A Practical Wearable Sensor-Based Human Activity Recognition Research Pipeline

Hui Liu[1][a], Yale Hartmann [1][b] and Tanja Schultz[1][c]

[1]*Cognitive Systems Lab, University of Bremen, Bremen, Germany*

{*hui.liu, yale.hartmann, tanja.schultz*}*@uni-bremen.de*

Keywords: human activity recognition, wearable healthcare, biodevices, biosignals, segmentation, annotation, feature extraction, digital signal proccesing, machine learning

Abstract: Many researchers devote themselves to studying various aspects of Human Activity Recognition (HAR), such as data analysis, signal processing, feature extraction, and machine learning models. In response to the fact that few documents summarize and form intuitive paradigms for the entire HAR research pipeline, based on the purpose of sharing our years of research experience, we propose a practical, comprehensive HAR research pipeline, called HAR-Pipeline, composed of nine research aspects, aiming to reflect the overall perspective of HAR research topics to the greatest extent and indicate the sequence and relationship between the tasks. Supplemented by the outcomes of our actual series of studies as examples, we demonstrate the proposed pipeline's rationality and feasibility.

1 INTRODUCTION

In this digital age, Human Activity Recognition (HAR) has been playing an increasingly important role in almost all aspects of life. HAR is often associated to the process of determining and naming human activities using sensory observations (Weinland et al., 2011). The recognition of human activities has been approached in two different ways, namely using external and wearable (internal) sensors (Lara and Labrador, 2012). External sensing technologies require the devices fixed in predetermined points of interest, so the inference of activities entirely depends on the voluntary interaction of the users with the sensors, while the devices for internal sensing are attached to the user, which leads to the research topic of wearable biosignal-based HAR. Whether based on external or internal sensing, HAR research involves various topics, such as hardware (equipment, sensor design, sensing technology, among others), software (acquisition, data visualization, signal processing, among others), and Machine Learning (ML) approaches (feature study, modeling, training, recognition, evaluation, among others). However, few documents summarize and form a paradigm for the entire framework of HAR research, which may be due to the fact that researchers usually focus on the research in one or several fields of HAR, such as modeling optimization, automatic segmentation, feature selection, and application scenarios, rather than the overall HAR process. Nevertheless, there are still articles trying to review the tasks in HAR research comprehensively. For example, (Bulling et al., 2014) put five blocks for body-worn inertial-sensor-based HAR into a chain, called Activity Recognition Chain (ARC), as the HAR research guideline, comprising stages for data acquisition, signal preprocessing and segmentation, feature extraction and selection, training, and classification. (Ke et al., 2013) reviewed video-based HAR and pointed out similar sub-tasks, but for video, the approaches and algorithms applied in the tasks are different from those for the biosignal-based research.

In our years of research, we have found that the chain model needs to be supplemented to a certain extent. In other words, the chain model is not necessarily a general research process that solves all potential problems. In fact, the overall research of HAR is not linear, of which there are many cycles and backtracking according to purposes and actual conditions. If researchers divide their research phases in the early stage to follow close to the chain consistency, it is very likely that when they discover insufficient or faulty early results in the later stage, they will find it challenging to rewind to the steps exactly.

[a] https://orcid.org/0000-0002-6850-9570
[b] https://orcid.org/0000-0002-1003-0188
[c] https://orcid.org/0000-0002-9809-7028

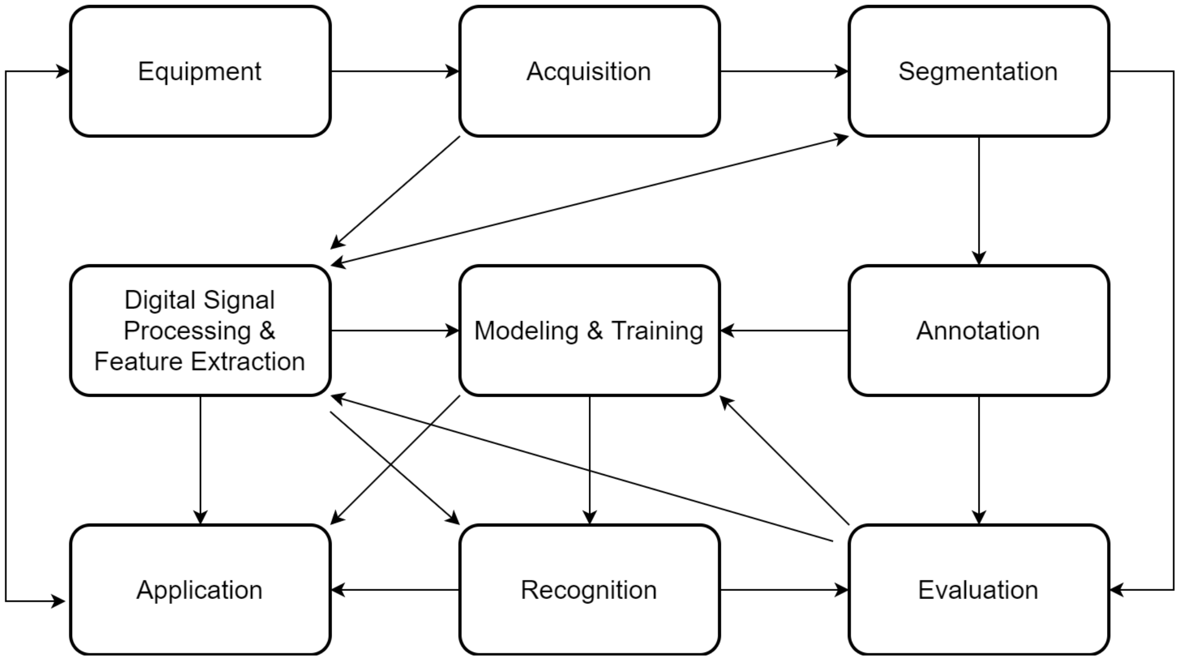

Figure 1: The proposed HAR-Pipeline for HAR research.

