Semantic Loss: A New Neuro-Symbolic Approach for Context-Aware Human Activity Recognition

LUCA ARROTTA, GABRIELE CIVITARESE, and CLAUDIO BETTINI, University of Milan, Italy

Deep Learning models are a standard solution for sensor-based Human Activity Recognition (HAR), but their deployment is often limited by labeled data scarcity and models' opacity. Neuro-Symbolic AI (NeSy) provides an interesting research direction to mitigate these issues by infusing knowledge about context information into HAR deep learning classifiers. However, existing NeSy methods for context-aware HAR require computationally expensive symbolic reasoners during classification, making them less suitable for deployment on resource-constrained devices (e.g., mobile devices). Additionally, NeSy approaches for context-aware HAR have never been evaluated on in-the-wild datasets, and their generalization capabilities in real-world scenarios are questionable. In this work, we propose a novel approach based on a semantic loss function that infuses knowledge constraints in the HAR model during the training phase, avoiding symbolic reasoning during classification. Our results on scripted and in-the-wild datasets show the impact of different semantic loss functions in outperforming a purely data-driven model. We also compare our solution with existing NeSy methods and analyze each approach's strengths and weaknesses. Our semantic loss remains the only NeSy solution that can be deployed as a single DNN without the need for symbolic reasoning modules, reaching recognition rates close (and better in some cases) to existing approaches.

CCS Concepts: • **Human-centered computing** → **Mobile computing**; **Mobile devices**.

Additional Key Words and Phrases: human activity recognition, neuro-symbolic, knowledge infusion, context-awareness, knowledge-based reasoning

ACM Reference Format:
Luca Arrotta, Gabriele Civitarese, and Claudio Bettini. 2023. Semantic Loss: A New Neuro-Symbolic Approach for Context-Aware Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 147 (December 2023), 29 pages. https://doi.org/10.1145/3631407

1 INTRODUCTION

The sensor-based Human Activity Recognition (HAR) research area is dominated by solutions based on purely data-driven Deep Learning (DL) models [18, 57]. While DL-based solutions are very effective, they still have some open research issues that limit their deployment in real-world scenarios. Among the major problems, there are labeled data scarcity [1] and the lack of transparency of the activity models [9].

In the literature, purely knowledge-based approaches have been proposed to tackle both problems [26]. Symbolic methods rely on domain knowledge (e.g., based on common-sense knowledge) to model constraints between sensor events and activities. The sensor data stream is then matched with symbolic rules to identify the most likely activities according to knowledge. Purely knowledge-based methods have two advantages: 1) they do not require labeled data, and 2) they are based on human-readable formalisms that make them interpretable and transparent. However, these approaches are too rigid since it is unlikely that logic constraints can cover all the possible patterns related to activity execution. Moreover, they are not suitable for sensors that generate continuous values (e.g., accelerometer) since raw data can not be mapped to a clear semantic.

Authors' address: Luca Arrotta, luca.arrotta@unimi.it; Gabriele Civitarese, gabriele.civitarese@unimi.it; Claudio Bettini, claudio.bettini@unimi.it, University of Milan, Via Celoria, 18, Milan, Italy.

2474-9567/2023/12-ART147
https://doi.org/10.1145/3631407

In the general machine learning community, Neuro-Symbolic AI (NeSy) methods are emerging to combine the strengths of data-driven and knowledge-based methods [33]. The idea of NeSy methods is to enhance DL models through domain knowledge. The potential advantages of NeSy are many. First, it may significantly improve the recognition rate by driving the classification with domain constraints. This may be especially true when only a limited amount of labeled data is available; hence, those constraints can not be learned directly from data. For the same reason, the use of domain knowledge can potentially improve the classification of those cases out of the training set distribution samples. Moreover, DL models enhanced through domain knowledge have the potential of being more interpretable and transparent, since their decisions are also influenced by the knowledge model [37]. An emerging approach for NeSy methods is *knowledge infusion*, where the constraints from a symbolic knowledge model are internally learned by the deep learning model [50]. Knowledge infusion enables the model to leverage knowledge constraints within the latent space, allowing it to both capitalize on this knowledge and effectively cope with intrinsic uncertainty in real-world data.

This work focuses on sensor-based HAR on mobile/wearable devices (e.g., smartphones, smartwatches). While the majority of existing works in this field only focus on inertial sensors, we also consider high-level context data (e.g., semantic position, weather) as also proposed by other research groups [8, 16, 45]. This research area is usually referred to as *Context-Aware Human Activity Recognition*. In this domain, running the recognition model directly on mobile/wearable devices is a desirable aspect when real-time recognition is a requirement. In fact, continuously transmitting sensor signals to a service provider can result in increased latency [3]. Moreover, onboard sensor processing may be preferred for privacy concerns, since this data may reveal sensitive information like personal habits or health conditions [14].

In the literature, a few NeSy approaches for context-aware HAR have been proposed by applying knowledge-based reasoning on high-level context data [5]. For instance, according to common-sense knowledge, the activity *shopping*[1] is associated with a semantic location context corresponding to a shop or a commercial area. This intuitive association can be represented using a symbolic formalism and infused in the deep learning model, reducing the amount of labeled data required to learn it.

To the best of our knowledge, these NeSy approaches for context-aware HAR have never been evaluated on public in-the-wild datasets, but only on small datasets acquired in a scripted fashion. Moreover, the existing approaches in the literature involve symbolic reasoning during both training and classification. In real-world deployments, where the DL model is deployed on resource-constrained devices (e.g., mobile/wearable devices), the adoption of symbolic reasoning during classification is not desirable since it is computationally demanding [15].

Our research question may be formulated as follows: *how can we effectively infuse knowledge in a context-aware HAR deep learning model by adopting symbolic reasoning only during training?* To address this question, in this work, we propose a novel NeSy method for Context-Aware HAR. Our approach is based on a custom loss function for the DL model combining a standard classification loss with a novel semantic loss function. The semantic loss component uses symbolic reasoning to drive the DL model in classifying activities considering both raw sensor data patterns and high-level knowledge constraints. Indeed, after the training phase, the classifier internally encodes such constraints, that are exploited to classify activities at run-time without requiring symbolic reasoning.

We designed different semantic loss functions and our experimental evaluation on scripted and in-the-wild datasets identifies the best candidates. The results show that our method outperforms in terms of recognition rates a classic *DL* approach based on a standard classification loss. We also compared our approach with two alternative NeSy strategies that use symbolic reasoning during classification, showing that our semantic loss often reaches recognition rates close (and sometimes better) to such state-of-the-art methods, while avoiding the significant cost of performing symbolic reasoning during inference. Furthermore, our results demonstrate that our semantic loss surpasses existing neuro-symbolic approaches in addressing uncertainty, showing significantly

[1]Considered separately from online shopping.

greater robustness in the presence of noisy data. Hence, we believe that our semantic loss reaches a good trade-off between efficiency and recognition rate.

To summarize, our contributions are the following:

- We formalize the NeSy Context-Aware HAR research problem and reformulate existing solutions using our notation.
- We propose a novel NeSy solution for Context-Aware HAR, based on a semantic loss function that does not require symbolic reasoning after deployment of the HAR system.
- We performed an extensive evaluation on scripted and in-the-wild datasets, comparing our solution with two existing NeSy methods that require symbolic reasoning during classification. Our results show that, especially considering in-the-wild settings, our semantic loss reaches recognition rates that are often close (and sometimes better) than the ones of the other approaches. Moreover, our semantic loss is significantly more robust to noisy data with respect to the other approaches.

2 RELATED WORK

2.1 Sensor-Based Human Activity Recognition

Most of the works proposed in the literature for sensor-based HAR on mobile/wearable devices rely on supervised Deep Learning (DL) methods [18, 57]. The combination of inertial and high-level context data has the potential to significantly improve the recognition rate compared to considering only inertial sensors as proposed by the majority of the works [13, 47].

Despite their success, existing DL solutions require a large amount of labeled data during the learning process. Unfortunately, the annotation process is error-prone, expensive, time-consuming, and tedious, especially considering large amounts of data. Moreover, the inner mechanisms of deep learning classifiers are opaque, thus not allowing humans to understand the rationale behind each model's prediction.

To mitigate the data scarcity problem, the HAR research community investigated transfer learning, unsupervised learning, and semi-supervised learning approaches [18]. Transfer learning methods usually take advantage of models trained on a source domain with a significant amount of labeled data. Such pre-trained models are then fine-tuned in a target domain using a small amount of labeled samples [48, 52]. On the other hand, semi-supervised approaches for HAR rely on small labeled datasets to initialize the model, which is then incrementally updated by leveraging the unlabeled data stream [1]. Semi-supervised methods for HAR include self-learning, co-learning, active learning, and label propagation. Finally, self-supervised learning strategies leverage large amounts of unlabeled data to pre-train a model capable of generating reliable feature representation of sensor data [30, 32, 35]. The pre-trained model is fine-tuned only using a limited amount of labeled data. Neuro-Symbolic AI (NeSy) could be potentially coupled with such techniques to further improve the recognition rates in data scarcity scenarios. For example, it could be possible to integrate domain constraints during the fine-tuning phase of self-supervised learning to further minimize the amount of required labeled data.

2.2 Neuro-Symbolic AI in the General ML Community

NeSy approaches integrate neural and symbolic AI architectures to combine their abilities to perform learning and knowledge-based reasoning [36]. This combination improves the capability of the deep learning classifier to learn from smaller amounts of training data, to better generalize on unseen data, and to increase its interpretability [25]. While in this paper we focus more on data scarcity and generalization capabilities, Section 6.3 also discusses the interpretability aspects of our approach.

In the last few years, several works explored the infusion of domain knowledge into deep neural networks only during the training phase [21], especially in the Computer Vision domain. A pioneer work in this area proposed a custom loss function capable of capturing whether the outputs of a neural network are valid according to logical

constraints [61]. This method was mainly validated considering image classification tasks in semi-supervised scenarios, where, during training, the semantic loss encourages the model to confidently assign a class also to unlabeled data, resulting in improved decision boundaries compared to supervised methods relying only on labeled data. More specifically, in these experiments, the semantic loss forces the model to condense all the predicted probabilities on a single class, increasing the confidence in its decision. Further experiments have demonstrated that this approach enhances recognition rates for learning tasks with highly structured outputs [4], such as finding the shortest weighted path in a grid in video games or predicting users' rankings over a set of items (i.e., preference learning). In both cases, the semantic loss penalizes predictions that do not align with logic constraints (e.g., when the shortest path found by the model is not valid).

Considering hierarchical multi-label classification tasks, researchers explored loss functions encoding semantic connections between classes and their hierarchy, with the objective of making misclassification less severe [4, 24, 27]. For instance, thanks to this approach, an image labeled with the class *boy* is more likely to be misclassified with the class *man* rather than with unrelated classes like *bicycle*.

Neuro-symbolic training strategies have also been explored to increase models' explainability. For instance, in [23], the authors designed a training procedure that aligns the explanations of a convolutional neural network's predictions with the ones of human experts, encoded through knowledge graphs. The focus of this work was on monument facade image classification.

Another promising research direction for knowledge infusion in deep learning models only during training is Knowledge Distillation, a technique that relies on the teacher-student learning paradigm. For instance, the work in [34] proposed a method for the NLP domain. In particular, the proposed approach consists of training a deep neural network to mimic the outputs of a teacher model trained with a loss function also taking into account logical rules. Such rules are encoded through soft logic constraints, i.e., constraints associated with weights between 0 and 1. For instance, in sentiment analysis, a constraint may consider the conjunction word "but" to ensure that the predicted sentiment for the entire sentence aligns with the sentiment of the clause that follows "but". Finally, recent studies have also explored infusing symbolic knowledge into graph neural networks [22] for vertex enrichment in drug-discovery applications.

It is interesting to note that the logic constraints considered in the existing works are relatively simple. Indeed, as reported in [34], even simple but effective rules lead to a substantial improvement in terms of recognition rate. To the best of our knowledge, we are the first to propose a NeSy approach infusing knowledge only during training that is specific for Context-Aware Human Activity Recognition.

