Quali-Mat: Evaluating the Quality of Execution in Body-Weight Exercises with a Pressure Sensitive Sports Mat

BO ZHOU, SUNGHO SUH, VITOR FORTES REY, CARLOS ANDRES VELEZ ALTAMIRANO, and PAUL LUKOWICZ, German Research Center for Artificial Intelligence, Germany and University of Kaiserslautern, Germany

While sports activity recognition is a well studied subject in mobile, wearable and ubiquitous computing, work to date mostly focuses on recognition and counting of specific exercise types. Quality assessment is a much more difficult problem with significantly less published results. In this work, we present Quali-Mat: a method for evaluating the quality of execution (QoE) in exercises using a smart sports mat that can measure the dynamic pressure profiles during full-body, body-weight exercises. As an example, our system not only recognizes that the user is doing push-ups, but also distinguishes 5 subtly different types of push-ups, each of which (according to sports science literature and professional trainers) has a different effect on different muscle groups. We have investigated various machine learning algorithms targeting the specific type of spatio-temporal data produced by the pressure mat system. We demonstrate that computationally efficient, yet effective Conv3D model outperforms more complex state-of-the-art options such as transfer learning from the image domain. The approach is validated through an experiment designed to cover 47 quantifiable variants of 9 basic exercises with 12 participants. Overall, the model can categorize 9 exercises with 98.6% accuracy / 98.6% F1 score, and 47 QoE variants with 67.3% accuracy / 68.1% F1 score. Through extensive discussions with both the experiment results and practical sports considerations, our approach can be used for not only precisely recognizing the type of exercises, but also quantifying the workout quality of execution on a fine time granularity. We also make the Quali-Mat data set available to the community to encourage further research in the area.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing design and evaluation methods**; • **Computing methodologies** → **Neural networks**; *Knowledge representation and reasoning*.

Additional Key Words and Phrases: quality of execution, sports activity recognition, ubiquitous sensing, pressure mapping, datasets, neural networks

ACM Reference Format:
Bo Zhou, Sungho Suh, Vitor Fortes Rey, Carlos Andres Velez Altamirano, and Paul Lukowicz. 2022. Quali-Mat: Evaluating the Quality of Execution in Body-Weight Exercises with a Pressure Sensitive Sports Mat. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 89 (June 2022), 45 pages. https://doi.org/10.1145/3534610

1 INTRODUCTION

1.1 Motivation

Monitoring physical exercises is a well studied topic in mobile, wearable and ubiquitous human activity recognition [21, 59, 70]. It can help athletes to track their performances, and the broader consumers to document their activity levels and achieve their fitness goals. It also enables professional trainers to better understand the trainee's

Authors' address: Bo Zhou; Sungho Suh; Vitor Fortes Rey; Carlos Andres Velez Altamirano; Paul Lukowicz, FirstName.LastName@dfki.de, German Research Center for Artificial Intelligence, Trippstadter Str. 122, Kaiserslautern, Germany, 67663 and University of Kaiserslautern, Erwin-Schrödinger Str. 52, Kaiserslautern, Germany, 67663.

2474-9567/2022/6-ART89 $15.00
https://doi.org/10.1145/3534610

progress, and social network competitions among friends to promote a healthy and active lifestyle. Fitness tracking based on inertial measurement unit (IMU) sensors is among the most studied methods, which have already found their entry into commercial applications in the forms of fitness trackers, and smartphone / smartwatch apps. Sensors are also integrated into fitness equipment such as treadmills and cycling bikes to provide users realtime information about their workouts (e.g. Technogym®, NordicTrack®, and Pelaton®). With the recent global pandemic and public gathering restrictions, some of the smart workout equipment have also found an increasing market in private homes [1], as physical exercise has significant positive impact on people's quality of life during the pandemic period [58]. However, such equipment is typically restricted to a certain type of exercise and requires physical space at home (e.g. stationary bikes or treadmills), which is not accessible for everyone. According to a recent survey [91], body-weight exercises are among the most popular home workouts thanks to the flexible choices of exercise types, location (e.g. both indoors and outdoors), and low cost, light weight equipment (e.g. resistant bands, mobile weights, fitness mat, etc.). While wearable devices such as smartwatches with IMUs can be used for detecting some body weight exercises [52], more comprehensive monitoring requires sensors to be placed on multiple body locations. This can be cumbersome in everyday settings.

Computer vision based methods can effectively extract poses and motions from typical upright positions. However, they provide poor results with non-typical positions such as horizontal and upside-down poses [15]. For applications inside private homes, the uncertainty of restricted camera ranges and angles, and poor indoor lighting conditions further increase the difficulty and complexity for computer vision based approaches as have been found in a recent study [35]. While it can be finicky to set up the proper camera angle before every workout; laying down a smart fitness mat is very intuitive and unobtrusive from the usability perspective, as it is an item that is typically already required for the exercises. As an alternative, Sundholm, et al. [83] has proposed using unobtrusive pressure mapping sensors in a sports mat format to detect body weight exercises that people commonly perform on sports mats such as push-ups, crunches, etc. The work was focused on categorizing and counting exercises.

In this work, we take this research a step further by moving from mere categorization or counting towards evaluating detailed quality of execution (QoE) during the exercises. This is in contrast not just to previous works on pressure sensing mats but also to much of the existing work in sports activity recognition in general. Most previous works in this area stay on the level of coarse analysis restricted to exercise categorization and counting as we further discuss in Section 2. On the other hand, in the sports science discipline, quality of execution has been proven to be more important than mere quantity [13, 33]. QoE evaluation of physical exercises outside the sports science and biomechanics disciplines still largely relies on subjective methods such as personal coaching and questionnaires [74]. This is partially due to the fact that the definition of quality is often ambiguous. For example, a verbal description of engaging a certain group of muscle is subjective to personal perception biases. However, most aspects of QoE can be quantified such as range of motion, stance and postural correctness, and minor variations (e.g. side or bilateral variations) of certain exercises [26, 34, 39].

1.2 Novelty and Contribution

In this paper, we have developed a method to automatically detect the quality of execution (QoE) thus beyond categorization in a wide range of popular full-body, body-weight exercises with appropriately selected and designed ubiquitous sensing hardware and machine learning models. To the best of our knowledge, the extent of category variety combined with the level of subtlety to differentiate the QoE we have targeted is beyond the state of the art which still largely remain on the categorization level. Specifically, we make the following contributions:

(1) We have designed and conducted an experiment involving subtle differences in the execution of common body weight exercises that sports science literature identifies as important for the quality and type of stimulation of various muscle groups with 12 participants and a smart fitness mat prototype. Overall, we

considered 9 exercise categories with a total of 47 variants. The recorded data set is not only used in the paper but is also made available to the community for further research.[1]

(2) We have explored a range of state-of-the-art machine learning algorithm candidates from classical curated features with probabilistic classifiers to various state-of-the-art deep learning models with respect to their performance on the specific type of signals produced by pressure mats in our application domain.
(3) Through the same leave out validation criteria, we have identified an appropriately adapted, efficient Conv3D model that outperforms all other methods. It can recognize the 9 exercise categories with 98.6% accuracy and macro F1 score, and 47 QoE variants with 67.3% accuracy and 68.1% macro F1 score in the leave-recording-out setting; and in the leave-participant-out setting, 96.9% accuracy and macro F1 score for categorizing 9 exercises and 56.1% accuracy and macro F1 score for differentiating 47 QoE variants.
(4) We demonstrate that the proposed methods can be used for evaluating the quality of execution in most cases with detailed discussions of both machine learning results and real-world sports considerations.

1.3 Paper Structure

Section 1 introduces the motivation and contribution. Section 2 reviews the related work in modalities and algorithms involving pressure sensors for activity recognition, and specifically sports activity recognition in the broader mobile, wearable and ubiquitous technologies. Section 3 explains the data acquisition procedure to produce a dataset for evaluating the QoE in body weight exercises. The implementation of machine learning algorithms are discussed in Section 4 with an exploratory approach. Section 5 presents the machine learning results of recognizing the exercises and variants, with discussions related to practical sports science concerns. Section 6 further discuses the results on the broader level of ubiquitous and wearable computing. The ablation considerations during the model design process and limitations are also discussed in this section. Section 7 concludes the entire paper.

2 RELATED WORK

To the best of our knowledge, our contributions stated in Section 1.2 offers novelty beyond the existing state of the art, specifically in detecting the quality of execution in sports exercises with mobile, wearable or ubiquitous technologies. Most existing works stay on the level of recognizing the category (e.g. push-up, sit-ups, etc.), not quality of execution (e.g. range of motion, stance, etc.) The few on evaluating QoE are not as comprehensive (e.g. 7 variations of push-ups in [6]) or lack scientific support from sports science knowledge, which we adhere to while defining the experiment protocols in Section 3.2.

Table 1 lists comparable studies in ubiquitous sports exercise recognition concerning categorization and QoE analysis. While higher accuracy numbers have been reported with random k-fold in [6, 32, 52, 82], we consider only proper leave-out splits in Table 1, as in [83] it was pointed out for repetitive exercises random k-fold with sliding windows introduces over-fitting; and classification based on time granularity during a small period in repetitive actions is over optimistic. For example, people tend to perform a cyclical motion with little variation during a few seconds, while the motion may be carried out differently in another recording session. Such splits would not represent practical recognition performance when the system is being tested with a new session or new user.

2.1 Activity Recognition with Pressure Sensors and Related Algorithms

Since high performance robust pressure mapping hardware with large scale (covering 1-by-2 meters), high resolution (<2cm pitch), high dynamic range (24bit) and high refresh rate (>20 frames per second) is scarce until recent commercially available development kits, there has not been much work investigating human activity

[1]The dataset is available on https://github.com/drbz-zhou/Quali-Mat.

Table 1. Comparable Work in Ubiquitous and Wearable Sports **Categorization (Cat.)** and **Quality of Execution (QoE)**

Study	Sensing Modality	Activities	Recognition Algorithms	Cat. Acc.(%)	QoE Acc.(%)	
Smart-Mat	pervasive pressure matrix (80x80cm)	10 full body exercises	statistical features, probabilistic classifier	88.7	none	[83]
W8-Scope	pervasive IMU on weight stack	10 gym exercises	statistical features, probabilistic classifier	93.8	none	[69]
Leg-Band	wearable pressure matrix (16x8cm)	5 gym leg exercises and non-workout	statistical features, probabilistic classifier	93.3	none	[100]
Capacitance	wearable passive capacitive sensors	7 gym exercises	neural network (TCN)	54.7	none	[9]
MM-fit	wearable IMUs (4) and RGBD camera	10 full body exercises	neural network (Autoencoder)	96.3	none	[82]
MyoGym	wearable IMU and sEMG	30 upper body exercises	statistical features, probabilistic classifier	71.6	none	[51]
IMUTube	wearable IMUs and video	11 and 7 mostly upper body exercises from 3 datasets	neural network (multi-modal pipeline) and virtual IMU augmentation	62~73	none	[52]
Sonar	active sonar from smartphones	8 full body exercises plus idle	neural networks (CNN, VGG16, LSTM)	81.3	none	[30]
GymCam	RGB camera	5 common weight lifting exercises	optical flow and statistical features	93.6	none	[48]
Push-ups	wearable IMUs	1 push-up exercise with 7 variants	statistical features, probabilistic classifier	none	7 classes 80.3	[6]
Rep-Penn	RGB camera	7 weighted and bodyweight exercises	pose joint heatmaps and neural networks	95.7	none	[96]
This Work	pervasive pressure matrix (1x2m)	body weight exercise **9 categories** and **47 variants**	neural networks (3D CNN and other models)	98.6	47 classes 67.5	

recognition with pressure sensor matrix. A fine-grain array of force sensitive resistors (FSRs) were placed inside gloves to recognize exercise activities in [3]. Textile pressure sensors were integrated inside a leg band to recognize gym leg exercises in [100]. In [67], a tactile floor covering 4 square meters (96×96 points at 2cm pitch) was used for detecting human and robot's positions in a smart factory setting. In [80], a pressure mat was used to analyse healthy cats' gait and jumps by statistical analysis on the signal characteristics.

As previously mentioned, using a prototype sensing mat of 80cm × 80cm with 1cm pitch between pressure sensitive points, Sundholm, et al. detected 10 exercises with 88.7% accuracy in leave-session-out settings [83]. The exercises defined were mostly drastically different from each other, and the machine learning algorithms were based on statistical features and probabilistic classifiers. Singh, et al. developed a transfer learning approach for identifying footsteps with a pressure mapping carpet in [78]. An InceptionV3 model with the trained weights from the ImageNet dataset was used as the feature extractor, and the last classification layer was replaced with a new blank dense layer for the specific classification tasks. However, the transfer learning approach with ImageNet models does not address the temporal nature of the dynamic pressure mapping data. In the mentioned work of [78], the time domain was simply collapsed by calculating the average over time per pixel, or selecting the frame with the highest average pixel intensity during the time window. In [17], Clever, et al. used a pressure mat on an adjustable hospital bed to estimate 3D poses of the patient lying in the bed using 2D convolutional neural networks. However, the study was limited to static, face-up lying poses without temporal sequences.

To extract useful information from pressure sensors that create an imagery of the surface pressure profile, image processing methods have been the general option. In statistical features, shape descriptors like image

moments were used such as in [83]. With deep learning, 2D convolutional layers, or transfer learning with models pre-trained with RGB pictures were investigated. For dynamic human activities, however, the time domain sequence is also very important. Temporal features such as fast Fourier or wavelet transforms were introduced for this aspect in [98]. In the broader activity recognition field, combinations of convolutional layers for the sensor channels and long short-term memory (LSTM) layers for the temporal domain were evaluated for motion sensors [64]. However, these methods isolate the spatial domain and temporal domain in different processes and the characteristics that involve simultaneous spatial-temporal domains are ignored, such as a ‘ pixel’ of a repetitive temporal pattern moving within a spatial region. In the video recognition discipline, 3D convolutional models were developed to address this problem, such as the C3D network [29], which could be a promising option for the dynamic pressure mapping sensor data.

2.2 Broader Mobile, Wearable and Ubiquitous Technologies in Sports Activity Recognition

Wearable motion sensors such as IMUs have been extensively studied in the sports activity recognition field especially in categorizing exercises [6, 21, 37, 40, 52]. However, studies with IMU-based systems have not yet provided such performance on body weight exercises. For example, in [52], with 13 exercises related with arm motions, wrist-worn IMUs achieve less than 70% accuracy with data augmentation from video-generated virtual IMU data. With filtering techniques to remove aspects like noisy postures, occlusion, foreground and background, the accuracy is still less than 80%, while those filtering layers may have removed crucial quality metrics such as incorrect postures. In [6], a quality of execution study focused on variations of push-ups with smartphone and smartwatch IMU sensors were able to distinguish the variations with above 80% accuracy. However, this study contains data only from push-ups, i.e. the accuracy is based on the assumption that the system can recognize push-ups with 100% accuracy, which is not realistic as other studies have suggested [52].

On the broader sports activity recognition landscape, except for approaches based on pressure mats and well-studied wearable IMUs, many novel sensing systems have been proposed as reviews such as [84, 89] showed. In [9], 3 passive capacitive sensors were placed on the body to recognize and count gym exercises. Surface electromyography (sEMG) is the golden standard to investigate muscle activation in sports science, in which high quality medical grade instrument is usually used [22, 79, 87]. Commercial sEMG devices have also been used to recognize physical exercises, such as the upper-body exercise data set recorded with the MYO®armband in [51]. Combinations of different sensors can also help recognize physical sports activities such as IMUs with photoplethysmography [8]. [82] presented a dataset combining RGB-depth camera, multiple on-body IMUs and heart rate during full-body workouts. In [69], a smartphone is used on the weight stacks of gym equipment to recognize different types of exercises using its built-in IMU. Smart textiles can also be combined with wearable garments to detect sports activities such as capacitive pressure sensors [38] or stretch sensors [7]. In [30], a smartphone was modified as an active sonar with the doppler effect, placing next to the user. Image recognition deep learning models were used to analyse the spectrogram and classify 8 types of exercises plus an idle state.

Estimating body poses from images has seen rapid improvement from early research on convolutional pose machines [90] to the current ready-to-deploy pose estimator models such as OpenPose[15]. Temporal relationships from video streams have also been evaluated with VideoPose3D [66] to improve the temporal consistency compared to individual frame-based methods. In [32], the OpenPose estimator was used to recognize five body weight exercises (Push-up, Squat, Plank, Forward Lunge, Sit-up). While 98% accuracy was reported with 10-fold cross-validation (without leave-out splits), the experiment requires a camera setup that the entire person is clearly visible during different exercises, which requires considerable empty space between the camera and the user, as well as equipment such as a tripod. Similar solutions based on computer vision and pose estimation are already commercially available in the form of smartphone apps (e.g. Onyx Home Workout [45], VAY Fitness Coach [2], etc.), which was evaluated in a user study with private home settings in [35]. The study found that inherent

limitations of computer vision significantly discourages the users, with all participants reporting pose detection issues and most participants complaining the space at their homes was not enough to accommodate the visibility needed for continuous pose tracking. In the research field of computer vision based pose tracking, atypical postures usually lead to false pose estimations, which often occur during body-weight exercises. This may be due to the fact that the mainstream datasets such as MS-COCO [56] consist mostly of upright and standard postures; and in atypical postures, self-occlusion often occur, that the users' own body part would block the view of the other parts. To date, atypical pose tracking is still a challenge for pose estimation [42, 43, 68]. While dedicated datasets and machine learning models in clearly defined spaces can solve niche scenarios such as pose estimations inside a car with wide angle cameras [61, 71]; it is unclear how such approaches can be implemented in the the unpredictable private home settings [96]. Apart from pose detection approaches, other computer vision methods such as optical flow have been used to detect motions. In the work GymCam [48], with such an approach, a camera overlooking a gym is used to recognize multiple users' weight lifting exercises. However, the method is restricted to moving the defined weights at the prescribed locations, which does not include exercises such as body weight exercises unconstrained by gym bench surface and weights. A combination of inertial sensors and computer vision methods was used to evaluate the key performance indicators during home exercise in [4], examining the movement range during kicking exercises.

3 STUDY DESIGN

3.1 Apparatus

The sensing hardware was a research prototype based on the system found in [99] which was designed by implementing the architecture proposed in [97]. It was developed during the EU project SimpleSkin (2013-2016). We describe the sensing principle in the following paragraph, while the open-access hardware resource can be used to reimplement our hardware. [2] On the other hand, a more cost-effective way to replicate our experiment setup is to procure a similar commercial product solution such as the fitness mat dev kit from SensingTex®(Spain), which has different technologies and specifications from our prototype but the same physical information.

The pressure sensing materials are pressure sensitive electrical resistance fabrics, produced by Sefar AG (Switzerland). The pressure sensitive middle layer is a carbon polymer fabric Carbotex®(Sefar®). It is sandwiched between two perpendicular layers of parallel conductive lines with 0.7cm width and 1.3cm spacing (1.5cm pitch). The thin fabric sensor matrix is protected with a thin felt layer (0.7mm). The pressure sensor matrix covers an area of approximately 1m-by-2m. A normal commercial foam yoga mat (standard 183cm-by-61cm size, 6mm thickness) is placed on top of the felt protection layer to provide user with a softer surface. Since our experiment is conducted in an open indoor space with hard ceramic tiles, the pressure mat is placed on top of a larger layer of normal soft foam sports floor mat, so that the participants can perform all activities on the foam surface comfortably. There are 128×64 points with 1.5cm pitch, each is sampled at 25 frames per second by eight parallel, 24-bit 16-channel high performance ADCs ADS1258, controlled by an FPGA EP3C25Q240. For the scanning process, the FPGA uses its GPIO output pins to turn on one of the 64 columns at a time, and the parallel ADCs sample the voltage of the 128 rows via a voltage divider structure. The scanning result is sent to a computer by an USB cable with the FTDI®8-bit FIFO protocol. The USB cable also supplies the power to the entire system. A C++ driver program was created to receive and reconstruct every 2D frame on a Windows computer.

3.2 Activity Definitions

We designed an exercise routine based on sports science literature and recommendations of the fitness trainers from the gym Unifit (Kaiserslautern, Germany) who gave out courses of body weight exercises. We focused on

[2]The hardware design sources can also be found in the online repository.

