Pfitzner, and B. Arnrich, "Data augmentation of kinematic time-series from rehabilitation exercises using GANs," in *Proc. IEEE Int. Conf. Omni-Layer Intell. Syst. (COINS)*, Barcelona, Spain, Aug. 2021, pp. 1–6, doi: 10.1109/COINS51742.2021.9524284.

[31] D. Yang, J. Tang, Y. Huang, C. Xu, J. Li, L. Hu, G. Shen, C.-J. M. Liang, and H. Liu, "TennisMaster: An IMU-based online serve performance evaluation system," in *Proc. 8th Augmented Hum. Int. Conf.*, New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 1–8, doi: 10.1145/3041164.3041186.

[32] R. Kianifar, A. Lee, S. Raina, and D. Kulić, "Automated assessment of dynamic knee valgus and risk of knee injury during the single leg squat," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–13, 2017, doi: 10.1109/JTEHM.2017.2736559.

[33] J. Richter, C. Wiede, B. Shinde, and G. Hirtz, "Motion error classification for assisted physical therapy—A novel approach using incremental dynamic time warping and normalised hierarchical skeleton joint data," in *Proc. 6th Int. Conf. Pattern Recognit. Appl. Methods (ICPRAM)*, vol. 2, Porto, Portugal: SciTePress, Feb. 2017, pp. 281–288, doi: 10.5220/0006108002810288.

[34] T. Decroos, K. Schütte, T. O. De Beéck, B. Vanwanseele, and J. Davis, "AMIE: Automatic monitoring of indoor exercises," in *Machine Learning and Knowledge Discovery in Databases* (Lecture Notes in Computer Science), vol. 11053, U. Brefeld, E. Curry, E. Daly, B. MacNamee, A. Marascu, F. Pinelli, M. Berlingerio, and N. Hurley, Eds. Cham, Switzerland: Springer, 2019, pp. 424–439, doi: 10.1007/978-3-030-10997-4_26.

[35] W. Zhang, C. Su, and C. He, "Rehabilitation exercise recognition and evaluation based on smart sensors with deep learning framework," *IEEE Access*, vol. 8, pp. 77561–77571, 2020, doi: 10.1109/ACCESS.2020.2989128.

[36] S. De Villa, A. Parra, A. Martin, J. Dominguez, and D. Casillas-Perez, "ML algorithms for the assessment of prescribed physical exercises," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Lausanne, Switzerland, Jun. 2021, pp. 1–6, doi: 10.1109/MeMeA52024.2021.9478725.

[37] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020, doi: 10.1109/ACCESS.2020.2980942.

[38] N. Hernandex-Cruz, J. Lundstram, J. Favela, I. McChesney, and B. Arnrich, "Literature review on transfer learning for human activity recognition using mobile and wearable devices with environmental technology," *Social Netw. Comput. Sci.*, vol. 1, p. 66, Feb. 2020, doi: 10.1007/s42979-020-0070-4.

[39] J.-K. Kim, K. B. Lee, J.-C. Kim, and S. G. Hong, "Patient identification based on physical rehabilitation movements using skeleton data," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2021, pp. 1572–1574, doi: 10.1109/ICTC52510.2021.9621049.

[40] S. H. Chowdhury, M. A. Amin, A. K. M. M. Rahman, M. A. Amin, and A. A. Ali, "Assessment of rehabilitation exercises from depth sensor data," in *Proc. 24th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2021, pp. 1–7, doi: 10.1109/ICCIT54785.2021.9689826.

[41] F. Amjad, M. H. Khan, M. A. Nisar, M. S. Farid, and M. Grzegorzek, "A comparative study of feature selection approaches for human activity recognition using multimodal sensory data," *Sensors*, vol. 21, no. 7, p. 2368, Mar. 2021, doi: 10.3390/s21072368.

[42] M. Cherrington, J. Lu, D. Airehrour, F. Thabtah, Q. Xu, and S. Madanian, "Feature selection: Multi-source and multi-view data limitations, capabilities and potentials," in *Proc. 29th Int. Telecommun. Netw. Appl. Conf. (ITNAC)*, Nov. 2019, pp. 1–6, doi: 10.1109/ITNAC46935.2019.9077992.

[43] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[44] H. Jain and G. Harit, "An unsupervised sequence-to-sequence autoencoder based human action scoring model," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5, doi: 10.1109/GlobalSIP45357.2019.8969424.

[45] D. Avola, L. Cinque, G. L. Foresti, and M. R. Marini, "An interactive and low-cost full body rehabilitation framework based on 3D immersive serious games," *J. Biomed. Informat.*, vol. 89, pp. 81–100, Jan. 2019, doi: 10.1016/j.jbi.2018.11.012.

[46] J. Hagelbäck, P. Liapota, A. Lincke, and W. Löwe, "The performance of some machine learning approaches in human movement assessment," in *Proc. 13th Multi Conf. Comput. Sci. Inf. Syst. (MCCSIS)*, 2019, pp. 35–42, doi: 10.33965/eh2019_201910l005.

[47] Q. Lei, H.-B. Zhang, J.-X. Du, T.-C. Hsiao, and C.-C. Chen, "Learning effective skeletal representations on RGB video for fine-grained human action quality assessment," *Electronics*, vol. 9, no. 4, p. 568, Mar. 2020, doi: 10.3390/electronics9040568.

[48] P. F. Dajime, H. Smith, and Y. Zhang, "Automated classification of movement quality using the Microsoft kinect V2 sensor," *Comput. Biol. Med.*, vol. 125, Oct. 2020, Art. no. 104021, doi: 10.1016/j.compbiomed.2020.104021.

[49] J. Palomares-Pecho, G. Silva-Calpa, C. Sierra-Franco, and A. B. Raposo, "End-user programming architecture for physical movement assessment: An interactive machine learning approach," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), vol. 12198. Cham, Switzerland: Springer, Jul. 2020, pp. 335–348, doi: 10.1007/978-3-030-49904-4_25.

