Motion Units: Generalized Sequence Modeling of Human Activities for Sensor-Based Activity Recognition

Hui Liu
Cognitive Systems Lab
University of Bremen
Bremen, Germany
hui.liu@uni-bremen.de

Yale Hartmann
Cognitive Systems Lab
University of Bremen
Bremen, Germany
yale.hartmann@uni-bremen.de

Tanja Schultz
Cognitive Systems Lab
University of Bremen
Bremen, Germany
tanja.schultz@uni-bremen.de

***Abstract*—This paper proposes an innovative activity modeling method for human activity recognition, which partitions the human activity into a sequence of shared, meaningful, and activity distinguishing states, called Motion Units, analog to phonemes in speech recognition. The partitions and generalization define a human activity dictionary, which endows this method with operability, universality, and expandability. Our preliminary experiments demonstrate on-par accuracy with other models while requiring fewer parameters and increasing separability between phases. Furthermore, the developed model was easily transferred with minor adjustments to two other datasets, demonstrating the proposed method's scalability. This framework enables expandable, interpretable, and scaleable modeling and recognition of human activities.**

***Index Terms*—human activity recognition, Hidden Markov Models, wearable sensors**

I. Introduction

Human Activity Recognition (HAR) has been playing an increasingly important role in daily life. In addition to video-based recognition, researches on sensor-based biosignal processing and topological modeling continued to emerge [1], [2]. Sensor-based activity recognition seeks profound high-level knowledge about human activities from multitudes of low-level sensor readings [3]. Many time series modeling technologies have proven their capabilities in HAR, such as Deep Neural Networks (DNN) and Hidden Markov Models (HMM). Both modeling technologies focus on the typical HAR problem, in which inputs are multichannel time series biosignals recorded from a set of sensors and outputs are predefined human activities. The research on DNN usually aims to refine the automatically learned features as the higher-level abstract representation of low-level raw time series signals through the deep architecture [3], [4]. In neural networks, the training and recognition procedure of a target activity must be divided into layers, which are often uninterpretable. In contrast, the concept of "states" in the HMM definition-tuple [5] may have the better explanatory power of the activities' internal structure.

Previous research on HAR sequential modeling often uses a fixed number of states, such as described in [6]. Similarly, the end-to-end HAR research framework [7] and the pilot real-time HAR system [8] modeled each activity with a single HMM state, i.e., the fundamental units of model training and recognition are "activities" themselves. Let us consider an analogical example in HMM-based speech recognition. Supposing "words" are modeled as the smallest recognizable units, the accuracy on a small dataset might be higher than a system relying on phonemes. However, this kind of recognizer has low training efficiency, low expandability, and low adaptability. The same holds for HAR.

HMM-based HAR has involved primitively expanding the number of states: [8] demonstrated the results of repeated experiments on the different number of states for each activity model, and [9] applied six states on all activities for a feature space study based on the best performance of repeated experiments. In [6], researchers used ten states for each activity. No matter the fixed number of states, each state's meaning is still unknown, similar to a DNN. Hence, there are two problems worth further exploring. First, could/should each activity contain a separate, explanatory number of states? If the answer is "yes," the explosion of possible combinations renders seeking a model based on repeated experiments unfeasible. Therefore, is there an approach to design HMMs of human activities more rule-based, normalized over blindly "trying"? As follows, we attempt to solve the two questions using the proposed modeling technology, *Motion Units*.

II. Methodology

In order to illustrate our idea of human activity modeling more clearly, we take a typical HMM-based word modeling in speech recognition as an example.

Figure 1 shows a three-state Bakis-model [10] constructing each phoneme (/d/ and /ɪ/). Each state, also called subphoneme, models parts of a phoneme (begin/middle/end). Following this topology, we can practically build a simple dictionary of the acoustic model without regard to contextual dependency for simplifying notations: {did: /d/-/ɪ/-/d/, dig: /d/-/ɪ/-/g/, gid: /g/-/ɪ-/d/, gig: /g/-/ɪ-/g/, digged: /d/-/ɪ/-/g/-/d/, gigged: /g/-/ɪ/-/g/-/d/, ...}.

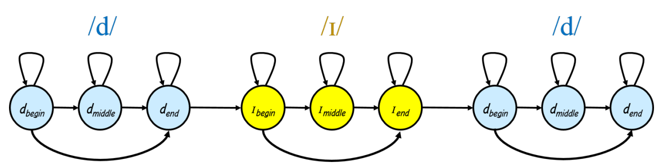

Fig. 1. A typical linear left-right HMM of the phoneme sequence "did."