2.3 Neuro-Symbolic Approaches for Human Activity Recognition

While most NeSy methods have been proposed for computer vision and NLP applications, only a few NeSy methods exist for HAR. Considering HAR in smart-home environments, the domain knowledge can be used to derive an initial activity model that is subsequently adapted to the user's habits through data-driven strategies [53]. In [10], unsupervised methods are used to extract frequent patterns from unlabeled data. These patterns are then associated with the corresponding activities through domain knowledge. However, while these approaches are effective on smart-home environmental sensors, they cannot be applied to the inertial sensor data provided by mobile and wearable devices, which are the focus of this work.

Neuroplex [60] is a Knowledge Distillation approach injecting probabilistic symbolic knowledge (i.e., finite state machines and logical rules) into a neural network responsible for detecting complex nursing events (e.g., *patient cleaning*). Indeed, these events can be identified by reasoning on spatially- and temporally-dependent low-level events derived from inertial sensors data using data-driven models. For instance, the complex event *patient cleaning* can be derived when the sequence of detected low-level events is composed of *patient oral care*

followed by *diaper exchange*. Differently from Neuroplex, we focus on infusing knowledge for the recognition of low-level events by taking advantage of additional context information collected by mobile/wearable devices.

Considering context-aware HAR, [13] proposed to use domain knowledge on high-level context data to refine the predictions of an activity classifier trained on inertial sensors data. Finally, a recent work proposes the infusion of domain knowledge on context data into the deep learning classifier during both the training and inference phases [5]. However, these approaches rely on ontological reasoning during classification, which may be critical for the deployment of resource-constrained mobile devices. Indeed, experimental work in the literature shows that running symbolic reasoning based on Description Logics (like we do in our paper) on Android mobile devices is up to 150 times slower than on machines with higher resources (e.g., servers) on the considered datasets [15]. In the HAR domain, the work in [13] reports that context-aware ontological reasoning on mobile devices takes on average 1.3 seconds for each data sample. This is due to the computational complexity of symbolic reasoners. Considering deterministic reasoners based on OWL2 ontologies (that is the most common approach considered in the HAR field [13]), reasoning tasks have polynomial complexity [39]. Hence, even if theoretically considered as *tractable*, these methods do not scale linearly with the size of the knowledge model (e.g., number of activities, context situations, and constraints) and may not be adequate for resource-constrained devices. On the other hand, probabilistic symbolic reasoners like the ones based on log-linear description logics [40] and used in a few context-aware HAR works [6, 12], have even higher complexity. While there are approximated methods to reduce the complexity, probabilistic symbolic reasoning is still computationally demanding. Since mobile devices collect samples with high periodicity (e.g., a few seconds) requiring to be locally processed, such approaches may be inefficient in terms of computational resources and energy consumption. Hence, we believe that removing symbolic reasoning from mobile applications is beneficial.

More details about existing NeSy context-aware HAR approaches will be presented in Section 3.3.

To the best of our knowledge, this is the first work that proposes a NeSy solution for Context-Aware HAR with the following characteristics: a) it infuses common-sense knowledge directly inside the DNN activity model during training, and b) it does not require symbolic reasoning during classification.

2.4 Learning Using Privileged Information

Another learning paradigm closely related to our semantic loss is *LUPI: Learning Using Privileged Information*, which is based on leveraging additional information exclusively accessible during the training phase. Originally proposed for Support Vector Machines (SVMs) models [56], LUPI aims to enhance the classifier's decisions by utilizing privileged information during training (e.g., correcting the hyperplane in the case of SVM).

LUPI has also been applied in the context of sensor-based Human Activity Recognition (HAR) models considering multi-device settings [2, 29, 42, 58]. During training, the subject is equipped with numerous wearable/mobile devices positioned on various parts of the body. However, during deployment, the model is constrained to a limited number of devices. In such scenarios, the *privileged* information encompasses sensor data from body positions that are unavailable after deployment. Similarly, the work in [38] proposes a knowledge distillation approach for video-based activity recognition, considering as *privileged information* data from mobile/wearable devices.

We believe that our semantic loss can also be considered as a LUPI approach, where the privileged information is the common-sense knowledge (obtained from symbolic reasoning) about the degree of consistency of each activity with respect to the user's context.

3 PRELIMINARIES

In this section, we formalize context-aware HAR and we formulate the NeSy Context-Aware HAR problem. Moreover, we take advantage of this formalization to re-formulate existing NeSy strategies for Context-Aware HAR.

3.1 Context-Aware Human Activity Recognition

Let D_u be the dataset of raw sensor data collected from the mobile devices (e.g., smartphone, smartwatch) of a user u. Given a set of users $U = \{u_1, \ldots, u_n\}$, let $D^\star = \{D_{u_1}, \ldots, D_{u_n}\}$ be the set of datasets of all the users. Let $A = \{a_1, \ldots, a_k\}$ be the set of considered activities. The dataset $D^\star$ is associated with a set of annotations L that describes the activities performed by each user u. Each annotation $\lambda \in L$ is a tuple $\lambda = \langle u, a, t_s, t_e \rangle$ where a is a label identifying the activity actually performed by u during the time interval $[t_s, t_e]$. Each user dataset D_u is partitioned in a set of non-overlapping fixed-length windows $W_u = \{w_1, \ldots, w_q\}$ with each window including z seconds of consecutive raw sensor data of D_u.

In this work, we use the notion of *context* as a specific high-level situation that occurs in the environment surrounding and including the user while sensor data are being acquired (e.g., *it is raining*, *location is a park*, *current speed is high*). Let $C = \langle C_1, \ldots, C_p \rangle$ be a set of possible contexts that are meaningful for the application domain.

For each window w of raw data we identify two subsets w^R and w^C. The subset w^C includes raw sensor data that we consider useful to derive high-level contexts in C through reasoning and/or abstraction, while w^R includes raw data that we consider appropriate to be directly processed by a data-driven model (e.g., data from inertial sensors). Note that these subsets can have a non-empty intersection and their union is the whole w. The composition of w^R and w^C strictly depends on the target application, the available data, the knowledge model, and the available external services to obtain high-level context information.

Considering, for example, location data, it may be appropriate to exclude raw GPS coordinates from w^R and use it to obtain semantic location or other higher-level location information that can be more easily correlated with activities. On the other hand, leaving raw GPS data in w^R may not lead to a better model (it may be difficult to find correlations with activities and even when found, it may be difficult for the model to generalize).

Given w^C, let $ca(w^C)$ be a *Context Aggregation* function that derives all the contexts $C^w \subset C$ that are true during w based on w^C. This function can rely on simple rules, available services, or context-aware middlewares [31]. For instance, the geographical coordinates provided by the location service of the user's smartphone can be used to derive her semantic location (e.g., at home, in a public park) by querying a dedicated web service.

Definition 3.1 (Context-aware HAR). Given a dataset $D^\star$ and the annotations set L, the problem of *context-aware Human Activity Recognition* is to provide to an unseen tuple $\langle w^R, C^w \rangle$, derived from a sensor data window w from user u, the probability distribution $P = \langle p_1, \ldots, p_k \rangle$, where p_i is the probability that u performed the activity a_i in contexts C^w, with $\sum_{i=1}^{k} p_i = 1$.

3.2 Neuro-Symbolic Context-Aware HAR

The *context-aware HAR* problem could be tackled by using purely data-driven models where context data are simply used as input. However, a more effective approach combines data-driven models with a knowledge model K that, based on a set of contexts C, encodes relationships between the activities in A and the contexts in C. For instance, according to common-sense knowledge, the activity *cooking* is usually performed in a kitchen or, anyway, in a room equipped with a cooker, microwave, or oven. This relationship between the activity and the typical environment in which it is performed can be used in the HAR process, thus reducing the amount of labeled data required to learn it.

Note that K can be built in several different ways: by domain experts using common-sense knowledge on HAR, re-using existing knowledge bases (e.g., ontologies), or considering semi-automatic approaches in charge of extracting knowledge from external sources (e.g., text, images, and videos from the web). In any case, building a comprehensive and robust knowledge model is a challenging task. Even the knowledge of a domain expert is limited and is not guaranteed to capture all the possible context situations in which an activity can be performed [19].

Even though knowledge models cannot capture all the possible scenarios, our experiments will show their advantages in mitigating data scarcity when properly combined with data-driven methods. Indeed, in addition to the available training data, common-sense knowledge has the potential to capture constraints/patterns that are not learnable because of insufficient data. While there may be cases in which some rigid constraints would wrongly indicate the inconsistency between a context and an activity due to incompleteness, the knowledge model is supposed to model most of the usual context situations, and it can be refined and extended. Hence, we expect these cases to be rare. Also note that knowledge representation frameworks, like ontologies, have an open-world assumption. Hence, if reasoning cannot find an explicit inconsistency between a given context and an activity, their relationship is considered consistent.

Formally, given a knowledge model K and a set of contexts C^w, let $SR(K, C^w)$ be a Symbolic Reasoning function that outputs, for each activity a_i, a likelihood value $l(a_i)$ (a value between 0 and 1) of a_i being consistent with the observed context C^w according to the constraints in K. Note that the majority of symbolic representation and reasoning approaches, including most ontologies, are based on formal logics that do not support uncertainty. In these cases $SR()$ will associate the value 1 to each a_i that is consistent with the observed context C^w according to the constraints in K, and the value 0 otherwise.

Definition 3.2 (Neuro-Symbolic Context-Aware HAR model). A *Neuro-Symbolic Context-Aware Human Activity Recognition* model combines a deep learning model DNN and the symbolic reasoning function $SR()$ to solve the context-aware HAR problem.

This very general definition is intended to capture in a single category approaches that combine in different ways the DNN and the $SR()$ modules as we will describe in Sections 3.3 and 4. Figure 1 graphically illustrates the high-level architecture of NeSy Context-Aware HAR shared by these approaches.

3.3 Formalization of Existing Neuro-Symbolic Approaches

In this section, we re-formulate existing Neuro-Symbolic AI (NeSy) approaches with the notation introduced in sections 3.1 and 3.2 to compare them with our novel NeSy approach in an appropriate way. In particular, we consider two state-of-the-art approaches for NeSy HAR: *context refinement* and *symbolic features*.

3.3.1 Context refinement. The goal of the *context refinement* method [13] is to a posteriori review the DNN predictions using the HAR knowledge encoded in K. As shown in Figure 2, the DNN is trained with the cross-entropy loss function $\mathcal{L}_{cross}$, which penalizes misclassifications on the training data. During classification, the output of the $SR()$ function is used to refine the probability distribution derived by DNN on a specific input. Intuitively, the likelihood values obtained by $SR()$ are used to reduce the probability of those activities that are less likely to be the correct predictions considering the current user's context.

More formally, given a probability distribution $P = \langle p_1, \ldots, p_k \rangle$ emitted by DNN on a tuple $\langle w^R, C^w \rangle$, and the likelihoods values provided by $SR(K, C^w)$, for each candidate activity a_i with $i = 1, \ldots, k$ we compute $p_i * l(a_i)$ and then normalize in order to obtain a *knowledge-refined probability distribution*.

Note that when symbolic reasoning is deterministic, $l(a_i)$ is a binary value and the above operation is equivalent to excluding some of the activities from the candidates and normalizing.