[50] S. S. Esfahlani, J. Butt, and H. Shirvani, "Fusion of artificial intelligence in neuro-rehabilitation video games," *IEEE Access*, vol. 7, pp. 102617–102627, 2019, doi: 10.1109/ACCESS.2019.2926118.

[51] S. S. Tabrizi, S. Pashazadeh, and V. Javani, "A deep learning approach for table tennis forehand stroke evaluation system using an IMU sensor," *Comput. Intell. Neurosci.*, vol. 2021, Apr. 2021, Art. no. 5584756, doi: 10.1155/2021/5584756.

[52] M. A. O'Reilly, T. E. Ward, E. Delahunt, B. Caulfield, and D. F. Whelan, "Technology in rehabilitation: Comparing personalised and global classification methodologies in evaluating the squat exercise with wearable IMUs," *Methods Inf. Med.*, vol. 56, no. 5, pp. 361–369, 2017, doi: 10.3414/ME16-01-0141.

[53] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Med.*, vol. 6, no. 7, Jul. 2009, Art. no. e1000097, doi: 10.1371/journal.pmed.1000097.

[54] C. Richter, E. King, S. Strike, and A. Franklyn-Miller, "Objective classification and scoring of movement deficiencies in patients with anterior cruciate ligament reconstruction," *PLoS ONE*, vol. 14, no. 7, Jul. 2019, Art. no. e0206024, doi: 10.1371/journal.pone.0206024.

[55] B. X. B. Yu, Y. Liu, and K. C. C. Chan, "Skeleton-based detection of abnormalities in human actions using graph convolutional networks," in *Proc. 2nd Int. Conf. Transdisciplinary AI (TransAI)*, Irvine, CA, USA, Sep. 2020, pp. 131–137, doi: 10.1109/TransAI49837.2020.00030.

[56] B. X. B. Yu, Y. Liu, K. C. C. Chan, Q. Yang, and X. Wang, "Skeleton-based human action evaluation using graph convolutional network for monitoring Alzheimer's progression," *Pattern Recognit.*, vol. 119, Nov. 2021, Art. no. 108095, doi: 10.1016/j.patcog.2021.108095.

[57] Y. Jiang, Y. Qin, I. Kim, and Y. Wang, "Towards an IoT-based upper limb rehabilitation assessment system," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jeju, South Korea, Jul. 2017, pp. 2414–2417, doi: 10.1109/EMBC.2017.8037343.

[58] R. Julianjatsono, R. Ferdiana, and R. Hartanto, "High-resolution automated Fugl-Meyer assessment using sensor data and regression model," in *Proc. 3rd Int. Conf. Sci. Technol. Comput.* Yogyakarta, Indonesia: Institute of Electrical and Electronics Engineers, Jul. 2017, pp. 28–32, doi: 10.1109/ICSTC.2017.8011847.

[59] H.-T. Jung, H. Kim, M. Y. Oh, T. Ryu, and Y. Kim, "Learning classifier to evaluate movement quality in unassisted pick-and-place exercises for post-stroke patients: A preliminary study," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2490–2493, doi: 10.1109/EMBC.2017.8037362.

[60] N. Eichler, H. Hel-Or, I. Shimshoni, D. Itah, B. Gross, and S. Raz, "3D motion capture system for assessing patient motion during Fugl-Meyer stroke rehabilitation testing," *IET Comput. Vis.*, vol. 12, no. 7, pp. 963–975, Jul. 2018, doi: 10.1049/iet-cvi.2018.5274.

[61] N. Eichler, H. Hel-Or, I. Shmishoni, D. Itah, B. Gross, and S. Raz, "Non-invasive motion analysis for stroke rehabilitation using off the shelf 3D sensors," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489593.

[62] H.-T. Jung, J. Park, J. Jeong, T. Ryu, Y. Kim, and S. I. Lee, "A wearable monitoring system for at-home stroke rehabilitation exercises: A preliminary study," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*. Las Vegas, NV, USA: Institute of Electrical and Electronics Engineers, Mar. 2018, pp. 13–16, doi: 10.1109/BHI.2018.8333358.

[63] J. He, S. Chen, Z. Guo, S. Pirbhulal, W. Wu, J. Feng, and G. Dan, "A comparative study of motion recognition methods for efficacy assessment of upper limb function," *Int. J. Adapt. Control Signal Process.*, vol. 33, no. 8, pp. 1248–1256, Aug. 2019, doi: 10.1002/acs.2941.

[64] H. Sarwat, H. Sarwat, M. I. Awad, and S. A. Maged, "Assessment of post-stroke patients using smartphones and gradient boosting," in *Proc. 15th Int. Conf. Comput. Eng. Syst. (ICCES)*, Cairo, Egypt, Dec. 2020, pp. 1–6, doi: 10.1109/ICCES51560.2020.9334654.

[65] H. Sarwat, M. I. Awad, S. A. Maged, and H. Sarwat, "Self-organizing maps to assess rehabilitation progress of post-stroke patients," in *Proc. 2nd Novel Intell. Lead. Emerg. Sci. Conf. (NILES)*, Giza, Egypt, Oct. 2020, pp. 116–121, doi: 10.1109/NILES50944.2020.9257927.

[66] C. Wang, L. Peng, Z.-G. Hou, J. Li, T. Zhang, and J. Zhao, "Quantitative assessment of upper-limb motor function for post-stroke rehabilitation based on motor synergy analysis and multi-modality fusion," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 943–952, Apr. 2020, doi: 10.1109/TNSRE.2020.2978273.

[67] X. Chen, Y. Shi, Y. Wang, and Y. Cheng, "Deep learning-based image automatic assessment and nursing of upper limb motor function in stroke patients," *J. healthcare Eng.*, vol. 2021, Aug. 2021, Art. no. 9059411, doi: 10.1155/2021/9059411.