Table 2. Quality of Execution in Workout Class Definitions (Part 1)

Variant	Exercise	Variation and Descriptions
Category I : Crunches		
1	Reference: Lie Down	knees up, feet down
2	Crunches I	lower range of motion, regulated by hands pointed to the sky with arms straight
3	Crunches II	medium range of motion, regulated by hands reaching to the knees with arms straight
4	Crunches III	higher range of motion, regulated by elbows reaching to the knees with arms straight
5	Side Crunches I	lower range of motion, regulated by hands pointed to the sky with arms straight
6	Side Crunches II	medium range of motion, regulated by both hands pointed to each knee alternatively
7	Side Crunches III	higher range of motion, regulated by elbow touching the opposite knee alternatively
Category II : Push-ups		
8	Reference: Hold Push-up	hands shoulder wide
9	Push-up I	hands shoulder wide and half range of motion, regulated by not fully extending elbows
10	Push-up II	hands shoulder wide and full range of motion, regulated by having elbows fully extended
11	Push-up III	hands wider than shoulder, full range of motion
12	Alternating Push-up	alternating single side push-up, hands wider than shoulder
Category III : Planking		
13	Reference: Standard Plank	depending on the person, either feet or knees can be on the floor, thighs and torso keep on a straight line
14	Slack Plank	torso not engaged so the spine bends downwards naturally
15	High Hip Plank	raise hip higher than standard straight plank
16	Plank Dip	Standard Plank but move body forward and backwards around the elbow support
17	Plank Push-up	hands are chest wide, change between plank position and push-up position
18	Chest Wide Plank Push-up	hands are chest wide (same as 17), only push-up without placing elbows on the mat
Category IV : Only Back Visible		
19	Reference: Back	lift legs and elbows, place hands on the back of the head
20	Leg-up Crunches	lift legs, crunching exercise till elbows touch the knees
21	Alternating Cycling	lift legs, alternately touch one elbow with the opposite knee, while the other opposite elbow-knee pairs are extended
22	Leg-only Cycling	lift elbows, empty cycle with only leg motions
23	Leg-lift I	raise legs from flat position to vertical position, with lower back pressing on the mat, this variation engages the abdominal muscles more effectively
24	Leg-lift II	same leg motion as 23, but have lower back suspended with an arched spine, and instead use hip as the anchoring point, this variation engages the abdominal muscles less effectively

9 categories of exercises in this study. The 9 categories were generalized groups within which the body print on the mat looks similar: I crunches, II push-ups, III planking, IV only back contacting the mat, V back and arms contacting the mat, VI standing, VII arch, VIII bridge and IX side planks. Within each category, we defined finer variations of the exercise, which in the real world can reflect details from correct exercise techniques to qualities, as well as a reference position where the participants stayed static. In total, **47 variants** were defined for the 9 categories, as listed in Table 2 and Table 3. Overall, all variations in each exercise category were decided as they lead to different levels of core muscle activation according to sports science studies with sEMG data [12, 14, 27, 63].

(1) Category I crunches mainly exercises the rectus abdominis muscles, with the side variations also exercising the oblique muscles. we defined a static pose of lying down with knees up, and 6 variations of 3 levels of height of the upper body position combined with either central motion or side motion [27]. The range of

Table 3. Quality of Execution in Workout Class Definitions (Part 2)

Variant	Exercise	Variation and Descriptions
Category V : Back and Arms Visible		
25	Reference: Hold Leg-up	with arms and back relaxed on the mat, while thighs are raised up vertically
26	Leg-raise I	raise legs vertically and upwards from the relaxed horizontal position with lower back suspended
27	Leg-raise II	similar as 26, but always keep thighs upwards without lowering legs to the horizontal position, this variation engages the torso muscles more than Leg-raise I
28	Leg Swing I	swing legs from the left and horizontal, then upwards vertical, to the right horizontal positions, with knees bent
29	Leg Swing II	similar as 28, but knees are extended straight
Category VI : Standing Up		
30	Reference: Stand	feet shoulder width, feet angle is decided by squeezing glutes
31	Shallow Squat	empty squat, half range of motion, regulated by the hands reach the knees with relaxed arms
32	Deep Squat	empty squat, full range of motion, regulated by the elbows reach the knees with folded arms
33	Toe Touch	bend downwards and reach toes with fingers repetitively, heels may leave the mat
34	Tip Toe	raise the heels repetitively
Category VII : Arch Step		
35	Reference I	alternately step forwards and then back, with one foot staying behind
36	Reference II	alternately step backwards and then front, with one foot staying in front
37	Lunge Forwards	step similar as 35, but with the knee behind bent downwards
38	Lunge Backwards	step similar as 36, but with the knee behind bent downwards
Category VIII : Bridge		
39	Reference: Bridge	with feet or heels, upper back, shoulders, arms and head staying on the mat, suspend the hip and lower back so that the thighs and torso are on the same slope line
40	Hip Thrust	same pose as 39, but move hip up and down with the help of glutes
41	Single Hip Thrust	same motion as 40, but with only one leg supporting the body and the other leg extended
Category IX : Side Plank		
42	Reference: Side Plank	body face to one side, with one foot or knee and one elbow of the same side supporting the body, torso and legs are straight on the same line
43	Slack Side Plank	similar pose as 42, but with the torso relaxed
44	Side Hip Thrust	similar pose as 42, and move hip up and down
45	Side Leg Raise	similar pose as 42, and raise and lower the leg that is not supporting the body
46	Arm Swing I	similar pose as 42, and swing the free arm from below the body to the vertical upwards position, face the front of the body
47	Arm Swing II	similar motion as 45, but face the moving hand

motion does not necessarily scale with muscle activation. The medium variant is considered to have more, sustained rectus abdominis muscle activation as the muscles are constantly supporting the upper body's weight. In the high range of motion, the abdominis muscle is at rest at the high-end position, but the high range requires more short burst muscle activation from the low-end position [81].

(2) Category II push-ups are well known for exercising the chest muscle groups (pectroralis major and deltold) and triceps, we defined wide or narrow hands positions as the narrower position is observed with more muscle activation [18]. The alternating variant has a similar effect of even narrower hand positions, and focuses more body weight on one side at a time[23]. While muscle activation also differs with different ranges of motions [6, 55], the variant we introduced is without fully extending the elbow, which exerts constant muscle activation, especially triceps, compared with when the arms are fully extended [72].

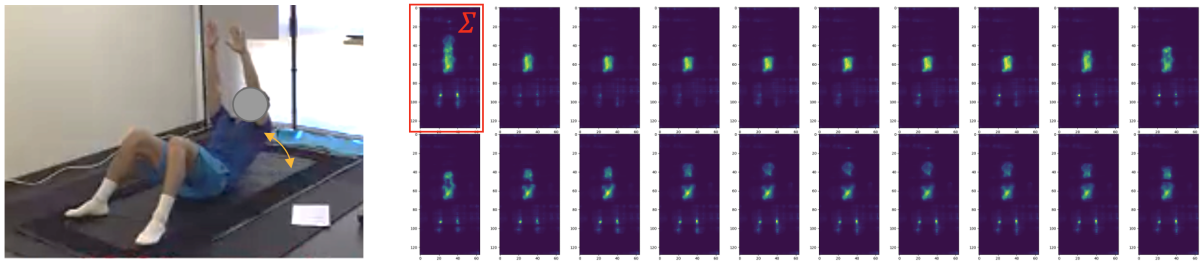
(3) Category III planking focuses on the stability of many trunk muscles from lumbar muscles, rectus abdominis and obliques to serratus muscles. Slack planking has little exercise effect on those muscles due to the lack of engagement [95]; while a high hip plank with a posterior pelvic tilt increases muscle activation [75]. Animated planking (plank dip) with contracting ankles introduce up to 60% more effective isometric contraction [16]. The combination of planking with push-ups were recommended by the licensed fitness trainers, which is an advanced exercise emphasising the serratus muscles and triceps in addition to all the benefits of push ups and planking.
(4) Category IV contains several different exercises with back supporting the whole body. Exercises involving raising the legs train mostly the lower rectus abdominis and transversus abdominis muscles [62]. Leg-up crunches and lying position cycling involve both the upper body and lower body for a higher level of core stability [44]. The fitness trainers suggested that leg lift can often be performed incorrectly if there is too much transverse plane rotation on the pelvis and the work from lower abdominal muscles is replaced by hip flexion, failing to reach the exercise goal [41]. And the common coaching recommendation is to perform Variant 23 Leg-lift I to keep lower back pressed on the ground as listed in Table 2.
(5) Category V contains more leg raising exercises when it is necessary to use arms on the side for ground support. The leg-raise exercises are different from the leg-lift variants in Category IV, as leg-raise involves raising the hip with the support of lumbar, thoracic, abdominal and serratus muscles. Leg-raise without lowering the legs to resting position exerts constant muscle activation of the above mentioned muscle groups, while lowering the legs will introduce more rectus abdominis muscle activation [41]. Leg swing mainly activates the oblique muscles [62].
(6) Category VI involves standing pose exercises. The muscle activation, mostly gluteus maximus and quadriceps, is stronger in deeper squats when unloaded [19]. Toe touch is a stretching exercise that improves the flexibility of the hamstrings, calves and lumbar spine [31]. Tip toe mainly exercises the calf muscles [53].
(7) Category VII contains lunge exercises, which have been shown to have more effective leg muscle activation than squats during body weight exercises [25]. Specifically, forward lunge is found to be even more effective than backward lunge [65]. We use stepping forward and backward without the lunge action to differ from the actual effective exercise.
(8) Category VIII bridges are common body-weight exercises for the gluteus maximus muscles [94]. The muscle activation intensity increases from static bridge to animated hip thrust, to single leg hip thrust [49].
(9) Category IX side planks work on the trunk muscles, especially the oblique muscles [5]. The static variants focus on isometric contraction, while the moving variants are intended for dynamic strengthening and stability of these muscles [20, 57].

The pressure map examples of every category and variant are shown as summary frames in Fig. 10, Fig. 11 and Fig. 12. Therefore, there are several classification goals of this study: recognizing the 9 categories (similar with the work in [83]), 47 variants, or the variants within each category.

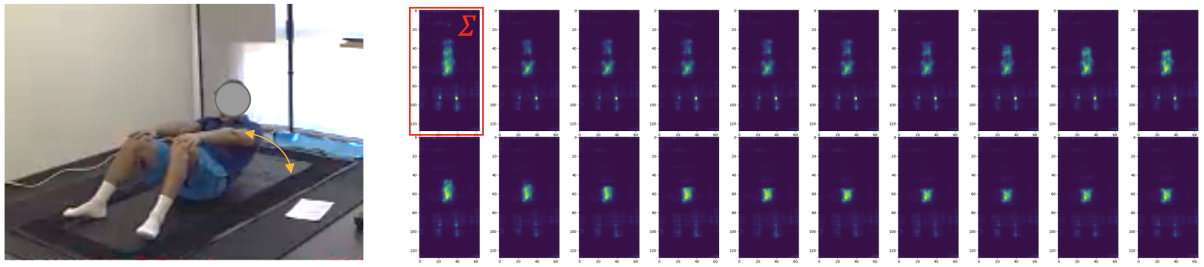
We emphasise that the variants are designed to be meaningfully different while having only trivial differences to distinguish the workout QoE. The 9 categories are similar with the 10 exercises from the work by Sundholm et al. [83]. While the pressure maps of variants within a category can look indistinguishable for the average human observer, such as the examples we show in Fig. 1.

3.3 Data Collection Details

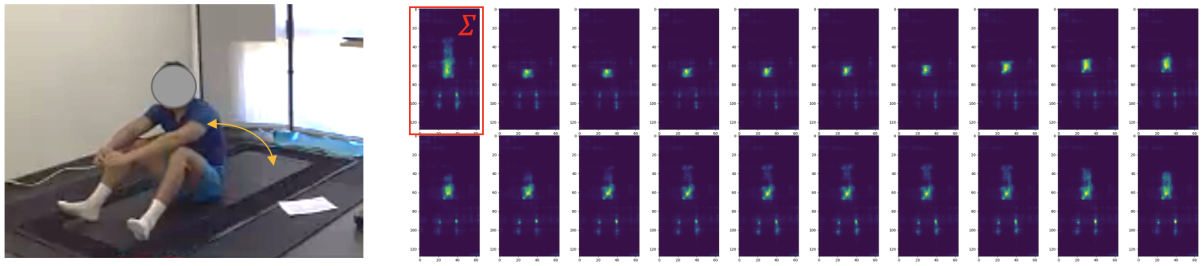
Twelve healthy participants aged between 24 and 33 took part in the data collection, out of which four were female and eight were male. The heights of the participants were between 162cm and 189cm. The experiment was set up in an indoor and spacious environment. Every participant performed three recording sessions across different days. Each recording session consisted of a routine of 47 exercises directed by the instructions in Table 2



(a) Class 2: Crunches I. Lower range of motion, regulated by hands pointed to the sky with arms straight.



(b) Class 3: Crunches II. Medium range of motion, regulated by hands reaching to the knees with arms straight.



(c) Class 4: Crunches III. Higher range of motion, regulated by elbows reaching to the knees with arms straight.

Fig. 1. Pressure map examples of 3 exercise variants of randomly selected 2-second windows. The photo depicts the end position of the range of motion from lying down. The first pressure map is the summary of the time window, and the following pressure maps are frames inside the selected window. Only every other frame, and only 19 of the 50 frames of the window are shown so that most of the time window can be fit in the page.

and Table 3. For each of the dynamic movements, the participants performed 10 continuous repetitions; for static classes (e.g. reference poses of each category and static planking), the participants were asked to remain in position for at least 10 seconds, holding it up to 30 seconds if possible. Before the first recording, the experiment conductor showed recorded video instructions of each exercise. The participants started with the exercise while fully rested and could take breaks at any time during the recording, to make sure each exercise was properly conducted in the sense that their movement was not distorted by fatigue. They wore appropriate sport apparels without shoes during the experiment.

The participants were of different fitness levels from beginner to enthusiast. We accepted different levels of execution depending on the participant's level. For example, a beginner participant could do all the exercises that require hips up in the air with knees touching the floor, instead of the supposed feet touching the floor, such as

push-up and planking. The enthusiast and intermediate participants could complete the entire routine without breaks; while the others required a long period of breaks in the middle of the recording session. Each recording session lasted between 25 minutes and 1 hour, depending on the participants' skill level and the resting time.

Personal and equipment hygiene procedures were carefully observed. The participants gave informed consent in accordance with the policies of the Ethics Team of the German Research Center for Artificial Intelligence (DFKI), which approved the experimental protocol. The dataset was manually annotated according to the video recordings to mark out the starting and finishing timestamps of each variant.

Without invading the participants' privacy, we show their body shape differences as pressure map examples from three selected activity variants in Fig. 2. We can see for example Participant 3 (189cm) and 4 (185cm) were the two tallest while Participant 10 (162cm) was the shortest. Participant 3 and 6 needed to have their shin and knee on the floor during side plank exercises due to their beginner level. Some participants (2, 3, 7, 12) also needed to use their knees to support instead of feet during the push-up exercises.

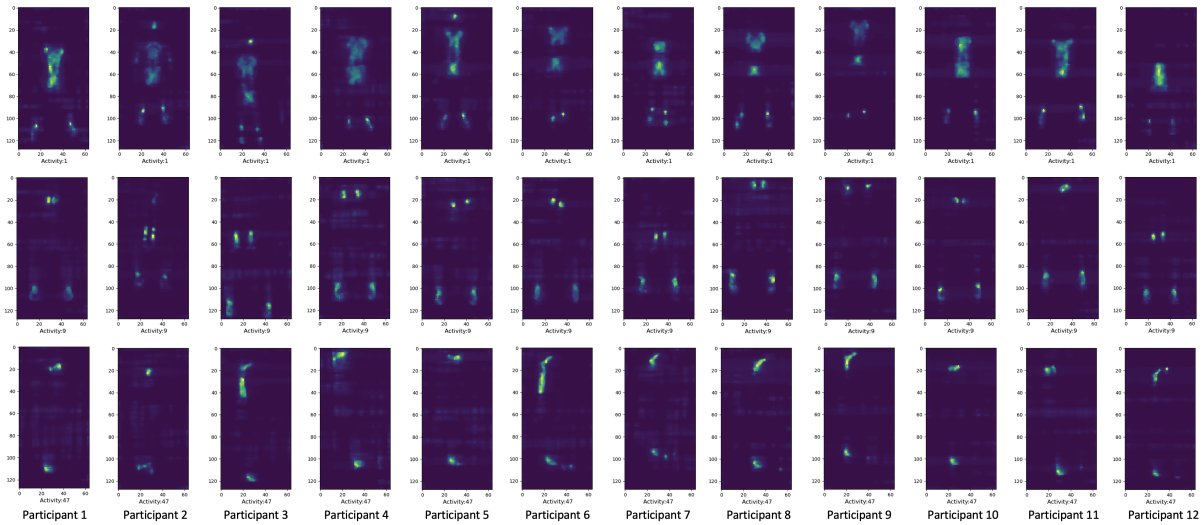


Fig. 2. Pressure map examples of every participant performing activity variant 1 lie down (first row), 9 push-up I (second row), 47 arm swing (third row). The participant's head aligns towards the top of the pressure map in the first row, and towards the bottom of the map in the second and third rows. Pressure maps are shown as summary frames within a time window.

4 EXPLORATORY APPROACH FOR MACHINE LEARNING ALGORITHMS

In this section we describe our exploratory approach for developing the machine learning classification algorithms. The data format of pressure mats is unique among other sensors: the output at every time step resembles an image; yet they are different from digital images from many aspects as listed in Table 4. The most fundamental difference is that for pressure maps, every pixel is a valuable sensor that contributes information on both its location and the point in time; while in digital images, individual pixels are much less significant. In addition, this work concerns time sequences of pressure maps which can be approximated as videos. However, the temporal sequence of pressure maps is in nature closer to physical sensors than to digital videos. For example, the time sequence of average value in a fixed region on the pressure map can be treated as a sensor channel [100], while this is typically not applicable for videos. Specifically, there has not been an established deep learning method for processing dynamic pressure maps.

Table 4. Characteristic Differences Between Pressure Maps And Digital Images

Aspects	Pressure Maps	Digital Images
Background	clean background (no pressure contact)	cluttered background
Content	pressure profiles with contours (of body parts)	details of local features within the shape
Occlusion	immune to soft material occlusion (e.g. clothing)	prone to opaque soft material occlusion (e.g. clothing)
Pixel value	single channel ADC readout (8-bit to 24-bit), every pixel offers valuable information	typically 8-bit 3-channel RGB, removing single pixels loses almost no information
Granularity	1.5 cm (this work)	<1mm
Scope	the object/person are usually only partially visible; the visible parts of the same object/person may not be connected	the object/person can be easily covered entirely in the image
Time series	individual pixels can be treated as sensor channels, more similar to multiple parallel channel sensors	objects/persons usually move to different pixels between frames, frames offer more useful information than time series of individual pixels

4.1 Dataset Preparation

The signals were pre-processed by standard spatial and temporal filtering to smooth out the jitters in the signal. Then within each frame, the raw 24-bit ADC output values of every pixel were normalized so that the mean value and standard deviation (SD) were both 0.5, thus mean + SD = 1 and mean - SD = 0. The data was saved with single precision floating point (FP32). The total dataset of 12 participants, each 3 sessions took 114GB of computer storage.

We first use leave-recording-out as the standard training and testing separation scheme. This would evaluate the system's performance on known users, which is suitable for private items, or shared items among a group of users such as gym members. To avoid over-fitting caused by action similarity in a short period of time, every session from a participant was recorded on different days and the participant wore different sport clothes. We targeted classifying both among the 9 categories and among the 47 variants. The time sequence of pressure maps were segmented with sliding windows, with lengths of 2 seconds (50 samples) and strides of 0.4 second (10 samples) within the continuous repetition period of every exercise variants. Event-based recognition is common in human activity recognition [11] after the prediction results on individual windows to smooth out the results in adjacent time windows, or majority voting within a longer event to improve the accuracy, with the assumption that people usually do not switch activities quickly every second. For example, in a similar work of detecting 10 exercises with a pressure mat [83], the event-based majority voting improved the accuracy by approximately 5%. However, we argue that this is not suitable in our use-case, as we want to distinguish minor variants and evaluate the quality of workout, which can happen spontaneously during the event in the real world. Therefore, instead of random K-fold, leave-recording-out validation with individual windows were evaluated: one recording session from every participants were left out, a random 10% portion of which was reserved for validation and early stopping, and the remaining 90% of the left out session was used for testing and generating the confusion matrix. No event-based or short period grouping was used. Since not all exercises lasted for the same duration, the dataset was balanced by randomly re-sampling the windows from the under-represented classes within each recording session.