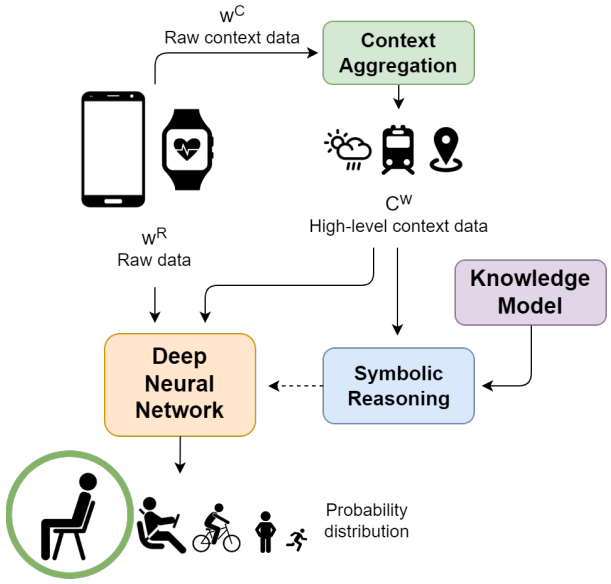

Fig. 1. The neuro-symbolic context-aware HAR approach

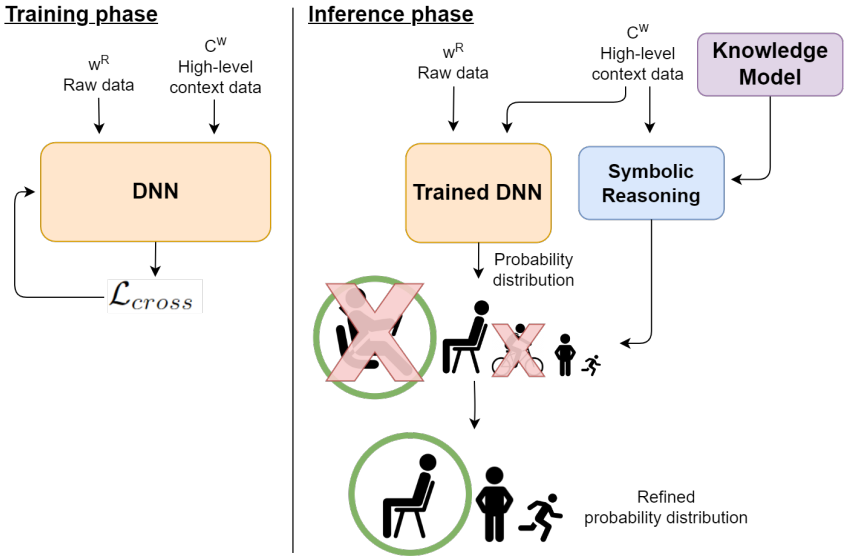

Fig. 2. The *context refinement* neuro-symbolic approach. In this example, two activities are excluded from the probability distribution since their likelihood, according to the Symbolic Reasoning module, is 0.

The objective of *context refinement* is to correct wrong decisions made by DNN, thus increasing its recognition rate. At the same time, it ensures that each classified activity is consistent with the surrounding context of the user. A drawback of this approach is that it requires the computation of $SR()$ during the inference phase of DNN,

i.e., each time an activity prediction is required. When the knowledge base is large in order to include a rich taxonomy of activities and to capture a wide range of context situations, even ontology languages with relatively low expressiveness (e.g., OWL2-DL) have polynomial time complexity for consistency computation. This has an impact on the cost of the run-time computation of $SR()$, especially when performed on mobile devices where energy consumption is a major issue.

Moreover, most ontology-based reasoning is not probabilistic and may encode rigid constraints about the relationships between contexts and activities, resulting in *context refinement* discarding activities that are occasionally performed in unusual context scenarios (e.g., the knowledge engineer may explicitly exclude that the activity *running* can be performed at *the mall*, as a semantic place).

In the following, we report a simplified running example of the *context-refinement* approach:

Example 3.1. *Consider an activity classifier trained offline in a supervised fashion by a service provider using a labeled dataset. After training, the classifier and a symbolic reasoner based on a standard ontology are deployed on Alice's smartphone to recognize her activities in real-time. Suppose that Alice is sitting, and the smartphone collects a window* $\langle w^R, C^w \rangle$ *of raw sensor data and high-level context data during the execution of this activity. Given this window, the classifier outputs the following probability distribution: Walking:* 50%*, Sitting:* 30%*, Standing:* 15%*, Running:* 5%*. We observe that the most likely activity is walking, which is not correct according to the ground truth. The high-level context* C^w *encodes the information that Alice's current speed is* 0*. By processing* C^w*, the deterministic symbolic reasoner infers that the likelihood of Walking and Running is* 0 *(since they can not be performed with null speed), while the likelihood of the other activities is* 1*. By multiplying each probability value with the corresponding likelihood and normalizing the resulting values, a new probability distribution is obtained: Sitting:* 67%*, Standing:* 33%*, Walking:* 0%*, Running:* 0%*. After refining the probability distribution, the most likely activity is sitting which corresponds with the actual activity performed by Alice.*

3.3.2 Knowledge infusion through symbolic features. The concept of introducing a knowledge infusion layer in a DNN was originally proposed in [50], and a first approach in this direction in the Context-Aware HAR domain, called *symbolic features*, was proposed in [5] some years later. The objective of the *symbolic features* is to directly incorporate the knowledge encoded in K into DNN, not only at the inference phase but also during the learning process. Hence, the *symbolic features* method allows the DNN also to learn the correlations between input data and context-consistent activities. As depicted in Figure 3, the information about the context-consistency of activities provided by $SR()$ is used to generate symbolic features that are infused within the hidden layers of DNN through a dedicated layer named *knowledge infusion* layer. More formally, given an input tuple $\langle w^R, C^w \rangle$, and the likelihoods values provided by $SR(K, C^w)$, the symbolic features consist of a vector f_s in which the i-th element is $l(a_i)$. Similarly to context refinement, please note that if symbolic reasoning is not probabilistic f_s is a binary vector.

Given the sequence of DNN's layers $\ell_1, \ldots, \ell_m$, and the symbolic features f_s generated through $SR()$, the *symbolic features* method adds to DNN a *knowledge infusion* layer ℓ_{ki}. This layer receives as input the symbolic features f_s and the features automatically extracted by a DNN's hidden layer ℓ_j with $1 < j < m$. Then ℓ_{ki} concatenates in the latent space the features received as input and generates a novel feature vector that is provided to the next layer ℓ_{j+1}. Also in this case, the DNN is trained through the cross-entropy loss function $\mathcal{L}_{cross}$.

This methodology is less rigid than *context refinement* in excluding some activities based on knowledge consistency since domain knowledge is infused into the data-driven model instead of just being used afterword, to modify the result of the neural network.

On the other hand, similarly to *context refinement*, the main problem of *symbolic features* is that they have to be computed even in the inference phase at each activity prediction, leading to computational cost and energy consumption when deployed on resource-constrained devices.

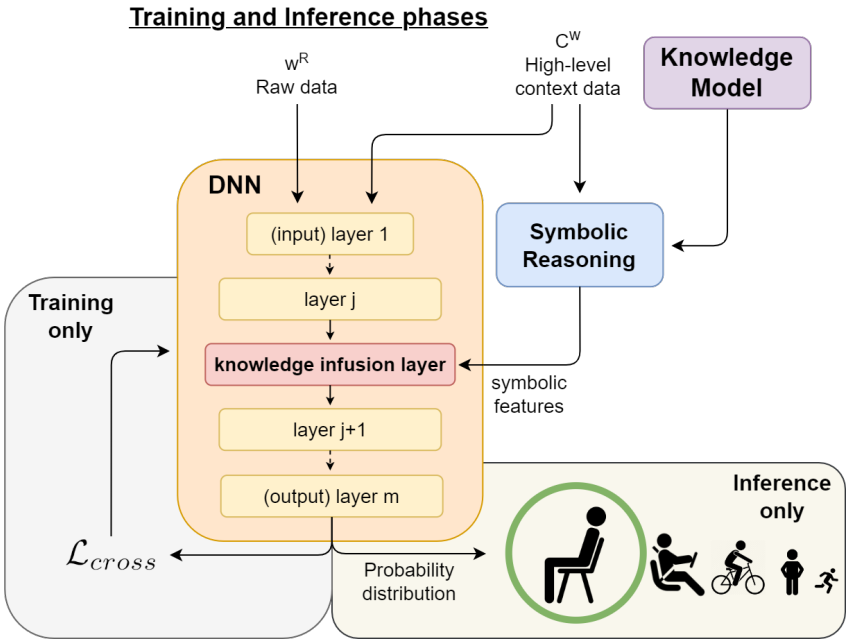

Fig. 3. The *symbolic features* neuro-symbolic approach

In the following, we report a simplified running example of the *symbolic features* approach:

Example 3.2. *A service provider trains, in a supervised way, an activity classifier using a labeled dataset and a symbolic reasoner based on a standard ontology. For each window* $\langle w^R, C^w \rangle$*, the symbolic reasoner analyzes* C^w *to obtain the likelihood values for each activity, that are used to generate symbolic features. For instance, when* C^w *includes home as semantic location, the symbolic feature corresponding to the driving activity is* 0*. The model is trained by providing windows of raw sensor data and high-level context data in the input layer, while symbolic features are given to the knowledge infusion layer. After training, the classifier and the reasoner are deployed on Alice's smartphone to recognize her activities in real-time. Suppose that Alice is sitting, and the smartphone collects a window* $\langle w^R, C^w \rangle$ *of raw sensor data and high-level context data during the execution of this activity. The high-level context* C^w *encodes the information that Alice's current speed is* 0*. By processing* C^w*, the symbolic reasoner generates a symbolic feature vector, where Walking and Running have value* 0 *(since they can not be performed with null speed), while the remaining activities have value* 1*. In order to perform classification, the window* $\langle w^R, C^w \rangle$ *is provided to the input layer, and the symbolic feature vector is provided to the knowledge infusion layer. Thanks to the information encoded in the symbolic features, the classifier will assign a lower probability value to Walking and Running, since it has learned during training that these activities are inconsistent according to the symbolic features.*

4 KNOWLEDGE INFUSION THROUGH SEMANTIC LOSS

In this section, we present our novel approach named *knowledge infusion through semantic loss* (or *semantic loss* for short) aimed to overcome the main drawbacks of the approaches presented in Section 3.3. Our method generates an activity classifier encoding knowledge-based constraints without requiring symbolic reasoning during the inference phase. Hence, a model based on *semantic loss* can be trained offline on a cloud-based server and then deployed on the users' mobile/wearable device to locally perform real-time activity recognition efficiently.

4.1 Methodology

In the following, we describe the mechanisms of our *semantic loss* approach. As depicted in Figure 4, the goal of

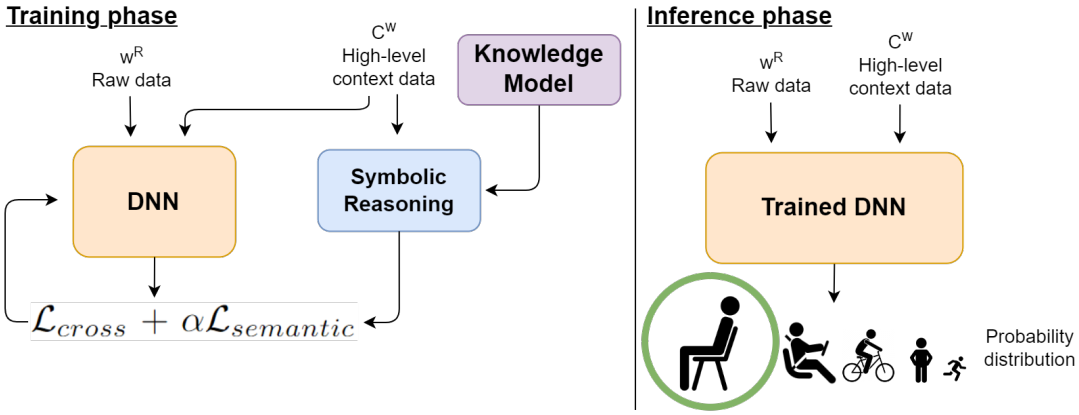

Fig. 4. Our neuro-symbolic approach based on semantic loss

semantic loss is to exploit the knowledge K to guide the learning process of DNN through a specifically designed loss function. As in the *symbolic features* method, DNN still learns the correlations between context-consistent activities and input data. At the same time, since no additional features are infused into DNN, the use of K and SR during classification is not required, thus solving one of the main limits of the existing solutions.

Specifically, the loss function $\mathcal{L} = \mathcal{L}_{cross} + \alpha\mathcal{L}_{semantic}$ that guides the training process of DNN is a combination of the cross-entropy loss function $\mathcal{L}_{cross}$ with a semantic loss function $\mathcal{L}_{semantic}$.

