I. Experimental Setup

Table I presents the best hyperparameters obtained through grid search on the validation set across all datasets.

II. Baseline HAR Models

We compared CLAUDIA to a diverse set of state-of-the-art CA-HAR models as baselines, encompassing graph-based and non-graph-based, deep learning, and machine learning techniques. We now provide brief introductions.

Previously proposed HAR models: were selected to represent a comprehensive array of approaches for CA-HAR, facilitating a thorough evaluation of CLAUDIA's performance.

ExtraMLP [1]: A deep learning model for CA-HAR, ExtraMLP utilizes Multi-Layer Perceptron (MLP) to analyze fine-grained handcrafted features from smartphones.

LightGBM [2]: A widely-used Gradient-Boosted Decision Trees (GBDT) variant, sampling data with large gradients to facilitate fast computation and high performance.

CRUFT [3]: A cutting-edge framework employing a dual-branch approach, with one branch utilizing an MLP for handcrafted feature analysis and the other employing a CNN-BiLSTM branch for analyzing raw accelerometer and gyroscope data. CRUFT leverages temporal correlations between instances and incorporates an uncertainty estimation module to handle noisy CA-HAR data collected in real world.

GCN [4]: A classic graph GNN towards better label encoding by connecting user nodes to context/activity nodes as a bipartite graph, GCN has previously demonstrated its ability to predict the purpose of user visits from geographic information. However, it considers all nodes and edges homogeneous.

HGCN [5]: Graph-based model connecting nodes within each instance with a hyperedge and addressing hypergraph properties using Hypergraph Convolution (HGC) Layers.

HHGNN [6]: Further addressing node heterogeneity by projecting different types of nodes with separate linear projection functions. Although different hyperedges are connected to different types of nodes, HHGNN treats them indiscriminately and uses a shared HGC for all hypergraphs.

State-of-the-art models originally proposed for user identification/authentication based on their gait patterns: These models are highly relevant to CA-HAR and user identification.

GaitAuth [7]: utilizes a CNN [8] to extract features from raw signals, followed by a Bi-LSTM [9] model to predict whether the input data was gathered from a given subject.

GaitIden [7]: employs parallel CNN and LSTM modules to automatically extract features from raw data, followed by a fully connected layer to identify the subject.

III. Main HAR Results

We present the detailed results for *WASH scripted* and *Extrasensory* in Table II and Table III, respectively.

IV. Analysis

Drilling down, Fig. 1 presents the confusion matrices and additional detail on phone placement identification for each dataset. Specifically, the diagonal of each confusion matrix represents the recall per class. On the *WASH scripted* dataset, we achieve 92% accurate classification over for each phone placement. Interestingly, the *On Table-Face Down* and *On Table-Face Up* phone placement labels are never confused for one another. However, on the two unscripted datasets, CLAUDIA occasionally confuses *In Pocket* for *In Hand*, as human hands are often close to their pockets when at rest.

References

[1] Y. Vaizman, N. Weibel, and G. Lanckriet, "Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification," *ACM IMWUT*, vol. 1, no. 4, pp. 1–22, 2018.

[2] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Adv. NIPS*, vol. 30, 2017.

[3] W. Ge and E. Agu, "Cruft: Context rec. under uncertainty using fusion and temporal learning," in *Proc. ICMLA*. IEEE, 2020, pp. 747–52.

[4] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc ICLR 2017*. OpenReview.net, 2017. [Online]. Available: https://openreview.net/forum?id=SJU4ayYgl

[5] S. Bai, F. Zhang, and P. H. Torr, "Hypergraph convolution and hypergraph attention," *Pattern Recognition*, vol. 110, p. 107637, 2021.

[6] W. Ge, G. Mou, E. O. Agu, and K. Lee, "Heterogeneous hyper-graph neural networks for context-aware human activity recognition," in *PerCom*, IEEE. IEEE, 2023, pp. 350–354.

[7] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, and Q. Li, "Deep learning-based gait recognition using smartphones in the wild," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3197–3212, 2020.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," *arXiv preprint arXiv:1402.1128*, 2014.

TABLE I
HYPERPARAMETER SETTINGS.

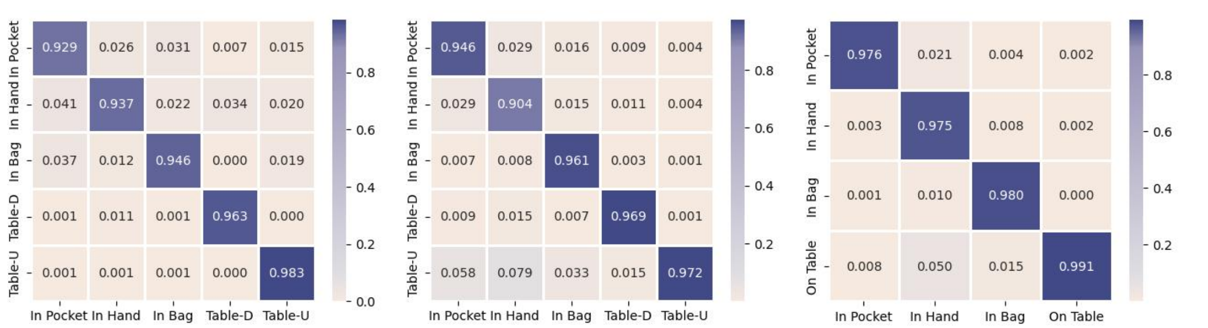
Hyperparameter	Search Space	*WASH scripted*	*WASH unscripted*	*Extrasensory*
lr	[1e-5, 3e-3)	5e-4	5e-5	1e-4
α	[1e-4, 10)	0.01	0.001	0.06
γ_1	[1e-4, 10)	1	1	0.01
γ_2	[1e-4, 10)	0.03	0.01	0.01
joint dim	[128, 8192]	256	6144	7168
lstm hidden	[8, 2048]	1536	128	16

TABLE II
MODEL PERFORMANCE ON THE *WASH scripted* DATASET. RESULTS ARE IN THE MCC/MACRO-F1 FORMAT (FOR BOTH METRICS, HIGHER NUMBERS INDICATE BETTER MODEL PERFORMANCE).