4.2 Curated Features

Previous works for pressure mapping data classification [3, 60, 83, 99] have established a workflow with conventionally calculated features and probabilistic classifiers. The features would abstract information from the spatial domain (e.g. image moments) and the temporal domain (e.g. frequency features). We implemented the feature set proposed in [98] of 17 spatial descriptors times 39 temporal features in Matlab. The details of each feature can be found in [98], here we only briefly list the features. The spatial descriptors include average value, variance, range, entropy, mean absolute deviation, the center of mass coordinate, the centroid coordinated of the contour shapes, area, and Hu's seven image moments. They are calculated for each frame, and every window is converted to time sequences of the 17 spatial descriptors. Then the 39 temporal features are applied on the time sequence: average, variance, range, skewness, kurtosis, waveform length, sum of values grater than mean, power spectrum density mean frequency, average spectrum density of 5 equally divided frequency bands. Also, wavelet transform with 4 filterbank iterations was calculated and for each filterbank, the mean, variance, range, skewness and kurtosis of the coefficient vectors were calculated as the last temporal features. In the end, 17×39=663 features are calculated for each window. Through leave-recording-out cross-validation trial with the Matlab's Classification Learner app, we decided the bagged tree classifier offers the best performance of 73.5% for 9 categories (chance level 11.1%) and 38.7% for 47 variants (chance level 2.13%).

4.3 Deep Learning Models

In our exploratory process, we designed several deep neural network candidate models through examining the problem task (recognizing 9 exercise categories and 47 variants) and the data format (3-dimensional, spatial-temporal dynamic pressure imagery). The models were designed according to explainable architectures and the state-of-the-art that are suitable for the task and data format. They were tested with the same leave-recording-out validation scheme and the best candidate is a simple yet effective 3D convolution model (Conv3D). A leave-person-out validation was also carried out to evaluate how the models would perform for strangers.

All deep learning models were trained with the Adam optimizer [50] and the binary cross-entropy loss function. We followed the procedures and best practices to help avoid over-fitting, as recommended by [10]. The initial learning rate is 0.001 for the first 20 epochs, and is decreased by a multiplier of 0.1 for every 10 more epochs. Early stopping is used with a patience of 50 epochs to avoid over-fitting. The deep learning models were evaluated with a dedicated workstation, with an AMD Ryzen 9 5950X 16-Core Processor, 128 GB of system RAM, and NVidia RTX A6000 GPU with 48GB VRAM. The software environment was Tensorflow version 2.4.0 with Python version 3.8.11, on the Windows 10 operating system.

*4.3.1 3D Convolutions (**Conv3D**).* 3D convolution has been used for machine learning in the video domain [28]. We use 3D convolution layers to correlate the temporal and spatial hierarchical information, and 3D average pooling layers to condense the spatial-temporal dimensions. The model architecture is illustrated in Fig. 3 and summarized in Table 6. Since the spatial dimensions and the temporal dimension originate from different physical concepts, the 3D kernels and pooling size tuple can be different for the two spatial dimensions and the time axis. We add increasing numbers of filters as the pooling layers decrease the data dimension to avoid information loss. This process is iterated 4 times until the original input shape (128, 64, 50) is reduced to (8, 4, 5, 80) with the last number 80 as the channels from the convolution filters.

Normally at this stage, the layer output will be flattened and taken by a fully connected layer as the classifier [86] or several fully connected layers [64]. Fully connected dense layers dramatically increase the hyper-parameters with large input shapes, which is usually not a concern for limited channels of sensor data. However, since our spatial-temporal data is relatively large, flattening the layer output at the stage of (8, 4, 5, 80) in size will result in dense layers of millions of learnable parameters, making the model more difficult to train with a relatively small dataset. Therefore, we further reduce the layer output to construct an efficient model. An additional 3D

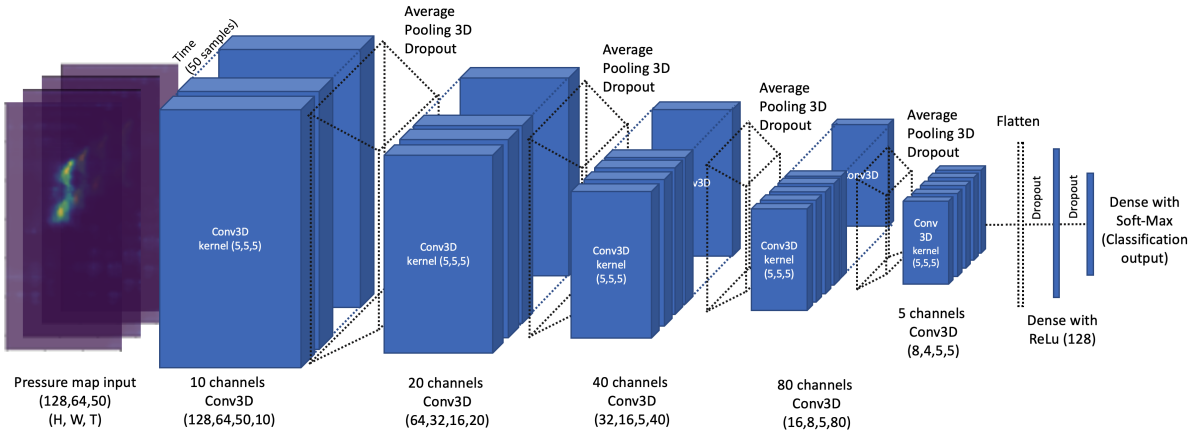


Fig. 3. Architecture illustration of the Conv3D model.

convolution layers with 5 filters is used to gradually simplify the latent features into less neuron outputs and reduce the channel size from 80 to 5. The outputs of the last 3D convolution layer can be considered linear combinations of the previous larger layers. The resulting intermediate tensor of size (8, 4, 5, 5) is flattened and a dense layer of 128 neurons with ReLu activation, and another dense layer with the number of classes and soft-max activation are used as the final classifier.

This way we can contain the model size under 1 million learnable parameters. In comparison, the smallest ImageNet models listed on Keras Applications is currently the MobileNetV2 [73], which has approximately 3.5 million parameters. Notably the Conv3D model deals with not only spatial image-like information, but time sequences of 50 consecutive frames. There are several benefits of smaller and efficient models such as the energy consumption savings and the possibility to be deployed on embedded microprocessors. For the validation process, the key benefit is that the model is easier to train with limited and niche dataset. [3]

4.3.2 Conv3D-Transformer4D. Transformer encoders with multi-head attentions [88] have gained much attention in the machine learning field thanks to its recent breakthroughs in natural language processing, out-performing LSTM based models for temporal sequences. With properly designed positional encoding, they are also suitable for spatial images [24] or temporal sequences of spatial coordinates [93]. We can also adapt the transformer model for our dataset. However, there is a problem to be solved first: following standard transformer encoder models, the (128, 64, 50) input will be encoded with 3D positional encoding, flattened, then fed to the inputs of the Q/K/V of the transformer encoder layer. However, matrix multiplication operations on the flattened vector with the size of 409600 will be performed inside the attention block, creating tensors with the size of 409600×409600. This is not a concern for text or image tasks as the input sizes are usually much smaller. But the matrix multiplication cannot be performed on the NVIDIA A6000 GPU with 48GB VRAM.

Therefore, we first reduce the input size through a few iterations of 3D convolution and pooling layers to support a realistic transformer encoder layer, resulting in a tensor size of (8, 4, 5, 5), with the last number 5 as the number of last convolution layer channels. Then we apply 4D sin/cos positional encoding, which is a direct extension of the 2D sine encoding methods from [93]. The transformer encoder layer is implemented according

[3] Although our dataset is 114GB, comparatively the 64×64 format of the ImageNet dataset is 6GB; our dataset is still very limited in terms of observation samples and negative classes.

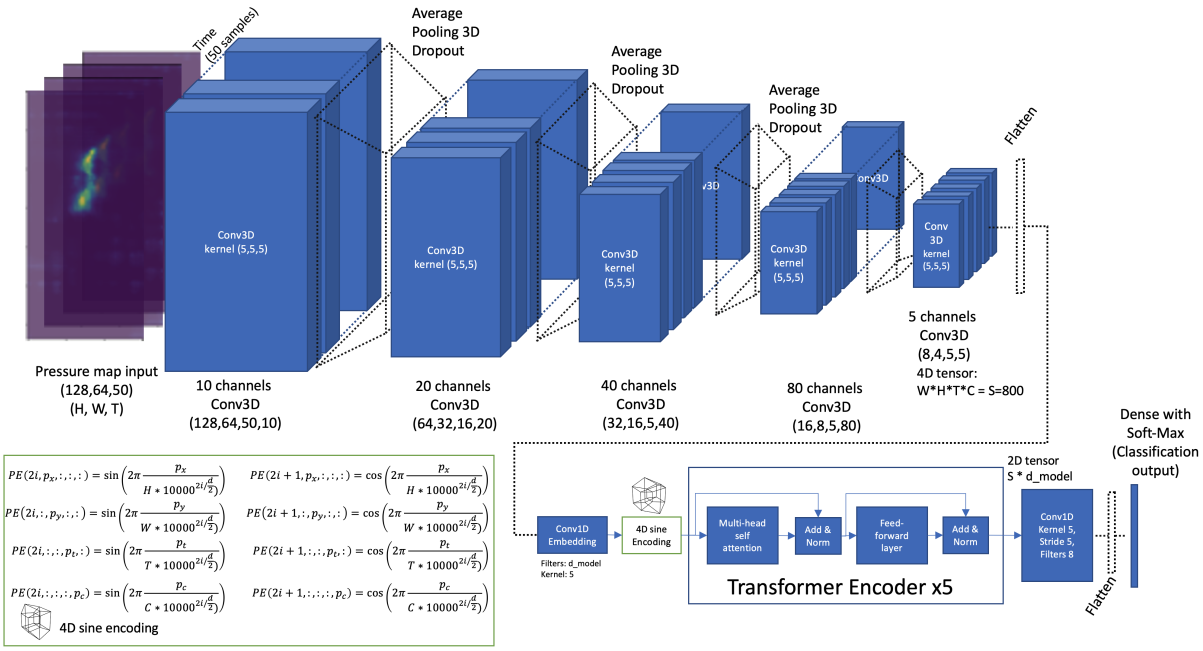


Fig. 4. Architecture illustration of the Conv3D-Transformer4D model including the 4D sine encoding scheme.

to [88]. Through trials, we have located the best performing configuration of the transformer encoders to be 4 heads, d_model = 32, d_ff = 32, and 5 encoder layers. The Conv3D-Transformer4D model is illustrated in Fig. 4.

4.3.3 TConv-ImageNet. In [78], a transfer learning approach was proposed for pressure maps to utilize image classification models with weights pre-trained by the ImageNet dataset to bring knowledge for spatial feature calculation, out-performing the curated spatial features such as image moments. However, transfer learning with image classification models does not solve the temporal characteristic of the data. In [78], the time dimension is simply collapsed by calculating the pixel-wise average in the time window, resulting in a single image per time sequence.

We propose to amend this transfer learning method by attaching a series of time-wise convolution-pooling operations to reduce the 50 time steps to 3 channels. So that instead of simple averaging, the information on the time domain can be extracted and preserved through learned convolution kernels. Then we use MobileNetV2 [73] as the base image classification model.

*4.3.4 Time-Distributed ImageNet-TConv (**TDImageNet-TConv**).* Instead of first condensing the time dimension, we can also first extract spatial features with image classification models, and then process the temporal sequence of these features with temporal models such as temporal convolution networks (TConv) [54] or LSTM. To realize this, we need to activate the image model for every time step to feed the pressure map of each time slice and generate the feature vector. We implemented a time-distributed layer to repeat the base image model while sharing the parameters across all time steps. The base image model was MobileNetV2 (without the last layer), which started with ImageNet weights and was also trained together with the rest of the network. The time axis is down sampled from 50 to 25 through average pooling to contain the model size. Within each time step, the image model output is flattened and fed to a dense layer of 128 neurons, which generates the feature vector of that time

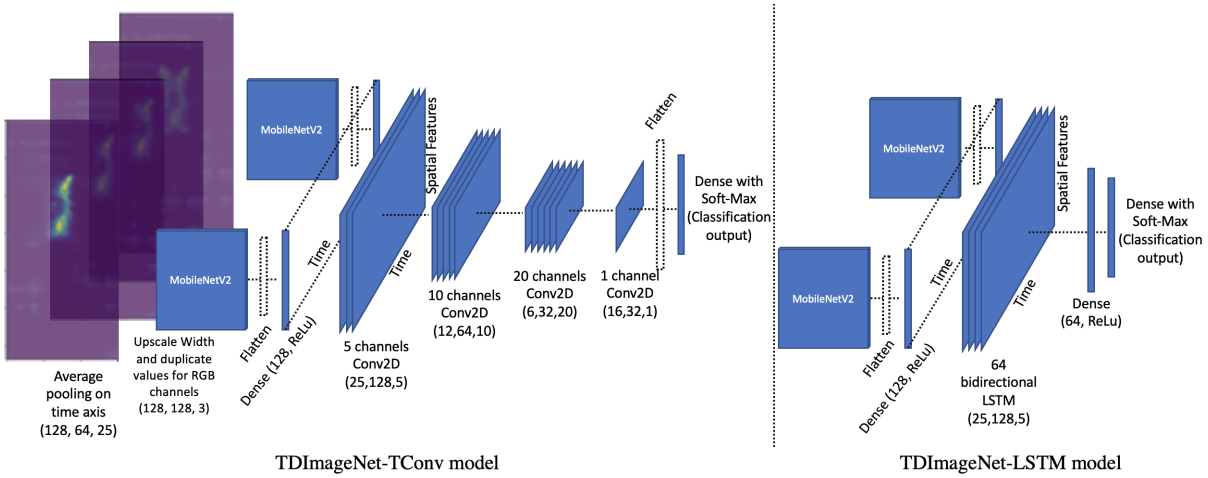


Fig. 5. Architecture illustration of the TDImageNet-TConv and TDImageNet-LSTM models.

step. Then the 128 features of 25 time steps are processed by a series of 2D convolution and pooling operations to reduce the information complexity before the last classification soft-max layer.

*4.3.5 Time-Distributed ImageNet-LSTM (**TDImageNet-LSTM**).* Instead of temporal convolution after the time-distributed feature vector outputs, we also implemented a model with a bidirectional LSTM layer to replace the temporal convolution as shown in Fig. 5.

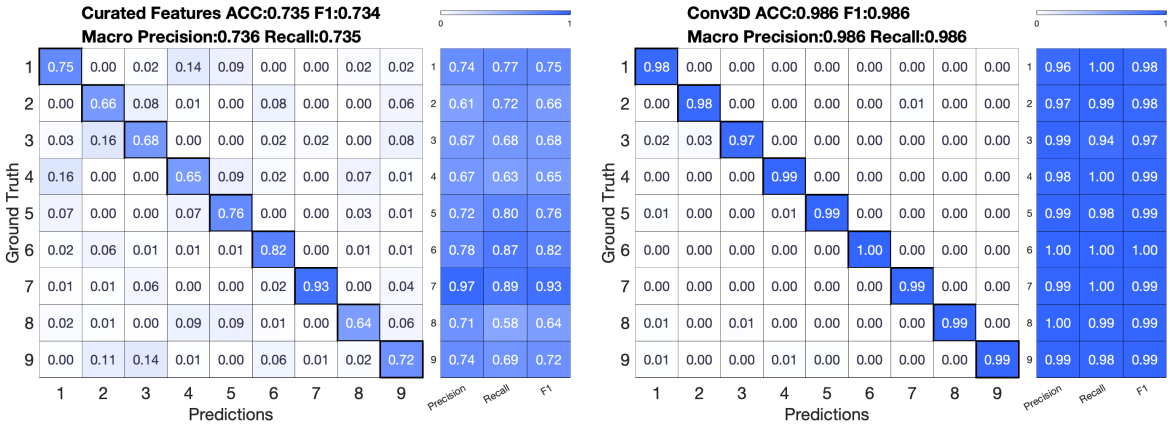
4.3.6 ConvLSTM. 2D convolutional LSTM layers [77] are compatible with 3D convolutional layers as both can be used for 3-dimentional input data. We use the ConvLSTM2D layers provided by Tensorflow through Keras. ConvLSTM2D essentially uses 2D convolution kernels to extract spacial information in a recurrent LSTM structure for the temporal information. We simply replace the 3D convolution layers in Conv3D model (Layers 1, 4, 8, 12, 15, 18 in Table 6) with bidirectional ConvLSTM2D, with the spatial dimensions aligned with the 2D convolution kernels and the temporal dimension aligned with the LSTM sequence axis. All except the last of the ConvLSTM2D layers return the sequences while the last ConvLSTM2D returns the last time step output.

5 RESULTS AND DISCUSSION

5.1 Distinguishing 9 Categories

First we present the results of the best performing model Conv3D against the curated feature set with bagged tree classifiers.

The classification results for 9 categories are presented in Fig. 6. It is clear that the combination of curated features and probabilistic classifiers struggle mostly between Category I (crunches, Fig. 10a) and IV (only back visible, Fig. 11a), Category II (push-ups, Fig. 10b) and III (planking, Fig. 10c). Category VIII (bridges, Fig. 12b) is misclassified with III (planking, Fig. 10c), IV (only back visible, Fig. 11a), IX (side plank, Fig. 12c) These misclassified pairs have the common characteristics that the overall global shapes are similar, and the further distinct visible traits are finer local details. For example, when the users are performing crunches (Category I), their back and feet are visible from the pressure map. But as the feet only takes very small area, the feature value may be overwhelmed by the larger back area, resulting in being misclassified as only back visible classes (Category IV).



(a) Curated Features with Bagged Tree classifier

(b) Conv3D

Fig. 6. Confusion matrix of recognizing 9 exercise categories.

This can be expected as the statistical spatial and temporal features are mathematically less sensitive to local variations.

On the other hand, the tiered convolution model Conv3D achieves recognition accuracy, macro F1, precision and recall of 98.6%, with some categories reaching 100% metrics. Notably the validation is performed as sliding windows without event based smoothing to boost the accuracy by considering adjacent time windows. We consider this result as highly accurate in the field of sensor based human activity recognition. The Conv3D model does not suffer from the globally similar, locally distinct problems described above for the conventional machine learning method. This might have been due to the fact that, the first levels of 3D convolution layers can learn local spatial-temporal characteristics, such as the shapes of the hand or back. This is because the kernel size (5, 5, 5) is much smaller compared to the data tensor size (128, 64, 50, 1). As the tensor size decreases in the spatial-temporal dimensions after several layers of average pooling, the kernel size (5, 5, 5) is then comparable with the tensor size, e.g. (8, 4, 5, 80). Thus the later 3D convolution layers can learn more global characteristics across the whole area of the mat.

When we leave one participant out in every validation iteration, the average accuracy and macro F1 score are reduced to 96.9%. This still outperforms every related work reported in Table 1.

5.2 Distinguishing 47 Variants and Comparing Different Deep Learning Models

Since any of the deep learning models we have implemented can easily reach over 90% accuracy for recognizing the 9 categories during our testing period, the task may have already been too simple for them. We use the 47 variants as the standard for comparing different deep learning models, because the differences are more obvious in the accuracy percentage.

For 47 variants inside the 9 categories, the curated features with the probabilistic classifier yield only 38.70% accuracy as listed in Table 5. Yet this is still an encouraging result considering:

(1) The chance level of 47 classes is 2.13%.
(2) The classification is window based, without event based smoothing among adjacent windows, as we assume for evaluating the quality of execution in exercises, the minor changes can happen during a bout of repetitions.