We consider the standard formula for the cross-entropy loss:

$$\mathcal{L}_{cross} = -\sum_{i=1}^{k} y_i \log(p_i) \tag{1}$$

where y_i is 1 only when a_i is the ground truth activity, while p_i is the probability of a_i obtained by the DNN.

Consistently with other works in the DL literature [17, 59], α is a trade-off parameter in charge of balancing the different loss terms. In particular, $\mathcal{L}_{semantic}$ determines the impact of the common-sense knowledge about context consistency on the DNN's output.

More formally, let $P = \langle p_1, \ldots, p_k \rangle$ be a probability distribution emitted by DNN on a tuple $\langle w^R, C^w \rangle$, and $l(a_i)$ be the likelihood value obtained by $SR(K, C^w)$ on the activity a_i. We denote with $\hat{p} \in P$ the maximum probability value of P, and with $\hat{a} \in A$ its corresponding activity.

In the following, we describe five alternative semantic loss functions we designed and tested for this work.

(1) The *AllConsistentActs (All)* semantic loss focuses on the whole probability distribution P. Intuitively, given P, this semantic loss has the objective of training the network to maximize the sum of the probability values in P that correspond to the context-consistent activities according to $SR()$ (i.e., the ones with likelihood greater than zero). Hence, we would expect that DNN learns to emit non-zero probabilities only for context-consistent activities during classification. Equation 2 formally defines the *All* semantic loss:

$$\mathcal{L}_{semanticAll}(P, SR) = 1 - \sum_{i} p_i \cdot l(a_i) \tag{2}$$

A potential drawback of this strategy is that, since it aggregates probability values with a sum, different combinations of these values may lead to the same penalty. Hence, the resulting penalties could be poorly informative for *DNN* to properly learn knowledge constraints. For this reason, the following alternative semantic losses only focus on the most likely activity $\hat{a}$.

(2) The *MinusProb-Prob (-PP)* semantic loss aims at associating low probability values with context-inconsistent activities and higher probability values with context-consistent activities. In particular, context-inconsistent predictions are penalized by their probability value. On the other hand, the penalty of context-consistent activities is inversely proportional to the probability $\hat{p}$ of the most likely activity according to the DNN, scaled by the likelihood $l(a_i)$ provided by *SR*. More formally,

$$\mathcal{L}_{semantic-PP}(P, SR) = \begin{cases} 1 - (\hat{p} \cdot l(\hat{a})) & \text{if } l(\hat{a}) > 0 \\ \hat{p} & \text{otherwise} \end{cases} \tag{3}$$

However, a potential drawback of this strategy is that penalty values for consistent activities with relatively low probability values are similar to penalty values for context-inconsistent activities with relatively high probability values.

(3) The goal of the *Zero-One (01)* semantic loss is to maximize the differences between penalties of context-consistent and context-inconsistent activities. Specifically,

$$\mathcal{L}_{semantic01}(SR) = \begin{cases} 0 & \text{if } l(\hat{a}) > 0 \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

The following strategies are refined versions of the *01* loss.

(4) The *MinusProb-One (-P1)* semantic loss aims at improving the confidence of *DNN* on context-consistent predictions. Indeed, while the penalty for context-inconsistent activities is fixed, the penalty for context-consistent activities is inversely proportional to the probability $\hat{p}$ of the most likely activity according to the DNN, scaled by the likelihood $l(a_i)$ provided by *SR*. Hence, context-consistent activities with low probability and/or likelihood values are penalized as well. More formally,

$$\mathcal{L}_{semantic-P1}(P, SR) = \begin{cases} 1 - (\hat{p} \cdot l(\hat{a})) & \text{if } l(\hat{a}) > 0 \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

(5) Finally, the idea of the *Zero-Prob (0P)* semantic loss is that context-consistent activities should not be penalized, while context-inconsistent activities should be penalized directly proportionally to their associated probability values. Hence, *DNN* should better learn that the higher the probability values of context-inconsistent activities, the higher the penalty. Therefore, *0P* aims at reducing the probability values on context-inconsistent activities. More formally,

$$\mathcal{L}_{semantic0P}(P, SR) = \begin{cases} 0 & \text{if } l(\hat{a}) > 0 \\ \hat{p} & \text{otherwise} \end{cases} \tag{6}$$

In the following, we report a simplified running example of our *semantic loss* approach:

Example 4.1. *A service provider trains, in a supervised way, an activity classifier using a labeled dataset and a symbolic reasoner based on a standard ontology. In particular, each window is fed-forward to the DNN. A loss function combining cross-entropy and AllConsistentActs is used to adjust the weights of the DNN. Suppose that, when feed-forwarding a window* $\langle w^R, C^w \rangle$*, the output probability distribution of the DNN is the following: Walking:* 50%*, Sitting:* 30%*, Standing:* 15%*, Running:* 5%*. Consider that the ground truth activity is Sitting and that the high-level context* C^w *encodes the information that the current speed is* 0*. By processing* C^w*, the symbolic reasoner outputs the*

likelihood values for each activity, where Walking and Running have value 0 *(since they can not be performed with null speed), while the remaining activities have value* 1*. Hence, by applying the formula in Equation 2, the value of the semantic loss is* $1 - (0.5 \cdot 0 + 0.3 \cdot 1 + 0.15 \cdot 1 + 0.05 \cdot 0) = 0.55$*. On the other hand, since the most likely activity does not correspond with the ground truth, the cross-entropy will generate* ≈ 1.73 *as a value. Supposing that* $\alpha = 5$*, the final value of the custom loss is* $1.73 + 5 \cdot 0.55 = 4.48$*, and it will be used to update the weights of the DNN. Hence, the knowledge constraints have a significant impact on determining how to update the weights of the DNN. After training, only the trained classifier is deployed on Alice's smartphone to recognize her activities in real-time. Suppose that Alice is sitting, and the smartphone collects a window* $\langle w^R, C^w \rangle$ *of raw sensor data and high-level context data during the execution of this activity. The high-level context* C^w *encodes the information that Alice's current speed is* 0*. By providing the window as input to the activity classifier, it will rely on the knowledge infused during training to assign a high probability to the sitting activity.*

5 EXPERIMENTAL EVALUATION

In this section, we describe the experimental evaluation that we carried out to assess the quality of our method based on semantic loss presented in Section 4, compared to the state-of-the-art NeSy approaches introduced in Section 3.3. First, we introduce the two datasets that we considered in this work. Then we describe our experimental setup: how we pre-processed the datasets, the models used and the evaluation methodology adopted. Finally, we present the results of our evaluation.

5.1 Datasets

The evaluation of context-aware HAR approaches requires datasets including both inertial sensor data and contextual information. However, there are a few publicly available datasets with such characteristics. Existing NeSy approaches for context-aware HAR have been evaluated only on scripted and non-public datasets [13]. In this work, we consider a scripted dataset that we collected in a parallel work and a publicly available in-the-wild dataset, both including sensor and context data.

5.1.1 DOMINO. *DOMINO* [7] is a HAR dataset we collected as parallel research in our research lab. *DOMINO* includes several context-dependent activities monitored through mobile devices that collected both inertial sensor data and high-level context data.

In particular, *DOMINO* includes data from 25 subjects wearing a smartwatch on their dominant hand's wrist and a smartphone in their pocket. Raw sensor data have been collected from the inertial sensors (accelerometer, gyroscope, and magnetometer) installed on both these mobile devices. At the same time, the dataset also includes high-level context data collected by combining public web services and the smartphone's built-in sensors. The measurements of the barometer and the GPS of the smartphone were discretized to provide information about the users' height and speed variations. Moreover, the dataset incorporates the output of the following web services: (1) *Google's Places API* provided the semantic places closest to the user; from this information, it was also derived the presence of the user in an indoor or an outdoor environment; (2) *OpenWeatherMap* provided current local weather conditions (e.g., sunny), while (3) *Transitland* provided transportation routes and stops close to the user; the combination of this information with location data was used to derive whether the user was following a public transportation route.

DOMINO was acquired in a scripted fashion: the volunteers were asked to perform a sequence of indoor/outdoor activities, but they were not told how to execute them. Also, the volunteers were monitored by the research staff during data acquisition. As a consequence, the variability of context situations is limited. Overall, *DOMINO* contains almost 9 hours of labeled data ($\approx$ 350 activities instances), including 14 different types of activities: *walking*, *running*, *standing*, *lying*, *sitting*, *stairs up*, *stairs down*, *elevator up*, *elevator down*, *cycling*, *moving by car*, *sitting on transport*, *standing on transport* and *brushing teeth*.

5.1.2 ExtraSensory. *ExtraSensory* [55] is a public dataset for context and activity recognition. It includes inertial and context data collected in the wild from mobile devices of up to 60 users. Inertial data were collected through each user's personal smartphone (including both iOS and Android devices) and from a smartwatch provided by the researchers. More specifically, the dataset includes raw data measured by the accelerometer, the gyroscope, and the magnetometer of the smartphone, and raw data collected by the accelerometer of the smartwatch. Besides providing raw sensor data, *ExtraSensory* also provides data as handcrafted feature vectors (138 features) extracted from the raw measurements collected through inertial and other smartphone sensors (e.g., microphone, luminosity sensor) in 20-second time windows.

Overall, *ExtraSensory* contains about $300k$ minutes of labeled data, including 51 different labels self-reported by the users and encoding both high-level context information (e.g., at home, with friends, phone in bag, phone is charging) and performed activities (e.g., sitting, bicycling).

Since it has been collected in the wild, different research groups in the HAR community used *ExtraSensory* to assess the generalization capabilities of activity recognition frameworks in real-world scenarios [20, 54]. Due to the complexity of the dataset, existing HAR methods evaluated on *ExtraSensory* achieved low recognition rates. For instance, by considering as input the raw inertial measurements provided by the accelerometer and the gyroscope of the smartphones, the CNN-based method proposed in [20] reached an average macro f1 score of ≈ 0.53, only considering 4 target activity classes: *idle* (*lying* or *sitting*), *walking*, *running*, and *cycling*. In another work, by considering the handcrafted features of *ExtraSensory*, an AdaBoost classifier reaches ≈ 0.63 of average macro f1 score on 5 target activities (i.e., *walking*, *standing*, *sitting*, *exercise*, and *sleeping*) [54]. Hence, this dataset represents a challenging benchmark.

5.2 Experimental Setup

In the following, we describe our experimental setup.

5.2.1 Data pre-processing. Consistently with existing works proposing NeSy approaches for Context-Aware HAR [13], for both datasets, we segmented sensor data into non-overlapping windows of $k = 4$ seconds. For each raw data window w, we considered in the subset w^R only the data from inertial sensors, while, considering our datasets, the rest of the data would be much more helpful in its aggregated high-level form (C^w).

In the following, we describe the specific pre-processing steps we adopted for each dataset.

DOMINO. Considering *DOMINO*, we planned to recognize all the 14 different available activities, by considering the raw inertial measurements collected by the accelerometer and the gyroscope of the smartphone and the smartwatch. Moreover, in our experiments, we considered 6 different context information types: the presence of the user in *indoor/outdoor* locations, her *semantic place* (e.g., home, office, gym, bar), her discretized *speed* (i.e., null, low, medium, high), her *proximity to public transportation routes*, her discretized *height variation* (i.e., negative, null, positive), and the *weather conditions* (e.g., sunny, rainy). Table 1 shows the number of samples involved during our experiments for each activity class of *DOMINO*.

ExtraSensory. Considering *ExtraSensory*, we planned to recognize 7 different activities: *bicycling*, *lying down*, *moving by car*, *on transport*, *sitting*, *standing*, and *walking*. Specifically, for the activity class *walking* we consider those samples labeled as *walking* and/or *strolling* in the original dataset. For *moving by car*, we consider samples labeled with *in a car*, *car driver*, and/or *car passenger*, even when coupled with the label *sitting*. Finally, we labeled as *on transport* those samples originally labeled with *sitting* or *standing* coupled with the label *on a bus*.