[68] L. Formstone, W. Huo, S. Wilson, A. McGregor, P. Bentley, and R. Vaidyanathan, "Quantification of motor function post-stroke using novel combination of wearable inertial and mechanomyographic sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1158–1167, 2021, doi: 10.1109/TNSRE.2021.3089613.

[69] J. Lee, H.-T. Jung, and S. I. Lee, "Estimating the quality of reaching movements in stroke survivors," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Athens, Greece, Jul. 2021, pp. 1–4, doi: 10.1109/BHI50953.2021.9508566.

[70] L. Li and A. Vakanski, "Generative adversarial networks for generation and classification of physical rehabilitation movement episodes," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 5, pp. 428–436, Oct. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6195368/

[71] N. Sadawi, A. Miron, W. Ismail, H. Hussain, and C. Grosan, "Gesture correctness estimation with deep neural networks and rough path descriptors," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2019, pp. 595–602, doi: 10.1109/ICDMW.2019.00090.

[72] A. Miron, N. Sadawi, W. Ismail, H. Hussain, and C. Grosan, "IntelliRehabDS (IRDS)—A dataset of physical rehabilitation movements," *Data*, vol. 6, no. 5, p. 46, May 2021, doi: 10.3390/data6050046.

[73] T. E. Ward, E. Delahunt, B. Caulfield, D. F. Whelan, and M. A. O'Reilly, "Technology in rehabilitation: Evaluating the single leg squat exercise with wearable inertial measurement units," *Methods Inf. Med.*, vol. 56, no. 2, pp. 88–94, 2017, doi: 10.3414/ME16-02-0002.

[74] N. Rao, P. M. Surana, R. Ragesh, and G. Srinivasa, "Analysis of joints for tracking fitness and monitoring progress in physiotherapy," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Kuala Lumpur, Malaysia, Sep. 2019, pp. 181–185, doi: 10.1109/ICSIPA45851.2019.8977749.

[75] C. Du, S. Graham, C. Depp, and T. Nguyen, "Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 281–285, doi: 10.1109/EMBC46164.2021.9629569.

[76] U. Akshaya Devi and N. Arulanand, "Smart rehabilitation exercise monitoring system for knee osteoarthritis patients," in *Proc. 12th Int. Conf. Adv. Comput., Control, Telecommun. Technol. (ACT)*, vol. 50. Kerala, India: Grenze Scientific Society, Aug. 2021, pp. 543–550. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117860537&partnerID=40&md5=0a3c16f8fea439804e69cb987c9ce422

[77] Z.-A. Zhu, Y.-C. Lu, C.-H. You, and C.-K. Chiang, "Deep learning for sensor-based rehabilitation exercise recognition and evaluation," *Sensors*, vol. 19, no. 4, p. 887, Feb. 2019, doi: 10.3390/s19040887.

[78] Y. Liao, A. Vakanski, and M. Xian, "A deep learning framework for assessing physical rehabilitation exercises," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 468–477, Feb. 2020, doi: 10.1109/TNSRE.2020.2966249.

[79] A. S. Whitford, E. Kim, E. Halilaj, K. Enseki, A. Popchak, and J. Hodgins, "Sensor-based evaluation of physical therapy exercises," in *Proc. 43rd Annual Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 7556–7561, doi: 10.1109/EMBC46164.2021.9629773.

[80] M. J. Raihan, M. A. R. Ahad, and A.-A. Nahid, "Automated rehabilitation exercise assessment by genetic algorithm-optimized CNN," in *Proc. Joint 10th Int. Conf. Inform., Electron. Vis. 5th Int. Conf. Imag., Vis. Pattern Recognit. (ICIEV-icIVPR)*, Aug. 2021, pp. 1–6, doi: 10.1109/ICIEVicIVPR52578.2021.9564240.

[81] Y. X. Zhi, M. Lukasik, M. H. Li, E. Dolatabadi, R. H. Wang, and B. Taati, "Automatic detection of compensation during robotic stroke rehabilitation therapy," *IEEE J. Transl. Eng. Health Med.*, vol. 6, 2018, Art. no. 2100107, doi: 10.1109/JTEHM.2017.2780836.

[82] J. Khoramdel, A. Moori, M. M. Moghaddam, and E. Najafi, "Compensatory movement detection in upper limb rehabilitation with deep learning methods," in *Proc. 9th RSI Int. Conf. Robot. Mechtron. (ICRoM)*, Nov. 2021, pp. 465–471, doi: 10.1109/ICRoM54204.2021.9663458.

[83] S. Miao, Y. Dang, Q. Zhu, S. Li, M. Shorfuzzaman, and H. Lv, "A novel approach for upper limb functionality assessment based on deep learning and multimodal sensing data," *IEEE Access*, vol. 9, pp. 77138–77148, 2021, doi: 10.1109/ACCESS.2021.3080592.

[84] A. Chytas, D. Fotopoulos, V. Kilintzis, E. Koutsiana, I. Ladakis, E. Kiana, T. Loizidis, and I. Chouvarda, "Characterization of upper limb functionality caused by neuromuscular disorders using novel motion features from a specialized gaming platform," in *Proc. 14th Int. Joint Conf. Biomed. Eng. Syst. Technol. (BIOSTEC)*, 2021, pp. 60–68, doi: 10.5220/0010244400600068.

[85] D. Avola, L. Cinque, G. L. Foresti, M. R. Marini, and D. Pannone, "VRheab: A fully immersive motor rehabilitation system based on recurrent neural network," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 24955–24982, Oct. 2018, doi: 10.1007/s11042-018-5730-1.