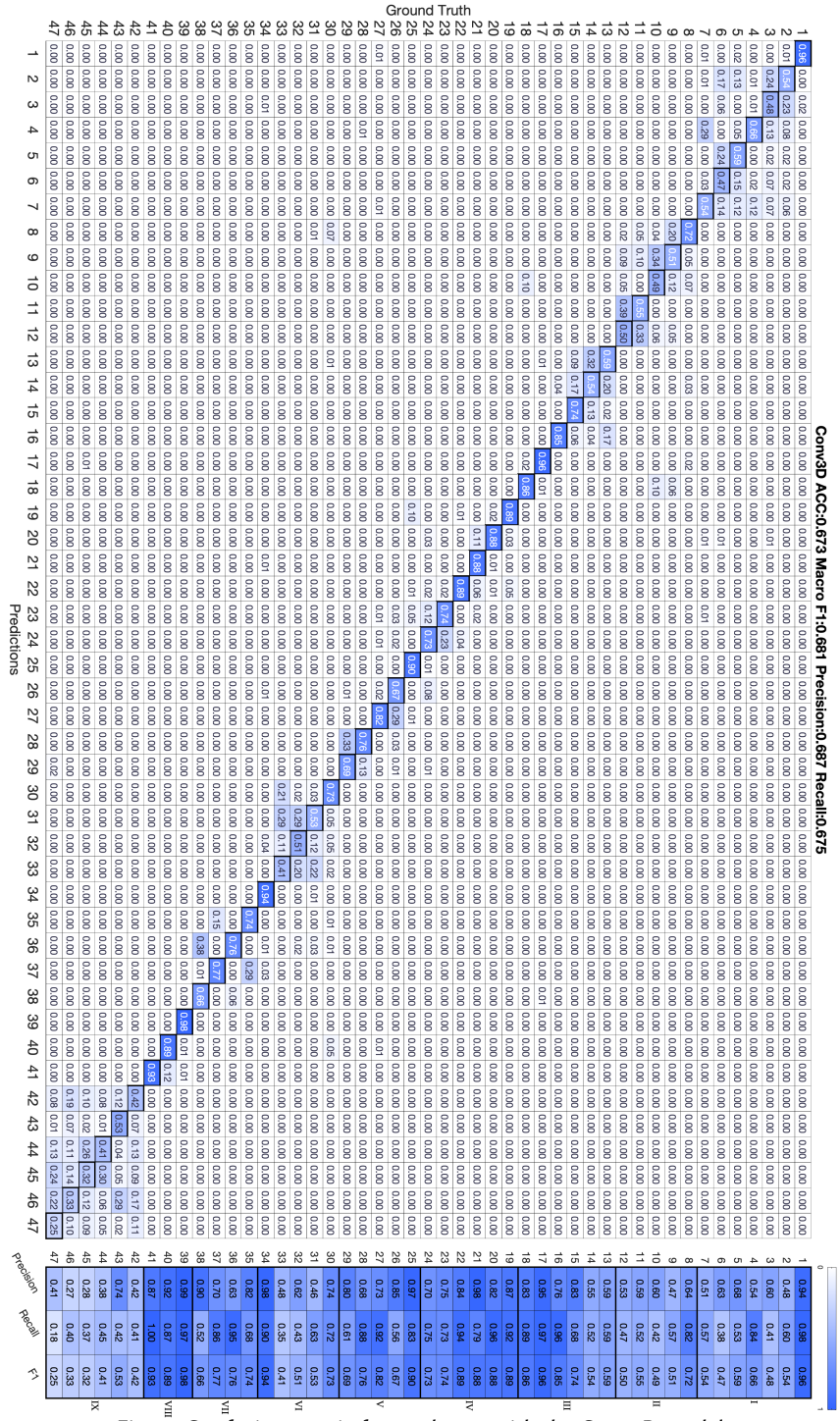


Fig. 7. Confusion matrix for 47 classes with the Conv3D model.

(3) The variants inside a category are designed to be very similar motions with only minor changes such as: range of motions; bent or straight knees. And often the difference are not in contact with the pressure sensor. For example, even though Variant 26 and 27 are both leg raises; the difference is in Variant 27 the legs are kept upward, while in Variant 26 the legs flip up and down without touching the floor.

Table 5. Model Comparison for Classifying 47 Exercise Variants

Model	Accuracy (%)	Epoch Time (s)	Parameters	Size (MB)
Conv3D	67.50	417	686,088	7.9
Conv3D-Transformer4D	52.97	637	506,476	7.8
TConv-ImageNet	48.71	340	2,690,443	28.2
TDImageNet-TConv	51.22	2670	4,857,092	56.4
TDImageNet-LSTM	48.17	2660	4,990,447	57.5
ConvLSTM	52.85	5915	2,631,609	30.3
663 features + bagged trees	38.70	-	-	-
chance level	2.13	-	-	-

Table 5 shows the comparison of accuracy and model complexity. First of all, all models have shown significantly better accuracy than the baseline probabilistic method with curated features and bagged tree classifiers. The time-distributed models TDImageNet-TConv and TDImageNet-LSTM need significantly more time for each training epoch. This is because the base image recognition models are activated 25 times for every window. Overall we can observe that although models trained with ImageNet can offer better accuracy than conventional features, they are significantly inferior to the Conv3D model. We attribute the improvement from conventional features to the fact that MobileNetV2 is better at extracting local information. However, since our data has both the spatial and temporal domains, the separation of using image classification models for extracting spatial information and other layers for temporal information seem to be the drawback. It is also possible that our dataset does not have enough activities and observations compared with ImageNet to train the models of millions of parameters, which is a common problem in sensor-based human activity recognition.

On the other hand, 3D convolution layers overcome the above mentioned possible roadblocks for image classification models. The 3D convolution kernels combine the spatial dimensions and temporal dimensions in a single operation. This will also benefit in situations like when a body part slides on the mat surface. Although transformer encoder is a promising and trending solution with a smaller model; our data format cannot fully utilize its potential without further separation of the images into patches and deeper exploration of spatio-temporal transformer architectures. The raw input size of (128, 64, 50), after flattening 409600, cannot fit inside the matrix multiplication with existing GPU hardware. If we put that into the context of natural language processing, this input size would mean the transformer encoder will process 409600 words in one shot, which is unrealistic. In order to fit the tensor inside, we need several layers of 3D convolution and pooling, which already are capable of extracting both local and global, spatial and temporal information.

ConvLSTM model also follows the idea of building a model dedicated to analyse the spatial-temporal format of this dataset, with the recurrent LSTM for the time sequence and the 2D convolution kernels inside the recurrent layer for the spatial information. Although ConvLSTM provides better accuracy (52.85%) than models involving pre-trained image classification components, it takes the most time to complete a single training epoch. It appears that the recurrent operation is not fully utilizing the parallel computing power of the GPU efficiently, which is common for recurrent models. [47]

The best performing deep learning model Conv3D yields 67.5% accuracy and 68.1% macro F1 score, with one of the smallest training epoch time and model size. Thus through our exploratory approach, we have located

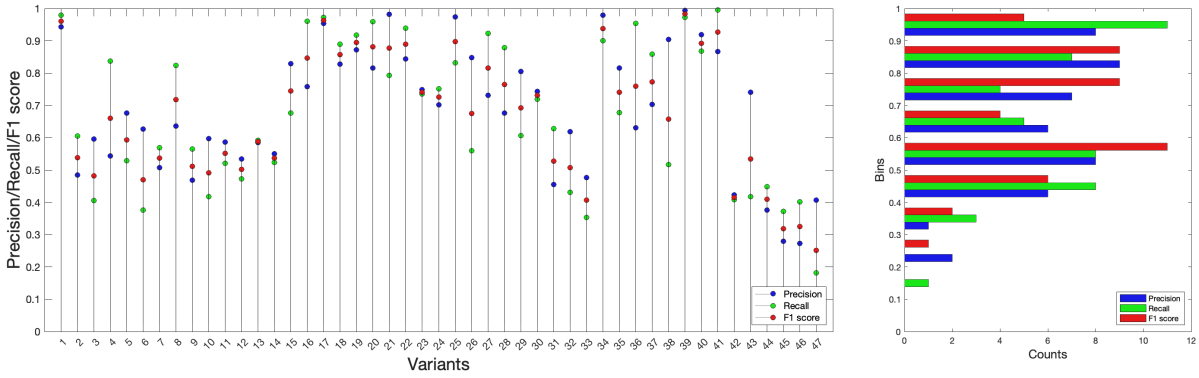


Fig. 8. The distributions of precision, recall and F1 scores of 47 variants (Conv3D model).

a simple, efficient yet effective model for this specific type of data and use case, without any transfer learning. Even though the Conv3D-Transformer4D has less parameters, is has longer epoch time. This is mostly due to the matrix multiplications inside the encoder layer. Thus, Conv3D-Transformer4D has more resource footprint than Conv3D. The confusion matrix of Conv3D with 47 classes is presented in Fig. 7. First of all, there is no noticeable cross category misclassifications, which is significant because the model for 47 classes does not have a pre-trained component to classify the 9 categories separately, as the model is summarized in Table 6. This implies that (1) a parallel model with 9 categories is not going to improve the result of the 47-class model and a single model for 47 classes is effective enough; (2) a single Conv3D model with 47 variants prediction output can also be used for classifying the exercise categories.

We also plot the detailed distributions of the precision, recall and f1 scores of every single variant in Fig. 8. Combining the confusion matrix previously shown in Fig. 7, some variant classes yield more than 80% in the metrics of concern, mainly from Categories IV (Only Back Visible), V (Back and Arms Visible), VII (Arch Step) and VIII (Bridge).

For some variant classes, the precision and recall may differ quite significantly, for example Variant 4, 26, 36, 38, etc. which is usually caused by that one class is often misclassified as other classes, but not the other way around. Since our dataset is balanced, this unequal precision and recall eventually cancels each other and thus the macro precision, recall and F1 score are similar with the overall accuracy.

The unilateral misclassification usually happens between two classes, such as Variants 36 Step Backwards and 38 Lunge Backwards., causing one class has high precision and low recall, while the other has the reversed metrics. The usefulness of such cases depends on the specific activities. In practice, for these two variants, Step Backwards is the lower quality of execution, and the focused error to correct from the fitness trainers' perspective. Thus a high recall (95%) would benefit the use case as most of the poor quality of execution exercises can be identified. As for the resulting lower recall of the correct variant Lunge Backwards, the trainers think there is no harm for reminding the trainee of the correct execution even if it is redundant. The same can be said for the pairs Variant 28 Leg Swing I and Variant 29 Leg Swing II, with Variant 28 as the lower quality of execution to be improved and Variant 29 to be promoted.

If the relationship is reversed, where the poorer quality variant has higher precision but lower recall against the desired variant, the result would be less desirable. Such as Variant 26 Leg-raise I (bad quality) and Variant 27 Leg-raise II (good quality). Since the system would misjudge some of the poor executions as of better quality.

With the leave-participant-out validation scheme for 47 variants, the performance metrics of the Conv3D model dropped significantly by around 11% to 56.1% for the accuracy, macro F1, precision and recall. Despite the performance drop, the metrics are still well above chance level and every other methods under the more optimistic leave-recording-out scheme listed in Table 5. The performance drop could be caused by the intrinsic behaviour difference across participants, especially with the diversity in our participant pool, combined with the activity definition that is meant to introduce and quantify minor differences. As we further discuss in Section 6, the leave-participant-out result is less significant in our envisioned use case.

5.3 Knowledge Recycling

We then take a closer look at the variants within each categories. Apart from examining the confusion matrix in Fig. 7, we also trained the Conv3D model with variants from every single categories separately, by changing the last dense layer size to the number of variants in a category. To save training time, and at the meanwhile reducing the carbon footprint of our research, we devised a knowledge recycling process.

(1) We divide the Conv3D model into feature encoding layers (layers 1-16 in Table 6 and the classifier layers (layer 17-22 in Table 6).
(2) The feature encoding layers are reused by keeping the parameters from the previous task as initialization with different classification task. For a different classification task, the classifier layers are modified according to the number of classes and are initialized randomly without prior weights.
(3) The order of classification tasks are: (1) 9 categories, (2-10) variants in individual categories, (11) 47 total variants.

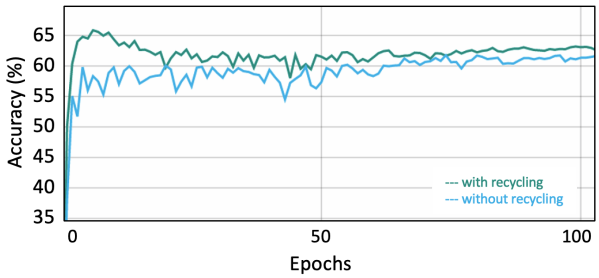
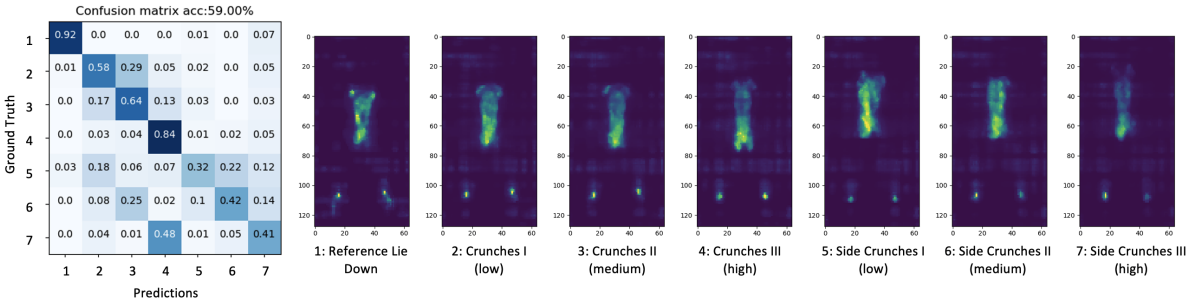
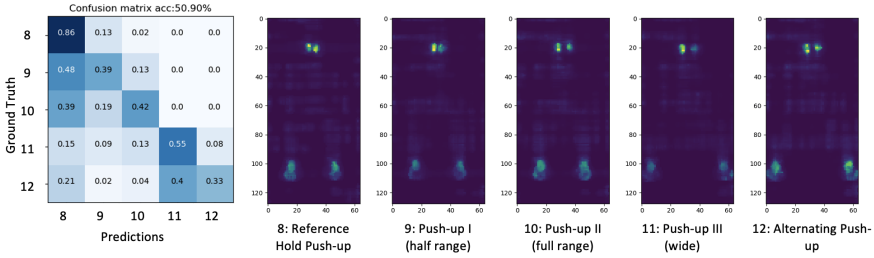


Fig. 9. Validation accuracy example during the training epochs of 47-class Conv3D with and without reusing the feature convolution layers.

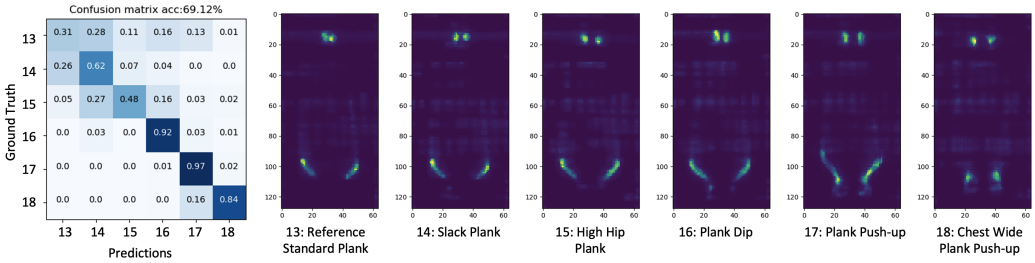
From our trials, the knowledge recycling approach and fresh initialization approach (vanilla) eventually reached similar accuracy in the final task; but with knowledge recycling, the models reached the best accuracy much earlier than those without recycling. For example, Fig. 9 shows the epoch validation accuracy values of two Conv3D models for 47 classes with and without knowledge recycling. The feature encoding layers from the model with recycling have already been trained previously with other classification tasks from 1-10 mentioned above. Benefiting from better weight initialization, this model reached the best performance within the first 10 epochs while the vanilla model needed more than 100 epochs to search for the optimal parameters. In the practical context, such a smart mat powered by AI could be provided as a service with the equipment to bridge fitness trainers and users, such as the current commercial solutions Peloton®bikes, VAHA®fitness mirror, etc. Knowledge recycling would bring the benefit of fast retraining a model when the coaches define new exercises and the corresponding data are added to the service. Reduced training time also directly lead to faster service update to the end users and less wasted resources (e.g. computation infrastructure and electricity) caused by training a blank network [76, 85].



(a) Category I: Crunches



(b) Category II: Push-ups



(c) Category III: Planking

Fig. 10. Confusion matrix of recognizing variants inside individual Categories I-III and the pressure map examples shown as per-pixel sum over the time window.

5.4 Variants within Categories

Fig. 10, Fig. 11, and Fig. 12 show the confusion matrix and pressure map examples for individual categories. The distributions of misclassifications of individually trained models are similar with those from the single model trained with total 47 variants. Thus we will mostly refer to the confusion matrix of individual categories. The following part of discussions should be considered together with the class definitions in Section 3.2, and the overall confusion matrix in Fig. 7.

5.4.1 Category I: Crunches. In Fig. 10a, the static reference pose is clearly distinguished from other classes. The participants' head points upwards in the pressure maps. There are misclassifications between low and medium range of motion (Crunches and Side Crunches I and II), but are well separated from the high range of motion variants (Crunches and Side Crunches III). The samples of Side Crunches are sometimes misclassified as the neutral crunches of corresponding range of motion, but the neutral crunches are not misclassified as side actions.

This is most obvious for the highest range of motion side crunches, which can be expected as the upper body does not touch the mat at the end position. The fitness trainers suggested that detecting the range of motions itself despite the misclassification between neutral and side actions is helpful. The trainees are typically aware of the differences between neutral and side actions; while for the range of motion they typically need external assistance such as a mirror or training partners to verify.

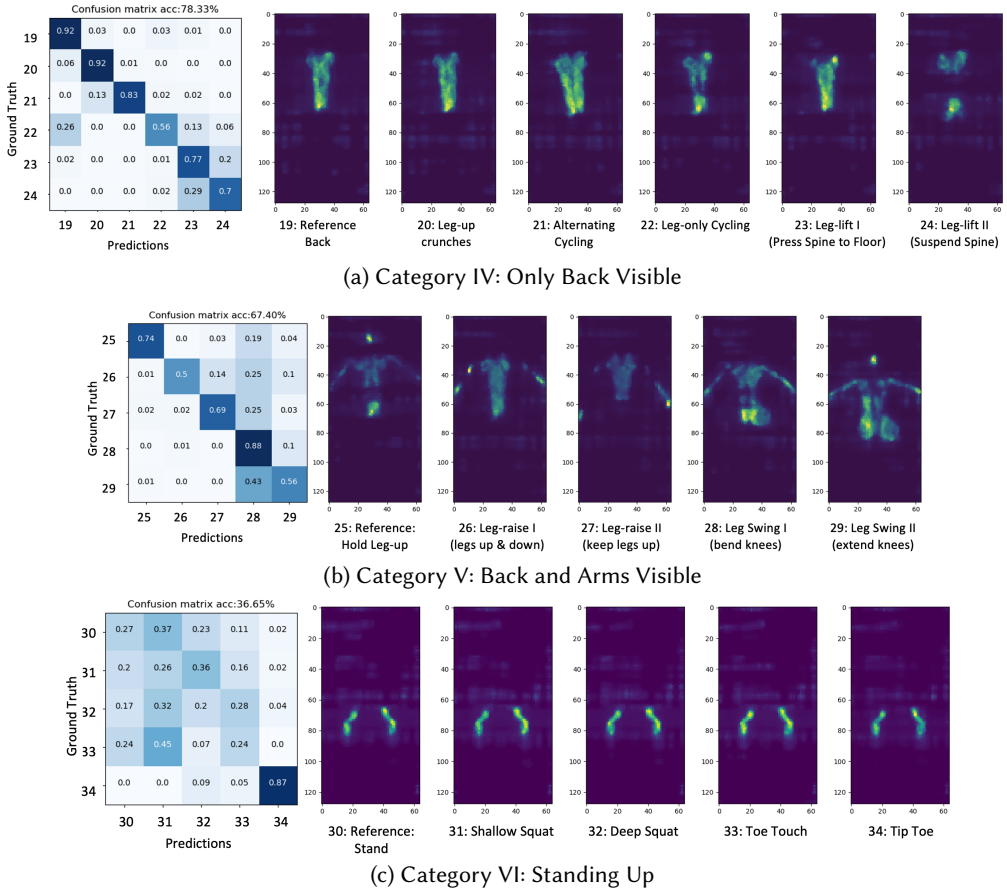
5.4.2 Category II: Push-ups. The participants' feet are in the upper part of the pressure maps in Fig. 10b. The shoulder-wide (Variant 8, 9, 10) and wider placements of hands (Variant 11, 12) are clearly separated from each other. The misclassifications may be attributed to the height differences of our participants, that some shorter participants' wider placements may be closer to the normal hands placements. Although the half (Variant 9) and full (Variant 10) range of motions are separated from each other, a major part of those samples are misclassified as the static hold push-up. Due to the difficulty of the last alternating push-up (Variant 12), some participants could not fully perform this exercise and some samples may end up similar with the wide push-up (Variant 11).