Researchers who are just entering the HAR field would be willing to understand how to construct a research plan for the research object at hand, while researchers who have already gone deep into certain aspects of HAR may also need to clearly understand what other tasks can be studied to improve their research results. Therefore, it is helpful to summarize and propose a detailed research pipeline for guidance. The HAR-Pipeline proposed in this paper is based on our years of research experience, whose completeness and operability have been evidenced through the outcomes of our actual series of studies.

Some following subsections are revised and improved based on the relevant chapters in the first author's doctoral thesis (Liu, 2021). Therefore, the description and examples are mainly related to the biosignal-based HAR of wearable sensing. From a macro perspective, we believe that the proposed pipeline also fits video-based HAR research to a large extent. For video-based HAR, several sub-tasks in the HAR-Pipeline need to be further supplemented. For example, besides classifying what activity is being done, video-based HAR systems sometimes need to recognize where the activity happens in the video sequence. However, such compensations do not affect the flow sheet of the nine topics in the HAR-Pipeline.

Although the pipeline outlines an overall HAR research perspective, the exploration of each link requires customized definitions and plans according to research objects and goals.

2 PIPELINE

Figure 1 illustrates the proposed HAR-Pipeline of end-to-end HAR research. The arrows between each component in Figure 1 indicate the processing order in the study. These nine components are essential and indispensable for the research of a complete end-to-end HAR system. The number of critical components will be reduced if the research aims at only one or several of these topics, but we can clearly understand the pre-stages and related topics linked to the target tasks according to the pipeline. The following subsections will expand on all components of the HAR-Pipeline.

2.1 Equipment: Devices and Sensors

Selecting the appropriate appliance for signal acquisition is essential during research preparation. For different sensing technologies, such as video-based sensing, biosensor-based sensing, smart home, among others, the related equipment involved is quite different, but each has a specific range of options. There are many considerations for choosing the applicable equipment, such as:

- Application scenarios
- Research requirements
- Signal transmission technologies
- Site situations
- Financial conditions

For biosignal-based HAR, almost all kinds of biosignal acquisition equipment are capable of particular HAR tasks, depending on different research purposes. As stated in Section 2.4, choosing equipment based on application consideration is a good starting point. The double-headed arrow between the “Equipment” and the ”Application” blocks in the pipeline stands for this relationship. For example, when the application scenario of the HAR research is daily life assistance or interactive entertainment, placing the mobile phone in a certain pocket of the clothes/trousers to sense the human inertial signals will become a very convenient, efficient, and reasonable equipment choice that fits the ultimate use case (Micucci et al., 2017) (Garcia-Gonzalez et al., 2020).

2.2 Software and Data Acquisition

HAR research relies on large amounts of data, which includes the laboratory data collections or real-world acquisition that meet in-house research goals, as well as the usage of external and public databases to verify models and methods.

If the research does not involve collecting in-house data, suitable public open-source datasets can be found and applied. In this case, other research teams have already done the “Acquisition” tasks in the HAR-Pipeline. For example, for biosignal-based HAR, many open-source datasets focusing on various application scenarios, sensor combinations, activity definitions, and body parts are online available, such as (Garcia-Gonzalez et al., 2020), *Opportunity* (Chavarriaga et al., 2013) (Roggen et al., 2010), *UniMiB SHAR* (Micucci et al., 2017), *Gait Analysis DataBase* (Loose et al., 2020), *ENABL3S* (Hu et al., 2018), *Upper-body movement* (Santos et al., 2020), *FORTH-TRACE* (Karagiannaki et al., 2016), *Real-World* (Sztyler and Stuckenschmidt, 2016), *PAMAP2* (Reiss and Stricker, 2012b) (Reiss and Stricker, 2012a), and *CSL-SHARE* (Liu et al., 2021a).

When researchers decide to record in-house data for specific research purpose, data collection will become an essential part of the entire HAR research after selecting the appropriate acquisition equipment. Usually, drivers for different programming languages are provided with the devices/products of sensor solutions sold on the industrial market, allowing users to access the devices and record the signals themselves. Many providers even offer multi-functional data acquisition software together with sensor products. However, if there are additional requirements and particular approaches for the data collection process, researchers sometimes have to implement customized programs or software.

2.3 Segmentation and Annotation

The task of segmentation in the HAR-Pipeline is to split a relatively long sequence of activities into several segments of single activity, which are suitable for model training and offline recognition, while annotation is the process of labeling each segment, such as “walk,” “jump,” or “cutting a cake,” according to different definitions of human activities in different datasets and application scenarios.