Before conducting our experiments, we performed some steps of data cleaning. First of all, we considered only those samples including inertial measurements recorded from the accelerometer and the gyroscope of the smartphone and from the accelerometer of the smartwatch. Indeed, for some users of *ExtraSensory*, gyroscope data from smartphones are not available. Moreover, not all of the dataset's users wore the smartwatch during

Table 1. Number of samples for each activity class in *DOMINO*

Activity	Number of samples
Brushing teeth	163
Cycling	323
Elevator down	171
Elevator up	110
Lying	387
Moving by car	188
Running	334
Sitting	1764
Sitting on transport	213
Stairs down	266
Stairs up	187
Standing	1875
Standing on transport	297
Walking	1378
Total	**7656**

data collection. Then, based on the available self-reported labels, we discarded the data collected while the smartphone's user was in a bag, or on a table. Indeed, we considered only phone positions that have been commonly considered in the literature (i.e., in the pocket and in hand). Finally, since the labels of *ExtraSensory* were self-reported by the users involved in the data collection, we discarded samples that we considered unreliable, due to the fact that they included self-reported labels not consistent with the recorded data. For instance, we discard segmentation windows including positive speed values but labeled with static physical activities like *lying*. As another example, we discarded those samples simultaneously labeled with *in a car* and *at home*. Table 2 shows the number of samples for each activity class of *ExtraSensory* after data cleaning. Note that, after our data cleaning process, we considered data overall from 31 subjects.

As inertial sensor data, we considered the raw data measured from the accelerometer and the gyroscope of the smartphone and from the accelerometer of the smartwatch.

Regarding context data, we considered the ones that can be easily derived from sensors of mobile/wearable devices. For instance, we considered as input context data the information about the user's semantic place (e.g., at the beach) since it could be derived by combining localization data and external web services, but not the position of the user's smartphone (i.e., in the pocket, in hand). In some cases, we discretized available information: for instance, the *speed* values observed thanks to the GPS were discretized into *null/low/medium/high speed*. Other high-level context information was obtained by directly considering available data, like *audio level*, *light level*, *screen brightness*, *battery plugged AC/USB*, *battery charging*, *on the phone*, *ringer mode normal/silent/vibrate*, and the time of the day (e.g., *Time 0-6*, *Time 18-24*). Moreover, we relied on the self-reported label *on a bus*, assuming that similar information could be derived by combining GPS data and web services like *Transitland*, as we did in *DOMINO*. Finally, we considered the semantic locations self-reported by the subjects (i.e., *home*, *workplace*, *school*,

Table 2. Number of samples for each activity class in *ExtraSensory*

Activity	Number of samples
Bicycling	2920
Lying down	3055
Moving by car	2150
On transport	610
Sitting	23905
Standing	14280
Walking	11230
Total	**58150**

gym, *restaurant*, *shopping*, *bar*, *beach*). As already mentioned, semantic location information can be derived, for instance, by combining location coordinates data with *Google's Places API*.

5.2.2 DNN's architecture. The *DNN* we used for our experiments receives as input three separate inputs for each segmentation window: a) the smartphone's inertial sensors data, b) the smartwatch's inertial sensors data, and c) the one-hot encoded high-level context data[2].

Similarly to existing works, we rely on convolutional neural networks to capture spatio-temporal dependencies of sensor data [28, 46, 62, 63]. Even though more sophisticated networks have been proposed in the literature, in this work we use a simple solution to focus on the contribution of knowledge. The exact structure of our own CNN model has been determined empirically. Specifically, inertial sensors' data from the smartphone are processed by three *convolutional layers* composed of 32, 64, and 96 filters with a kernel size equal to 24, 16, and 8, respectively. These layers are separated by *max pooling* layers with a pool size of 4. After the three *convolutional layers*, we add a *global max pooling* layer, followed by a *fully connected* layer that includes 128 neurons. The smartwatch inertial sensors' data are provided to another component of *DNN* that presents the same sequence of layers used to automatically extract features from the smartphone's inertial data. The only difference is that, in this case, the three *convolutional layers* present a kernel size of 16, 8, and 4, respectively. Finally, the high-level context data is provided to a single *fully connected* layer composed of 8 neurons. The features extracted by these three independent flows are then combined thanks to a *concatenation* layer, which is followed by a *dropout* layer with a dropout rate of 0.1, and a *fully connected* layer with 256 neurons, useful to extract meaningful correlations between the concatenated features. The last layer of the network is a *softmax* layer that is in charge of providing a probability distribution over the possible activities.

In our experiments, we use this DNN architecture in four different ways:

- As a purely data-driven *baseline*, without further modifications
- Enhanced with our *semantic loss* (see Section 4)
- Enhanced by combining in the *concatenation layer* the *symbolic features* and the features automatically extracted from input data (see Section 3.3.2)
- As the *DNN* module of the *context refinement* approach (see Section 3.3.1)

[2]Note that, we did not include raw context data as input since it is intuitively easier to learn correlations between activities and high-level context (e.g., semantic place) rather than between activities and raw context (e.g., geographical coordinates).

5.2.3 Knowledge model and symbolic reasoning. Ontologies are currently the most widely used formalism to represent and reason about common knowledge and context data[11]. Compared to simple rules, the ontology representation that we adopt has the advantage of enabling hierarchical and relational reasoning; for example, the relationship between a location context and an activity class is inherited by activities in a subclass (more specialized activities). This means that the ontology captures implicit rules and enables reasoning based on rule chaining.

There are standard tools, languages, and reasoners for representing knowledge with ontologies. We adopt these standards by using Protegè[3] as the tool to design and visualize the ontology, OWL as the ontology language, and Pellet [51] as the reasoner. OWL2-DL (a sublanguage of OWL) offers a clear semantic in terms of the underlying description logic (a subclass of first-order logic), and an automatic polynomial time decision procedure for consistency and other inferences. Ontologies adopt an *open world assumption*, hence if some relationship or fact cannot be derived as false it may be true. Note however that some strict constraints can be formulated, for example stating that the activity *sitting on transport* can only take place while the user is following a public transportation route. Hence, if location context data reveals that the subject is not following one of these routes, that activity is considered inconsistent. There are no fuzzy values for consistency in these formalisms.

For our main experiments we decided to adopt this well-established knowledge representation framework and, by using the above-mentioned tool, we extended the knowledge model K proposed in the paper where the NeSy *context refinement* method was introduced [13] encoding domain-based relationships between activities and contexts according to common-sense knowledge. Our extension is intended to better cover the taxonomy of activities and their relationship with context data for the datasets considered in our work. For example, context information about speed and movement provided in the ExtraSensory dataset required to revise the part of the ontology regarding this context.

We use the ontology *consistency checking* as the Symbolic Reasoning function $SR()$ defined in our formalization. In particular, for each activity, we evaluate if it is *consistent* considering the available context data. Context-consistent activities are associated with 1 as likelihood, while context-inconsistent activities are associated with 0.

Since our definition of symbolic reasoning on the knowledge model admits also fuzzy or probabilistic methods to evaluate context consistency, we include an experiment considering a probabilistic ontology. In particular, we slightly extended the knowledge model proposed in [12], which is the probabilistic extension of the model originally proposed in [13]. This model relies on a probabilistic ontology composed of *soft constraints* (i.e., rules associated with weight) and *hard constraints* (i.e., rigid rules that are always true). For instance, the soft constraint *running can be performed indoors* has a lower weight than the soft constraint *running can be performed outdoors*. An example of a hard constraint is *running implies a non-null speed*. In this case, the $SR()$ function is the ELOG reasoner [41], which is based on a log-linear probabilistic logic.

5.2.4 Cross-validation. We evaluated the approaches presented in Sections 3.3 and 4 by adopting the *leave-k-users-out* cross-validation technique. At each fold, k users are used to populate the test set, while the remaining users are used to populate training (90%) and validation (10%) sets. We also simulated several data scarcity scenarios by downsampling the available training data at each fold (e.g., 1%, 50%).

Considering the *DOMINO* dataset, we considered $k = 1$ (leave-one-user-out cross-validation). On the other hand, as also done by other works in the literature [20], for the *ExtraSensory* dataset we choose $k = 5$. At each iteration, we used the test set to evaluate the recognition rate of the different approaches in terms of the F1 score.

For the sake of robustness, we run each experiment 5 times, computing the average f1 score and the 95% confidence interval. Overall, the training process was based on a maximum of 200 epochs and a batch size of 32

[3] https://protege.stanford.edu/

Table 3. Comparison between the Semantic Loss types on the different datasets

	Dataset (training set percentage) DOMINO (100%)	ExtraSensory (10%)
Baseline	0.9024	0.5199
MinusProb-Prob (-PP)	0.9139 $\alpha = 5$	0.5402 $\alpha = 4$
Zero-One (01)	0.9042 $\alpha = 1$	0.5270 $\alpha = 7$
Zero-Prob (0P)	0.9162 $\alpha = 3$	0.5288 $\alpha = 9$
AllConsistentActs (ALL)	0.9094 $\alpha = 1$	**0.5872** $\alpha = 30$
MinusProb-One (-P1)	**0.9261** $\alpha = 7$	0.5298 $\alpha = 5$

samples. We considered an *early stopping* strategy, stopping the learning process when the loss computed on the validation set did not improve for 5 consecutive epochs.

5.3 Results

In the following, we show how our *semantic loss* approach outperforms a purely data-driven classifier in terms of recognition rate both in scripted and in-the-wild scenarios. We also compare our method with the Neuro-Symbolic AI (NeSy) approaches presented in Section 3.3. Our main results consider the OWL ontology as the knowledge model since it is a widely used knowledge and context representation framework. The results using an experimental probabilistic knowledge model are presented in Section 5.3.5.

Although our method does not include symbolic reasoning during classification, it often reaches recognition rates that are close (and sometimes better) to the ones of the other approaches, especially considering the more realistic scenarios of *ExtraSensory*.

5.3.1 Semantic loss types comparison. Table 3 compares the recognition rates (in terms of overall f1 score) of the five semantic loss functions presented in Section 4 on *DOMINO* and *ExtraSensory*. To better emphasize the differences in the recognition rates, on *ExtraSensory* we decided to show the results obtained by considering a data scarcity scenario in which only 10% of the training data are available. Indeed, the number of training samples in *DOMINO* is nearly equal to the number contained in only 10% of the training samples in *ExtraSensory*. Moreover, Table 3 also includes the best α value for each semantic loss type[4] and the results obtained by the purely data-driven *baseline* that is based on a standard classification loss.