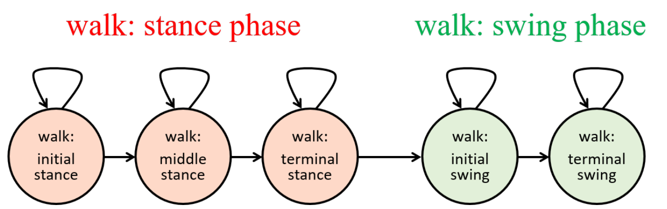

Fig. 2. A linear left-right HMM for one gait in the activity "walk." Red: states/sub-phases in the stance phase; green: states/sub-phases in the swing phase.

Phoneme and sub-phoneme are conventional in the research aspects of linguistic and speech signal processing. How to analogically define the intern topology of human activity?

A. Phase and State Partitioning of Human Activities

Some recognition systems only focus on upper body activities, such as Airwriting [11], which belongs to the research fields of gesture/posture recognition rather than activity recognition. Most sensor-based HAR research, including the above-listed pieces, uses and sometimes only uses body-worn sensors placed below the waist, such as pant's pockets, thighs, knees, shanks, and feet, because the lower body plays a decisive role in position translation, an essential part of most human activities. Therefore, the research on gait analysis provides us the first clue. [12] and [13] commonly distinguish two phases into seven sub-phases during one full gait cycle. The initial contact event is often used as the start/end event of a gait cycle and may explicitly be modeled as seen in [14] and [15]. However, this does not mean that directly applying "seven" or "eight" as the number of states to all activities will achieve the best accuracy, as clarified in previous work [9]. For example, some of these events have a too short duration that does not fit a single HMM state, while others may require specific sensor positions to distinguish. There is still a gap to bridge from biological research and sports science to informatics modeling.

In collaboration with kinesiologists of the Institute of Sport and Sports Science at Karlsruhe Institute of Technology, we decided to model one gait cycle as five states, representing three and two states respectively in the "two phases" gait analysis, based on the investigation of sports and gait science knowledge (e.g., [12]–[15]), the phase duration analysis, and the inspiration from speech recognition (e.g., Figure 1).

Figure 2 depicts the modeling scheme of a typical gait-based activity "walk." A complete gait is divided into two phases as observed from one leg: the stance (ground-contacting) phase and the swing phase. In the stance phase, we adopted three states (initial/middle/terminal) by analogizing the sub-phoneme in speech recognition, while in the swing phase, we designed only two states (initial/terminal). These states can go by the name of sub-phases analogically. We have also investigated that during a regular "walk" activity, the duration of the stance phase varies between 200ms and 800ms, and the swing phase between 200ms and 600ms, which provides an essential reference for the window length selection in subsequent tasks of training and recognition. Other gait-based activities, such as "run," "go upstairs/downstairs," and "V-cut to the left/right," follow closely to the pattern of "two-phase-five-state" gait modeling. In the study of partitioning, model generalization (see Section II-B) has not been applied yet, so each state of each activity has its unique name, i.e., currently, in the entire HMM-dictionary, no activities share states.

We continued to design the HMM modeling of more activities progressively through the experience accumulation from the modeling procedure of gait-based activities. The "two-phase-five-state" topology fits mainly the gait-based activities. Consequently, we must investigate the sports science knowledge for each new activity and analyze the data to derive the quantities and topology. For example, static activities like "stand" and "sit" can be described using only one state; in vertical shifting activities where the feet stay in place like "stand up" or "sit down" are modeled by two states (initial/terminal); "jump" is divided into three phases (takeoff/shift-up/land) with five sub-phases (see Section II-C).

In the following, we use **I**, **M**, and **T** to denote the sub-phases "initial," "middle," and "terminal" respectively, while the phase names like "stance," "swing," "shift," "takeoff," and "land" are abbreviated as **St**, **Sw**, **Sh**, **To**, and **La** respectively. For example, **TSw** represents the state "terminal swing."

It is worth mentioning that the number of generated states based on sports science knowledge and signal duration analysis is not necessarily the final optimal solution; thereupon, a certain amount of repeated experiments for fine-tuning may still help. Compared with the blindly "trying" or uniformly use of a fixed number of states for all activities, the phase and state partitioning is more interpretable and expandable, which serves as a benchmark model for the following model generalization research.