In many cases, segmentation and annotation are performed simultaneously. However, we separate them into two sub-tasks in the pipeline instead of merging them together for the following reasons:

- Segmentation and annotation have different post-stage topics linked to them. As Figure 1 shows, segmentation is undoubtedly a prerequisite for annotation, and its output will be the input for signal processing and feature extraction, while annotation generates labels for two follow-up tasks: training and evaluation.
- The generation of annotated labels indeed accompanies most segmentation work, but segmentation can also become a research object by itself, such as ML-based automatic segmentation. The segmented data provided to the signal processing and feature extraction sub-tasks do not require the participation of annotation.
- The annotation itself can also be a research object, such as the definition and disambiguation of single motion, motion sequence, among others.

Segmentation can be performed manually. In the segmentation work of video-based HAR or biosignal collection supplemented by the video camera(s) recording the whole process, the acquired dataset will be segmented by dedicated persons relying on the video. Another approach of manual segmentation for biosignal-based HAR is the use of data visualization. If the collected signals have good recognizable discrimination, we can also segment the data by directly visualizing the signals. Such being the case, video is not necessarily required. If we thoroughly know what happened during the data collection, e.g., through detailed text records, the process will be more efficient. Taking Figure 2 as an example, since the sensors are marked clearly in the visualization, if we accurately get informed what activities happened during the data acquisition, we can segment and annotate the data manually only based on the data visualization, without applying any video information.

The advantages of manual segmentation are apparent. It is straightforward and intuitive, and the result should be close to the human’s understanding of

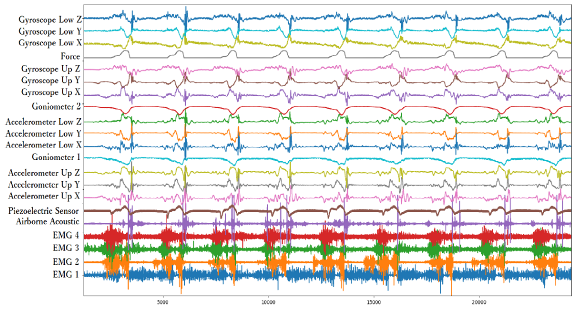

Figure 2: An example of multisensorial data visualization for data segmentation.

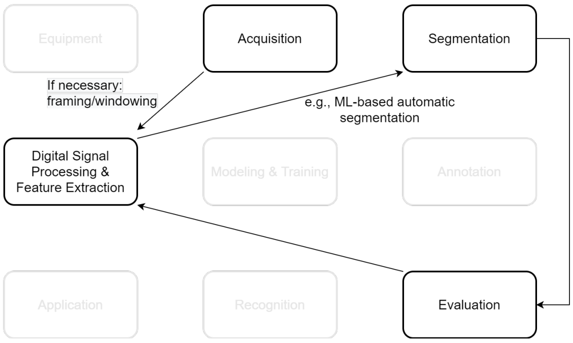

Figure 3: The segmentation-study-loop in the HAR-Pipeline.

"activity." The segmented data has, therefore, strong rationality and readability. Moreover, manual segmentation can often be accompanied by annotation — marking each segment with a predefined label. However, the shortcomings of manual segmentation cannot be ignored.

First, manual segmentation has high requirements for the operators: concentration, patience, attentiveness, and even the need to receive some training in advance to adapt to the segmentation requirements for specific research tasks. Even so, manual segmentation is still unavoidably subjective, resulting in poor repeatability and errors due to human factors. (Kahol et al., 2003) uses a comparative example to corroborate the subjectivity during the manual segmentation: for the segmentation work of the same data piece, the first annotator set 10 boundaries, while the second annotator set 21. The synchronization mechanism between video signals and biosignals will also affect the quality of the segmentation results on biosignal data — often, acoustic or optical signals are used to confirm the starting/ending synchronization points in time. Last but not least, manual segmentation is more expensive in terms of time and labor cost.

Besides manual segmentation, modern ML methods, like *Gaussian Mixture Models* (GMMs), *Principal Component Analysis* (PCA), and *Probabilistic Principal Component Analysis* (PPCA), are being used to segment human activity automatically or semi-supervised (Barbič et al., 2004). (Guenterberg et al., 2009) applied signal energy to segment data. Moreover, appropriate features of long-term signals instead of segments can be extracted for the segmentation task, such as the research in (Ali and Aggarwal, 2001). The research on segmentation algorithms forms a segmentation-study-loop in the HAR-Pipeline, as shown in Figure 3.

In addition, some datasets use non-conventional segmentation methods to save time, such as simple statistical analysis-based segmentation and manual intervention-based semi-automatic segmentation, which can enable researchers to obtain expected data segments as early as possible for subsequent research without having to stay too long at this stage.

The data in the *UniMiB SHAR* dataset (Micucci et al., 2017) are automatically and uniquely segmented into three-second windows around a magnitude peak during the activities. This automatic segmentation mechanism is effortless to execute and has no demand on equipment or ML algorithms. However, the resulted segments are not always correct due to a certain number of too-long or too-shot activity segments, as well as misinterpreted peaks that do not belong to the assigned activity. The use of fixed-value window length has a good simulation for real-time HAR systems.