Each semantic loss strategy leads to an improvement in the recognition rates compared to the *baseline*, with *-P1* achieving the best improvements on *DOMINO* ($\approx$ +2.5%) and *All* on *ExtraSensory* ($\approx$ +6.5%). Before running the experiments, we expected similar results for *01*, *-P1*, and *0P* since all these strategies aim at maximizing the distance in penalties between consistent and not-consistent activities. While this insight is confirmed on *ExtraSensory*, on *DOMINO* the *01* approach proved to be not very effective in improving the recognition rate. On this dataset, we observed that, besides increasing the difference between the penalties applied to context-consistent and

[4] α values have been determined empirically by performing a grid search in the range $[1, 35]$

Table 4. *DOMINO*: Results in terms of macro f1 score and 95% confidence interval

	Training set percentage									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Baseline	0.5946 (±0.008)	0.7529 (±0.010)	0.8268 (±0.006)	0.8556 (±0.011)	0.8835 (±0.011)	0.8917 (±0.010)	0.8915 (±0.006)	0.9007 (±0.007)	0.8965 (±0.002)	0.9024
Semantic loss -P1	0.6144 (±0.024) $\alpha = 7$	0.7712 (±0.004) $\alpha = 8$	0.8469 (±0.002) $\alpha = 9$	0.8679 (±0.010) $\alpha = 7$	0.8892 (±0.006) $\alpha = 7$	0.8889 (±0.007) $\alpha = 7$	0.9049 (±0.006) $\alpha = 8$	0.8997 (±0.003) $\alpha = 7$	0.9021 (±0.008) $\alpha = 6$	0.9261 $\alpha = 7$
Symbolic features	0.7268 (±0.008)	0.8590 (±0.012)	**0.9107** (±0.011)	0.9152 (±0.009)	0.9198 (±0.011)	0.9237 (±0.009)	**0.9265** (±0.004)	0.9254 (±0.008)	**0.9277** (±0.007)	**0.9408**
Context refinement	**0.8192** (±0.009)	**0.8811** (±0.007)	0.9078 (±0.009)	**0.9178** (±0.005)	**0.9281** (±0.012)	**0.9305** (±0.006)	0.9225 (±0.004)	**0.9274** (±0.005)	0.9232 (±0.002)	0.9221

Table 5. *ExtraSensory*: Results in terms of macro f1 score and 95% confidence interval

	Training set percentage								
	1%	2.5%	5%	7.5%	10%	25%	50%	75%	100%
Baseline	0.3127 (±0.023)	0.4279 (±0.008)	0.4867 (±0.013)	0.5167 (±0.016)	0.5199 (±0.011)	0.5842 (±0.016)	0.6096 (±0.007)	0.5813 (±0.032)	0.6053
Semantic loss All	0.3366 (±0.027) $\alpha = 29$	0.4895 (±0.010) $\alpha = 30$	0.5256 (±0.016) $\alpha = 26$	0.5650 (±0.016) $\alpha = 26$	0.5872 (±0.014) $\alpha = 30$	0.6331 (±0.013) $\alpha = 29$	**0.6323** (±0.011) $\alpha = 18$	0.6131 (±0.011) $\alpha = 16$	**0.6244** $\alpha = 17$
Symbolic features	0.3418 (±0.010)	0.4720 (±0.016)	0.5877 (±0.025)	0.6359 (±0.008)	0.6534 (±0.012)	0.6404 (±0.010)	0.6216 (±0.007)	**0.6268** (±0.007)	0.6205
Context refinement	**0.6324** (±0.014)	**0.6540** (±0.003)	**0.6797** (±0.003)	**0.6656** (±0.004)	**0.6622** (±0.007)	**0.6483** (±0.010)	0.6258 (±0.007)	0.6067 (±0.023)	0.6190

context-inconsistent predictions, it is also crucial to consider the probability values emitted by *DNN*, especially in the case of a context-consistent prediction, as proved by the *-P1* semantic loss. Finally, the improvement of the *All* strategy on *DOMINO* is limited, probably because learning knowledge constraints considering the whole probability distribution is unnecessarily too hard on simple scripted scenarios. On the other hand, this strategy significantly outperforms the others in the more realistic settings included in *ExtraSensory*.

5.3.2 Comparison with other approaches. Tables 4 and 5 compare our best *semantic loss* method (i.e., *-P1* on *DOMINO* and *All* on *ExtraSensory*) with: i) the purely data-driven *baseline*, ii) the *symbolic features* strategy, and iii) the *context refinement* strategy. More specifically, we considered different percentages of available training data for each dataset, thus comparing the approaches in different data scarcity scenarios. Note that, during the experimental evaluation, we empirically determined the optimal α values of the *semantic loss* for each training set percentage.

Overall, on each dataset, the NeSy approaches outperform the *baseline*, considering almost all the data scarcity scenarios. This result suggests that traditional symbolic AI approaches have the potential to enhance the predicting capabilities of purely data-driven deep learning models.

Focusing on the scripted scenarios of *DOMINO* (Table 4), the improvement of the *semantic loss* is lower than the other approaches, especially considering data scarcity scenarios. For instance, considering 10% of training data,

semantic loss leads to a recognition rate boost over the *baseline* of ≈ +2% on *DOMINO*. On the other hand, *symbolic features* and *context refinement* lead to improvements of ≈ +13% and ≈ +22%, respectively. These performance differences become progressively smaller while increasing training data availability. Indeed, when all the available training data are considered, both *semantic loss* and *context refinement* outperform the *baseline* by ≈ +2%, while *symbolic features* leads to an improvement of ≈ +4%.

On the other hand, different insights are observed when focusing on the realistic scenarios of *ExtraSensory* (Table 5). Indeed, on this dataset, the differences between the three NeSy approaches are smaller. For instance, considering 10% of training data, the recognition rate improvements of *semantic loss*, *symbolic features*, and *context refinement* are ≈ +7%, ≈ +13%, and ≈ +14%, respectively.

In general, the *semantic loss* achieves improvements that lie between ≈ +2% and ≈ +7%, sometimes outperforming the recognition rates of the other NeSy techniques. Indeed, the *semantic loss* outperforms *context refinement* from 50% to 100% of training data, and it also outperforms *symbolic features* on 100% of training data. Overall, *context refinement* is more effective than methods based on knowledge infusion (i.e., *symbolic features* and *semantic loss*) when the availability of labeled data is drastically low. However, when slightly more training data are available (e.g., 25% on *ExtraSensory*), all the NeSy approaches lead to similar improvements.

Our results indicate that our *semantic loss* is effective in capturing relationships between high-level context data and activities with respect to learning them directly from the training set by using purely data-driven models. This is especially true on the *ExtraSensory* dataset, where the improvement of *semantic loss* compared to the *baseline* is larger. Indeed, *DOMINO* covers a significantly lower variability of context situations compared to *ExtraSensory*, and the relationships between context and activities can be captured more easily by the *DNN*. On the other hand, the in-the-wild nature of *ExtraSensory* implies a significantly more complex learning task that can be partially lightened by knowledge reasoning.

Note that, due to the complexity of the dataset, we achieved relatively low recognition rates on *ExtraSensory* (e.g., the max F1 score is ≈ 0.68). As described in Section 5.1, our results are in line with other works on the same dataset [20, 54].

Since the computational complexity of symbolic reasoning is not adequate for real-world deployment on resource-constrained devices like smartphones and smartwatches, the choice of the optimal solution should consider a trade-off between recognition rate and efficiency. We believe that our *semantic loss* method is a much more promising approach since it still improves the *baseline* while not requiring symbolic reasoning at all after training.

5.3.3 Activity-level results. Figure 5 compares the confusion matrices obtained by the three considered NeSy approaches and the baseline on *ExtraSensory*, considering the data scarcity scenario where only 10% of training data are available[5]. From these confusion matrices, it emerges the contribution of domain knowledge in improving the recognition of different activities. For instance, the *baseline* often confuses *on transport* with *moving by car* due to their similar patterns (in terms of inertial measurements and speed), even though context information (e.g., whether the user is following a public transportation route) should help in distinguishing them.

Indeed, even though high-level context data are provided as input to the *baseline*, it is not feasible to learn from the training set all the possible correlations between all the possible context conditions and the performed activities. Hence, enhancing the data-driven model with symbolic AI approaches based on domain knowledge has a key role in enhancing the capabilities of the deep learning model and mitigating this problem, thus significantly reducing the confusion between these two activities.

Finally, we observed that each approach performed poorly on the *lying down* activity, which was often confused with *sitting*. We noticed that it is consistent with other papers in the literature that used the ExtraSensory dataset [20]. This is likely due to the fact that both *lying down* and *sitting* are static activities with similar sensor

[5]We show a representative run among the 5 repetitions of the experiment.

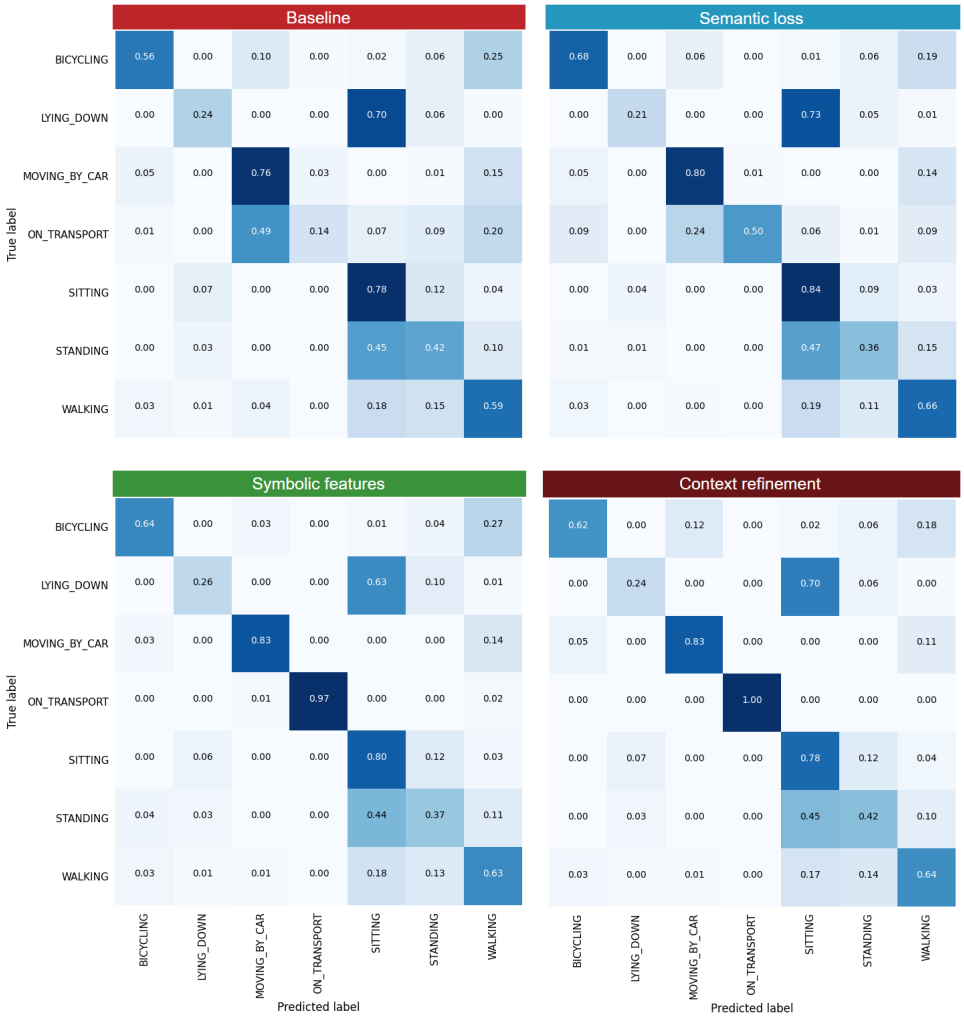

Fig. 5. Comparison between the confusion matrices of the *baseline* and the three considered Neuro-Symbolic AI approaches trained with 10% of training data on the *ExtraSensory* dataset

patterns, hence the exact posture is difficult to recognize. Moreover *sitting* is over-represented in the dataset, while *lying down* is underrepresented. For these reasons, the model often outputs *sitting* even if the correct activity is *lying down*.

5.3.4 Robustness to noise. In order to show that our semantic loss is robust to uncertainty even considering a standard ontology, we performed another set of experiments by introducing noise in the test data. In particular, we performed different experiments considering 5%, 10%, and 15% of noisy data in the test set. More specifically, for each perturbed data sample, we modified the semantic location context with another one (plausibly not too distant from the real one) that the knowledge model considers inconsistent with the ground truth activity. This perturbation simulates noise in GPS data acquired from mobile devices, often impacting the actual semantic

location where the user is located. For instance, a subject at home may be wrongly located at a coffee shop that is in a nearby building.

Table 6 shows the results of this experiment. We observe that noise has the most negative impact on context

Table 6. Average results with 5 different runs in terms of macro f1 score, considering 10% of training data and different percentages of dirty samples in the test set

	Original test set	**5% of dirty test set (delta)**	**10% of dirty test set (delta)**	**15% of dirty test set (delta)**
Baseline	0.5199	0.5089 (- 1.10%)	0.4983 (- 2.16%)	0.4954 (- 2.45%)
Semantic loss	0.5872	**0.5566** (- 3.06%)	**0.5229** (- 6.43%)	**0.5196** (- 6.76%)
Symbolic features	0.6534	0.5498 (- 10.36%)	0.5200 (- 13.34%)	0.5043 (- 14.91%)
Context refinement	**0.6622**	0.5430 (- 11.92%)	0.5057 (- 15.65%)	0.4801 (- 18.21%)

refinement, thus confirming that it is the most rigid approach. Indeed, by discarding activities that are not consistent with the current context, this approach is the one suffering more from noisy context data. Surprisingly, the symbolic features approach also significantly degrades the recognition rate. This is probably due to the fact that this method heavily relies on the infused symbolic features during classification, which leads to misleading features when activities are wrongly considered inconsistent. On the other hand, these results show that our semantic loss is significantly more robust to noise compared to the other NeSy methods. Finally, we observed that the baseline method is the approach most robust to uncertainty, due to better generalization capabilities. Nonetheless, our semantic loss still outperforms the baseline in each considered setting, hence confirming the advantage of infusing knowledge in deep learning models.