B. Model Generalization and Motion Units (MUs)

The next step is to study the generalizability of these states to simplify the overall modeling further, for which we also start with the analogy of speech recognition. In Figure 1 and the example dictionary, the Bakis-models of the phoneme /d/, /ɪ/, and /g/ are used repeatedly in the modeling (without regard to the variations due to contextual dependency), which will be generalized in the construction of the dictionary to simplify the entire HMM modeling topology, expand the vocabulary efficiently, and help train the models practically. Even taking account of contextual dependency and co-articulation, each language has its typical set of primarily fixed phonemes, which is undoubtedly beneficial for the model generalization. The model generalization in HAR is not as intuitive and convenient as in speech recognition, as the following three characteristics determine: First of all, an HAR system usually focuses on a unique application scenario. Therefore, its included activities are also application-specific, some of which even distinctive. Secondly, the human activity itself is an infinitely extendable connotation, inside which we can always discover, define and even invent new activities. Last but not least, even though we can find several basic and indispensable, but seemingly identical activities in different HAR researches, such as "stand," "sit," "walk," "run," and "go upstairs/downstairs," their exact definitions and descriptions diverge due to different application scenarios, equipment, and research requirements.

To explain our proposed model generalization scheme concisely, let us take some gait-based activities introduced in Section II-A as an example. In non-generalized modeling, "walk," "go upstairs," and "run" all contain two phases and five states that are different from each other, i.e., we must use fifteen distinct states to model these three activities. Not to mention repetitions like multiple gait cycles in each activity. Which of these states can be generalized? The most straightforward consideration is from inertial biosignals' general knowledge: the **ISw** states and the **TSw** states in these three activities' swing phases cannot be merged. "Walk" and "go upstairs" have different translational directions, while "walk" and "run" have different movement speeds. We concentrate on the generalizability of the remaining three states in the stance phase. There are two critical approaches:

Theoretical research (e.g., based on sports and biosignal knowledge). Taking our in-house collected dataset CSL19 as an example, besides inertial sensors, we also use EMG sensors on the thigh and shank, respectively. Based on the EMG-signal knowledge and the statistical analysis of the activities "walk" and "go upstairs," the **ISt** states, regarded as "muscle initialization for the movement about to happen," are mergeable, while the **TSt** states are not because the driving muscles and the translational inertial signals are quite different. What about the **MSt** states?

Experimental research. Repeated experiments or clustering analysis can help investigate whether the theoretically indeterminable states like **MSt** are globally, partially, or hardly generalizable. Notably, the complexity of the repeated experiments in this step is not large-scaled. The preliminary partitioning and theoretical generalization design have already provided an appropriate baseline model. Moreover, although we have studied the generalizability of the **ISt** and **TSt** states of "walk" and "go upstairs" in theoretical research, it is still necessary to use repeated experiments or clustering analysis to verify the theoretical design's reliability.

The final states are designed based on the approaches described above, regardless of whether they are used repeatedly in the applied modeling dictionary. We call them *Motion Units* (MUs), and they are the generalized smallest recognizable units composing each human activity in the HAR system.

C. MU-DNA (Directional Nomenclature & Anchored)

MUs can play an essential role in efficiently modeling new activities and should be given interpretable and meaningful names. To name an MU with a specific motion trial, we propose a Six-Directional Nomenclature (6DN). In brief, 6DN means the six directions front (forwards), back(wards), up(wards), down(wards), left(wards), and right(wards) in the torso coordinate system, and their various combinations. We can use the letters **F**, **B**, **U**, **D**, **L**, and **R** to abbreviate them, respectively.

The theoretical and experimental research in the example of "walk" versus "go upstairs" results in them sharing the first two MUs, **walk-ISt-F** and **walk-MSt-F**. The remaining third, fourth, and fifth MUs are named as follows, respectively: "walk": **walk-TSt-F**, **walk-ISw-F**, and **walk-TSw-F**; "go upstairs": **walk-TSt-FU**, **walk-ISw-FU**, and **walk-TSw-FU**. It is noticeable that "walk" in the above-mentioned MU names represents the primary category of these two activities (they are both gait-based activities; "go upstairs" is a sub-activity to "walk" with a unique direction), **St** and **Sw** indicate the phases, and **I**, **M**, and **T** denote the sub-phases, as described in Section II-A. **F** (front) and **FU** (front+up) are related to more exclusive states within 6DN to distinguish different MUs according to movement directions.