In the CSL-SHARE dataset, a pushbutton was applied for a semi-automated segmentation and annotation solution, of which the applicability and correctness have been verified during numerous experiments (see Section 3). The so-called protocol-for-pushbutton mechanism loads a predefined activity sequence protocol during each data recording session and prompts the user to perform the activities one after the other. Each activity is displayed on the screen one by one while the user controls the activity recording by pushing, holding, and releasing the pushbutton. For example, the user sees the instruction "please hold the pushbutton and do: sit-to-stand" and prepares for it, then pushes the button and starts to do the activity "sit-to-stand." They keep holding the pushbutton while standing up from sitting, then release it to finish this activity. With the release, the system displays the next activity instruction, e.g., "stand," the process continues until the predefined acquisition protocol is fully processed.

The protocol-for-pushbutton mechanism was implemented to reduce the time and labor costs of manual annotation. The resulting segmentations required little to no manual correction, and lay a good foundation for subsequent research. Nevertheless, this mechanism has some limitations (Liu et al., 2021a):

- The mechanism can only be applied during acquisition and is incapable of segmenting archived data;
- Clear activity start-/endpoints need to be defined, which is impossible in cases like field studies;
- Activities requiring both hands are not possible due to participants holding the pushbutton;
- The pushbutton operation may consciously or subconsciously affect the activity execution;
- The participant forgetting to push or release the button results in subsequent segmentation errors.

2.4 Biosignal Processing, Feature Extraction, and Feature Study

HAR research is inextricably linked with signal processing. Compared to traditional electronic signals, biosignals have some unique properties that highlight the research topics of biosignal processing.

Some signal processing jobs can occur before segmentation (directly after or even during the acquisition), such as filtering, amplification, noise reduction, and artifact removal. Another common example is normalization, which can also be applied to the whole collected biosignals instead of segments. Real-time systems need to use accumulated normalization because what we obtain from a real-time recording is always the continuous influx of short-term signal streams. Besides, feature extraction may also occur before segmentation, such as in the research of feature-based segmentation (Ali and Aggarwal, 2001), as described in Figure 3.

Due to the characteristics of biosignals and the demand for training and decoding, the segmented biosignals need to be preprocessed before further steps, a typical application of *Digital Signal Processing* (DSP). Usually, the biosignals are firstly windowed using a specific window function with overlap. Taking the common-used biosignals as an example, a mean normalization is usually applied to the inertial and the EMG signals to reduce the impact of Earth acceleration and set the EMG signals' baseline to zero. Then, the EMG signals are rectified, a widely adopted signal processing method for muscle activities (Liu and Schultz, 2018).

Because multimodal biosignals or video data for HAR systems are usually large-scale data, it is not common to use the raw data directly. Therefore, subsequently, features are extracted from each of the resulting windows.

Figure 4 illustrates with a schema the windowing and feature extraction on multichannel signals to

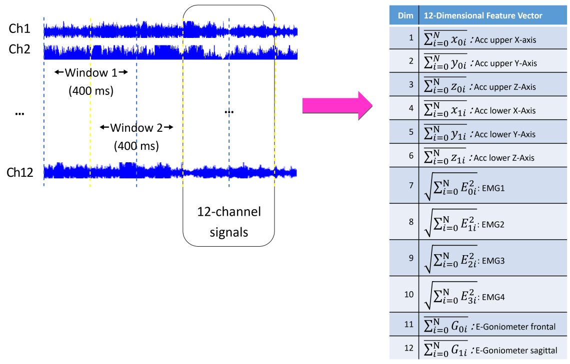

Figure 4: Example of building a feature vector: windowing and feature extraction for 400 ms window size.

build feature vectors. The 12-channel signals are windowed through a shifting window with a length of 400 ms and an overlap of 200 ms. Usually, the overlap between two adjacent windows can have a length chosen between 0 and the window length: the smaller the overlap length, the longer the training time of the model. Based on the windowing function, features will be extracted from each channel and form a feature vector of a window, which will be used for the follow-up tasks of training and recognition. The feature vector in the example of Figure 4 has a minimal dimension of twelve when only one feature is extracted from each signal channel.

Figure 4 implies two typical features applied in many pieces of HAR research works, namely average and *Root Mean Square* (RMS), from the statistical domain. Besides, there are also various applicable features of time series in the time domain and the frequency domain. (Figueira et al., 2016) summarized many features for HAR research in statistical, temporal, and spectral domains. Hence, numerous features can be extracted from various types of signals. The use of existing open-source feature libraries, such as the *Time Series Feature Extraction Library* (TSFEL) (Barandas et al., 2020) and the *Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests* (tsfresh) (Christ et al., 2018), will significantly broaden the types of functional features and improve the efficiency of feature calculation.

Features of different signals can be combined by early or late fusion, i.e., the feature vectors of single signal streams are either concatenated to form one multi-signal feature vector (early fusion), or recognition is performed on single signal feature vectors, and the combination is done on decision level (late fusion).