5.3.5 Results with a probabilistic knowledge base. Table 7 summarizes the results that we obtained on both datasets by using a probabilistic knowledge model slightly adapted from the one proposed in [12].

Table 7. Average results with 5 different runs in terms of macro f1 score, considering a data scarcity scenario simulated by using 10% of training data and the probabilistic version of each method

	DOMINO		**ExtraSensory**	
	Deterministic	Probabilistic	Deterministic	Probabilistic
Baseline	0.5946	0.5946	0.5199	0.5199
Semantic loss	0.6144	**0.6372**	0.5872	**0.6013**
Symbolic features	0.7268	**0.7365**	**0.6534**	0.6408
Context refinement	0.8192	**0.8399**	0.6622	**0.6793**

For the sake of simplicity, we show the results considering the data scarcity scenario where only 10% of labeled data are available. Our results indicate that, in general, introducing fuzziness only slightly improves the recognition rate obtained by the approach based on a standard ontology. The maximum improvement is $\approx$ 2% on the DOMINO dataset. The only case where the probabilistic approach is slightly worse than the deterministic one

is by using symbolic features on the ExtraSensory dataset. This is likely due to the fact that, on this dataset, it often happens that the ground truth activity is not always the one corresponding to the symbolic feature with the highest likelihood. This aspect significantly complicates the learning process since this method heavily relies on symbolic features, as already discussed for the results presented in Section 5.3.4. On the other hand, considering the deterministic case, consistent activities are always associated with a symbolic feature with a value of 1, thus avoiding this problem.

We believe that the small improvement in the recognition rate does not justify the effort of designing and managing probabilistic ontologies. Indeed, such models require significant effort in deciding the weights that should be associated with soft constraints, that should capture general aspects of activities execution. Hence, we believe that relying on standard ontologies to capture the most common situations is an appropriate choice when coupled with the proposed semantic loss method since it reduces the modeling effort while maintaining good accuracy.

6 DISCUSSION

6.1 Strengths and Weaknesses of Neuro-Symbolic Approaches

In the following, we discuss the strengths and weaknesses of the Neuro-Symbolic AI (NeSy) approaches presented in Sections 3.3 and 4. This information is also summarized in Table 8.

Table 8. Comparison of pros and cons of NeSy methods

	context refinement	symbolic features	semantic loss
improving recognition rate	x	x	x
mitigating data scarcity	x	x	x
retraining not required when knowledge is revised	x		
handling data uncertainty			x
symbolic reasoning not required after deployment			x

Compared to other methods, *context refinement* often reaches the highest recognition rates, especially when the amount of available training data is limited. However, this method may be less effective when based on an imperfect knowledge model. Indeed, *context refinement* always discards activities only relying on the user's surrounding context considering rigid constraints. For instance, a user could ride a bicycle even in unusual context scenarios (e.g., on a pedestrian-only road). Hence, when the knowledge model does not cover all the possible contexts in which an activity can be performed, combining the information from inertial data with knowledge would be more convenient in refining the probability distribution. Moreover, our results show that context refinement performs poorly in the presence of uncertainty in context data.

While the *symbolic features* method is less accurate than *context refinement*, it is slightly better in capturing the intrinsic uncertainty in sensor data by learning correlations between features and contexts, as opposed to the latter's direct application of rigid rules.

However, both approaches require the use of the symbolic reasoning module at each activity prediction, making them less suitable for deployment on mobile devices. Moreover, both approaches are significantly less effective than semantic loss in the presence of uncertainty in context data.

On the other hand, our *semantic loss* can be trained offline on a server with high computational capabilities and then deployed and used on a mobile device without the need for computationally expensive symbolic reasoning

tasks. Indeed, *semantic loss* is still able to significantly improve the recognition rate. Additionally, it is the most robust NeSy approach when context data is noisy.

6.2 Revising/Updating the Knowledge Model

In this work, we assumed that the knowledge model is static and never updated. However, this is not necessarily true in real-world settings. Indeed, we expect that the model can be *extended* by including new knowledge and/or *revised*.

If the knowledge is *extended* by including new activities or new context sources, all the NeSy models have to be modified to accommodate for new inputs and/or new output classes. New representative training data are also required. On the other hand, the knowledge can be *revised* to refine existing constraints between contexts and activities. For instance, domain experts may realize that the existing constraints are not adequate and should be improved. In this scenario, an advantage of context-refinement is that it does not require retraining the DNN, since symbolic reasoning is applied only during classification. However, re-training is required for the approaches based on knowledge infusion.

In our scenario, the model is pre-trained offline by a service provider with storage and computational capabilities and then deployed on mobile devices for inference. Hence, we believe that in this scenario, the service provider could easily re-train the model from scratch by taking into account the new knowledge model and possibly new representative data points.

When this is not possible or convenient, we believe that continual learning approaches (e.g., based on the teacher-student paradigm) could be adopted to incrementally train the underlying deep learning model to retain previous knowledge and learn new constraints, without the need for re-training from scratch. We believe that existing continual learning approaches could be effective when the knowledge model is *extended*, while it is more challenging when it is *revised* since incremental learning should allow the model to select which constraints to retain and which to update.

A more in-depth investigation on how to incrementally train neuro-symbolic approaches upon changes in the knowledge model is the subject of future work.

6.3 Interpretability

In the literature, Neuro-Symbolic AI methods are well-known for improving the interpretability of deep learning models [25]. Indeed, the decisions of a NeSy model are driven by infused knowledge. Hence, the knowledge model itself can be used to interpret the output of the classifier. Moreover, eXplainable AI methods (XAI) such as model induction (e.g., LIME [44]) or the saliency-based ones (e.g., GradCAM [49]) can be used to further inspect how the deep learning model reaches each decision.

In this paper, the knowledge infused into the model is about the relationships between high-level context data and activities. To better inspect the interpretability aspects of our model, we applied an XAI model induction approach named RISE [43] to visualize the importance of high-level context features on the supervised baseline (i.e., without knowledge infusion) and on our semantic loss model[6]. As an example, Figure 6 shows the average importance of high-level context features on the *ExtraSensory* dataset for the activity *on transport* on the baseline model. Figure 7 shows the same result for the semantic loss model.

We observe that the baseline model considers important many features that are not directly related to the activity, like *screen brightness*. On the other hand, the semantic loss model, consistently with the infused knowledge, considers particularly important only the context *on a bus*. Taking into account our results in Figure 5, it is clear

[6]For this evaluation, we randomly split the dataset into 70% for train, 10% for validation, and 20% for test; the models were trained on the train set and feature importance was computed on the predictions made on the test set.

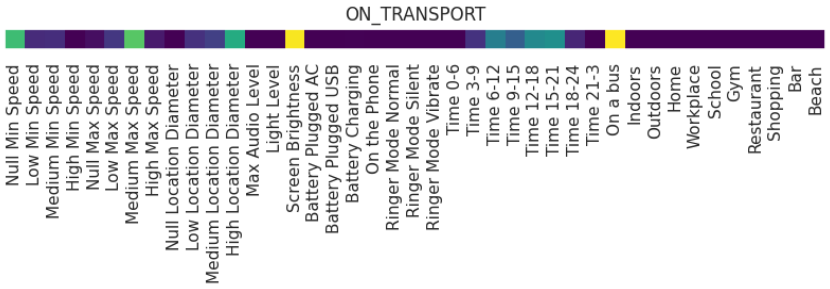

Fig. 6. Average feature importance for ON TRANSPORT obtained using XAI methods on the **baseline** model. The brighter the color, the more important the corresponding feature was for classification.

Fig. 7. Average feature importance for ON TRANSPORT obtained using XAI methods on the **semantic loss** model. The brighter the color, the more important the corresponding feature was for classification.

that this improvement led the classifier to achieve better results since it focuses on context features that are actually relevant considering the knowledge model.

Fig. 8. Example explanation for a WALKING sample based on the x-axis measurements from the smartwatch's accelerometer. The brightness of the color indicates the level of importance of each measurement for classification.

However, in this work, the classifier's decision is not based only on context data, but also on inertial sensor data that are inherently challenging to explain. While it is possible to highlight the portion of the signal that was important for the classifier (e.g., see Figure 8), this is difficult for humans to interpret and our knowledge model does not affect the interpretability of such signals. Therefore, our knowledge infusion approach leads to a deep learning model that is *partially* interpretable.

7 CONCLUSION AND FUTURE WORK

In this work, we presented a novel Neuro-Symbolic AI approach for context-aware HAR based on a combination of a classical loss function with a *semantic loss*. Our method infuses domain knowledge inside a deep learning classifier, improving its recognition rate. Compared to existing neuro-symbolic approaches, our method avoids symbolic reasoning during classification, thus making the model deployment feasible even on devices with limited computational resources. The advantage of our approach is particularly evident in realistic in-the-wild settings. Moreover, with respect to existing NeSy approaches for Context-Aware HAR, our semantic loss is also promising in coping with uncertainty in context data.

Besides the research directions previously mentioned in Section 6, in the following we discuss other plans for future work. First, context-aware HAR requires continuously obtaining context data. However, this may be computationally intensive since it may involve costly operations on mobile/wearable devices (e.g., continuously calling web services). Since high-level contexts may not change so rapidly, we will design strategies to obtain new information periodically (e.g., with a low periodicity, when GPS data exhibits significant changes, etc.). Thanks to these strategies, it could also be possible to run our method when mobile devices are not connected to the internet for short periods.

Moreover, we will evaluate how considering knowledge models that encode different levels of detail affects the performance of the NeSy approaches we compared in this work. Indeed, we expect to observe different results when considering a knowledge model that defines only usual contexts in which activities can take place, compared to a knowledge model that instead considers both usual and unusual scenarios.

We also want to explore other strategies to infuse knowledge inside deep learning models. For instance, the symbolic reasoning function may be approximated by a dedicated deep learning model (e.g., through a Graph Neural Network that learns domain constraints from a knowledge graph). This model could significantly reduce symbolic reasoning time, hence making the *context refinement* and *symbolic features* methods more practical in real-world deployments.

Another interesting line of research is to explore whether our approach can be adopted in pervasive computing domains different from HAR, where context information may have a major role (e.g., sensor-based healthcare systems, emotion recognition, anomaly detection).

Finally, we want to study how to quantitatively evaluate the intrinsic interpretability of the DNN components of NeSy approaches.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for the valuable feedback which significantly contributed to improving this work.

Part of this research was supported by the MUSA – Multilayered Urban Sustainability Action – project, funded by the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strengthening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D".

REFERENCES

[1] Zahraa S Abdallah, Mohamed Medhat Gaber, Bala Srinivasan, and Shonali Krishnaswamy. 2018. Activity recognition with evolving data streams: A review. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–36.

[2] Kohei Adachi, Paula Lago, Yuichi Hattori, and Sozo Inoue. 2022. Using LUPI to Improve Complex Activity Recognition. In *Sensor-and Video-Based Activity and Behavior Computing: Proceedings of 3rd International Conference on Activity and Behavior Computing (ABC 2021)*. Springer, 39–55.

[3] Preeti Agarwal and Mansaf Alam. 2020. A lightweight deep learning model for human activity recognition on edge devices. *Procedia Computer Science* 167 (2020), 2364–2373.

[4] Kareem Ahmed, Stefano Teso, Kai-Wei Chang, Guy Van den Broeck, and Antonio Vergari. 2022. Semantic probabilistic layers for neuro-symbolic learning. *Advances in Neural Information Processing Systems* 35 (2022), 29944–29959.