5.4.3 Category III: Planking. Similar with Category II, the participants' feet point upwards in the pressure maps in Fig. 10c. The dynamic activities Plank Dip (Variant 16), Plank Push-up (Variant 17), and Wide Plank Push-up (Variant 18) are well separated from each other. Although the static planking are often misclassified, Slack Plank (Variant 14) is still recognizable with 62% accuracy, which is often the wrong pose the fitness trainers try to correct. The slack plank is by design difficult for current vision based pose estimation models including OpenPose[15] and VideoPose3D [66] as their skeleton systems have only one single vector to represent the torso.

5.4.4 Category IV: Only Back Visible. As shown in Fig. 11a, the model can very well distinguish all the exercise variants in this category. The least recognizable exercise is leg-only cycling (Variant 22), without motion from the shoulders. It is mostly misclassified with the static reference pose, which in practice is still helpful because this variant is considered the less effective exercise without engaging much of the core muscles according to the fitness trainers. For the two leg-lift variants 23 and 24, it was also pointed out by the trainers that pressing the spine to the floor during the exercise provides better muscle engagement compared to leaving the lower back off the floor, and is often difficult to observe externally as the lower back is obstructed by the user's own body and clothing. This is again a weakness for video-based pose estimators because of their over-simplified torso representation. Our approach can distinguish these two quality of execution variants with a relatively high accuracy over 70%, which can be helpful in improving exercise QoE.

5.4.5 Category V: Back and Arms Visible. The variants in Fig. 11b were all exercises with leg motions when the users' legs were not on the floor, with no intended motions from the body parts that were touching the floor. Considering this, the model can still recognize these variants with 67.4% accuracy. For the two variants leg-raise I and II, the fitness trainers commented that in general keeping the legs up during the leg-raise exercise (Variant 27) requires more constant engagement of the core muscles. Thus, it is helpful in practice that these two variants can be well distinguished. Also leg swings with bent knees (Variant 28) is considered the easiest of all these exercises. This variant with high accuracy but also high false positive can also be useful for an fitness assistant application to encourage users to push themselves. In Fig. 7, however, when the model is trained with the entire dataset with all 47 variants, leg swing variants and leg raise variants are clearly separated from each other.

5.4.6 Category VI: Standing Up. Fig. 11c shows the variants performed with a standing position. Except for Tip Toe (Variant 34), the sensing system and model have reached their limit in recognizing any minor differences as all of the relevant motions are not directly connected to the floor. Tip toe is clearly recognizable because the heel leaves the floor during a repetition. This then may require the combination with further kinematic sensing approaches such as IMUs or videos. However, in Fig. 7, Variant 30 can still be recognized with 74% precision and 72% recall when the model is being trained with all variants together. This may be contributed by the fact that



(a) Category IV: Only Back Visible

(b) Category V: Back and Arms Visible

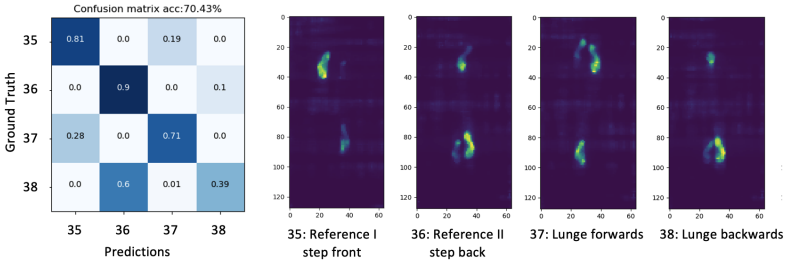
(c) Category VI: Standing Up

Fig. 11. Confusion matrix of recognizing variants inside individual Categories IV-VI and the pressure map examples shown as per-pixel sum over the time window.

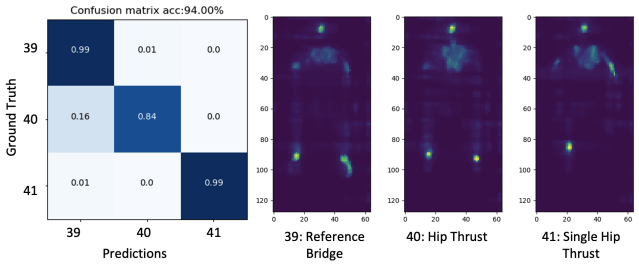
the data samples within Category VI is not enough to have the model learn temporal differences between static and dynamic activities, while the entire dataset from other categories can help the model learn such information.

5.4.7 Category VII: Arch Step. In Fig. 12a, we can observe that the motions of stepping forwards (Variants 35, 37) and backwards (Variants 36, 38) are clearly separated from each other with zero misclassifications. The physical difference between stepping (Variants 35, 36) and lunges (Variants 37, 38) is that in lunges, the knee of the leg behind should bend towards the floor, forcing the leg muscles to engage more to keep the balance. This then explains the misclassifications between these two types of motions as the major difference is not in direct contact with the floor. However, in Fig. 7, these misclassifications are improved and more equally balanced.

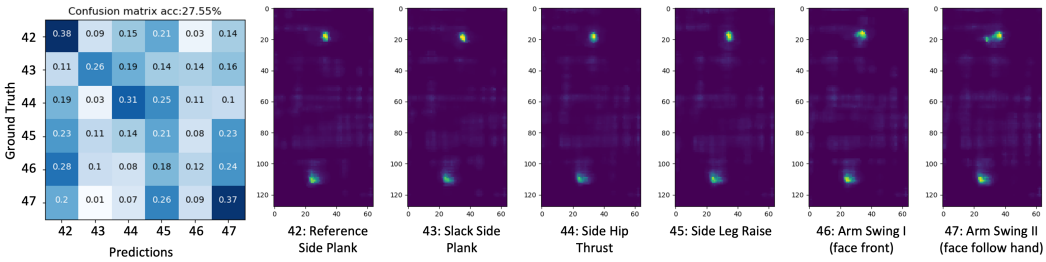
5.4.8 Category VIII: Bridge. Fig. 12b shows the results of the bridge exercises. Single hip thrust (Variant 41) is clearly separated from the other activities, as only one foot is visible at a time. The dynamic hip thrust samples are sometimes misclassified with the static bridge variants. The fact that these motions were mostly off the ground, and not sufficiently propagated to the pressure mat might have cause this misclassification. Overall, the model



(a) Category VII: Arch Step



(b) Category VIII: Bridge



(c) Category IX: Side Plank

Fig. 12. Confusion matrix of recognizing variants inside individual Categories VII-IX and the pressure map examples shown as per-pixel sum over the time window.

can distinguish the static bridge, hip thrust and single leg hip thrust with high precision and recall for evaluating the quality of execution.

5.4.9 Category IX: Side Plank. This category is the most difficult to recognize as shown in Fig. 12c. It was expected during the design of the experiment as all of the relevant motions are off the ground. The body parts that are touching the ground are rigid elbows and feet on very small areas, and the variations are trivial. For example, the only difference between Variant 46 and Variant 47 is whether the face of the user is following the moving hand. Although side plank as a whole can be distinguished from other categories with 98% accuracy as shown in Fig. 6, the system can no longer distinguish further quality of execution details. The fitness trainers also noted that the relatively high precision of Variant 43 Slack Side Plank as shown in Fig. 7 and Fig. 8 (74%) is quite helpful as this is the most common mistake they will look for from the trainees during side plank exercises.

6 FURTHER DISCUSSIONS AND LIMITATIONS

6.1 Further Results Discussion

We have investigated the variants in every exercise category to discuss the effectiveness of QoE evaluation such as posture correctness and range of motion in Section 5. Some of the mis-classifications in the models trained with only the data samples from individual categories were significantly reduced when the model was trained with the entire dataset, such as Variant 28 and 30. This might be due to the fact that, although the data from another category are visually different, they may still contribute low level information that are not sufficient or obvious in the smaller scope of the category. Although ensemble learners or hierarchical deep learning models [92] have been shown to be effective for structured class distributions such as our dataset, our results imply that this is not needed in this application. Thus, with a single monolithic Conv3D model, it is not necessary to design a nested structure of smaller learner models. In practice, this will be useful as the model can evaluate the quality of execution and categorize the exercise type at the same time, while the users are performing a variety of exercises, which is common in body weight exercises performed on sports mats.

Although transfer learning from the image domain has previously proven helpful for limited pressure sensor data [78], in this work we have shown that a smaller network built specifically for the spatial-temporal pressure maps data significantly outperforms transfer learning implementations. The fundamental differences of pressure matrix data from digital images and videos as we listed in Table 4, together with the larger dataset than previous works with pressure matrix [78, 83], may have contributed to the better performance of the Conv3D model.

6.2 Further Use Case Discussion

The fitness trainers involved in the study design process gave further comments apart from the specifically targeted suggestions on individual categories in Section 5.4. We envisioned a scenario where their trainees can use such a smart mat at their home, in combination with directed personal training sessions at the gym; and the coaches can access their home workout data to plan for new training sessions. Overall, the trainers thought the accuracy of 67.3% of detecting 47 variants is sufficient for them to have a statistical idea of how the trainees are keeping to the personal training programs on the quality of execution level. Furthermore, the variants within some categories can be recognized with much better precision and recall (above 80% or 90%), those categories mostly cover exercises while the person is lying on the back, stepping or lunging. The trainers thought this can be very useful in detailed performance and quality analysis. Although the variants of crunches in Category I yield less than 60% classification metrics, the trainers agreed that most of the misclassifications are between side and neutral crunches of within the same range of motion. It might have been overly aggressive in designing the variant details; while in practice, identifying the range of motion in this category alone is already quite useful. The trainers also noted that being able to detect the slack torso poses from correctly engaged torso in various exercises is very significant, as (1) these are the common mistakes the trainers seek to correct, (2) the slack can be difficult to visually spot due to angle, self obstruction and clothing draping from the body.

When asked about their opinions regarding the prediction performance degradation from leave-recording-out to leave-person-out, the fitness trainers expressed the leave-recording-out result could be more suitable for their envisioned use cases, where the trainees would attend personal training sessions to provide the training data needed for leave-recording-out to evaluate their home workout outside the coaching sessions. The similar setting is also suitable for their fitness classes where the members are known to the gym and trainers. The 96.9% categorization accuracy and F1 score are sufficient for high level exercise logging. While exercise category logging such as the fitness tracking functions on current smart watches can utilize leave-person-out results; we consider quality of execution shall still involve professional coaches on an individual basis. Even in the traditional fitness coaching models, the trainers usually tailor personalized training programs for individual trainees instead of a single program that fits all.

However, the trainers emphasised in the end that their opinions may be restricted by various factors such as region and the scope of their trainees, and thus suggested us to primarily refer to the sports science literature as we did in Section 3.2 and Section 5. Our envisioned use case is also subject to the actual consumer market, profitability, interactions and user experience design, and many other factors. Therefore, we would leave gauging the usefulness of our approach in evaluating the quality of execution to the broader sports science and ubiquitous computing communities with our objective findings.

6.3 Towards Effective Machine Learning Models

The knowledge recycling can be helpful in practical use cases when new exercises are introduced into a growing dataset. The models can be retrained without many epochs and thus reducing the total machine learning carbon footprint. During our model exploration, using max pooling instead of average pooling has lead to worse performance. This may be also due to the characteristic of the pressure matrix data, that every pixel is a valuable sensor. While in digital images, individual pixels are often manipulated without causing information loss on the total image. On the other hand, max pooling is more aggressive, as it picks only one input from the pooling kernel; it may not be as suitable for pressure matrix as average pooling, which uses all the input values from the pooling kernel.

In our model, the last classifier layers have approximately 100,000 parameters thanks to Layer 15 which significantly reduces the tensor size. If we remove Layer 15 and directly add the classifier layers after Layer 14, the classifier layer would have over 1.6 million parameters, most of which is from the second-last dense layer. And the model would have totally 2.17 million parameters. As mentioned above, the larger model will be more difficult to train considering the limited dataset. Layer 15 does not necessarily reduce information, as each output is a linear combination of a learned kernel from the previous layer with the output shape of (8, 4, 5, 80). We have tried further reducing the tensor dimension before attaching the dense layers by adding another conv3d layer after Layer 16 with 1 channel and output shape of (8, 4, 5, 1), then reshaping this tensor to (8, 4, 5) and adding another conv2d layer with 2 channels, with the output shape (8, 4, 2). This way, the classifier layers have less than 4000 parameters and the total model has 100,000 less parameters compared to the model in Table 6. However, this even smaller model yields approximately 5% less accuracy than the Conv3D model (which still outperforms any other alternatives). This might be a point for compromise between efficiency and accuracy, which is relevant for edge or mobile machine learning in the future.

6.4 Advantages and Limitations

Overall, we can observe that the accuracy for a certain variant is less when the relevant motions of the exercise variant is away from the pressure mat. This is also observed by the study in [17] using pressure maps to estimate poses, that the pose estimation uncertainty increases when the body part is off the pressure sensing surface.

Although pressure sensors, motion sensors and computer vision all have their corresponding advantages in the sports recognition domain; the physical information derived from the ground surface is fundamentally different from either the motion on a person's body or the visuals from external angles. Humans' interaction with the supporting surfaces is one of the fundamental physical aspects. This aspect is especially important in our targeted use case where most of the exercises are carried out with many parts of the body in contact with the ground. Apart from the practical concerns that we have discussed in Section 1.1, some activities and motions of interest that wearable motion or vision methods might be prone to errors, are trivial tasks for the floor pressure mat. For example, whether the lower back is suspended or pressed on the floor during the leg raise exercise (Variants 26 and 27) is difficult to observe even with human eyes, and the current vision-based pose estimation skeleton models do not have joint points in the torso. As another one of its trivial capabilities, the pressure mat also provides the absolute measurement of distance with the granularity of the sensing point's pitch to evaluate the

hands or gait widths for example. On the other hand, approximating distance is still a difficult task for either motion sensors and monocular cameras, which might involve complex inverse kinematics and 3D projections.

A combination with computer vision methods might offer complimentary coverage, as vision-based pose estimation is more suitable for upright positions, which are difficult to distinguish from the floor pressure. While floor based pressure mat shows reliable result in most flat oriented positions and when many parts of the body are in contact with the floor, as these postures often cause self occlusion by the user[42]. It could also be interesting to investigate if the lying pose estimation methods with pressure mats introduced in [17] can be expanded with poses related to exercises and fused with synchronized video streams to improve the pose tracking results.

Our classification results were based on individual windows, treating them as independent from each other. While time order based or event based smoothing between adjacent windows usually improves the recognition results such as the work in [83], we consider this not applicable in practical quality of execution evaluations, as quality can change during a series of repetitive movements. It may still be applicable if short period window smoothing is applied across a few seconds without considering an event of defined starting and finishing time, which may help further improve the accuracy. But this will be difficult for other future studies to compare with our study. Thus we adhere to the standard leave-out data split scheme.

The knowledge recycling process showed we can reduce training epochs by better weight initialization when new exercises are being added to the overall learning task. However, the scale of the dataset from our experiment is not enough to investigate problems such as catastrophic forgetting which could happen when a deep learning model is reused and new learning tasks are added [46].

While we designed a study that mimics a normal, complete body weight exercise routine so that the participants could finish each recording within one hour, there are many more categories of exercises, or variants of the exercises defined in this work beyond our dataset. For example, Dhahbi et al. [23] analysed 46 variants for push-ups alone in a sports mechanics study. In [36], several variations of hip thrust exercises were investigated, involving positions and angles of the feet. There are also other exercise categories such as quadruped positions.

The dataset contains mostly medium tempo exercises. We would like to further investigate the quality of execution in faster exercises such as high-intensity interval training (HIIT) in our future work. One challenge of faster paced exercises is the difficulty to establish a quality oriented data collection protocol. For example, at the planning phase of our dataset, we have considered variants of different speeds such as fast or slow push-ups, which were investigated in [6]. However, through trials we have found it was difficult for the participants to control the speed while performing the exercise variant correctly according to the desired quality metrics, especially for beginners.

Quality of execution in physical sports is a vast topic, and our Quali-Mat is only a beginning towards evaluating the quality of execution based on sensor and machine learning.

7 CONCLUSION

In conclusion, Quali-Mat explored evaluating the quality of execution (QoE) in body weight exercises with a pressure sensing sports mat and efficient machine learning models. A Quali-Mat dataset containing 12 participants, with each 3 recording sessions were gathered, performing 47 variants of 9 categories from common full-body exercises, which were defined according to quantifiable quality metrics. Quali-Mat showed that an appropriately designed neural network model specifically for the spatial-temporal pressure mat data can achieve 98.6% accuracy and macro F1 score for recognizing the 9 exercise categories outperforming the state of the art mobile, ubiquitous and wearable sports activity categorization studies. Beyond the scope of the categorization among the majority of the state of the art, Quali-Mat achieved 67.3% accuracy and 68.1% macro F1 score for distinguishing the 47 QoE variants, which were designed to have only trivial differences within a certain type of exercise. Overall,

Quali-Mat allows quantifying workout quality, even for some variants that the relevant motions are not on the floor.

ACKNOWLEDGMENTS

This work has been supported by BMBF (German Federal Ministry of Education and Research) in the project SocialWear (01IW20002).

REFERENCES

[1] 2021. Home Fitness Equipment Global Market Report 2021: COVID-19 Implications and Growth to 2030. *The Business Research Company* (2021).
[2] VAY AG. 2022. VAY Fitness Coach. https://www.vay.ai. [Online; accessed 26-Jan-2022].
[3] AH Akpa, Masashi Fujiwara, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto. 2019. A smart glove to track fitness exercises by reading hand palm. *Journal of Sensors* 2019 (2019).
[4] Justin Amadeus Albert, Lin Zhou, Pawel Glöckner, Justin Trautmann, Lisa Ihde, Justus Eilers, Mohammed Kamal, and Bert Arnrich. 2020. Will You Be My Quarantine: A Computer Vision and Inertial Sensor Based Home Exercise System. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 380–383.
[5] Omar Baritello, Josefine Stoll, Eduardo Martinez-Valdes, Steffen Müller, F Mayer, and J Müller. 2019. Neuromuscular activity of trunk muscles during side plank exercise and an additional motoric-task perturbation. *German Journal of Sports Medicine/Deutsche Zeitschrift fur Sportmedizin* 70, 6 (2019).
[6] Sebastian Baumbach and Andreas Dengel. 2017. Measuring the performance of push-ups-qualitative sport activity recognition. In *International Conference on Agents and Artificial Intelligence*, Vol. 2. SCITEPRESS, 374–381.
[7] Ganapati Bhat, Ranadeep Deb, Vatika Vardhan Chaurasia, Holly Shill, and Umit Y Ogras. 2018. Online human activity recognition using low-power wearable devices. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE, 1–8.
[8] Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, Simone Orcioni, and Claudio Turchetti. 2017. Human activity recognition using accelerometer and photoplethysmographic signals. In *International conference on intelligent decision technologies*. Springer, 53–62.
[9] Sizhen Bian, Vitor F Rey, Peter Hevesi, and Paul Lukowicz. 2019. Passive capacitive based approach for full body gym workout recognition and counting. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*. IEEE, 1–10.
[10] Jason Brownlee. 2018. *Better deep learning: train faster, reduce overfitting, and make better predictions*. Machine Learning Mastery.
[11] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 1–33.
[12] Jeannette M Byrne, Nicole S Bishop, Andrew M Caines, Kalynn A Crane, Ashley M Feaver, and Gregory EP Pearcey. 2014. Effect of using a suspension training system on muscle activation during the performance of a front plank exercise. *The Journal of Strength & Conditioning Research* 28, 11 (2014), 3049–3055.
[13] John Cairney, Kerry R McGannon, and Michael Atkinson. 2018. Exercise is medicine: Critical considerations in the qualitative research landscape.
[14] Joaquin Calatayud, Jose Casaña, Fernando Martín, Markus D Jakobsen, Juan C Colado, Pedro Gargallo, Álvaro Juesas, Víctor Muñoz, and Lars L Andersen. 2017. Trunk muscle activity during different variations of the supine plank exercise. *Musculoskeletal Science and Practice* 28 (2017), 54–58.
[15] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence* 43, 1 (2019), 172–186.
[16] Jung-Hoon Choi, Da-Eun Kim, and Heon-Seock Cynn. 2021. Comparison of trunk muscle activity between traditional plank exercise and plank exercise with isometric contraction of ankle muscles in subjects with chronic low back pain. *The Journal of Strength & Conditioning Research* 35, 9 (2021), 2407–2413.
[17] Henry M Clever, Ariel Kapusta, Daehyung Park, Zackory Erickson, Yash Chitalia, and Charles C Kemp. 2018. 3d human pose estimation on a configurable bed from a pressure image. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 54–61.
[18] Robert M Cogley, Teasha A Archambault, Jon F Fibeger, Mandy M Koverman, et al. 2005. Comparison of muscle activation using various hand positions during the push-up exercise. *Journal of strength and conditioning research* 19, 3 (2005), 628.
[19] Paul Comfort and Peter Kasim. 2007. Optimizing squat technique. *Strength and Conditioning Journal* 29, 6 (2007), 10.
[20] Paul Comfort, Stephen J Pearson, and David Mather. 2011. An electromyographical comparison of trunk muscle activity during isometric trunk and dynamic strengthening exercises. *The Journal of Strength & Conditioning Research* 25, 1 (2011), 149–154.
[21] Emily E Cust, Alice J Sweeting, Kevin Ball, and Sam Robertson. 2019. Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. *Journal of sports sciences* 37, 5 (2019), 568–600.