Usually, the modeling, training and recognition research of the HAR-Pipeline takes feature-related research as the premise, including feature dimensionality study (feature vector stacking and feature space

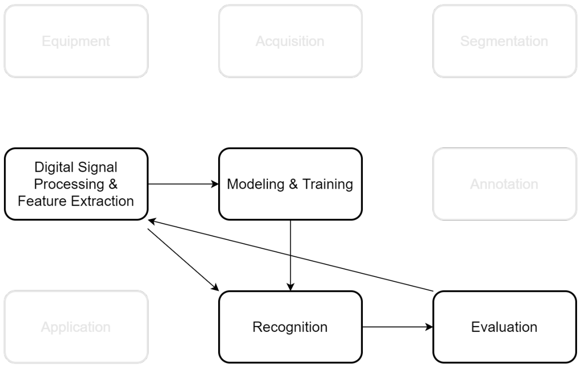

Figure 5: The feature-study-loop in the HAR-Pipeline.

reduction) and feature selection. Figure 5 depicts the iterative feature-study-loop in the HAR-Pipeline.

For feature space reduction research, commonly used methods include *Principal Component Analysis* (PCA) and *Linear Discrimination Analysis* (LDA). The former does not require the annotated labels, while the latter does. For feature selection, methods such as *Minimum Redundancy Maximum Relevance* (mRMR) (Peng et al., 2005) and *Analysis of Variance* (ANOVA) (St et al., 1989) (Girden, 1992) can be applied practically. The greedy forward feature selection approach based on complete training and recognition operation can provide more convincing results, but it often costs days to run an experiment. It is noteworthy that the preliminary studies of features do not necessarily provide the optimal solutions of the entire HAR system, but should offer a strong baseline as the point of departure for the iterative improvement process. Figure 5 manifests the iterative process.

2.5 Activity Modeling, Training, Recognition and Evaluation

Various ML methods for modeling have been applied to model human activities from sensor data effectively for later training and recognition, such as *Deep Neural Networks* (DNNs) and *Hidden Markov Models* (HMMs).

Many pieces of research works have shown the capability of *Convolutional Neural Networks* (CNNs) (Lee et al., 2017) (Ronaoo and Cho, 2015) and *Recurrent Neural networks* (RNNs) (Inoue et al., 2018) (Singh et al., 2017) for HAR research. Recently, *Residual Neural Network* (ResNet) models (He et al., 2016), which proved to be a compelling improvement of DNN for image processing, have also been used to research human activity recognition. A small amount of literature has already occurred in this direction (Tuncer et al., 2020), (Keshavarzian et al., 2019), (Long et al., 2019). However, in many cases, researchers do not know each layer's specific physical meaning in neural networks. In contrast, the concept of "states" in the HMM definition-tuple (Rabiner, 1989) may have the better explanatory power of the activities' internal structure. In addition to the interpretability, HMM has other advantages for HAR study, such as the generalizability and reusability of models and states (Liu et al., 2021b).

HMMs are widely used for various activity recognition tasks, such as (Lukowicz et al., 2004) and (Amma et al., 2010). The former applies HMMs to an assembly and maintenance task, while the latter presents a wearable system that enables 3D handwriting recognition based on HMMs. In this so-called *Airwriting* system, the users write text in the air as if they were using an imaginary blackboard, while the handwriting gestures are captured wirelessly by accelerometers and gyroscopes attached to the back of the hand.

Based on approaches for obtaining adequate activities' knowledge, such as human activity duration analysis (Liu and Schultz, 2022), the activity models for training and recognition can be built by any appropriate modeling method for HAR systems such as CNNs, RNNs, or HMMs. After training the model by taking the feature vector sequence from the "DSP & Feature Extraction" task (see Section 2.4) and the labels from the "Annotation" task (see Section 2.3), the research follows the decoding of the activities based on the prepared feature vector sequence and provides the recognition results of the most probable activities. In some research occasions, Top-N mode can be applied to generate N recognition results sorted by probabilities. In other words, the recognition result of the Top-N mode is not just one activity but N most probable activities.

A series of criteria and indicators will be applied to evaluate the prediction results using the ground truth provided by annotation: recognition accuracy, precision, recall, F-score, confusion matrix, among others. The evaluation results will contribute to improving the modeling, training, segmentation studies, and feature studies. Besides the segmentation-study-loop (see Figure 3) and the feature-study-loop (see Figure 5) introduced above, the parameter-tuning-loop aims to adjust the important parameters in training, such as the number of Gaussians for each emission model for HMM-based HAR. New modification of the modeling, such as state amount and state generalization, can also happen during the re-training process. The loop for parameter tuning and modeling optimization in Figure 6 guides the iterative experiments.

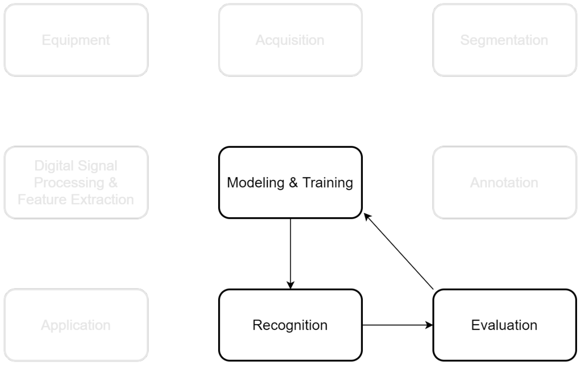

Figure 6: The parameter-tuning-loop or modeling-optimization-loop in the HAR-Pipeline.