Based on 6DN, we can quickly model new activities preliminarily. For example, "V-cut to the left" is also a gait-based activity that can be analogically modeled as {**run-ISt-F**, **run-MSt-F**, **run-TSt-FL**, **run-ISw-FL**, **run-TSw-FL**}, and "jump upwards" is modeled as {**jump-ITo-U**, **jump-TTo-U**, **jump-Sh-U**, **jump-ILa-D**, **jump-TLa-D**}. For another example of expandability, a football-specific activity "beat out a shot by diving right-forward" may be modeled by combining "jump" and the directions **FUR** and **FDR** of 6DN.

All activities described by 6DN can be decomposed into one or several axial translational movements, which will cause the body or body part to leave its original position. If an activity or an MU does not involve translational movement, we can regard them as "an **Anchored** activity/MU" and do not need to use 6DN to name them. An obvious example is the activity "stand" and its single MU **stand**, to which adding any direction is superfluous.

In summary, for any activity and its attached MUs, they either have translational movement (named with 6DN) or not (Anchored). We abbreviate such a "Motion Units' Directional Nomenclature/Anchored" pattern as the "MU-DNA" of human activities. Moreover, considering the literal aesthetics, we can also abbreviate "Motion Units' Generalization" as "MU-Gene," making it picturesque to think of MUs as MU-Gene and MU-DNA (different from the biological meaning).

Rapid and straightforward modeling does not make up the entirety of model research. Adequate repeated verification experiments and parameter tuning are still essential.

III. EXPERIMENTS AND EVALUATION

The real-world applicability of the introduced framework is evaluated by comparing five models (single-state, fixed-state, partition, and Motion Unit topologies, as well as a skyline) across three datasets (CSL19 [16], UniMiB SHAR [17], and ENABL3S [18]). All datasets focus on Activities of Daily Living (ADL) like “walk,” “go upstairs,” and “sit.” Additionally, the CSL19 dataset includes several sport-related activities, while the UniMiB SHAR dataset includes different types of falls.

The upper and lower reference in the experiments is created using a six-state (listed as “Fixed” in Table I) and single-state (listed as “Single” in Table I) topology, as these have been proven competent [7]–[9]. The first topology (listed as “Partition” in Table I) based on our framework introduces cycles as well as phase and state partitioning as described in Section II-A. The second topology (listed as “Motion Units” in Table I) shares states between activities and implements MUs as described in Section II-B.

The partitioning and MU topologies have been developed for the CSL19 dataset and then applied directly to the UniMiB SHAR and ENABL3S datasets, where both datasets’ unique activities extended the CSL19 topology. In the UniMiB SHAR and ENABL3S datasets, the precise number of gait cycles per segment varies. Therefore, a single gait cycle is modeled, and an initial and terminal “random”-state consumes all other data for the partitioning and MU experiments.

Lastly, a skyline based on [16] for the CSL19 and UniMiB SHAR datasets is added. The ENABL3S skyline is created using the partitioning model. Note that the UniMiB SHAR accuracy is 4% lower than the reference as the LDA-based feature space reduction was omitted. The skylines are allowed up to thirty Gaussians per Gaussian Mixture Model (GMM) in each HMM state, while all other models are limited to seven Gaussians per GMM. This behavior ensures a good skyline and a comparable number of parameters in the topology experiments, as the number of parameters scales proportionally with the number of states in the HMM topology.

The implemented recognizer and chosen hyperparameters are based on [16]. The recognizer for the CSL19 dataset uses 100ms Hamming windows with an overlap of 50ms. Mean and root-mean-square features are calculated per window, and the whole sequence normalized. On the UniMiB SHAR dataset, 400ms Hamming windows with an overlap of 320ms and normalization is used. A 20-dimensional combination of features, including min, slope, and spectral spread, was chosen from the TSFEL [19] library with an HMM-based greedy forward feature selection. The hyperparameters for the ENABL3S dataset were determined by grid search in person-independent 10-fold cross-validation and are as follows: 50ms Hamming windows with 20ms overlap with normalization. As features, mean, slope, spectral centroid, and spectral kurtosis are calculated for each sensor channel.