[86] Bai, Song, and Li, "Design and analysis of cloud upper limb rehabilitation system based on motion tracking for post-stroke patients," *Appl. Sci.*, vol. 9, no. 8, p. 1620, Jan. 2019, doi: 10.3390/app9081620.

[87] S.-H. Lee, J. Cui, L. Liu, M.-C. Su, L. Zheng, and S.-C. Yeh, "An evidence-based intelligent method for upper-limb motor assessment via a VR training system on stroke rehabilitation," *IEEE Access*, vol. 9, pp. 65871–65881, 2021, doi: 10.1109/ACCESS.2021.3075778.

[88] J. Richter, C. Wiede, L. Lehmann, and G. Hirtz, "Motion evaluation by means of joint filtering for assisted physical therapy," in *Proc. IEEE 7th Int. Conf. Consum. Electron. Berlin (ICCE-Berlin)*, Berlin, Germany, Sep. 2017, pp. 10–14, doi: 10.1109/ICCE-Berlin.2017.8210575.

[89] W. Wei, C. McElroy, and S. Dey, "Human action understanding and movement error identification for the treatment of patients with Parkinson's disease," in *Proc. IEEE Int. Conf. Healthcare Inform. (ICHI)*, New York, NY, USA, Jun. 2018, pp. 180–190, doi: 10.1109/ICHI.2018.00028.

[90] W. Wei, C. McElroy, and S. Dey, "Towards on-demand virtual physical therapist: Machine learning-based patient action understanding, assessment and task recommendation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1824–1835, Sep. 2019, doi: 10.1109/TNSRE.2019.2934097.

[91] S. I. Lee, C. P. Adans-Dester, M. Grimaldi, A. V. Dowling, P. C. Horak, R. M. Black-Schaffer, P. Bonato, and J. T. Gwin, "Enabling stroke rehabilitation in home and community settings: A wearable sensor-based approach for upper-limb motor training," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–11, 2018, doi: 10.1109/JTEHM.2018.2829208.

[92] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. I. Badia, "Learning to assess the quality of stroke rehabilitation exercises," in *Proc. 24th Int. Conf. Intell. User Interfaces*, New York, NY, USA, Mar. 2019, pp. 218–228, doi: 10.1145/3301275.3302273.

[93] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. I. Bermúdez Badia, "Co-design and evaluation of an intelligent decision support system for stroke rehabilitation assessment," in *Proc. ACM Hum.-Comput. Interact.*, Oct. 2020, vol. 4, no. CSCW2, pp. 1–27, doi: 10.1145/3415227.

[94] M. H. Lee, D. P. Siewiorek, A. Smailagic, and A. Bernardino, "Interactive hybrid approach to combine machine and human intelligence for personalized rehabilitation assessment," in *Proc. ACM Conf. Health, Inference, Learn.* New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 160–169, doi: 10.1145/3368555.3384452.

[95] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. Badia, "Towards personalized interaction and corrective feedback of a socially assistive robot for post-stroke rehabilitation therapy," in *Proc. 29th IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Naples, Italy, Aug. 2020, pp. 1366–1373, doi: 10.1109/RO-MAN47096.2020.9223462.

[96] A. Kanade, M. Sharma, and M. Manivannan, "Tele-EvalNet: A low-cost, teleconsultation system for home based rehabilitation of stroke survivors using multiscale CNN-LSTM architecture," 2021, *arXiv:2112.03168*.

[97] S. Miao, C. Shen, X. Feng, Q. Zhu, M. Shorfuzzaman, and Z. Lv, "Upper limb rehabilitation system for stroke survivors based on multi-modal sensors and machine learning," *IEEE Access*, vol. 9, pp. 30283–30291, 2021, doi: 10.1109/ACCESS.2021.3055960.

[98] B. Bacic, "Towards the next generation of exergames: Flexible and personalised assessment-based identification of tennis swings," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489602.

[99] B. Bačič and P. A. Hume, "Computational intelligence for qualitative coaching diagnostics: Automated assessment of tennis swings to improve performance and safety," *Big Data*, vol. 6, no. 4, pp. 291–304, Dec. 2018, doi: 10.1089/big.2018.0062.

[100] F. Hülsmann, J. P. Göpfert, B. Hammer, S. Kopp, and M. Botsch, "Classification of motor errors to provide real-time feedback for sports coaching in virtual reality—A case study in squats and Tai Chi pushes," *Comput. Graph. (Pergamon)*, vol. 76, pp. 47–59, Nov. 2018, doi: 10.1016/j.cag.2018.08.003.

[101] S.-M. Lim, H.-C. Oh, J. Kim, J. Lee, and J. Park, "LSTM-guided coaching assistant for table tennis practice," *Sensors*, vol. 18, no. 12, p. 4112, Nov. 2018, doi: 10.3390/s18124112.

[102] H. Hegazy and M. Abdelsalam, M. Hussien, S. Elmosalamy, Y. Hassan, A. Nabil, and A. Atia, "Multi-sensor fusion for online detection and classification of table tennis strokes," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 2, pp. 201–210, Apr. 2021, doi: 10.22266/ijies2021.0430.18.

[103] I. Ghosh, S. R. Ramamurthy, and N. Roy, "StanceScorer: A data driven approach to score badminton player," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2020, pp. 1–6, doi: 10.1109/PerComWorkshops48775.2020.9156220.

[104] L. Liu, S. Qiu, Z. Wang, J. Li, and J. Wang, "Canoeing motion tracking and analysis via multi-sensors fusion," *Sensors*, vol. 20, no. 7, p. 2110, Apr. 2020, doi: 10.3390/s20072110.

[105] K. Soekarjo, D. Orth, E. Warmerdam, and J. Van Der Kamp, "Automatic classification of strike techniques using limb trajectory data" in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (CEUR Workshop Proceedings), vol. 11330. Cham, Switzerland: Springer, 2018, pp. 137–148, doi: 10.1007/978-3-030-17274-9_11.