2.6 Application

The purpose of most HAR research is to contribute to a practical application environment, such as auxiliary medical care, rehabilitation technology, safety assurance, and interactive entertainment. In many instances, researchers must undergo adjustments from offline recognition to real-time online performance.

From the final optimized models to reach the application level, there are many tasks to accomplish, such as user demand analysis, software interface development, user customization, network and server technology, among others, some of which should not be of concern to HAR researchers. However, it should be pointed out that the "Application" block in the HAR-Pipeline may not only be the end of the entire HAR research, just as we did not set the "Equipment" block as the starting point. The arrow from "Application" to "Equipment" displays that, in practice, the application considerations play a decisive role in equipment selection (and other related tasks).

3 RESEARCH EXAMPLES FOLLOWING THE PIPELINE

3.1 From Application to Equipment

In our research, we planned to build an end-to-end wearable sensor-based HAR system for assisting the early treatment of gonarthrosis, which is under the framework of a collaborative research project. Therefore, we used a knee bandage provided by one of the project partners as a wearable carrier of sensors, aiming to develop an HAR-based mobile technology system that senses its users' movements utilizing proximity sensors.

The wearable sensor carrier determines which biosignal acquisition devices and sensors we should consider and compare, and how we integrate the selected devices and sensors into the knee bandage. Related research procedures of equipment are explicated in (Liu and Schultz, 2018). Finally, we chose *biosignalsplux*[1], providing expandable solutions of hot-swappable sensors and automatic synchronization. One hub from the *biosignalsplux* research kit records biosignals from 8 channels, each up to 16 bits, simultaneously. The selected accelerometers, gyroscopes, and electrogoniometer offer relatively slow signals, while the nature of the EMG and microphone signals requires higher sampling rates. Low-sampled channels are up-sampled to be synchronized and aligned with high-sampled channels.

3.2 From Equipment to Software Development, Data Acquisition, Segmentation, and Annotation

We developed a software called *Activity Signal Kit* (ASK) (Liu and Schultz, 2018) with a *Graphical User Interface* (GUI) and multi-functionalities using the driver library provided by the company of the devices and sensor products. The ASK software connects and synchronizes recording devices automatically. In the subsequent three data acquisition events, we used two or three *biosignalsplux* hubs as recording devices. Therefore, ASK collects up to 24-channel sensor data from all hubs simultaneously and continuously. All recorded data are archived orderly with dates and timestamps for subsequent application.

The novel protocol-for-pushbutton mechanism of segmentation and annotation (see Section 2.3) has been implemented in the ASK software. Moreover, the baseline ASK software provides the functionalities of signal processing, feature extraction, modeling, training, and recognition by applying our in-house developed HMM-based decoder *BioKIT* (Telaar et al., 2014). As a summary, the ASK baseline software has the following features:

- Connects to wearable biosignal recording devices;
- Enables multisensorial acquisition and archiving;
- Implements protocol-for-pushbutton mechanism of practical segmentation and annotation;
- Processes biosignals and extracts feature vectors for iterative feature studies (see Figure 5);
- Facilitates modeling research with the training-recognition-evaluation iteration (see Figure 6).

A series of upgraded and expanded versions of the ASK baseline software, such as the real-time end-to-end HAR system and its on-the-fly add-on, have

[1] biosignalsplux.com

been developed based on modeling and recognition achievements after the proof-of-concept modeling experiments (Liu and Schultz, 2019).

After finishing the ASK software development and the first testing cycle, we applied it to collecting a pilot one-subject seven-activity dataset to validate the HAR-Pipeline and the software's practicability and robustness. The experimental results are described in (Liu and Schultz, 2018). After ensuring that the data collection function of the ASK Software runs efficiently without errors and obstacles, we continued to record a larger four-subject dataset of 18 activities using the ASK software. The *CSL-SHARE* (Cognitive Systems Lab Sensor-based Human Activity REcordings) dataset (called *CSL19* in the earlier pieces of literature) is a follow-up to the two datasets mentioned above and was recorded in a controlled laboratory environment. It contains 20 participants' 22 activities in 17 channels of 4 sensor types. Standing on the dataset's robustness according to our numerous experimental results, we have shared the *CSL-SHARE* dataset as an open-source wearable sensor-based dataset, contributing research materials to the researchers in related fields (Liu et al., 2021a).

This "pilot—advanced—comprehensive—share" data accumulation process well reflects the application of the HAR-Pipeline: After each dataset was collected, it went through each pipeline component for verification, laying the foundation for the improved collection work in the next stage.

3.3 From Data to Feature Study

Based on the segmented and annotated data, we continued to extract various features by utilizing the joint-developed feature library. An example of feature visualization on the above-introduced *CSL-SHARE* dataset is given in (Barandas et al., 2020). Subsequently, we studied the feature selection and feature dimensionality (feature vector stacking and feature space reduction) on our in-house collected datasets and one external open-source dataset to create a good benchmark for the subsequent modeling study, such as presented in (Hartmann et al., 2020) (Hartmann et al., 2021), and (Hartmann et al., 2022).