[22] Déborah de Araújo Farias, Jeffrey M Willardson, Gabriel A Paz, Ewertton de S Bezerra, and Humberto Miranda. 2017. Maximal strength performance and muscle activation for the bench press and triceps extension exercises adopting dumbbell, barbell, and machine modalities over multiple sets. *The Journal of Strength & Conditioning Research* 31, 7 (2017), 1879–1887.

[23] Wissem Dhahbi, Helmi Chaabene, Anis Chaouachi, Johnny Padulo, David G Behm, Jodie Cochrane, Angus Burnett, and Karim Chamari. 2018. Kinetic analysis of push-up exercises: a systematic review with practical recommendations. *Sports biomechanics* (2018).

[24] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[25] Priscilla Dwelly, Gretchen Oliver, Heather Adams-Blair, David Keeley, and Hiedi Hoffman. 2009. Improved muscle activation in performing a body weight lunge compared to the traditional back squat. In *ISBS-Conference Proceedings Archive*.

[26] Richard A Ekstrom, Robert A Donatelli, and Kenji C Carp. 2007. Electromyographic analysis of core trunk, hip, and thigh muscles during 9 rehabilitation exercises. *Journal of orthopaedic & sports physical therapy* 37, 12 (2007), 754–762.

[27] Rafael F Escamilla, Clare Lewis, Duncan Bell, Gwen Bramblet, Jason Daffron, Steve Lambert, Amanda Pecson, Rodney Imamura, Lonnie Paulos, and James R Andrews. 2010. Core muscle activation during Swiss ball and traditional abdominal exercises. *journal of orthopaedic & sports physical therapy* 40, 5 (2010), 265–276.

[28] Christoph Feichtenhofer. 2020. X3d: Expanding architectures for efficient video recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 203–213.

[29] Christoph Feichtenhofer, Axel Pinz, and Richard P Wildes. 2017. Spatiotemporal multiplier networks for video action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4768–4777.

[30] Biying Fu, Florian Kirchbuchner, and Arjan Kuijper. 2020. Unconstrained workout activity recognition on unmodified commercial off-the-shelf smartphones. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 1–10.

[31] RL Gajdosik, CK Hatcher, and S Whitsell. 1992. Influence of short hamstring muscles on the pelvis and lumbar spine in standing and during the toe-touch test. *Clinical Biomechanics* 7, 1 (1992), 38–42.

[32] Preetham Ganesh, Reza Etemadi Idgahi, Chinmaya Basavanahally Venkatesh, Ashwin Ramesh Babu, and Maria Kyrarini. 2020. Personalized system for human gym activity recognition using an RGB camera. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 1–7.

[33] Carol Ewing Garber, Bryan Blissmer, Michael R Deschenes, Barry A Franklin, Michael J Lamonte, I-Min Lee, David C Nieman, and David P Swain. 2011. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. (2011).

[34] Carol Ewing Garber, Bryan Blissmer, Michael R Deschenes, Barry A Franklin, Michael J Lamonte, I-Min Lee, David C Nieman, David P Swain, et al. 2011. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Medicine and science in sports and exercise* 43, 7 (2011), 1334–1359.

[35] Andrew Garbett, Ziedune Degutyte, James Hodge, and Arlene Astell. 2021. Towards Understanding People's Experiences of AI Computer Vision Fitness Instructor Apps. In *Designing Interactive Systems Conference 2021*. 1619–1637.

[36] César L Collazo García, Javier Rueda, Bruno Suárez Luginick, and Enrique Navarro. 2020. Differences in the electromyographic activity of lower-body muscles in hip thrust variations. *The Journal of Strength & Conditioning Research* 34, 9 (2020), 2449–2455.

[37] NF Ghazali, N Shahar, NA Rahmad, NA J Sufri, MA As' ari, and HF M Latif. 2018. Common sport activity recognition using inertial sensor. In *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE, 67–71.

[38] Thomas Holleczek, Alex Rüegg, Holger Harms, and Gerhard Tröster. 2010. Textile pressure sensors for sports applications. In *SENSORS, 2010 IEEE*. IEEE, 732–737.

[39] Con Hrysomallis. 2011. Balance ability and athletic performance. *Sports medicine* 41, 3 (2011), 221–232.

[40] Yu-Liang Hsu, Shih-Chin Yang, Hsing-Cheng Chang, and Hung-Che Lai. 2018. Human daily and sport activity recognition using a wearable inertial sensor network. *IEEE Access* 6 (2018), 31715–31728.

[41] Hai Hu, Onno G Meijer, Paul W Hodges, Sjoerd M Bruijn, Rob L Strijers, Prabath WB Nanayakkara, Barend J van Royen, Wenhua Wu, Chun Xia, and Jaap H van Dieën. 2012. Understanding the active straight leg raise (ASLR): An electromyographic study in healthy subjects. *Manual therapy* 17, 6 (2012), 531–537.

[42] Ying Huang, Bin Sun, Haipeng Kan, Jiankai Zhuang, and Zengchang Qin. 2019. FollowMeUp Sports: New benchmark for 2D human keypoint recognition. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer, 110–121.

[43] Jihye Hwang, John Yang, and Nojun Kwak. 2020. Exploring Rare Pose in Human Pose Estimation. *IEEE Access* 8 (2020), 194964–194977.

[44] Nwannadi Vivian Ifeyinwa, Ikele Chioma Nneka, Ikele Ikenna, Uneke Chibuike Solomon Theophilus, Ugwu Sandra Ugonne, Ojukwu Chidiebele Petronilla, Mgbeojedo Ukamaka Gloria, Okemuo Adaora Justina, Emmanuel Grace Nneoma, and Ekemezie Wendy. 2021. Analysis of the Effects of Double Straight Leg Raise and Abdominal Crunch Exercises on Core Stability. *International Journal of Medical Science and Dental Research* (2021).

[45] Onyx Inc. 2022. Onyx Home Workout. https://www.onyx.fit. [Online; accessed 26-Jan-2022].
[46] Ronald Kemker, Marc McClure, Angelina Abitino, Tyler Hayes, and Christopher Kanan. 2018. Measuring catastrophic forgetting in neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
[47] Viacheslav Khomenko, Oleg Shyshkov, Olga Radyvonenko, and Kostiantyn Bokhan. 2016. Accelerating recurrent neural network training using sequence bucketing and multi-gpu data parallelization. In *2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP)*. IEEE, 100–103.
[48] Rushil Khurana, Karan Ahuja, Zac Yu, Jennifer Mankoff, Chris Harrison, and Mayank Goel. 2018. GymCam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–17.
[49] Dongsu Kim, Jongchan Jung, and Yijung Chung. 2021. The Effects of Performing Bridge Exercise and Hip Thrust Exercise using Various Knee Joint Angles on Trunk and Lower Body Muscle Activation in Healthy Subjects. *Physical Therapy Rehabilitation Science* 10, 2 (2021), 205–211.
[50] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[51] Heli Koskimäki, Pekka Siirtola, and Juha Röning. 2017. Myogym: introducing an open gym data set for activity recognition collected using myo armband. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 537–546.
[52] Hyeokhyen Kwon, Bingyao Wang, Gregory D Abowd, and Thomas Plötz. 2021. Approaching the Real-World: Supporting Activity Recognition Training with Virtual IMU Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–32.
[53] Christopher R Lattimer, Claude Franceschi, and Evi Kalodiki. 2018. Optimizing calf muscle pump function. *Phlebology* 33, 5 (2018), 353–360.
[54] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. 2017. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 156–165.
[55] Gregory J Lehman, Brandon MacMillan, Ian MacIntyre, Michael Chivers, and Mark Fluter. 2006. Shoulder muscle EMG activity during push up variations on and off a Swiss ball. *Dynamic Medicine* 5, 1 (2006), 1–7.
[56] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.
[57] Paul Macadam, John Cronin, and Bret Contreras. 2015. An examination of the gluteal muscle activity associated with dynamic hip abduction and hip external rotation exercise: a systematic review. *International journal of sports physical therapy* 10, 5 (2015), 573.
[58] Grazia Maugeri, Paola Castrogiovanni, Giuseppe Battaglia, Roberto Pippi, Velia D'Agata, Antonio Palma, Michelino Di Rosa, and Giuseppe Musumeci. 2020. The impact of physical activity on psychological health during Covid-19 pandemic in Italy. *Heliyon* 6, 6 (2020), e04315.
[59] Eleonora Mencarini, Amon Rapp, Lia Tirabeni, and Massimo Zancanaro. 2019. Designing wearable systems for sports: A review of trends and opportunities in human–computer interaction. *IEEE Transactions on Human-Machine Systems* 49, 4 (2019), 314–325.
[60] Vangelis Metsis, Georgios Galatas, Alexandros Papangelis, Dimitrios Kosmopoulos, and Fillia Makedon. 2011. Recognition of sleep patterns using a bed pressure mat. In *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. 1–4.
[61] Pramod Murthy, Onorina Kovalenko, Ahmed Elhayek, C Couto Gava, and Didier Stricker. 2017. 3d human pose tracking inside car using single rgb spherical camera. In *ACM Chapters Computer Science in Cars Symposium CSCS 2017-. ACM Chapters Computer Science in Cars Symposium (CSCS-17)*.
[62] Yuki Nakai, Masayuki Kawada, Takasuke Miyazaki, and Ryoji Kiyama. 2019. Trunk muscle activity during trunk stabilizing exercise with isometric hip rotation using electromyography and ultrasound. *Journal of Electromyography and Kinesiology* 49 (2019), 102357.
[63] Thomas W Nesser, Neil Fleming, and Matthew J Gage. 2015. Activation of Selected Core Muscles during Pressing. *International Journal of Kinesiology and Sports Science* 3, 4 (2015), 56–61.
[64] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.
[65] Samho Park, TianZong Huang, Junyoung Song, and Myungmo Lee. 2021. Comparative Study of the Biomechanical Factors in Range of Motion, Muscle Activity, and Vertical Ground Reaction Force between a Forward Lunge and Backward Lunge. *Physical Therapy Rehabilitation Science* 10, 2 (2021), 98–105.
[66] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli. 2019. 3d human pose estimation in video with temporal convolutions and semi-supervised training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7753–7762.
[67] Tobias Peter, Simone Bexten, Veit Müller, Viola Hauffe, and Norbert Elkmann. 2020. Object Classification on a High-Resolution Tactile Floor for Human-Robot Collaboration. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vol. 1. IEEE, 1255–1258.

[68] Rafal Pytel, Osman Semih Kayhan, and Jan C van Gemert. 2021. Tilting at windmills: Data augmentation for deep pose estimation does not help with occlusions. In *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 10568–10575.
[69] Meera Radhakrishnan, Archan Misra, and Rajesh K Balan. 2021. W8-Scope: Fine-grained, practical monitoring of weight stack-based exercises. *Pervasive and Mobile Computing* (2021), 101418.
[70] Alen Rajšp and Iztok Fister. 2020. A systematic literature review of intelligent data analysis methods for smart sport training. *Applied Sciences* 10, 9 (2020), 3013.
[71] Luis Gustavo Tomal Ribas, Marta Pereira Cocron, Joed Lopes Da Silva, Alessandro Zimmer, and Thomas Brandmeier. 2021. In-Cabin vehicle synthetic data to test Deep Learning based human pose estimation models. In *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 610–615.
[72] Jun G San Juan, David N Suprak, Sean M Roach, and Marc Lyda. 2015. The effects of exercise type and elbow angle on vertical ground reaction force and muscle activity during a push-up plus exercise. *BMC musculoskeletal disorders* 16, 1 (2015), 1–9.
[73] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.
[74] Jacob D Sartor, Amy E Latimer-Cheung, Shane N Sweet, Brooke H Thompson, and Jennifer R Tomasone. 2021. Exploring the relationship between quality and quantity of participation in an online community-based exercise program. *Journal of Exercise, Movement, and Sport (SCAPPS refereed abstracts repository)* 52, 1 (2021).
[75] Brad J Schoenfeld, Bret Contreras, Gul Tiryaki-Sonmez, Jeffrey M Willardson, and Fabio Fontana. 2014. An electromyographic comparison of a modified version of the plank with a long lever and posterior tilt versus the traditional plank exercise. *Sports biomechanics* 13, 3 (2014), 296–306.
[76] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. 2020. Green ai. *Commun. ACM* 63, 12 (2020), 54–63.
[77] Xingjian Shi, Zhourong Chen, Hao Wang, Dit Yan Yeung, Wai Kin Wong, and Wang Chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems* 2015 (2015), 802–810.
[78] Monit Shah Singh, Vinaychandran Pondenkandath, Bo Zhou, Paul Lukowicz, and Marcus Liwickit. 2017. Transforming sensor data to the image domain for deep learning—An application to footstep detection. In *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2665–2672.
[79] Lindsay V Slater and Joseph M Hart. 2017. Muscle activation patterns during different squat techniques. *Journal of strength and conditioning research* 31, 3 (2017), 667–676.
[80] Sarah M Stadig and Anna K Bergh. 2015. Gait and jump analysis in healthy cats using a pressure mat system. *Journal of Feline Medicine and Surgery* 17, 6 (2015), 523–529.
[81] Eric Sternlicht, Stuart G Rugg, Matt D Bernstein, and Scott D Armstrong. 2005. Electromyographical analysis and comparison of selected abdominal training devices with a traditional crunch. *The Journal of Strength & Conditioning Research* 19, 1 (2005), 157–162.
[82] David Strömbäck, Sangxia Huang, and Valentin Radu. 2020. MM-Fit: Multimodal Deep Learning for Automatic Exercise Logging across Sensing Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–22.
[83] Mathias Sundholm, Jingyuan Cheng, Bo Zhou, Akash Sethi, and Paul Lukowicz. 2014. Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 373–382.
[84] Juri Taborri, Justin Keogh, Anton Kos, Alessandro Santuz, Anton Umek, Caryn Urbanczyk, Eline van der Kruk, and Stefano Rossi. 2020. Sport biomechanics applications using inertial, force, and EMG sensors: a literature overview. *Applied bionics and biomechanics* 2020 (2020).
[85] Neil C Thompson, Kristjan Greenewald, Keeheon Lee, and Gabriel F Manso. 2021. Deep Learning's Diminishing Returns: The Cost of Improvement is Becoming Unsustainable. *IEEE Spectrum* 58, 10 (2021), 50–55.
[86] Robin Tibor Schirrmeister, Lukas Gemein, Katharina Eggensperger, Frank Hutter, and Tonio Ball. 2017. Deep learning with convolutional neural networks for decoding and visualization of eeg pathology. *arXiv e-prints* (2017), arXiv–1708.
[87] Roland van den Tillaar and Gertjan Ettema. 2013. A comparison of muscle activity in concentric and counter movement maximum bench press. *Journal of human kinetics* 38 (2013), 63.
[88] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.
[89] Tianyi Wang, Yanglei Gan, Scott D Arena, Lubomir T Chitkushev, Guanglan Zhang, and Reza Rawassizadeh. 2021. Advances for Indoor Fitness Tracking, Coaching, and Motivation: A Review of Existing Technological Advances. *IEEE Systems, Man, and Cybernetics Magazine* 7, 1 (2021), 4–14.
[90] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. 2016. Convolutional pose machines. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 4724–4732.
[91] Jan Wilke, Lisa Mohr, Adam S Tenforde, Pascal Edouard, Chiara Fossati, Marcela González-Gross, Celso Sanchez Ramirez, Fernando Laiño, Benedict Tan, Julian David Pillay, et al. 2020. Restrictercise! Preferences regarding digital home training programs during confinements associated with the COVID-19 pandemic. *International journal of environmental research and public health* 17, 18 (2020),

6515.
[92] Zhicheng Yan, Hao Zhang, Robinson Piramuthu, Vignesh Jagadeesh, Dennis DeCoste, Wei Di, and Yizhou Yu. 2015. HD-CNN: hierarchical deep convolutional neural networks for large scale visual recognition. In *Proceedings of the IEEE international conference on computer vision*. 2740–2748.
[93] Sen Yang, Zhibin Quan, Mu Nie, and Wankou Yang. 2021. TransPose: Keypoint localization via transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 11802–11812.
[94] James W Youdas, Mallory MP Boor, Arynn L Darfler, Margaret K Koenig, Katherine M Mills, and John H Hollman. 2014. Surface electromyographic analysis of core trunk and hip muscles during selected rehabilitation exercises in the side-bridge to neutral spine position. *Sports Health* 6, 5 (2014), 416–421.
[95] James W Youdas, Kendra C Coleman, Erin E Holstad, Stephanie D Long, Nicole L Veldkamp, and John H Hollman. 2018. Magnitudes of muscle activation of spine stabilizers in healthy adults during prone on elbow planking exercises with and without a fitness ball. *Physiotherapy theory and practice* 34, 3 (2018), 212–222.
[96] Qingtian Yu, Haopeng Wang, Fedwa Laamarti, and Abdulmotaleb El Saddik. 2021. Deep Learning-Enabled Multitask System for Exercise Recognition and Counting. *Multimodal Technologies and Interaction* 5, 9 (2021), 55.
[97] Bo Zhou, Jingyuan Cheng, Mathias Sundholm, and Paul Lukowicz. 2014. From smart clothing to smart table cloth: Design and implementation of a large scale, textile pressure matrix sensor. In *International conference on architecture of computing systems*. Springer, 159–170.
[98] Bo Zhou and Paul Lukowicz. 2019. TPM Feature Set: a Universal Algorithm for Spatial-Temporal Pressure Mapping Imagery Data.
[99] Bo Zhou, Monit Shah Singh, Sugandha Doda, Muhammet Yildirim, Jingyuan Cheng, and Paul Lukowicz. 2017. The carpet knows: Identifying people in a smart environment from a single step. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 527–532.
[100] Bo Zhou, Mathias Sundholm, Jingyuan Cheng, Heber Cruz, and Paul Lukowicz. 2016. Never skip leg day: A novel wearable approach to monitoring gym leg exercises. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–9.

A APPENDIX

A.1 Model Details

The details of the best performing model are listed in Table 6.

A.2 Leave Person Out

While leave session out emulates the system's performance within a known group of users, it does not provide insight on how the system will perform on new users that were not included in the training dataset. We thus performed leave-person-out validation with the best performing model (Conv3D), following the similar leave-out preparation procedure described in Section 4.1, but change leaving a recording out to leave a person's data out. Every participant's data was kept in the testing data in every iteration. The process was repeated once for each participant. The accumulated confusion matrix, precision, recall and F1 score of every class from 9 categories and 47 variants are shown in Fig. 13 and Fig. 15. The individual's accuracy distribution is also shown in Fig. 14.

Compared with the leave-recording-out results from Section 5, the metrics including accuracy, macro precision, recall and F1 score remain highly accurate, with slight degradations of less than 2%, and still outperform every state of the art listed in Table 1. Therefore, even when tested with strangers, the system can classify the exercise categories with 96.9% accuracy.

When it comes to distinguishing 47 variants concerning the quality of execution of exercises, the performance metrics have dropped significantly by around 11% to 56.1%. We suspect this is caused by the combination of two factors: (1) the classification task is to distinguish minor variations that is already difficult to distinguish by human observers; (2) in activity recognition, cross-person validation is known to introduce similar variations caused by differences from personal behaviour, body shape, etc.