[106] K. Makino, Y. Kitano, and H. Nishizaki, "Classification of swing motion of tennis using a recurrent-based neural network," in *Proc. 12th Int. Conf. Human Syst. Interact. (HSI)*, Richmond, VA, USA, Jun. 2019, pp. 237–242, doi: 10.1109/HSI47298.2019.8942630.

[107] G. B. Ross, B. Dowling, N. F. Troje, S. L. Fischer, and R. B. Graham, "Objectively differentiating movement patterns between elite and novice athletes," *Med. Sci. Sports Exerc.*, vol. 50, no. 7, pp. 1457–1464, Jul. 2018, doi: 10.1249/MSS.0000000000001571.

[108] A. Boyle, G. B. Ross, and R. B. Graham, "Machine learning and deep neural network architectures for 3D motion capture datasets," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Montreal, QC, Canada, Jul. 2020, pp. 4827–4830, doi: 10.1109/EMBC44109.2020.9176426.

[109] G. B. Ross, B. Dowling, N. F. Troje, S. L. Fischer, and R. B. Graham, "Classifying elite from novice athletes using simulated wearable sensor data," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 814, Aug. 2020, doi: 10.3389/fbioe.2020.00814.

[110] H. Brock, Y. Ohgi, and J. Lee, "Learning to judge like a human: Convolutional networks for classification of ski jumping errors," in *Proc. ACM Int. Symp. Wearable Computers*, vol. F130534. New York, NY, USA: Association for Computing Machinery, Sep. 2017, pp. 106–113, doi: 10.1145/3123021.3123038.

[111] D. Leightley, J. S. McPhee, and M. H. Yap, "Automated analysis and quantification of human mobility using a depth sensor," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 939–948, Jul. 2017, doi: 10.1109/JBHI.2016.2558540.

[112] M. A. O'Reilly, D. F. Whelan, T. E. Ward, E. Delahunt, and B. M. Caulfield, "Classification of deadlift biomechanics with wearable inertial measurement units," *J. Biomech.*, vol. 58, pp. 155–161, Jun. 2017, doi: 10.1016/j.jbiomech.2017.04.028.

[113] M. A. O'Reilly, D. F. Whelan, T. E. Ward, E. Delahunt, and B. Caulfield, "Classification of lunge biomechanics with multiple and individual inertial measurement units," *Sports Biomech.*, vol. 16, no. 3, pp. 342–360, Jul. 2017, doi: 10.1080/14763141.2017.1314544.

[114] R. Tekriwal and B. J. Pandian, "ANN based assistance for exercise patterns using accelerometer data," *Energy Proc.*, vol. 117, pp. 424–431, Mar. 2017, doi: 10.1016/j.egypro.2017.05.163.

[115] C.-C. Hsiao, P.-C. Yu, R.-G. Lee, and J. Jiang, "Deep-learning LSTM mechanism and wearable devices based virtual fitness-coach information system for barbell bench press," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*. Toronto, ON, Canada, Oct. 2020, pp. 3880–3885, doi: 10.1109/SMC42975.2020.9283422.

[116] J. Lee, H. Joo, J. Lee, and Y. Chee, "Automatic classification of squat posture using inertial sensors: Deep learning approach," *Sensors*, vol. 20, no. 2, p. 361, Jan. 2020, doi: 10.3390/s20020361.

[117] C. Zhang, L. Liu, Q. Xiang, J. Li, and X. Ren, "Sports pose estimation based on LSTM and attention mechanism," in *6GN for Future Wireless Networks*, vol. 337, X. Wang, V. C. M. Leung, K. Li, H. Zhang, X. Hu, and Q. Liu, Eds. Cham, Switzerland: Springer, 2020, pp. 538–550, doi: 10.1007/978-3-030-63941-9_45.

[118] A. Hannan, M. Z. Shafiq, F. Hussain, and I. M. Pires, "A portable smart fitness suite for real-time exercise monitoring and posture correction," *Sensors*, vol. 21, no. 19, p. 6692, Oct. 2021, doi: 10.3390/s21196692.

[119] P. Müller, F. Rauterberg, P. Achenbach, T. Tregel, and S. Göbel, "Physical exercise quality assessment using wearable sensors," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), vol. 12945. Cham, Switzerland: Springer, 2021, pp. 229–243, doi: 10.1007/978-3-030-88272-3_17.

[120] C. Zhang, L. Liu, M. Yao, W. Chen, D. Chen, and Y. Wu, "HSiPu2—A new human physical fitness action dataset for recognition and 3D reconstruction evaluation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Nashville, TN, USA, Jun. 2021, pp. 3166–3175, doi: 10.1109/CVPRW53098.2021.00354.

[121] A. Vakanski, H.-P. Jun, D. Paul, and R. Baker, "A data set of human body movements for physical rehabilitation exercises," *Data*, vol. 3, no. 1, p. 2, Mar. 2018, doi: 10.3390/data3010002.

[122] A. Miron and C. Grosan, "Classifying action correctness in physical rehabilitation exercises," 2021, *arXiv:2108.01375*.

[123] M. Capecci, M. G. Ceravolo, F. Ferracuti, S. Iarlori, A. Monteriu, L. Romeo, and F. Verdini, "The KIMORE dataset: KInematic assessment of MOvement and clinical scores for remote monitoring of physical REhabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1436–1448, Jul. 2019, doi: 10.1109/TNSRE.2019.2923060.

[124] A. Nadian-Ghomsheh. (Jul. 2018). *Vision Based Online Exercise Monitoring System With Visual Feedback for Stroke Rehabilitation Therapy*. [Online]. Available: https://doi.org/10.6084/m9.figshare.6741713.v1

[125] E. Dolatabadi, Y. X. Zhi, B. Ye, M. Coahran, G. Lupinacci, A. Mihailidis, R. Wang, and B. Taati, "The Toronto rehab stroke pose dataset to detect compensation during stroke rehabilitation therapy," in *Proc. 11th EAI Int. Conf. Pervasive Comput. Technol. Healthcare*, New York, NY, USA, May 2017, pp. 375–381, doi: 10.1145/3154862.3154925.