3.4 From Feature Study to Human Activity Modeling Research

On different datasets, we applied various types of HMM modeling topologies to study the human activity modeling according to the study-loop illustrated in Figure 6. Each activity can be modeled using a single HMM state (Liu and Schultz, 2018), (Liu and Schultz, 2019), or a fixed number (greater than one) of HMM states (Hartmann et al., 2020), (Liu and Schultz, 2019), (Rebelo et al., 2013). Both topologies work, but with shortcomings (Xue and Liu, 2021).

Regarding the fact that no matter the fixed number of states, each state's meaning is still unknown, (Liu et al., 2021b) explores two problems: Could/should each activity contain a separate, explanatory number of states? Is there an approach to design HMMs of human activities more rule-based, normalized over blindly "trying"? A novel modeling technology, *Motion Units*, endowed with operability, universality, and expandability, was proposed to solve the questions.

3.5 From Modeling Research to Application

A wearable real-time HAR system *Activity Signal Kit Echtzeit-Decoder* (ASKED) (Liu and Schultz, 2019) was further implemented based on the modeling experimental results on the pilot dataset, which verifies the data recording, feature extraction, training, and recognition functionality in the ASK baseline software (see Section 3.2).

Balance of accuracy versus speed was first studied to improve real-time recognition performance. A shorter step-size of windows shift results in a shorter delay of the recognition outcomes, but the interim recognition results may fluctuate due to temporary search errors. On the other hand, longer delay due to larger step-sizes contradicts a real-time system's characteristics, though it generates more accurate interim recognition results. According to the activity duration analysis (Liu and Schultz, 2022), the experimental results, and the user experience, a balancing setting of 400 ms window length with 200 ms window overlap performs the best, providing satisfactory recognition results with a barely noticeable delay.

More introduction to our real-time HAR system, including its engaging on-the-fly add-on functionality, can be found in (Liu and Schultz, 2019).

4 CONCLUSIONS

Based on the purpose of sharing our years of research experience, in this paper, we propose a practical, comprehensive HAR research pipeline, called HAR-Pipeline, composed of nine research aspects, aiming to reflect the entire perspective of HAR research topics to the greatest extent and indicate the sequence and relationship between the tasks. Supplemented by our actual series of studies as examples, we exhibited the proposed pipeline's feasibility.

REFERENCES

Ali, A. and Aggarwal, J. (2001). Segmentation and recognition of continuous human activity. In *Proceedings IEEE Workshop on Detection and Recognition of Events in Video*, pages 28–35. IEEE.

Amma, C., Gehrig, D., and Schultz, T. (2010). Airwriting recognition using wearable motion sensors. In *First Augmented Human International Conference*, page 10. ACM.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). TSFEL: Time series feature extraction library. *SoftwareX*, 11:100456.

Barbič, J., Safonova, A., Pan, J.-Y., Faloutsos, C., Hodgins, J. K., and Pollard, N. S. (2004). Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface*, pages 185–194. Citeseer.

Bulling, A., Blanke, U., and Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):1–33.

Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Tröster, G., Millán, J. d. R., and Roggen, D. (2013). The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033–2042.

Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, 307:72–77.

Figueira, C., Matias, R., and Gamboa, H. (2016). Body location independent activity monitoring. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4: BIOSIGNALS*, pages 190–197. INSTICC, SciTePress.

Garcia-Gonzalez, D., Rivero, D., Fernandez-Blanco, E., and Luaces, M. R. (2020). A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors*, 20(8):2200.

Girden, E. R. (1992). *ANOVA: Repeated measures*. Sage.

Guenterberg, E., Ostadabbas, S., Ghasemzadeh, H., and Jafari, R. (2009). An automatic segmentation technique in body sensor networks based on signal energy. In *Proceedings of the Fourth International Conference on Body Area Networks*. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering.

Hartmann, Y., Liu, H., Lahrberg, S., and Schultz, T. (2022). Interpretable high-level features for human activity recognition. In *BIOSIGNALS 2022 — 15th International Conference on Bio-inspired Systems and Signal Processing*. INSTICC, SciTePress. forthcoming.

Hartmann, Y., Liu, H., and Schultz, T. (2020). 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*, pages 135–140. INSTICC, SciTePress.

Hartmann, Y., Liu, H., and Schultz, T. (2021). 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*, pages 215–222. INSTICC, SciTePress.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *CVPR 2016 - IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Hu, B., Rouse, E., and Hargrove, L. (2018). Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals. *Frontiers in Robotics and AI*, 5:14.

Inoue, M., Inoue, S., and Nishida, T. (2018). Deep recurrent neural network for mobile human activity recognition with high throughput. *Artificial Life and Robotics*, 23(2):173–185.

Kahol, K., Tripathi, P., Panchanathan, S., and Rikakis, T. (2003). Gesture segmentation in complex motion sequences. In *Proceedings of 2003 International Conference on Image Processing*, volume 2, pages II–105. IEEE.

Karagiannaki, K., Panousopoulou, A., and Tsakalides, P. (2016). The forth-trace dataset for human activity recognition of simple activities and postural transitions using a body area network.

Ke, S.-R., Thuc, H. L. U., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H., and Choi, K.-H. (2013). A review on video-based human activity recognition. *Computers*, 2(2):88–131.