In each experiment, the HMM recognizer is evaluated using person-independent leave-one-subject-out cross-validation with the full sequence. Additionally, an independent evaluation is performed on the force-aligned and as vectors relabeled data (excluding the random states’ data) for each a k-nearest neighbors algorithm (KNN), linear discriminant analysis (LDA), and nearest centroid classifier. Figure 3 indicates this duality: the HMM recognizer are evaluated with sequence data (blue lines), and independently trained HMMs provide alignments (colored points) for the LDA, KNN, and centroid classifiers. The figure additionally illustrates that the HMM learns correct meanings, as the states, such as **ISt** and **MSt**, are correctly and automatically labeled [13].

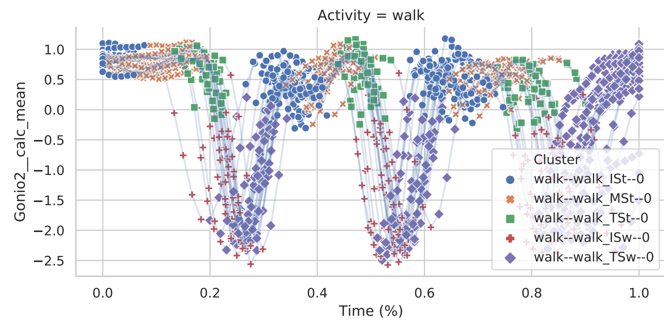

Fig. 3. HMM force-alignments based on the “Partition” topology for the activity “walk” in the CSL19 dataset. Twenty randomly sampled sequences are plotted on a percentage-based time axis. The lines show each sequence, while colors/shapes denote each vector’s state as determined by the alignment.

Table I lists the experimental results and reads as follows: the number of states denotes the sum of individual states across all activities and denotes the number of target classes for the LDA to learn. The number of Gaussians denotes the sum of unique Gaussians and, in combination with the number of units, indicates the number of trainable parameters. The HMM column denotes recognition accuracy over activities, while the LDA accuracy provides an indirect measure of state separability and is not directly comparable to the HMM accuracy. The KNN and nearest centroid classifiers are omitted for brevity as they highly correlate with the LDA evaluation.

Our experiments show that phase and state partitioning can retain recognition accuracy compared to a fixed-state topology while requiring fewer training parameters. Table I shows a decrease of 0.3%, 0.7%, and 2.4% absolute on the CSL19, UniMiB SHAR, and ENABL3S datasets, respectively. At the same time, the trainable parameters decrease by 40%, 22%, and 34% relative, and state separability, as indicated by the LDA accuracy, increases significantly for most datasets. The MUs in the following experiment further decrease the number of trainable parameters while retaining (UniMiB SHAR and ENABL3S) or slightly decreasing accuracy (3.3% on CSL19). In continuation of the partitioning experiments, state separability is increased further.

Too few trainable parameters might explain the distinct performance drops. Recall that the only difference between skyline and partitioning settings on the ENABL3S dataset is the allowed Gaussians (thirty and seven respectively), which is penalized with a drop of 2.4% absolute accuracy. A similar lack of parameters might explain the difference between partitioning and MU topology on the CSL19 dataset, but further investigation is required as several hyperparameters differ.

When comparing these results to state-of-the-art, it is es-

TABLE I
MOTION UNITS: PRELIMINARY RESULTS

Model	CSL19				UniMiB SHAR				ENABL3S			
	#States	#Gauss.	HMM	LDA	#States	#Gauss.	HMM	LDA	#States	#Gauss.	HMM	LDA
Skyline	94	1347	0.936	0.232	141	1234	0.727	0.247	32	582	0.948	0.473
Fixed	132	924	0.939	0.214	92	631	0.715	0.134	42	294	0.948	0.386
Partition	94	658	0.936	0.234	81	557	0.708	0.261	32	224	0.924	0.467
Motion Units	57	399	0.903	0.299	77	529	0.707	0.284	28	196	0.923	0.465
Single	22	154	0.892	0.081	17	119	0.640	0.169	7	49	0.917	0.336

sential to note that higher performances might be achieved without the restriction in trainable parameters. The results of [16] for the UniMiB SHAR and CSL19 datasets have been replicated without feature space reduction as the skyline in Table I. The MU-based recognizer has slightly lower accuracy (3.3% on CSL19 and 2.0% on UniMiB SHAR) but uses significantly fewer parameters while increasing state separability and interpretability. On the ENABL3S dataset, [20] achieves 93.60% with a user-adaptive system, and [21] achieves 92.74% (reported as 7.26% error) with a CNN-based approach, both with a leave-one-out cross-validation scheme. Our skyline (94.8%) with partitioning as well as the MU-based recognizer (92.3%) compare well here.