[126] H. Pirsiavash, C. Vondrick, and A. Torralba, "Assessing the quality of actions," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 556–571, doi: 10.1007/978-3-319-10599-4_36.

[127] P. Parmar and B. T. Morris, "Learning to score Olympic events," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 76–84, doi: 10.1109/CVPRW.2017.16.

[128] S. Tabrizi, V. Javani, and S. Pashazadeh. (Oct. 2020). *Data Acquired by a Single Object Sensor (IMU) of Table Tennis Forehand Strokes*. [Online]. Available: https://doi.org/10.17632/b7bc9y232m.3

[129] A. Ebert, C. Marouane, C. Ungnadner, and A. Klein, "An open, labeled dataset for analysis and assessment of human motion," in *Wireless Mobile Communication and Healthcare* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), P. Perego, A. M. Rahmani, and N. TaheriNejad, Eds. Cham, Switzerland: Springer, 2018, pp. 99–106, doi: 10.1007/978-3-319-98551-0_12.

[130] D. Leightley, M. H. Yap, J. Coulson, Y. Barnouin, and J. S. McPhee, "Benchmarking human motion analysis using Kinect one: An open source dataset," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2015, pp. 1–7, doi: 10.1109/APSIPA.2015.7415438.

[131] J. Wang, T. Zhu, J. Gan, L. Chen, H. Ning, and Y. Wan, "Sensor data augmentation by resampling for contrastive learning in human activity recognition," 2021, *arXiv:2109.02054*.

[132] N. De Giorgis, E. Puppo, P. Alborno, and A. Camurri, "Evaluating movement quality through intrapersonal synchronization," *IEEE Trans. Hum.-Mach. Syst.*, vol. 49, no. 4, pp. 304–313, Aug. 2019, doi: 10.1109/THMS.2019.2912498.

[133] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021, doi: 10.1109/tpami.2019.2929257.

[134] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019, doi: 10.1109/ACCESS.2019.2912200.

[135] *Kaia Health AI Lab | Motion Coach^{TM}*. Accessed: Aug. 18, 2022. [Online]. Available: https://kaiahealth.com/technology/

[136] *Onyx: Home Workout*. Accessed: Aug. 18, 2022. [Online]. Available: https://www.onyx.fit

[137] *VAY Fitness Coach*. Accessed: Aug. 18, 2022. [Online]. Available: https://www.vay.ai

[138] *Tempo Studio*. Accessed: Aug. 18, 2022. [Online]. Available: https://tempo.fit/shop/studio

[139] World Health Organization. (Oct. 2020). *Rehabilitation 2030 Initiative*. [Online]. Available: https://www.who.int/initiatives/rehabilitation-2030

[140] D. A. Sharma, M. F. Chevidikunnan, F. R. Khan, and R. A. Gaowgzeh, "Effectiveness of knowledge of result and knowledge of performance in the learning of a skilled motor activity by healthy young adults," *J. Phys. Therapy Sci.*, vol. 28, no. 5, pp. 1482–1486, 2016, doi: 10.1589/jpts.28.1482.

[141] E. Ebrahimi, S. V. Babu, C. C. Pagano, and S. Jörg, "An empirical evaluation of visuo-haptic feedback on physical reaching behaviors during 3D interaction in real and immersive virtual environments," *ACM Trans. Appl. Perception*, vol. 13, no. 4, pp. 1–21, Jul. 2016, doi: 10.1145/2947617.

[142] L. Brennan, E. Dorronzoro Zubiete, and B. Caulfield, "Feedback design in targeted exercise digital biofeedback systems for home rehabilitation: A scoping review," *Sensors*, vol. 20, no. 1, p. 181, Jan. 2020, doi: 10.3390/s20010181.

[143] K. M. Wilson, W. S. Helton, N. R. de Joux, J. R. Head, and J. J. S. Weakley, "Real-time quantitative performance feedback during strength exercise improves motivation, competitiveness, mood, and performance," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 61, no. 1, pp. 1546–1550, Sep. 2017, doi: 10.1177/1541931213601750.

[144] K. M. Wilson, N. R. de Joux, J. R. Head, W. S. Helton, J. S. Dang, and J. J. S. Weakley, "Presenting objective visual performance feedback over multiple sets of resistance exercise improves motivation, competitiveness, and performance," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 62, no. 1, pp. 1306–1310, Sep. 2018, doi: 10.1177/1541931218621299.

[145] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.

[146] S. Khairat, D. Marc, W. Crosby, and A. Al Sanousi, "Reasons for physicians not adopting clinical decision support systems: Critical analysis," *JMIR Med. Informat.*, vol. 6, no. 2, p. e8912, Apr. 2018, doi: 10.2196/medinform.8912.

[147] J. Parker, G. Mountain, and J. Hammerton, "A review of the evidence underpinning the use of visual and auditory feedback for computer technology in post-stroke upper-limb rehabilitation," *Disab. Rehabil., Assistive Technol.*, vol. 6, no. 6, pp. 465–472, Nov. 2011, doi: 10.3109/17483107.2011.556209.

[148] D. Simonsen, M. B. Popovic, E. G. Spaich, and O. K. Andersen, "Design and test of a Microsoft kinect-based system for delivering adaptive visual feedback to stroke patients during training of upper limb movement," *Med. Biol. Eng. Comput.*, vol. 55, no. 11, pp. 1927–1935, Nov. 2017, doi: 10.1007/s11517-017-1640-z.