Keshavarzian, A., Sharifian, S., and Seyedin, S. (2019). Modified deep residual network architecture deployed on serverless framework of iot platform based on human activity recognition application. *Future Generation Computer Systems*, 101:14–28.

Lara, O. D. and Labrador, M. A. (2012). A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3):1192–1209.

Lee, S.-M., Yoon, S. M., and Cho, H. (2017). Human activity recognition from accelerometer data using convolutional neural network. In *BIGCOMP 2017 - IEEE International Conference on Big Data and Smart Computing*, pages 131–134. IEEE.

Liu, H. (2021). *Biosignal Processing and Activity Modeling for Multimodal Human Activity Recognition*. PhD thesis, University of Bremen.

Liu, H., Hartmann, Y., and Schultz, T. (2021a). CSL-SHARE: A multimodal wearable sensor-based human activity dataset. *Frontiers in Computer Science*, 3:90.

Liu, H., Hartmann, Y., and Schultz, T. (2021b). Motion Units: Generalized sequence modeling of human activities for sensor-based activity recognition. In *EUSIPCO 2021 — 29th European Signal Processing Conference*, pages 1506–1510.

Liu, H. and Schultz, T. (2018). ASK: A framework for data acquisition and activity recognition. In *Proceedings of the 11th International Joint Conference on Biomedi-*

cal Engineering Systems and Technologies - Volume 3: BIOSIGNALS, pages 262–268. INSTICC, SciTePress.

Liu, H. and Schultz, T. (2019). 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*, pages 47–55. INSTICC, SciTePress.

Liu, H. and Schultz, T. (2022). How long are various types of daily activities? statistical analysis of a multimodal wearable sensor-based human activity dataset. In *HEALTHINF 2022 — 15th International Conference on Health Informatics*. INSTICC, SciTePress. forthcoming.

Long, J., Sun, W., Yang, Z., and Raymond, O. I. (2019). Asymmetric residual neural network for accurate human activity recognition. *Information*, 10(6):203.

Loose, H., Tetzlaff, L., and Bolmgren, J. L. (2020). A public dataset of overground and treadmill walking in healthy individuals captured by wearable imu and semg sensors. In *Biosignals 2020 - 13th International Conference on Bio-Inspired Systems and Signal Processing*, pages 164–171. INSTICC, SciTePress.

Lukowicz, P., Ward, J. A., Junker, H., Stäger, M., Tröster, G., Atrash, A., and Starner, T. (2004). Recognizing workshop activity using body worn microphones and accelerometers. In *In Pervasive Computing*, pages 18–32.

Micucci, D., Mobilio, M., and Napoletano, P. (2017). UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10):1101.

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77(2), pages 257–286.

Rebelo, D., Amma, C., Gamboa, H., and Schultz, T. (2013). Human activity recognition for an intelligent knee orthosis. In *BIOSIGNALS 2013 - 6th International Conference on Bio-inspired Systems and Signal Processing*, pages 368–371.

Reiss, A. and Stricker, D. (2012a). Creating and benchmarking a new dataset for physical activity monitoring. In *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 1–8.

Reiss, A. and Stricker, D. (2012b). Introducing a new benchmarked dataset for activity monitoring. In *ISWC 2012 - 16th International Symposium on Wearable Computers*, pages 108–109. IEEE.

Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., et al. (2010). Collecting complex activity datasets in highly rich networked sensor environments. In *INSS 2010 - 7th International Conference on Networked Sensing Systems*, pages 233–240. IEEE.

Ronaoo, C. A. and Cho, S.-B. (2015). Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors. *Journal of the Korean Information Science Society*, pages 858–860.

Santos, S., Folgado, D., and Gamboa, H. (2020). Upper-body movements: Precise tracking of human motion using inertial sensors.

Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., and Holzinger, A. (2017). Human activity recognition using recurrent neural networks. In *CD-MAKE 2017 - International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pages 267–274. Springer.

St, L., Wold, S., et al. (1989). Analysis of variance (anova). *Chemometrics and intelligent laboratory systems*, 6(4):259–272.

Sztyler, T. and Stuckenschmidt, H. (2016). On-body localization of wearable devices: An investigation of position-aware activity recognition. In *PerCom 2016 - 14th IEEE International Conference on Pervasive Computing and Communications*, pages 1–9. IEEE Computer Society.

Telaar, D., Wand, M., Gehrig, D., Putze, F., Amma, C., Heger, D., Vu, N. T., Erhardt, M., Schlippe, T., Janke, M., et al. (2014). Biokit—real-time decoder for biosignal processing. In *INTERSPEECH 2014 - 15th Annual Conference of the International Speech Communication Association*.

Tuncer, T., Ertam, F., Dogan, S., Aydemir, E., and Pławiak, P. (2020). Ensemble residual network-based gender and activity recognition method with signals. *The Journal of Supercomputing*, 76(3):2119–2138.

Weinland, D., Ronfard, R., and Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Computer vision and image understanding*, 115(2):224–241.

Xue, T. and Liu, H. (2021). Hidden Markov Model and its application in human activity recognition and fall detection: A review. In *CSPS 2021 - 10th International Conference on Communications, Signal Processing, and Systems*. forthcoming.