Despite the performance degradation for 47 classes, the misclassifications are still mostly contained within every category. This can be considered significant since the model for recognizing 47 variants were not exposed to the information about the 9 categories. The miscassifications and the distribution of individual class precision, recall and F1 also reflect the trend shown in the leave-recording-out result in Fig. 7. This result sill outperforms every

Table 6. Model Summary of Conv3D for 47 classes

Layer (type)	Output Shape	Kernel	Parameters
Feature Encoding Layers			
1: conv3d	(None, 128, 64, 50, 10)	(5, 5, 5)	1260
2: average pooling3D	(None, 64, 32, 16, 10)	(2, 2, 3)	0
3: dropout	(None, 64, 32, 16, 10)		0
4: conv3d	(None, 64, 32, 16, 20)	(5, 5, 5)	25020
5: batch normalization	(None, 64, 32, 16, 20)		80
6: average pooling3d	(None, 32, 16, 5, 20)	(2, 2, 3)	0
7: dropout	(None, 32, 16, 5, 20)		0
8: conv3d	(None, 32, 16, 5, 40)	(5, 5, 5)	100040
9: batch normalization	(None, 32, 16, 5, 40)		160
10: average pooling3d	(None, 16, 8, 5, 40)	(2, 2, 1)	0
11: dropout	(None, 16, 8, 5, 40)		0
12: conv3d	(None, 16, 8, 5, 80)	(5, 5, 5)	400080
13: batch normalization	(None, 16, 8, 5, 80)		320
14: average pooling3d	(None, 8, 4, 5, 80)	(2, 2, 1)	0
15: conv3d	(None, 8, 4, 5, 5)	(5, 5, 5)	50005
16: batch normalization	(None, 8, 4, 5, 5)		20
Classifier Layers			
17: dropout	(None, 8, 4, 5, 5)		0
18: flatten	(None, 800)		0
19: dense	(None, 128)		102528
20: dropout	(None, 128)		0
21: batch normalization	(None, 128)		512
22: dense	(None, 47)		6063
Total params			686,088
Trainable params			685,542
Non-trainable params			546

other method we have explored in Table 5 which reports the metrics with the more optimistic leave-recording-out validation scheme.

A.3 Visual Differences of 47 Variations

In this section we include visual examples of all the exercise variants listed in Table 2 and Table 3. Class 2-4 are already presented in Fig. 1, therefore they are excluded here. The static poses, especially the reference classes of each category (Cat), are shown in Fig. 16. For the dynamic motions, the first pressure map is the summary of the time window (as in Fig. 1), and the following pressure maps are frames inside the selected window. Only every other frame, and only 19 of the 50 frames of the window are shown so that most of the time window can be fit in the page. For some exercise variants, while the pressure imagery does not have significant changes, the motion can be distinguished from the overall pressure changes caused by he momentum transferred from the user to the mat. In such cases, a time plot of the average pressure during the 50 frames window is presented replacing the second row of pressure maps. An illustration video can also be found on the online repository.

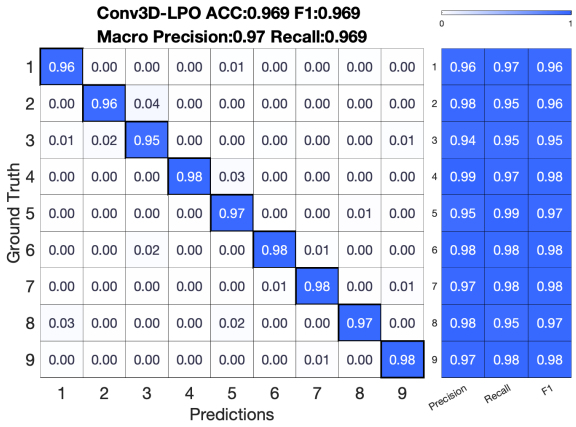


Fig. 13. Confusion matrix of Leave-Person-Out for 9 exercise categories with the Conv3D model.

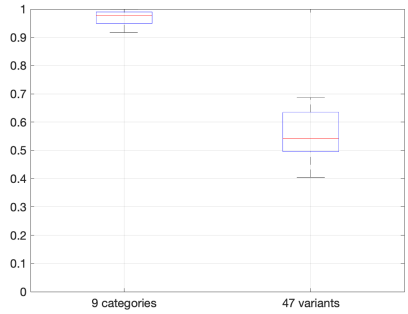


Fig. 14. Boxplot of the test individual's accuracy distribution in Leave-Person-Out

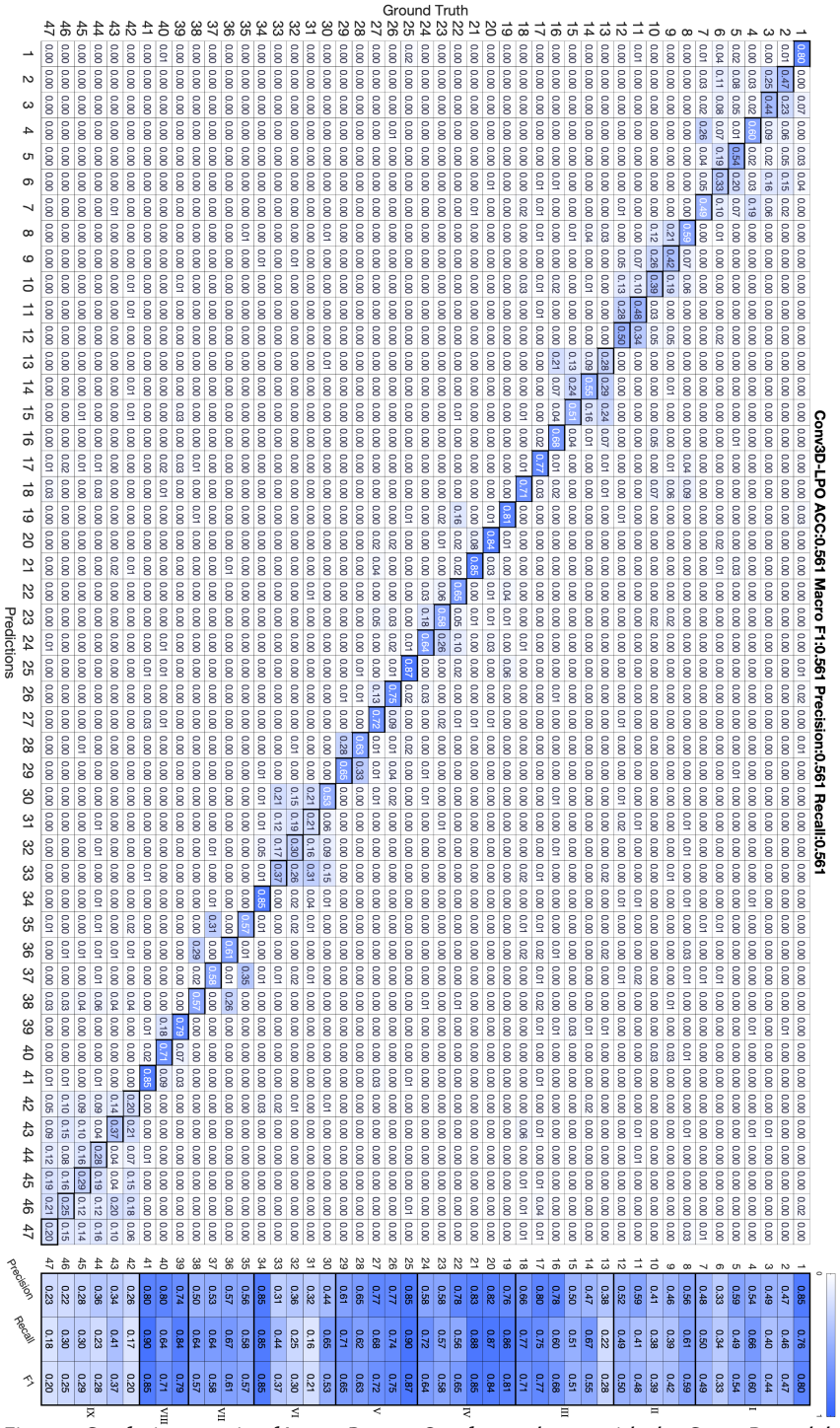
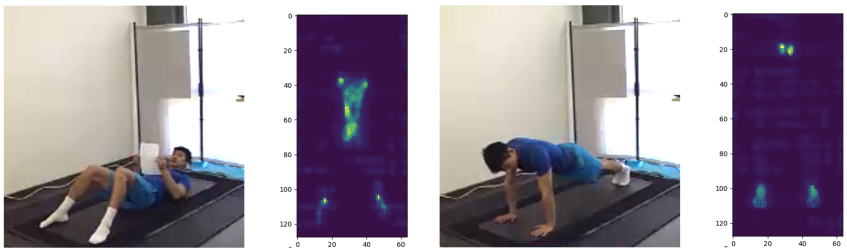
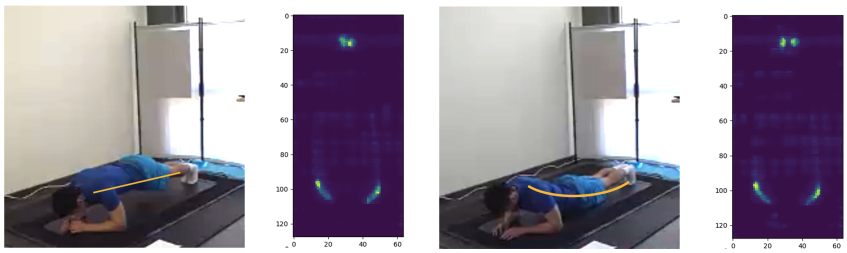


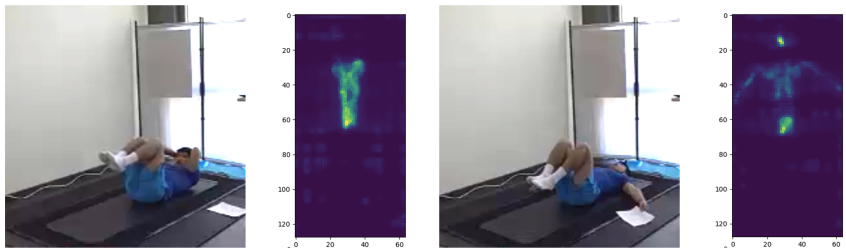
Fig. 15. Confusion matrix of Leave-Person-Out for 47 classes with the Conv3D model.



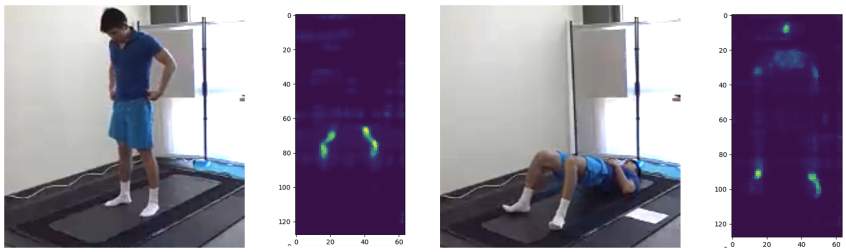
(a) Class 1: Cat I f ref. Lie Down (b) Class 8: Cat II ref. Hold Push-up



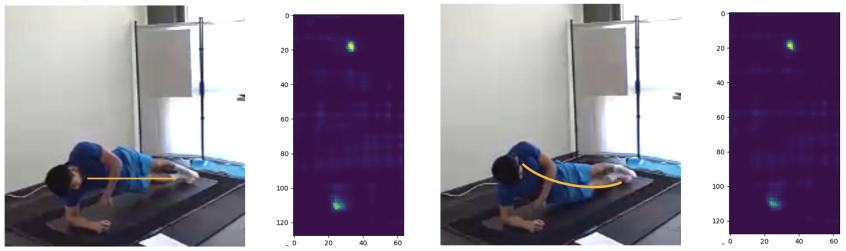
(c) Class 13: Cat III ref. Standard Plank (d) Class 14: Slack Plank



(e) Class 19: Cat IV ref. Back (f) Class 25: Cat V ref. Hold Leg-up

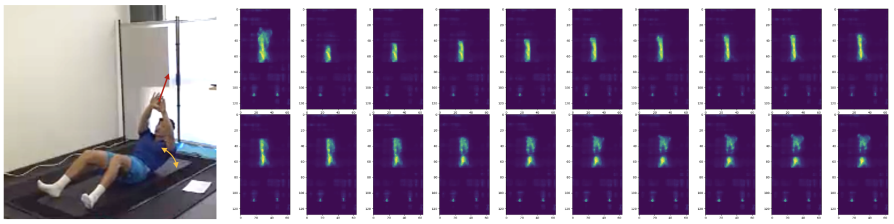


(g) Class 30: Cat VI ref. Stand (h) Class 39: Cat VIII ref. Bridge

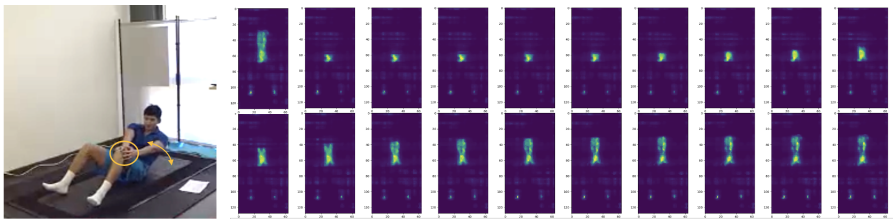


(i) Class 42: Cat IX ref. Side Plank (j) Class 43: Slack Side Plank

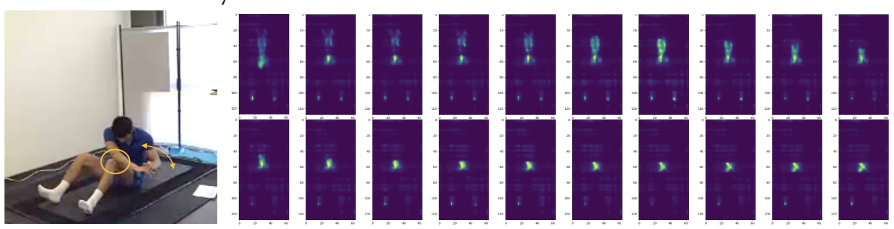
Fig. 16. Pressure map examples of the static reference and variant poses.



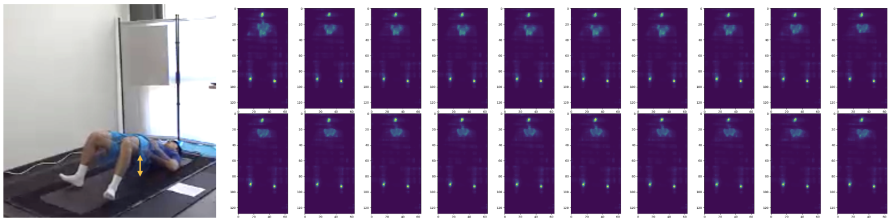
(a) Class 5: Side Crunches I. Lower range of motion, regulated by hands pointed to the sky with arms straight.



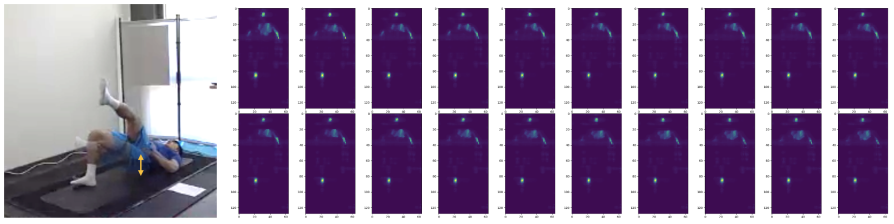
(b) Class 6: Side Crunches II. Medium range of motion, regulated by both hands pointed to each knee alternatively.



(c) Class 7: Side Crunches III. Higher range of motion, regulated by elbow touching the opposite knee alternatively.

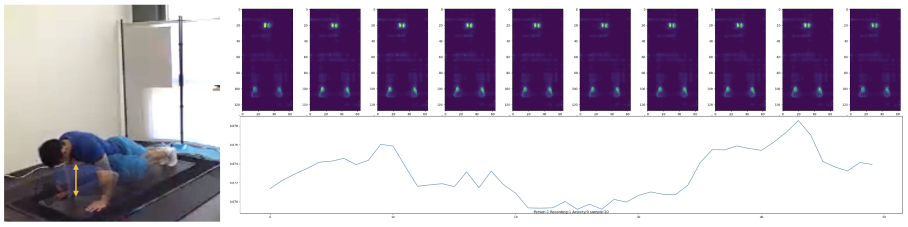


(d) Class 40: Hip Thrust. Same pose as Class 39, but move hip up and down with the help of glutes.

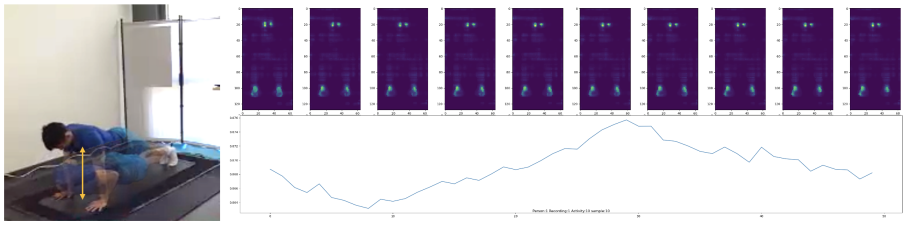


(e) Class 41: Single Hip Thrust. Same motion as 40, but with only one leg supporting the body and the other leg extended.

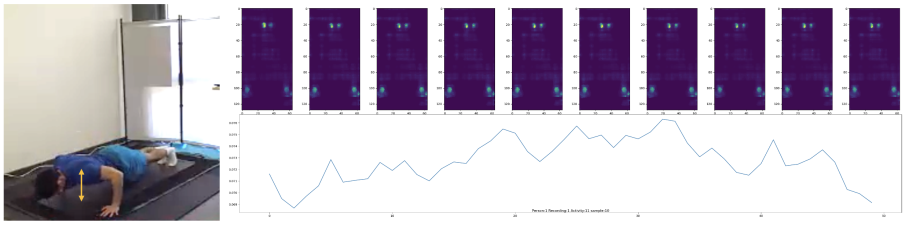
Fig. 17. Pressure map examples of Classes 5-7 from Cat I Crunches and Class 40-41 from Cat VIII Bridge.



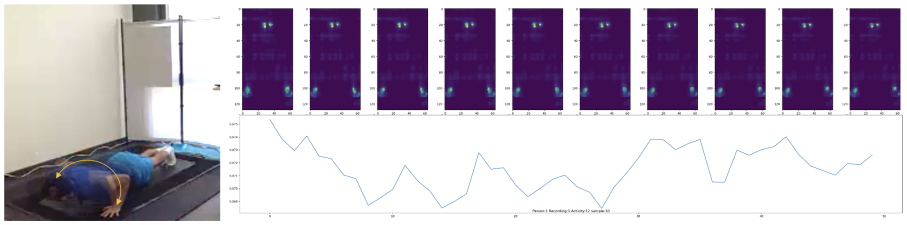
(a) Class 9: Push-up I. Hands shoulder wide and half range of motion, regulated by not fully extending elbows.



(b) Class 10: Push-up II. Hands shoulder wide and full range of motion, regulated by having elbows fully extended.

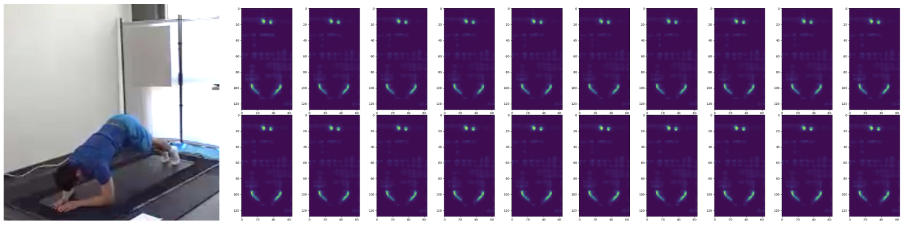


(c) Class 11: Push-up III. Hands wider than shoulder, full range of motion.

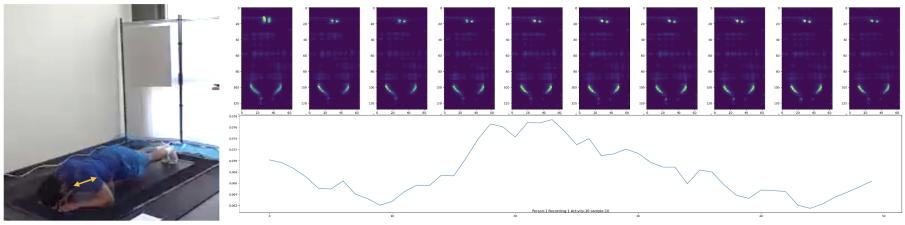


(d) Class 12: Alternating Push-up. Alternating single side push-up, hands wider than shoulder.