[149] S. Tretriluxana and J. Tretriluxana, "Differential effects of feedback in the virtual reality environment for arm rehabilitation after stroke," in *Proc. 8th Biomed. Eng. Int. Conf. (BMEiCON)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2016, doi: 10.1109/BMEiCON.2015.7399562.

[150] S. An and U. Y. Ogras, "MARS: mmWave-based assistive rehabilitation system for smart healthcare," *ACM Trans. Embedded Comput. Syst.*, vol. 20, no. 5s, pp. 1–22, Sep. 2021, doi: 10.1145/3477003.

[151] A. T. Burge, A. E. Holland, C. F. McDonald, M. J. Abramson, C. J. Hill, A. L. Lee, N. S. Cox, R. Moore, C. Nicolson, P. O'Halloran, A. Lahham, R. Gillies, and A. Mahal, "Home-based pulmonary rehabilitation for COPD using minimal resources: An economic analysis," *Respirology*, vol. 25, no. 2, pp. 183–190, 2020, doi: 10.1111/resp.13667.

[152] M. P. Shenoy and P. D. Shenoy, "Identifying the challenges and cost-effectiveness of telerehabilitation: A narrative review," *J. Clin. DIAGNOSTIC Res.*, pp. 1–4, 2018, doi: 10.7860/JCDR/2018/36811.12311.

[153] G. E. Shields, A. Wells, P. Doherty, A. Heagerty, D. Buck, and L. M. Davies, "Cost-effectiveness of cardiac rehabilitation: A systematic review," *Heart*, vol. 104, no. 17, pp. 1403–1410, Sep. 2018, doi: 10.1136/heartjnl-2017-312809.

[154] M. Wu and J. Luo, "Wearable technology applications in healthcare: A literature review," *On-Line J. Nursing Inform.*, vol. 23, no. 3, 2019. [Online]. Available: https://www.himss.org/resources/wearable-technology-applications-healthcare-literature-review

[155] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "On the personalization of classification models for human activity recognition," *IEEE Access*, vol. 8, pp. 32066–32079, 2020, doi: 10.1109/ACCESS.2020.2973425.

[156] E. Halilaj, A. Rajagopal, M. Fiterau, J. L. Hicks, T. J. Hastie, and S. L. Delp, "Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities," *J. Biomech.*, vol. 81, pp. 1–11, Nov. 2018, doi: 10.1016/j.jbiomech.2018.09.009.

[157] T. Hakim, "A comprehensive review of skeleton-based movement assessment methods," 2020, *arXiv:2007.10737*.

[158] S. Rich U. Uy and P. A. Abu, "Analysis of detecting compensation for robotic stroke rehabilitation therapy using imbalanced learning and outlier detection," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Fukuoka, Japan, Feb. 2020, pp. 432–437, doi: 10.1109/ICAIIC48513.2020.9064992.

[159] M. Rana and V. Mittal, "Wearable sensors for real-time kinematics analysis in sports: A review," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1187–1207, Jan. 2021, doi: 10.1109/JSEN.2020.3019016.

[160] E. Mencarini, A. Rapp, L. Tirabeni, and M. Zancanaro, "Designing wearable systems for sports: A review of trends and opportunities in human—Computer interaction," *IEEE Trans. Hum.-Mach. Syst.*, vol. 49, no. 4, pp. 314–325, Aug. 2019, doi: 10.1109/THMS.2019.2919702.

[161] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 1–33, Jan. 2014, doi: 10.1145/2499621.

[162] P. Kumar, R. Saini, M. Yadava, P. P. Roy, D. P. Dogra, and R. Balasubramanian, "Virtual trainer with real-time feedback using Kinect sensor," in *Proc. IEEE Region Symp. (TENSYMP)*, Cochin, India, Jul. 2017, pp. 1–5, doi: 10.1109/TENCONSpring.2017.8070063.

[163] P. Caserman, C. Krug, and S. Göbel, "Recognizing full-body exercise execution errors using the teslasuit," *Sensors*, vol. 21, no. 24, p. 8389, Dec. 2021, doi: 10.3390/s21248389.

[164] A. Farnoosh, Z. Wang, S. Zhu, and S. Ostadabbas, "A Bayesian dynamical approach for human action recognition," *Sensors*, vol. 21, no. 16, p. 5613, Aug. 2021, doi: 10.3390/s21165613.

[165] P. Maceira-Elvira, T. Popa, A.-C. Schmid, and F. C. Hummel, "Wearable technology in stroke rehabilitation: Towards improved diagnosis and treatment of upper-limb motor impairment," *J. Neuroeng. Rehabil.*, vol. 16, no. 1, p. 142, Nov. 2019, doi: 10.1186/s12984-019-0612-y.

[166] Q. Wang, P. Markopoulos, B. Yu, W. Chen, and A. Timmermans, "Interactive wearable systems for upper body rehabilitation: A systematic review," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, pp. 1–21, Mar. 2017, doi: 10.1186/s12984-017-0229-y.

[167] M. O'Reilly, B. Caulfield, T. Ward, W. Johnston, and C. Doherty, "Wearable inertial sensor systems for lower limb exercise detection and evaluation: A systematic review," *Sports Med.*, vol. 48, no. 5, pp. 1221–1246, May 2018, doi: 10.1007/s40279-018-0878-4.

[168] M. N. Kamel Boulos and S. P. Yang, "Mobile physical activity planning and tracking: A brief overview of current options and desiderata for future solutions," *mHealth*, vol. 7, p. 13, Jan. 2021, doi: 10.21037/mhealth.2020.01.01.

[169] B. Barshan and A. Yurtman, "Classifying daily and sports activities invariantly to the positioning of wearable motion sensor units," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 4801–4815, Jun. 2020, doi: 10.1109/JIOT.2020.2969840.

[170] A. Yurtman, B. Barshan, and B. Fidan, "Activity recognition invariant to wearable sensor unit orientation using differential rotational transformations represented by quaternions," *Sensors*, vol. 18, no. 8, p. 2725, Aug. 2018, doi: 10.3390/s18082725.