Category	UI/UA		non-GNN			GNN			Our Model	
Model	**GaitAuth**	**GaitIden**	**ExtraMLP**	**LightGBM**	**CRUFT**	**GCN**	**HGCN**	**HHGNN**	**CLAUDIA**	**Improv.(%)**
Lying Down	.211/.533	.203/.529	.183/.513	.185/.520	.202/.526	.220/.540	.212/.536	.221/.540	.231/.543	4.5/0.6
Sitting	.554/.747	.552/.739	.530/.718	.550/.739	.492/.690	.586/.762	.417/.645	.522/.716	.622/.784	6.1/2.9
Walking	.859/.929	.810/.902	.885/.942	.869/.934	.871/.935	.873/.936	.715/.852	.874/.937	.902/.951	1.9/1.0
Sleeping	.569/.745	.585/.756	.505/.701	.596/.763	.519/.712	.525/.715	.475/.683	.477/.684	.557/.736	-6.5/-3.5
Talking On Phone	.511/.710	.461/.677	.393/.632	.528/.723	.601/.768	.378/.625	.395/.638	.493/.698	.617/.779	2.7/1.4
Bathroom	.347/.612	.280/.561	.316/.576	.348/.598	.362/.614	.330/.590	.285/.570	.313/.577	.453/.673	25.1/9.6
Standing	.330/.587	.313/.581	.356/.598	.345/.597	.380/.622	.390/.634	.304/.591	.342/.591	.442/.662	13.3/4.4
Jogging	.655/.802	.645/.795	.632/.786	.826/.907	.671/.812	.728/.848	.536/.722	.653/.799	.825/.906	-0.1/-0.1
Running	.627/.786	.657/.808	.715/.840	.713/.841	.592/.761	.774/.877	.537/.727	.683/.821	.824/.905	6.5/3.2
Stairs-Going Down	.260/.544	.283/.570	.460/.680	.351/.608	.371/.618	.376/.621	.383/.630	.419/.652	.460/.684	0.0/0.6
Stairs-Going Up	.192/.525	.168/.515	.316/.589	.274/.567	.280/.568	.307/.585	.298/.582	.308/.589	.382/.639	20.9/8.5
Typing	.768/.872	.714/.839	.612/.772	.770/.873	.744/.859	.653/.800	.525/.716	.558/.737	.785/.882	1.9/1.0

TABLE III
MODEL PERFORMANCE ON THE *Extrasensory* DATASET.

Category	UI/UA		non-GNN			GNN			Our Model	
Model	**GaitAuth**	**GaitIden**	**ExtraMLP**	**LightGBM**	**CRUFT**	**GCN**	**HGCN**	**HHGNN**	**CLAUDIA**	**Improv.(%)**
Lying Down	.950/.975	.888/.944	.936/.968	.911/.955	.968/.984	.941/.971	.931/.965	.962/.981	.992/.996	2.5/1.2
Sitting	.877/.939	.753/.876	.797/.898	.755/.876	.897/.948	.781/.888	.801/.900	.870/.935	.964/.982	7.5/3.6
Walking	.667/.813	.528/.727	.554/.739	.560/.746	.709/.838	.580/.758	.578/.754	.718/.846	.898/.948	25.1/12.1
Sleeping	.961/.981	.918/.959	.950/.975	.934/.967	.977/.988	.952/.976	.962/.981	.979/.989	.995/.997	1.6/0.8
Talking	.764/.872	.632/.798	.681/.822	.697/.833	.834/.912	.740/.859	.614/.777	.858/.927	.951/.975	10.8/5.2
Bath-Shower	.664/.811	.483/.695	.496/.695	.729/.848	.654/.801	.648/.798	.659/.805	.780/.880	.895/.945	14.7/7.4
Toilet	.523/.720	.422/.662	.429/.652	.524/.716	.599/.767	.446/.666	.641/.796	.670/.813	.833/.911	24.3/12.1
Standing	.741/.860	.570/.759	.621/.787	.573/.760	.797/.891	.639/.802	.563/.748	.766/.875	.928/.963	16.4/8.1
Running	.736/.856	.627/.792	.616/.774	.656/.802	.733/.850	.724/.845	.693/.826	.866/.930	.936/.967	8.0/4.1
Stairs-Going Down	.517/.718	.412/.656	.441/.661	.791/.886	.618/.780	.462/.676	.633/.791	.695/.831	.835/.913	5.6/3.0
Stairs-Going Up	.561/.747	.407/.649	.446/.664	.654/.801	.651/.801	.513/.710	.614/.778	.706/.837	.863/.929	22.3/10.9
Exercising	.708/.838	.581/.760	.665/.808	.735/.852	.792/.887	.673/.812	.636/.788	.830/.909	.946/.973	14.0/7.0



(a) *WASH scripted* (b) *WASH unscripted* (c) *Extrasensory*

Fig. 1. Confusion matrix of three CA-HAR datasets. (Fig. 1a and 1b used the abbreviations of phone placement names, where 'Table-D' and 'Table-U' represent 'On Table-Face Down' and 'On Table-Face Up', respectively.)