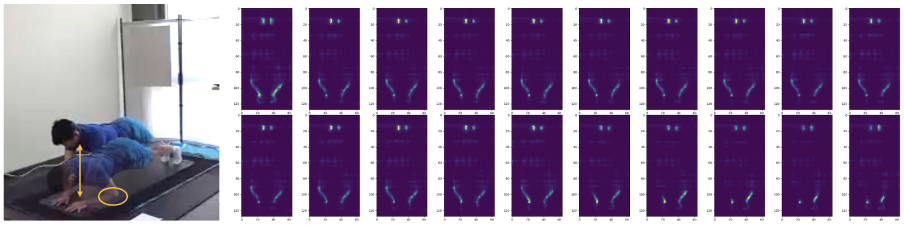
Fig. 18. Pressure map examples of Classes 9-12 from Cat II Push-ups.



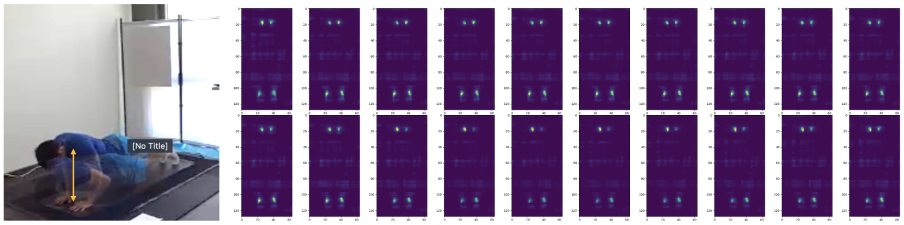
(a) Class 15: High Hip Plank. Raise hip higher than standard straight plank.



(b) Class 16: Plank Dip. Standard Plank but move body forward and backwards around the elbow support.

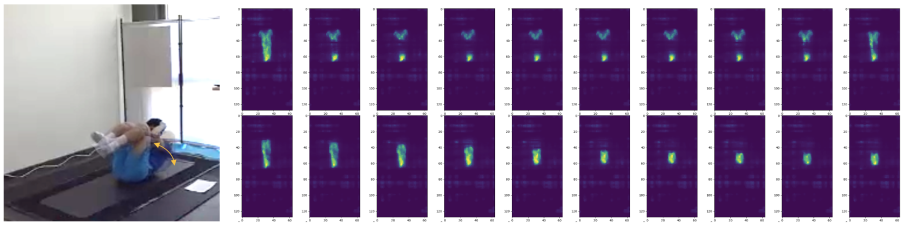


(c) Class 17: Plank Push-up. Hands are chest wide, change between plank position and push-up position.

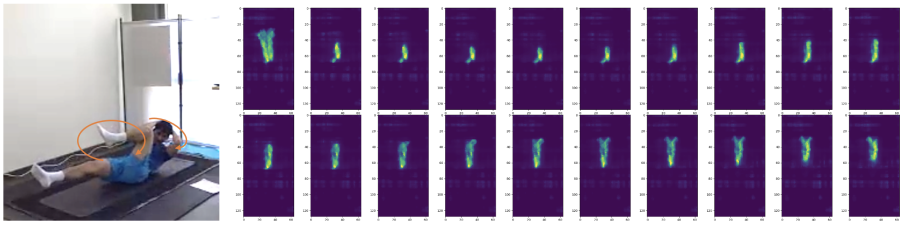


(d) Class 18: Chest Wide Plank Push-up. Hands are chest wide (same as 17), only push-up without placing elbows on the mat.

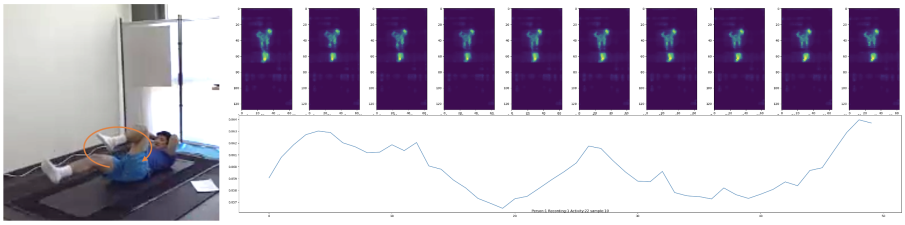
Fig. 19. Pressure map examples of Classes 15-18 from Cat III Planking.



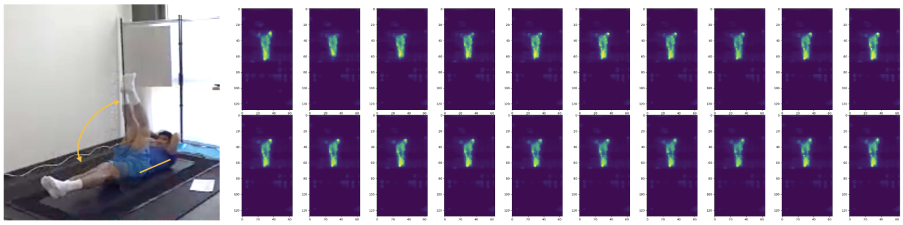
(a) Class 20:Leg-up Crunches. Lift legs, crunching exercise till elbows touch the knees.



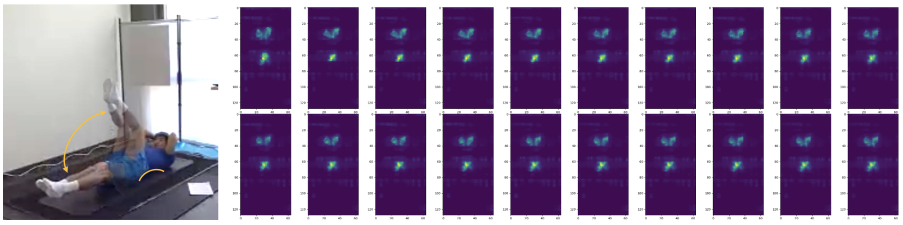
(b) Class 21: Alternating Cycling. Lift legs, alternately touch one elbow with the opposite knee, while the other opposite elbow-knee pairs are extended.



(c) Class 22: Leg-only Cycling. Lift elbows, empty cycle with only leg motions.

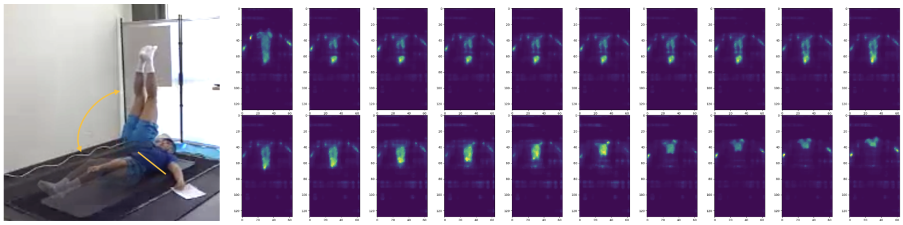


(d) Class 23: Leg-lift I. Raise legs from flat position to vertical position, with lower back pressing on the mat, this variation engages the abdominal muscles more effectively.

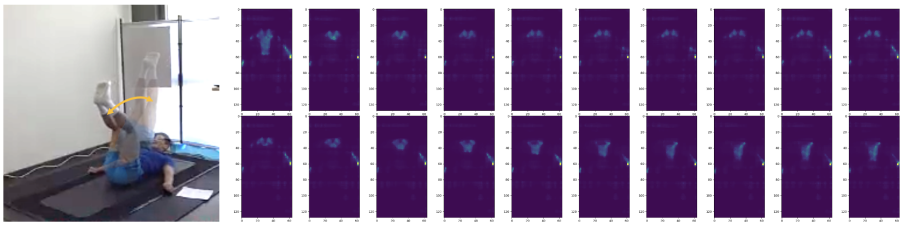


(e) Class 24: Leg-lift II. Same leg motion as Class 23, but have lower back suspended with an arched spine, and instead use hip as the anchoring point, this variation engages the abdominal muscles less effectively.

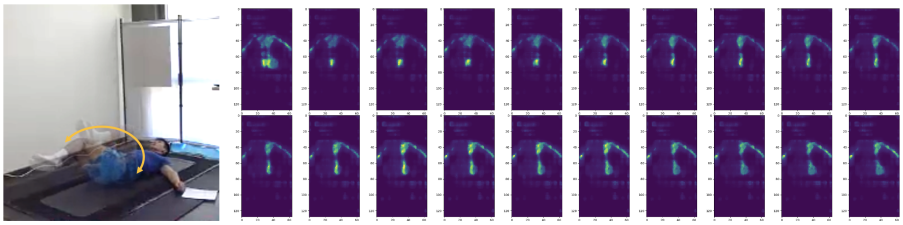
Fig. 20. Pressure map examples of Classes 20-24 from Cat IV Only Back Visible (to the mat).



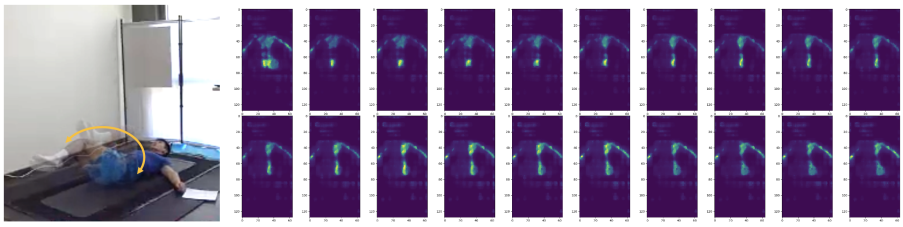
(a) Class 26: Leg-raise I. Raise legs vertically and upwards from the relaxed horizontal position with lower back suspended.



(b) Class 27: Leg-raise II. Similar as 26, but always keep thighs upwards without lowering legs to the horizontal position, this variation engages the torso muscles more than Leg-raise I.

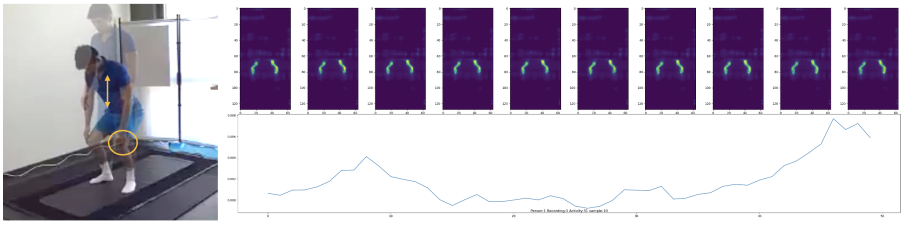


(c) Class 28: Leg Swing I. Swing legs from the left and horizontal, then upwards vertical, to the right horizontal positions, with knees bent.

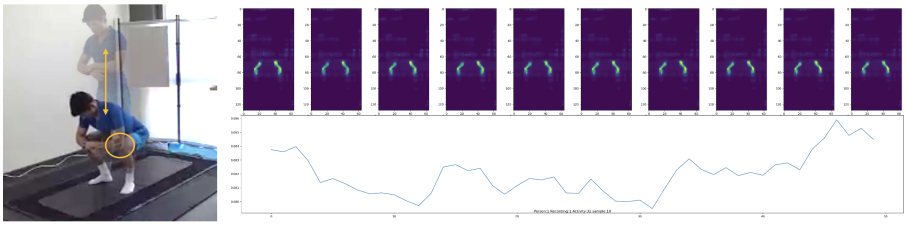


(d) Class 29: Leg Swing II. Similar as 28, but knees are extended straight.

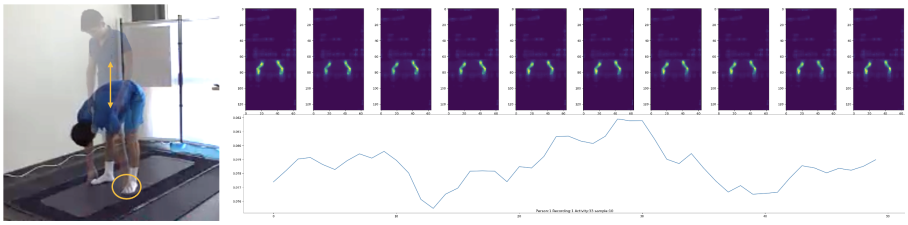
Fig. 21. Pressure map examples of Classes 26-29 from Cat V Back and Arms Visible (to the mat).



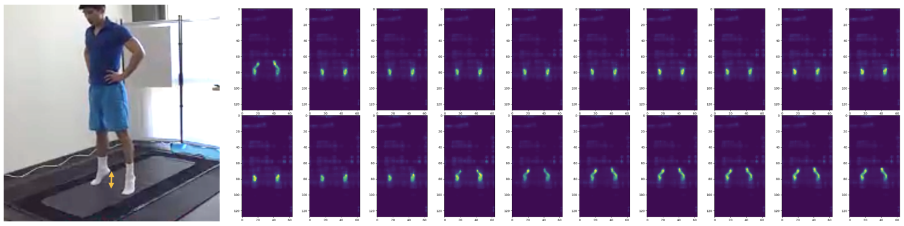
(a) Class 31: Shallow Squat. Empty squat, half range of motion, regulated by the hands reach the knees with relaxed arms.



(b) Class 32: Deep Squat. Empty squat, full range of motion, regulated by the elbows reach the knees with folded arms.

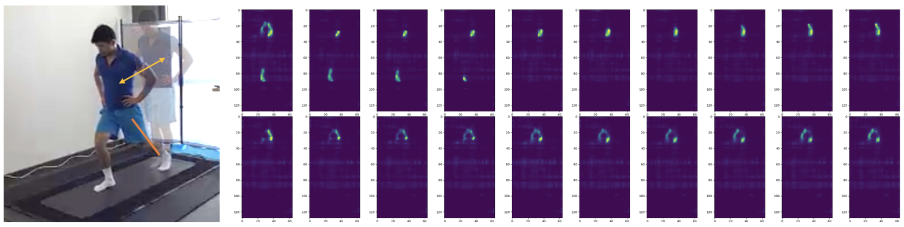


(c) Class 33: Toe Touch. Bend downwards and reach toes with fingers repetitively, heels may leave the mat.

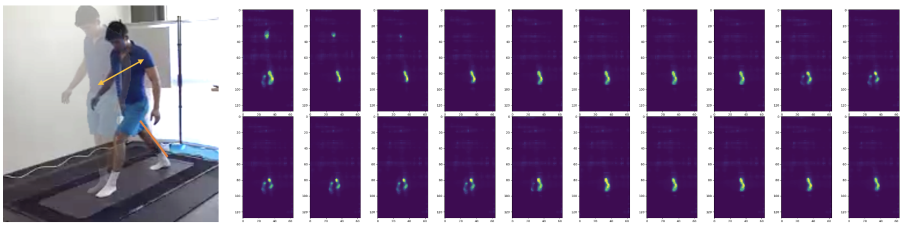


(d) Class 34: Tip Toe. Raise the heels repetitively.

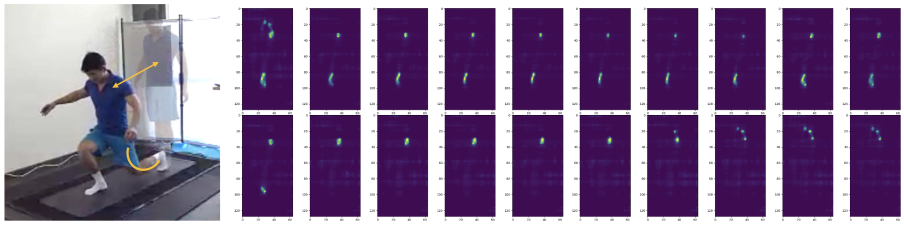
Fig. 22. Pressure map examples of Classes 31-34 from Cat VI Standing Up.



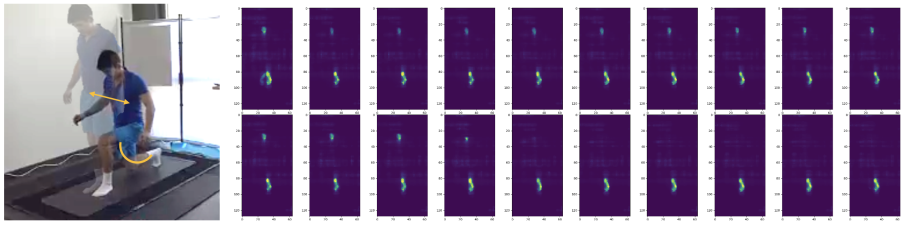
(a) Class 35: Arch Step Reference I. Alternately step forwards and then back, with one foot staying behind



(b) Class 36: Arch Step Reference II. Alternately step backwards and then front, with one foot staying in front.

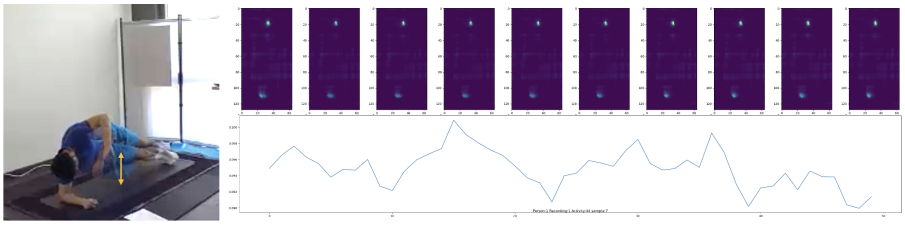


(c) Class 37: Lunge Forwards. Step similar as 35, but with the knee behind bent downwards.

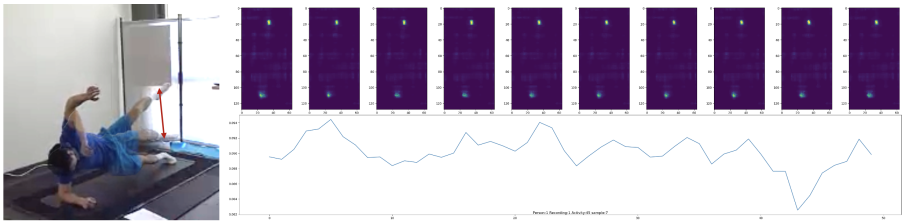


(d) Class 38: Lunge Backwards. Step similar as 36, but with the knee behind bent downwards.

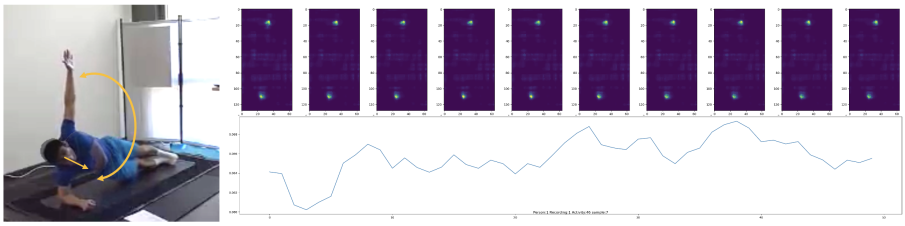
Fig. 23. Pressure map examples of Classes 35-36 from Cat VII Arch Step.



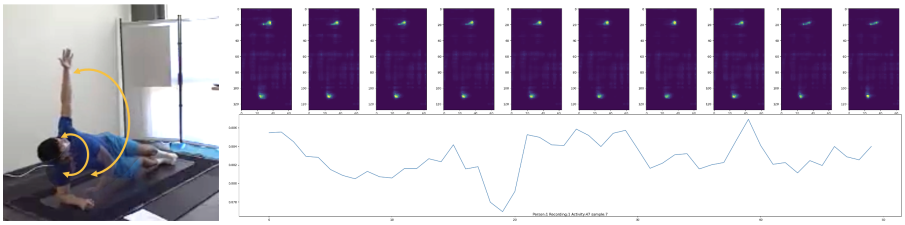
(a) Class 44: Side Hip Thrust. Similar pose as 42, and move hip up and down.



(b) Class 45: Side Leg Raise. Similar pose as 42, and raise and lower the leg that is not supporting the body.



(c) Class 46:Arm Swing I. Similar pose as 42, and swing the free arm from below the body to the vertical upwards position, face the front of the body.



(d) Class 47: Arm Swing II. Similar motion as 45, but rotate head and face the moving hand.

Fig. 24. Pressure map examples of Classes 44-47 from Cat IX Side Plank.