[5] Luca Arrotta, Gabriele Civitarese, and Claudio Bettini. 2022. Knowledge Infusion for Context-Aware Sensor-Based Human Activity Recognition. In *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 1–8.

[6] Luca Arrotta, Gabriele Civitarese, and Claudio Bettini. 2023. Probabilistic knowledge infusion through symbolic features for context-aware activity recognition. *Pervasive and Mobile Computing* 91 (2023), 101780.

[7] Luca Arrotta, Gabriele Civitarese, Riccardo Presotto, and Claudio Bettini. 2023. DOMINO: A Dataset for Context-Aware Human Activity Recognition using Mobile Devices. In *2023 24th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 346–351.

[8] Yusra Asim, Muhammad Awais Azam, Muhammad Ehatisham-ul Haq, Usman Naeem, and Asra Khalid. 2020. Context-aware human activity recognition (CAHAR) in-the-Wild using smartphone accelerometer. *IEEE Sensors Journal* 20, 8 (2020), 4361–4371.

[9] Martin Atzmueller, Naveed Hayat, Matthias Trojahn, and Dennis Kroll. 2018. Explicative human activity recognition using adaptive association rule-based classification. In *2018 IEEE International Conference on Future IoT Technologies (Future IoT)*. IEEE, 1–6.

[10] Gorka Azkune and Aitor Almeida. 2018. A scalable hybrid activity recognition approach for intelligent environments. *IEEE Access* 6 (2018), 41745–41759.

[11] Claudio Bettini, Oliver Brdiczka, Karen Henricksen, Jadwiga Indulska, Daniela Nicklas, Anand Ranganathan, and Daniele Riboni. 2010. A survey of context modelling and reasoning techniques. *Pervasive and mobile computing* 6, 2 (2010), 161–180.

[12] Claudio Bettini, Gabriele Civitarese, Davide Giancane, and Riccardo Presotto. 2020. Procaviar: Hybrid data-driven and probabilistic knowledge-based activity recognition. *IEEE Access* 8 (2020), 146876–146886.

[13] Claudio Bettini, Gabriele Civitarese, and Riccardo Presotto. 2020. Caviar: Context-driven active and incremental activity recognition. *Knowledge-Based Systems* 196 (2020), 105816.

[14] Claudio Bettini and Daniele Riboni. 2015. Privacy protection in pervasive systems: State of the art and technical challenges. *Pervasive and Mobile Computing* 17 (2015), 159–174.

[15] Carlos Bobed, Roberto Yus, Fernando Bobillo, and Eduardo Mena. 2015. Semantic reasoning on mobile devices: Do androids dream of efficient reasoners? *Journal of Web Semantics* 35 (2015), 167–183.

[16] Liang Cao, Yufeng Wang, Bo Zhang, Qun Jin, and Athanasios V Vasilakos. 2018. GCHAR: An efficient Group-based Context—Aware human activity recognition on smartphone. *J. Parallel and Distrib. Comput.* 118 (2018), 67–80.

[17] Chaofan Chen, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K Su. 2019. This looks like that: deep learning for interpretable image recognition. *Advances in neural information processing systems* 32 (2019).

[18] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2021. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–40.

[19] Gabriele Civitarese, Timo Sztyler, Daniele Riboni, Claudio Bettini, and Heiner Stuckenschmidt. 2019. POLARIS: Probabilistic and ontological activity recognition in smart-homes. *IEEE Transactions on Knowledge and Data Engineering* 33, 1 (2019), 209–223.

[20] Federico Cruciani, Anastasios Vafeiadis, Chris Nugent, Ian Cleland, Paul McCullagh, Konstantinos Votis, Dimitrios Giakoumis, Dimitrios Tzovaras, Liming Chen, and Raouf Hamzaoui. 2020. Feature learning for human activity recognition using convolutional neural networks: A case study for inertial measurement unit and audio data. *CCF Transactions on Pervasive Computing and Interaction* 2, 1 (2020), 18–32.

[21] Tirtharaj Dash, Sharad Chitlangia, Aditya Ahuja, and Ashwin Srinivasan. 2022. A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Scientific Reports* 12, 1 (2022), 1040.

[22] Tirtharaj Dash, Ashwin Srinivasan, and Lovekesh Vig. 2021. Incorporating symbolic domain knowledge into graph neural networks. *Machine Learning* 110, 7 (2021), 1609–1636.

[23] Natalia Díaz-Rodríguez, Alberto Lamas, Jules Sanchez, Gianni Franchi, Ivan Donadello, Siham Tabik, David Filliat, Policarpo Cruz, Rosana Montes, and Francisco Herrera. 2022. EXplainable Neural-Symbolic Learning (X-NeSyL) methodology to fuse deep learning representations with expert knowledge graphs: The MonuMAI cultural heritage use case. *Information Fusion* 79 (2022), 58–83.

[24] Marc Fischer, Mislav Balunovic, Dana Drachsler-Cohen, Timon Gehr, Ce Zhang, and Martin Vechev. 2019. DL2: training and querying neural networks with logic. In *International Conference on Machine Learning*. PMLR, 1931–1941.

[25] Manas Gaur, Keyur Faldu, and Amit Sheth. 2021. Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable? *IEEE Internet Computing* 25, 1 (2021), 51–59.

[26] KS Gayathri, KS Easwarakumar, and Susan Elias. 2017. Probabilistic ontology based activity recognition in smart homes using Markov Logic Network. *Knowledge-Based Systems* 121 (2017), 173–184.

[27] Eleonora Giunchiglia and Thomas Lukasiewicz. 2020. Coherent hierarchical multi-label classification networks. *Advances in neural information processing systems* 33 (2020), 9662–9673.

[28] Sojeong Ha, Jeong-Min Yun, and Seungjin Choi. 2015. Multi-modal convolutional neural networks for activity recognition. In *2015 IEEE International conference on systems, man, and cybernetics*. IEEE, 3017–3022.

[29] Massinissa Hamidi and Aomar Osmani. 2021. Human activity recognition: A dynamic inductive bias selection perspective. *Sensors* 21, 21 (2021), 7278.

[30] Harish Haresamudram, Irfan Essa, and Thomas Plötz. 2022. Assessing the state of self-supervised human activity recognition using wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–47.
[31] Karen Henricksen, Jadwiga Indulska, Ted McFadden, and Sasitharan Balasubramaniam. 2005. Middleware for distributed context-aware systems. In *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2005, Agia Napa, Cyprus, October 31-November 4, 2005, Proceedings, Part I*. Springer, 846–863.
[32] Shruthi K Hiremath, Yasutaka Nishimura, Sonia Chernova, and Thomas Plötz. 2022. Bootstrapping Human Activity Recognition Systems for Smart Homes from Scratch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.
[33] P Hitzler and M Sarker. 2022. Neuro-Symbolic AI= Neural+ Logical+ Probabilistic AI. *Neuro-Symbolic Artificial Intelligence: The State of the Art* 342 (2022), 173.
[34] Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. 2016. Harnessing deep neural networks with logic rules. *arXiv preprint arXiv:1603.06318* (2016).
[35] Yash Jain, Chi Ian Tang, Chulhong Min, Fahim Kawsar, and Akhil Mathur. 2022. ColloSSL: Collaborative self-supervised learning for human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–28.
[36] Md Kamruzzaman Sarker, Lu Zhou, Aaron Eberhart, and Pascal Hitzler. 2021. Neuro-Symbolic Artificial Intelligence: Current Trends. *arXiv e-prints* (2021), arXiv–2105.
[37] Xuhong Li, Haoyi Xiong, Xingjian Li, Xuanyu Wu, Xiao Zhang, Ji Liu, Jiang Bian, and Dejing Dou. 2022. Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems* 64, 12 (2022), 3197–3234.
[38] Yang Liu, Keze Wang, Guanbin Li, and Liang Lin. 2021. Semantics-aware adaptive knowledge distillation for sensor-to-vision action recognition. *IEEE Transactions on Image Processing* 30 (2021), 5573–5588.
[39] Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, Carsten Lutz, et al. 2009. OWL 2 web ontology language profiles. *W3C recommendation* 27, 61 (2009).
[40] Mathias Niepert, Jan Noessner, and Heiner Stuckenschmidt. 2011. Log-linear description logics. In *IJCAI*. 2153–2158.
[41] Jan Noessner and Mathias Niepert. 2011. ELOG: a probabilistic reasoner for OWL EL. In *International Conference on Web Reasoning and Rule Systems*. Springer, 281–286.
[42] Nobuyuki Oishi, Daniel Roggen, Philip Birch, and Paula Lago. 2023. Learning Using Privileged Information for Wearable-based Human Activity Recognition. In *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 233–234.
[43] Vitali Petsiuk, Abir Das, and Kate Saenko. 2018. Rise: Randomized input sampling for explanation of black-box models. *arXiv preprint arXiv:1806.07421* (2018).
[44] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.
[45] Daniele Riboni and Claudio Bettini. 2011. COSAR: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing* 15 (2011), 271–289.
[46] Charissa Ann Ronao and Sung-Bae Cho. 2016. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications* 59 (2016), 235–244.
[47] Saguna Saguna, Arkady Zaslavsky, and Dipanjan Chakraborty. 2013. Complex activity recognition using context-driven activity theory and activity signatures. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6 (2013), 1–34.
[48] Andrea Rosales Sanabria, Franco Zambonelli, and Juan Ye. 2021. Unsupervised domain adaptation in activity recognition: A GAN-based approach. *IEEE Access* 9 (2021), 19421–19438.
[49] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.
[50] Amit Sheth, Manas Gaur, Ugur Kursuncu, and Ruwan Wickramarachchi. 2019. Shades of knowledge-infused learning for enhancing deep learning. *IEEE Internet Computing* 23, 6 (2019), 54–63.
[51] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. 2007. Pellet: A practical owl-dl reasoner. *Journal of Web Semantics* 5, 2 (2007), 51–53.
[52] Elnaz Soleimani and Ehsan Nazerfard. 2021. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *Neurocomputing* 426 (2021), 26–34.
[53] Abdul Syafiq Abdull Sukor, Ammar Zakaria, Norasmadi Abdul Rahim, Latifah Munirah Kamarudin, Rossi Setchi, and Hiromitsu Nishizaki. 2019. A hybrid approach of knowledge-driven and data-driven reasoning for activity recognition in smart homes. *Journal of Intelligent & Fuzzy Systems* 36, 5 (2019), 4177–4188.
[54] Pratik Tarafdar and Indranil Bose. 2021. Recognition of human activities for wellness management using a smartphone and a smartwatch: a boosting approach. *Decision Support Systems* 140 (2021), 113426.
[55] Yonatan Vaizman, Katherine Ellis, and Gert Lanckriet. 2017. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE pervasive computing* 16, 4 (2017), 62–74.

[56] Vladimir Vapnik and Akshay Vashist. 2009. A new learning paradigm: Learning using privileged information. *Neural networks* 22, 5-6 (2009), 544–557.

[57] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2019. Deep learning for sensor-based activity recognition: A survey. *Pattern recognition letters* 119 (2019), 3–11.

[58] Anjana Wijekoon, Nirmalie Wiratunga, Sadiq Sani, Stewart Massie, and Kay Cooper. 2018. Improving kNN for human activity recognition with privileged learning using translation models. In *International Conference on Case-Based Reasoning*. Springer, 448–463.

[59] Di Wu, Si-Jia Zheng, Chang-An Yuan, and De-Shuang Huang. 2019. A deep model with combined losses for person re-identification. *Cognitive Systems Research* 54 (2019), 74–82.

[60] Tianwei Xing, Luis Garcia, Marc Roig Vilamala, Federico Cerutti, Lance Kaplan, Alun Preece, and Mani Srivastava. 2020. Neuroplex: learning to detect complex events in sensor networks through knowledge injection. In *Proceedings of the 18th conference on embedded networked sensor systems*. 489–502.

[61] Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, and Guy Broeck. 2018. A semantic loss function for deep learning with symbolic knowledge. In *International conference on machine learning*. PMLR, 5502–5511.

[62] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Twenty-fourth international joint conference on artificial intelligence*.

[63] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. 2017. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics* 28, 1 (2017), 162–169.