IV. CONCLUSION AND FUTURE WORK

This paper introduced an innovative, interpretable, and easily extensible modeling technique for HMM-based HAR, by utilizing phase/state partitioning and state generalization across activities, similar to phonemes in speech recognition.

The preliminary experiments show that modeling activities with phase/state partitioning and state generalization allow for fewer parameters while maintaining recognition accuracy and improving class separability. Furthermore, applying the same topology to three datasets showed that the framework scales extraordinarily across different sensors and sensor positions.

The proposed method bridges the gap from movement science to machine learning and opens several new topics. The easy extension allows expansion to further sensor setups, new activities, and additional datasets. The higher separability between phases enables an extensive feature study and a neural network as HMM emission model. Most importantly, this lays the foundation for well-defined meaning-carrying and activity-composing units for HAR: Motion Units.

REFERENCES

[1] S. Ramasamy Ramamurthy and N. Roy, “Recent trends in machine learning for human activity recognition—a survey,” *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, pp. 1–11, jul 2018.

[2] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, “Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey,” vol. 1, no. 1, apr 2020.

[3] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[4] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, “Deep convolutional neural networks on multichannel time series for human activity recognition,” in *IJCAI*, vol. 15, 2015, pp. 3995–4001.

[5] L. R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” in *PROCEEDINGS OF THE IEEE*, 1989, pp. 257–286.

[6] D. Rebelo, C. Amma, H. Gamboa, and T. Schultz, “Human activity recognition for an intelligent knee orthosis,” in *BIOSIGNALS 2013 - 6th International Conference on Bio-inspired Systems and Signal Processing*, 2013, pp. 368–371.

[7] H. Liu and T. Schultz, “ASK: A framework for data acquisition and activity recognition,” in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOSIGNALS*, INSTICC. SciTePress, 2018, pp. 262–268.

[8] H. Liu and T. Schultz, “A wearable real-time human activity recognition system using biosensors integrated into a knee bandage,” in *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 1: BIODEVICES*, INSTICC. SciTePress, 2019, pp. 47–55.

[9] Y. Hartmann, H. Liu, and T. Schultz, “Feature space reduction for multimodal human activity recognition,” in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4: BIOSIGNALS*, INSTICC. SciTePress, 2020, pp. 135–140.

[10] R. Bakis, “Continuous speech recognition via centisecond acoustic states,” *The Journal of the Acoustical Society of America*, vol. 59, no. S1, pp. S97–S97, 1976.

[11] C. Amma, D. Gehrig, and T. Schultz, “Airwriting recognition using wearable motion sensors,” in *First Augmented Human International Conference*, 2010, p. 10, ACM.

[12] M. W. Whittle, “Clinical gait analysis: a review,” *Human movement science*, vol. 15, no. 3, pp. 369–387, 1996.

[13] M. W. Whittle, *Gait analysis: an introduction*. Butterworth-Heinemann, 2014.

[14] M. A. B. Arous *et al.*, “Knee kinematics feature selection for surgical and nonsurgical arthroplasty candidate characterization.” in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOSIGNALS*, 2018, pp. 176–181.

[15] N. Mezghani *et al.*, “Identification of knee frontal plane kinematic patterns in normal gait by principal component analysis,” *Journal of Mechanics in Medicine and Biology*, vol. 13, no. 03, p. 1350026, 2013.

[16] Y. Hartmann, H. Liu, and T. Schultz, “Feature space reduction for human activity recognition based on multi-channel biosignals,” in *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies*, INSTICC. SciTePress, 2021, pp. 215–222.

[17] D. Micucci, M. Mobilio, and P. Napoletano, “UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones,” *Applied Sciences*, vol. 7, no. 10, p. 1101, 2017.

[18] B. Hu, E. Rouse, and L. Hargrove, “Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals,” *Frontiers in Robotics and AI*, vol. 5, p. 14, 2018.

[19] M. Barandas *et al.*, “TSFEL: Time series feature extraction library,” *SoftwareX*, vol. 11, p. 100456, 2020.

[20] K. Zhang, J. Wang, C. W. De Silva, and C. Fu, “Unsupervised cross-subject adaptation for predicting human locomotion intent,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 3, pp. 646–657, 2020.

[21] U. H. Lee, J. Bi, R. Patel, D. Fouhey, and E. Rouse, “Image transformation and CNNs: A strategy for encoding human locomotor intent for autonomous wearable robots,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5440–5447, oct 2020.