[171] V. Camomilla, E. Bergamini, S. Fantozzi, and G. Vannozzi, "Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review," *Sensors*, vol. 18, no. 3, p. 873, Mar. 2018, doi: 10.3390/s18030873.

[172] I. Weygers, M. Kok, M. Konings, H. Hallez, H. De Vroey, and K. Claeys, "Inertial sensor-based lower limb joint kinematics: A methodological systematic review," *Sensors*, vol. 20, no. 3, p. 673, Jan. 2020, doi: 10.3390/s20030673.

[173] C. Chen, R. Jafari, and N. Kehtarnavaz, "A survey of depth and inertial sensor fusion for human action recognition," *Multimedia Tools Appl.*, vol. 76, no. 3, pp. 4405–4425, Feb. 2017, doi: 10.1007/s11042-015-3177-1.

[174] N. Jaouedi, N. Boujnah, and M. S. Bouhlel, "A new hybrid deep learning model for human action recognition," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 32, no. 4, pp. 447–453, May 2020, doi: 10.1016/j.jksuci.2019.09.004.

[175] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, "Human action recognition from various data modalities: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–20, 2022, doi: 10.1109/TPAMI.2022.3183112.

[176] M. Nekoui, F. O. Tito Cruz, and L. Cheng, "FALCONS: FAst Learner-grader for CONtorted poses in Sports," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*. Seattle, WA, USA, Jun. 2020, pp. 3941–3949, doi: 10.1109/CVPRW50498.2020.00458.

[177] E. Barsoum, J. Kender, and Z. Liu, "HP-GAN: Probabilistic 3D human motion prediction via GAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1499909–1499149, doi: 10.1109/CVPRW.2018.00191.

[178] T. Plötz and Y. Guan, "Deep learning for human activity recognition in mobile computing," *Computer*, vol. 51, no. 5, pp. 50–59, May 2018, doi: 10.1109/MC.2018.2381112.

[179] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–40, May 2021, doi: 10.1145/3447744.

FOTOS FRANGOUDES received the B.S. degree in computer science from the University of Cyprus, Nicosia, Cyprus, in 2011, and the M.S. degree in computer science with a focus on game development from the University of Southern California, Los Angeles, CA, USA, in 2013.

He has been a Research Associate with the CYENS—Centre of Excellence, Immersive Technologies for Intelligent and Creative Applications (ITICA) Group, Nicosia, since 2019. His work focuses on the development of interactive media applications across the extended reality spectrum. His research interests include analysis of human motion and physiological data.

MARIA MATSANGIDOU received the B.S. degree in communication and internet studies from the Cyprus University of Technology, Limassol, Cyprus, in 2012, the M.A. degree in communication and new journalism from the Open University of Cyprus, Nicosia, Cyprus, in 2013, the M.A. degree in social and developmental psychology from the University of Cyprus, in 2014, and the Ph.D. degree in human–computer interaction from the University of Kent, Kent, U.K., in 2018.

She has been a Postdoctoral Research Fellow with the CYENS—Centre of Excellence, Smart, Ubiquitous and Participatory Technologies for Healthcare Innovation (HealthXR) Group, Nicosia, since 2019. She is also the local coordinator of several Horizon 2020 and Erasmus+ programs. She has been involved in studies on the use of virtual reality technology in severe dementia, cancer, patients with amputees, burn injuries, motor impairment, and mental health disorders.

EIRINI C. SCHIZA (Member, IEEE) received the B.S. degree in computer science and the M.S. degree in computer science—multimedia applications and virtual environments from the University of Sussex, Brighton, U.K., in 2010 and 2011, respectively, and the Ph.D. degree from the Johan Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands, in 2018.

She has been a Postdoctoral Research Fellow with the CYENS—Centre of Excellence, Smart, Ubiquitous and Participatory Technologies for Healthcare Innovation (HealthXR) Group, Nicosia, Cyprus, since 2018, and the eHealth Laboratory and the CIL Laboratory, Department of Computer Science, University of Cyprus, Nicosia, since 2013. Her research interests include eHealth and supporting technologies, medical diagnostic systems, computational intelligence, multimedia applications, and virtual environments.

KLEANTHIS NEOKLEOUS received the B.S. and M.S. degrees in mechanical engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2004, the M.S. degree in space science and technology from the Lulea University of Technology, Lulea, Sweden, in 2005, the double M.S. degree in space systems and automation from Czech Technical University, Prague, Czech Republic, in 2006, and the Ph.D. degree in computer science from the University of Cyprus, Nicosia, Cyprus, in 2011.

He has been the Multidisciplinary Research Group Leader with the CYENS—Centre of Excellence, Immersive Technologies for Intelligent and Creative Applications (ITICA) Group, Nicosia, since 2019. His research interests include computational intelligence and eHealth.

CONSTANTINOS S. PATTICHIS (Fellow, IEEE) is currently a Professor with the Department of Computer Science; the Director of the Biomedical Engineering Research Centre, University of Cyprus; and the Leader of the HealthXR Smart, Ubiquitous, and Participatory Technologies for Healthcare Innovation, CYENS—Centre of Excellence. He has 30 years of experience in eHealth and connected health, medical imaging, biosignal analysis, intelligent systems, and explainable AI, and more recently in mHealth interventions based on X Reality applications. He has been involved in numerous projects in these areas funded by EU and other bodies, with a total funding managed close to 15 million Euro. He has published 130 journal publications, 236 conference papers, 30 chapters in books, an editor of three books, 22 journal special issues, and 19 conference proceedings in these areas.

Dr. Pattichis is a fellow of IET, the International Academy of Medical and Biomedical Engineering (IAMBE), and European Alliance for Medical and Biological Engineering and Science (EAMBES).